The Annals of Applied Statistics
2017, Vol. 11, No. 2, 481–503
DOI: 10.1214/16-AOAS960
© Institute of Mathematical Statistics, 2017

SPACE AND CIRCULAR TIME LOG GAUSSIAN COX PROCESSES WITH APPLICATION TO CRIME EVENT DATA

BY SHINICHIRO SHIROTA[1] AND ALAN E. GELFAND

Duke University

We view the locations and times of a collection of crime events as a space–time point pattern. Then, with either a nonhomogeneous Poisson process or with a more general Cox process, we need to specify a space–time intensity. For the latter, we need a *random* intensity which we model as a realization of a spatio-temporal log Gaussian process. Importantly, we view time as circular not linear, necessitating valid separable and nonseparable covariance functions over a bounded spatial region crossed with circular time. In addition, crimes are classified by crime type. Furthermore, each crime event is recorded by day of the year, which we convert to day of the week marks.

The contribution here is to develop models to accommodate such data. Our specifications take the form of hierarchical models which we fit within a Bayesian framework. In this regard, we consider model comparison between the nonhomogeneous Poisson process and the log Gaussian Cox process. We also compare separable vs. nonseparable covariance specifications.

Our motivating dataset is a collection of crime events for the city of San Francisco during the year 2012. We have location, hour, day of the year, and crime type for each event. We investigate models to enhance our understanding of the set of incidences.

1. Introduction. The times of crime events can be viewed as circular data; that is, working at the scale of a day, we can imagine event times as wrapped around a circle of circumference 24 hours [which, without loss of generality, can be rescaled to $[0, 2\pi)$]. Furthermore, over a specified number of days, we can view the set of event times, consisting of a random number of crimes, as a point pattern on the circle. Suppose, additionally, that we attach to each crime event its spatial location over a bounded domain. Then, for a bounded spatial region, we have a space–time point pattern over this domain, again with time being circular.

The contribution here is to develop suitable models for such data, motivated by a set of crime events for the city of San Francisco in 2012. The challenges we address involve the following: (i) clustering in time—event times are not uniformly distributed over the 24 hour circle; (ii) spatial structure—evidently, some parts of the city have higher incidence of crime events than others; (iii) crime type—characterization of point pattern varies with type of crime so that different models

Received September 2015; revised July 2016.

[1]Supported in part by the Nakajima Foundation.

Key words and phrases. Derived covariates, hierarchical model, marked point pattern, Markov chain Monte Carlo, separable and nonseparable covariance functions, wrapped circular variables.

are needed for different crime types; (iv) incorporating covariate information—we anticipate that introducing suitable *constructed* spatial and temporal covariates will help to explain the observed point patterns; (v) the need for spatio-temporal random effects—the constructed spatial and temporal covariates will not adequately explain the space–time point patterns; (vi) the availability of marks—in addition to a location and a time within the day, each event has an associated day of the year which we convert to a day of the week. We propose a range of point pattern models to address these issues; fortunately, our motivating dataset is rich enough to investigate them.

We focus on the problem of building a log Gaussian Cox process (LGCP) which includes, as a special case, a nonhomogeneous Poisson process (NHPP), over space and circular time. We need to build a suitable intensity surface which is driven by a realization of a log Gaussian process incorporating a valid covariance function over space and time.

An initial model for a set of points in space and circular time is a nonhomogeneous Poisson process (NHPP) with an intensity $\lambda(\boldsymbol{s}, t)$ over, say, $D \times S^1$, where D is the spatial region of interest and time lies on the unit circle, S^1. We illuminate such intensities below, but we also note that an NHPP will not prove rich enough for our data, and so we propose a space by circular time log Gaussian Cox process (LGCP). This leads to consideration of space–time dependence. Does time of day affect the spatial pattern of crime? Does location affect the point pattern of event times? Hence, we consider both separable and nonseparable models. We develop a parametric nonseparable space by circular time correlation function building on Gneiting's specification [see Gneiting (2002)]. We note very recent work from Porcu, Bevilacqua and Genton (2016) which presents valid covariance functions on R^1 crossed with spheres.

Typically, time is modeled linearly, leading to a large literature on point patterns over bounded time intervals [see, e.g., Daley and Vere-Jones (2003, 2008)]. Adding space, Brix and Diggle (2001) offer development of a space–time LGCP. Rodrigues and Diggle (2012) consider a space–time process convolution model for modeling of space–time crime events. Liang et al. (2014) consider process convolution for space with a dynamic model for time. Taddy (2010) proposes a Bayesian semiparametric framework for modeling correlated time series of marked spatial Poisson processes with application to tracking intensity of violent crime.

In fact, in this context, it is important to articulate the difference between viewing time in a *linear* manner vs. a *circular* manner. With linear time there is a past and a future. We can condition on the past and predict the future, and we can incorporate seasonality and trend in time. With circular time, as with angular data in general, we only obtain a value once we supply an orientation, for example, the customary midnight with time, although, below, we argue to start the day at 02:00. Then we have no temporal ordering of our crime events except within a defined 24 hour window. We are only interested in modeling the intensity over space and

circular time. For us, prediction would consider the number of events of a particular crime type, in a specified neighborhood, over a window of hours during the day, adjusting for day of the week. For a decision maker, the value would be to facilitate making daily spatial staffing decisions during 24 hour cycles. We do not assert that one modeling approach is better than the other. Rather, the modeling approaches address different questions and yield different inference. We do note that our approach is novel in considering space with circularity of time.

There is a useful literature modeling crime data as linear in time using past locations to predict future locations. In this regard, see Mohler et al. (2011), Mohler (2013) and Chainey, Tompson and Uhlig (2008). Mohler et al. (2011) employ self-exciting point process models, similar to those used in earthquake modeling [e.g., Ogata (1998)], arguing that crime, when viewed linearly in time, can exhibit "contagion-like" behavior. Mohler (2013) considers a self-exciting process with the background rate driven by a log Gaussian Cox process to disentangle contagion from other types of correlation. Chainey, Tompson and Uhlig (2008) focus on hotspot assessment. This is purely spatial analysis, which may be implemented across various time periods for comparison.

Wrapping time to a circle takes us to the realm of directional or angular data where we find applications to, for example, wind directions, wave directions and animal movement directions. For a review of directional data methodology see, for example, Fisher (1993), Jammalamadaka and SenGupta (2001), Mardia (1972) and Mardia and Jupp (2000). Traditional models for directional data have employed the von Mises distribution, but recent work has elaborated the virtues of the wrapping and projection approaches, particularly in space and space–time settings [see Jona-Lasinio, Gelfand and Jona-Lasinio (2012) and Wang and Gelfand (2014)].

For event times during a day, wrapping time seems natural. Again, these times only arise given an orientation. However, crimes at 23:55 and 00:05 are as temporally close as crimes at 23:45 and 23:55. Another example analogous to our setting might be to model the arrival times (over 24 hours) of patients to a hospital (and, to add space, we might consider the addresses of the arrivals).

Our data consists of a set of crime events in San Francisco (SF) during the year 2012. Each event has a time of day and a location. In fact, we also have a classification into crime type, and we also have assignment of each crime to a district, arising by suitable partitioning of the city. Last, we know the day of the year for the event, enabling consideration of day of the week effects.

There is a substantial literature which employs regression models to explain the incidence of crime using a variety of socio-economic variables. In particular, for spatially referenced covariates, we can imagine employing census unit risk factors such as percent of home ownership, median family income and measures of neighborhood quality, along with racial and ethnic composition. Such covariates could be developed as tiled spatial surfaces over San Francisco using census data at, say, tract or block scale. As an alternative, we introduce illustrative point-referenced constructed covariates in space and time. In particular, with regard to space, we

imagine high-risk locations, so-called *crime attractors*, that is, places of high population and high levels of human activity such as commercial centers and malls. Then we view risk as exposure in terms of distance from such landmarks. We adopt these covariates in both the NHPP and LGCP models. As a temporal risk factor, we introduce into the mean a function which reflects the fact that, depending upon the type of crime, evening and late night hours may experience higher incidence of crime than morning and afternoon hours. Again, we adopt this covariate in both the NHPP and LGCP models. Exploratory data analysis in Section 2 reveals a day of the week effect with regard to daily crime time.

Finally, we address model comparison. In particular, within the Bayesian framework, how do we compare a NHPP model with a LGCP model? We adopt the strategy proposed in Leininger and Gelfand (2016). Briefly, the idea is to develop a cross-validation, employing a fitting/training point pattern and a testing/validation point pattern. Using the validation point pattern, with regard to model adequacy, we look at empirical coverage vs. nominal coverage of credible predictive intervals. In particular, these intervals are associated with the posterior distribution of predictive residuals for cell counts for randomly selected sets. With regard to model comparison, we look at rank probability scores [see, e.g., Gneiting and Raftery (2007) and Czado, Gneiting and Held (2009)] for the posterior distributions associated with predictive residuals, again for cell counts for randomly selected sets.

The format of the paper is as follows. In Section 2 we provide the details of our crime event dataset along with some exploratory analysis. In Section 3 we consider model construction with associated theoretical background. Employing day of the week marks, in Section 4 we introduce our model specifications and fitting strategies. In Section 5 we provide the inference results for both a simulated dataset and the crime event dataset. Finally, in Section 6 we present a brief summary along with proposed future work.

2. The dataset. Our dataset consists of crime events in the city of San Francisco in 2012. We have three crime type categories: (1) assault, (2) burglary/robbery[2] and (3) drug. Each crime event has a time (date, day of week, time of day) and location (latitude and longitude) information. Spatial coordinates (latitude and longitude) were transformed into eastings and northings. Each crime event is also classified into a district. In particular, there are 10 districts in San Francisco: (1) Bayview, (2) Central, (3) Ingleside, (4) Mission, (5) Northern, (6) Park, (7) Richmond, (8) Southern, (9) Taraval, (10) Tenderloin (see Figure 1).

[2]In the original dataset, burglary and robbery events are reported separately with hourly histograms and spatial density maps provided in the supplementary materials [Shirota and Gelfand (2017)]. Burglary and robbery are not universally accepted as being behaviorally similar. However, we aggregate these crimes due to their similar definition and to increase the number of events in our point patterns.

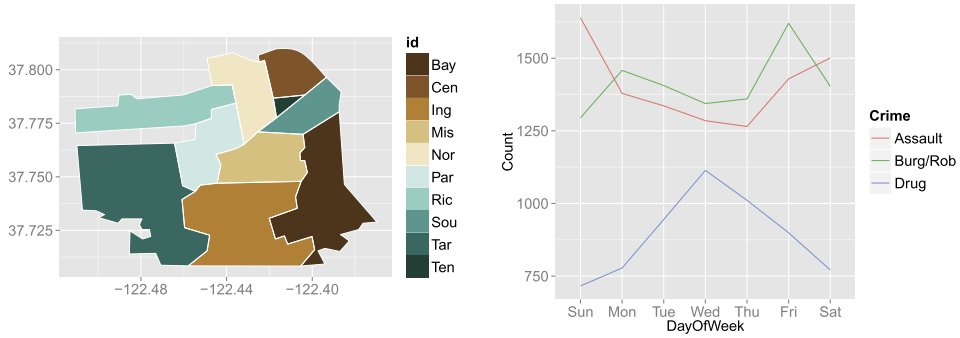

FIG. 1. *The map of San Francisco (left) and crime counts on each day of week (right).*

Figure 1 shows the counts of crime events for day of week.[3] Counts for crime types show different patterns. Assault events happen more on weekends, but burglary/robbery events happen most on Friday. Interestingly, it seems drug events happen most on Wednesday.

Figure 2 shows the data by type and by day of the week (3×7 plots) in the form of "rose" diagrams. This figure reveals differences among crime types and also differences across day of the week. For example, drug-related crime events are

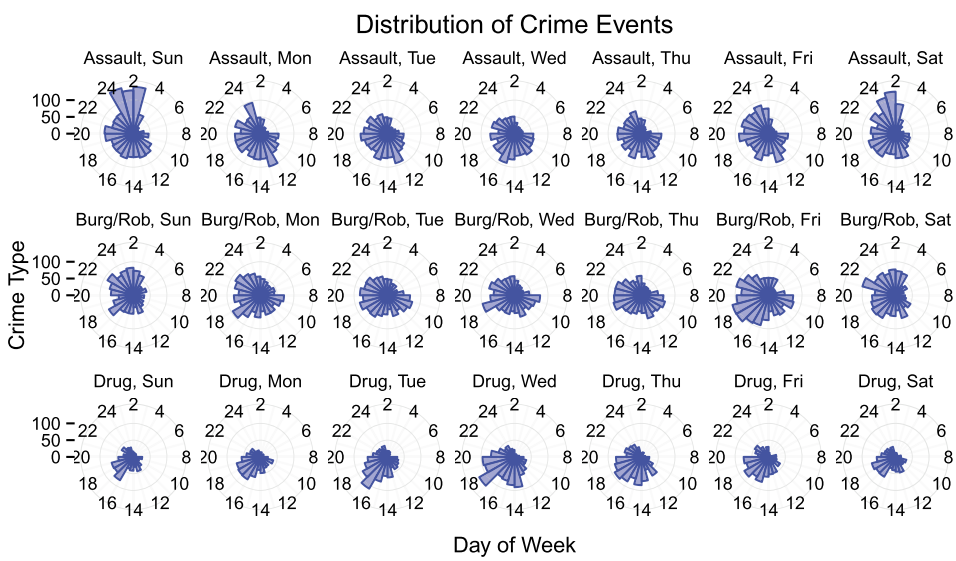

FIG. 2. *Histograms of crime events by type and by day of the week.*

[3]Here, and in the sequel, we take day of the week as 02:00 to 02:00. This definition interprets crime events on, for example, Saturday night as including the early hours of Sunday morning.

observed more from 5 to 7 pm, while burglary/robbery crime events are observed later in the day. Overall, the circular time dependence of crime events is seen, for example, large counts from evening to late night and small counts from early morning through the middle of the day. In the point pattern model construction below, we model each crime type separately and, within crime type, incorporate day of week as a mark.

3. Modeling and theory. Observations on a circle lead us to the world of directional data, as illustrated in Figure 2. Once an orientation has been chosen, the circular observations are specified using the angle from the orientation to the corresponding point on the unit circle. However, here, we are only concerned with point patterns on a circle. For the nonhomogeneous Poisson process and log Gaussian Cox process models we only need to specify the intensity functions for the processes, and so, in what follows, we only consider specifications for intensities for space–time point patterns over $D \times S^1$ where S^1 is the unit circle.

3.1. *The nonhomogeneous Poisson process* (*NHPP*) *and Log Gaussian Cox process* (*LGCP*). Again, since the crime events are random both in number and in space–time location, it is natural to think of them as a random point pattern over space and time. Here, we make the assumption that events are located in space and time, conditionally independent given their intensity, anticipating that the intensity surface will explain the observed clustering in space and time. Then we consider the two most common models for such a setting: the NHPP and the LGCP. The LGCP dates at least to Møller, Syversveen and Waagepetersen (1998). As a spatial process, it is defined so that the log of the intensity is a Gaussian process (GP), that is,

$$\log \lambda(\boldsymbol{s}) = X(\boldsymbol{s})^T \beta + Z(\boldsymbol{s}), \qquad Z(\boldsymbol{s}) \sim \mathcal{GP}(\mathbf{0}, C). \tag{3.1}$$

Here, $Z(\boldsymbol{s})$ is a zero mean stationary, isotropic GP over D with covariance function C, which provides spatial random effects for the intensity surface, pushing up and pulling down the surface, as appropriate. If we remove $Z(\boldsymbol{s})$ from the log intensity, then we obtain the associated NHPP. NHPP's have a long history in the literature [see, e.g., Illian et al. (2008)]. In fact, if Y is a Cox process with intensity $\Lambda(\boldsymbol{s})$, then, conditional on $\Lambda(\boldsymbol{s}) = \lambda(\boldsymbol{s})$, Y is an NHPP with intensity $\lambda(\boldsymbol{s})$. Evidently, an LGCP provides a very flexible intensity specification. Below, we will argue that, with regard to our crime data, we prefer the additional flexibility of a *space–time* Log Gaussian Cox Process (LGCP) to the associated NHPP. Our model is in the spirit of the space–time LGCP introduced in Brix and Diggle (2001).

3.2. *Circular covariance functions for Gaussian processes*. Again, we consider a three-dimensional Gaussian process with a two-dimensional location, and one-dimensional circular time. In general, we seek

$$Z(\boldsymbol{s}, t) \sim \mathcal{GP}(0, C), \qquad (\boldsymbol{s}, t) \in \mathbb{R}^2 \times S^1. \tag{3.2}$$

We need to specify valid correlation functions over $\mathbb{R}^2 \times S^1$.

Gneiting (2013) proposes families of circular correlation functions (CCF's) based on truncation of familiar spatial correlation functions. He shows that the completely monotone functions are strictly positive definite on spheres of any dimension, for example, powered exponential, Matérn, generalized Cauchy and Dagum families. One of the examples in Gneiting (2013) is the powered exponential family,

$$C_{\mathrm{PE}}(u) = \exp\bigl(-(\phi u)^{\alpha}\bigr), \qquad u \in [0, \pi], \alpha \in (0, 1]. \tag{3.3}$$

If $\alpha \in (0, 1]$, then this function is a strictly positive definite function for any dimension, but if $\alpha > 1$, then (5) is no longer positive definite, even in one dimension.

Another example in Gneiting (2013) is the generalized Cauchy family,

$$C_{\mathrm{GC}}(u) = \bigl(1 + (\phi u)^{\alpha}\bigr)^{-\tau/\alpha}, \qquad \text{for } u \in [0, \pi]\alpha \in (0, 1], \tau > 0, \tag{3.4}$$

where τ is a shape parameter which doesn't affect the positive definiteness as long as $\tau > 0$. This function is positive definite for any dimension if $\alpha \in (0, 1]$. Again, for $\alpha > 1$, (3.4) is also not positive definite, even in one dimension.

It may be surprising that restriction of familiar spatial correlation functions to the spherical domain maintains positive definiteness on the sphere. However, this enables convenient choices and, in fact, we adopt the generalized Cauchy family as the circular correlation function in the analysis below.

3.3. *Space and linear time covariance functions.* Next, we turn to valid covariance functions over $R^2 \times S^1$. We consider both the separable case, which is immediate, and also the nonseparable case.

Separable covariance functions. In the context of the LGCP model, we need to specify the covariance function for the latent Gaussian process $Z(\boldsymbol{s}, t)$. Separable space–time covariance functions are often adopted due to convenient specification and computational simplification [Banerjee, Carlin and Gelfand (2015)]. The separable specification arises if the space–time covariance function is written as a product of a valid space and a valid time covariance function, that is,

$$C_{s,t}(\boldsymbol{h}, u) = C_s(\boldsymbol{h}) C_t(u). \tag{3.5}$$

In our setting, we can define a valid space–time covariance function merely by choosing as C_s any valid covariance function on R^2 and multiplying it by any of the foregoing valid CCF's. The resulting covariance matrix for a set of $(\boldsymbol{s}, t)$'s with N $\boldsymbol{s}$'s by M t's will have a Kronecker product form $C_s \otimes C_t$ where C_s and C_t are $N \times N$ and $M \times M$ covariance matrices. The simplified inverse, determinant and Cholesky decomposition result, making the separable specification computationally efficient and tractable in high-dimensional cases.

In this regard, we note that the point pattern data arises as a set $(\boldsymbol{s}_i, t_i)$, $i = 1, 2, \dots, n$, where n is the total number of points. As we don't have a factorization

in space and time, why is the separable form helpful? Below, we clarify the need for grid approximation (discretization) for both space and time in order to evaluate the LGCP likelihood and in order to obtain manageable computation for the model fitting. Then N and M will become the number of grid centroids for space and for time, respectively. For the separable case, we can then take advantage of the Kronecker factorization. For the nonseparable case, we require the Cholesky decomposition and the inverse of an $NM \times NM$ matrix, making the computation much more demanding.

Nonseparable covariance functions. It is evident that the separable covariance specification is restrictive for real data applications because it precludes space–time interaction of the sort we mentioned in the Introduction. Various versions of nonseparable covariance functions have been proposed for the case where space is again R^2 and time is linear. Cressie and Huang (1999) propose specifications of nonseparable stationary covariance functions based on Fourier transformation on $\mathbb{R}^d$ with criteria which guarantee positive definiteness. Since their specification requires a closed-form solution for the d-dimensional Fourier transformation, the class of functions is relatively small. Gneiting (2002) proposed a flexible parametric family of nonseparable covariance functions, extending the results of Cressie and Huang (1999). Gneiting's class takes the form

$$C(\boldsymbol{h}, u) = \frac{\sigma^2}{\psi(\|u\|^2)^{d/2}} \varphi\left(\frac{\|\boldsymbol{h}\|}{\psi(\|u\|^2)}\right), \qquad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}^l, \tag{3.6}$$

where $\varphi(t)$, $t \geq 0$, is a completely monotone function and $\psi(t)$, $t \geq 0$, is a positive function with a completely monotone derivative. In our modeling below, we utilize Gneiting's specification. However, these covariance functions are specified on $\mathbb{R}^d \times \mathbb{R}^1$; our need is to provide valid nonseparable covariance functions on $\mathbb{R}^d \times S^1$.

3.4. *Space by circular time covariance functions.* Finally, we turn to the space–time covariance functions we seek. For the spatial correlation function, we assume the exponential correlation function,

$$C_s(h) = \exp(-\phi_s \|h\|). \tag{3.7}$$

For the circular time correlation function, we consider the truncated generalized Cauchy correlation function [again, see, Gneiting (2013)].

Then we arrive at the following proposed nonseparable covariance function over space by circular time:

$$\begin{aligned} C(\boldsymbol{h}, u) = {} & \frac{\sigma^2}{(1 + (\phi_t u)^\alpha)^{\delta + \gamma(d/2)}} \\ & \times \exp\left(-\frac{\phi_s \|\boldsymbol{h}\|}{(1 + (\phi_t u)^\alpha)^{\gamma/2}}\right), \qquad \gamma \in (0, 1], \alpha \in (0, 1], \end{aligned} \tag{3.8}$$

where $(\boldsymbol{h}, u) \in \mathbb{R}^2 \times [0, \pi]$ and γ is the nonseparability parameter. We can show that this is a valid covariance function on $\mathbb{R}^2 \times S^1$ following the proof in Gneiting (2002), working with the valid circular choices according to Gneiting (2013). We prove this result in the supplemental material.

As noted in the Introduction, motivated by processes observed on the Earth over time, recently, Porcu, Bevilacqua and Genton (2016) proposed nonseparable covariance functions on spheres crossed with linear time (as well as cross-covariance functions for multivariate random fields defined over a sphere). Hence, in space and time, their nonseparable covariance functions are specified on $S^d \times \mathbb{R}$, where S^d is the d-dimensional unit sphere. One of their contributions is similar to ours, that is, following work by Gneiting (2002) presenting nonseparable space–time covariance functions with further work by Gneiting [Gneiting (2013)] presenting spatial correlation functions on a sphere, they obtain fairly general nonseparable forms [see Table 2 in the online supplementary material of Porcu, Bevilacqua and Genton (2016)] over this product space. Here, we follow a similar path but take space as R^2 with time as circular, and, for our application, we employ the particular class of the generalized Cauchy family over this product space. As a result, the case of the real line crossed with the circle provides the common domain.

For model fitting with (3.8), we need to implement calculations for an $NM \times NM$ matrix. For the whole of the city of San Francisco, N is very large. Thus, for convenience, we take a smaller region and adopt a more local investigation of nonseparability for space–time crime patterns. Figure 3 shows the spatial locations of all of the crime events and the point patterns for the Tenderloin and Mission districts where relatively more events are observed than in the other districts. Then we create a rectangular region around this area (see Figure 4 below). For this region, in the interest of comparison, we implement the truncated generalized Cauchy correlation function for circular time in both the separable and nonseparable cases.

4. Model specification, fitting and checking. We specify the intensity for the NHPP as $\log \lambda(\boldsymbol{s}, t) = \log \lambda_0(\boldsymbol{s}) + \log \kappa_0(t)$. For the LGCP, we add $Z(\boldsymbol{s}, t)$, a mean 0 GP with covariance function chosen according to the previous section.

Covariate specification. We employ constructed space and time covariates. For the spatial covariates, we identify a set of landmarks. These landmarks are referred to as *crime attractors* [Brantingham and Brantingham (1995)] and are selected from centers of commercial activity, that is, places with high population density and high human exposure. Examples might include malls, market streets and amusement centers. For a given landmark, we employ a directional Gaussian kernel function as the distance measure from crime location to landmark, that is, inverse distance measures risk; the smaller the distance the larger the risk.

Formally, the covariate level at location $\boldsymbol{s}$ associated with landmark k is

$$g_k(\boldsymbol{s}, \boldsymbol{s}_k^*) = \exp\left(-\frac{1}{2}(\boldsymbol{s} - \boldsymbol{s}_k^*)' \Sigma_k^{-1} (\boldsymbol{s} - \boldsymbol{s}_k^*)\right), \tag{4.1}$$

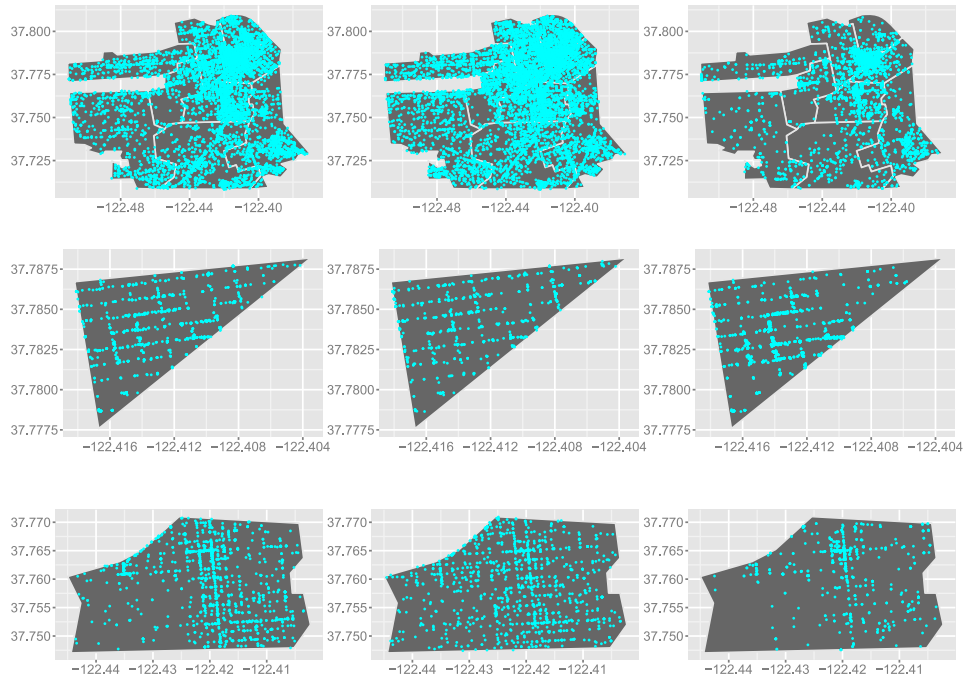

FIG. 3. *Crime event locations in SF* (*upper*), *Tenderloin* (*middle*) *and Mission* (*lower*) *for assault* (*left*), *burglary/robbery* (*middle*) *and drug* (*right*).

where $\boldsymbol{s}_k^*$ is the location of landmark k and $\Sigma_k = \begin{pmatrix} \sigma_1^2 & \rho_k\sigma_1\sigma_2 \\ \rho_k\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ is a positive definite matrix for landmark k. Here, σ_1^2 and σ_2^2 are scales for Easting and Northing coordinates, and ρ_k is the correlation of the kernel for landmark k. In fact, σ_1 is the centroid to centroid Easting distance between adjacent grid cells, and σ_2 the centroid to centroid Northing distance between adjacent grid cells. ρ_k is treated as an unknown, along with the β_k's, a regression coefficient assigned to g_k. Thus, we let $\log \lambda_0(\boldsymbol{s}) = \sum_{k=1}^K \beta_k g_k(\boldsymbol{s}, \boldsymbol{s}_k^*)$.

Figure 4 shows the contour plot for assault events and the landmarks. Illustratively, we create two landmarks, $L_1 = (-122.408, 37.784)$ (Union Square Shopping Center, henceforth "Union Square") and $L_2 = (-122.419, 37.764)$ (BART Station, 16th and Mission, henceforth "BART Station"). The left-hand side is the contour plot of the kernel density estimate for the observed assault events, obtained by `ggplot2` in R package. The right-hand side of Figure 4 overlays the subregion used with the nonseparable covariance.

To form a temporal covariate, we need a function whose support is the unit circle. Since crime events occur more frequently in the evening and night hours, and less in the morning and afternoon hours, the most elementary constructed covariate

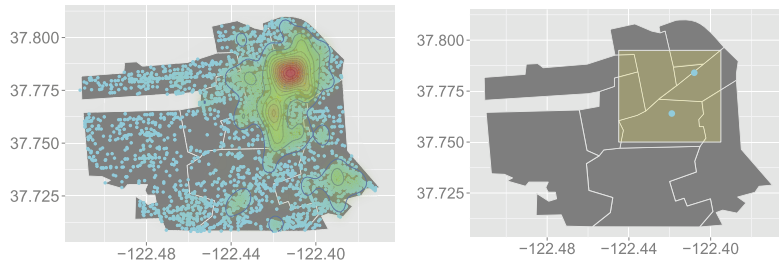

FIG. 4. *The contour plot of the kernel density estimate for the assault events* (*left*) *and landmarks with the subregion* (*right*, *color rectangular region*).

which reflects this would have two levels. Here, we let

$$\kappa(t) = \mu\big(1 + \delta 1\big(t \in [4\pi/3, 2\pi)\big)\big). \tag{4.2}$$

On the 24 hour scale, this choice of κ would be interpreted as adopting level μ for times between 02:00 and 18:00 and level $\mu(1+\delta)$ for times between 18:00 and 02:00 in the morning. μ and δ become model parameters; alternative windows could be explored. In fact, as demonstrated in Figure 2, the pattern and incidence of crime events vary across day of week, and so we introduce a different temporal covariate for each day of the week (writing μ_w and δ_w for $w = \text{Sun}, \dots, \text{Sat}$). Recalling Section 3, we make these covariates consistent by defining day of the week as 02:00 to 02:00. This specification yields $\kappa(t, w)$ of the form in (4.2). Combined with $g_k(\boldsymbol{s}, \boldsymbol{s}_k^*)$ in (4.1), this enables our specification of $\lambda(\boldsymbol{s}, t, w)$ below.

Model specification. Our baseline space by circular time LGCP model with *separable* space–time covariance function is defined below. We employ the foregoing binary time of day covariate and landmark distance covariates. The model is defined on grid points and on each day of week. After discretization, let J be the total number of space–time grids cells, that is, we have $(\boldsymbol{s}_j, t_j)$, for $j = 1, \dots, J$ and $w = \text{Sun}, \dots, \text{Sat}$,

$$y_{\boldsymbol{s}_j, t_j, w} | \lambda(\boldsymbol{s}_j, t_j, w) \sim \mathcal{NHPP}\big(\lambda(\boldsymbol{s}_j, t_j, w)\big), \qquad \boldsymbol{s}_j \in \mathbb{R}^2 t_j \in S^1, \tag{4.3}$$

$$\lambda(\boldsymbol{s}_j, t_j, w) = \lambda_0(\boldsymbol{s}_j)\kappa(t_j, w)\exp\left(-\frac{\sigma^2}{2} + Z(\boldsymbol{s}_j, t_j)\right), \tag{4.4}$$

$$Z(\boldsymbol{s}_j, t_j) \sim \mathcal{GP}(0, C), \qquad C = \sigma^2 C_s \otimes C_t, \tag{4.5}$$

$$\log \lambda_0(\boldsymbol{s}_j) = \sum_{k=1}^{K} \beta_k g_k\big(\boldsymbol{s}_j, \boldsymbol{s}_k^*\big), \tag{4.6}$$

$$g_k\big(\boldsymbol{s}_j, \boldsymbol{s}_k^*\big) = \exp\left(-\frac{1}{2}\big(\boldsymbol{s}_j - \boldsymbol{s}_k^*\big)' \Sigma_k^{-1} \big(\boldsymbol{s}_j - \boldsymbol{s}_k^*\big)\right), \tag{4.7}$$

$$\kappa(t_j, w) = \sum_{v=\text{Sun}}^{\text{Sat}} \mathbf{1}(w = v)\mu_v\big(1 + \delta_v \mathbf{1}(t_j \in \boldsymbol{I})\big), \tag{4.8}$$

where $y_{\boldsymbol{s}_j, t_j, w}$ is the count for grid cell j on day w, $t \in S^1$ is the circular time variable, $\boldsymbol{s}$ is the location coordinates and $\boldsymbol{I} = [4\pi/3, 2\pi)$. The space and circular time Gaussian process $Z(\boldsymbol{s}, t)$ is common across all the days of week, while the temporal covariate $\kappa(t, w)$ is dependent on the day of week.

Prior specifications. We assume gamma priors for the timescale parameters μ and δ for convenience and normal priors for $\boldsymbol{\beta}$'s with large variance. As for the parameters of Gaussian processes, we assume uniform distributions for ϕ_s and ϕ_t. The range associated with ϕ_s is chosen such that the correlation between locations at the maximum distance for the study region is 0.05. The maximum circular distance in time is π, and so we chose ϕ_t to provide correlation 0.05 at that distance. We used these priors in both the separable and nonseparable cases.

For the spatial Gaussian processes, ϕ_s and σ^2 are not identifiable [Zhang (2004)], and so we need to adopt an informative prior distribution for one of them. Here, we are informative about ϕ and adopt an inverse gamma distribution for σ^2 with relatively large variance. Finally, we assume a uniform prior on the domain of definition for the *separability* parameter γ.

4.1. *Model fitting*. In fitting of the LGCP model, we have a stochastic integral of the form $\int_{D\times S^1} \lambda(\boldsymbol{s}, t)\, d\boldsymbol{s}\, dt$ in the exponential of the likelihood. We use grid cell approximation for this integral as well as for the product term in the likelihood, yielding

$$L\big(\lambda(\cdot)\big) \propto \exp\left(-\sum_{w=\text{Sun}}^{\text{Sat}} \sum_{j=1}^{J} \lambda\big(\boldsymbol{s}_j^*, t_j^*, w\big)\Delta_{s,t,w}\right) \prod_{w=\text{Sun}}^{\text{Sat}} \prod_{j=1}^{J} \lambda^{n_{j,w}}\big(\boldsymbol{s}_j^*, t_j^*, w\big), \tag{4.9}$$

where $n_{j,w}$ is the number of events in grid j on day w, J is the number of grid cells, N is the total number of points in the point pattern and $(\boldsymbol{s}_j^*, t_j^*)$ are the centroids of the grids cells. Fitting this approximation is straightforward because we only require evaluation of $\lambda(\boldsymbol{s}_j^*, t_j^*, w)$ over the grid cells.

Sampling of parameters related to the NHPP model can be implemented through the Metroplis Hastings (MH) algorithm. However, for the LGCP model, sampling the large number of Z's from the Gaussian process is difficult to implement efficiently with standard Gibbs sampling. The customary Metropolis–Hastings algorithm often gets stuck in local modes, and so a more sophisticated MCMC algorithm is required. A now common approach for the LGCP model is to utilize the Metropolis adjusted Langevin algorithm (MALA) [see Møller, Syversveen and Waagepetersen (1998) and Girolami and Calderhead (2011)].

Here, for the Gaussian process outputs and hyperparameters, we use elliptical slice sampling [Murray, Adams and Graham (2010) and Murray and Adams

(2010)] as discussed in Leininger and Gelfand (2016). Our sampling algorithm is based on algorithms 1 and 2 in Leininger (2014), pages 50–51. Let $\boldsymbol{Z}$ denote the vector of Gaussian process variables we need to sample, with $\boldsymbol{Z}$ having covariance matrix $C_{\boldsymbol{\theta}}$. Let $\boldsymbol{Z} = L_{\boldsymbol{\theta}}\boldsymbol{\nu}$, where $C_{\boldsymbol{\theta}} = L_{\boldsymbol{\theta}}L_{\boldsymbol{\theta}}'$ and $\boldsymbol{\nu} \sim N(0, I)$. We sample $\boldsymbol{\nu}^* = \boldsymbol{\nu}\cos(\omega) + \boldsymbol{\eta}\sin(\omega)$, where $\boldsymbol{\eta} \sim N(0, I)$ through the elliptical slice sampling algorithm. Given a sampled $\boldsymbol{\nu}$, we sample the hyperparameters $\boldsymbol{\theta}$ by proposing the $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ and $\boldsymbol{Z}^* = L_{\boldsymbol{\theta}^*}\boldsymbol{\nu}$ such that

$$u < \frac{L(\boldsymbol{Z}^*)\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{L(\boldsymbol{Z})\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}, \qquad u \sim \mathcal{U}[0, 1],$$

where $L(\cdot)$ is the likelihood in (4.9). We use random walk Metropolis Hastings for proposing the candidates, adaptively tuning the variance of the proposal density [see Andrieu and Thoms (2008)]. In this algorithm, sampling the hyperparameters does not involve direct evaluation of the prior distribution of the Gaussian process $[N(\boldsymbol{Z}|\boldsymbol{0}, C_{\boldsymbol{\theta}})]$ due to the transformation of variables.

4.2. *Model adequacy and model comparison.* Cross-validation is a standard approach for assessing model adequacy and is available for point pattern models with conditionally independent locations given the intensity, as for both the NHPP and LGCP [see Leininger and Gelfand (2016)].

We implement cross-validation by obtaining a training (fitting) dataset and a testing (validation) using p-thinning as proposed by Leininger and Gelfand (2016). Let p denote the retention probability, that is, we delete $\boldsymbol{s}_i \in \mathcal{S}$ with probability $1 - p$. This produces a training point pattern $\mathcal{S}^{\text{train}}$ and test point pattern $\mathcal{S}^{\text{test}}$, which are independent, conditional on $\lambda(\boldsymbol{s})$. In particular, $\mathcal{S}^{\text{train}}$ has intensity $\lambda(\boldsymbol{s})^{\text{train}} = p\lambda(\boldsymbol{s})$. We set $p = 0.5$ and estimate $\lambda(\boldsymbol{s})^{\text{train}}$ $\boldsymbol{s} \in D$. Then we convert the posterior draws of $\lambda^{\text{train}}(\boldsymbol{s})$ into predictive draws of $\lambda^{\text{test}}(\boldsymbol{s})$ using $\lambda^{\text{test}}(\boldsymbol{s}) = \frac{1-p}{p}\lambda^{\text{train}}(\boldsymbol{s})$.

Let $\{B_k\}$ be a collection of subsets of D. For the choice of $\{B_k\}$, Leininger and Gelfand (2016) suggest to draw random subsets of the same size uniformly over D. Specifically, for $q \in (0, 1)$, if the area of each B_k is $q|D|$, then q is the *relative* size of each B_k. They argue that making the subsets disjoint is time-consuming and unnecessary. Based on the p-thinning cross-validation, we consider two model performance criteria: (1) predictive interval coverage (PIC) and (2) rank probability score (RPS). PIC offers assessment of model adequacy, and RPS enables model comparison.

Predictive interval coverage. After the model is fitted to $\mathcal{S}^{\text{train}}$, the posterior predictive intensity function can supply posterior predictive point patterns and therefore samples from the posterior predictive distribution of $N(B_k)$. For the ℓth posterior sample, $\ell = 1, \dots, L$, the associated predictive residual is defined as

$$R_\ell^{\text{pred}}(B_k) = N^{\text{test}}(B_k) - N^{(\ell)}(B_k),$$

where $N^{\text{test}}(B_k)$ is the number of points of the test data in B_k. If the model is adequate, then the empirical predictive interval coverage rate, that is, the proportion of intervals which contain 0, is expected to be roughly the nominal level of coverage; below, we choose 90% nominal coverage. Empirical coverage much less than the nominal suggests model inadequacy; predictive intervals are too optimistic. Empirical coverage much above, for example, 100%, is also undesirable. It suggests that the model is introducing more uncertainty than needed.

Rank probability score. Gneiting and Raftery (2007) propose the continuous rank probability score (CRPS). This score is derived as a proper scoring rule and enables a criterion for assessing the precision of a predictive distribution for continuous variables. In our context, we seek to compare a predictive distribution to an observed count. Czado, Gneiting and Held (2009) discuss rank probability scores (RPS) for count data. Intuitively, a good model will provide a predictive distribution that is very concentrated around the observed count. While the RPS has a challenging formal computational form, it is directly amenable to Monte Carlo integration. In particular, for a given B_k, we calculate the RPS as

$$\begin{aligned}\mathrm{RPS}\big(F, N^{\text{test}}(B_k)\big) &= \frac{1}{L}\sum_{\ell=1}^{L}\big|N^{(\ell)}(B_k) - N^{\text{test}}(B_k)\big| \\ &\quad - \frac{1}{2L^2}\sum_{\ell=1}^{L}\sum_{\ell'=1}^{L}\big|N^{(\ell)}(B_k) - N^{(\ell')}(B_k)\big|.\end{aligned}$$

Summing over the collection of B_k gives a model comparison criterion. Smaller values of the sum are preferred.

5. Data analysis. In this section we implement our model first for simulated data and then for the SF crime dataset. We investigate our ability to recover the true model through the simulation study where we use the same geographic region and time period as in the real SF data. We approximate the likelihood by taking space and circular time grid cells. We take 238 spatial grid cells for the full region and 64 spatial grid cells for the subregion. As for the time grid, we take 48 time grid cells, that is, 02:00–02:30, 02:30–03:00, ..., 1:30–02:00. As representative points to evaluate the intensity, we take the centroids of the cells. We consider a separable covariance specification for the full region and nonseparable covariance for the subregion around the Tenderloin and the northern part of the Mission districts. We use the constructed spatial and temporal covariates as discussed in Section 4. For the directional Gaussian kernels associated with the landmarks, we need to estimate the correlation parameters $\rho_k, k = 1, 2$. We plug in the posterior means of these parameters obtained under the NHPP models for each crime category (see Table 1).

TABLE 1
Posterior means of the correlation parameters ρ_1 and ρ_2 in the directional Gaussian kernels for the NHPP model

	Full region			Subregion		
	Assault	**Burg/Rob**	**Drug**	**Assault**	**Burg/Rob**	**Drug**
ρ_1	0.097	0.088	0.209	0.151	0.232	0.284
ρ_2	−0.142	−0.344	0.054	−0.537	−0.893	0.946

Following Section 4.1, the priors for the parameters are chosen as $\mu_w, \delta_w \sim \mathcal{G}(2, 0.05)$, $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, 100\mathbf{I}_K)$, $\sigma^2 \sim \mathcal{IG}(2, 0.05)$, $\phi_s \sim \mathcal{U}[0, 0.3]$, $\phi_t \sim \mathcal{U}[0, 6]$ and $\gamma \sim \mathcal{U}[0, 1)$. We generate 150,000 posterior samples and take 100,000 as the burn-in period for the full region, and 100,000 samples and take 50,000 as burn-in period for the subregion. We also report the inefficiency factor (IF),[4] which suggests the relative number of correlated draws necessary to attain the same variance of the posterior sample mean from the uncorrelated draws [Chib (2001)].

5.1. *Simulation study results*. The simulation data are generated from our model specification in Section 4.1 with posterior means ρ_1 and ρ_2 for assault events (i.e., $\rho_1 = 0.097$ and $\rho_2 = -0.142$) and without day of week effects, that is, $\mu_w = \mu$ and $\delta_w = \delta$ for $w = \text{Sun}, \dots, \text{Sat}$. Specifically, we generate a point pattern under a separable covariance function for the full region and a second point pattern under a nonseparable covariance function for the subregion. The resulting numbers of points are 9,935 for the full region and 6,451 for the subregion. A LGCP model with separable covariance (LGCP-Sep) and NHPP model with the spatial and time covariates are implemented for simulated data over the full region.

Table 2 shows the estimation summary for the simulation data over the full region. The NHPP inference shows high precision but poor accuracy. For LGCP-Sep, although consistency for ϕ_s and σ^2 is not guaranteed [see Zhang (2004)], we see good recovery of the parameters. For the NHPP model with the spatial and time covariates, since the space–time Gaussian processes are not included in the model, the estimation results have biases relative to the true values based on the LGCP-Sep. The true and posterior intensity surface at two time grids: (1) 12:00–12:30 and (2) 24:00–00:30 are shown in the supplementary material. Compared with the true surface, we see some preference for the LGCP-Sep.

Table 3 shows the estimation summary for the simulated data over the subregion. True values of parameters are recovered well by the LGCP model with nonseparable covariance (LGCP-NonSep). Although the true value of γ is included in

[4]The inefficiency factor is the ratio of the numerical variance of the estimate from the MCMC sample relative to that from hypothetical uncorrelated samples, and is defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$, where ρ_s is the sample autocorrelation at lag s. These values on tables are calculated with samples obtained at each 50th iteration.

TABLE 2
Estimation summary for NHPP and the space by circular time LGCP with separable covariance for the simulation data over the full region

	NHPP					LGCP-Sep (True)			
	True	**Mean**	**95%CI**	**IF**		**True**	**Mean**	**95%CI**	**IF**
μ	1	0.081	[0.079, 0.084]	1	μ	1	1.317	[0.705, 2.485]	72
δ	0.5	1.125	[1.049, 1.208]	1	δ	0.5	0.391	[0.185, 0.649]	66
β_1	3	3.048	[2.966, 3.128]	4	β_1	3	3.990	[3.165, 4.949]	72
β_2	3	3.287	[3.204, 3.362]	4	β_2	3	2.673	[1.599, 3.907]	74
ρ_1	0.097	0.551	[0.501, 0.596]	1	σ^2	3	3.256	[2.850, 3.886]	73
ρ_2	−0.142	−0.424	[−0.480, −0.362]	1	ϕ_s	0.02	0.019	[0.018, 0.021]	60
					ϕ_t	0.1	0.101	[0.085, 0.117]	65
					$\sigma^2\phi_s$	0.06	0.064	[0.056, 0.075]	68
					$\sigma^2\phi_t$	0.3	0.328	[0.290, 0.375]	56

the 95% CI, the posterior for γ has large variance. Since the true model is LGCP-NonSep with $\gamma = 0.8$, the estimation of ϕ_t by LGCP-Sep yields some bias. The true and posterior intensity surface for two time grids: (1) 12:00–12:30 and (2) 24:00–00:30 are shown in the supplemental material. The estimated intensity surfaces for LGCP-Sep and LGCP-NonSep are similar to each other. Additionally, we also compare the values of RPS and PIC of LGCP-Sep with those of LGCP-NonSep. Model comparison will change with the selection of time grid but, altogether, the differences are ignorable. Details are provided in the supplementary material.

TABLE 3
Estimation summary for the space by circular time LGCP for simulation data over the subregion: LGCP with separable covariance (left) and LGCP with nonseparable covariance (right)

	LGCP-Sep					LGCP-NonSep (True)			
	True	**Mean**	**95%CI**	**IF**		**True**	**Mean**	**95%CI**	**IF**
μ	1	0.778	[0.196, 2.283]	73	μ	1	0.533	[0.144, 1.796]	73
δ	0.5	0.196	[0.000, 0.502]	70	δ	0.5	0.249	[0.076, 0.491]	58
β_1	3	2.909	[1.913, 4.150]	73	β_1	3	2.961	[2.261, 3.946]	72
β_2	3	2.600	[1.903, 3.213]	71	β_2	3	2.472	[1.895, 3.132]	70
σ^2	1	0.842	[0.567, 1.442]	72	σ^2	1	0.913	[0.535, 1.301]	72
ϕ_s	0.02	0.029	[0.014, 0.043]	70	ϕ_s	0.02	0.024	[0.014, 0.039]	69
ϕ_t	0.1	0.168	[0.117, 0.236]	58	ϕ_t	0.1	0.125	[0.084, 0.181]	61
$\sigma^2\phi_s$	0.02	0.023	[0.017, 0.032]	64	$\sigma^2\phi_s$	0.02	0.022	[0.013, 0.032]	67
$\sigma^2\phi_t$	0.1	0.139	[0.091, 0.217]	64	$\sigma^2\phi_t$	0.1	0.111	[0.075, 0.163]	61
					γ	0.8	0.498	[0.031, 0.951]	67

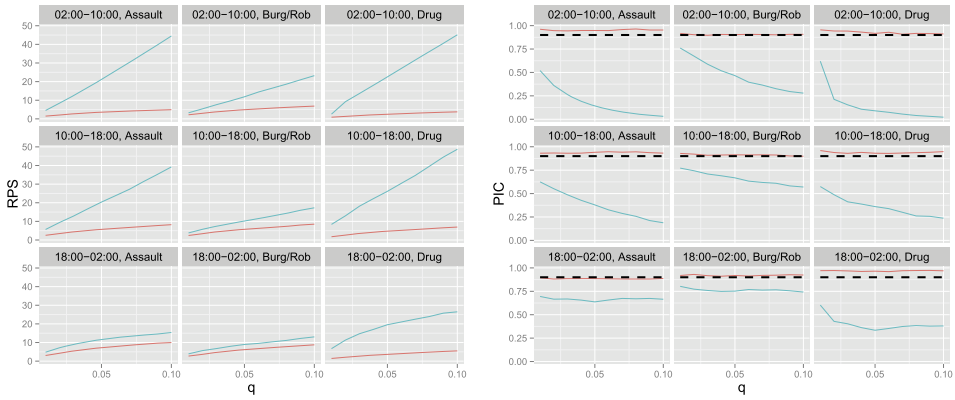

FIG. 5. *The rank probability score (left) and predictive interval coverage (right) for the full region: NHPP (blue line) and LGCP with separable covariance (red line). q is the relative area of B_k to that of D.*

5.2. *Real data application.*

Full region. The numbers of crime events in 2012 for the full region are 9834 for assault, 9884 for burglary/robbery and 6234 for drug. We implement LGCP-Sep and NHPP, assessing validation with 90% PIC and comparison using RPS. We separate $\mathcal{S}^{\text{test}}$ and $\mathcal{S}^{\text{train}}$ with $p = 0.5$. We calculate RPS and 90% PIC for three time ranges: (1) 02:00–10:00, (2) 10:00–18:00 and (3) 18:00–02:00. As for the choice of B_k, from the grid approximation over space and time, each B_k is chosen as a sum of grid units over space for each time interval. As above, the area of B_k is approximately equal to $q|D|$, where $|D|$ is the total area and here we choose the relative size $q \leq 0.1$. We randomly choose 1,000 sets of B_k uniformly over D and, following Section 4.2, compute average RPS and 90% PIC over these sets. Figure 5 shows RPS and PIC for the three crime categories and three time ranges. Figure 5 reveals that the LGCP-Sep model outperforms the NHPP model for all of the crime types and all of the time ranges. Figure 5 also demonstrates that the LGCP-Sep 90% predictive intervals capture nominal coverage very well while the PIC's for the NHPP are too small.

There may be interest in assessing *local* model adequacy to see if there are areas where the model is not performing well. Due to the flexibility of the LGCP-Sep model, we would not expect to see any local anomalies. However, we can assess this using our posterior samples and our collection of B_k's, enabing calculation of local PIC's. These are presented in Figure 6 for each crime type over the spatial region. We see that most are above the nominal, and a few below, but there is no evident pattern, no clustering. Then we can conclude that any over- or under-fitting occurs at random over the region.

Table 4 shows the estimation results for the space by circular time LGCP for the three crime categories in 2012. With day of week-specific μ_w and δ_w, μ_0 and δ_0

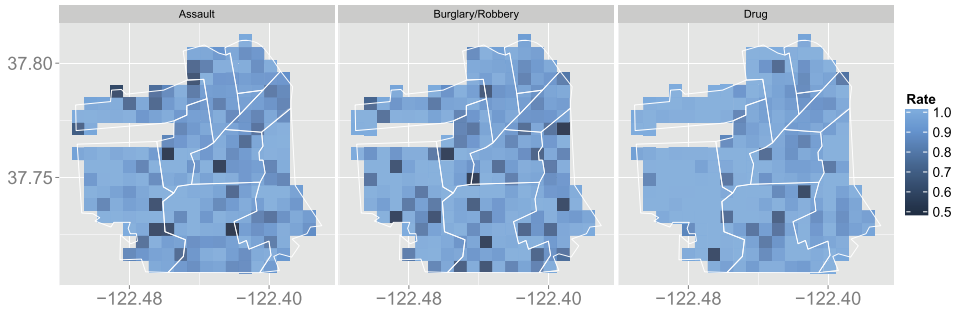

FIG. 6. *Local in-sample PIC for the three crime types.*

are set to the means of them over the days of week, yielding $\mu_w - \mu_0$ and $\delta_w - \delta_0$ as deviations. See Figure 7 below for inference on the δ_w across day of the week. The spatial covariates $\boldsymbol{\beta}$ are *positively* significant. In particular, β_1 and β_2 for drug crimes show larger values than those for the other crime types. This result suggests that drug events are more concentrated around landmarks L_1 and L_2.

Figure 7 shows the posterior mean and 95% CI of $\sum_{j=1}^{J} \lambda(\boldsymbol{s}_j^*, t_j^*, w)\Delta_{s,t,w}$ against counts on each day of week. For a given w, $\sum_{j=1}^{J} \lambda(\boldsymbol{s}_j^*, t_j^*, w_j^*)\Delta_{s,t,w}$ is approximately the expected number of crime events on day w a year. The left panel demonstrates that the posterior mean of $\sum_{j=1}^{J} \lambda(\boldsymbol{s}_j^*, t_j^*, w)\Delta_{s,t,w}$ traces the observed counts on days of week accurately. The right panel displays the posterior mean and 95% CI of δ_w. Although the variance of δ_w is large, this figure shows that δ_w varies with day of week; for assault, weekend δ's are larger. Since all of the δ_w's are positive, regardless of day of week or type of crime, we find elevated risk in the evening hours. Interestingly, although drug counts on Wednesday are

TABLE 4
Estimation results for space by circular time LGCP for the full region with separable covariance: assault (left), burglary/robbery (middle) and drug (right)

	Assault			**Burglary/Robbery**			**Drug**		
	Mean	**95%CI**	**IF**	**Mean**	**95%CI**	**IF**	**Mean**	**95%CI**	**IF**
μ_0	36.94	[20.62, 60.12]	70	36.18	[21.45, 57.79]	70	32.43	[19.04, 58.32]	69
δ_0	0.342	[0.201, 0.613]	68	0.188	[0.069, 0.424]	69	0.344	[0.137, 0.672]	70
β_1	1.654	[1.146, 2.023]	69	2.646	[2.038, 3.542]	73	3.874	[2.146, 5.045]	74
β_2	1.202	[0.614, 1.778]	70	0.470	[0.048, 0.823]	65	3.745	[2.117, 4.399]	71
σ^2	5.598	[5.064, 6.471]	73	5.756	[5.331, 6.219]	72	8.424	[7.868, 8.984]	67
ϕ_s	0.011	[0.010, 0.013]	70	0.005	[0.005, 0.007]	70	0.027	[0.025, 0.030]	68
ϕ_t	0.137	[0.123, 0.151]	58	0.178	[0.163, 0.196]	59	0.161	[0.140, 0.182]	63
$\sigma^2\phi_s$	0.066	[0.059, 0.072]	61	0.033	[0.028, 0.037]	65	0.231	[0.207, 0.259]	63
$\sigma^2\phi_t$	0.769	[0.681, 0.864]	61	1.027	[0.887, 1.164]	65	1.362	[1.182, 1.540]	60

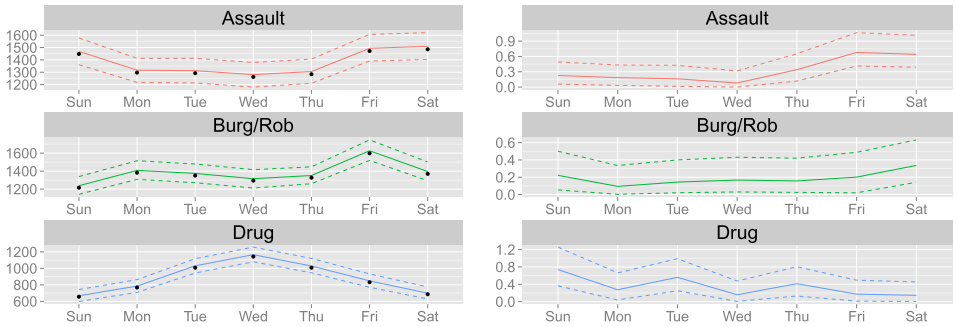

FIG. 7. *Posterior mean and* 95% *CI of* $\sum_{j=1}^{J} \lambda(s_j^*, t_j^*, w)\Delta_{s,t,w}$ (*left*: *dotted points are crime counts*) *and* δ_w (*right*) *on each day of week for the full region*: *dashed lines are* 95% *CI*.

larger than those on other days, δ_{Wed} for drug is smaller than those for the other days. Additionally, in the supplemental material, we provide figures for the posterior mean intensity surfaces for the three crime categories for three time grids: (1) 08:00–08:30, (2) 16:00–16:30 and (3) 24:00–00:30. The figures reveal different intensity patterns for each category and time grid.

Subregion. Finally, we turn to the nonseparable case, providing results for the subregion and comparison with the separable case. The number of points are 5579 for assault, 5407 for burglary/robbery and 4415 for drug crimes. Figure 8 shows the RPS and PIC for three models for the subregion: (1) NHPP, (2) LGCP-Sep and (3) LGCP-NonSep. Although both LGCP models fit considerably better than the NHPP model, the model performance of LGCP-Sep is difficult to distinguish from

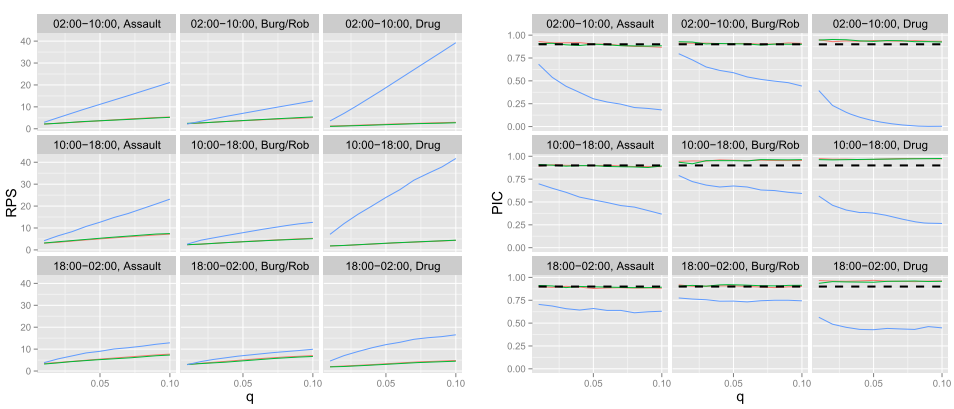

FIG. 8. *The rank probability score* (*left*) *and predictive interval coverage* (*right*) *for the subregion*: *NHPP* (*blue line*), *LGCP with separable covariance* (*red line*) *and LGCP with nonseparable covariance* (*green line*). q *is the relative area of* B_k *to that of* D.

TABLE 5
Estimation results for the space by circular time LGCP for the subregion with nonseparable covariance: assault (left), burglary/robbery (middle) and drug (right)

	Assault			Burglary/Robbery			Drug		
	Mean	**95%CI**	**IF**	**Mean**	**95%CI**	**IF**	**Mean**	**95%CI**	**IF**
μ_0	38.49	[19.89, 66.82]	67	67.93	[12.39, 51.44]	71	0.281	[0.237, 0.764]	71
δ_0	0.360	[0.196, 0.625]	67	0.347	[0.000, 0.980]	71	0.284	[0.001, 0.764]	70
β_1	1.875	[0.819, 2.586]	72	2.236	[1.819, 2.551]	65	4.204	[3.643, 4.975]	71
β_2	1.889	[1.103, 2.726]	72	1.498	[0.449, 2.464]	66	2.562	[1.451, 3.972]	66
σ^2	5.065	[4.435, 5.508]	70	7.244	[6.661, 8.382]	72	2.331	[1.761, 2.935]	67
ϕ_s	0.014	[0.012, 0.016]	65	0.003	[0.003, 0.004]	60	0.075	[0.059, 0.094]	66
ϕ_t	0.147	[0.117, 0.191]	69	0.236	[0.196, 0.277]	59	0.604	[0.354, 0.815]	68
γ	0.104	[0.006, 0.246]	55	0.069	[0.003, 0.194]	48	0.236	[0.016, 0.533]	53
$\sigma^2\phi_s$	0.072	[0.063, 0.083]	57	0.033	[0.028, 0.037]	65	0.231	[0.207, 0.259]	63
$\sigma^2\phi_t$	0.742	[0.617, 0.933]	66	1.027	[0.887, 1.164]	65	1.362	[1.182, 1.540]	60

that of LGCP-NonSep. This result is consistent with our findings in the simulation study.

Table 5 shows the estimation results. The estimated values of μ_0 for assault and burglary/robbery crimes are higher and have larger variances than that for drug crimes. From our model specification, μ_0 and σ^2 have strong positive correlation because both parameters are scale parameters for the intensity, that is, $\mu_0 \exp(-\sigma^2/2)$. In fact, the results reveal larger values of μ_0 and σ^2 for assault and burglary/robbery events, and a smaller value of μ_0 and σ^2 for drug events. $\gamma > 0$ express the degree of nonseparability. It varies with crime type but has high uncertainty so that the differences between types are not distinguished. As in the simulation study, the difference between LGCP-Sep and LGCP-NonSep is very small with respect to RPS. Figure 9 shows the posterior mean and 95% CI of the $\sum_{j=1}^{J} \lambda(\boldsymbol{s}_j^*, t_j^*, w)\Delta_{s,t,w}$ against counts on each day of the week. The right figure exhibits the posterior mean and 95% CI of the δ_w. Although the results are similar to those for the full region, δ_{Wed} for drug is larger than that for the full region, while now δ_{Tue} for drug is smaller than that for the full region.

6. Summary and future work. We have looked at times and locations of crime events for the city of San Francisco. We have argued that these data should be treated as point patterns in space and time where time should be treated as circular. We introduced derived spatial covariates (using distance from landmarks) and temporal covariates (using day of the week). We have looked at NHPP and LGCP models for such data. For the latter, we have proposed valid space and circular time Gaussian processes, both separable and nonseparable, for use in the LGCP. We have shown through a simulation example that we can recover the underlying model and intensity surface. We have discussed criteria for model adequacy

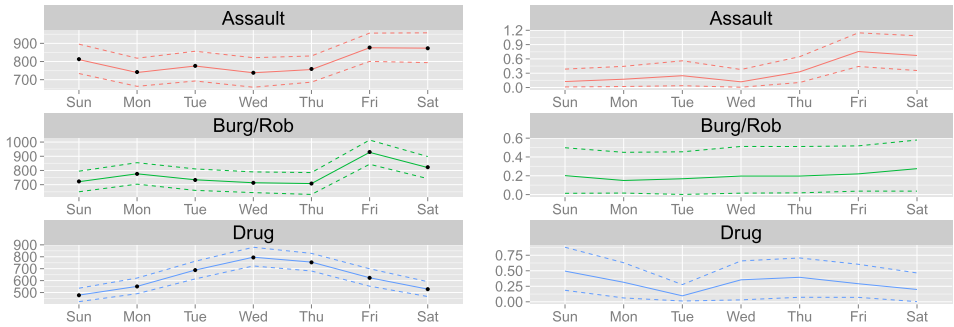

FIG. 9. *Posterior mean and* 95% *CI of* $\sum_{j=1}^{J} \lambda(s_j^*, t_j^*, w)\Delta_{s,t,w}$ (*left*: *dotted points are crime counts*) *and* δ_w (*right*) *on each day of week for the subregion*: *dashed lines are* 95% *CI*.

(PIC) and model comparison (RPS). We have shown that the LGCP outperforms the NHPP for the SF crime data. However, strong support for nonseparability for the subregion is not seen through our model estimation.

Future work will focus on more efficient computation. It will find us trying to develop appropriate approximations to enable us to fit the nonseparable model to larger regions. It will also consider alternative approaches to the likelihood approximation, following strategies proposed by Adams, Murray and MacKay (2009).

Acknowledgments. The authors thank Giovanna Jona Lasinio for suggesting this problem, for useful conversations and for providing the San Francisco dataset. The computational results are obtained by using Ox [Doornik (2007)].

SUPPLEMENTARY MATERIAL

Supplement to "Space and circular time log Gaussian Cox processes with application to crime event data" (DOI: 10.1214/16-AOAS960SUPP; .pdf). In this online supplement article, we provide (1) proof of the validity of our proposed nonseparable covariance function on $\mathbb{R}^2 \times S^1$ and (2) additional figures and tables to see posterior mean intensity estimates under different models.

REFERENCES

ADAMS, R. P., MURRAY, I. and MACKAY, D. J. C. (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the* 26*th International Conference on Machine Learning* MIT Press, Cambridge, MA.

ANDRIEU, C. and THOMS, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* **18** 343–373. MR2461882

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. MR3362184

BRANTINGHAM, P. and BRANTINGHAM, P. (1995). Criminality of place: Crime generators and crime attractors. *Eur. J. Crim. Policy Res.* **3** 5–26.

BRIX, A. and DIGGLE, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 823–841. MR1872069

CHAINEY, S., TOMPSON, L. and UHLIG, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Secur. J.* **21** 4–28.

CRESSIE, N. and HUANG, H. C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494

CZADO, C., GNEITING, T. and HELD, L. (2009). Predictive model assessment for count data. *Biometrics* **65** 1254–1261.

DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd ed. Springer, New York. MR1950431

DALEY, D. J. and VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd ed. Springer, Berlin. MR2371524

DOORNIK, J. (2007). *Ox: Object Oriented Matrix Programming*. Timberlake Consultants Press.

FISHER, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge Univ. Press, Cambridge. MR1251957

GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 123–214. MR2814492

GNEITING, T. (2002). Nonseparable, stationary covariance functions for space–time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475

GNEITING, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. MR3102554

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548

ILLIAN, J., PENTTINEN, A., STOYAN, H. and STOYAN, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns. Statistics in Practice*. Wiley, Chichester. MR2384630

JAMMALAMADAKA, S. R. and SENGUPTA, A. (2001). *Topics in Circular Statistics*. World Scientific, River Edge, NJ. MR1836122

JONA-LASINIO, G., GELFAND, A. E. and JONA-LASINIO, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *Ann. Appl. Stat.* **6** 1478–1498. MR3058672

LEININGER, T. J. (2014). Bayesian analysis of spatial point patterns. PhD dissertation. MR3232305

LEININGER, T. J. and GELFAND, A. E. (2016). Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. *Bayesian Anal.* **12** 1–30. MR3597565

LIANG, W. W. J., COLVIN, J. B., SANSÓ, B. and LEE, H. K. H. (2014). Modeling and anomalous cluster detection for point processes using process convolutions. *J. Comput. Graph. Statist.* **23** 129–150. MR3173764

MARDIA, K. V. (1972). *Statistics of Directional Data*. Academic Press, London. MR0336854

MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. Wiley, Chichester. MR1828667

MOHLER, G. O. (2013). Modeling and estimation of multi-source clustering in crime and security data. *Ann. Appl. Stat.* **7** 1825–1839. MR3127957

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

MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.* **25** 451–482. MR1650019

MURRAY, I. and ADAMS, R. P. (2010). Slice sampling covariance hyperparameters of latent Gaussian models. In *Advances in Neural Information Processing Systems* 23. MIT Press, Cambridge, MA.

MURRAY, I., ADAMS, R. P. and GRAHAM, M. M. (2010). Elliptical slice sampling. In *Proceedings of the* 13*th International Conference on Artifical Intelligence and Statistics* (*AISTAT*). AISTAT Press.

OGATA, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50** 379–402.

PORCU, E., BEVILACQUA, M. and GENTON, M. G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J. Amer. Statist. Assoc.* **111** 888–898. MR3538713

RODRIGUES, A. and DIGGLE, P. J. (2012). Bayesian estimation and prediction for inhomogeneous spatiotemporal log-Gaussian Cox processes using low-rank models, with application to criminal surveillance. *J. Amer. Statist. Assoc.* **107** 93–101. MR2949344

SHIROTA, S. and GELFAND, A. E. (2017). Supplement to "Space and circular time log Gaussian Cox processes with application to crime event data." DOI:10.1214/16-AOAS960SUPP.

TADDY, M. A. (2010). Autoregressive mixture models for dynamic spatial Poisson processes: Application to tracking intensity of violent crime. *J. Amer. Statist. Assoc.* **105** 1403–1417. MR2796559

WANG, F. and GELFAND, A. E. (2014). Modeling space and space–time directional data using projected Gaussian processes. *J. Amer. Statist. Assoc.* **109** 1565–1580.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303

DEPARTMENT OF STATISTICAL SCIENCE
DUKE UNIVERSITY
DURHAM, NORTH CAROLINA 27708-0251
USA
E-MAIL: ss571@stat.duke.edu
alan@stat.duke.edu