Modeling and Anomaly Detection for Event Occurrences Following an Inhomogeneous Spatio-Temporal Poisson Process

Waley W. J. Liang, Jacob B. Colvin, Bruno Sansó, and Herbert K. H. Lee University of California, Santa Cruz

February 18, 2010

Abstract

This paper presents a model that describes spatial variations of the intensity of events that occur at random geographical locations. An inhomogeneous Poisson process is used to model the intensity over a spatial region with multiplicative spatial and temporal covariate effects. Dynamic temporal effects are incorporated into the model allowing changes in the intensity structure over time. Additionally, anomaly detection in the event rates is developed based on exceedance probabilities. The methods are demonstrated on data of major crimes in Cincinnati during 2006.

1 Introduction

We consider the problem of inference for time evolving point patterns. We focus on sequences of inhomogeneous Poisson processes with time-varying intensities using a Bayesian approach, proposing a model-based flexible tool and illustrating its use for anomalies detection. As a motivating example we analyze a dataset of major crimes (those involving serious violence against another person) in Cincinnati during 2006. These are depicted by the blue points in Figure 1. We can see that crimes are more concentrated just above the river (where we refer as the downtown area) than they are at other locations. Even at locations where crimes are sparse, small clusters of occurrences are noticeable. Clearly, there is an inhomogeneous underlying structure for the intensity of crimes over Cincinnati and our task is to design and implement a model to describe such intensity. Since each crime occurs at a specific location over the spatial region, the collection of these occurrences can be described by a spatial point process (PP). In general, a spatial PP is a random collection of events occurring in a spatial region where each event is referenced by the location at which it occurs. For example, the arrival of buses on a street, occurrence of mountain fires, and incidence of diseases can all be described by a spatial PP. Typically, the spatial locations are recorded in longitude, latitude, and elevation, though sometimes only one coordinate is of interest.

For concreteness we will we will work with PPs on two-dimensional domains, but the results clearly generalize to other dimensional spaces. Moreover, the crimes shown in Figure 1 occur not only at different locations but also at different times. Hence, we consider a spatial PP indexed by time.

In this paper we focus on inhomogeneous Poisson processes, for which specification of the intensity function is required. The intensity function defines the rate of occurrence for the Poisson process. This is assumed to vary smoothly in space and time. We use an innovative convolution of a time-varying discrete gamma process. To reduce the computational burden we use compactly supported kernels for the convolution. The result is a model that is flexible but computationally tractable, is able to realistically capture the space and time evolution of observed data, and is able to incorporate information from covariates.

In the next section, we give a brief overview of the basics of spatial-temporal point processes, references to some of the existing inferential procedures, and the properties of a Poisson process. In Section 3, we specify the intensity function of the Poisson process, covariates, and priors for the model parameters. In Section 4, extension to the dynamic model and inference based on Forward Filtering Backward Sampling is revealed. Detection of anomalous events based on exceedance probabilities is developed in Section 5. In Section 6, plotting of the results on the crime data and discussions are given.



Figure 1: Major crime locations in Cincinnati in 2006.

2 Inhomogeneous spatio-temporal Poisson process

A spatio-temporal point process N is defined as a random measure over a region $\{S \times \mathcal{T}\}$, where $S \subseteq \mathbb{R}^p$ and $\mathcal{T} \subseteq \mathbb{R}^+$. Each point in the process is referenced by a spatial location $\mathbf{s} \in S$ and time $t \in \mathcal{T}$. In this paper, we focus on developing an inhomogeneous spatio-temporal Poisson process model on a two-dimensional spatial domain. Following the description by Kottas and Sansó (2007), an inhomogeneous spatial Poisson process is a random countable subset Π of $S \subseteq \mathbb{R}^2$. It has the property that for any finite collection A_1, \dots, A_k of pairwise disjoint subsets of S, the random variables $N(A_1), \dots, N(A_k)$, i.e., the number of points of Π lying in each A_j , are independent and each follows a Poisson distribution with mean $\int_{A_j} \lambda(\mathbf{s}) d\mathbf{s}$. Here, $\lambda(\cdot)$ is the intensity function of the inhomogeneous Poisson process. Denote the observation locations by $\mathbf{s}_1, \dots, \mathbf{s}_n \in S$. The likelihood of $\lambda(\cdot)$ is

$$L(\lambda) \propto \exp\{-\Lambda(S)\} \prod_{i=1}^{n} \lambda(\mathbf{s}_i) ,$$

where $\Lambda(S) = \int_{S} \lambda(\mathbf{s}) d\mathbf{s}$. This model can be extended to allow the intensity to change over time. Suppose that the data corresponds to T different time intervals. In each time interval twe have an intensity function $\lambda(\mathbf{s}, t)$. Let t_i be the time that corresponds to the observation at location \mathbf{s}_i and assume that the intensity is constant within each time interval. The likelihood of this extended model is

$$L(\lambda) \propto \prod_{i=1}^{n} \exp\left\{-\int_{\mathcal{S}} \lambda(\mathbf{s}, t_i) d\mathbf{s}\right\} \prod_{i=1}^{n} \lambda(\mathbf{s}_i, t_i) .$$
(1)

Covariate information (spatial and/or temporal) can be incorporated into $\lambda(\mathbf{s}, t)$ and their effects can be estimated along with other parameters of λ based on observed data. Ripley (1981, 1988), Diggle (1983, 1985), Guttorp and Thompson (1990); Guttorp (1995), Stoyan, Kendall and Mecke (1995), Stoyan and Stoyan (1995) give details of spatial point process in which stationarity or homogeneity is assumed and inference based on non-parametric methods is the focus. For likelihood-based, including simulation-based, inference (e.g., MCMC methods), see Møller (1999), Lieshout (2000), Diggle (2003), Møller and Waagepetersen (2003, 2004), Baddeley, Gregori, Mateu, Stoica and Stoyan (2006), and Waagepetersen and Guan (2009). These recent texts often relax the restrictive assumption of stationarity or homogeneity. For spatial-temporal point processes, the papers by Diggle (2005) and by Jensen, Jónsdóttir, Schmiegel and Barndorff-Nielsen (2006) provide introductions to the subject and some likelihood-based inference methods. Kottas and Sansó (2007) propose a Bayesian non-parametric approach to modeling the intensity of an inhomogeneous Poisson process using mixtures of Dirichlet processes (DPM). Taddy (2009) extends the model to consider temporally-varying intensities. The DPM is computationally efficient and offers a lot of flexibility, however it can not be easily adapted to handle covariates without substantial loss of computational efficiency.

3 A nonparametric model

Based on the above Poisson process likelihood, we need to specify the intensity function $\lambda(\mathbf{s}, t)$, which depends only on (\mathbf{s}, t) . In our model, it is decomposed as $\lambda(\mathbf{s}, t) = \tau \lambda(\mathbf{s})\mu(\mathbf{s})\nu(t)$, where τ is a scalar parameter, $\tau \lambda(\mathbf{s})$ is the baseline intensity, $\mu(\mathbf{s})$ is the effect due to spatially varying covariates, and $\nu(t)$ is the effect due to temporal covariates. The decomposition assumes that spatial and temporal effects due to covariates have a multiplicative effect on the baseline intensity of the process.

3.1 Specification of the baseline intensity

A flexible and computationally efficient way of specifying spatial processes is given by process convolutions (Higdon, Swall and Kern, 1999; Higdon, 2001). While these are mostly used in the context of inhomogeneous Gaussian processes, see for example Lemos and Sansó (2009), Lee *et al.* (2008) have used them to model spatial processes with positive values. Following these ideas $\lambda(\mathbf{s})$, is specified

$$\lambda(\mathbf{s}) = \int k(\mathbf{s} - \mathbf{u} | \boldsymbol{\psi}) z(\mathbf{u}) d\mathbf{u},$$

where $z(\cdot)$ is a gamma process and $k(\cdot|\psi)$ is a kernel function with parameter vector ψ . A discrete approximation can be obtained by considering a grid of regularly spaced background points $\mathbf{u}_1, \dots, \mathbf{u}_m \in \mathcal{S}$ and the intensity can be approximated as

$$\lambda(\mathbf{s}) = \sum_{j=1}^{m} k(\mathbf{s} - \mathbf{u}_j | \boldsymbol{\psi}) z(\mathbf{u}_j).$$
(2)

Assuming that $z(\mathbf{u}_i)$ follows a gamma distribution ensures a nonnegative intensity.

The use of a convolution as defined Equation (2) within the likelihood in Equation (1) entails a computational problem. In fact computing the likelihood requires the evaluation of n products of sums. To tackle this problem we assume that the kernel $k(\cdot)$ is a function with compact support. More specifically we define

$$k(\mathbf{s} - \mathbf{u}_j | \boldsymbol{\psi}) = \frac{k_B(\mathbf{s} - \mathbf{u}_j | \boldsymbol{\psi})}{\sum_{j=1}^m k_B(\mathbf{s} - \mathbf{u}_j | \boldsymbol{\psi})},$$

where $k_B(\cdot|\boldsymbol{\psi})$ is a bivariate anisotropic Bézier kernel (Brenning, 2001) function with parameter vector $\boldsymbol{\psi}$. Denote $\mathbf{s} = (s_1, s_2)$, $\mathbf{u} = (u_1, u_2)$, and $\boldsymbol{\psi} = (l_1, l_2, w)$, the form of the kernel is

$$k_B(\mathbf{s} - \mathbf{u} | \boldsymbol{\psi}) = \left[1 - \left(\frac{s_1 - u_1}{l_1} \right)^2 + \left(\frac{s_2 - u_2}{l_2} \right)^2 \right]^w \mathbb{1}_{\left(\sqrt{((s_1 - u_1)/l_1)^2 + ((s_2 - u_2)/l_2)^2} \right) \le 1} \right],$$

where **u** is the center of the kernel, l_1 and l_2 are the ranges of the support in the s_1 and s_2 directions such that the kernel is strictly zero beyond these ranges, $\mathbb{1}_A$ is the indicator

function of the set A, and w is a parameter that governs the smoothness of the resulting process. A Bézier kernel has a compact elliptical support, which ensures that $z(\mathbf{u}_j)$ s are more directly associated to the local crime intensity. Moreover, the effective number of terms in the sum of Equation (2) is substantially smaller than m. For n observations, we let $\boldsymbol{\lambda} = (\lambda(\mathbf{s}_1), \dots, \lambda(\mathbf{s}_n))', \mathbf{z} = (z(\mathbf{u}_1), \dots, z(\mathbf{u}_m))',$ and \mathbf{K} a $n \times m$ matrix such that $\mathbf{K}_{ij} = k(\mathbf{s}_i - \mathbf{u}_j | \boldsymbol{\psi})$. Then we have the conditionally linear representation of Equation (2), $\boldsymbol{\lambda} = \mathbf{K}\mathbf{z}$. Thanks to the use of a Bézier kernel we can use sparse matrix computations in this expression.

3.2 Specification of covariates

Let $\mathbf{X}(\mathbf{s})$ be a vector of spatial covariates, such as topographical features, population density or spatial income distribution, and let $\mathbf{F}(t)$ be a vector of temporal covariates, such as seasonal effects, typically given as

$$\mathbf{F}'(t) = \left(\sin\left(\frac{2\pi t}{T}\right), \cos\left(\frac{2\pi t}{T}\right)\right), \quad t \in \{1, \cdots, T\}.$$

The spatial and temporal effects on the process intensity are given by

$$\mu(\mathbf{s}) = \exp\{\mathbf{X}'(\mathbf{s})\boldsymbol{\beta}\}, \quad \boldsymbol{\beta} \in \mathbb{R}^p$$

$$\nu(t) = \exp\{\mathbf{F}'(t)\boldsymbol{\phi}\}, \quad \boldsymbol{\phi} \in \mathbb{R}^q,$$

where p and q denote the number of spatial and temporal covariates, respectively. The full likelihood for $\lambda(\mathbf{s}, t)$ can be written as

$$L(\lambda) \propto \exp\{-\tau \Lambda_T \Lambda_0\} \tau^n \prod_{i=1}^n \mu(\mathbf{s}_i) \nu(t_i) \sum_{j=1}^m k(\mathbf{s}_i - \mathbf{u}_j | \boldsymbol{\psi}) z(\mathbf{u}_j)$$

where

$$\Lambda_T = \sum_{t=1}^T \exp\{\mathbf{F}'(t)\phi\}$$

$$\Lambda_0 = \sum_{j=1}^m z(\mathbf{u}_j) \int_S k(\mathbf{s} - \mathbf{u}_j | \boldsymbol{\psi}) \exp\{\mathbf{X}'(\mathbf{s})\boldsymbol{\beta}\} d\mathbf{s}$$

$$\approx \sum_{j=1}^m z(\mathbf{u}_j) \sum_{\mathbf{s} \in S} k(\mathbf{s} - \mathbf{u}_j | \boldsymbol{\psi}) \exp\{\mathbf{X}'(\mathbf{s})\boldsymbol{\beta}\} \Delta \mathbf{s}$$

The above approximation follows from the idea that the observation window can be discretized and the intensity within each grid cell is assumed to be constant. The area of a grid cell is denoted by Δs .

3.3 Bayesian inference

We follow a Bayesian approach to estimate the parameters of the model. To proceed, the following priors are imposed on the parameters: $z(\mathbf{u}_j) \sim \Gamma(A_z, B_z), P(\tau) \propto \frac{1}{\tau}, \beta_v \sim N(0, \sigma_{\beta_v}^2)$ and $\phi_l \sim N(0, \sigma_{\phi_l}^2)$, where $A_z, B_z, \sigma_{\beta_v}$, and σ_{ϕ_l} are constants. The resulting posterior distributions are given in the Appendix. Markov Chain Monte Carlo (MCMC) is used to explore the posterior distributions of these parameters. Specifically, Gibbs sampling is used to obtain samples of τ since its posterior distribution can be obtained in closed form. The Metropolis-Hastings algorithm is used to obtain posterior samples of $z(\mathbf{u}_j), \beta_v$, and ϕ_l with the following proposal distributions:

$$z^{*}(\mathbf{u}_{j}) \sim \frac{z^{h-1}(\mathbf{u}_{j})}{b_{z}} Beta\left(\frac{a_{z}b_{z}}{2}, \frac{a_{z}(1-b_{z})}{2}\right),$$

$$\beta_{v}^{*} \sim N(\beta_{v}^{h-1}, \sigma_{\beta_{v}}^{2}),$$

$$\phi_{l}^{*} \sim N(\phi_{l}^{h-1}, \sigma_{\phi_{l}}^{2}),$$

where a_z , b_z , $\sigma_{\beta_v^*}$, $\sigma_{\phi_l^*}$ are constants, and h denotes the MCMC iteration index. The multiplicative beta proposal is implemented to improve the acceptance rate of $z(\mathbf{u}_j)$ and it seems to induce a fast convergence of the MCMC.

4 A Dynamic approach

4.1 Model specification

The model specified in Section 3 incorporates a multiplicative temporal effect on the intensity of the Poisson process using the function $\mathbf{F}(t)$. This implies that temporal covariates can only re-scale the baseline intensity, but not its shape. In other words, the shape of the intensity at any time strongly depends on the entire data set, with little influence of events occurring at the time being considered. Changes in the shape of the space-time intensity due to time can only be captured using a non-separable model. To this end we modify the baseline intensity by making the convolved gamma process $z(\mathbf{u}_j)$ time dependent. Given that the convolving kernel has compact support, we expect that the time evolution of the latent process will be determined mostly by the occurrences of events close to the nodes. To obtain a simple yet flexible model for the time evolution of $z_t(\mathbf{u}_j)$ we use the methods of West and Harrison (1997) and fit a first-order polynomial DLM for the log of $z_t(\mathbf{u}_j)$. This will account for smooth time variations in the shape of the intensity. Let $y_t(\mathbf{u}_j) = \log(z_t(\mathbf{u}_j))$.

$$egin{array}{rcl} \mathbf{Y}_t &=& oldsymbol{ heta}_t + \mathbf{v}_t, \ oldsymbol{ heta}_t &=& oldsymbol{ heta}_{t-1} + \mathbf{w}_t \ , \end{array}$$

where

$$\begin{aligned} \mathbf{Y}_t &= (y_t(\mathbf{u}_1), \cdots, y_t(\mathbf{u}_m))', \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{V}_t), \\ \boldsymbol{\theta}_t &= (\theta_t(\mathbf{u}_1), \cdots, \theta_t(\mathbf{u}_m))', \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t) , \end{aligned}$$

 \mathbf{v}_t and \mathbf{w}_t are the observational error and evolution error, respectively. More elaborate dynamics that include, for example, seasonalities or additional covariates, can be accounted for by using observational and evolution matrices other than the identity. For the observation error variance we take $\mathbf{V}_t = \sigma^2 \mathbf{I}_m$. For the evolution error variance we use a discount factor approach (Section 6.3, West and Harrison (1997)). A discount factor is a parameter, $\delta \in (0, 1]$ that determines the amount of information loss between consecutive time steps in the evolution equation.

4.2 Model fitting

The model specified in the previous section is fitted using an elaboration of the Monte Carlo method developed for the model of Section 3. The full conditionals for the convolution parameters are modified by noting that

$$p(z_t(\mathbf{u}_j)) = \frac{1}{z_t(\mathbf{u}_j)} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (\log(z_t(\mathbf{u}_j)) - \theta_t(\mathbf{u}_j))^2\right),$$

which leads to the expression

$$\pi(z_t(\mathbf{u}_j)|\cdots) \propto \exp\{-\tau\Lambda\} \prod_{t=1}^T \prod_{i=1}^m \sum_{j=1}^m k(\mathbf{s}_i - \mathbf{u}_j|\boldsymbol{\psi}) z_t(\mathbf{u}_j) p(z_t(\mathbf{u}_j)) ,$$

where

$$\Lambda = \sum_{t=1}^{T} \exp\{\mathbf{F}'(t)\boldsymbol{\phi}\} \sum_{j=1}^{m} z_t(\mathbf{u}_j) \sum_{\mathbf{s}\in\mathcal{S}} k(\mathbf{s}-\mathbf{u}_j|\boldsymbol{\psi}) \exp\{\mathbf{X}'(\mathbf{s})\boldsymbol{\beta}\}\Delta\mathbf{s} .$$

Following is a rough sketch of the MCMC procedure at each iteration.

- 1. Sample $z_t(\mathbf{u}_j)$ from (20) via the Metropolis-Hasting algorithm.
- 2. Based on the most recently sampled $z_t(\mathbf{u}_j)s$, compute \mathbf{Y}_t and sample $\boldsymbol{\theta}_t$ using the Forward Filtering Backward Sampling algorithm.
- 3. Sample all other parameters based on the most recently sampled $\boldsymbol{\theta}_t$ and treating $\exp(\boldsymbol{\theta}_t)$ as a sample of the underlying discretized gamma process at time t.
- 4. Repeat steps 1 through 3 for the next MCMC iteration.

The proposed dynamic model can be used for predictions of the intensity at any time given data up to the previous time. For example, if we have data from January to November, we can predict the intensity for December using the one-step ahead forecast distribution. For details, refer to Chapter 4 of West and Harrison (1997).

5 Anomaly Detection

The proposed model can be used to detect events which are anomalous relative to either its spatial or temporal structure. Following the general approach of Cressie (1991) for examining residuals, we impose a grid over the spatial domain. We consider the number of events in a grid cell in a particular month to be anomalous if it is much larger than would be expected under the Poisson process model. This expected value could either be from data including the month in question (i.e., finding anomalous residuals) or from predictions (checking if new data match predicted values). We typically focus on one-sided tests because of the non-negative character of counts. We compute a measure of anomaly via exceedance probabilities (EP) based on the Poisson model. A one sided EP is calculated as

$$P = Pr(N_t(\mathbf{s}) > \lambda_t(\mathbf{s}))$$

where $N_t(\mathbf{s})$ is the observed crime count over region \mathbf{s} at time t and $\lambda_t(\mathbf{s})$ is the expected number of crimes at time t over region \mathbf{s} as fitted or predicted by the poisson process model. A two-sided EP can be defined analogously. $N_t(\mathbf{s})$ is identified as anomalous if $P < \alpha$, where α is a pre-specified significance level. Typical values of α include 0.01, 0.05, and 0.1. We define two types of anomalies as follows.

• Spatial Anomaly: A spatial anomaly occurs when the number of observations is significantly larger than the expected number of counts in a grid cell for all times. This is caused by changes in the intensity function that are not properly modeled by the kernel grid, where the kernel smoothing estimate of the crime intensity function is unable to account for the nonlinear changes in the underlying crime intensity function. This is an indication that the assumed spatial continuity is not present in the observed data, i.e., a particular grid cell is very unlike its neighbors. Spatial EP's are calculated as

$$Pr(N(\mathbf{s}) > \lambda(\mathbf{s})) = \sum_{\lambda} Pr(N(\mathbf{s}) > \lambda(\mathbf{s})|\lambda) \times Pr(\lambda) .$$

• **Temporal Anomaly**: A temporal anomaly occurs when the observed event count in a grid cell is significantly larger than the expected number at a particular time while other time intervals show no significant divergence from model assumptions. Temporal EP's are calculated as

$$Pr(N_t(\mathbf{s}) > \lambda_t(\mathbf{s})) = \sum_{\lambda_t} \Pr(N_t(\mathbf{s}) > \lambda_t(\mathbf{s}) | \lambda_t) \times Pr(\lambda_t) .$$

6 Example

We now apply the models to the crime data shown in Figure 1. This dataset is publicly available at http://www.cincinnati-oh.gov/police/pages/-4258-/. We discretize the observation window S using a 20 × 20 grid. The center of each grid cell represents the coordinate

of the grid cell and also serves as a background location \mathbf{u} of the gamma process. One additional row of \mathbf{u} is placed outside the observation window on each side to eliminate edge effects. The covariates being considered are the population density and vacancy rates over the observation window. The spatial covariate vector is constructed as

$$\mathbf{X}'(\mathbf{s}) = (x_1(\mathbf{s}), x_2(\mathbf{s})) , \qquad (3)$$

where $x_1(\mathbf{s})$ represents population density and $x_2(\mathbf{s})$ represents vacancy rate at location \mathbf{s} . Vacancy rates can be used as a proxy for the economic health of the neighborhood. The given covariate values vary in different parts of \mathcal{S} and they have been pre-processed so that each of the covariates has a fixed value within each grid cell. As we are interested in monthly effects, the temporal covariate vector is constructed as

$$\mathbf{F}'(t) = \left(\sin\left(\frac{2\pi t}{12}\right), \cos\left(\frac{2\pi t}{12}\right)\right), \quad t \in \{1, \cdots, 12\}.$$
(4)

Using the priors defined in Section 3.3, the parameter values are chosen as following. The standard deviation (SD) of the spatial effect prior, σ_{β_v} , is chosen to be 0.05, for both population density and vacancy rate effects. The posterior mean of the population density effect is estimated to be 4.5%, and that of the vacancy rate effect is estimated to be 6.8%, which seem fairly reasonable. The SD of the multiplicative temporal effect prior, σ_{ϕ_l} , is chosen to be 0.5 and 4 for the sine and cosine terms, respectively. This ensures that this temporal effect is higher in the middle of the year (when it is warmer and more people are out) than at the beginning and end of year, and when averaged, is around 1. The prior shape and rate parameters, A_z and B_z of $z(\mathbf{u})$, are chosen to be 75 and 13000, respectively, to ensure a reasonable posterior mean (around 320) for τ under the assumption that τ represents the monthly baseline counts over the entire spatial domain, and $\tau \lambda(\mathbf{s}) \mu(\mathbf{s})$ represents the proportion of τ that corresponds to grid s. For the dynamic model, the discount factor δ and the observational SD σ are chosen to be 0.95 and $\sqrt{0.5}$, respectively. They should ideally be treated as unknown parameters and estimated via MCMC. But after trying different values of δ within [0.9, 1] (this is the range of values in which a discount factor almost always takes), model results stayed about the same. So, for the purpose of a simpler model, we decided to use the median value, 0.95. The value of σ is chosen to retain the same general structure in the intensity across all months while allowing some shape changes without significantly affecting the convergence of the spatial and temporal effects parameters. The proposal distribution parameters are chosen as $a_z = 2$, $b_z = 0.5$, $\sigma_{\beta_1^*} = \sigma_{\beta_2^*} = 0.1$, $\sigma_{\phi_1^*} = 0.07$, and $\sigma_{\phi_{\alpha}^*} = 4$, to ensure good mixing of the posterior samples. The following sections summarize the results obtained from the nonparametric model and the dynamic model.

6.1 Results of the nonparametric model

In Figure 2, the posterior samples of τ are shown in the first row. It has a posterior mean of approximately 320, which is close to the observed average number of crimes per month. The second row displays histograms for the population density and vacancy rate effects. Based

on these histograms, the interpretation is that on average, every increase of 1000 people per square mile would raise the intensity by 4.5%, and every increase of 10% in the vacancy rate would raise the intensity by 6.8%. The third row shows the monthly effect, which suggests that there are more crimes during the middle of the year than near the start and end of the year. Figure 3 displays the baseline intensity, the baseline intensity multiplied by spatial effects (population density and vacancy rates), and the estimated intensity for June. Notice that their shapes look roughly the same, with the difference being in the scale.

6.2 Results of the dynamic model

Under the dynamic model using data from January to December, the posterior mean of τ decreases from its previous value of 320 (under the nonparametric model) to 1.40, and the spatial effects also decrease slightly compared to their previous values. This is because some of the mean and spatial effects that were described by these parameters are now being captured by the $z_t(\mathbf{u}_i)s$. On the other hand, the multiplicative temporal effect stave about the same. The estimated intensities of February, June and December are shown in Figure 4. The shapes of these intensities are quite noticeably distinct, however, they still have some resemblance, due to the fact that the number and location of crimes are fairly similar across the months. In Figure 5, we show the results of dynamic fitting. For each month, we show the fit using only the data from January up through that month, but not the data from later in the year, as would be done in practice (fitting the model based on the current data, not waiting until the end of the year to fit everything). In the figure, the posterior means of $\{z_t(\mathbf{u}_i), i = 73, 77, 145, 188\}$ are plotted v.s. month in blue, and the number of observations within the support of the Bézier kernel centered at the corresponding \mathbf{u}_i s are included in red. These $z_t(\mathbf{u}_i)$ s are picked arbitrarily for illustration, but are representative of the spatial variability. Notice that $z_t(\mathbf{u}_{73}), z_t(\mathbf{u}_{145})$, and $z_t(\mathbf{u}_{188})$ are learned very quickly, and each roughly follows the trend of the number of observation from January to December, while it takes about 5 months for $z_t(\mathbf{u}_{77})$ to learn the trend. Nevertheless, the behaviors of the picked $z_t(\mathbf{u}_i)$ s conform to our initial assumption that they depend on the number of observations around them, and are good indications that our dynamic model is working properly. Figure 6 displays the estimated intensities for February and June, and a predicted intensity for December under the dynamic model using data from January to November.

6.3 Anomaly detection results

Anomaly detection is applied the the dynamic model using data from January to December. The left panel of Figure 7 shows an example of a spatial anomaly, where the observed crime counts are nearly all above the upper level of the 95% interval of expected counts. This grid cell is cell number 30, which is in the lower middle part of the downtown area (the place where the intensity has the highest peak). The middle panel shows an example of a temporal anomaly, where all of the observed crime counts are within or near the boundary of the 95% interval of expected counts except for July whose count is way above the upper level of the interval. This grid cell is cell number 109, which is in the upper left part of the downtown



Factor for 1000 people/sqmi increase Factor for 10% vacancy rate increase 600 800 500 600 400 Frequency Frequency 300 400 200 200 100 0 0 1.036 1.040 1.048 1.06 1.044 1.05 1.07 1.08 exp(1000*beta1) exp(0.1*beta2)

Seansonal Model Component(month of the year)



Figure 2: Results of the nonparametric model. First row: posterior samples of τ . Second row: spatial effect (left - population density; right - vacancy rate). Third row: monthly effect (90% credible interval is indicated by the black dash lines)



Estiamted baseline * spatial effect



Estiamted baseline * spatial effect * time effect for June



Figure 3: Intensities estimated from the nonparametric model. Top row: Baseline intensity - $\tau\lambda(\mathbf{s})$. Middle row: Baseline × spatial effects - $\tau\lambda(\mathbf{s})\mu(\mathbf{s})$. Bottom row: Intensity for June - $\tau\lambda(\mathbf{s})\mu(\mathbf{s})\nu(6)$.

Estiamted baseline * spatial effect * time effect for Febuary



Estiamted baseline * spatial effect * time effect for June



Estiamted baseline * spatial effect * time effect for December



Figure 4: Intensities of February (top), June (middle), and December (bottom) estimated from the dynamic model using data from January to December.



Figure 5: Results of the dynamic model, fitting each month using data from January through that month. Blue: $\{z_t(\mathbf{u}_j), j = 73, 77, 145, 188\}$ is plotted separately v.s. time t (month). Red: number of observations within the support of the Bézier kernel centered at $z_t(\mathbf{u}_i)$.

area. The right panel shows a histogram of the EP for all of the grid cells over all months. Note that most values are above the significance level of 0.01, 0.05, and even 0.1, indicating that most data are consistent with the model, and then there is a secondary mode near zero for the locations and times that are anomalous. The left panel of Figure 8 provides an image plot of which areas show the most spatially anomalous behavior, consistently over time, under the dynamic model using data from January through December. The color of a grid cell represents the EP level such that red, brown, and yellow correspond to EP values within [0, 0.01), [0.01, 0.05), and [0.05, 0.1), respectively. The middle panel shows anomalies for the month of December under the dynamic model using data from January through December (anomalous residuals). The right panel shows the December EP when December is predicted under the dynamic model (observations not matching predictions).

7 Conclusions

In this paper, we develop a methodology for modeling the intensity of events that occur at random geographical locations. Specifically, we use a Poisson process model with an intensity function constructed by convolution of a gamma process with a bivariate Bézier kernel and allow dependence on spatial and temporal covariates. A dynamic temporal effect is incorporated into the model to allow changes not only in the height but also in the shape of the intensity over time. A Bayesian approach is used to estimate the parameters of the model. Additionally, we illustrate how to detect spatial and/or temporal anomalies via



Estiamted baseline * spatial effect * time effect for June



Predicted baseline * spatial effect * time effect for December



Figure 6: Estimated intensities of February (top), June (middle), and predicted intensity of December (bottom) under the dynamic model using data from January to November.



Figure 7: Left and middle panels: Plots of observed crime counts (green dots), expected crime counts (blue curve), and 2.5% and 97.5% percentiles (red curves) v.s. time (month) for grid cell 30 and 109. Right panel: Histogram of EPs for all grid cells over all months



Figure 8: Left panel: an image plot of spatial EP under the dynamic model using the January - December data. Middle panel: an image plot of temporal EP for December under the dynamic model using the January - December data. Right panel: an image plot of predicted temporal EP for December under the dynamic model using the January - November data. Red, brown, and yellow correspond to EP values within [0,0.01), [0.01,0.05), and [0.05,0.1), respectively.

exceedance probabilities.

While we applied the methods herein on a crime dataset, the methodology is more widely applicable to spatio-temporal point processes. Part of the initial motivation was to develop methods for analysis and prediction of improvised explosive device attacks in Iraq, another spatio-temporal point process (although that data is highly classified). In that setting, there are also patterns over space and time, with important covariate information.

Acknowledgments

We thank Luis Acevedo-Arreguin and Matt Taddy for their help on initial exploratory work on this project, and William Hanley and Gardar Johannesson of Lawrence Livermore National Laboratory for connecting us with this project. This project was partially funded by subcontracts B565602 and B573750 with Lawrence Livermore National Laboratory.

Appendix - Full conditional distributions

$$\pi(z(\mathbf{u}_{j})|\cdots) \propto z(\mathbf{u}_{j})^{A_{z}-1} \exp\{-(\tau \Lambda_{T} \Lambda_{0} + B_{z} z(\mathbf{u}_{j}))\} \prod_{i=1}^{n} \sum_{r=1}^{m} k(\mathbf{s}_{i} - \mathbf{u}_{r} | \boldsymbol{\psi}) z(\mathbf{u}_{r})$$

$$\pi(\tau|\cdots) \propto \tau^{n-1} \exp\{-\tau (\Lambda_{T} \Lambda_{0})\}$$

$$\pi(\beta_{v}|\cdots) \propto \exp\{-\tau \Lambda_{T} \Lambda_{0} + \beta_{v} \sum_{i=1}^{n} \mathbf{X}_{v}'(\mathbf{s}_{i}) - \beta_{v}^{2}/(2\sigma_{\beta_{v}}^{2})\}$$

$$\pi(\phi_{l}|\cdots) \propto \exp\{-\tau \Lambda_{T} \Lambda_{0} + \phi_{l} \sum_{i=1}^{n} \mathbf{F}_{l}'(t_{i}) - \phi_{l}^{2}/(2\sigma_{\phi_{l}}^{2})\},$$

where $\mathbf{X}'_{v}(\mathbf{s}_{i})$ denotes the v^{th} component of $\mathbf{X}'(\mathbf{s}_{i})$ and $\mathbf{F}'_{l}(t_{i})$ denotes the l^{th} component of $\mathbf{F}'(t_{i})$.

References

- Baddeley, A., Gregori, P., Mateu, J., Stoica, R. and Stoyan, D. (2006) Case studies in spatial point process modeling. In Springer Lecture Notes in Statistics 185. New York: Springer-Verlag.
- Brenning, A. (2001) Geostatistics without stationarity assumptions within geographical information systems. *Freiberg Online Geoscience*, 6, 1–108.
- Cressie, N. (1991) Statistics for Sspatial Data. John Wiley and Sons, Inc.
- Diggle, P. (1985) A kernel method for smoothing point process data. *Applied Statisistics*, **34**, 138–147.

Diggle, P. J. (1983) Statistical analysis of spatial point patterns. London: Academic Press.

- Diggle, P. J. (2003) Statistical analysis of spatial point patterns, 2nd edition. London: Arnold.
- Diggle, P. J. (2005) Spatio-temporal point processes: Methods and applications. In Statistical Methods of Spatio-Temporal Systems (eds. B. Finkenstadt and V. Isham), 146. Boca Raton: Chapman & Hall/CRC.
- Guttorp, P. (1995) Stochastic Modeling of Scientific Data. London: Chapman & Hall/CRC.
- Guttorp, P. and Thompson, M. (1990) Nonparametric estimation of intensities for sampled counting processes. *Journal of the Royal Statistical Society B*, **52**, 157–173.
- Higdon, D. (2001) Space and space-time modeling using process convolutions. Tech. Rep. 2001-03, Institute of Statistics and Decision Sciences, Duke University, Durham, NC.
- Higdon, D. M., Swall, J. and Kern, J. C. (1999) Non-stationary spatial modeling. In *Bayesian Statistics 6* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), 761–768. Oxford University Press.
- Jensen, E., Jónsdóttir, K., Schmiegel, J. and Barndorff-Nielsen, O. (2006) Spatio-temporal modelling - with a view to biological growth. In *Statistical Methods of Spatio-Temporal Systems* (eds. B. Finkenstadt and V. Isham), 4776. Boca Raton: Chapman & Hall/CRC.
- Kottas, A. and Sansó, B. (2007) A Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning* and Inference, 137, 3151–3163.
- Lee, H. K., Sansó, B., Zhou, W. and Higdon, D. (2008) Inference for a proton accelerator using convolution models. *Journal of the American Statistical Association*, **103**, 604–613.
- Lemos, R. T. and Sansó, B. (2009) Spatio-temporal model for mean, anomaly and trend fields of north atlantic sea surface temperature. *Journal of the American Statistical Association*, 104, 5–18.
- Lieshout, M. N. M. V. (2000) Markov point processes and their applications. London: Imperial College Press.
- Møller, J. (1999) Markov chain monte carlo and spatial point processes. In Stochastic geometry: likelihood and computation, Monographs on Statistics and Applied Probability 80 (eds. W. S. K. O. E. Barndorff-Nielsen and M. N. M. van Lieshout), 141172. Boca Raton: Chapman & Hall/CRC.
- Møller, J. and Waagepetersen, R. P. (2003) An introduction to simulation-based inference for spatial point processes. In Spatial statistics and computational methods, Springer Lecture Notes in Statistics 173 (ed. J. Møller), 143198. New York: Springer-Verlag.

- Møller, J. and Waagepetersen, R. P. (2004) Statistical Inference and Simulation for Spatial Point Processes. Boca Raton: Chapman & Hall/CRC.
- Ripley, B. D. (1981) Spatial statistics. New York: Wiley.
- Ripley, B. D. (1988) *Statistical inference for spatial processes*. Cambridge: Cambridge University Press.
- Stoyan, D., Kendall, W. S. and Mecke, J. (1995) Stochastic geometry and its applications, 2nd edition. Chichester: Wiley.
- Stoyan, D. and Stoyan, H. (1995) Fractals, random shapes and point fields. Chichester: Wiley.
- Taddy, M. (2009) An auto-regressive mixture model for dynamic spatial poisson processes: Application to tracking the intensity of violent crime. URL http://faculty.chicagobooth.edu/matt.taddy/research/crimepoints.pdf.
- Waagepetersen, R. and Guan, Y. (2009) Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society B*, **71**, 685–702.
- West, M. and Harrison, J. (1997) Bayesian Forecasting and Dynamic Models. Springer.