Original Article

Self-exciting point process models of civilian deaths in Iraq

Erik Lewis[a], George Mohler[b], P. Jeffrey Brantingham[c,*] and Andrea L. Bertozzi[a]
[a]Department of Mathematics, University of California – Los Angeles, CA 90095-1555, USA.
E-mails: elewis@math.ucla.edu; bertozzi@math.ucla.edu
[b]Department of Mathematics and Computer Science, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053-0290, USA.
E-mail: gmohler@scu.edu
[c]Department of Anthropology, University of California, 341 Haines Hall, Box 951553, Los Angeles, CA 90095-1553, USA.
E-mail: branting@ucla.edu

*Corresponding author.

Abstract Our goal in this article is to characterize temporal patterns of violent civilian deaths in Iraq. These patterns are expected to evolve on time-scales ranging from years to minutes as a result of changes in the security environment on equally varied time-scales. To assess the importance of multiple time-scales in evolving security threats, we develop a self-exciting point process model similar to that used in earthquake analysis. Here the rate of violent events is partitioned into a background rate and a foreground self-exciting component. Background rates are assumed to change on relatively long time-scales. Foreground self-excitation, in which events trigger an increase in the rate of violence, is assumed to be short-lived. We explore the model using data from Iraq Body Count on civilian deaths between 2003 and 2007. Our results indicate that self-excitation makes up as much as 37–50 per cent of all violent events and that self-excitation lasts at most between two and six weeks, depending upon the district in question. Appropriate security responses may benefit from taking these different time-scales of violence into consideration.
Security Journal advance online publication, 19 September 2011; doi:10.1057/sj.2011.21

Keywords: modeling violence; point process; rational choice theory; routine activity theory; density estimation

Introduction

Security threats evolve on time-scales ranging from decades to minutes. At one extreme, populations, communities and the urban built environment experience turnover and reorganization on time-scales of years to decades (Dugan, 1999; Sampson, 2011). Individual attitudes and cultural norms display inertia over similar time-scales (Wuthnow, 2005; Akers, 2008). Security threats tracing their origins to long time-scales may emerge almost imperceptibly. For example, radical ideologies may arise out of slowly worsening material conditions over generations (Berman, 2009; Kilcullen, 2009). Gangs may become embedded within the social fabric of a community since gang members also play important non-gang roles in their neighborhoods (Sanchez-Jankowski, 1991; McCauley and Moskalenko, 2008). Violent and property crime may become an ambient feature of a local neighborhood, generating and reinforcing persistent structural characteristics such as high concentrated disadvantage and low collective efficacy (Dugan, 1999; Sampson *et al*, 2002; Hipp, 2010).

At the other extreme, motivated offenders move and mix within social and physical environments changing on time-scales of minutes to days (Cohen and Felson, 1979; Brantingham and Brantingham, 1995). Situational security threats tracing their origins to these short time-scales may emerge suddenly, with no prior warning. For example, heated arguments that escalate into violent attacks (Luckenbill, 1977; Felson and Steadman, 1983), car jackings of temporarily unoccupied vehicles (Miethe and Sousa, 2010), or theft of property in busy public spaces (Felson and Clarke, 1998) capitalize on fleeting, situational opportunities that disappear as quickly as they appear.

Between these extremes, habitual, routine activities and the organization of collective action within a local community may change on intermediate time-scales of days to months (Sampson and Wooldredge, 1987; Osgood *et al*, 1996). Security threats emerging on intermediated time-scales may evolve fast enough to be perceived as a growing problem, but not so fast as to illicit an immediate, concentrated response. For example, drug markets may grow slowly at first and then ramp up in size quickly as the number of transactions passes some critical threshold (Taniguchi *et al*, 2009). Similarly, criminal social networks may grow relatively slowly, constrained both by the geography of how co-offenders are distributed in space and the difficulty in finding people to fill certain functional roles (Malm *et al*, 2008).

Deciding which of these various time-scales is most important is extremely difficult. In some contexts, it is reasonable to argue that certain security threats will not disappear if the long-term 'root causes' are not dealt with. Drug violence, for example, may be unlikely to abate unless the problems of demand are addressed (Chabat, 2002). Similarly, active insurgencies may be expected to continue unless the long-standing grievances of population are rectified (Kilcullen, 2009). However, the strategic options for dealing with security threats emerging on such long time-scales are likely to themselves require planning and investment that last years to decades. The outcomes of such strategies may be particularly uncertain because of the time-scales involved.

Conversely, in other contexts, short-term situational factors may be most relevant to dealing with immediate security threats. Address the physical or social environmental characteristics that generate crime opportunities and you may eliminate crime. The risk that property is stolen from a public space, for example, may be mitigated by manipulating levels of actual or perceived surveillance and ownership (Felson and Clarke, 1998; Keizer *et al*, 2008; Wortley and McFarlane, 2011). Similarly, regulate the timing, locations and quantities of legal public alcohol consumption and one may greatly reduce the likelihood of non-domestic assaults (Livingston, 2008). Security threats arising at these short time-scales may be tractable following situational crime prevention or problem-oriented policing strategies. However, such solutions are by definition local (situational) and may be vulnerable to simple innovations by offenders or other hostile actors (Cornish and Clarke, 1987). Indeed, Brantingham *et al* (2005) have suggested that situational crime prevention measures may have a 'half-life' the duration of which is dependent upon how quickly offenders can adapt.

While there are good theoretical and practical reasons to treat long, intermediate and short time-scale processes as conceptually separate, it is clear in reality that they all operate simultaneously. At a theoretical level, there are at least two possible ways to conceive of overlapping time-scales in relation to actual crime or other hostile events. First, consider a cluster of crimes in a particular spatial region occurring during a defined period of time. We might imagine that some fraction of the crimes in the cluster are exclusively the result of long-term processes, while others are exclusively the result of intermediate or short-term processes (see Short *et al*, 2009). Crimes may occur near one another in space and time despite having their origin in processes operating at different time-scales. For example, it is well-known that burglars differ greatly in their baseline offending rates; some burglars commit at most one or two crimes per year, while others commit 50 or more per year (Wright and Decker, 1994; Canela-Cacho *et al*, 1997). For a system with only two active burglars, only ~4 per cent (2/52) of the crimes in a given year would be attributable to the occasional offender, while ~96 per cent (50/52) would be linked to the prolific offender. The two crimes attributable to the occasional offender will be embedded in the spatio-temporal pattern of crimes committed by the prolific offender. Second, consider the alternative focusing only on a single crime. We might attribute some fraction of the cause of that one crime to a long-term process, while the remainder might be attributed to intermediate or short-term processes. The time-scales involved might be highly skewed towards long- or short-term processes, or they might be balanced among them. A given insurgent attack, for example, might be viewed simultaneously as the product of long-standing grievances between sectarian groups, social network structures that supply insurgents with bomb making *materiel*, and situational factors such as the spontaneous gathering of potential victims. At present, the relative contributions of these different time scales to such attacks are entirely speculative.

Regardless of the conceptualization one adopts, the appropriate security response may benefit from taking into account processes operating on multiple time-scales. It is not immediately clear how to do this, however. Here we propose an approach to modeling the dynamics of crime or hostile events on multiple time-scales based on self-exciting point processes commonly used in the study of earthquakes (Ogata, 1988; Zhuang *et al*, 2002; Mohler *et al*, 2011). They are readily extended to deal with spatial dynamics as needed (Vere-Jones, 2009). The central idea in self-exciting point process models is that any cluster of discrete events can be divided into those that are background and those that are foreground events, akin to the first conceptualization of time-scales above. Background events are typically attributed to processes that either do not change (that is, they are stationary in time), or change only over relatively long time-scales, though they may vary significantly through space. For example, it is often assumed that spontaneous earthquakes occur at a stationary rate determined by time-invariant structures in the fault system. Background earthquakes are by definition statistically independent of one another in space and time. Thus one background event neither increases nor decreases the chance of another background event occurring. The dynamics of background events are thus well-specified by a Poisson process (Ogata, 1988; Vere-Jones, 2009). By contrast, foreground events are attributed to processes that show significant local temporal (and spatial) dependencies. For example, major earthquakes generate aftershocks during a narrow window of time and nearby to the location of the initiating event. Foreground events are not only statistically dependent on prior background events, but in fact may be thought of as a byproduct of prior background events. The dynamics of foreground events are thus no longer described by a simple Poisson process, but rather require models that capture the statistical dependence among some events. Following the terminology established in seismology, the resulting process is considered 'self-exciting' and the associated models are called self-exciting point process models (see also Holden, 1987).

Similar ideas have been proposed in relation to crime event patterning (Brantingham and Brantingham, 1984; Townsley *et al*, 2003; Johnson *et al*, 2007; Johnson, 2008). That background or ambient crime rates may be relatively stationary in time, but heterogeneous in space is well-documented in environmental criminology (see Andresen *et al*, 2009). Such differences may represent persistent structural biases in social and economic access (Morenoff *et al*, 2001), or stable variability in the distribution of crime opportunities independent of social inequality (Farrell *et al*, 2010). Foreground interdependencies among crimes, by contrast, have been recognized only relatively recently (Hough *et al*, 1983; Trickett *et al*, 1992). Interdependencies among crimes may be understood in the context of rational choice and routine activity theory of crime (Cohen and Felson, 1979; Cornish and Clarke, 1986). In some instances, individuals committing an initial property or violent crime may return within days to the same or a nearby place to replicate the successes of the previous event. In other instances, an act of violence by any individual or group may incite reprisals and counter reprisals, leading to a self-excited cycle of violence. In one case, it has been found that self-exciting foreground processes account for ~13 per cent of burglaries, leaving ~87 per cent linked to stationary background processes (Mohler *et al*, 2011). Here we pursue a similar analysis of violent events in an extreme security setting. Specifically, we ask: (1) whether temporal patterns of violent events in extreme security settings can be partitioned into background and foreground components; and (2) what models best describe long-term background and short-term foreground processes. We evaluate these questions using data on violent civilian deaths at the height of the Iraqi conflict between 2003 and 2007. Townsley *et al* (2008) established that short-term event dependencies lasting up to two weeks characterized insurgent IED attacks in Iraq over a similar time frame. Using data from Iraq Body Count (2008), we evaluate the temporal dynamics of violent civilian deaths from a wider array of causes including coalition and insurgent actions, garden-variety and organized crime, sectarian violence and so-called blood feuds. We expect foreground events driven by self-exciting processes to be far more important than background events because many of the normal institutional and sociocultural controls on behavior are severely impaired or completely absent. In such settings hostile actors may exploit vulnerable targets or seek retribution with few constraints.

This article is organized as follows. In the first section, we introduce point process models as well as three modifications to standard self-exciting point processes to account for variations in the background rate. In the next section, we introduce data on civilian deaths in Iraq and the assumptions necessary to proceed with analysis. In the subsequent section, we analyze four different regions of Iraq comparing the effectiveness of each model in capturing the process of violence. In the final section, we discuss the implications of our findings and possible directions for future work.

Statistical Point-Processes

A point process is a stochastic model commonly used to describe the occurrence of discrete events in time and/or space (Schoenberg *et al*, 2006). A convenient way to view a point process is in terms of a list of times $t_1, t_2, \ldots, t_n$ at which hostile events $1, 2, \ldots, n$ occur. In the case of an *orderly* point process, events cannot occur simultaneously and, thus, multiple events can be arranged into an unambiguous sequence. For example, if two insurgent attacks arise from an orderly point process then we should be able to say not only which event occurs before the other, but also how much time separates the two events.

A point process is characterized by its conditional intensity $\lambda(t)$, which represents the mean instantaneous rate at which events are expected to occur given the history of the process up to time t (Ogata, 1988). For completeness we write

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

where $\mathcal{H}_t$ represents the history of events prior to time t, and the expectation is in terms of the number of events $N[t, t+\Delta t)$ occurring between time t (inclusive) and $t+\Delta t$ (exclusive).

An important example of a point process is a Poisson process, which we introduce here as a useful benchmark for evaluating self-excitation since it represents complete randomness. In general, a point process is classified as a Poisson process if events occurring at two different times are statistically independent of one another, meaning that an event at time t_1 neither increases nor decreases the probability of an event occurring at any subsequent time. Independence also characterizes aggregates of events if the process is Poisson. For example, if $[t_1, t_2)$ is an interval of time between t_1 (inclusive) and t_2 (exclusive) and $N[t_1, t_2)$ represents the number of events occurring between time t_1 and t_2, then N is a Poisson process if the distribution of number of events in the interval follows a Poisson probability density function, and the events in $N[t_1, t_2)$ are statistically independent of events occurring in any subsequent interval $N[t_3, t_4)$. In other words, the collection of events occurring between times t_1 and t_2 neither increase nor decrease the likelihood of a collection of events occurring between times t_3 and t_4. Finally, note that the conditional intensity $\lambda(t)$ of a Poisson process is deterministic, meaning that events are linked causally to the conditional intensity and only this conditional intensity. Most commonly, Poisson processes are treated as stationary (that is, non-changing in time), which means that $\lambda(t)$ is a positive constant real number.

A self-exciting point process stands in stark contrast to a Poisson process. A point process N is said to be self-exciting if any one event at time t_1 increases the likelihood of an event occurring at time t_2, or if a collection of events in $N[t_1, t_2)$ increase the likelihood of another collection of events occurring during a subsequent interval $N[t_3, t_4)$. In statistical terms, a self-exciting point process exhibits covariance between collections of events in time

$$\mathrm{Cov}\left[N[t_1, t_2), N[t_2, t_3)\right] > 0$$

The covariance between collections of events for a Poisson process is $\mathrm{Cov}[N[t_1, t_2), N[t_2, t_3)] = 0$. Note that a self-inhibiting process, where a collection of events in one time interval *decreases* the likelihood of events in a subsequent time interval has $\mathrm{Cov}[N[t_1, t_2), N[t_2, t_3)] < 0$ (see Holden, 1987). For more on point processes, see Daley and Vere-Jones (2003, 2008).

The Hawkes Process Model

We investigate a specific class of point processes termed a Hawkes Process (Hawkes, 1971), which simultaneously captures the idea that any given event, or collection events can be causally linked to a background Poisson process and foreground self-exciting process. The conditional intensity for a Hawkes process is (Hawkes and Oakes, 1974)

$$\lambda(t) = \mu + k_0 \sum_{t_k < t} g(t - t_k) \tag{1}$$

In equation (1), μ represents the background rate of events, which in most applications is assumed to be constant in time (Zhuang *et al*, 2002). The parameter μ by itself represents a stationary Poisson process. The second half of the sum describes self-excitation and has components k_0 and g. The parameter k_0 reflects how much excitation is generated by a collection of prior events, while g represents the density of prior events necessary to trigger excitation. In mechanistic terms, if self-excitation is triggered, then k_0 is the amount by which the conditional intensity increases above its background rate. The expected number of events in any time interval will increase above the background rate if $k_0 > 0$. However, when exactly these events are added is dependent upon the specific functional form of the triggering density g (Hawkes, 1971; Ogata, 1988). We use an exponential distribution for the triggering density g, similar to that used by Egesdal *et al* (2010), giving the conditional intensity

$$\lambda(t) = \mu + k_0 \sum_{t_k < t} w e^{-w(t - t_k)} \tag{2}$$

Here, w defines a rate of decay for the triggering density, controlling how quickly the overall rate $\lambda(t)$ returns to its baseline level μ. In a sense, w defines how long self-excitation lasts following an event. If w is large, self-excitation will last only a short while and additional events will only be added above background shortly after the initial event. Conversely, if w is small, then self-excitation will last for a much longer period of time and events added above background over a protracted period. Such time-limited self-excitation is well documented in the criminological literature with typical durations lasting between two and six weeks (Johnson, 2008; Short *et al*, 2009).

In behavioral terms, k_0 might be thought of as the strength of the incentive to replicate a past success. For example, if an insurgent considered an attack on day t_1 as particularly successful they might want to add two more attacks to their plans. Alternatively, k_0 might represent the strength of the drive to seek retribution for a previous attack. For example, if an attack on day t_1 was seen as particularly extreme, then the victims might decide to add two reprisals to their normal response to sectarian conflict. Similarly, the inverse of w (that is, w^{-1}) may be thought of as the average time-window over which insurgents plan to add events. For example, two events may be added over the next two days (that is, $w = 1/2 = 0.5$), or the next 14 days (that is, $w = 1/14 = 0.0714$). We may think of the former as a concentrated, and the latter as a persistent self-exciting point process.

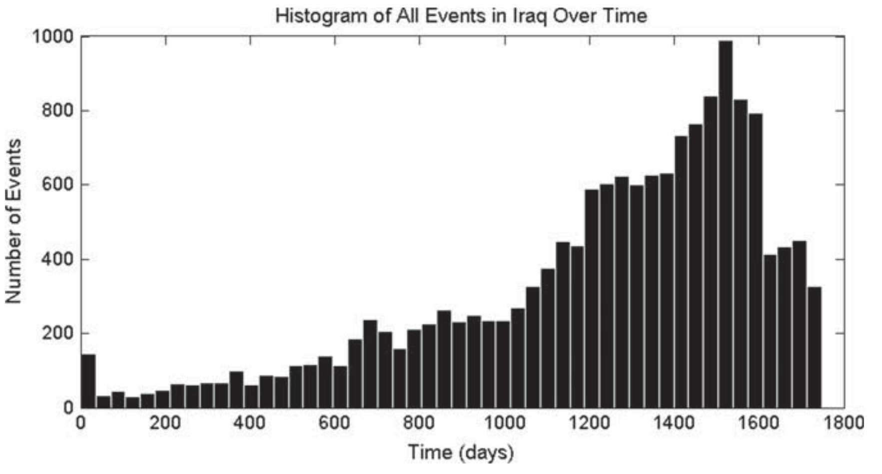

Figure 1: A histogram of all events occurring in Iraq with time on the horizontal axis and number of events on the vertical axis. Over 1747 days, between 20 March 2003 and 31 December 2007, there are a total of 15 977 events in 50 bins.

Inspection of our data on violent death in Iraq indicates that a stationary Poisson background rate μ is unrealistic (Figure 1). An attempt to fit the model in equation (2) would require that the upward trend in the data be driven exclusively by self-excitation, which on behavioral grounds seems unjustified. As a consequence, we consider several alternative models for non-stationary background rate μ, retaining in the exponential form of the triggering density g used by Egesdal *et al* (2010). In other words, we model μ as a background rate that changes as a result of long time-scale processes. The triggering density remains a short time-scale, foreground phenomenon superimposed on the long time-scale process.

Given the observed trend in the Iraqi data, the simplest choice for a non-stationary μ is a step function that allows for the background rate to jump to different stationary levels at different points in time. We choose a step function parameterized by three values μ_1, μ_2 and μ_3, representing three different background rates over separate intervals of time. Our baseline model becomes:

$$\lambda(t) = \mu_{step}(t) + k_0 \sum_{t_k < t} w e^{-w(t-t_k)} \tag{3}$$

where

$$\mu_{step}(t) = \begin{cases} \mu_1 & \text{for } 0 \leqslant t < t_1 \\ \mu_2 & \text{for } t_1 \leqslant t < t_2 \\ \mu_3 & \text{for } t_2 \leqslant t < T \end{cases}$$

We choose the two points at which the background rate jumps, t_1 and t_2, based on visual inspection of where the largest jumps in activity occur. Values of t_1 and t_2 are then held constant while fitting the other model parameters. Appropriate jump times in the background rate could be determined via model fitting along with other model parameters. However, as discussed below, this greatly adds to model complexity. Moreover, our interest is not specifically in whether different step functions with different inflection points outperform one another, but rather whether a step function with self-excitation offers a better explanation than the same step function without self-excitation.

We consider a second model with a stationary background rate up until time t_c at which point there is a linear increase in the background rate. The exact onset of this change depends on which city we are considering, but it typically begins between 400 and 1000 days into the data. In each case, we choose t_c based on visual inspection of the data, with the same justifications as offered above. The second model becomes:

$$\lambda(t) = \mu_l(t) + k_0 \sum_{t_k < t} w e^{-w(t-t_k)} \tag{4}$$

where

$$\mu_l(t) = \begin{cases} \mu_c & \text{for } 0 \leqslant t < t_c \\ \mu_s(t - t_c) + \mu_c & \text{for } t_c \leqslant t < T \end{cases}$$

Finally, we propose a third model based on non-parametric estimation of μ. We use variable bandwidth kernel smoothing to construct a smoothed version of the data (Silverman, 1998; Zhuang *et al*, 2002; Schoenberg, 2003). The method works by summing Gaussian normal distributions for each point in the data. The mean of each Gaussian curve is the recorded day of the event. The standard deviation of the distribution is chosen depending on the local density of events. Thus, in regions of the time series where there are many events, the standard deviation of the local Gaussian curves will be small. In regions where there are fewer events, the standard deviation will be larger. Allowing the standard deviation to vary depending on density avoids potential problems of over smoothing in regions with many events, and under-smoothing in regions with few events. It has been shown that variable bandwidth kernel smoothing is more robust where events do not approximate a uniform distribution in time (Silverman, 1998).

Since we need the rate function $\lambda(t)$ to integrate to the total number of events, we introduce a third parameter p to allow for this. Our final model becomes:

$$\lambda(t) = pn\mu_{sm}(t) + (1-p)k_0 \sum_{t_k < t} w e^{-w(t-t_k)} \tag{5}$$

where μ_{sm} is the smoothed background rate, n is the total number of events in the data set, p is the proportion of the process attributed to the smoothed background component, and $(1-p)$ is the proportion attributed to self-excitation. Note that the smoothed background rate unavoidably includes both true background events as well as any foreground events. In principle, it would be ideal to have some *a priori* information that would allow us to separate out background events and use only these for data smoothing. In practice, this is not feasible because there is always inherent uncertainty about which events are truly background events. We simply assume therefore that the smoothed data are a close proxy of the true background rate.

To estimate parameters, we use maximum likelihood estimation (MLE) (Fisher, 1922; Rubin, 1972; Ozaki, 1979). MLE seeks to determine, for a specified probabilistic model, the set of model parameters that are most *likely* to have given rise to a set of known observations. MLE is conceptually the inverse of probabilistic modeling, which seeks to determine the most *probable* observations given a specified model and defined parameter values. For example, probabilistic modeling would estimate the most *probable* number of heads to observe in 10 flips of an unfair coin, given the chance a head in one flip is $q = 0.7$. By contrast, MLE would seek the most *likely* value for q given that seven heads were observed in 10 flips. The likelihood that it is a fair coin $q = 0.5$ is $L = 0.117$, but the *maximum* likelihood $L = 0.2678$ that it is an unfair coin with $q = 0.7$. For simple models with a small number of parameters, it is often possible to calculate the maximum likelihood explicitly (Myung, 2003). For more complex models, one must calculate maximum likelihood values numerically. In practice, this means that a set of test parameter values are chosen and the log likelihood function is calculated. Small adjustments are made to the parameter values and the log likelihood function is recalculated. The process is repeated until the function maximum is found. For example, if we want to estimate parameters for the linear model in equation (4), we iterate over all possible values of μ_c, μ_s, k_0 and w, subject to the constraint that all four parameters are positive, until we find highest point or peak of the log likelihood function. For completeness, the log likelihood function may be written as

$$\log L(\mu_c, \mu_s, k_0, w \mid t_1, .., t_n) = \sum_{t_i : 1 \leqslant i \leqslant n} \log(\lambda(t_i)) - \int_0^T \lambda(t)\,dt.$$

MLE proves to be very efficient at finding accurate parameter values even for highly non-linear functions (Myung, 2003).

To compare different models, we use Akaike's Information Criterion (AIC) (Akaike, 1973, 1974). For a given model, the AIC is equal to $2\kappa - 2\log(L)$ where κ is the number of free parameters in the model and L is the maximum value of the likelihood function. AIC rewards model parameterizations that are more likely, in the sense of maximum likelihood discussed above. However, AIC penalizes models with more parameters, under the assumption that simpler models are preferable to more complex models. Overall, smaller AIC values imply a better models (Bozdogan, 1987). We note that the AIC is a relative scale used to compare different models, and is not a test for absolute goodness of fit.

In Egesdal *et al* (2010), a self-exciting model for gang violence is compared to a stationary Poisson process with the background rate equal to the average number of events over the time interval in consideration. This is a reasonable starting assumption for gang retaliations given that many gang rivalries are relatively stable over time-scales of years, though individual membership may be highly fluid (Klein and Maxson, 2006). However, due to the secular trend apparent in the Iraq data, a stationary Poisson process is a very unlikely model for the background rate of violence. The appropriate comparison, therefore, is between the AIC values for each self-exciting model and those for the equivalent non-stationary Poisson model with self-excitation removed. In other words, we seek to evaluate the relative importance of long time-scale background processes with the short time-scale processes included against long time-scale background processes absent the short time-scale foreground processes. To implement this comparison we use MLE to estimate parameters for each full model with self-excitation, calculating the resulting AIC value. We then remove self-excitation by setting $k_0=w=0$ and then re-estimate parameters for the background rate using MLE. This procedure puts the models on more even footing with respect to most likely parameterizations.

Data

Since Operation Iraqi Freedom began on 20 March 2003, nearly 100 000 Iraqi civilians have died of violent causes (Hicks *et al*, 2011). These deaths have occurred as a result of garden-variety and organized crime, sectarian civil strife, local insurgent actions, and transnational terrorist attacks. Coalition forces and Iraqi security forces have also contributed to the total. We evaluate the dynamics of background and foreground processes underlying these deaths using data from Iraq Body Count, an organization seeking to record all civilian deaths in Iraq (Iraq Body Count, 2008). The number of fatalities linked to any event is not an estimate by the organization, but a count corroborated by at least two recognized news sources. In the data we consider, from 20 March 2003 to 31 December 2007, there are 15 977 unique violent events, corresponding to as many as 89 766 individual deaths. Each entry in the data contains a start date, end date, minimum number of deaths, maximum number of deaths, town and possibly a district of where the event occurs.

Several simplifying assumptions are necessary to make temporal analysis of the IBC data straightforward. First, we take events, not number of deaths, as the unit of analysis. For example, on 28 December 2007, 14 people were killed by a car bomb in Al-Tayaran Square in Baghdad. We consider this one event, not 14.

Second, we consider only the date of events, ignoring the type of event and any information about suspected perpetrators. For example, we do not distinguish between IED attacks and gunfire, nor do we ignore events in the data that are unrelated to IED attacks. This means that no events are discarded because they are seen as unrelated to our analysis.

Third, we only consider the start date and not the end date for events. Data entries for which the time of occurrence is a range of days reflect uncertainty as to when the individuals actually died. For example, if a mass grave is discovered, there is uncertainty as to how and when the deaths occurred. The effect from this simplification depends on the region in consideration. Overall, 93.45 per cent of the events have the same start and end date.

Fourth, we group data by location according to the smallest known geographic region available. For some cities, like Baghdad, we have events down to each local district, but this is not true of most other cities in Iraq.

Finally, we assume that events recorded on the same day are statistically independent. Since we only have events recorded to the day, we cannot determine whether events occurring on the same day are correlated with one another. Thus, when we compute the sum for $\lambda(t)$ in equation (2) we only consider events occurring on dates strictly before the day on which a focal event occurs. Technically this means the process is no longer a simple point process. However, we reason that we can analyze the data without allowing events occurring on the same day to effect one another. To evaluate the nature of this final assumption, we provide a histogram of the frequency of days with a given number of events

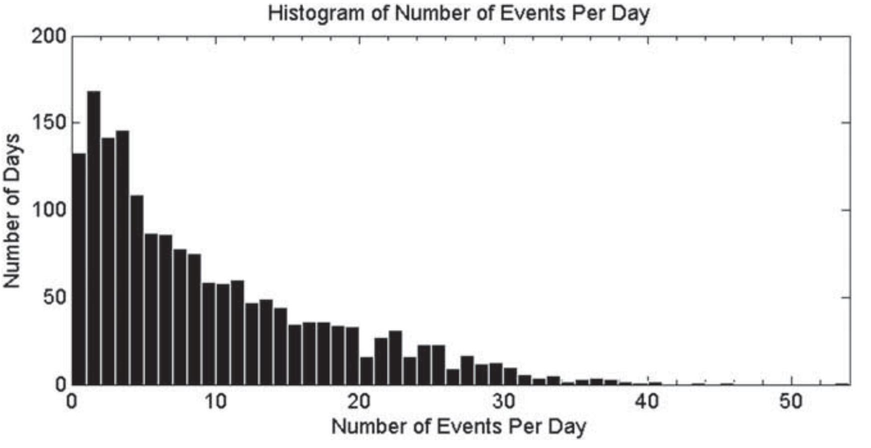

Figure 2: A histogram of the number of events per day over all of Iraq between 20 March 2003 and 31 December 2007. The most events occurring on a single day is 53. The mean number of events on a single day is 9.15.

(Figure 2). Out of a possible 1748 days, there are 133 days with no events, while there are 169 days with one event. This means there are 1446 days with more than one event. Few of these events presumably happened at exactly the same moment in time, which means that they are potentially self-exciting. Small attacks that are meant to draw in first responders as victims for a larger, follow-on attack, for example, may represent an extremely fast self-exciting process present in the data. However, we are unable to detect this or other processes operating at time-scales shorter than one day.

Analysis and Results

We examine temporal patterns of violent deaths in four different Iraqi regions including Karkh, Najaf, Mosul and Fallujah. There is considerable uncertainty about the demographic and sectarian characteristics of each of these regions during the course of the conflict. If Baghdad is a useful guide, there appears to have been a comprehensive internal reorganization of neighborhoods along sectarian lines between 2003 and 2009 (Izady, 2011). Our choices of areas are therefore based on gross differences in the data alone. Karkh is a district at the heart of Baghdad and is chosen because it contains the largest number of events of any spatial region in Iraq. Najaf is a mid-sized city in central Iraq. It displays obvious temporal clustering, but far fewer events than other region we analyze. Mosul, in the North, and Fallujah, in the west, have intermediate numbers of events.

In each case, we must define *a priori* several parameters necessary to implement point process models with non-stationary background rates. These parameters include: (1) the times t_1 and t_2 at which the step function changes; (2) the time t_c at which a linear increasing background rate starts; and (3) for kernel smoothing, a minimum bandwidth $b_{\min}$ and nearest neighbor distance (Silverman, 1998). Values for these parameters are given in Table 1.

Table 1: Fixed model parameters determined by visual inspection

City	t_1[a]	t_2[a]	t_c[b]	*kth neighbor*[c]	b_{min}[d]
Karkh	661	1385	400	200	80
Najaf	661	1385	1050	50	30
Mosul	1050	1350	975	200	150
Fallujah	575	1025	400	80	30

[a]Data series days at which step function jumps.

[b]Data series day at which piecewise linear function starts to increase.

[c]The *k*th nearest neighbor used in kernel smoothing.

[d]Minimum bandwidth used in kernel smoothing.

Karkh

The first region we consider is Karkh, a district of Baghdad. Over the entire data set there are 2278 events spanning 1742 days. Table 2 shows that AIC values for models with self-excitation are universally smaller than those for models without self-excitation. For example, the AIC value for the model in equation (4) with a linear increase in the background rate beginning at t_c is 905.7, while for the same model with no self-excitation the AIC is 1624.7. Adding self-excitation results in a more complicated model, but this added complexity is offset by better model performance. The step function and the linear models with self-excitation perform almost identically. Both adequately capture the abrupt jump in the number of events around 1385 days into the data. However, the smoothed background rate models, with and without self-excitation, dramatically outperform all of the alternatives with AIC values of 829.6 and 855.0, respectively. It is perhaps not surprising that a smoothed background rate model captures the complex trajectory of violent death in Karkh better than simpler models. But it is also clear that a foreground self-exciting process, superimposed on this complex background process, offers an even better model of the associated dynamics in spite of the greater model complexity. This superior performance is visually confirmed in Figure 3.

Estimating the parameters for the smoothed background rate model with self-excitation, we get $(1-\hat{p})\hat{k}_0=0.364$, $\hat{w}=0.064$ and $\hat{p}=0.6368$ (Table 2). These parameter values have straightforward behavioral interpretations. First, like $\hat{k}_0$ in the non-smoothed models, $(1-\hat{p})\hat{k}_0$ in the smoothed model signifies the average number of direct offspring events that are caused by a preceding event. The three non-smoothed models suggest that every 10 events cause on average another nine offspring events. The smoothed background rate model with self-excitation implies on average about four offspring for every 10 preceding events. The lower magnitude of self-excitation in the smoothed background rate model is understandable given that a more complex background rate trajectory is allowed. Conversely, the simpler models must attribute more of the overall trend to short time-scale foreground processes (Figure 3).

Second, we can interpret w^{-1} as an average time over which we expect an excited event to happen following an initiating event. The three poorer performing models give $\hat{w}^{-1}=19.01$ days. In other words, each initiating event is expected to produce one offspring on average within 19 days. The smoothed background rate model suggests a slightly shorter time window with $\hat{w}^{-1}=15.08$ days. Three initiating events occurring on the same day may be

Table 2: Parameter estimates for every event in Karkh and comparisons of models with and without self-excitation

City	$\hat{\mu}$	$\hat{k}_0$	$\hat{w}^{-1}$	*Hawkes AIC*	*NoSE AIC*[a]	*Best fit*
Karkh (μ)	0.060	0.959	19.23	906.0	3330.6	Hawkes
Karkh (μ_{step})	0.053, 0.143, 0.053[b]	0.935	19.23	905.6	3331.1	Hawkes
Karkh (μ_l)	0.0552, 0.0001[c]	0.915	18.59	905.7	1624.7	Hawkes
Karkh (μ_{sm})	0.6368[d]	0.364[e]	15.08	829.6	855.0	Hawkes

[a]Models without self-excitation, implemented by setting k_0 and w to zero and re-estimating parameters for the background rate.

[b]Step function parameter estimates $\hat{\mu}_1$ for $0 \leq t < t_1$, $\hat{\mu}_2$ for $t_1 \leq t < t_2$ and $\hat{\mu}_3$ for $t_2 \leq t < T$, respectively.

[c]Piecewise linear $\hat{\mu}_c$ for $0 \leq t < t_c$ and $\hat{\mu}_s(t - t_c)$ for $t_c \leq t < T$.

[d]Kernel smoothing parameter estimate $\hat{p}$.

[e]Reported value is for $(1 - p)\, k_0$, which is comparable to k_0 in other models.

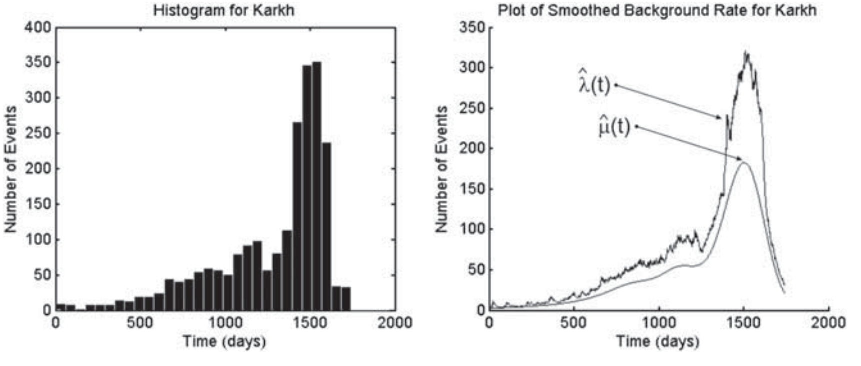

Figure 3: A histogram of all events in Karkh (left). The estimated fit of the data for the smoothed background rate model $\hat{\lambda}(t)$ (right). The smoothed background rate $\hat{p}\hat{\mu}_{sm}(t)$ is plotted on the right as well for reference. To plot both of these on the same scale as the histogram, we multiply the estimates by T and then divide by the number of bins (30).

expected to generate one offspring within two weeks. In either case, the briefness of the time window for the production of self-excited offspring events could correspond to the amount of time a hostile actor needs to prepare for another attack.

Finally, since $\int_0^T \hat{\lambda}(t)\mathrm{d}t$ is equal to the number of events occurring in the time interval $[0, T]$, we have that $\int_0^T \hat{\mu}(t)\mathrm{d}t$ is an estimate for the total number of background events in our data set and $\int_0^T \hat{\lambda}(t) - \hat{\mu}(t)\mathrm{d}t$ is the number of events generated by foreground self-excitation. For the first row of Table 2, where the background rate is stationary (that is, $\hat{\mu}(t) = 0.060$), the number of background events is estimated at 137, which represents only 6 per cent of the total number of events. Self-excited events would therefore have to make up 94 per cent of the sample, or 2141 events. In addition to the large AIC value, this result illustrates the poor fit of a stationary (simple Poisson) background rate model. For the smoothed background rate model, the estimated proportion of background events is simply $\hat{p} = 0.6368$ and therefore the number of events is $n\hat{p} = 1451$ events (63.68 per cent), where $n = 2278$ is the total number of events in Karkh. Thus $n(1 - \hat{p}) = 827$ (36.32 per cent) events are inferred to be the product of self-excitation.

We note the consistency of the estimates for k_0 and w^{-1} across the first three models. The estimates stand in contrast to those for the smoothed background rate model, which has the lowest AIC. This consistency is also not the case for other areas in Iraq. We expect that large sample size for Karkh is at least partially responsible for the consistency across models.

Najaf

The second region we consider is Najaf, a medium-sized city 180 miles south of Baghdad. A total of 149 events were recorded in Najaf spanning 1718 days (Figure 4). AIC values for each model, with and without self-excitation, occupy a much narrower range compared with Karkh (Table 3). We attribute this clustering of AIC values to the small sample size. The AIC value of 949.8 for the linear model with self-excitation is marginally better than the other models. This stands in contrast to the results from the other districts analyzed where the smoothed background rate typically outperforms all other models. However, this is not necessarily surprising given the sparseness of the data. As a result, long time-scale, secular trends are not immediately apparent and one might expect models with stationary, or near-stationary background rates to do reasonably well. In fact, we note that all of the various models are closer to one another in AIC values than in any other region (Table 3).

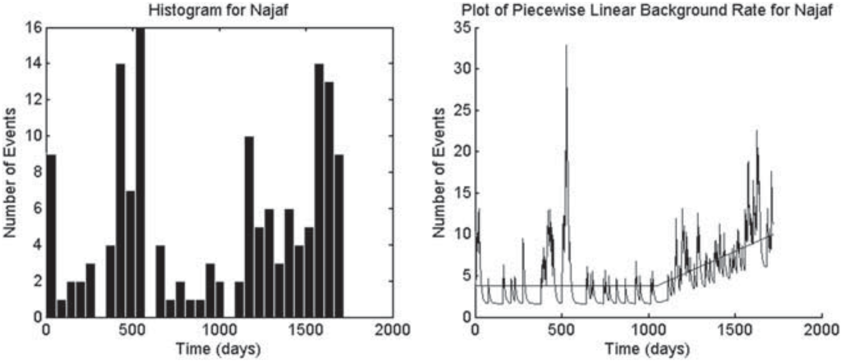

Figure 4: A histogram of all events in Najaf (left). The estimated fit for $\hat{\lambda}(t)$ with a linear background rate is plotted on the right (the jagged curve). The fit for the data without self excitation is plotted on the right as well (the straight line).

Table 3: Parameter estimates for every event in Najaf and comparisons of models with and without self-excitation[a]

City	$\hat{\mu}$	$\hat{k}_0$	$\hat{w}^{-1}$	*Hawkes AIC*	*NoSE AIC*	*Best fit*
Najaf (μ)	0.037	0.590	9.709	963.7	1028.4	Hawkes
Najaf (μ_{step})	0.032, 0.034, 0.078	0.521	8.772	958.2	1007.5	Hawkes
Najaf (μ_l)	0.0286, 0.0001	0.497	8.354	949.8	1004.8	Hawkes
Najaf (μ_{sm})	0.5054	0.501	8.606	959.2	1004.9	Hawkes

[a]Parameter details as in Table 2.

Following the same intuition developed in the previous example, we see for the linear model that $\hat{k}_0 = 0.497$. Thus every 10 events produce on average five offspring events (Table 3). However, we expect to wait only approximately $w^{-1} = 8.35$ days for an offspring event to occur. The timescale for self-excitation is about half as long as the timescale for Karkh. The estimated number of background events from the linear model is $\int_0^T \hat{\mu}(t)\mathrm{d}t = 94$, which is 63.1 per cent of all the events. An estimated 55 events (36.9 per cent) are therefore the results of self-excitation. Overall these proportions are very similar to that observed in Karkh, suggesting a common partitioning of violence into background and foreground process.

Mosul

The third region we consider is Mosul, the second largest city in Iraq. There are 1300 events in Mosul occurring over 1718 days. The smoothed background model with self-excitation performs better than all other models, yielding an AIC value of 2545.1 (Table 4). The step function model ranks second. Figure 5 shows what looks like a steady increase in events that does not drop off to the same degree as in other areas. One might expect such a pattern to be a perfect candidate for a linear background rate, yet it performs only marginally better than a model with a stationary background rate.

The smoothed background rate model estimates $(1 - \hat{p})\hat{k}_0 = 0.3856$ (Table 4), meaning that on average every 10 events generate approximately four offspring events. Values for

Table 4: Parameter estimates for every event in Mosul and comparisons of models with and without self-excitation[a]

City	$\hat{\mu}$	$\hat{k}_0$	$\hat{w}^{-1}$	*Hawkes AIC*	*NoSE AIC*	*Best fit*
Mosul (μ)	0.0533	1.0024	58.82	2570.5	3370.4	Hawkes
Mosul (μ_{step})	0.0969, 0.4169, 0.5639	0.7123	68.28	2558.4	2611.2	Hawkes
Mosul (μ_l)	0.0950, 0.0008	0.7354	49.02	2570.4	2626.9	Hawkes
Mosul (μ_{sm})	0.6344	0.3856	41.08	2545.1	2551.8	Hawkes

[a]Parameter specifications as in Table 2.

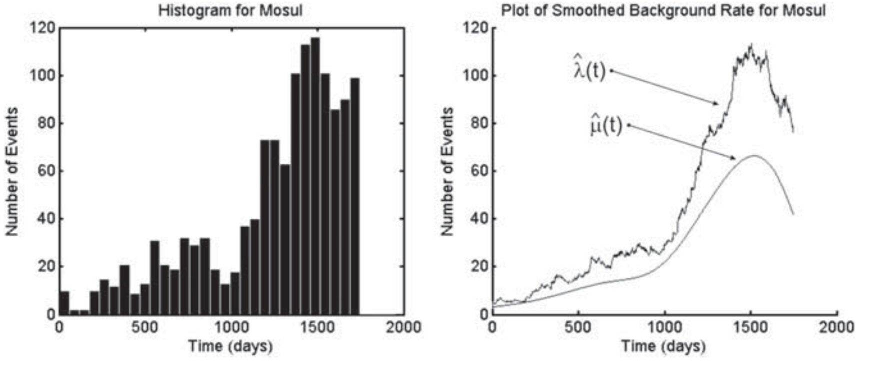

Figure 5: A histogram of the all events in Mosul (left). The estimated fit of the data for the smoothed background rate model $\hat{\lambda}(t)$ is plotted on the right. The smoothed background rate $\hat{\mu}_{sm}(t)$ is plotted on the right for reference.

$\hat{w}^{-1}$ vary much more widely across models. The value of $\hat{w}^{-1}$ for the smoothed background model is 41.08 days, suggesting a much longer time-scale for self-excited events compared to Karkh and Najaf. Note that we estimate w not w^{-1}, which means that small differences in $\hat{w}$ are amplified when looking at $\hat{w}^{-1}$. The estimate for the number of background events in the smoothed background model is $n\hat{p}=657$ events, which is 50.5 per cent of the total number of events, slightly less than in Karkh and Najaf. Accordingly, $n(1-\hat{p})=643$ events (49.5 per cent) are attributable to self-excitation.

Fallujah

The last region we consider is Fallujah. There are 501 events in this region over 1748 days. Here too the smoothed background rate model with self-excitation yields the smallest AIC value of 1929.8 (Table 5). The step function model outperforms the linear model, most likely due to the drop-off in events near the end of the data set. The close fit of the smoothed background model is represented in Figure 6.

Similar to Karkh and Mosul, the smoothed background rate model estimates $(1-\hat{p})\hat{k}_0=0.402$ suggesting every 10 events produce approximately four offspring events on average. The value of $\hat{w}^{-1}$ from the smoothed background rate model has self-excitation occurring on a

Table 5: Parameter estimates for every event in Fallujah and comparisons of models with and without self-excitation[a]

City	$\hat{\mu}$	$\hat{k}_0$	$\hat{w}^{-1}$	*Hawkes AIC*	*NoSE AIC*	*Best fit*
Fallujah (μ)	0.0394	0.8788	23.52	1952.0	2277.0	Hawkes
Fallujah (μ_{step})	0.0605, 0.0350, 0.1717	0.6739	19.85	1944.5	2011.5	Hawkes
Fallujah (μ_l)	0.0447, 0.0002	0.7758	20.33	1949.8	2054.8	Hawkes
Fallujah (μ_{sm})	0.6020	0.4017	17.20	1929.8	1946.2	Hawkes

[a]Parameter specifications as in Table 2.

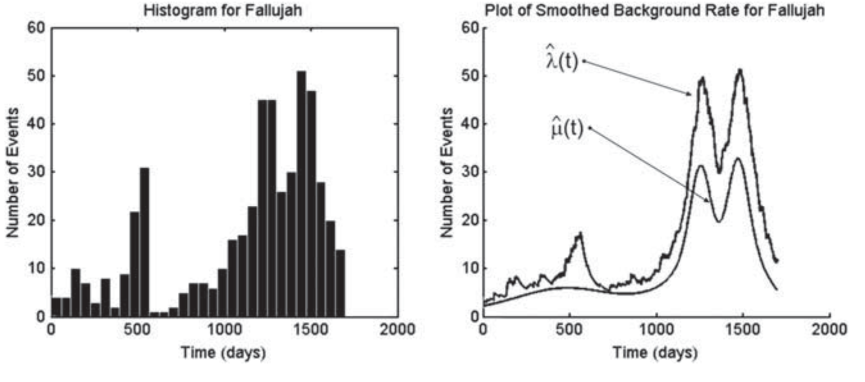

Figure 6: A histogram of the all events in Fallujah (left). The estimated fit of the data for the smoothed background rate model $\hat{\lambda}(t)$ is plotted on the right. The smoothed background rata $\hat{\mu}_{sm}(t)$ is plotted on the right as well for reference.

time-scale of 17 days. The estimated number of background events is $n\hat{p}=302$ which is about 60 per cent of all the events in Fallujah. The inferred number of self-excited events is $n(1-\hat{p})=199$ (40 per cent).

Discussion and Conclusion

Self-excited point process models partition the rate of events occurring in time (and space) into background and foreground components. Background processes generate events that are statistically independent of one another, meaning that one background event neither increases nor decreases the likelihood of a subsequent event. The most well-known background point process is a simple Poisson process, which generates events at a stationary or constant rate in time. By contrast, foreground processes are viewed as generating events that are statistically dependent. If events at one time (and place) increase the likelihood of subsequent events, then the process is said to be self-exciting. It is most common to assume that foreground processes operate over relative fast temporal and spatial scales, creating short-term temporal correlations between events. Background processes are therefore reasonably seen as either stationary in time, or changing only over only long time-scales.

Most self-exciting point process models have treated background rates as stationary Poisson processes and superimpose a short-term self-exciting process (Ogata, 1988; Egesdal *et al*, 2010). Such models may be appropriate where there is good reason to believe that the broad structural characteristics of the physical or social environment are relatively invariant. Where the environment is changing, even at relatively long time-scales, models that assume a stationary background rate may be inaccurate. Such is the case in the present study, where we have sought to model the time-course of civilian violent deaths over more than four years of the Iraqi conflict. It seems unreasonable to assume that the background rate of civilian deaths remained constant between March 2003 and December 2007 because of the deep structural changes to the security environment that occurred over this time. Moreover, assuming a stationary background rate would also require that virtually all of the rise and fall in violence is a result of short-term self-exciting processes. For example, a stationary background rate model for Karkh, a district in Baghdad, would imply that only 105 of 2278 events (4.6 per cent) are attributable background processes. The remaining 95.4 per cent of events would necessarily be attributed foreground self-exciting processes. While it is true that the relative proportion of foreground events in the Iraqi case seems unprecedented by comparison with garden-variety crime (see below), attributing all but 5 per cent to self-excitation seems unwarranted.

The models presented here are unique in that they consider a special class of point-processes where the background rate is non-stationary. We consider three models including: (1) a step function background rate; (2) a linear increasing background rate; and (3) a smoothed background rate. We also include a stationary background rate model for comparison. The models were fit to observed violent events derived from Iraq Body Count using maximum likelihood estimation and models were compared using AIC. In no instance did a stationary background rate model outperform a model with a non-stationary background rate. Moreover, AIC values suggest that models which include a self-exciting component outperform the corresponding model where self-excitation has been removed. These results confirm that both long and short time-scale processes are simultaneously important components of the dynamics of violence in Iraq. Focusing on either background or foreground processes to the exclusion of the other neglects a substantial fraction of the causal dynamics. Indeed, for models favored by AIC, we find that that background events make up between 50 and 63 per cent of all observed violent events. Accordingly, self-excited events are estimated to comprise 37–50 per cent of all events. By comparison, self-excitation accounts for approximately 13 per cent of events in a study of residential burglary in Los Angeles (Mohler *et al*, 2011). Clearly, self-excitation drives a substantial proportion of the total violence in Iraq.

Our comparisons of different regions within Iraq suggest both commonalities and differences in the short time-scale, foreground dynamics of violence. In three of the four regions examined here, including Karkh, Mosul and Fallujah, for every 10 initiating events we expect four self-excited offspring events (that is, $\hat{k}_0 = 0.4017$). In Najaf, every 10 initiating events are expected to generate approximately five offspring events (that is, $\hat{k}_0 = 0.501$). It is possible therefore that the number of daughter events generated by an initiating event is characteristic across all regions of Iraq. By contrast, there is considerable variability in the time-scales over which self-excited events occur. Examining the models favored by AIC suggests that self-excited daughter events may happen on average within one week of an initiating event (Najaf), within approximately two weeks (Karkh, Fallujah), or within slightly more than one month (Mosul). Such regional differences may stem from unique situational characteristics in each area. For example, locations that see fewer daughter events per initiating event may suggest that local populations and institutions responded more effectively to short-term, immediate security threats by damping opportunities for follow-on violence. Similarly, self-exciting violence that seems to play out over short time windows may suggest better organization on the part of the perpetrators of violence. Given the lack of direct ethnographic scale evidence from the settings analyzed here, such observations are necessarily speculative. Nevertheless, the criminological literature leads us to believe there are also good generic behavioral reasons to suppose that the violence on-the-ground in Iraq was driven by a combination of background and foreground self-excited components (Jacobs and Wright, 2006; Short *et al*, 2008; Townsley *et al*, 2008; Andresen *et al*, 2009; Mohler *et al*, 2011).

Our results may hold implications for designing strategies to deal with security threats that evolve on multiple time-scales. At a coarse level, the fraction of events attributable to background and foreground processes may provide a rough gauge of the appropriate mix of strategies to aim at long and short time-scales, respectively. However, this says nothing about differences in the costs and logistical complexities of tackling security problems arising at different times. Addressing the causes of violence arising on long time-scales may be particularly challenging (Kilcullen, 2009; Bjelopera and Randol, 2010). Strategies to deal with short time-scale security problems may be more accessible and may yield more immediate results (Clarke and Newman, 2006), both in terms of the volume of impact and the local spatio-temporal dynamics of violence. If our conclusion that 37–50 per cent of events in the IBC data is the result of foreground self-exciting processes, then security strategies that interrupt even a fraction of these events may have a large impact on the volume of violence.

Our analyses also indicate the time-scales at which security responses may be most effective. For example, the fastest self-excitation time-scale $\hat{w}^{-1}$ is observed in Najaf where foreground events occur on average within eight days of initiating events. To interrupt such a process, one must be able to respond to the initiating events within eight days. In Mosul, one might have greater flexibility to interrupt self-excitation given that such events occur on average within 41 days of initiating events. However, we can only speculate on the potential impact of getting inside these short-term temporal processes. One possible outcome is that the average number of daughters generated by any one initiating event k_0 may decline.

Future work could include adding a conditional magnitude to the intensity function, where magnitude would be measured by the severity of the attack. This is routine in the use of models to study earthquake dependencies where large magnitude earthquakes are more likely to generate aftershocks than smaller magnitude earthquakes. A similar dynamic may characterize violent acts, where large magnitude attacks cause more combatants to 'pile in'. To our knowledge, this has not been done with respect to criminal or other violent activity. It may be difficult to determine the magnitude of a violent attack, but in the present case the number of deaths occurring in each event may serve as a proxy for the magnitude. Another possible direction for the future work involves estimating the background intensity more accurately. Considering a semi-parametric or non-parametric estimation of the background rate, or possibly using other data sets like troop levels could prove effective. In earthquake research, different choices of the triggering function g in equation (1) have been analyzed and compared (Ogata, 1998). Similar work could be done here to determine accuracy of triggering densities.

Acknowledgements

We would like to thank Rick Schoenberg and two anonymous referees for their insights. This work was supported by ARO grants (STIR) W911NS-09-1-0559 and W911NF1010472, ARO MURI grant 50363-MA-MUR, ONR grant N000141010221 and AFOSR-MURI grant FA9550-10-1-0569.

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In: B.N. Petrov and F. Caski (eds.) *Proceedings of the Second International Symposium on Information Theory*. Budapest, Hungary: Akademiai Kiado.

Akaike, H. (1974) A new look at statistical model identification. *IEEE Transactions on Automatic Control* AC19(6): 716–723.

Akers, R. (2008) *Social Learning and Social Structure: A General Theory of Crime and Deviance*. Boston, MA: Northeastern University Press.

Andresen, M.A., Brantingham, P.J. and Kinney, J.B. (2009) *Classics in Environmental Criminology*. Boca Raton, FL: CRC Press.

Berman, E. (2009) *Radical, Religious, and Violent: The New Economics of Terrorism*. Cambridge, MA: The MIT Press.

Bjelopera, J.P. and Randol, M.A. (2010) American Jihadist terrorism: Combating a complex threat. Congressional Research Service.

Bozdogan, H. (1987) Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52(3): 345–370.

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

Brantingham, P., Brantingham, P. and Taylor, W. (2005) Situational crime prevention as a key component in embedded crime prevention. *Canadian Journal of Criminology and Criminal Justice/La Revue canadienne de criminologie et de justice pénale* 47(2): 271–292.

Brantingham, P.J. and Brantingham, P.L. (1984) *Patterns in Crime*. New York: Macmillan.
Canela-Cacho, J.A., Blumstein, A. and Cohen, J. (1997) Relationship between the offending frequency (λ) of imprisoned and free offenders. *Criminology* 35(1): 133–176.
Chabat, J. (2002) Mexico's war on drugs: No margin for maneuver. *The Annals of the American Academy of Political and Social Science* 582(1): 134–148.
Clarke, R.V. and Newman, G.R. (2006) *Outsmarting the Terrorists*. Westport, CT: Praeger Security International.
Cohen, L.E. and Felson, M. (1979) Social-change and crime rate trends: A routine activity approach. *American Sociological Review* 44(4): 588–608.
Cornish, D.B. and Clarke, R.V. (eds.) (1986) *The Reasoning Criminal: Rational Choice Perspectives on Criminal Offending*. New York: Springer-Verlag.
Cornish, D.B. and Clarke, R.V. (1987) Understanding crime displacement: An application of rational choice theory. *Criminology* 25(4): 933–947.
Daley, D. and Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes*. New York: Springer Verlag.
Daley, D. and Vere-Jones, D. (2008) *An Introduction to the Theory of Point Processes*, 2nd edn. New York: Springer Verlag.
Dugan, L. (1999) The effect of criminal victimization on a household's moving decision. *Criminology* 37(4): 903–930.
Egesdal, M., Fathauer, C., Louie, K. and Neuman, J. (2010) Statistical and stochastic modeling of gang rivalries in Los Angeles. *SIAM Undergraduate Research Online* 3: 72–94.
Farrell, G., Tilley, N., Tseloni, A. and Mailley, J. (2010) Explaining and sustaining the crime drop: Clarifying the role of opportunity-related theories. *Crime Prevention & Community Safety* 12(1): 24–41.
Felson, M. and Clarke, R.V. (1998) *Opportunity Makes the Thief: Practical Theory for Crime Prevention*. London: Home Office Policing and Reducing Crime Unit.
Felson, R.B. and Steadman, H.J. (1983) Situational factors in disputes leading to criminal violence. *Criminology* 21(1): 59–74.
Fisher, R. (1922) On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A* 222(602): 309–368.
Hawkes, A.G. (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1): 83.
Hawkes, A.G. and Oakes, D. (1974) Cluster process representation of a self-exciting process. *Journal of Applied Probability* 11(3): 493–503.
Hicks, M.H.-R., Dardagan, H., Guerrero Serdán, G., Bagnall, P.M., Sloboda, J.A. and Spagat, M. (2011) Violent deaths of Iraqi civilians, 2003–2008: Analysis by perpetrator, weapon, time, and location. *PLoS Medicine* 8(2): e1000415.
Hipp, J.R. (2010) A dynamic view of neighborhoods: The reciprocal relationship between crime and neighborhood structural characteristics. *Social Problems* 57(2): 205–230.
Holden, R.T. (1987) Time series analysis of a contagious process. *Journal of the American Statistical Association* 82(400): 1019–1026.
Hough, M., Mayhew, P. and Britain, G. (1983) *The British Crime Survey: First Report*. London: Her Majesty's Stationery Office.
Iraq Body Count. (2008) http://www.iraqbodycount.org. Data with spatial information additions supplied by IBC in February 2008.
Izady, M. (2011) Baghdad, Iraq, ethnic composition, 2003–2009, http://gulf2000.columbia.edu/maps.shtml.
Jacobs, B.A. and Wright, R. (2006) *Street Justice: Retaliation in the Criminal Underworld*. Cambridge, UK: Cambridge University Press.
Johnson, S. (2008) Repeat burglary victimisation: A tale of two theories. *Journal of Experimental Criminology* 4(3): 215–240.
Johnson, S.D. *et al* (2007) Space-time patterns of risk: A cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology* 23: 201–219.
Keizer, K., Lindenberg, S. and Steg, L. (2008) The spreading of disorder. *Science* 322(5908): 1681–1685.
Kilcullen, D. (2009) *The Accidental Guerrilla: Fighting Small Wars in the Midst of a Big One*. Oxford, UK: Oxford University Press.
Klein, M.W. and Maxson, C.L. (2006) *Street Gang Patterns and Policies*. New York: Oxford University Press.
Livingston, M. (2008) Alcohol outlet density and assault: A spatial analysis. *Addiction* 103(4): 619–628.
Luckenbill, D.F. (1977) Criminal homicide as a situated transaction. *Social Problems* 25(2): 176–186.

Malm, A.E., Kinney, B.J. and Pollard, N. (2008) Social network and distance correlates of criminal associates involved in illicit drug production. *Security Journal* 21(1): 77–94.
McCauley, C. and Moskalenko, S. (2008) Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence* 20(3): 415–433.
Miethe, T.D. and Sousa, W.H. (2010) Carjacking and its consequences: A situational analysis of risk factors for differential outcomes. *Security Journal* 23(4): 241–258.
Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P. and Tita, G.E. (2011) Self-exciting point process modeling of crime. *Journal of the American Statistical Association* 106(493): 100–108.
Morenoff, J., Sampson, R. and Raudenbush, S. (2001) Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology* 39(3): 517–558.
Myung, I.J. (2003) Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology* 47(1): 90–100.
Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point-processes. *Journal of the American Statistical Association* 83(401): 9–27.
Ogata, Y. (1998) Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics* 50(2): 379–402.
Osgood, D.W., Wilson, J.K., O'malley, P.M., Bachman, J.G. and Johnston, L.D. (1996) Routine activities and individual deviant behavior. *American Sociological Review* 61(4): 635–655.
Ozaki, T. (1979) Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics* 31: 145–155.
Rubin, I. (1972) Regular point processes and their detection. *IEEE Transactions on Information Theory* 18(5): 547.
Sampson, R.J. (2011) The community. In: J.Q. Wilson and J. Petersilia (eds.) *Crime and Public Policy*, 2nd edn. New York: Oxford University Press.
Sampson, R.J., Morenoff, J.D. and Gannon-Rowley, T. (2002) Assessing 'Neighborhood Effects': Social processes and new directions in research. *Annual Review of Sociology* 28: 443–478.
Sampson, R.J. and Wooldredge, J.D. (1987) Linking the micro- and macro-level dimensions of lifestyle-routine activity and opportunity models of predatory victimization. *Journal of Quantitative Criminology* 3(4): 371–393.
Sanchez-Jankowski, M.S. (1991) *Islands in the Street: Gangs and American Urban Society*. Berkeley, CA: University of California Press.
Schoenberg, F.P. (2003) Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association* 98(464): 789–795.
Schoenberg, F.P., Brillinger, D.R. and Guttorp, P. (2006) Point processes, spatial–temporal. In: A. El Shaarawi and W. Piegorsch (eds.) *Encyclopedia of Environmetrics*. New York: Wiley.
Short, M.B., D'Orsogna, M.R., Brantingham, P.J. and Tita, G. (2009) Measuring and modeling repeat and near-repeat burglary effects. *Journal of Quantitative Criminology* 25(3): 325–339.
Short, M.B. *et al* (2008) A statistical model of criminal behavior. *Mathematical Models and Methods in Applied Sciences* 18(S1): 1249–1267.
Silverman, B. (1998) *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.
Taniguchi, T.A., Rengert, G.F. and Mccord, E.S. (2009) Where size matters: Agglomeration economies of illegal drug markets in Philadelphia. *Justice Quarterly* 26(4): 670–694.
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.
Townsley, M., Johnson, S.D. and Ratcliffe, J.H. (2008) Space time dynamics of insurgent activity in Iraq. *Security Journal* 21(3): 139–146.
Trickett, A., Osborn, D.R., Seymour, J. and Pease, K. (1992) What is different about high crime areas. *British Journal of Criminology* 32(1): 81–89.
Vere-Jones, D. (2009) Some models and procedures for space-time point processes. *Environmental and Ecological Statistics* 16: 173–195.
Wortley, R. and McFarlane, M. (2011) The role of territoriality in crime prevention: A field experiment. *Security Journal* 24(2): 149–156.
Wright, R.T. and Decker, S.H. (1994) *Burglars on the Job: Streetlife and Residential Breaking*. Boston, MA: Northeastern University Press.
Wuthnow, R. (2005) Democratic renewal and cultural inertia: Why our best efforts fall short. *Sociological Forum* 20(3): 343–367.
Zhuang, J., Ogata, Y. and Vere-Jones, D. (2002) Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association* 97(458): 369–380.