Functional Principal Component Analysis of Spatio-Temporal Point Processes with Applications in Disease Surveillance

Yehua Li and
Department of Statistics and Statistical Laboratory, Iowa State University, Ames, IA 50011

Yongtao Guan
Department of Management Science, University of Miami, Coral Gables, FL 33124

Yehua Li: yehuali@iastate.edu; Yongtao Guan: yguan@bus.miami.edu

Abstract

In disease surveillance applications, the disease events are modeled by spatio-temporal point processes. We propose a new class of semiparametric generalized linear mixed model for such data, where the event rate is related to some known risk factors and some unknown latent random effects. We model the latent spatio-temporal process as spatially correlated functional data, and propose Poisson maximum likelihood and composite likelihood methods based on spline approximations to estimate the mean and covariance functions of the latent process. By performing functional principal component analysis to the latent process, we can better understand the correlation structure in the point process. We also propose an empirical Bayes method to predict the latent spatial random effects, which can help highlight hot areas with unusually high event rates. Under an increasing domain and increasing knots asymptotic framework, we establish the asymptotic distribution for the parametric components in the model and the asymptotic convergence rates for the functional principal component estimators. We illustrate the methodology through a simulation study and an application to the Connecticut Tumor Registry data.

Keywords

Composite likelihood; Functional data; Latent process; Point process; Semi-parametric methods; Splines; Spatio-temporal data; Strong mixing

1 Introduction

Spatio-temporal point patterns commonly arise from many fields including ecology, epidemiology and seismology (e.g., Brix and Møller, 2001; Diggle, 2006; Schoenberg, 2003). The log-Gaussian Cox processes (LGCPs), first introduced by Møller et al. (1998) in the spatial case and later on extended to the spatio-temporal setting by Brix and Møller (2001) and Brix and Diggle (2001), provide a wide class of useful models for modeling such kind of data. For a typical spatio-temporal LGCP, its intensity function is assumed to be a log-linear model of some latent spatio-temporal Gaussian process, where the mean of the process may depend on some observed covariates. Borrowing ideas from recent developments in functional data analysis (Ramsay and Silverman, 2005), we model the latent temporal process at any fixed spatial location as a functional process with a standard

functional principal component expansion. We allow the functional principal component scores at different locations to be spatially correlated. The proposed model can accommodate both nonparametric temporal trend and spatio-temporal correlations in the point process.

In functional data analysis (FDA), the data considered are collections of curves, which are usually modeled as independent realizations of a stochastic process. Some recent papers on this topic include Yao, Müller and Wang (2005ab), Hall and Hosseini-Nasab (2006), Hall, Müller and Wang (2006) and Li and Hsing (2010ab). Di et al. (2009) and Zhou et al. (2010) studied multi-level functional data, where functional data at the lower level of the hierarchy are allowed to be temporally correlated. All the aforementioned papers considered only Gaussian type of functional data. Recently, Hall, Müller and Yao (2008) studied generalized longitudinal data, where the non-Gaussian longitudinal trajectories are linked to some Gaussian latent processes through a nonlinear link function and these latent random processes are modeled as functional data. For such non-Gaussian longitudinal data, Hall et al. proposed a nonparametric estimation procedure based on a delta method, which is an approximation by ignoring the higher order influence of the latent processes. There has also been some recent work on functional data modeling of point processes, including Bouzas et al. (2006), Illian et al. (2006) and Wu et al. (2013). These authors considered data with independent replicates of the point process and modeled a summary measure of the point process (e.g. the intensity function or the *L*-function) as functional data.

To develop FDA tools for spatio-temporal point processes, we encounter many new challenges and our proposed method is hence different from those in the literature in a number of ways. First, in most FDA papers in the literature, the data consist of *n* independent units (subjects). In our settings, however, there is only one realization of the spatio-temporal process, and the data are correlated both spatially and temporally. Second, unlike the scenarios considered in the classic FDA literature where the functional trajectories can be directly observed, the functional data in our setting are latent processes that determine the rate of events. To estimate the covariance structure of the process, we propose a novel method based on composite likelihood and spline approximation. We develop asymptotic properties of our estimators under an increasing domain asymptotic framework. Third, we perform spatial prediction of the latent principal component scores using an empirical Bayes method. These predicted spatial random effects can be put into maps to highlight hot areas with unusually high event rates or increasing trends in event rates. Such information can be valuable to government agencies when making public health polices.

Our work is motivated by cancer surveillance data collected by the Connecticut Tumor Registry (CTR). The CTR is a population-based resource for examining cancer patterns in Connecticut, and its computerized database includes all reported cancer cases diagnosed in Connecticut residents from 1935 to the present. Our primary interest here is to study the spatio-temporal pattern of pancreatic cancer incidences based on 8,230 pancreatic cancer cases in the CTR database from 1992 to 2009. The residential addresses and time of diagnosis are both available and are assumed to be generated by a spatio-temporal point process.

The rest of the paper is organized in the following way. We introduce the model assumptions in Section 2 and propose our estimation procedures in Section 3. Then we study the asymptotic properties of the proposed estimators in Section 4. The proposed methods are tested by a simulation study in Section 5 and are applied to the CTR data in Section 6. Assumptions for our asymptotic theory are collected in the appendix. All technical proofs and implementation details, including variance estimation, model selection and model diagnostic, are provided in the online Supplementary Material.

2 Model Assumptions

Let N denote a spatio-temporal point process that is observed on $W = D \otimes T$, where $D \subset \mathbb{R}^2$ is a spatial domain and T is a time domain. Let X be an L^2 Gaussian random field on W. We assume that conditional on X, N is a Poisson process with an intensity function $\lambda(s, t)$ given by

$$g\{\lambda(\boldsymbol{s}, t)\} = \boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{s}, t)\boldsymbol{\beta} + X(\boldsymbol{s}, t), \quad (1)$$

where g is a known link function such that $g^{-1}(\cdot)$ is nonnegative, $\boldsymbol{Z}(s, t)$ is a d-dimensional covariate vector, and X represents spatio-temporal random effects that cannot be explained by $\boldsymbol{Z}$.

In this paper, we will focus on the log link function, i.e. $g(\cdot) = \log(\cdot)$. The model given in (1) then becomes an LGCP model. In point process literature, the effect of the covariate $\boldsymbol{Z}(s, t)$ is often assumed to be parametric (Møller and Waagepetersen, 2007), although nonparametric approaches have also been recently proposed (Guan, 2008a; Guan and Wang, 2010). Similarly, a parametric model is generally used for the covariance structure of the latent process $X(s, t)$. For example, Brix and Diggle (2001) assumed a covariance structure from a class of Ornstein-Uhlenbeck processes, while Diggle et al. (2005) used a parametric covariance model that is stationary both in space and time. However, we are not aware of any existing literature that models the latent process nonparametrically in a spatio-temporal log-Gaussian Cox process as what we will do next.

For a fixed location s, $X(s, t)$ can be considered as an L^2 Gaussian process on T, and hence by the standard Karhunen-Loève expansion (Ash and Gardner, 1975),

$$X(\boldsymbol{s}, t) = \mu(t) + \sum_{j=1}^{p} \xi_j(\boldsymbol{s})\psi_j(t), \quad (2)$$

where $\mu(t) = \mathrm{E}\{X(s, t)\}$ with the expectation taken over all locations, $\psi_j(\cdot)$'s are orthonormal functions, and $\xi_j(\cdot)$'s are independent spatial Gaussian random fields. We assume that $\xi_j(s)$ is a zero-mean random field with variance ω_j and covariance function, $C_j(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathrm{cov}\{\xi_j(\boldsymbol{s}_1), \xi_j(\boldsymbol{s}_2)\}$, for $j = 1, 2, \cdots, p$. The functions $\psi_j(\cdot)$'s are called the eigenfunctions of process X. We assume that $\psi_j(\cdot)$'s are kept in a descending order of ω_j's, i.e. $\omega_1 \quad \omega_2 \quad \cdots \quad \omega_p > 0$. The number of principal components p can be ∞ in theory, but is often assumed to be finite for practical considerations.

The general covariance function of $X(\boldsymbol{s}, t)$ is

$$R(\boldsymbol{s}_1, \boldsymbol{s}_2, t_1, t_2) = \mathrm{cov}\{X(\boldsymbol{s}_1, t_1), X(\boldsymbol{s}_2, t_2)\} = \sum_{j=1}^{p} C_j(\boldsymbol{s}_1, \boldsymbol{s}_2) \psi_j(t_1) \psi_j(t_2), \quad (3)$$

which implies that $X(\boldsymbol{s}, t)$ is not necessarily stationary in t. Note that the above model coincides with the spatial coregionalization model commonly used for multivariate Gaussian random fields (Gelfand et al., 2004) and is not separable when $p > 1$. To connect with the FDA literature, it is helpful to consider the covariance function of the latent process $X(\boldsymbol{s}, t)$ at the same location $\boldsymbol{s}$. By setting $\boldsymbol{s}_1 = \boldsymbol{s}_2$, (3) is simplified to

$$R_T(t_1, t_2) = \mathrm{cov}\{X(\boldsymbol{s}, t_1), X(\boldsymbol{s}, t_2)\} = \sum_{j=1}^{p} \omega_j \psi_j(t_1) \psi_j(t_2), \quad \boldsymbol{s} \in D, \quad t_1, t_2 \in T, \quad (4)$$

where ω_j and $\psi_j(\cdot)$'s are the eigenvalues and eigenfunctions of $R_T(\cdot, \cdot)$. If a consistent estimator $\hat{R}_T(\cdot, \cdot)$ exists, one can then estimate $\{\omega_j, \psi_j(\cdot)\}$ by an eigenvalue decomposition of $\hat{R}_T$ using a standard functional data analysis approach (Ramsay and Silverman, 2005).

We estimate the proposed model through the use of the first- and second-order intensity functions of N. Let $N(d\boldsymbol{s}, dt)$ denote the number of events in an infinitesimal window $(d\boldsymbol{s}, dt)$, and let $|d\boldsymbol{s}|$ and $|dt|$ denote the volumes of $d\boldsymbol{s}$ and dt, respectively. The marginal first-order intensity function, which characterizes the probability to observe an event at a given location and time, is defined as

$$\lambda(\boldsymbol{s}, t) = \lim_{|d\boldsymbol{s}|, |dt| \to 0} \frac{\mathrm{E}\{N(d\boldsymbol{s}, dt)\}}{|d\boldsymbol{s}||dt|} = \mathrm{E}[\exp\{\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{s}, t)\boldsymbol{\beta} + X(\boldsymbol{s}, t)\}] = \exp\{\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{s}, t)\boldsymbol{\beta} + \gamma(t)\}, \quad (5)$$

where $\gamma(t) = \mu(t) + (1/2)\sum_{j=1}^{p} \omega_j \psi_j^2(t)$. In derivation of (5), we use the fact that $\mathrm{E}\{\exp(Y)\}$ $= \exp(\mu + \sigma^2/2)$ for a $Y \sim \mathrm{Normal}(\mu, \sigma^2)$ and the covariance of $X(\boldsymbol{s}, t)$ in (4). Non-stationarity in the first-order intensity function can be modeled by including proper spatio-temporal covariates $\boldsymbol{Z}(\boldsymbol{s}, t)$. For example, in our disease surveillance application, non-stationarity in the cancer rate caused by spatially varying population level is accommodated by including population density as a covariate.

The second-order intensity function, which characterizes the correlation within the process, is defined as

$$
\begin{aligned}
\lambda_2(\boldsymbol{s}_1, \boldsymbol{s}_2, t_1, t_2) &= \lim_{\substack{|d\boldsymbol{s}_1|, |d\boldsymbol{s}_2|, \\ |dt_1|, |dt_2| \to 0}} \frac{\mathrm{E}\{N(d\boldsymbol{s}_1, dt_1)N(d\boldsymbol{s}_2, dt_2)\}}{|d\boldsymbol{s}_1||dt_1||d\boldsymbol{s}_2||dt_2|} \\
&= \mathrm{E}[\exp\{\boldsymbol{Z}^T(\boldsymbol{s}_1, t_1)\boldsymbol{\beta} + \boldsymbol{Z}^T(\boldsymbol{s}_2, t_2)\boldsymbol{\beta} + X(\boldsymbol{s}_1, t_1) + X(\boldsymbol{s}_2, t_2)\}] \\
&= \lambda(\boldsymbol{s}_1, t_1)\lambda(\boldsymbol{s}_2, t_2)\exp\{\textstyle\sum_{j=1}^{p} C_j(\boldsymbol{s}_1, \boldsymbol{s}_2)\psi_j(t_1)\psi_j(t_2)\},
\end{aligned}
\quad (6)
$$

Li and Guan   Page 5

for (s_1, t_1) (s_2, t_2), where the last equality is a result of the Gaussian assumption for the principal component random processes $\xi_j(s)$. The Gaussian assumption is also commonly made in other settings such as generalized linear mixed models and spatial hierarchical models (Banerjee et al., 2003).

Given the first- and second-order intensity functions, the pair correlation function (e.g., Møller and Waagepetersen, 2004) for the point process is

$$\mathscr{G}_2(s_1, s_2, t_1, t_2) = \frac{\lambda_2(s_1, s_2, t_1, t_2)}{\lambda(s_1, t_1)\lambda(s_2, t_2)} = \exp\left\{\sum_{j=1}^{p} C_j(s_1, s_2)\psi_j(t_1)\psi_j(t_2)\right\}. \quad (7)$$

If $C_j(s_1, s_2)$ is stationary, i.e., it only depends on the spatial lag $s_1 - s_2$, then $\mathscr{G}_2(s_1, s_2, t_1, t_2)$ is a function of $s_1 - s_2$ and the time points (t_1, t_2). Hence, the point process is second-order intensity reweighted stationary in space (Braddley et al., 2000).

3 Estimation procedure

3.1 Estimation of the mean components

The Poisson maximum likelihood (Schoenberg, 2005) method is a general approach to fit parametric models for the intensity function of a point process, where the point process can be purely spatial, temporal or spatio-temporal. Asymptotic properties of the resulting estimator such as consistency and asymptotic normality were considered in Guan and Loh (2007). In the spatio-temporal case, let $\lambda(s, t; \theta)$ be such a model under consideration where θ is some unknown parameter. Then, θ can be estimated by maximizing

$$\ell(\theta) = \sum_{(s,t) \in N \cap W} \log\left\{\lambda(s, t; \theta)\right\} - \int_D \int_T \lambda(s, t; \theta)\, dt\, ds. \quad (8)$$

In our setting, we will apply the above method to estimate β and $\gamma(t)$. Observe that $\lambda(s, t; \theta)$ $= \exp\{Z^{\mathrm{T}}(s, t)\beta + \gamma(t)\}$. We then modify (8) as

$$\ell(\beta, \gamma) = \sum_{(s,t) \in N \cap W} \left\{Z^{T}(s, t)\beta + \gamma(t)\right\} - \int_D \int_T \exp\{Z^{T}(s, t)\beta + \gamma(t)\}\, dt\, ds. \quad (9)$$

To further parameterize $\gamma(t)$, we propose to approximate it by regression splines (Zhou et al., 1998; Zhu, et al., 2008). For simplicity, we assume the time domain to be $T = [0, 1]$. Let $\kappa_j = j/(J_1 + 1), j = 0, \ldots, J_1 + 1$ be equally spaced knots on T, then we can define $K_1 = J_1 + r_1$ normalized B-spline basis functions of order r_1, which form the basis of a functional space $\mathscr{S}_{K_1}^{r_1}$. The B-spline basis functions are

$$B_j(t) = (\kappa_j - \kappa_{j-r_1})[\kappa_{j-r_1}, \ldots, \kappa_j](\kappa - t)_+^{r_1 - 1}, \quad j = 1, \ldots, K_1,$$

where $[\kappa_{j-r1}, \dots, \kappa_j]\varphi(\kappa)$ denotes the r_1th order divided difference of the function $\varphi(\kappa)$ on $r_1 + 1$ points $\kappa_{j-r1}, \dots, \kappa_j$, $\kappa_j = \kappa_{\min\{\max(j,0), J_1+1\}}$ for $j = 1 - r_1, \dots, K_1$, and $(x)_+ = \max(x, 0)$.

Denote the estimators of $\boldsymbol{\beta}$ and $\gamma(t)$ as

$$\{\hat{\boldsymbol{\beta}}, \hat{\gamma}(t)\} = \arg\max_{(\boldsymbol{\beta}, \gamma) \in \mathbb{R}^d \otimes \mathscr{S}_{K_1}^{r_1}} \ell(\boldsymbol{\beta}, \gamma). \quad (10)$$

Let $\boldsymbol{B}(t) = \{B_1(t), \dots, B_{K1}(t)\}^T$ be the vector of spline basis, and write $\gamma(t) = \boldsymbol{B}^T(t)\boldsymbol{v}$. For convenience of developing asymptotic theory, we denote

$\tilde{\boldsymbol{B}}(t) = \sqrt{K_1}\boldsymbol{B}(t)$, $\mathbb{X}(\boldsymbol{s},t) = \{\boldsymbol{Z}^T(\boldsymbol{s},t)\tilde{B}^T(t)\}^T$, and $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, K_1^{-1/2}v^T)^T$. Then $\hat{\boldsymbol{\theta}}$ is the solution of the estimating equation

$$\boldsymbol{0} = \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) = \sum_{(\boldsymbol{s},t) \in N \cap W} \mathbb{X}(\boldsymbol{s},t) - \int_D \int_T \mathbb{X}(\boldsymbol{s},t)\exp\{\mathbb{X}^T(\boldsymbol{s},t)\boldsymbol{\theta}\}dtd\boldsymbol{s}. \quad (11)$$

The estimating equation can be solved numerically by a Newton-Raphson algorithm, where the integral in the equation is evaluated numerically. Asymptotic properties of these estimators are studied in Section 4. The number of spline basis functions K_1 is often deemed as a tuning parameter in spline smoothing. Selection of this tuning parameter is discussed in Section W.5 of the Supplementary Material.

3.2 Estimation of the eigenvalues and eigenfunctions

As mentioned before, the eigenvalues and eigenfunctions, $\{\omega_j, \psi_j(\cdot)\}$, can be estimated by an eigen-decomposition of the covariance function R_T. By (6), the second-order intensity of the point process across time at a given spatial location is

$$\lambda_{2,\boldsymbol{s}}(t_1, t_2) = \lambda_2(\boldsymbol{s}, \boldsymbol{s}, t_1, t_2) = \lambda(\boldsymbol{s}, t_1)\lambda(\boldsymbol{s}, t_2)\exp\{R_T(t_1, t_2)\}. \quad (12)$$

We will approximate R_T by tensor product splines. Let $\{B_j(t); j = 1, \dots, K_2\}$ be B-spline functions with order r_2 defined on J_2 equally spaced knots on $[0, 1]$. The tensor product spline basis functions are given by $B_{jj'}(t_1, t_2) = B_j(t_1)B_{j'}(t_2), j, j' = 1, \dots, K_2$. Denote $\boldsymbol{B}_{[2]}(t_1, t_2) = (B_{11}, B_{12}, \dots, B_{1K_2}, B_{21}, \dots, B_{K_2K_2})^T(t_1, t_2)$, and the functional space spanned by $\boldsymbol{B}_{[2]}$ as $\mathscr{S}_{[2],K_2}^{r_2}$. Then the spline approximation for the covariance function is $R_T(t_1, t_2) \approx \boldsymbol{B}_{[2]}^T(t_1, t_2)\boldsymbol{b}$.

We estimate $\boldsymbol{b}$ and hence R_T by generalizing the composite likelihood approach of Guan (2006) and Waagepetersen (2007). Let $\lambda_2(\boldsymbol{s}_1, \boldsymbol{s}_2, t_1, t_2; \boldsymbol{\eta})$ be a parametric model for the second-order intensity function of a point process depending on some parameter vector $\boldsymbol{\eta}$. Then, $\boldsymbol{\eta}$ can be estimated by maximizing

$$\ell_c(\boldsymbol{\eta}) = \sum_{(\boldsymbol{s}_1,t_1)\in N\cap W} \sum_{\substack{(\boldsymbol{s}_2,t_2)\in \\ N\cap\{D\otimes T-(\boldsymbol{s}_1,t_1)\}}} w(\boldsymbol{s}_1,\boldsymbol{s}_2,t_1,t_2)\log\left\{\lambda_2(\boldsymbol{s}_1,\boldsymbol{s}_2,t_1,t_2;\boldsymbol{\eta})\right\}$$

$$-\int_D\int_D\int_T\int_T w(\boldsymbol{s}_1,\boldsymbol{s}_2,t_1,t_2)\lambda_2(\boldsymbol{s}_1,\boldsymbol{s}_2,t_1,t_2;\boldsymbol{\eta})\,dt_1\,dt_2\,d\boldsymbol{s}_1\,d\boldsymbol{s}_2, \tag{13}$$

where $D\otimes T - (\boldsymbol{s}_1,t_1) = \{(\boldsymbol{s}_2,t_2): (\boldsymbol{s}_2,t_2)\in D\otimes T, \text{ and } (\boldsymbol{s}_2,t_2) \quad (\boldsymbol{s}_1,t_1)\}$ and $w(\boldsymbol{s}_1,\boldsymbol{s}_2,t_1,t_2)$ is a prespecified weight function. For our purpose, it is sufficient to estimate $\lambda_{2,s}(t_1,t_2)$ as a result of (12). Thus, we may want to consider pairs of events that occurred at the same location. However, for an orderly spatial point process, the probability of observing two events at the same location is zero. Instead, we define a small spatial neighborhood for every event location s, denoted as $D_{s,\delta} = \{\boldsymbol{u}\in D, \|\boldsymbol{u}-\boldsymbol{s}\| < \delta\}$, and consider pairs of the given event and any other events within the neighborhood. This can be achieved by defining the weight function as

$$w(\boldsymbol{s}_1,\boldsymbol{s}_2,t_1,t_2)=I(\|\boldsymbol{s}_1-\boldsymbol{s}_2\|<\delta),$$

where $I(\cdot)$ is an indicator function. We assume that the spatial covariance functions $C_j(\cdot)$'s are continuous at 0, and we choose δ to be small so that $\lambda_2(\boldsymbol{s}_1,\boldsymbol{s}_2,t_1,t_2) \approx \lambda_{2,s_1}(t_1,t_2)$ for any $s_2\in D_{s_1,\delta}$. The role of δ in our estimation procedure will be discussed in Section 4, after developing the asymptotic theory of the proposed covariance estimator, and a practical criterion to choose δ is provided in the online Supplementary Material.

With the above modifications, the composite likelihood criterion in (13) becomes

$$\ell_c(\boldsymbol{\beta},\gamma,R_T) = \sum_{(\boldsymbol{s}_1,t_1)\in N\cap W} \sum_{\substack{(\boldsymbol{s}_2,t_2)\in \\ N\cap\{D_{\boldsymbol{s}_1,\delta}\otimes T-(\boldsymbol{s}_1,t_1)\}}} \log\{\lambda_{2,\boldsymbol{s}_1}(t_1,t_2)\}$$

$$-\int_D\int_{D_{\boldsymbol{s}_1,\delta}}\int_T\int_T \lambda_{2,\boldsymbol{s}_1}(t_1,t_2)\,dt_1\,dt_2\,d\boldsymbol{s}_2\,d\boldsymbol{s}_1, \tag{14}$$

where $D_{\boldsymbol{s}_1,\delta}\otimes T - (\boldsymbol{s}_1,t_1) = \{(\boldsymbol{s}_2,t_2): \|\boldsymbol{s}_2-\boldsymbol{s}_1\| \quad \delta, \text{ and } (\boldsymbol{s}_2,t_2) \quad (\boldsymbol{s}_1,t_1)\}$. We propose to estimate the covariance function as the maximizer of the composite likelihood restricted in the spline space, i.e.,

$$\hat{R}_T = \arg\max_{R_T\in s_{[2],K_2}^{r_2}} \ell_c(\hat{\boldsymbol{\beta}},\hat{\gamma},R_T), \tag{15}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}$ are the estimators defined in (10).

The covariance estimator can be rewritten as $\hat{R}_T(t_1,t_2)=\boldsymbol{B}_{[2]}^T(t_1,t_2)\hat{\boldsymbol{b}}$, where $\boldsymbol{b}$ is the solution of

$$0= \frac{\partial \ell_c}{\partial \boldsymbol{b}}(\hat{\boldsymbol{\beta}}, \hat{\gamma}, \boldsymbol{b}) = \sum_{(\boldsymbol{s}_1, t_1) \in N} \sum_{\substack{(\boldsymbol{s}_2, t_2) \in \\ N \cap \{D_{\boldsymbol{s}_1, \delta} \otimes T - (\boldsymbol{s}_1, t_1)\}}} \boldsymbol{B}_{[2]}(t_1, t_2)$$

$$- \int_D \int_{D_{\boldsymbol{s}_1, \delta}} \int_T \int_T \boldsymbol{B}_{[2]}(t_1, t_2) \hat{\lambda}(\boldsymbol{s}_1, t_1) \hat{\lambda}(\boldsymbol{s}_2, t_2) \exp\left\{\boldsymbol{B}_{[2]}^T(t_1, t_2)\boldsymbol{b}\right\} dt_1 \, dt_2 \, ds_2 \, ds_1, \qquad (16)$$

where $\hat{\lambda(\boldsymbol{s}, t)} = \exp\{\boldsymbol{Z}^T(\boldsymbol{s}, t)\hat{\boldsymbol{\beta}} + \hat{\gamma}(t)\}$. When the neighborhood $D_{\boldsymbol{s}, \delta}$ is sufficiently small and the number of knots of the spline basis is sufficiently large, (16) is an approximately unbiased estimating equation.

The estimates of the eigenvalues and eigenfunctions are obtained by solving the eigen-decomposition problems

$$\int_T \hat{R}_T(t_1, t_2) \hat{\psi}_j(t_1) \, dt_1 = \hat{\omega}_j \hat{\psi}_j(t_2), j = 1, \ldots, p. \quad (17)$$

Since our estimator $\hat{R_T}(\cdot, \cdot)$ is constrained in a functional subspace spanned by tensor products of a spline basis $\boldsymbol{B}(\cdot)$, the estimated eigenfunction function is spanned by the same spline basis. Hence, the functional eigen-decomposition problem in (17) can be translated into a multivariate problem. Notice that our estimator $\hat{R}_T$ is inherently symmetric because the same pairs of events contribute equally in estimating $R_T(t_1, t_2)$ and $R_T(t_2, t_1)$. We can arrange the coefficient vector $\hat{\boldsymbol{b}}$ into a symmetric matrix $\hat{\boldsymbol{G}}$, so that $\hat{R_T}(t_1, t_2) = \boldsymbol{B}^T(t_1)\hat{\boldsymbol{G}}\boldsymbol{B}(t_2)$. Define an inner product matrix $\mathcal{J} = \int_T \boldsymbol{B}(t)\boldsymbol{B}^T(t)dt$, then the eigen-decomposition problem in (17) is equivalent to the multivariate generalized eigenvalue decomposition

$$\hat{\boldsymbol{\phi}}_j^T \mathcal{J} \hat{\boldsymbol{G}} \mathcal{J} \hat{\boldsymbol{\phi}}_j = \hat{\omega}_j, \text{subject to} \hat{\boldsymbol{\phi}}_j^T \mathcal{J} \hat{\boldsymbol{\phi}}_{j'} = I(j = j'), \quad (18)$$

where $I(\cdot)$ is an indicator function. Then, $\hat{\psi_j}(t) = \boldsymbol{B}^T(t)\hat{\varphi_j}, j = 1, \ldots, p.$

In the procedures described above, selecting the tuning parameters K_2 and δ as well as selecting the number of principal components p are important issues, which are addressed in Section W.5 of the Supplementary Material.

3.3 Estimation of the spatial correlation

In the previous section, we estimate the eigenfunctions ψ_j's and eigenvalues ω_j's using pairs of events that occurred within a close distance to avoid the complications of spatial correlation. With ψ_j's and ω_j's consistently estimated, we now estimate the spatial correlation functions using another composite likelihood that includes pairs of events further apart. Suppose the spatial covariance functions are of a parametric form $C_j(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\vartheta}_j)$, where $\boldsymbol{\vartheta}_j$'s are unknown parameters. We will focus on stationary covariance models such as the flexible class of Matérn covariance models (Stein, 1999). Stationarity in space is commonly assumed in spatial statistics including spatio-temporal log Gaussian Cox processes (e.g., Brix and Møller, 2001; Diggle et al., 2005). In what follows, we use $C_j(\boldsymbol{s}_1 - \boldsymbol{s}_2; \boldsymbol{\vartheta}_j)$ instead in order to reflect the assumption of stationarity.

Define $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^T, \ldots, \boldsymbol{\vartheta}_p^T)^T$. To estimate $\boldsymbol{\vartheta}$, we again modify the composite likelihood (13) through the use of a proper weight function w. Specifically, we use

$$w(\boldsymbol{s}_1, \boldsymbol{s}_2, t_1, t_2) = \lambda^{-1}(\boldsymbol{s}_1, t_1)\lambda^{-1}(\boldsymbol{s}_2, t_2)I(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \leq \varrho),$$

where ϱ is a pre-specified spatial distance. By (6),

$$w(\boldsymbol{s}_1, \boldsymbol{s}_2, t_1, t_2)\lambda_2(\boldsymbol{s}_1, \boldsymbol{s}_2, t_1, t_2) = \exp\left\{\sum_{j=1}^{p} C_j(\boldsymbol{s}_1 - \boldsymbol{s}_2; \vartheta_j)\psi_j(t_1)\psi_j(t)\right\}I(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \leq \varrho).$$

Thus, we avoid integrating the covariate process $\boldsymbol{Z}(\boldsymbol{s}, t)$ over the entire spatial temporal domain while evaluating the integral in (13). Let $\ell_{\text{c,spat}}(\boldsymbol{\vartheta}, \boldsymbol{\beta}, \gamma, \boldsymbol{\omega}, \boldsymbol{\psi})$ be the resulting modified composite likelihood, where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_p)^T$ and $\boldsymbol{\psi}(t) = (\psi_1, \ldots, \psi_p)^T(t)$. We substitute $(\boldsymbol{\beta}, \gamma, \boldsymbol{\omega}, \boldsymbol{\psi})$ with their estimators described in Section 3.2 and therefore define the estimator of $\boldsymbol{\vartheta}$ as

$$\hat{\boldsymbol{\vartheta}} = \arg\max_{\boldsymbol{\vartheta}} \ell_{\text{c,spat}}(\boldsymbol{\vartheta}, \hat{\boldsymbol{\beta}}, \hat{\gamma}, \hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\psi}}). \quad (19)$$

The proposed weight function w excludes the pairs of events with a distance greater than ϱ, since the spatial correlation tends to diminish as the spatial lag increases and including events that are too far away may provides little information about the correlation function. The parameter ϱ therein can be considered as a tuning parameter. A reasonable choice of ϱ is the range of the spatial correlation, which can be estimated by fitting a pilot parametric model to the data or by checking the empirical pair correlation function of the point pattern (e.g., Guan, 2008b).

3.4 Prediction of the spatial random effects

To predict the random fields $\xi_k(\boldsymbol{s})$, we use a maximum a posteriori (MAP) predictor as in Møller et al. (1998). For ease of presentation, we assume that the spatial domain D is a rectangle $[0, L]^2$. We partition D into smaller rectangles, $D_{ij} = [(i - 1)/M, i/M] \times [(j - 1)/M, j/M]$, $i, j = 1, \ldots, ML$. We take each D_{ij} sufficiently small so that $\xi_k(\boldsymbol{s})$ is approximately a constant for $\boldsymbol{s} \in D_{ij}$, and denote this value as $\xi_{ij,k}$, for $k = 1, \ldots, p$. Given $\boldsymbol{\xi}_{ij} = (\xi_{ij,1}, \ldots, \xi_{ij,p})^T$, the conditional log-likelihood for the events in $D_{ij} \otimes T$ is

$$
\begin{aligned}
\ell_{ij} = & \sum_{\substack{(\boldsymbol{s},t) \in \\ N \cap \{D_{ij} \otimes T\}}} \left\{ \boldsymbol{Z}^T(\boldsymbol{s}, t)\boldsymbol{\beta} + \mu(t) + \sum_{k=1}^{p} \xi_{ij,k}\psi_k(t) \right\} \\
& - \int_{D_{ij}} \int_T \exp\left\{ \boldsymbol{Z}^T(\boldsymbol{s}, t)\boldsymbol{\beta} + \mu(t) + \sum_{k=1}^{p} \xi_{ij,k}\psi_k(t) \right\} dt d\boldsymbol{s}.
\end{aligned}
\quad (20)
$$

By the model, $\boldsymbol{\xi}_k$ has a prior distribution Normal$(\boldsymbol{0}, \Sigma_k)$, where Σ_k is the covariance matrix for the k-th principal component by interpolating $C_k(\boldsymbol{s}_1 - \boldsymbol{s}_2; \vartheta_k)$ on the discrete grid points.

Collect the grid point values of the *k*th principal component score into $\boldsymbol{\xi}_k = \{\xi_{ij,k}; i, j = 1, \ldots, M\}$, then the log posterior density of $\boldsymbol{\xi} = \{\boldsymbol{\xi}_k; k = 1, \ldots, p\}$ is

$$p(\boldsymbol{\xi}) \propto \sum_{i=1}^{M}\sum_{j=1}^{M}\ell_{ij} - \frac{1}{2}\sum_{k=1}^{p}\boldsymbol{\xi}_k^T {\sum_k}^{-1}\boldsymbol{\xi}_k. \quad (21)$$

We substitute $\boldsymbol{\beta}$, $\psi_k(t)$ and ω_k with their estimators defined above, and $\mu(t)$ with $\hat{\mu(t)} = \hat{\gamma(t)} - 1/2\hat{R_T}(t, t)$. The empirical Bayes estimator $\hat{\boldsymbol{\xi}}$ is then obtained by maximizing the posterior (21). We can also draw samples from the posterior (21) using the Metropolis-adjusted Langevin algorithm (MALA) described in Møller et al. (1998), and estimate the prediction error by the posterior variance.

Choosing the partition in spatial prediction is a compromise between prediction bias and computation feasibility. The latent processes are defined in a continuous space, hence a finer spatial grid leads to smaller bias. On the other hand, using a finer grid increases the dimension of the latent random vector $\boldsymbol{\xi}_k$ and makes it harder to simulate from the posterior distribution in (21). Specifically, a higher dimension of $\boldsymbol{\xi}_k$ make it harder for the Markov chain to mix and, as a result, the Markov chain takes a longer time to converge. In many real applications such as the CTR data considered in this paper, there are natural choices for the partition of the spatial domain, e.g. we used the census tracts to partition the state of Connecticut.

4 Asymptotic properties

To distinguish from other possible values in the parameter space, we denote the true parameters (functions) as β_0, γ_0, R_{T0}, ω_{j0} and ψ_{j0}. We study the asymptotic properties of the proposed estimators under an increasing domain asymptotic framework as in Guan (2006). We consider a sequence of spatial domains D_n with expanding areas, but the time domain T remains fixed. Specifically, we assume

$$C_1 n^2 \leq |D_n| \leq C_2 n^2 \text{ and } C_1 n \leq |\partial D_n| \leq C_2 n \text{ for some } 0 < C_1 < C_2 < \infty, \quad (22)$$

where $|D_n|$ denotes the perimeter of D_n. Condition (22) is satisfied by many commonly encountered shape sequences. For example, let $D \subset (0, 1] \times (0, 1]$ be the interior of a simple closed curve with nonempty interior. If we multiply D by n to obtain D_n, then D_n satisfies (22). This formulation allows for a wide variety of shapes as the sequence of observation windows, including both rectangular and elliptical regions.

For any function $f(\boldsymbol{x})$ defined on a compact set $\mathcal{I}$, where $\mathcal{I} \subset \mathbb{R}$ or $\mathbb{R}^2$, define the supremum and L^2 norm of g to be $\|f\|_\infty = \sup_{\boldsymbol{x} \in I}|f(\boldsymbol{x})|$ and $\|f\| = \{\int_I f^2(\boldsymbol{x})d\boldsymbol{x}\}^{1/2}$, respectively. For any m dimensional vector $\boldsymbol{a}$, define its L^2 norm $\|\boldsymbol{a}\| = (\boldsymbol{a}^T\boldsymbol{a})^{1/2}$, and its L^∞ norm $\|\boldsymbol{a}\|_\infty = \max_{j=1}^{m}|a_j|$. For any real valued $m_1 \times m_2$ matrix $\boldsymbol{A} = (a_{ij})$, define its L^2 norm as $\|\boldsymbol{A}\| = \sup_{\boldsymbol{x} \in \mathbb{R}^{m_2}}\|\boldsymbol{A}\boldsymbol{x}\|/\|\boldsymbol{x}\|$, its L^∞ norm as $\|\boldsymbol{A}\|_\infty = \max_{i=1}^{m_1}\sum_{j=1}^{m_2}|a_{ij}|$, and its Frobenius norm as $\|\boldsymbol{A}\|_F = \{\mathrm{tr}(\boldsymbol{A}^T\boldsymbol{A})\}^{1/2}$.

Theorem 1

Let T be a fixed time domain, D_n satisfies condition (22), then under Assumptions 1-5 in Appendix A,

$$\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \| + \| \hat{\gamma} - \gamma_0 \| = O_p\{K_1^{-r_1+1/2} + (K_1/|D_n|)^{1/2}\}.$$

The convergence rate in Theorem 1 is not optimal. A more detailed study in Theorem 2 below reveals that $\hat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}_0$ in a parametric convergence rate and is asymptotically normal, and $\hat{\gamma}(t)$ converges to $\gamma_0(t)$ with the usual nonparametric asymptotic convergence rate. To facilitate this result, we first define the residual process in the spatio-temporal pattern (Baddeley et al., 2005) as

$$\varepsilon(d\boldsymbol{s}, dt) = N(d\boldsymbol{s}, dt) - \exp\{\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{s}, t)\boldsymbol{\beta}_0 + \gamma_0(t)\} d\boldsymbol{s}\,dt. \quad (23)$$

We also define

$$\boldsymbol{\mu}_Z(t; \boldsymbol{\beta}, \gamma) = E[\boldsymbol{Z}(\boldsymbol{s}, t)\exp\{\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{s}, t)\boldsymbol{\beta} + \gamma(t)\}]/q(t; \boldsymbol{\beta}, \gamma) \quad (24)$$

where $q(t; \boldsymbol{\beta}, \gamma)$ is defined in Assumption 3, and

$$\sum_Z(\boldsymbol{\beta}, \gamma) = E\left[\int_T \{\boldsymbol{Z}(\boldsymbol{s}, t) - \boldsymbol{\mu}_Z(t; \boldsymbol{\beta}, \gamma)\}^{\otimes 2} \exp\{\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{s}, t)\boldsymbol{\beta} + \gamma(t)\} dt\right], \quad (25)$$

where $\boldsymbol{x}^{\otimes 2} = \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$ for any vector $\boldsymbol{x}$.

Theorem 2

Under the same conditions as in Theorem 1, we have the following weak convergence result

$$|D_n|^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \rightarrow Normal\{\boldsymbol{0}, \sum_{Z,0}^{-1}\Omega\sum_{Z,0}^{-1}\},$$

where $\Sigma_{Z,0}$ is a shorthand for $\Sigma_Z(\boldsymbol{\beta}_0, \gamma_0)$ and

$$\Omega = \lim_{n\rightarrow\infty} \frac{1}{|D_n|}\mathrm{var}\left[\int_T \int_{D_n} \{\boldsymbol{Z}(\boldsymbol{s}, t) - \boldsymbol{\mu}_Z(t; \boldsymbol{\beta}_0\gamma_0)\}\varepsilon(d\boldsymbol{s}, dt)\right].$$

A tighter asymptotic convergence rate for $\hat{\gamma}$ is $\| \hat{\gamma} - \gamma_0 \| = O_p(K_1^{1/2}|D_n|^{-1/2} + K_1^{-r_1})$.

For statistical inference, we need to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$. We follow Heinrich and Prokešová (2010) to derive a consistent moment estimator for $\mathrm{cov}(\hat{\boldsymbol{\beta}})$. Details of the derivations are given in the Web Appendix B. We also outline a strategy on how to estimate

the variance of $\hat{\gamma}(\cdot)$, in light of the fact that both $\hat{\beta}$ and $\hat{\gamma}(\cdot)$ are obtained by solving the estimating equation (11).

Next, we study the asymptotic properties of the estimated covariance function and those of the estimated eigenvalues and eigenfunctions. The radius of the local neighborhood in the composite likelihood (14) should depend on n. However, we will continue to use δ for ease of exposition.

Theorem 3

Assume that condition (22) and Assumptions 1-9 in Appendix A are true. Then,

$$\| \hat{R}_T - R_{T0} \| = O_p\{K_2(|D_n||D_\delta|)^{-1/2} + \delta + K_2^{-r_2} + (K_1/|D_n|)^{1/2} + K_1^{-r_1}\}.$$

Theorem 3 implies that the radius parameter δ plays a similar role to the bandwidth used in nonparametric regressions. As such, there is a trade-off between bias and variance when choosing the optimal δ. Specifically, increasing δ will include more data into the estimation and hence reduce the variance of $\hat{R}_T$, but it will increase the bias due to the use of pairs of events that are much further apart; the opposite can be said when decreasing δ. A practical method to select δ is proposed in Section W.5.2.

Following Theorem 1 in Hall and Hosseini-Nasab (2006), the following asymptotic properties for the functional principal component estimators in (17) and (18) are immediate.

Corollary 1

Letting $\Delta_n = \| \hat{R}_T - R_{T0} \|$, then $\sup_j |\hat{\omega}_j - \omega_j| \leq \Delta_n$. If p is finite, define $\omega_{p+1} = 0$. Put $\tau_j = \min_{1 \leq k \leq j}(\omega_k - \omega_{k+1})$, $J = \inf\{j \geq 1 : \omega_j - \omega_{j+1} \leq 2\Delta_n\}$, then $\| \hat{\psi}_j - \psi_j \| \leq C\Delta_n/\tau_j$, for $1 \leq j \leq J - 1$.

Theorems 2 and 3 and Corollary 1 show that our estimators $\hat{\beta}$, $\hat{\gamma}(\cdot)$ and $\{\hat{\omega}_j, \hat{\psi}_j(\cdot); j = 1, \ldots, p\}$ are consistent, and hence by plugging in these consistent estimators the method described in Section 3.3 also provides a consistent estimator for the spatial correlation parameter ϑ. Using the theory in Guan (2006), we have the following corollary.

Corollary 2

Under condition (22), the assumptions in Appendix A and the regularity conditions in Theorem 1 of Guan (2006), $\hat{\vartheta}$ defined by (19) is consistent to the true correlation parameter ϑ.

5 Simulation study

Let the spatial region be $D = [0, 2]^{\otimes 2}$, and the time window be $T = [0, 1]$. We assume that $Z(\boldsymbol{s})$ is a one dimensional covariate, which is generated as an isotropic Gaussian random field on D with an exponential covariance structure. In particular, $cov\{Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)\} = \exp(-\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|/\rho)$, and we set the scale parameter to be $\rho = 0.2$. The random field $X(\boldsymbol{s}, t)$ is generated with $\mu(t) = 3 + 2t^2$ and $p = 2$ principal components, where $(\omega_1, \omega_2) = (2, 1)$, $\psi_1(t) = 1$, and $\psi_2(t) = \sqrt{2}\cos(2\pi t)$. Both principal component scores, $\xi_j(\boldsymbol{s})$, $j = 1, 2$, are generated from Gaussian random fields with isotropic exponential covariance structures $C_j(\boldsymbol{s}_1 - \boldsymbol{s}_2; \vartheta_j)$

$= \omega_j \exp(-\|s_1 - s_2\|/\vartheta_j)$, and the scale parameters ϑ_j are also set to be 0.2. Both $Z(s)$ and $\xi_j(s)$ are simulated on a regular grid with increments 0.01, using the *RandomFields* package in **R**. The events are generated using rejection sampling.

In this setting, the covariance function is $R_T(t_1, t_2) = 2 + 2\cos(2\pi t_1)\cos(2\pi t_2)$. The two principal components have clear interpretations. The first principal component is a random intercept. If $\xi_1(s)$ is high in a location s, the event intensity is high at that location. The second principal component can be interpreted as a periodic random effect. On average, there are 1,661 events in the defined spatio-temporal domain.

The simulation is repeated 200 times, and the proposed model selection and estimation procedures are applied to each simulated data set. We first choose the tuning parameters as described in Section W.5 of the Supplementary Material. The AIC (W.9) picks $K_1 = 10$ for most of the simulated data sets, and the cross-validation procedure in Section W.5.2 chooses $K_2 = 7$ and $\delta = 0.01$ most frequently. Therefore, we fix the value of these tuning parameters for further estimation. Under our choice of δ, we include, on average, one neighboring event for every event in the composite likelihood (14). Next, we apply our second AIC (W.11) to choose the number of principal components, and it chooses the correct number, $p = 2$, of principal components 57% of the time. We find that AIC tends to choose an over-fitted model, and 88% of the time, it chooses the number of principal components to be between 2 and 4. Such an over-fitting tendency is consistent with what has been discovered in the literature. Since under-fitting is usually a more serious problem than over-fitting, the performance of AIC seems satisfactory.

The estimation results are summarized in Figure 1. In panel (a) of Figure 1, we show the box-plots of $\hat{\beta}$, $\hat{\omega_1}$ and $\hat{\omega_2}$. As we can see, $\hat{\beta}$ is almost unbiased to the true value $\beta_0 = 1$, which is consistent with our asymptotic theory. The estimated eigenvalues are slightly biased but nevertheless close to the truth. Although these estimators are consistent, some bias is often reported in FDA literature in a finite-sample setting, see Li and Hsing (2010b). This is true even when direct measurements are made on the curves. In our setting, $X(s, t)$ are latent processes which makes estimation of these parameters considerably harder. In that sense, the behavior of these estimators are reasonable. In panels (b) – (d) of Figure 1, we summarize the estimation results for $\gamma(t)$ and the two eigenfunctions, where we compare the mean, 5% and 95% pointwise percentiles of the functional estimators with the true functions. The plots suggest that these estimators behave reasonably well. We also provide, in panel (e) of Figure 1, the box-plots of $\hat{\vartheta_1}$ and $\hat{\vartheta_2}$, which are the spatial correlation parameters for the two principal components estimated using the composite likelihood method in Section 3.3. The box-plots show that these estimates are reasonably close to the true value 0.2. Since the second principal component is less prominent in the data, its spatial correlation is harder to estimate. Consequently, $\hat{\vartheta_2}$ is more variable then $\hat{\vartheta_1}$.

We also perform spatial prediction for the latent processes $\xi_1(s)$ and $\xi_2(s)$ as described in Section 3.4. Plots of the prediction results in a typical run is provided in Section W.6 of the online Supplementary Material. These predicted maps can provide useful information on hot spatial regions due to clustering.

6 Data Analysis

We apply the proposed methodology to historical cancer incidence records collected by the CTR. The CTR is the oldest cancer registry in the United States. Since the Surveillance, Epidemiology and End Results (SEER) Program was launched by the National Cancer Institute in 1973, it has always been a program participating SEER site. The CTR has reciprocal reporting agreements with cancer registries in all adjacent states (and Florida, a popular winter destination for retirees) to identify Connecticut residents with cancer diagnosed and/or treated in these states. For each identified CTR case, both the date of diagnosis and residential address at the time of diagnosis were recorded, along with a list of demographic and diagnostic variables. The longitude and latitude of a diagnosis are recorded in the Universal Transverse Mercator (UTM) coordinate system.

Pancreatic cancer is the fourth most common cause of cancer-related deaths in both men and women in the United States. We consider the CTR data of 8,230 pancreatic cancer incidences that were diagnosed from 1992 to 2009. Our primary interest is to study the spatio-temporal pattern of pancreatic cancer incidences, after having accounted for heterogeneities in both population density and socioeconomic status (SES) scores at the census block group level. There are 2,616 block groups within the state of Connecticut. The SES score is an aggregated measure to reflect poverty level in a neighborhood, where a higher SES score indicates a more deprived neighborhood (Wang et al., 2009).

Similar to model (1), we assume that the conditional intensity for the cancer incidences is

$$\lambda(\boldsymbol{s}, t) = \lambda_0(\boldsymbol{s}) \exp\{Z(\boldsymbol{s})\boldsymbol{\beta} + X(\boldsymbol{s}, t)\},$$

where $\lambda_0(\boldsymbol{s})$ and $Z(\boldsymbol{s})$ are the population density and SES score at $\boldsymbol{s}$, and $X(\boldsymbol{s}, t)$ is a latent process with the same structure as in (2). We assume that $\lambda_0(\boldsymbol{s})$ and $Z(\boldsymbol{s})$ are constants within a block group.

We first the estimate the parameters in the first order intensity. The AIC in (W.9) picks $K_1 = 9$ cubic B-splines to model the function $\gamma(t)$. We apply the proposed method in Section 3.1 to estimate β and $\gamma(t)$, and use the method described in Section W.4 of the Supplementary Material to estimate the standard error of the estimators. The estimated coefficient for the SES score is $\hat{\beta} = 1.63 \times 10^{-2}$, with standard error 3.87×10^{-3}. We therefore conclude that the SES score is positively associated with the pancreatic cancer rate in a neighborhood. The estimated temporal trend function $\gamma(t)$ and the 95% confidence bands are presented in Figure 2. The plot suggests that the pancreatic cancer rate was increasing over the years in the study period.

To estimate the covariance function R_T, the cross-validation procedure in Section W.5.2 picks $K_2 = 9$ and $\delta = 1000$ UTM units. Therefore, block groups with a UTM distance less than 1000 are considered neighbors, the pancreatic cancer incidences in neighboring block groups are considered neighboring events, and all such pairs of neighboring events are used in the composite likelihood (14). Note that 1000 UTM units is about 1 kilometer, which is a

small distance in the scale of this application. The AIC defined in (W.11) suggests that there are two principal components in $X(s, \cdot)$. The first two eigenvalues are 8.742 and 0.345 which explain a total of 93% of variation in the covariance function. The first two eigenfunctions are shown in the second plot of Figure 2. As we can see, $\hat{\psi_1}(t)$ given by the solid curve is almost a constant over time, indicating that the first principal component score $\xi_1(s)$ is a spatial random intercept. When $\xi_1(s)$ is higher, the cancer rate at s is also higher than the average rate. On the other hand, $\hat{\psi_2}(t)$ represents an increasing trend in time, even though it does not increase linearly. Hence, when $\xi_2(s)$ is higher, the cancer rate at s increases faster than average.

We also model the spatial correlation in $\xi_1(s)$ and $\xi_2(s)$ by the exponential correlation function, and estimate correlation range parameters by the composite likelihood method in Section 3.3. The estimated range is about 2400 UTM units for both principal components. We perform spatial prediction for $\xi_1(s)$ and $\xi_2(s)$ at the census tract level, by simulating samples from the posterior (21). We use the posterior means as the predicted values of the principal component scores, and the posterior standard deviations as the prediction errors.

In the two panels of Figure 3, the predicted values of $\xi_1(s)$ and $\xi_2(s)$ are highlighted by gray levels in maps of Connecticut respectively, where black represents the highest positive values and white represents the lowest negative values. By the interpretation of the two principal components described above, we believe the dark census tracts in panel (a) of Figure 3 have higher pancreatic cancer rates than average, while the dark tracts in panel (b) have higher increasing rate in pancreas cancer than others. The posterior standard deviations of the two latent random fields are also given in the two panels of the Figure 4. These maps help us to understand the uncertainty in the spatial prediction. We perform z-tests on the predicted principal component scores and find that about 27% of census tracts have ξ_1 significant different from 0, however none of the predictions for ξ_2 are significant. The latter result is simply because the relatively large amount of prediction error for ξ_2. We think that the signal in the second principal component is much weaker compared with the first component and the test on a local signal (i.e. at any single census tract) does not have enough power.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix A: Assumptions for the theoretical results

A.1 Notation

For any subset $E \subset \mathbb{R}^2$, let $\mathcal{F}(E)$ be the σ-algebra generated by $N \cap (E \otimes T)$ and $\{\mathbf{Z}(\mathbf{s}, t), \xi_j(\mathbf{s}, t), j = 1, \cdots, p : (\mathbf{s}, t) \in (E \otimes T)\}$. To quantify the spatial dependence, we introduce the strong mixing coefficient (Rosenblatt, 1956),

$$\alpha_{k,l}(h) = \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}(E_1), A_2 \in \mathcal{F}(E_2), |E_1| \leq k, |E_2| \leq l, d(E_1, E_2 \geq h\}, \quad \text{(A.1)}$$

where $d(E_1, E_2)$ denotes the minimal spatial distance between E_1 and E_2.

Define

$$\lambda_4(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4, t_1, t_2, t_3, t_4) = \lim_{\substack{|d\mathbf{s}_j|, |dt_j| \to 0 \\ j=1,\dots,4}} \frac{E\{N(d\mathbf{s}_1, dt_1)N(d\mathbf{s}_2, dt_2)N(d\mathbf{s}_3, dt_3)N(d\mathbf{s}_4, dt_4)\}}{|d\mathbf{s}_1||dt_1||d\mathbf{s}_2||dt_2||d\mathbf{s}_3||dt_3||d\mathbf{s}_4||dt_4|},$$

$$\mathcal{G}_4(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, t_1, t_2, t_3, t_4) = \frac{\lambda_4(\mathbf{s}, \mathbf{s}+\mathbf{u}_1, \mathbf{s}+\mathbf{u}_3, \mathbf{s}+\mathbf{u}_2+\mathbf{u}_3, t_1, t_2, t_3, t_4)}{\lambda(\mathbf{s}, t_1)\lambda(\mathbf{s}+\mathbf{u}_1, t_2)\lambda(\mathbf{s}+\mathbf{u}_3, t_3)\lambda(\mathbf{s}+\mathbf{u}_2+\mathbf{u}_3, t_4)}, \quad \text{(A.2)}$$

and

$$\mathcal{M}(t_1, t_2) = \frac{1}{|D_n||D_\delta|} \int_{D_n} \int_{D_{\mathbf{s}_1,\delta}} \lambda(\mathbf{s}_1, t_1)\lambda(\mathbf{s}_2, t_2) d\mathbf{s}_2 d\mathbf{s}_1, \quad \text{(A.3)}$$

where $|D_\delta|$ is the common area for all $D_{\mathbf{s},\delta}$, e.g. $|D_\delta| = \pi\delta^2$ if $D_{\mathbf{s},\delta}$ is a disc centered at $\mathbf{s}$. Put $\mu_{\mathcal{M}}(t_1, t_2) = E\{\mathcal{M}(t_1, t_2)\}$.

A.2 Assumptions

We make the following assumptions in order to derive our asymptotic theory.

- *Assumption 1.* Define the class of Hölder continuous functions on $[0, 1]$ as
  $C_1^{r,a}[0,1] = \{f : \sup_{t_1 t_2 \in [0,1]} |f^{(r)}(t_1) - f^{(r)}(t_2)|/|t_1 - t_2|^a < \infty\}$ for some nonnegative integer r and some $a > 0$. We assume $\gamma_0 \in C_1^{r_1,a}[0,1]$ where $r_1 \geq 2$ is the order of the spline estimator and $a > 0$.

- *Assumption 2.* We assume that the processes N, $\mathbf{Z}$ and ξ_j, $j = 1, \dots, p$, are strictly stationary in $\mathbf{s}$ and satisfy the following mixing condition (Guyon, 1995):

  $$\sum_{m \geq 1} m\alpha_{k,1}(m) < \infty \text{ and } \alpha_{k,\infty}(m) = o(m^{-2}) \text{ for some } k > 0 \text{ and } m > 0.$$

  We also assume that $E[\mathbf{Z}(\mathbf{s}, t) \exp\{\mathbf{Z}^{\mathrm{T}}(\mathbf{s}, t)\beta + \gamma(t)\}]^C < \infty$ for some $C > 2$, and $\sup^{t_1, t_2} \int_{\mathbb{R}^2} |\mathcal{G}_2(\mathbf{u}, t_1, t_2) - 1| d\mathbf{u} < \infty$.

- *Assumption 3.* Define $q(t; \beta, \gamma) = E[\exp\{\mathbf{Z}^{\mathrm{T}}(\mathbf{s}, t)\beta + \gamma(t)\}]$, which does not depend on $\mathbf{s}$ by the stationarity of $\mathbf{Z}(\mathbf{s}, t)$ for any fixed t. Assume that $0 < C_1 \leq \min_t q(t; \theta, \gamma)$

$\max_t q(t; \boldsymbol{\theta}, \gamma)$ $C_2 < \infty$, for all

$(\boldsymbol{\beta}, \gamma) \in \mathcal{N}_{C_0} = \{(\boldsymbol{b}, g): \boldsymbol{b} \in \mathbb{R}^d, g \in C_1^{r,a}[0,1], \| \boldsymbol{b} - \boldsymbol{\beta}_0 \| + \| g - \gamma_0 \|_\infty < C_0 \}$.

- *Assumption 4.* Let $\boldsymbol{\mu}_Z(t; \boldsymbol{\beta}, \gamma)$ and $\Sigma_Z(\boldsymbol{\beta}, \gamma)$ be defined in (24) and (25), and $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ be the functionals to take the maximum and minimum eigenvalues of a matrix. Assume that $\boldsymbol{\mu}_Z(t; \boldsymbol{\beta}, \gamma) \in C_1^{r_1, a}[0,1]$ and $0 < C_3$ $\lambda_{\min}(\Sigma_Z)$ $\lambda_{\max}(\Sigma_Z)$ $C_4 < \infty$, for all $(\boldsymbol{\beta}, \gamma) \in \mathcal{N}_{C_0}$. We also assume $\boldsymbol{\mu}_Z$ is continuous in $\boldsymbol{\beta}$ and γ, with $\| \boldsymbol{\mu}_Z(\bullet; \boldsymbol{\beta}_1, \gamma_1) - \boldsymbol{\mu}_Z(\bullet; \boldsymbol{\beta}_2, \gamma_2) \|$ $C(\| \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 \| + \| \gamma_1 - \gamma_2 \|)$, and similar for Σ_Z.

- *Assumption 5.* Let $C > 0$ be a genuine constant. $K = C|D_n|^{\upsilon_1}$, $1/(4r_1) < \upsilon_1 < 1/2$.

- *Assumption 6.* The spatial covariance functions $C_j(\boldsymbol{u}) = \text{cov}\{\xi_j(\boldsymbol{s}), \xi_j(\boldsymbol{s} + \boldsymbol{u})\}$ are Lipschitz continuous at $\boldsymbol{0}$. There exists a constant M_0 such that $|C_j(\boldsymbol{u}) - \omega_j|$ $M_0 \|\boldsymbol{u}\|$, for $j = 1, \ldots, p$.

- *Assumption 7.* Define the class of bivariate Hölder continuous functions on $[0, 1]^{\otimes 2}$ as $C_2^{r,a} = \left\{ f : \sup_{\boldsymbol{t}_1, \boldsymbol{t}_2 \in [0,1]^{\otimes 2}} |f^{(u_1, u_2)}(\boldsymbol{t}_1) - f^{(u_1, u_2)}(\boldsymbol{t}_2)| / \| \boldsymbol{t}_1 - \boldsymbol{t}_2 \|^a < \infty, \text{ for } u_1, u_2 \right.$ $0, u_1 + u_2$ $r \}$. We assume that $R_{T_0} \in C_2^{r_2, a}$, where r_2 2 is the order of the tensor product spline function and $a > 0$.

- *Assumption 8.* Assume that $0 < C_5$ $, \mu_m(t_1, t_2)$ $C_6 < \infty$ for all $t_1, t_2 \in T$,

$$\sup_{\boldsymbol{u}_1, \boldsymbol{u}_2, t_1, \ldots, t_4} \int_{\mathbb{R}^2} |\mathcal{G}_4(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, t_1, t_2, t_3, t_4) - \mathcal{G}_2(\boldsymbol{u}_1, t_1, t_2) \mathcal{G}_2(\boldsymbol{u}_2, t_3, t_4)| d\boldsymbol{u}_3 < \infty.$$

- *Assumption 9.* Assume that $\delta \to 0$, $|D_n||D_\delta| \to \infty$ and $K_2 = C(|D_n||D_\delta|)^{u_2}$, $1/(4r_2) < u_2 < 1/2$.
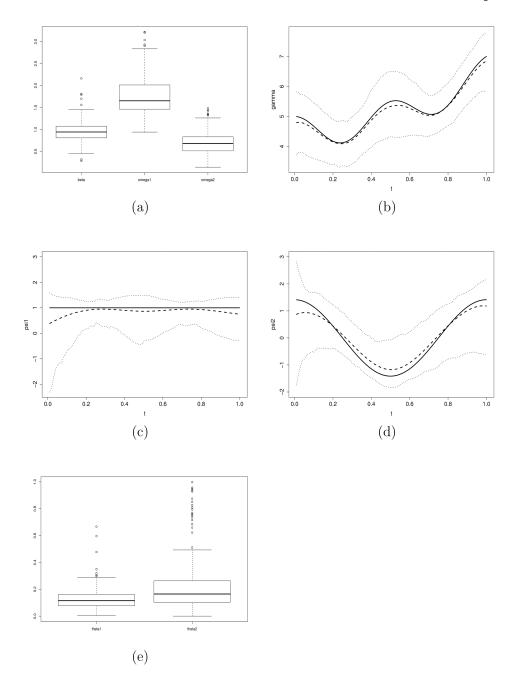
References

Ash, RB.; Gardner, MF. Topics in stochastic processes. Academic press; 1975.

Baddeley AJ, Møller J, Waagepetersen R. Non- and semi- parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica. 2000; 54:329–350.

Baddeley A, Turner R, Møller J, Hazelton M. Residual analysis for spatial point processes (with discussion). Journal of the Royal Statistical Society, Series B. 2005; 67:617–666.

Banerjee, S.; Gelfand, AE.; Carlin, BP. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC; 2003.

Brix A, Diggle PJ. Spatiotemporal prediction for log-Gaussian Cox processes. Journal of the Royal Statistical Society, Series B. 2001; 63:823–841.

Brix A, Møller J. Space-time multi type log Gaussian Cox processes with a view to modelling weeds. Scandinavian Journal of Statistics. 2001; 28:471–488.

Bouzas PR, Valderrama M, Aguilera AM, Ruiz-Fuentes N. Modeling the mean of a doubly stochastic Poisson process by functional data analysis. Computational Statistics & Data Analysis. 2006; 50:2655–2667.

Demko S. Inverses of band matrices and local convergence of spline projection. SIAM Journal on Numerical Analysis. 1977; 14:616–619.

Di C, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel functional principal component analysis. The Annals of Applied Statistics. 2009; 3:458–488. [PubMed: 20221415]

Diggle P, Rowlingson B, Su T. Point process methodology for on-line spatio-temporal disease surveillance. EnvironMetrics. 2005; 16:423–434.

Diggle PJ. Spatio-temporal point processes, partial likelihood, foot and mouth disease. Statistical Methods in Medical Research. 2006; 15:325–336. [PubMed: 16886734]
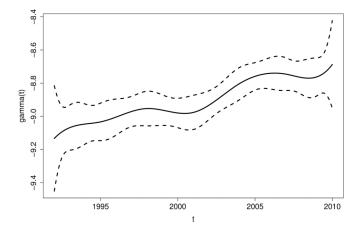
Gelfand AE, Schmidt AM, Banerjee S, Sirmans CF. Nonstationary multivariate process modeling through spatially varying coregionalization. TEST. 2004:263–312.

Guan Y. A composite likelihood approach in fitting spatial point process models. Journal of the American Statistical Association. 2006; 101(476):1502–1512.

Guan Y, Loh JM. A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. Journal of the American Statistical Association. 2007:1377–1386.

Guan Y. On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. Journal of the American Statistical Association. 2008a; 103:1238–1247.

Guan Y. A KPSS test for stationarity for spatial point processes. Biometrics. 2008b; 64:800–806. [PubMed: 18218065]

Guan Y. On nonparametric variance estimation for second-order statistics of in-homogeneous spatial point processes with known parameteric intensity form. Journal of the American Statistical Association. 2009; 104:1482–1491. [PubMed: 20336182]

Guan Y, Sherman M, Calvin JA. A nonparametric test for spatial isotropy using subsampling. Journal of the American Statistical Association. 2004; 99:810–821.

Guan Y, Wang H. Sufficient dimension reduction for spatial point processes directed by Gaussian random fields. Journal of the Royal Statistical Society, Series B. 2010:367–387.

Guyon, X. Random fields on a network: modeling, statistics, and applications. Springer-Verlage; New York: 1995.

Hall P, Hosseini-Nasab M. On properties of functional principal components analysis. Journal of the Royal Statistical Society, Series B. 2006; 68:109–126.

Hall P, Müller HG, Wang JL. Properties of principal component methods for functional and longitudinal data analysis. Annals of Statistics. 2006; 34:1493–1517.

Hall P, Müller HG, Yao F. Modelling sparse generalized longitudinal observations with latent Gaussian processes. Journal of the Royal Statistical Society, Series B. 2008; 70:703–723.

Heinrich L, Prokešová M. On estimating the asymptotic variance of stationary point processes. Methodology and Computing in Applied Probability. 2010; 12(3):451–471.

Huang J, Yang L. Identification of nonlinear additive autoregressive models. Journal of the Royal Statistical Society, Series B. 2004; 66:463–477.

Illian, J.; Benson, E.; Crawford, J.; Staines, H. Case studies in spatial point process modeling, vol. 185 of Lecture Notes in Statistics. Springer; New York: 2006. Principal component analysis for spatial point processes –assessing the appropriateness of the approach in an ecological context; p. 135-150.

Li Y, Hsing T. Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. Annals of Statistics. 2010a; 38:3028–3062.

Li Y, Hsing T. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. Annals of Statistics. 2010b; 38:3321–3351.

Møller J, Syversveen AR, Waagepetersen RP. Log-gaussian cox processes. Scandinavian Journal of Statistics. 1998; 25:451–482.

Møller J, Waagepetersen R. Modern statistics for spatial point processes. Scandinavian Journal of Statistics. 2007; 34:643–684.

Ramsay, JO.; Silverman, BW. Functional Data Analysis. 2nd. Springer-Verlag; New York: 2005.

Rosenblatt M. A central limit theorem and a strong mixing condition. Proceedings of the National Academy of Science. 1956; 42:43–47.

Schoenberg FP. Multidimensional residual analysis of point process models for earthquake occurrences. Journal of the American Statistical Association. 2003; 98:789–795.

Schoenberg FP. Consistent parametric estimation of the intensity of a spatialtemporal point process. Journal of Statistical Planning and Inference. 2005; 128:79–93.

Schumaker, LL. Spline Functions. Wiley; New York: 1981.

Stein, ML. Interpolation of Spatial Data. Springer; New York: 1999.

Stone C. The dimensionality reduction principle for generalized additive models. Annals of Statistics. 1986; 14:590–606.

Stone C. The use of polynomial splines and their tensor products in multivariate function estimation. Annals of Statistics. 1994; 22:118–184.

Tanaka U, Ogata Y, Stoyan D. Parameter estimation and model selection for Neymann-Scott point processes. Biometrical Journal. 2007; 49:1–15.

Waagepetersen RP. An estimating function approach to inference for inhomogeneous Neyman-Scott Processes. Biometrics. 2007; 63:252–258. [PubMed: 17447951]

Wang R, Gross CP, Halene S, Ma X. Neighborhood socioeconomic status influences the survival of elderly patients with myelodysplastic syndromes in the Unite States. Cancer Causes and Control. 2009; 20(8):1369–1376. [PubMed: 19455395]

Wu S, Müller HG, Zhang Z. Functional data analysis for point processes with rare events. Statistica Sinica. 2013; 23:1–23.

Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association. 2005a; 100:577–590.

Yao F, Müller HG, Wang JL. Functional linear regression analysis for longitudinal data. Annals of Statistics. 2005b; 33:2873–2903.

Zhou L, Huang J, Martinez JG, Maity A, Baladandayuthapani V, Carroll RJ. Reduced rank mixed effects models for spatially correlated hierarchical functional data. Journal of the American Statistical Association. 2010; 105:390–400. [PubMed: 20396628]

Zhou S, Shen X, Wolfe DA. Local asymptotics for regression splines and confidence regions. Annals of Statistics. 1998; 26(5):1760–1782.

Zhu Z, Fung WK, He X. On the asymptotics of marginal regression splines with longitudinal data. Biometrika. 2008; 95(4):907–917.

Figure 1.

Estimation results in the simulation study. Panel (a) shows the boxplots of $\hat{\beta}, \hat{\omega_1}$ and $\hat{\omega_2}$. Panels (b) – (d) show the estimation results for $\hat{\gamma}(t)$, $\hat{\psi_1}(t)$ and $\hat{\psi_2}(t)$ respectively, where the solid curve in each panel is the true curve, the dashed curve is the mean of the estimator, and the two dotted curves are the pointwise 5% and 95% percentiles of the estimator. Panel (e) shows box plots of the estimated spatial correlation parameters ϑ_1 and ϑ_2 for the two principal components.
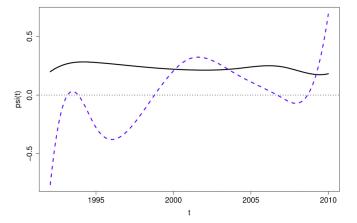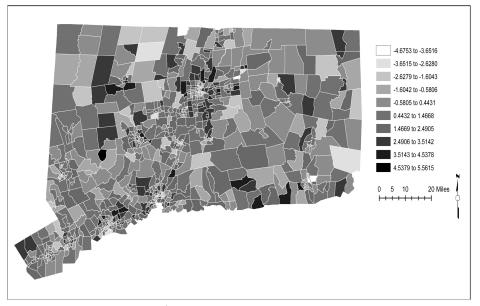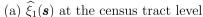
Figure 2.
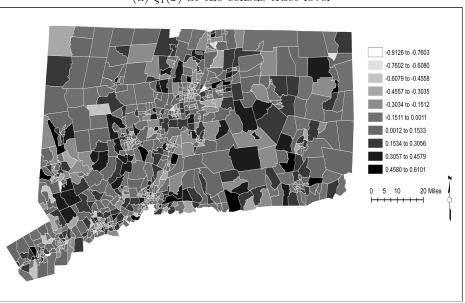Estimation results for the Connecticut Tumor Registry data. The first plot is the estimated temporal trend $\hat{\gamma}(t)$ (the dashed curves are the 95% confidence band), the second plot shows the first two estimated eigenfunctions (the solid curve is the $\hat{\psi_1}(t)$ and the dashed curve is $\hat{\psi_2}(t)$).

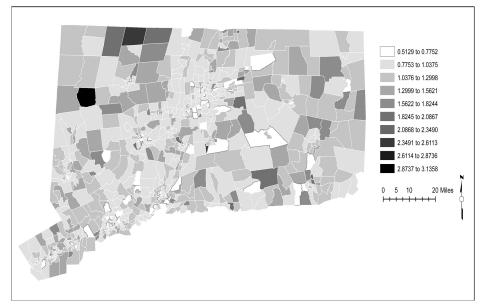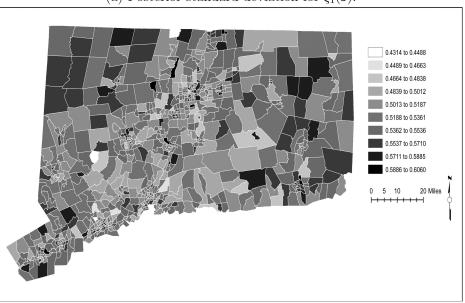(a) $\widehat{\xi}_1(\boldsymbol{s})$ at the census tract level

(b) $\widehat{\xi}_2(\boldsymbol{s})$ at the census tract level

Figure 3.
Estimated scores for the first two principal components in the Connecticut Tumor Registry data. The principal component scores are estimated at census tract level and highlighted by grey levels on the map of Connecticut.

(a) Posterior standard deviation for $\xi_1(\boldsymbol{s})$.

(b) Posterior standard deviation for $\xi_2(\boldsymbol{s})$.

Figure 4.
Estimated prediction errors for the first two principal components in the Connecticut Tumor Registry data. The prediction errors are the square root of the posterior variance of the scores at census tract level.