

MODELING AND ESTIMATION FOR SELF-EXCITING SPATIO-TEMPORAL MODELS OF TERRORIST ACTIVITY

BY NICHOLAS J. CLARK¹ AND PHILIP M. DIXON

Iowa State University

Spatio-temporal hierarchical modeling is an extremely attractive way to model the spread of crime or terrorism data over a given region, especially when the observations are counts and must be modeled discretely. The spatio-temporal diffusion is placed, as a matter of convenience, in the process model allowing for straightforward estimation of the diffusion parameters through Bayesian techniques. However, this method of modeling does not allow for the existence of self-excitation, or a temporal data model dependency, that has been shown to exist in criminal and terrorism data. In this manuscript we will use existing theories on how violence spreads to create models that allow for both spatio-temporal diffusion in the process model as well as temporal diffusion, or self-excitation, in the data model. We will further demonstrate how Laplace approximations similar to their use in Integrated Nested Laplace Approximation can be used to quickly and accurately conduct inference of self-exciting spatio-temporal models allowing practitioners a new way of fitting and comparing multiple process models. We will illustrate this approach by fitting a self-exciting spatio-temporal model to terrorism data in Iraq and demonstrate how choice of process model leads to differing conclusions on the existence of self-excitation in the data and differing conclusions on how violence spread spatio-temporally in that country from 2003–2010.

1. Introduction. A typical spatio-temporal model consists of three levels, a data model, a process model, and a parameter model. A common way to model data then is to assume the data model, $Y(\cdot)$, is conditionally independent given the process model $X(\cdot)$. For example, if observations take place at areal regions, s_i , at discrete time periods, t , and $Y(s_i, t)$ are counts, a common model is $Y(s_i, t) | X(s_i, t) \sim \text{Pois}(\exp(X(s_i, t)))$. The spatio-temporal diffusion structure is commonly then placed on the process model which commonly is assumed to have a Gaussian joint distribution of $X \sim \text{Gau}(\mathbf{0}, Q^{-1}(\theta))$. The majority of analysis of these models is done using Bayesian techniques requiring a further parameter model for θ . The challenge in these models is, then, determining an appropriate structure for $Q^{-1}(\theta)$ or $Q(\theta)$. If both the covariance and the precision is chosen to be too dense inference quickly becomes impossible due to the size of $Q^{-1}(\theta)$. In spatio-temporal models it is quite common for the dimension of Q to be larger

Received March 2017; revised August 2017.

¹Supported in part by an Omar Nelson Bradley Fellowship.

Key words and phrases. Laplace approximations, nonseparable space–time, spatial-temporal dependence, Bayesian, terrorism.

than, say, $10^4 \times 10^4$. A thorough overview giving many examples of this method of modeling is given in [Cressie and Wikle \(2012\)](#).

In modeling terrorism or crime data one possibility is to use an extremely general spatio-temporal process model to capture variance not explained through the use of covariates. For example, [Python et al. \(2016\)](#) use a Matern class covariance function over space and an AR(1) process over time. They then use covariates to test the impact of infrastructure, population, and governance. The general spatio-temporal process models used, in this case, has an extremely sparse precision structure greatly aiding in computations.

While diffusion in spatio-temporal models is often modeled through a latent process, more recent models describing the spread of violence have incorporated self-excitation, or spatio-temporal diffusion that exists linearly in the data model itself. Self-excitation is the theory that in terrorism, or crime, the probability of an event occurring is a function of previous successful events. For instance [Mohler et al. \(2011\)](#) demonstrate that burglars are more likely to rob locations that have previously and successfully been robbed. [Mohler \(2013\)](#) derived a class of models that allowed for temporal diffusion in both the process model as well as the data model and demonstrated how the two processes were identifiable.

In the modeling of terrorism data [Lewis et al. \(2012\)](#), [Porter and White \(2012\)](#), and [Mohler \(2013\)](#) have all successfully used the self-excitation approach to model. Most recently, [Tench, Fry and Gill \(2016\)](#) used a likelihood approach for temporal modeling of IEDs in Northern Ireland using self-excitation. However, in these papers the existence and analysis of self-excitation was the primary objective and any process model dependency was either ignored or treated as a nuisance. The one exception is in [Mohler \(2013\)](#) where a temporal only model was assumed for the process model and inference was conducted on both the process model dependency and the data model dependency.

In this manuscript, we will consider a spatial and a spatial-temporal process model that allows for self-excitation. We will present two self-exciting models for terrorist activity that have different process models corresponding to different notions of how terrorism evolves in time and space as well as temporal dependency in the data model to account for self-excitation. These two models are specific cases of more general spatio-temporal models that allow dependency in both the process model as well as the data model.

We will further show how Laplace approximations similar to their use in Integrated Nested Laplace Approximation (INLA), an approximate Bayesian method due to [Rue, Martino and Chopin \(2009\)](#), can be used to conduct inference for these types of models. We will show, via simulation, how INLA, when appropriately modified, can accurately be used to make inference on process level parameters for self-exciting models and aid analysts in determining the appropriate process model when scientific knowledge cannot be directly applied as in [Cressie and Wikle \(2012\)](#). Finally, we will apply this technique to terrorism data in Iraq. We will show that choice of process model, in this case, results in differing conclusions on the impact of self-excitation in the model.

2. Self-exciting spatio-temporal models. The use of self-exciting models in both criminal and terrorism modeling has become increasingly popular over the last decade after being originally introduced in Short et al. (2008). Self-excitement, in a statistical model, directly models copycat behavior by letting an observed event increase the intensity (or excites a model) over a specified time or location. Self-exciting models are closely related to Hawkes processes, which are counting processes, where the probability of an event occurring is directly related to the number of events that previously occurred. In a self-exciting model, the criminal intensity at a given spatio-temporal location, (x, y, t) , is a mixture of a background rate, ν , and self-excitement function, $f(\mathcal{H}_{x,y,t})$ that is dependent on the observed history at that location, $\mathcal{H}_{x,y,t}$.

A common temporal version of a discretized Hawkes process is

$$(2.1) \quad Y_t \sim \text{Pois}(\lambda_t),$$

$$\lambda_t = \nu + \sum_{j < t} \kappa(t - j)y_j,$$

$$(2.2) \quad t \in \{1, 2, \dots, T\}.$$

In this example, in order for the process to have finite expectation in the limit, $\kappa(t - j)$ must be positive and $\sum_{i=1}^{\infty} \kappa(i) < 1$. $\kappa(t - j)$ can be thought of as the probability that an event at time j triggers an event at time i . Laub, Taimre and Pollett (2015) provides an excellent overview of the mathematical properties of the continuous Hawkes process and the discrete process when $\kappa(t - j)$ is taken to be an exponential decay function.

In Mohler et al. (2011), ν was treated as separable in space and time and was nonparametrically estimated using stochastic declustering; while in Mohler (2013), the spatial correlations were ignored and an AR(1) process was used for $\kappa(t)$ and an exponential decay was assumed for the self-excitation. In terrorism modeling Lewis et al. (2012) used a piecewise linear function for $\kappa(t)$.

Here we will first define a general model that allows spatial or spatial-temporal correlation to exist in the process model and positive temporal correlation to exist in the data model to allow for self-excitation. First define $s_i \in \mathcal{R}^2, i \in \{1, 2, \dots, s\}$ as locations in a fixed, areal, region. We further define $t \in \{1, 2, \dots, n\}$ as discrete time. The general form of a spatial-temporal self-exciting model is then given in (2.3):

$$(2.3) \quad Y(s_i, t) | \mu(s_i, t) \sim \text{Pois}(\mu(s_i, t)),$$

$$\mu(s_i, t) = \exp(X(s_i, t)) + \eta Y(s_i, t - 1),$$

$$X \sim \text{Gau}(\mathbf{0}, Q^{-1}(\theta)).$$

Comparing the above to the Hawkes process, we now have ν as a function of space–time and denote it as $X(s_i, t)$. We use the simplest form of self-excitation

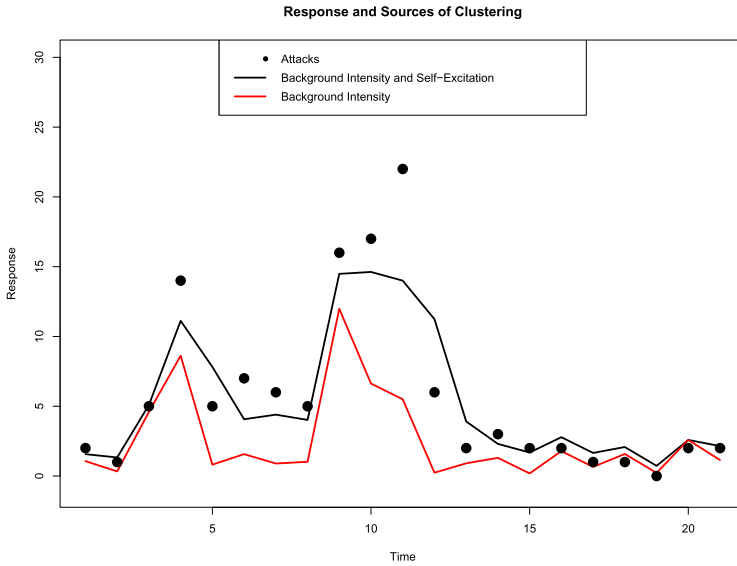


FIG. 1. This figure shows an example of the expectation of two processes, one with self-excitation and one without. The bottom line is the expectation of a process with no self-excitation, the top has self-excitation of $\eta = 0.4$. The data realizations are from the process with self-excitation.

letting $\kappa(t - j)$ be a point-mass function such that $\kappa(k) = \eta$ for $k = 1$ and $\kappa(k) = 0$ for $k \neq 1$. In all cases $Y(s_i, t)$ will be discrete, observable, count data.

To contrast (2.3) with a typical spatial model, Figure 1 depicts the expectation for one areal location (s_i, t) without self-excitation and with self-excitation as shown in Figure 1. In this figure, the lower line shows $\mu(s_i, t)$ with $\eta = 0$, and the upper line has $\eta = 0.4$. The impact of self-excitation is clearly present in time 10–13.

2.1. *Spatially correlated self-exciting model.* In the first example of (2.3) we assume the background intensity rate, $X(s_i, t)$ has only spatial correlation. This model is motivated through the assumption that the latent dependency, $X(s_i, t)$, is as a continuous measure of violent tendency at region s_i at time period t , and regions that are closer together in space are assumed to share common characteristics.

Next, define $N(s_i)$ as the neighborhood of location s_i where two regions are assumed to be neighbors if they share a common border. $|N(s_i)|$ is the number of neighbors of location s_i . The model for $Y(s_i, t)$ or the number of observed violent events at a given space–time location is then given by

$$\begin{aligned}
 Y(s_i, t) | \mu(s_i, t) &\sim \text{Pois}(\mu(s_i, t)), \\
 (2.4) \quad \mu(s_i, t) &= \exp(X(s_i, t)) + \eta Y(s_i, t - 1),
 \end{aligned}$$

$$X(s_i, t) = \theta_1 \sum_{s_j \in N(s_i)} X(s_j, t) + \varepsilon(s_i, t),$$

$$\varepsilon(s_i, t) \sim \text{Gau}(0, \sigma^2).$$

Letting \mathbf{H} denote the spatial neighborhood matrix such that $H_{i,j} = H_{j,i} = 1$ if s_i and s_j are neighbors, the full distribution of the joint distribution of the latent state is $\mathbf{X} \sim \text{Gau}(\mathbf{0}, (\mathbf{I}_{ns,ns} - \mathbf{I}_{n,n} \otimes \theta_1 \mathbf{H})^{-1} \mathbf{L} (\mathbf{I}_{ns,ns} - \mathbf{I}_{n,n} \otimes \theta_1 \mathbf{H})^{-1})$ where $\mathbf{L} = \text{diag}(\sigma^2, \dots, \sigma^2)$. The evolution in the latent field is equivalent to the spatial evolution in what is commonly referred to as a Simultaneous or Spatial Auto-regressive (SAR) model. Alternatively, the Conditional Auto-regressive (CAR) model of Besag (1974) could be used to model the latent state modifying the joint density above.

The difference between the SAR and (2.4) is in the self-excitement parameter, η . In (2.4), temporal correlation is present, but it is present through the data model specification rather than through a temporal evolution in the latent state. Therefore, temporal correlation is a function solely of the self-excitation in the process. η gives the probability of an event at time $t - 1$ creating an event at time t . In order for the system to be well behaved, η is constrained to (0,1). In order for the joint distribution of \mathbf{X} to be valid, $\theta_1 \in (\psi_{(1)}^{-1}, \psi_{(n)}^{-1})$ where ψ_i is the i th smallest eigenvalues of \mathbf{H} .

The critical assumption in this model is that the propensity of a given location to be violent is spatially correlated with its adjacent spatial neighbors and only evolves over time as a function of excitation. If terrorism is diffusing according to this model, regions that are geographically adjacent are behaving in a similar manner. The existence of self-excitation would indicate that individuals within a region are being inspired through the actions of others. While combating terrorism is complex, if terrorism is diffusing in this manner, one suggestion would be to identify the root causes within a geographic area as well as quick action against any malicious actor to discourage copycat behavior.

2.2. *Reaction–diffusion self-exciting model.* Alternatively, a model similar to Short et al. (2008) can be used to motivate the process model for the latent state resulting in a nonseparable spatio-temporal, \mathbf{X} . Here, $X(s_i, t)$ corresponds to a continuous measure of violence due to terrorists or criminals at location s_i at time t . This is still a latent variable as we are not directly measuring $X(s_i, t)$. However, now, in order for an area to increase in violent tendency, a neighboring area must decrease as the actors causing the violence move from location to location. Furthermore, if terrorists are removed from the battlefield at a rate proportional to the total number of terrorists present and if terrorists move to fill power vacuums, the process model is similar to the reaction–diffusion partial differential equation [see Cressie and Wikle (2012) for more on the reaction–diffusion model]

$$(2.5) \quad \frac{\partial X(s_i, t)}{\partial t} = \frac{\kappa}{|N(s_i)|} \Delta X(s_i, t) - \alpha X(s_i, t).$$

In order to generalize this partial differential equation (PDE) to an irregular lattice, we make use of the graphical Laplacian, Γ , in place of Δ in (2.5). Γ is a matrix that extends the notion of second derivatives to irregular graphs and can be defined as a matrix of the same dimension as the number of geographical regions with entries given by

$$\Gamma(s_i, s_j) \begin{cases} -N(s_i), & j = i, \\ 1, & j \in N(s_i), \\ 0, & \text{otherwise.} \end{cases}$$

With the addition of a random noise term assumed to be Gaussian, the full model can be seen as an example of (2.3):

$$\begin{aligned} (2.6) \quad & Y(s_i, t) | \mu(s_i, t) \sim \text{Pois}(\mu(s_i, t)), \\ & \mu(s_i, t) = \exp(X(s_i, t)) + \eta Y(s_i, t - 1), \\ & X(s_i, t) = \frac{\kappa}{|N(s_i)|} \sum_{s_j \in N(s_i)} X(s_j, t - 1) \\ & \quad + (1 - \kappa - \alpha) X(s_i, t - 1) + \varepsilon(s, t), \\ & \varepsilon(s, t) \sim \text{Gau}(0, \sigma^2). \end{aligned}$$

In contrast to the Spatially Correlated Self-Exciting (SCSE) model, the process model dependency exists in both space and time. In order to derive properties of this model, we first let $\mathbf{M} = \kappa \text{diag}(\frac{1}{|N_{s_i}|})\Gamma + (1 - \alpha)\mathbf{I}_{s,s}$ and now note that this is equivalent to a Vector Auto-Regressive (VAR) model $\mathbf{X}_t = \mathbf{M}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \sim \text{Gau}(\mathbf{0}, \sigma^2\mathbf{I})$.

The VAR(1) model requires all the eigenvalues of \mathbf{M} to be between -1 and 1 . This can be satisfied by first noting that 0 is always an eigenvalue of $\text{diag}(\frac{1}{|N_{s_i}|})$ trivially corresponding to the eigenvector of all 1 's. The largest eigenvalue is at most 2 as shown in Chung (1997). Due to the structure of $(1 - \alpha)\mathbf{I}_{s,s}$ this implies maximum eigenvalue of \mathbf{M} is $(1 - \alpha)$ and minimum is $-2\beta + (1 - \alpha)$. Therefore, the parameter spaces for α and κ are $\alpha \in (0, 1)$ and $\kappa \in (\frac{-\alpha}{2}, \frac{2-\alpha}{2})$.

Just as in the SCSE model, if ε has a Gaussian distribution, the Reaction–Diffusion Self-Exciting (RDSE) model has an exact solution for the latent Gaussian field, \mathbf{X} .

Letting Σ_s be the spatial covariance at a fixed period of time which is assumed to be invariant to time, then we can solve for Σ_s by using the relationship $\Sigma_s = \mathbf{M}\Sigma_s\mathbf{M}^T + \sigma^2\mathbf{I}$. As demonstrated by Cressie and Wikle (2012), this leads to $\text{vec}(\Sigma_s) = (\mathbf{I}_{s^2, s^2} - \mathbf{M} \otimes \mathbf{M})^{-1} \text{vec}(\sigma^2\mathbf{I}_{s,s})$ where $\text{vec}(\cdot)$ is the matrix operator that stacks each column of the matrix on top of one or another. Recall that s is the size of the lattice that is observed at each time period. The joint distribution for all

X is then $X \sim \text{Gau}(\mathbf{0}, Q_{\text{rd}}^{-1}(\theta))$ where

$$(2.7) \quad Q_{\text{rd}}^{-1}(\theta) = \begin{bmatrix} \Sigma_s & | & M\Sigma_s & | & \dots & | & M^n \Sigma_s \\ \hline \Sigma_s M^T & | & \Sigma_s & | & \dots & | & M^{n-1} \Sigma_s \\ \dots & | & \dots & | & \dots & | & \dots \\ \hline \Sigma_s (M^T)^n & | & \Sigma_s (M^T)^{n-1} & | & \dots & | & \Sigma_s \end{bmatrix}.$$

However, practically, this involves inverting a potentially large matrix $\mathbf{I}_{s^2, s^2} - \mathbf{M} \otimes \mathbf{M}$. Therefore, it is easier to deal with the inverse of (2.7) given in (2.8):

$$(2.8) \quad Q_{\text{rd}}(\theta) = \begin{bmatrix} \mathbf{I}_{n,n} & | & -M & | & \mathbf{0} & | & \dots & | & \dots \\ \hline -M^T & | & M^T M + \mathbf{I}_{n,n} & | & -M & | & \mathbf{0} & | & \dots \\ \mathbf{0} & | & -M^T & | & M^T M + \mathbf{I}_{n,n} & | & -M & | & \dots \\ \dots & | & \dots & | & \dots & | & \dots & | & \dots \\ \hline \mathbf{0} & | & \dots & | & -M^T & | & M^T M + \mathbf{I}_{n,n} & | & -M \\ \hline \mathbf{0} & | & \dots & | & \dots & | & -M^T & | & \mathbf{I}_{n,n} \end{bmatrix} \frac{1}{\sigma^2}.$$

The primary difference between the SCSE model and the RDSE model is that the process model correlation in the SCSE is only spatial while in the RDSE it is spatio-temporal. In the toy examples below, we show the expectation for $X(s_i, t)$ for both the SCSE and the RDSE models on a 4×4 lattice structure. We fixed both models with a value of $X(s_1, 1) = 10$ as the upper left-hand observation at time 1. As seen in the RDSE model, the high count at time 1 spreads to neighboring regions in time 2 and time 3, whereas the process model has no temporal spread in the SCSE but has a high level of spatial spread.

Spatially Correlated Latent Process Conditional on $(s_1, 1) = 10$

Time 1				Time 2			
10	5	2	1	0	0	0	0
5	4	3	2	0	0	0	0
2	3	2	1	0	0	0	0
1	1	1	1	0	0	0	0

Reaction-Diffusion Latent Process Conditional on $(s_1, 1) = 10$

Time 1				Time 2				Time 3			
10	1	1	0	3	3	0	0	2	1	1	0
1	1	0	0	3	1	0	0	1	1	0	0
1	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

Practically, if data follows the RDSE model, it implies a high terrorism count in one region will manifest into a high terrorism count in a neighboring region at a later time period. In combating terrorism, the RDSE model might suggest isolating geographical regions to mitigate the risk of spread while addressing self-excitation through direct action against malicious actors who are inspiring others.

3. Model fitting. In both the RDSE and the SCSE, spatio-temporal diffusion exists in both the process model and the data model. If the diffusion was solely in the process model, a technique for inference would be INLA.

INLA was first proposed in [Rue, Martino and Chopin \(2009\)](#) to specifically address the issue of Bayesian inference of high-dimensional latent Gaussian random fields (LGRFs). An example of this for count data is

$$\begin{aligned}
 Y(s_i) &\sim \text{Pois}(\mu(s_i, t)), \\
 \mu(s_i, t) &= \exp(\lambda(s_i, t)), \\
 \lambda(s_i, t) &= \beta_0 + \mathbf{Z}^t \boldsymbol{\beta} + X(s_i, t), \\
 \mathbf{X} &\sim \text{Gau}(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta})).
 \end{aligned}
 \tag{3.1}$$

INLA is often preferable over Markov chain Monte Carlo (MCMC) for these types of models. An issue with traditional MCMC techniques for these models is that the dimension of X is often very large. Therefore, while MCMC has $O_p(N^{-1/2})$ errors, the N in the errors is the simulated sample size for the posterior. Just getting $N = 1$ may be extremely difficult due to the vast number of elements of X that need to be estimated. In general, MCMC will take hours or days in order to successfully simulate from the posterior making the computational cost of fitting multiple process models extremely high. In [Python et al. \(2016\)](#), terrorism data was fit using a grid over the entire planet using INLA, though without self-excitation in the model.

To address the issues with MCMC use in LGRFs, [Rue, Martino and Chopin \(2009\)](#) developed a deterministic approach based on multiple Laplacian approximations. A LGRF is any density that can be expressed as

$$\begin{aligned}
 \pi(\boldsymbol{\theta}, \mathbf{X} | \mathbf{Y}) &\propto \pi(\boldsymbol{\theta}) |\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \\
 &\times \exp \left[\frac{-1}{2} \mathbf{X}^t \mathbf{Q}(\boldsymbol{\theta}) \mathbf{X} + \sum_s \log(\pi(Y(s_i) | X(s_i), \boldsymbol{\theta})) \right].
 \end{aligned}
 \tag{3.2}$$

In order to conduct inference on this model, we need to estimate $\pi(\boldsymbol{\theta} | \mathbf{y})$, $\pi(\theta_i | \mathbf{y})$ and $\pi(x_i | \mathbf{y})$. The main tool [Rue, Martino and Chopin \(2009\)](#) employ is given in their equation (3) as

$$\tilde{\pi}(\boldsymbol{\theta} | \mathbf{Y}) \propto \frac{\pi(\mathbf{X}, \boldsymbol{\theta}, \mathbf{Y})}{\tilde{\pi}_G(\mathbf{X} | \boldsymbol{\theta}, \mathbf{Y})} \Big|_{\mathbf{X}=\mathbf{x}^*(\boldsymbol{\theta})}.
 \tag{3.3}$$

In Rue, Martino and Chopin (2009), they note that the denominator of (3.3) almost always appears to be unimodal and approximately Gaussian. The authors then propose to use a Gaussian approximation to $\pi(\mathbf{X}|\boldsymbol{\theta}, \mathbf{Y})$, which is denoted above as $\tilde{\pi}_G$. Moreover, (3.3) should hold no matter what choice of \mathbf{X} is used, so a convenient choice for \mathbf{X} is the mode for a given θ , which Rue, Martino and Chopin (2009) denote as $x^*(\boldsymbol{\theta})$.

Now, $\pi(\boldsymbol{\theta}|\mathbf{Y})$ can be explored by calculating the marginal for choices of θ , which if chosen carefully can greatly decrease the computational time. These explored values can then be numerically integrated out to get credible intervals for $\pi(\theta_i|\mathbf{Y})$.

Following the exploration of $\theta|Y$, and computation of $\theta_i|Y$, INLA next proceeds to approximate $\pi(X(s_i)|\boldsymbol{\theta}, \mathbf{Y})$. The easiest way to accomplish this is to use the marginals that can be derived straightforwardly from $\tilde{\pi}_G(\mathbf{X}|\boldsymbol{\theta}, \mathbf{Y})$ from (3.3). In this paper we will use this technique for simplicity of computation, however, if the latent states are of interest in the problem (and they often are), this can be problematic as it fails to capture any skewness of the posterior of \mathbf{X} . One way to correct this is to re-apply (3.3) in the following manner:

$$(3.4) \quad \tilde{\pi}_{LA}(X(s_i)|\theta, y) \propto \frac{\pi(\mathbf{X}, \theta, \mathbf{Y})}{\tilde{\pi}_G(\mathbf{X}_{-s_i}|X(s_i), \boldsymbol{\theta}, \mathbf{Y})} \Big|_{x_{-i}=\mathbf{x}_{-i}^*(x_i, \theta)}.$$

In (3.4) \mathbf{X}_{-s_i} is used to represent \mathbf{X} with latent variable $X(s_i)$ removed. This is a reapplication of Tierney and Kadane’s marginal posterior density and gives rise to the nested term in INLA.

3.1. *Laplace approximation for spatio-temporal self-exciting models.* While INLA is an attractive technique due to computational speed and implementation, it is not immediately usable for the SCSE and the RDSE as the structure in (2.3) is

$$(3.5) \quad \mu(s_i, t) = \exp(X(s_i, t)) + \eta Y(s_i, t - 1), \quad \eta \in (0, 1).$$

In this structure, $X(\cdot)$ and $Y(\cdot)$ are not linearly related and a Gaussian prior for η is clearly not appropriate due to the parameter space constraints.

However, Laplace approximations can still be used by conducting inference on η at the same time inference is conducted on the the set of latent model parameters. In both the SCSE model and the RDSE model, the full conditional for the latent state is

$$(3.6) \quad \pi(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}\mathbf{X}^T \mathbf{Q}(\boldsymbol{\theta})\mathbf{X} + \sum_{s_i, t} \log \pi(Y(s_i, t)|X(s_i, t), \eta, Y(s_i, t - 1))\right).$$

Here we will let $\boldsymbol{\theta} = (\theta_1, \sigma^2, \eta)^T$ and $\mathbf{Q}_{sc}(\boldsymbol{\theta}) = (\mathbf{I}_{sn,sn} - \theta_1 \mathbf{I}_{t,t} \otimes \mathbf{H})$ for the SCSE model and use $\mathbf{Q}_{rd}(\boldsymbol{\theta})$ for the RDSE model defined in (2.8).

While θ in (3.6) does not contain η we next do a Taylor series expansion of $\log \pi(Y(s_i, t)|X(s_i, t), \eta, Y(s_i, t-1))$, as a function of $X(s_i, t)$ and, for each s_i, t , expand the term about a guess for the mode, say $\mu_0(s_i, t)$. First we write $\mathbf{B}^*(\theta|\mu_0)$ as a vector of the same length as $X(s_i, t)$ where each element is given by

$$(3.7) \quad B(s_i, t|\mu_0) = \left(\frac{\partial \log \pi(Y(s_i, t))}{\partial X(s_i, t)} \Big|_{X(s_i, t)=\mu(s_i, t)} - \mu(s_i, t) \frac{\partial^2 \log \pi(Y(s_i, t))}{\partial X(s_i, t)^2} \Big|_{X(s_i, t)=\mu(s_i, t)} \right).$$

Next, we further define $\mathbf{Q}^*(\theta|\mu_0)$ as the updated precision matrix:

$$(3.8) \quad \mathbf{Q}^*(\theta|\mu_0) = \mathbf{Q}(\theta) + \text{diag} \left(-\frac{\partial^2 \log \pi(Y(s_i, t))}{\partial X(s_i, t)^2} \Big|_{X(s_i, t)=\mu(s_i, t)} \right),$$

where $\mathbf{Q}(\theta)$ is either $\mathbf{Q}_{\text{sc}}(\theta)$ or $\mathbf{Q}_{\text{rd}}(\theta)$ depending on the context. Then we can write

$$(3.9) \quad \pi(X|Y, \theta) \propto \exp \left(-\frac{1}{2} \mathbf{X}^T (\mathbf{Q}^*(\theta|\mu_0)) \mathbf{X} + \mathbf{X}^T (\mathbf{B}^*(\theta|\mu_0)) \right).$$

While in (3.8), $\mathbf{Q}(\theta)$, the original precision matrix, does not contain η , $\mathbf{Q}^*(\theta)$, the updated precision matrix, does depend on the self-excitation parameter.

Next, we find the values of $\mu(s_i)$ that maximize (3.9). This is done through the use of an iterative maximization algorithm by solving for μ_1 in $(\mathbf{Q}^*(\theta|\mu_0))\mu_1 = \mathbf{B}^*(\theta|\mu_0)$. For a fixed θ , this converges rapidly, due to the sparsity of both \mathbf{Q}_{sc} and \mathbf{Q}_{rd} .

In (3.3), for a fixed θ , we can then find $x^*(\theta)$. When the denominator of (3.3) is evaluated at $x^*(\theta)$ it becomes $|\mathbf{Q}^*(\theta) \frac{1}{2\pi}|^{1/2}$ which is equivalent to the hyperparameter inference recommended by Lee and Nelder (1996) as pointed out by R. A. Rigby in Rue, Martino and Chopin (2009).

In order to best explore $\pi(\theta|Y)$, the posterior mode is first found through a Newton–Raphson-based method. In order to do this we approximate the Hessian matrix based on finite difference approximation to the second derivatives.

After locating the posterior mode of $\pi(\theta|Y)$, the parameter space can be explored using the exploration strategy laid out in Section 3.1 of Rue, Martino and Chopin (2009).

Now, for the set of diffusion parameters θ , which contain η , we have a method of estimating $\pi(\theta|Y)$. Inference for any further data model covariates can now be conducted in the same manner as done in Rue, Martino and Chopin (2009).

3.2. Model comparison and goodness of fit. In order to conduct model comparison, we will use the deviance information criterion (DIC) originally proposed by Spiegelhalter et al. (2002). Goodness of fit will be conducted through the use of posterior predictive P -values, outlined by Gelman, Meng and Stern (1996).

To approximate the DIC, we first find the effective number of parameters for a given θ . As noted in Rue, Martino and Chopin (2009), we can estimate this by using $n - \text{tr}(\mathbf{Q}(\theta)\mathbf{Q}^*(\theta)^{-1})$ for both the SCSE model and the RDSE model. This gives the effective number parameters for a given θ , which can then be averaged over $\pi(\theta|\mathbf{Y})$ to get the effective number of parameters for the model.

Second, we calculate the deviance of the mean

$$(3.10) \quad -2 \sum_{s_i, t} \log \pi(Y(s_i, t) | \hat{X}(s_i, t), \theta^*),$$

where θ^* is the posterior mode and $\hat{X}(s_i, t)$ is the expectation of the latent state fixing $\theta = \theta^*$. DIC can then be found through deviance of the mean plus two times the effective number of parameters as in Chapter 7 of Gelman et al. (2014) and initially recommended by Spiegelhalter et al. (2002).

In order to assess goodness of fit in analyzing the terrorism data in Section 5, we will use posterior predictive P -values as described by Gelman, Meng and Stern (1996). Here, we pick critical components of the original data set that we wish to see if the fitted model can accurately replicate, for instance the number of zeroes in the data set which we can designate as $T(\mathbf{Y})$. Next, for an index $m = 1, \dots, M$, we then draw a value of θ_m according to $\pi(\theta|\mathbf{Y})$ and simulate a set of observations $Y^*(s_i, t)_m$ of the same dimension as \mathbf{Y} and compute $T(\mathbf{Y}_m^*)$. This process is repeated M times and a posterior predictive P -value is computed as $\frac{1}{M} \sum_{m=1}^M I[T(\mathbf{Y}_m^*) > T(\mathbf{Y})]$ where $I[\cdot]$ is the indicator function. While not a true P -value, both high and low values of the posterior predictive P -value should cause concern over the fitted models ability to replicate features of the original data set.

4. Simulation. In order to validate the Laplace-based methodology for SCSE models, we conducted simulation studies using data on an 8×8 spatial grid assuming a rook neighborhood structure. In order to decrease the edge effect, we wrapped the grid on a torus so each node had four neighbors. For each grid location we simulated 100 observations, creating a spatio-temporal model that had 6400 observations, meaning in (3.3), $\mathbf{Q}(\theta)$ had a dimension of 6400×6400 .

In the first simulation, we used (2.4) fixing the parameters at values that generated data that appeared to resemble the data from Iraq used in Section 5. The generating model we used was

$$(4.1) \quad \begin{aligned} Y(s_i, t) | \mu(s_i, t) &\sim \text{Pois}(\mu(s_i, t)), \\ \mu(s_i, t) &= \exp(-1 + X(s_i, t)) + 0.2Y(s_i, t - 1), \\ X(s_i, t) &= 0.22 \sum_{s_j \in N(s_i)} X(s_j, t) + \varepsilon(s_i, t), \\ \varepsilon(s_i, t) &\sim \text{Gau}(0, 0.4). \end{aligned}$$

The spatial parameter for model was $\theta_1 = 0.22$ which suggests a positive correlation between spatially adjacent locations. An η value of 0.2 would suggest that each event that occurs at one time period increases the expected number of events at the next time period by 0.2. Here we fix σ^2 at 0.4 and use a value of $\beta_0 = -1$ to reflect that, in real world applications, the latent process likely is not zero mean.

Once the data were generated, we found $\pi(\boldsymbol{\theta}|\mathbf{Y})$ by applying (3.3). Here, we note that the numerator of (3.3) is $\pi(\mathbf{X}, \boldsymbol{\theta}, \mathbf{Y}) = \pi(\mathbf{Y}|\mathbf{X}, \eta, \boldsymbol{\theta})\pi(\eta)\pi(\mathbf{X}|\theta_1, \sigma^2)\pi(\theta_1)\pi(\sigma^2)$, which requires a prior specification for η, θ_1 , and σ . In order to reflect an a-priori lack of knowledge, we choose vague priors for all parameters. In this model, we use a half-Cauchy with scale parameter of 25 for σ and a Uniform(ψ_1^{-1}, ψ_n^{-1}) where $\psi_{(i)}$ is the i th largest eigenvector of the spatial neighborhood. As we used a shared-border, or rook, neighbor structure wrapped on a torus, the parameter space is $(-0.25, 0.25)$ as each spatial location has four neighbors. The choice of the half-Cauchy is in line with the recommendations for vague priors for variance components of hierarchical models as outlined in Gelman (2006) and rigorously defended in Polson and Scott (2012). We let the prior for η be Uniform(0, 1).

Using a gradient descent method with step-halving we found the posterior mode of $\pi(\boldsymbol{\theta}|\mathbf{Y})$ to be $\sigma^2 = 0.32$, $\theta_1 = 0.22$, and $\eta = 0.20$. Using the z -based parameterization described in Section 3.1 we next explored the parameterization $\log \pi(\boldsymbol{\theta}|\mathbf{Y})$ and found credible intervals of $\pi(\sigma^2|\mathbf{Y}) = (0.29, 0.36)$, $\pi(\theta_1|\mathbf{Y}) = (0.22, 0.23)$ and $\pi(\eta|\mathbf{Y}) = (0.18, 0.21)$. Fixing $\boldsymbol{\theta}$ at the posterior mode, we then found an approximate 95% credible interval for β_0 to be $(0.07, -1.67)$. If further refinement for β_0 was required, we could proceed to use (3.4). This was not done here as β_0 was not the subject of our primary inference.

The posterior maximum and credible interval for σ^2 appear to be slightly lower than expected, but the remaining parameter credible intervals covered the generating parameter.

Next, we simulated from the RDSE model letting $\beta_0 = 0$, $\alpha = 0.1$, $\kappa = 0.2$, $\sigma^2 = 0.25$, and $\eta = 0.4$.

In fitting the model, we again use vague priors for all of the parameters. Again, we place a half-Cauchy prior on σ^2 as described above. In order to conform to the parameter space of α and κ , we let $\pi(\alpha) \sim \text{Unif}(0, 1)$ and $\pi(\kappa|\alpha) \sim \text{Unif}(-\frac{\alpha}{2}, 1 - \frac{\alpha}{2})$.

Again using the Laplace approximation technique of Section 3, we found the posterior mode of $\pi(\boldsymbol{\theta}|\mathbf{Y})$ to be at $\alpha = 0.085$, $\kappa = 0.19$, $\sigma^2 = 0.21$, $\eta = 0.35$. 95% credible intervals for the posterior marginals were $\alpha \in (0.07, 0.10)$, $\kappa \in (0.14, 0.24)$, $\sigma^2 \in (0.18, 0.24)$, and $\eta \in (0.32, 0.40)$. At the posterior mode of $\boldsymbol{\theta}$, the posterior marginal for β_0 was approximately $(-0.03, 0.01)$. Critically, if there is self-excitement in the data, in all simulations it was differentiable from the latent diffusion. This is a spatial-temporal analogue to the finding in Mohler (2013) where a temporal AR(1) process was differentiable from self-excitement.

In our simulations, the approximations described in this manuscript performed reasonably well for inference on the spatio-temporal diffusion parameters in most cases. However, when σ^2 is large, or when η is large, we have found that the approximations create bias in one or more of the parameters likely due to the high effective number of parameters. However, all approximate likelihood based methods will likely struggle in these cases as well. As noted in Rue, Martino and Chopin (2009), the approximation error in Laplace-based methods is related to the number of effective latent variables over the total sample size.

5. Spatio-temporal diffusion of violence in Iraq (2003–2010).

5.1. *Statistical models and data.* One region where the reasons for the diffusion of terror and crime still remains unclear is in Iraq during 2003 to 2010. While violence undoubtedly spread throughout the country, it remains unclear how or why, spatio-temporally, the spread occurred. Part of the uncertainty is that there still is not agreement over whether violence was due to insurgency, civil war or organized crime. For example, Hoffman (2006) refers to the violence in Iraq as an insurgency, Fearon (2007) argues that the spread of violence was due to a civil war, and Williams (2009) argues that there was a large presence of organized crime in the country.

A few previous studies have examined the presence or absence of self-excitation. In Lewis et al. (2012), the authors concluded that self-excitation was present in select cities in Iraq during this time period. The presence of the self-excitation finding was echoed in Braithwaite and Johnson (2015) where the authors also noted a correlation between locations that shared microscale infrastructure similarities. This would suggest repeat or near repeat actions were causing the increase in violence in a region.

However, in both of these cases, the latent spatio-temporal diffusion was a priori and assumed to be known. In fact, this is likely not the case. In a classic work on the subject, Midlarsky, Crenshaw and Yoshida (1980) discuss how heterogeneity between locations can cause correlation in violence or individuals who cause violence can actually physically move from one location to another. In particular, if violence is strictly due to crime we would expect self-excitement and limited diffusion between geographical regions. Whereas, if violence is due to insurgencies, we would expect more movement of actors as they seek to create widespread disruption in the country. The former theory is reflected in the spatial correlation in the SCSE model, and the latter theory would correspond to the reaction–diffusion component of the second model.

The overarching goal of this analysis, thus, is to determine whether in Iraq the growth of violence in fixed locations was due to the presence or absence of self-excitation. Furthermore, we want to determine whether the latent diffusion of violence is due to the movement of population such as in the reaction–diffusion model,

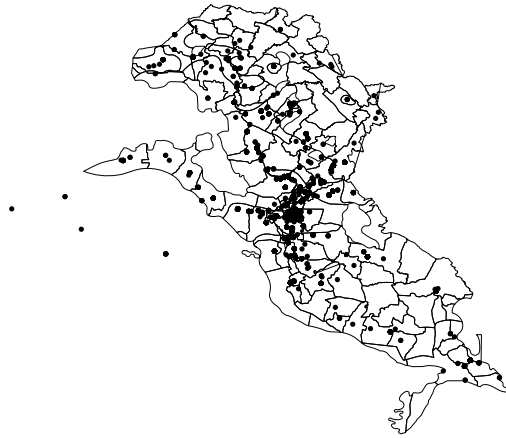


FIG. 2. *Spatial depiction of 6263 events in Iraq.*

or whether there is static spatial correlation. We will answer this while controlling for exogenous factors that may also explain the rise in violence in a region.

In order to address this question as well as demonstrate how Laplace-based approximations can be used to fit real world data to models of the class of (3.1), we used data from the Global Terrorism Database (GTD) introduced in [LaFree and Dugan \(2007\)](#) to examine the competing theories. The GTD defines terrorist events as events that are intentional, entail violence and are perpetrated by sub-national actors. Additionally, the event must be aimed at obtaining a political, social, religious or economic goal and must be conducted in order to coerce or intimidate a larger actor outside of the victim. The majority of lethal events in Iraq from 2003–2010 fit the above category.

The GTD uses a variety of open media sources to capture both spatial and temporal data on terroristic events. The database contains information on what the event was, where it took place, when it took place and what terrorist group was responsible for the event. From 2003 to 2010 in Iraq, the database contains 6263 terrorist events; the spatial structure is shown in Figure 2.

As seen in this map, the majority of the violence is in heavily populated areas such as Baghdad and in the regions north up the Tigris river to Mosul and west through the Euphrates river. In order to model this data, we aggregated the point data to 155 political districts intersected by ethnicities and aggregated the data monthly for 96 months meaning that $\Sigma(\theta)$ in (2.3) is a $14,880 \times 14,880$ matrix. Population for each political district was taken from the Empirical Studies of Conflict Project website, available from <https://esoc.princeton.edu/files/ethnicity-study-ethnic-composition-district-level>.

We considered covariates controlling for the population density within a fixed region as well as for the underlying ethnicity. We will make the simplifying assumption that both of these are static over time. Previous statistical analysis on

terrorism considered macro level covariates, such as democracy in Python et al. (2016) that differ country to country but would not change within a single country as analyzed here. Other studies considered more micro level covariates such as road networks that were found to be statistically related to terrorism in Braithwaite and Johnson (2015). Here we take the viewpoint that the vast majority of the incidents in Iraq were directed against individuals rather than terrorist events directed at fixed locations. Therefore, we would expect a higher population density to provide more targets for a potential terrorist to attack. Furthermore, covariates such as road networks, number of police or number of US soldiers would all be highly collinear with population density. We do, though, consider a covariate for ethnicity in a region. Specifically, we add an indicator if the region is predominately Sunni. The disenfranchisement of the Sunnis and the high level of violence in Sunni dominated areas has been well established; see, for example, Baker III and Hamilton (2006). Previous research in Linke, Witmer and O’Loughlin (2012) focusing on Granger Causality also suggested indicators for majority Sunni/Shia may be appropriate in any analysis of violence in Iraq:

$$\begin{aligned}
 Y(s_i, t) | \mu(s_i, t) &\sim \text{Pois}(\mu(s_i, t)), \\
 \mu(s_i, t) &= \exp[\beta_0 + \beta_1 \log \text{Pop}(s_i) + \beta_2 \text{Sunni}(s_i) + X(s_i, t)] \\
 &\quad + \eta Y(s_i, t - 1), \\
 X &\sim \text{Gau}(\mathbf{0}, \mathbf{Q}^{-1}(\theta)).
 \end{aligned}
 \tag{5.1}$$

The complete statistical model is given in (5.1). We next fit this model letting $\mathbf{Q} = \mathbf{Q}_{\text{sc}}$ and $\mathbf{Q} = \mathbf{Q}_{\text{rd}}$. We further consider fixing $\eta = 0$ to test the presence or absence of self-excitement in the data for both process models.

5.2. *Results.* We fit all four models using the Laplace approximation method described in Section 3.1. For each of the parameters we used vague proper priors to ensure posterior validity. In the SCSE model and the spatially correlated models we used a half-Cauchy with scale parameter of 5 for σ and a $\text{Uniform}(\psi_1^{-1}, \psi_n^{-1})$ where ψ are the eigenvalues of \mathbf{H} . Using the neighborhood structure corresponding to the geographical regions described above, this corresponded to a $\text{Uniform}(-0.3, 0.13)$. For each of the exogenous covariates, we used independent $\text{Gaussian}(0, 1000)$ priors. For the SCSE model we further assumed a $\text{Uniform}(0, 1)$ prior for η .

In fitting the RDSE model, we again used a half-Cauchy with scale parameter of 5 for σ . For the decay parameter, α , we used a $\text{Uniform}(0, 1)$ prior and for the diffusion parameter, κ , we chose a $\text{Uniform}(\frac{-\alpha}{2}, 1 - \frac{\alpha}{2})$ in order to ensure we were in the allowable parameter space.

All four models took approximately 30 minutes to an hour depending on starting values to converge using a Newton–Raphson-based algorithm to find the maximum. Gaussian approximations to the 95% credible intervals for the parameters

TABLE 1
95% Credible Intervals for Model Parameters

	Spatial Correlation Only	SCSEM	Reaction–Diffusion Only	RDSEM
β_0	(−20.2, −18.4)	(−16, −15.5)	(−22.4, −18.0)	(−22, −18.6)
β_1	(1.2, 1.4)	(0.9, 1.1)	(1.1, 1.4)	(1.1, 1.5)
β_2	(1.6, 1.9)	(1.1, 1.3)	(0.10, 0.33)	(0.17, 0.53)
η	–	(0.35, 0.37)	–	(0, 0.04)
σ^2	(1.9, 2.7)	(2.1, 2.5)	(0.20, 0.30)	(0.18, 0.26)
θ_1	(0.08, 0.10)	(0.09, 0.1)	–	–
α	–	–	(0.001, 0.007)	(0.001, 0.007)
κ	–	–	(0.03, 0.07)	(0.03, 0.06)

are given for all four models are shown in Table 1. As can be seen in comparing the SCSE model to the RDSE model, the presence or absence of self-excitation appears to be dependent on the choice of structure of \mathcal{Q} . Furthermore, the impact of majority Sunni is also dependent on whether the reaction–diffusion or spatially correlated model was used.

Using the methodology described in Section 3.2, we next calculated DIC as well as posterior predictive P -values based on the maximum observed value and the number of zeroes in the data set. In the original data set, the maximum number of events observed for all regions and months was 26 and the data set had 13,445 month/district observations that were zero. The model assessment and selection results are shown in Table 2.

Clearly from Table 2 the models with an underlying reaction–diffusion process model outperform those with spatial correlation only. Furthermore, the addition of self-excitation in the model appears to have minimal impact. In particular, without self-excitation, the spatial correlation model tends to under count the number of violent activities while the SCSE model tends to over count. There really is not much difference between the RDSE model and the reaction–diffusion model, so we prefer the simpler reaction–diffusion only model. While β_2 is only minimally

TABLE 2
Model Assessment and Selection Statistics for Iraq Data

Model	DIC	P -Values Maximum Value	P -Value Zeros
Spatially Correlated Model without Self-Excitation	9370	0.02	1
Spatially Correlated Self-Exciting Model	8722	0.97	0
Reaction–Diffusion Only Model	8664	0.53	0.81
Reaction–Diffusion Self-Exciting Model	8699	0.45	0.89

significant in the reaction–diffusion model, the models perform better including the covariate than disregarding it entirely.

5.3. Significance. Under all measures of performance, the reaction–diffusion model, (2.6), without self-excitation outperforms the other models under consideration. This process model as well as values of the covariates and the lack of self-excitement in the data offer several insights into the causes of the spread of violence in Iraq.

Not surprisingly, the reaction–diffusion model has a positive relationship between log population and violence. As the majority of attacks are directed at individuals, it would clearly follow that regions that have higher population will offer more targets as well as more potential combatants. Further, higher populated areas also would have had higher number of Iraqi government officials as well as US military presence. The positive, though small, relationship between Sunni and violence is also not surprising as, in general, predominately Sunni areas were generally more disenfranchised following the transition to a new Iraqi government after the downfall of Saddam Hussein.

More significantly, though, was that the reaction–diffusion model fit the data better than the SCSE model or the RDSE model. This suggests that increases in violence can be attributed, at least in part, to movement between high violence and low violence areas rather than repeat or near repeat actors in a fixed location.

While the κ parameter may appear small in Table 1, it still has an impact on the process. For the sake of simplicity, we can demonstrate this on a three node system. For this system, we consider a central node that has a high level of violence surrounded by two nodes that have a low level of violence. In this setup, we will fix $\beta_0 = -19$, $\beta_1 = 1.3$ and consider each node as having a population of 1000 and let $\kappa \in \{0.03, 0.07\}$. The resultant system over time is shown in Figure 3. Even in this simple system, there is a noticeable increase in violence as the center node diffuses throughout the entire system.

The implications of the reaction–diffusion model being preferable over the SCSE model or the spatial correlation only model can be seen by going back to the original PDE that inspired the model, (2.5). The underlying assumption in that model is that the rate of violence spreading to a region is determined through the levels of violence in neighboring regions. From a military planning perspective, this would suggest that if there is a peaceful region surrounded by areas of high violence, the peaceful region should be isolated to prevent the movement of malicious actors. This strategy would be consistent with published military strategy as outlined in *U.S. Army and Marine Corps (2006)*.

Finally, this offers insight into the nature of the conflict that was fought in Iraq. For instance, *Zhukov (2012)* discusses how insurgencies diffuse throughout a population by either physical movement of actors or through movement of ideas, whereas *Short et al. (2008)* suggest that criminal violence would be expected to have an element of self-excitation. When accounting for the possibility of spatial

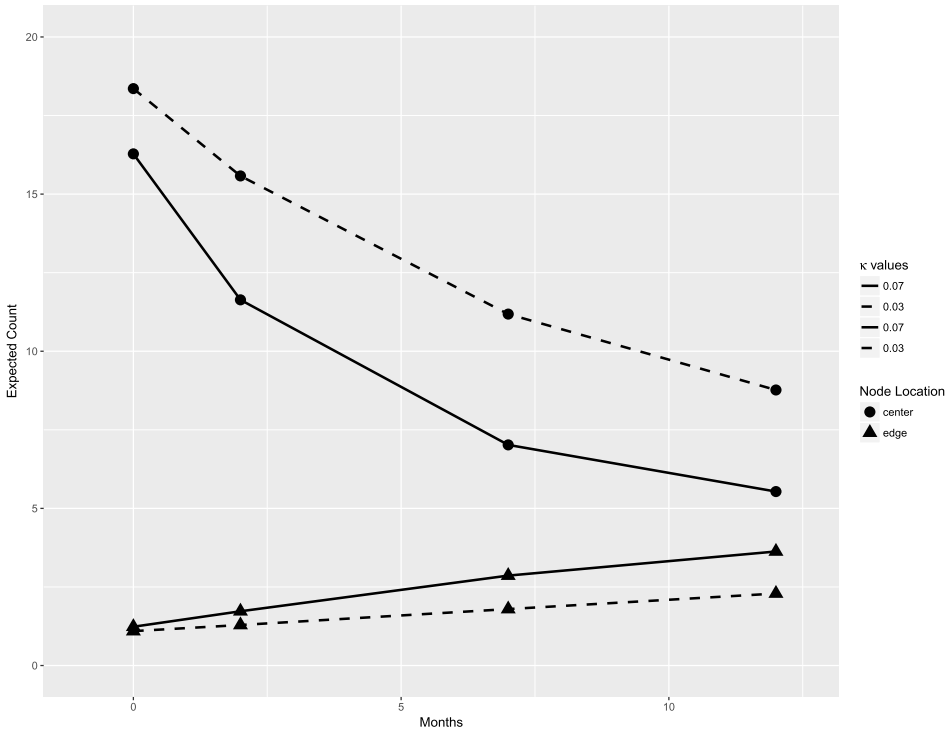


FIG. 3. This plot shows the expected changes in violence in a simple three node system where the center node starts with a high level of violence and the other two nodes start with a low level of violence for $\kappa = 0.03$, depicted as dashed lines in above figure, and $\kappa = 0.07$, depicted as straight lines. As seen here after 12 months for $\kappa = 0.07$ the nodes are essentially at equilibrium.

diffusion, this self-excitation does not appear to be present in the Iraqi data set. Therefore, as a diffusion-based model fits the data better, this would suggest the spread of violence was due to the physical movement of an insurgent population or ideology rather than a criminal element that would be expected to stay more static at a location.

6. Discussion. In this paper, we develop statistical models that allow for spatio-temporal diffusion in the process model and temporal diffusion in the data model. We relate the models to existing theory in how violence diffuses in space and time. We further developed a Laplace approximation for spatio-temporal models that contain self-excitement. This modification allows for a quick and accurate fit to commonly used models in both the analysis of terrorism and criminology. A critical difference between classical INLA and our proposal is that in our proposal, inference is not only performed on the hyperparameters during the exploration of $\pi(\theta|Y)$ in (3.3), but also on the self-excitation parameter η . While η is not generally thought of as a hyperparameter, when the linear expansion of the log

likelihood is done in (3.9), η enters into \mathbf{Q}^* and \mathbf{B}^* in a similar manner as the hyperparameters.

While we only considered two process models and self-excitation that only exists for one time period, the methodology outlined above can easily be extended to allow self-excitation to have an exponential decay similar to the modeling technique of Mohler et al. (2011). As shown above, the absence of testing multiple process models may result in premature conclusions about how violence is spreading over regions. While self-excitation may be present in one model, its significance may be dulled through the use of alternative process models resulting in differing conclusions.

Although self-excitation has become increasingly popular, alternate approaches based on Besag's auto-logistic model, as used in Weidmann and Ward (2010) are possible, though care must be taken if count data is used as Besag's auto-Poisson does not permit positive dependency. As shown in Kaiser and Cressie (1997), a Winsorized Poisson must be used if positive dependency is desired, as it most certainly is in terrorism modeling. In this case, the data model dependency would linearly be associated with the log of $\mu(s_i, t)$.

Though the motivation for the models in this paper was the spatio-temporal spread of violence, the novel concept of combining latent process dependency and data model dependency has the potential to be used in other fields. For example, in the modeling of thunderstorms, self-excitation may be present temporally, while process model dependency may also be appropriate due to small-scale, unobservable, spatial or spatio-temporal dependency. Laplace approximations, as demonstrated in this paper, allow for quick and relatively accurate methods to fit multiple types of self-exciting spatio-temporal models for initial inference.

Acknowledgments. The authors thank the Editor, Thomas Brendan Murphy, the Associate Editor, and a reviewer whose feedback during the review process greatly helped to improve this manuscript.

REFERENCES

- BAKER III, J. A. and HAMILTON, L. H. (2006). *The Iraq Study Group Report*. Vintage Books.
- BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236.
- BRAITHWAITE, A. and JOHNSON, S. D. (2015). The battle for Baghdad: Testing hypotheses about insurgency from risk heterogeneity, repeat victimization, and denial policing approaches. *Terrorism and Political Violence* **27** 112–132.
- CHUNG, F. R. K. (1997). *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics **92**. Amer. Math. Soc., Providence, RI. MR1421568
- CRESSIE, N. and WIKLE, C. K. (2012). *Statistics for Spatio-Temporal Data*. Wiley, New York.
- FEARON, J. D. (2007). Iraq's civil war. *Foreign Aff.* **86** 2.
- GELMAN, A., MENG, X.-L. and STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6** 733–760.
- GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1** 515–534.

- GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2014). *Bayesian Data Analysis 2*. Chapman & Hall/CRC, Boca Raton, FL.
- HOFFMAN, B. (2006). Insurgency and counterinsurgency in Iraq. *Studies in Conflict & Terrorism* **29** 103–121.
- KAISER, M. S. and CRESSIE, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statist. Probab. Lett.* **35** 423–432.
- LAFREE, G. and DUGAN, L. (2007). Introducing the global terrorism database. *Terrorism and Political Violence* **19** 181–204.
- LAUB, P. J., TAIMRE, T. and POLLETT, P. K. (2015). Hawkes Processes. Preprint. Available at [arXiv:1507.02822](https://arxiv.org/abs/1507.02822).
- LEE, Y. and NELDER, J. A. (1996). Hierarchical generalized linear models. *J. Roy. Statist. Soc. Ser. B* **58** 619–678. [MR1410182](https://doi.org/10.2307/2346478)
- LEWIS, E., MOHLER, G., BRANTINGHAM, P. J. and BERTOZZI, A. L. (2012). Self-exciting point process models of civilian deaths in Iraq. *Security Journal* **25** 244–264.
- LINKE, A. M., WITMER, F. D. and O'LOUGHLIN, J. (2012). Space–time granger analysis of the war in Iraq: A study of coalition and insurgent action-reaction. *Int. Interact.* **38** 402–425.
- MIDLARSKY, M. I., CRENSHAW, M. and YOSHIDA, F. (1980). Why violence spreads. *International Studies Quarterly* **24** 262–298.
- MOHLER, G. (2013). Modeling and estimation of multi-source clustering in crime and security data. *Ann. Appl. Stat.* **7** 1525–1539. [MR3127957](https://doi.org/10.1214/12-AAS017)
- MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and TITA, G. E. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106** 100–108. [MR2816705](https://doi.org/10.1198/016214510000000000)
- POLSON, N. G. and SCOTT, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Anal.* **7** 887–902. [MR3000018](https://doi.org/10.1214/11-BA703)
- PORTER, M. D., WHITE, G. (2012). Self-exciting hurdle models for terrorist activity. *Ann. Appl. Stat.* **6** 106–124.
- PYTHON, A., ILLIAN, J., JONES-TODD, C. and BLANGIARDO, M. (2016). A Bayesian Approach to Modelling Fine-Scale Spatial Dynamics of Non-State Terrorism: World Study, 2002–2013. Preprint. Available at [arXiv:1610.01215](https://arxiv.org/abs/1610.01215).
- RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392.
- SHORT, M. B., D'ORSOGNA, M. R., PASOUR, V. B., TITA, G. E., BRANTINGHAM, P. J., BERTOZZI, A. L. and CHAYES, L. B. (2008). A statistical model of criminal behavior. *Math. Models Methods Appl. Sci.* **18** 1249–1267.
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 583–639.
- TENCH, S., FRY, H. and GILL, P. (2016). Spatio-temporal patterns of IED usage by the Provisional Irish Republican Army. *European J. Appl. Math.* **27** 377–402.
- U.S. ARMY AND MARINE CORPS (2006). Counterinsurgency, FM 3-24, MCWP 3-33.5.
- WEIDMANN, N. B. and WARD, M. D. (2010). Predicting conflict in space and time. *Journal of Conflict Resolution* **54** 883–901.
- WILLIAMS, P. (2009). *Criminals, Militias, and Insurgents: Organized Crime in Iraq*. Strategic Studies Institute, Carlisle, PA.
- ZHUKOV, Y. M. (2012). Roads and the diffusion of insurgent violence: The logistics of conflict in Russia's North Caucasus. *Political Geography* **31** 144–156.

DEPARTMENT OF STATISTICS AND STATISTICAL LEARNING
IOWA STATE UNIVERSITY
SNEDECOR HALL
AMES, IOWA 50011-1210
USA
E-MAIL: nclark1@iastate.edu
pdixon@iastate.edu