Modeling Disease Incidence Data with Spatial and Spatio-Temporal Dirichlet Process Mixtures

Athanasios Kottas[*,1], **Jason A. Duan**[2], and **Alan E. Gelfand**[3]

[1] Department of Applied Mathematics and Statistics, 1156 High Street, University of California, Santa Cruz, CA 95064, USA
[2] School of Management, 135 Prospect Street, Yale University, New Haven, CT 06520, USA
[3] Institute of Statistics and Decision Sciences, Box 90251, Duke University, Durham, NC 27708, USA

Received 1 November 2006, revised 12 June 2007, accepted 27 June 2007

Summary

Disease incidence or mortality data are typically available as rates or counts for specified regions, collected over time. We propose Bayesian nonparametric spatial modeling approaches to analyze such data. We develop a hierarchical specification using spatial random effects modeled with a Dirichlet process prior. The Dirichlet process is centered around a multivariate normal distribution. This latter distribution arises from a log-Gaussian process model that provides a latent incidence rate surface, followed by block averaging to the areal units determined by the regions in the study. With regard to the resulting posterior predictive inference, the modeling approach is shown to be equivalent to an approach based on block averaging of a spatial Dirichlet process to obtain a prior probability model for the finite dimensional distribution of the spatial random effects. We introduce a dynamic formulation for the spatial random effects to extend the model to spatio-temporal settings. Posterior inference is implemented through Gibbs sampling. We illustrate the methodology with simulated data as well as with a data set on lung cancer incidences for all 88 counties in the state of Ohio over an observation period of 21 years.

Key words: Areal unit spatial data; Dirichlet process mixture models; Disease mapping; Dynamic spatial process models; Gaussian processes.

1 Introduction

Data on disease incidence (or mortality) are typically available as rates or summary counts for contiguous geographical regions, e.g., census tracts, post or zip codes, districts, or counties, and collected over time. Hence, though cases occur at point locations (residences), the available responses are associated with entire subregions in the study region. We denote the disease incidence counts (number of cases) by y_{it}, where $i = 1, \ldots, n$ indexes the regions B_i, and $t = 1, \ldots, T$ indexes the time periods. In practice, we may have covariate information associated with the region, e.g., percent African American, median family income, percent with some college education. In some cases, though we only know the areal unit into which a case falls, we may have covariate information associated with the case, e.g., sex, race, age, previous comorbidities. Moreover, any of this covariate information could be time dependent. We devote Section 2.3 below to a discussion of how to accommodate such information in our modeling framework. However, the focus here is on flexible modeling of areal unit spatial random effects and so, to avoid obscuring our primary contribution, we do not consider covariates elsewhere.

* Corresponding author: e-mail: thanos@ams.ucsc.edu, Phone: +831 459 5536, Fax: +831 459 4829

A primary inferential objective in the analysis of disease incidence data is summarization and explanation of spatial and spatio-temporal patterns of disease (*disease mapping*); also of interest is spatial smoothing and temporal prediction (forecasting) of disease risk. The field of *spatial epidemiology* has grown rapidly in the past fifteen years with the introduction of spatial and spatio-temporal hierarchical models; see, e.g., Elliott et al. (2000), and Banerjee, Carlin and Gelfand (2004) for reviews and further references.

Working with counts, the typical assumption (for rare diseases) is that the y_{it}, conditionally on parameters R_{it}, are independent $\text{Po}(y_{it} \mid E_{it}R_{it})$ (we will write $\text{Po}(\cdot \mid m)$ for the Poisson probability mass function/distribution with mean m). Here, E_{it} is the expected disease count, and R_{it} is the relative risk, for region i at time period t. (Below we will use an alternative and, we assert, preferable, specification, writing $n_{it}p_{it}$ for the Poisson mean, where n_{it} is the specified number of individuals at risk in region i at time t and p_{it} is the corresponding disease rate.) With R^* denoting an overall disease rate, E_{it} is obtained as R^*n_{it}, using either *external* or *internal* standardization, the former developing R^* from reference tables, the latter computed from the given data set, e.g., $R^* = \sum_{i,t} y_{it} / \sum_{i,t} n_{it}$. The relative risks are explained through different types of random effects. For instance, a specification with random effects additive in space and time is $\log R_{it} = \mu_{it} + u_i + v_i + \delta_t$, where μ_{it} is a component for the regional covariates (e.g., $\mu_{it} = \boldsymbol{x}'_{it}\beta$ for regression coefficients β), u_i are regional random effects (typically, the u_i are assumed i.i.d. $N(0, \sigma_u^2)$), v_i are spatial random effects, and δ_t are temporal effects (say, with an autoregressive prior).

The most commonly used prior for the v_i is based on some form of a conditional autoregressive (CAR) structure (e.g., Clayton and Kaldor, 1987; Besag, York and Mollie, 1991; Bernardinelli, Clayton and Montomoli, 1995; Waller et al., 1997; Knorr-Held and Besag, 1998). For instance, the widely-used specification of Besag et al. (1991) is characterized through local dependence structure by considering for each region i a set, ϑ_i, of *neighbors*, which, for example, can be defined as all regions contiguous to region i. Then the (improper) joint prior *density* for the v_i is built from the prior full conditionals $v_i \mid \{v_j : j \neq i\}$. These are normal distributions with mean $m_i^{-1} \sum_{j \in \vartheta_i} v_j$ and variance λm_i^{-1}, where λ is a precision hyperparameter and m_i is the number of neighbors of region i.

A different hierarchical formulation, discussed in Böhning, Dietz and Schlattmann (2000), involves replacing the normal mixing distribution with a discrete distribution taking values φ_j (that represent the relative risks for k underlying time-space clusters) with corresponding probabilities p_j, $j = 1, \ldots, k$. Hence, marginalizing over the random effects, the distribution for each region i and time period t emerges as a discrete Poisson mixture, $\sum_{j=1}^{k} p_j \, \text{Po}(y_{it} \mid E_{it}\varphi_j)$. See, also, Militino, Ugarte and Dean (2001) for use of such discrete Poisson mixtures in the simpler setting without a temporal component. In this setting, related is the Bayesian work of Knorr-Held and Rasser (2000) based on spatial partition structures, which divide the study region into a number of clusters (i.e., sets of contiguous regions) with constant relative risk, assuming, in the prior model, random number, size, and location for the clusters.

When spatio-temporal interaction is sought, the additive form $v_i + \delta_t$ is replaced by v_{it}. The latter has been modeled using independent CAR models over time, dynamically with independent CAR innovations, or as a CAR in space and time (see, e.g., Waller et al., 1997; Knorr-Held, 2000).

Rather than modeling the spatial dependence through the finite set of spatial random effects, one for each region, an alternative prior specification arises by modeling the underlying continuous-space relative risk (or rate) surface and obtaining the induced prior models for the relative risks (or rates) through aggregation of the continuous surface. This approach is less commonly used in modeling for disease incidence data (among the exceptions are Best, Ickstadt and Wolpert, 2000, and Kelsall and Wakefield, 2002). However, it, arguably, offers a more coherent modeling framework, since by modeling the underlying continuous surfaces, it avoids the dependence of the prior model on the data collection procedure, i.e., the number, shapes, and sizes of the regions chosen in the particular study. It replaces the specification of a proximity matrix, which spatially connects the subregions, with a covariance function, which directly models dependence between arbitrary pairs of locations (and induces a covariance between arbitrary subregions using block averaging).

In this paper, we follow this latter approach, our main objective being to develop a flexible nonparametric model for the needed risk (or rate) surfaces. In particular, denote by D the union of all regions in the study area and let $z_{t,D} = \{z_t(s) : s \in D\}$ be the latent disease rate surface for time period t, on the logarithmic scale. Hence, $z_t(s) = \log p_t(s)$, where $p_t(s)$ is the probability of disease at time t and spatial location s. (With rare diseases, the logarithmic transformation is practically equivalent to the logit transformation). We propose spatial and spatio-temporal nonparametric prior models for the vectors of log-rates $z_t = (z_{1t}, \ldots, z_{nt})$, which we define by block averaging the surfaces $z_{t,D}$ over the regions B_i, i.e., $z_{it} = |B_i|^{-1} \int_{B_i} z_t(s)\, ds$, where $|B_i|$ is the area for region B_i. We develop the spatial prior model by block averaging a Gaussian process (GP) to the areal units determined by the regions B_i, and then centering a Dirichlet process (DP) prior (Ferguson, 1973) around the resulting n-variate normal distribution. We show that the model is equivalent to the prior model that is built by block averaging a spatial DP (Gelfand, Kottas and MacEachern, 2005). To model the z_t, we can specify them to be independent replications under the DP or we can add a further dynamic level to the model with z_t evolving from z_{t-1} through independent DP innovations. We use the former in our simulation example in Section 4.1; we use the latter with our real data example in Section 4.2.

With regard to the existing literature, our approach is, in spirit, similar to that of Kelsall and Wakefield (2002) where an isotropic GP was used for the log-relative risk surface. However, as exemplified in Section 2.2, we relax both the isotropy and the Gaussianity assumptions. In addition, we develop modeling for disease incidence data collected over space and time. Moreover, as we show in Section 2.1, our nonparametric model has a more general mixture representation than that of Böhning et al. (2000) as it incorporates spatial dependence and it allows, through the DP prior, a random number of mixture components.

The plan of the paper is as follows. Section 2 develops the methodology for the spatial and spatio-temporal modeling approaches. Section 3 discusses methods for posterior inference with more details given in the Appendix. Section 4 includes illustrations motivated by a previously analyzed dataset involving lung cancers for the 88 counties in Ohio over a period of 21 years. In fact, in Section 4.1 we develop a simulated dataset for these counties which is analyzed using both our modeling specification as well as a GP model, revealing the benefit of our approach. We also reanalyze a subset of the original data in Section 4.2. Finally, Section 5 provides a summary and discussion of possible extensions.

2 Bayesian Nonparametric Models for Disease Incidence Data

The spatial model is discussed in Section 2.1. Section 2.2 reviews spatial DPs and demonstrates how their use provides foundation for the modeling approach. Section 2.3 discusses how to include different types of covariate information. Lastly, Section 2.4 develops a nonparametric spatio-temporal modeling framework.

2.1 The spatial prior probability model

Here, we treat the log-rate surfaces $z_{t,D}$ as independent realizations (over time) from a stochastic process over D. We build the model by viewing the counts y_{it} and the log-rates z_{it} as aggregated versions of underlying (continuous-space) stochastic processes. The finite-dimensional distributional specifications for the y_{it} and the z_{it} are induced through block averaging of the corresponding spatial surfaces.

For the first stage of our hierarchical model, we use the standard Poisson specification working with the $n_{it}p_{it}$ form for the mean. We note that this parametrization seems preferable to the $E_{it}R_{it}$ form, since it avoids the need to develop the E_{it} through standardization; the overall log-rate emerges as the intercept in our model. Inference for the R_{it} is readily available from our models using posterior samples of the p_{it} with any particular choice of E_{it}.

Hence, the y_{it} are assumed conditionally independent, given $z_{it} = \log p_{it}$, from Po $(y_{it} \mid n_{it} \exp(z_{it}))$. This specification can be derived through aggregation of an underlying Cox process under the following assumptions and approximations. For the time period t, assume that the disease incidence cases, over region D, are distributed according to a Cox process with intensity function $n_t(s)\, p_t(s)$, where $\{n_t(s) : s \in D\}$ is the population density surface and $p_t(s)$ is the disease rate at time t and location s. If we assume a uniform population density over each region at each time period (this assumption is, implicitly, present in standard modeling approaches for disease mapping), we can write $n_t(s) = n_{it} |B_i|^{-1}$ for $s \in B_i$. Hence, aggregating the Cox process over the regions B_i, we obtain, conditionally on $\boldsymbol{z}_{t,D}$, that the y_{it} are independent, and each y_{it} follows a Poisson distribution with mean $\int_{B_i} n_t(s)\, p_t(s)\, \mathrm{d}s = n_{it} p^*_{it}$, where $p^*_{it} = |B_i|^{-1} \int_{B_i} p_t(s)\, \mathrm{d}s$. If we approximate the distribution of the p^*_{it} with the distribution of the $\exp(z_{it})$, we can write $y_{it} \mid z_{it} \overset{\text{ind.}}{\sim} \text{Po}\,(y_{it} \mid n_{it} \exp(z_{it}))$ for the first stage distribution. We note that the stochastic integral for p^*_{it} is not accessible analytically. Moreover, using Monte Carlo integration to approximate the p^*_{it} is computationally infeasible (Short, Carlin and Gelfand, 2005). Also, Kelsall and Wakefield (2002) use a similar approximation working with relative risk surfaces. Brix and Diggle (2001) do so as well, using a stochastic differential equation to model $p_t(s)$.

To build the prior model for the log-rates $\boldsymbol{z}_t$, we begin with the familiar form, $z_t(s) = \mu_t(s) + \theta_t(s)$, for the log-rate surfaces $\boldsymbol{z}_{t,D}$. Here, $\mu_t(s)$ is the mean structure and $\boldsymbol{\theta}_{t,D} = \{\theta_t(s) : s \in D\}$ are spatial random effects surfaces. As discussed in Section 2.3, the surfaces $\{\mu_t(s) : s \in D\}$ can be elaborated through covariate surfaces over D. In the absence of such covariate information, we might set $\mu_t(s) = \mu_0$, for all t, and use a normal prior for μ_0. Alternatively, we could set $\mu_t(s) = \mu_0 + \mu_t$, where the μ_t are i.i.d. $N(0, \sigma^2_\mu)$ with random hyperparameter σ^2_μ. In what follows for the spatial prior model, we illustrate with the common μ_0 specification.

To develop the model for the spatial random effects, first, let the $\boldsymbol{\theta}_{t,D}$, $t = 1, \ldots, T$, given σ^2 and ϕ, be independent realizations from a mean-zero isotropic GP with variance σ^2 and correlation function $\rho(||s - s'||; \phi)$ (say, $\rho(||s - s'||; \phi) = \exp(-\phi\, ||s - s'||)$ as in the examples in Section 4). Hence by aggregating over the regions B_i, we obtain $z_{it} = \mu_0 + \theta_{it}$, where $\theta_{it} = |B_i|^{-1} \int_{B_i} \theta_t(s)\, \mathrm{d}s$ is the block average of the surface $\boldsymbol{\theta}_{t,D}$ over region B_i. The induced distribution for $\boldsymbol{\theta}_t = (\theta_{1t}, \ldots, \theta_{nt})$ is a mean-zero n-variate normal with covariance matrix $\sigma^2 R_n(\phi)$, where the (i, j)-th element of $R_n(\phi)$ is given by $|B_i|^{-1} |B_j|^{-1} \int_{B_i} \int_{B_j} \rho(||s - s'||; \phi)\, \mathrm{d}s\, \mathrm{d}s'$. Next, consider a DP prior for the spatial random effects $\boldsymbol{\theta}_t$ with precision parameter $\alpha > 0$ and centering (base) distribution $N_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi))$ (we will write $N_p(\cdot \mid \lambda, \Sigma)$ for the p-variate normal density/distribution with mean vector λ and covariance matrix Σ). We denote this DP prior by DP $(\alpha, N_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi)))$. The choice of the DP in this context yields data-driven deviations from the normality assumption for the spatial random effects.

Note that the above structure implies for the vector of counts $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})$ a nonparametric Poisson mixture model given by $\int \prod_{i=1}^n \text{Po}\,(y_{it} \mid n_{it} \exp(\mu_0 + \theta_{it}))\, \mathrm{d}G(\boldsymbol{\theta}_t)$, where the mixing distribution $G \sim \text{DP}\,(\alpha, N_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi)))$. Under this mixture specification, the distribution for the vectors of log-rates, $\boldsymbol{z}_t = \mu_0 1_n + \boldsymbol{\theta}_t$, is discrete (a property induced by the discreteness of DP realizations), a feature of the model that could be criticized. Moreover, although posterior simulation is feasible, it requires more complex MCMC algorithms than the standard Gibbs sampler for DP-based hierarchical models (e.g., Escobar and West, 1998). Thus, to overcome both concerns above, we replace the DP prior for the $\boldsymbol{z}_t$ with a DP mixture prior,

$$\boldsymbol{z}_t \mid \mu_0, \tau^2, G \overset{\text{ind.}}{\sim} \int N_n(\boldsymbol{z}_t \mid \mu_0 1_n + \boldsymbol{\theta}_t, \tau^2 I_n)\, \mathrm{d}G(\boldsymbol{\theta}_t), \quad G \sim \text{DP}\,(\alpha, N_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi)))\,.$$

That is, we now write $z_{it} = \mu_0 + \theta_{it} + u_{it}$, with u_{it} i.i.d. $N(0, \tau^2)$. Introduction of a heterogeneity effect in addition to the spatial effect is widely employed in the disease mapping literature dating to Besag et al. (1991), though with concerns about balancing priors for the effects (see, e.g., Banerjee et al., 2004, and references therein). Here, in responding to the above concerns, we serendipitously achieve this benefit.

The mixture model for the $\boldsymbol{y}_t$ now becomes $f(\boldsymbol{y}_t \mid \mu_0, \tau^2, G) = \int \prod_{i=1}^n p(y_{it} \mid \mu_0, \tau^2, \theta_{it})\, \mathrm{d}G(\theta_t)$, where $p(y_{it} \mid \mu_0, \tau^2, \theta_{it}) = \int \text{Po}\,(y_{it} \mid n_{it} \exp(z_{it}))\, N(z_{it} \mid \mu_0 + \theta_{it}, \tau^2)\, \mathrm{d}z_{it}$ is a Poisson-lognormal mix-

ture. Equivalently, the model can be written in the following semiparametric hierarchical form

$$\begin{array}{lll} y_{it} \mid z_{it} & \overset{\text{ind.}}{\sim} & \text{Po}\,(y_{it} \mid n_{it} \exp(z_{it})), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T \\ z_{it} \mid \mu_0, \theta_{it}, \tau^2 & \overset{\text{ind.}}{\sim} & N(z_{it} \mid \mu_0 + \theta_{it}, \tau^2), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T \\ \boldsymbol{\theta}_t \mid G & \overset{\text{i.i.d.}}{\sim} & G, t = 1, \ldots, T \\ G \mid \sigma^2, \phi & \sim & \text{DP}\,(\alpha, N_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi))) \,. \end{array} \tag{1}$$

The model is completed with independent priors $p(\mu_0)$, $p(\tau^2)$ and $p(\sigma^2)$, $p(\phi)$ for μ_0, τ^2, and for the hyperparameters σ^2, ϕ of the DP prior (the precision parameter α is fixed). Note that the spatio-temporal extension of model (1), developed in Section 2.4, involves temporal effects μ_t, which are added to $\mu_0 + \theta_{it}$ in the second-stage specification.

In practice, we work with a marginalized version of model (1),

$$p(\mu_0)\, p(\tau^2)\, p(\sigma^2)\, p(\phi)\, p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T \mid \sigma^2, \phi) \prod_{i=1}^{n} \prod_{t=1}^{T} \text{Po}\,(y_{it} \mid n_{it} \exp(z_{it}))\, N(z_{it} \mid \mu_0 + \theta_{it}, \tau^2)\,, \tag{2}$$

which is obtained by integrating the random mixing distribution G over its DP prior (Blackwell and MacQueen, 1973). The resulting joint prior distribution for the $\boldsymbol{\theta}_t$, $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T \mid \sigma^2, \phi)$, is given by

$$N_n(\boldsymbol{\theta}_1 \mid 0, \sigma^2 R_n(\phi)) \prod_{t=2}^{T} \left\{ \frac{\alpha}{\alpha + t - 1}\, N_n(\boldsymbol{\theta}_t \mid \mathbf{0}, \sigma^2 R_n(\phi)) + \frac{1}{\alpha + t - 1} \sum_{j=1}^{t-1} \delta_{\boldsymbol{\theta}_j}(\boldsymbol{\theta}_t) \right\}, \tag{3}$$

where δ_a denotes a point mass at a. Hence, the $\boldsymbol{\theta}_t$ are generated according to a Pólya urn scheme; $\boldsymbol{\theta}_1$ arises from the base distribution, and then for each $t = 2, \ldots, T$, $\boldsymbol{\theta}_t$ is either set equal to $\boldsymbol{\theta}_j$, $j = 1, \ldots, t-1$, with probability $(\alpha + t - 1)^{-1}$ or is drawn from the base distribution with the remaining probability.

Note that we have defined the prior model for the spatial random effects $\boldsymbol{\theta}_t$ by starting with a GP prior for the surfaces $\boldsymbol{\theta}_{t,D}$, block averaging the associated GP realizations over the regions to obtain the $N_n(\mathbf{0}, \sigma^2 R_n(\phi))$ distribution, and, finally, centering a DP prior for the $\boldsymbol{\theta}_t$ around this n-variate normal distribution. This approach might suggest that the DP prior is dependent, in an undesirable fashion, on the specific choice of the regions (e.g., their number and size). The next section addresses this potential criticism by connecting the model in (1) with the spatial DP (SDP) from Gelfand et al. (2005).

2.2 Formulation of the model through spatial Dirichlet processes

We first review SDPs, which provide nonparametric prior models for random fields $\boldsymbol{W}_D = \{W(s) : s \in D\}$ over a region $D \subseteq R^d$. Central to their development is the DP constructive definition (Sethuraman, 1994). According to this definition, a random distribution arising from DP (α, G_0), where G_0 denotes the base distribution, is almost surely discrete and admits the representation $\sum_{\ell=1}^{\infty} \omega_\ell \delta_{\varphi_\ell}$, where $\omega_1 = z_1$, $\omega_\ell = z_\ell \prod_{r=1}^{\ell-1} (1 - z_r)$, $\ell = 2, 3, \ldots$, with $\{z_r, r = 1, 2, \ldots\}$ i.i.d. from Beta$(1, \alpha)$, and, independently, $\{\varphi_\ell, \ell = 1, 2, \ldots\}$ i.i.d. from G_0. Under the standard setting for DPs, φ_ℓ is either scalar or vector valued.

To model $\boldsymbol{W}_D$, φ_ℓ is extended to a realization of a random field, $\varphi_{\ell,D} = \{\varphi_\ell(s) : s \in D\}$, and thus G_0 is extended to a spatial stochastic process G_{0D} over D. A stationary GP is used for G_{0D}. The resulting SDP provides a (random) distribution for $\boldsymbol{W}_D$, with realizations G_D given by $\sum_{\ell=1}^{\infty} \omega_\ell \delta_{\varphi_{\ell,D}}$. The interpretation is that for any collection of spatial locations in D, say, $(s_1, \ldots, s_M)$, G_D induces a random probability measure $G^{(M)}$ on the space of distribution functions for $(W(s_1), \ldots, W(s_M))$. In fact, $G^{(M)} \sim \text{DP}\,(\alpha, G_0^{(M)})$, where $G_0^{(M)}$ is the M-variate normal distribution for $(W(s_1), \ldots, W(s_M))$ induced by G_{0D}. It can be shown that the random process G_D yields non-Gaussian finite dimensional

distributions, has nonconstant variance, and is nonstationary, even though it is centered around a stationary GP G_{0D}.

In practice, modeling with SDPs requires some form of replication from the spatial process (although missingness across replicates can be handled). Assuming T replicates, the data can be collected in vectors $\boldsymbol{y}_t = (y_t(s_1), \ldots, y_t(s_n))'$, $t = 1, \ldots, T$, where $(s_1, \ldots, s_n)$ are the observed locations. Working with continuous real-valued measurements, the SDP is used as a prior for the spatial random effects surfaces, say, $\zeta_{t,D} = \{\zeta_t(s) : s \in D\}$, in the standard hierarchical spatial modeling framework, $Y_t(s) = \mu_t(s) + \zeta_t(s) + \varepsilon_t(s)$. Here, $\varepsilon_t(s)$ are i.i.d. $N(0, \tau^2)$, and $\mu_t(s)$ is the mean structure. For instance, with X_t a $p \times n$ matrix of covariate values (whose (i, j)-th element is the value of the i-th covariate at the j-th location for the t-th replicate) and β a $p \times 1$ vector of regression coefficients, we could write $X_t'\beta$ for the mean structure associated with $\boldsymbol{y}_t$. Hence, the $\boldsymbol{y}_t$, given β, τ^2, and $G^{(n)}$, are independent from the DP mixture model $\int N_n(\boldsymbol{y}_t \mid X_t'\beta + \zeta_t, \tau^2 I_n)\, \mathrm{d}G^{(n)}(\zeta_t)$, where $\zeta_t = (\zeta_t(s_1), \ldots, \zeta_t(s_n))$, $G^{(n)} \sim \mathrm{DP}(\alpha, G_0^{(n)})$ (induced by the SDP prior for the $\zeta_{t,D}$), with $G_0^{(n)}$ an n-variate normal (induced at $(s_1, \ldots, s_n)$ by the base GP of the SDP prior). Details for posterior inference and spatial prediction can be found in Gelfand et al. (2005).

The hierarchical nature of the modeling framework enables extensions by replacing the first stage Gaussian distribution (the kernel of the DP mixture) with any other distribution. In this spirit, and returning to the setting for disease incidence data, the SDP can be proposed as the prior for the spatial random effects surfaces $\boldsymbol{\theta}_{t,D}$ to replace the isotropic GP prior that we used to build the DP model in Section 2.1. Therefore, now the model is developed by assuming that the $\boldsymbol{\theta}_{t,D}$, $t = 1, \ldots, T$, given G_D, are independent from G_D, where G_D, given σ^2 and ϕ, follows a SDP prior with precision parameter α and base process $G_{0D} = \mathrm{GP}(\boldsymbol{0}, \sigma^2\rho(|| s - s' ||; \phi))$ (i.e., the same isotropic GP used in Section 2.1).

Next, we block average the $\boldsymbol{\theta}_{t,D}$ over the regions B_i with respect to their distribution that results by marginalizing G_D over its SDP prior. Recall that for any set of spatial locations s_r, $r = 1, \ldots, M$, over D, the random distribution $G^{(M)}$ induced by G_D follows a DP with base distribution $G_0^{(M)}$ induced by G_{0D}. Because we can choose M arbitrarily large and the set of locations s_r to be arbitrarily dense over D, using the Pólya urn characterization for the DP, we obtain that, marginally, the $\boldsymbol{\theta}_{t,D}$ arise according to the following Pólya urn scheme. First, $\boldsymbol{\theta}_{1,D}$ is a realization from G_{0D}, and then, for each $t = 2, \ldots, T$, $\boldsymbol{\theta}_{t,D}$ is identical to $\boldsymbol{\theta}_{j,D}$, $j = 1, \ldots, t-1$, with probability $(\alpha + t - 1)^{-1}$ or is a new realization from G_{0D} with probability $\alpha(\alpha + t - 1)^{-1}$. Hence, if we block average $\boldsymbol{\theta}_{1,D}$, we obtain the $N_n(\boldsymbol{0}, \sigma^2 R_n(\phi))$ distribution for $\boldsymbol{\theta}_1$. Then, working with the conditional specification for $\boldsymbol{\theta}_{2,D}$ given $\boldsymbol{\theta}_{1,D}$, if we block average $\boldsymbol{\theta}_{2,D}$, $\boldsymbol{\theta}_2$ arises from $N_n(\boldsymbol{0}, \sigma^2 R_n(\phi))$ with probability $\alpha(\alpha + 1)^{-1}$ or $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1$ with probability $(\alpha + 1)^{-1}$. Analogously, for any $t = 2, \ldots, T$, the induced conditional prior $p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{t-1}, \sigma^2, \phi)$ is a mixed distribution with point masses at $\boldsymbol{\theta}_j$, $j = 1, \ldots, t-1$, and continuous piece given by the $N_n(\boldsymbol{0}, \sigma^2 R_n(\phi))$ distribution; the corresponding weights are $(\alpha + t - 1)^{-1}$, $j = 1, \ldots, t-1$, and $\alpha(\alpha + t - 1)^{-1}$.

Thus, the prior distribution for the $\boldsymbol{\theta}_t$ in (3) can be obtained by starting with a SDP prior for the $\boldsymbol{\theta}_{t,D}$ (centered around the same isotropic GP prior used in Section 2.1 for the $\boldsymbol{\theta}_{t,D}$), and then block averaging the (marginal) realizations from the SDP prior over the regions. As in Section 2.1, we extend $z_t = \mu_0 1_n + \boldsymbol{\theta}_t$ to $z_t = \mu_0 1_n + \boldsymbol{\theta}_t + \mathbf{u}_t$, where the $\mathbf{u}_t$ are independent $N_n(\boldsymbol{0}, \tau^2 I_n)$. Hence, model (2) is equivalent to the marginal version of the model above, i.e., with G_D marginalized over its SDP prior.

This argument provides formal justification for model (1)–(3). The SDP is a nonparametric prior for the continuous-space stochastic process of spatial random effects; regardless of the number and geometry of regions chosen to partition D, it induces the appropriate corresponding version of the model in (2).

2.3 Introducing covariates

Here, we indicate how covariate information can be incorporated in the context of the model given in (1). Our approach is to consider how we would handle the idealized situation of point-referenced case/

non-case data and then propagate the effect of the assumptions and approximations in Section 2.1 (the approach is similar in spirit to that of Wakefield and Shaddick, 2006). In particular, illustrating with a single covariate surface $\{X_t(s) : s \in D\}$, suppose $z_t(s) = \beta_{0t} + \beta_{1t}X_t(s) + \theta_t(s)$. If $X_t(s)$ is an areal unit level covariate, i.e., $X_t(s) = X_{it}$, for all $s \in B_i$, then $p^*_{it} = \exp(\beta_{0t} + \beta_{1t}X_{it})\ |B_i|^{-1}\int_{B_i} \exp(\theta_t(s))\,\mathrm{d}s$. So, for such covariates, no approximation beyond that of Section 2.1 is required.

Next, associate with each of the n_{it} individuals at risk in areal unit i at time t an (unknown) location s_{ij}, $j = 1, 2, \ldots, n_{it}$, and covariate level $X_t(s_{ij})$ (suppressing time t in the notation for s_{ij}). At each location there is a Bernoulli trial with probability $p_t(X_t(s_{ij}))$. (Here, we write $p_t(X_t(s_{ij}))$, instead of $p_t(s_{ij})$, to emphasize the dependence on the covariate.) Since incidence rates are usually very small, we can envision a Poisson approximation to the sum of the n_{it} Bernoulli trials in areal unit i at time t with expectation equal to $p_{it} = \sum_{j=1}^{n_{it}} p_t(X_t(s_{ij}))$.

Suppose that $X_t(s)$ is categorical, in fact, for convenience, binary. Then, though we do not know where they occur, we do know that n_{0it} of the $X_t(s_{ij})$ are 0 and n_{1it} of the $X_t(s_{ij})$ are 1. So, in the absence of spatial effects, $\sum_{j=1}^{n_{it}} p_t(X_t(s_{ij})) = n_{0it}p_t(0) + n_{1it}p_t(1) = n_{it}p^*_{it}$ where $p^*_{it} = n_{it}^{-1}(n_{0it}p_t(0) + n_{1it}p_t(1))$. With spatial effects and with locations assigned at random, we obtain $\sum_{j=1}^{n_{it}} p_t(X_t(s_{ij})) = \sum_{\{s_{ij}:X_t(s_{ij})=0\}} \exp(\beta_{0t} + \theta_t(s_{ij})) + \sum_{\{s_{ij}:X_t(s_{ij})=1\}} \exp(\beta_{0t} + \beta_{1t} + \theta_t(s_{ij}))$. Again, we know the number of 0 s and 1 s but can only assume they are randomly assigned to the s_{ij}. Hence, for $\ell = 0, 1$, $\sum_{\{s_{ij}:X_t(s_{ij})=\ell\}} \exp(\theta_t(s_{ij})) \approx n_{it}^{-1}n_{\ell it} \sum_{j=1}^{n_{it}} \exp(\theta_t(s_{ij})) \approx n_{\ell it}\,|B_i|^{-1}\int_{B_i} \exp(\theta_t(s))\,\mathrm{d}s$, and, thus, $\sum_{j=1}^{n_{it}} p_t(X_t(s_{ij})) \approx n_{it}p^*_{it}$, with

$$p^*_{it} = \frac{n_{0it}}{n_{it}} \exp(\beta_{0t})\,|B_i|^{-1}\int_{B_i} \exp(\theta_t(s))\,\mathrm{d}s + \frac{n_{1it}}{n_{it}} \exp(\beta_{0t} + \beta_{1t})\,|B_i|^{-1}\int_{B_i} \exp(\theta_t(s))\,\mathrm{d}s\,.$$

Finally, making the same approximation (i.e., $\exp(\theta_{it}) \approx |B_i|^{-1}\int_{B_i} \exp(\theta_t(s))\,\mathrm{d}s$), we can write $p^*_{it} \approx \exp(\beta_{0t} + \theta_{it})\{1 + n_{it}^{-1}n_{1it}[\exp(\beta_{1t}) - 1]\} \approx \exp(\beta_{0t} + \theta_{it})[1 + n_{it}^{-1}n_{1it}\beta_{1t}] \approx \exp(\beta_{0t} + n_{it}^{-1}n_{1it}\beta_{1t} + \theta_{it})$.

Lastly, with a continuous covariate, we may envision two scenarios – (i) that it is available for each of the n_{it} individuals at risk in areal unit i at time t or (ii) more generally, that it is available as a surface known over the entire study region. Again, the quantity of interest is $\sum_{j=1}^{n_{it}} p_t(X_t(s_{ij})) = \sum_{j=1}^{n_{it}} \exp(\beta_{0t} + \beta_{1t}X_t(s_{ij}) + \theta_t(s_{ij})) = n_{it}p^*_{it}$ where $p^*_{it} = n_{it}^{-1} \exp(\beta_{0t}) \sum_{j=1}^{n_{it}} \exp(\beta_{1t}X_t(s_{ij}) + \theta_t(s_{ij}))$. In case (i), let $V_{it} = n_{it}^{-1}\sum_{j=1}^{n_{it}} X_t(s_{ij})$ while in case (ii) let $V_{it} = |B_i|^{-1}\int_{B_i} X_t(s)\,\mathrm{d}s$; under our assumptions, in either case, V_{it} can be calculated. Then, as earlier, we approximate the distribution of p^*_{it} by the distribution of $\exp(z_{it})$. In either case, we obtain $p^*_{it} \approx \exp(\beta_{0t} + \beta_{1t}V_{it} + \theta_{it})$.

2.4 A spatio-temporal modeling framework

To extend the spatial model of Section 2.1 to a spatio-temporal setting, we cast our modeling in the form of a dynamic spatial process model (see Gelfand, Banerjee and Gamerman, 2005, for parametric hierarchical modeling in this context, and for related references). We now view the log-rate process $\boldsymbol{z}_{t,D} = \{z_t(s) : s \in D\}$ as a temporally evolving spatial process.

To develop a dynamic formulation, we begin by writing $z_t(s) = \mu_0 + \mu_t + \theta_t(s)$ and add temporal structure to the model through *transition equations* for the $\theta_t(s)$, say, $\theta_t(s) = \nu\theta_{t-1}(s) + \eta_t(s)$, where, in general, $|\nu| < 1$, and the innovations $\eta_{t,D} = \{\eta_t(s) : s \in D\}$ are independent realizations from a spatial stochastic process. We can now define the nonparametric prior for the block averages $\eta_{it} = |B_i|^{-1}\int_{B_i} \eta_t(s)\,\mathrm{d}s$ of the $\eta_{t,D}$ surfaces following the approach of Section 2.1 or, equivalently, of Section 2.2. Proceeding with the latter, we assume that the $\eta_{t,D}$, given G_D, are independent from G_D, and assign a SDP prior to G_D with parameters α and $G_{0D} = \mathrm{GP}(\boldsymbol{0}, \sigma^2\rho(||s - s'||;\phi))$. Marginalizing G_D over its prior, the induced prior, $p(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_T \mid \sigma^2, \phi)$, for the $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{nt})$ is given by (3) (with $\boldsymbol{\eta}_t$ replacing $\boldsymbol{\theta}_t$). Block averaging the surfaces in the transition equations, we obtain $\boldsymbol{\theta}_t = \nu\boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t$, where $\boldsymbol{\theta}_{t-1} = (\theta_{1,t-1}, \ldots, \theta_{n,t-1})$. (We set $\boldsymbol{\theta}_1 = \boldsymbol{\eta}_1$, i.e., $\boldsymbol{\theta}_0 = 0$.) Adding, as before, the i.i.d. $N(0, \tau^2)$

terms to the z_{it}, we obtain the following general form for the spatio-temporal hierarchical model

$$\begin{array}{lcl} y_{it} \mid z_{it} & \overset{\text{ind.}}{\sim} & \text{Po}\,(y_{it} \mid n_{it} \exp(z_{it})), \quad i = 1, \ldots, n, \;\; t = 1, \ldots, T \\ z_{it} \mid \mu_0, \mu_t, \theta_{it}, \tau^2 & \overset{\text{ind.}}{\sim} & N(z_{it} \mid \mu_0 + \mu_t + \theta_{it}, \tau^2), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T \\ \boldsymbol{\theta}_t & = & \nu \boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t \\ \boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_T \mid \sigma^2, \phi & \sim & p(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_T \mid \sigma^2, \phi)\,. \end{array} \tag{4}$$

The specification for the μ_t will depend on the particular application. For instance, the μ_t could be i.i.d., say, from a $N(0, \sigma_\mu^2)$ distribution (with random σ_μ^2), or they could be explained through a parametric function, say, a polynomial trend, $\sum_{j=1}^{m} \beta_j t^j$. In what follows, we work with the former specification, which is implemented in the example of Section 4.2.

3 Posterior Inference and Prediction

Regarding the spatial model of Section 2.1, as is evident from expression (3), the DP prior induces a clustering in the $\boldsymbol{\theta}_t$ (in their prior and hence also in the posterior for model (2)). Let T^* be the number of distinct $\boldsymbol{\theta}_t$ in $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T)$ and denote by $\boldsymbol{\theta}^* = \{\boldsymbol{\theta}_j^* : j = 1, \ldots, T^*\}$ the vector of distinct values. Defining the vector of configuration indicators, $\boldsymbol{w} = (w_1, \ldots, w_T)$, such that $w_t = j$ if and only if $\boldsymbol{\theta}_t = \boldsymbol{\theta}_j^*$, $(\boldsymbol{\theta}^*, \boldsymbol{w}, T^*)$ yields an equivalent representation for $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T)$. Denote by $\boldsymbol{\psi}$ the vector that includes $(\boldsymbol{\theta}^*, \boldsymbol{w}, T^*)$ and all other parameters of model (2). Draws from the posterior $p(\boldsymbol{\psi} \mid \text{data})$, where $\text{data} = \{(y_{it}, n_{it}) : i = 1, \ldots, n, t = 1, \ldots, T\}$, can be obtained using the Gibbs sampler discussed in the Appendix.

The multivariate density estimate for the vector of log-rates associated with the subregions B_i is given by the posterior predictive density for a *new* $\boldsymbol{z}_0 = (z_{10}, \ldots, z_{n0})$,

$$p(\boldsymbol{z}_0 \mid \text{data}) = \int\int N_n(\boldsymbol{z}_0 \mid \mu_0 \mathbf{1}_n + \boldsymbol{\theta}_0, \tau^2 I_n)\, p(\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}^*, \boldsymbol{w}, T^*, \sigma^2, \phi)\, p(\boldsymbol{\psi} \mid \text{data})\, \mathrm{d}\boldsymbol{\theta}_0\, \mathrm{d}\boldsymbol{\psi}\,. \tag{5}$$

Here, $\boldsymbol{\theta}_0 = (\theta_{10}, \ldots, \theta_{n0})$ is the vector of spatial random effects corresponding to $\boldsymbol{z}_0$, and

$$p(\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}^*, \boldsymbol{w}, T^*, \sigma^2, \phi) = \frac{\alpha}{\alpha + T}\, N_n(\boldsymbol{\theta}_0 \mid \mathbf{0}, \sigma^2 R_n(\phi)) + \frac{1}{\alpha + T} \sum_{j=1}^{T^*} T_j \delta_{\boldsymbol{\theta}_j^*}(\boldsymbol{\theta}_0)\,, \tag{6}$$

where T_j is the size of the j-th cluster $\boldsymbol{\theta}_j^*$. Hence, the model has the capacity to capture, through the mixing in the $\boldsymbol{\theta}_j^*$, non-standard features in the distribution of log-rates over the regions.

Turning to the spatio-temporal model of Section 2.4, let $\boldsymbol{\psi} = (\mu_0, \tau^2, \nu, \sigma^2, \phi, \sigma_\mu^2, \boldsymbol{z}, \boldsymbol{\eta}, \boldsymbol{\mu})$ be the parameter vector corresponding to model (4), where $(\boldsymbol{z}, \boldsymbol{\eta}, \boldsymbol{\mu}) = \{(\boldsymbol{z}_t, \boldsymbol{\eta}_t, \mu_t) : t = 1, \ldots, T\}$. We have

$$\begin{aligned} p(\boldsymbol{\psi} \mid \text{data}) \propto\; & p(\mu_0)\, p(\tau^2)\, p(\nu)\, p(\sigma^2)\, p(\phi)\, p(\sigma_\mu^2)\, p(\boldsymbol{\eta} \mid \sigma^2, \phi) \prod_{t=1}^{T} N(\mu_t \mid 0, \sigma_\mu^2) \\ & \times \prod_{t=1}^{T} N_n\left(\boldsymbol{z}_t \mid (\mu_0 + \mu_t)\, \mathbf{1}_n + \sum_{\ell=1}^{t} \nu^{t-\ell} \boldsymbol{\eta}_\ell, \tau^2 I_n\right) \prod_{i=1}^{n} \prod_{t=1}^{T} \text{Po}\,(y_{it} \mid n_{it} \exp(z_{it}))\,. \end{aligned}$$

The Gibbs sampler given in the Appendix can be used to obtain draws from $p(\boldsymbol{\psi} \mid \text{data})$. For instance, of interest might be inference for the log-rates z_{it} corresponding to specific time periods t and/or regions i. Moreover, of interest is temporal forecasting for disease rates at future time points. In particular, the posterior forecast density for the vector of log-rates $\boldsymbol{z}_{T+1}$ at time $T+1$,

$$\begin{aligned} p(\boldsymbol{z}_{T+1} \mid \text{data}) = \int\int\int & N_n(\boldsymbol{z}_{T+1} \mid (\mu_0 + \mu_{T+1})\, \mathbf{1}_n + \textstyle\sum_{\ell=1}^{T+1} \nu^{T+1-\ell} \boldsymbol{\eta}_\ell, \tau^2 I_n)\, N(\mu_{T+1} \mid 0, \sigma_\mu^2) \\ & \times p(\boldsymbol{\eta}_{T+1} \mid \boldsymbol{\eta}, \sigma^2, \phi)\, p(\boldsymbol{\psi} \mid \text{data})\, \mathrm{d}\mu_{T+1}\, \mathrm{d}\boldsymbol{\eta}_{T+1}\, \mathrm{d}\boldsymbol{\psi}\,, \end{aligned}$$

where $p(\boldsymbol{\eta}_{T+1} \mid \boldsymbol{\eta}, \sigma^2, \phi)$ can be expressed as in (6) by replacing $\boldsymbol{\theta}_0$ with $\boldsymbol{\eta}_{T+1}$ and using the, analogous to $(\boldsymbol{\theta}^*, \boldsymbol{w}, T^*)$, clustering structure in the $(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_T)$.

Figure 1 Map of the 88 counties in the state of Ohio. The scale of the x and y axis indicates longitude and latitude, respectively.

4 Data Illustrations

Our data consists of the number of annual lung cancer deaths in each of the 88 counties of Ohio from 1968 to 1988. The population of each county is also recorded. Figure 1 depicts the geographical locations and neighborhood structure of the 88 counties in Ohio.

For both models (1) and (4) we work with an exponential correlation function, $\rho(||s - s'||; \phi) = \exp(-\phi\,||s - s'||)$. For both data examples, we use a discrete uniform prior for ϕ with values in $[0.001, 0.1]$; σ^{-2} and τ^{-2} have gamma $(1, 1)$ priors; μ_0 is assigned a normal prior with mean 0 and large variance (there was very little sensitivity to choices between 10^2 and 10^8 for the prior variance); and α is set equal to 1 (results were practically identical under $\alpha = 5$ and $\alpha = 10$). Finally, the results of Section 4.2 under model (4) are based on a gamma$(1, 1)$ prior for σ_μ^{-2}, and a uniform prior for ν on $(-1, 1)$.

4.1 Simulation example

We illustrate the fitting of our spatial model in (1)–(3) with a simulated data set. We use exactly the same geographical structure with the 88 Ohio counties, but generate the areal incidence rate from a two-component mixture of multivariate normal distributions whose correlation matrix is calculated by block averaging isotropic GPs. The GPs cover the entire area of Ohio. The induced correlation matrix of the 88 blocks is computed by Monte Carlo integration. The simulated counts are obtained as follows. For $i = 1, \ldots, 88$ and $t = 1, \ldots, T$ (with $T = 40$), we first generate z_{it} independent $N(\mu + \theta_{it}, \tau^2)$ and, then, y_{it} independent Po $(n_i \exp(z_{it}))$, where n_i is the population of county i in 1988 (based on the aggregated data over all gender/race/age groups). The distribution of the spatial random effects $\boldsymbol{\theta}_t = (\theta_{1t}, \ldots, \theta_{nt})$ arises through a mixture of two block-averaged GPs. In particular, for $\ell = 1, 2$, let $\boldsymbol{\theta}^{(\ell)} = (\theta_1^{(\ell)}, \ldots, \theta_n^{(\ell)}) \sim N_n((-1)^\ell\, \mu_\theta 1_n, \sigma_\ell^2 R)$, with the (i, j)-th element of the correlation matrix R given by $|B_i|^{-1}|B_j|^{-1}\int_{B_i}\int_{B_j} \exp(-\phi\,||s - s'||)\,\mathrm{d}s\,\mathrm{d}s'$. Then, each $\boldsymbol{\theta}_t$ is independently

sampled from $0.5\boldsymbol{\theta}^{(1)} + 0.5\boldsymbol{\theta}^{(2)}$. The values of the parameters are $\mu = -6.5$, $\mu_\theta = 0.5$, $\sigma_1^2 = \sigma_2^2 = 1/32$, $\tau^2 = 1/256$, and $\phi = 0.05$. Under these choices, marginally, each θ_{it} has a bimodal distribution of the form $0.5\,N(-\mu_\theta, \sigma_1^2) + 0.5\,N(\mu_\theta, \sigma_2^2)$. This simulation experiment was designed to illustrate the inferential scope of our modeling framework, and to compare with a related parametric model (discussed below). For real epidemiological data, such a scenario might arise due to unknown or unobservable genetic predispositions, which could render a portion of the population more vulnerable to a particular disease than the rest of the population.

The Bayesian goodness of fit is illustrated with posterior predictive densities for the log-rates, which are estimated using (5). In Figure 2 (left and middle columns) we compare the true densities of the model from which we simulated the data with the posterior predictive densities under model (1), for four selected counties. They are "Delaware" and "Franklin" in central Ohio, "Hamilton" in southwest, and "Stark" in northeast (see Figure 1). "Franklin" includes Columbus and "Hamilton" includes Cincinnati so these are highly populated counties. "Delaware" is more suburban and "Stark" is very rural. We also plot posterior predictive densities from a parametric model based on a GP $(\mathbf{0}, \sigma^2 \exp(-\phi\,||s - s'||))$ for the spatial random effects surfaces. This specification results as a limiting version of model (1) (for $\alpha \to \infty$) where the $\boldsymbol{\theta}_t$, given σ^2 and ϕ, are i.i.d. $N_n(\mathbf{0}, \sigma^2 R_n(\phi))$. The SDP model clearly outperforms the GP model.

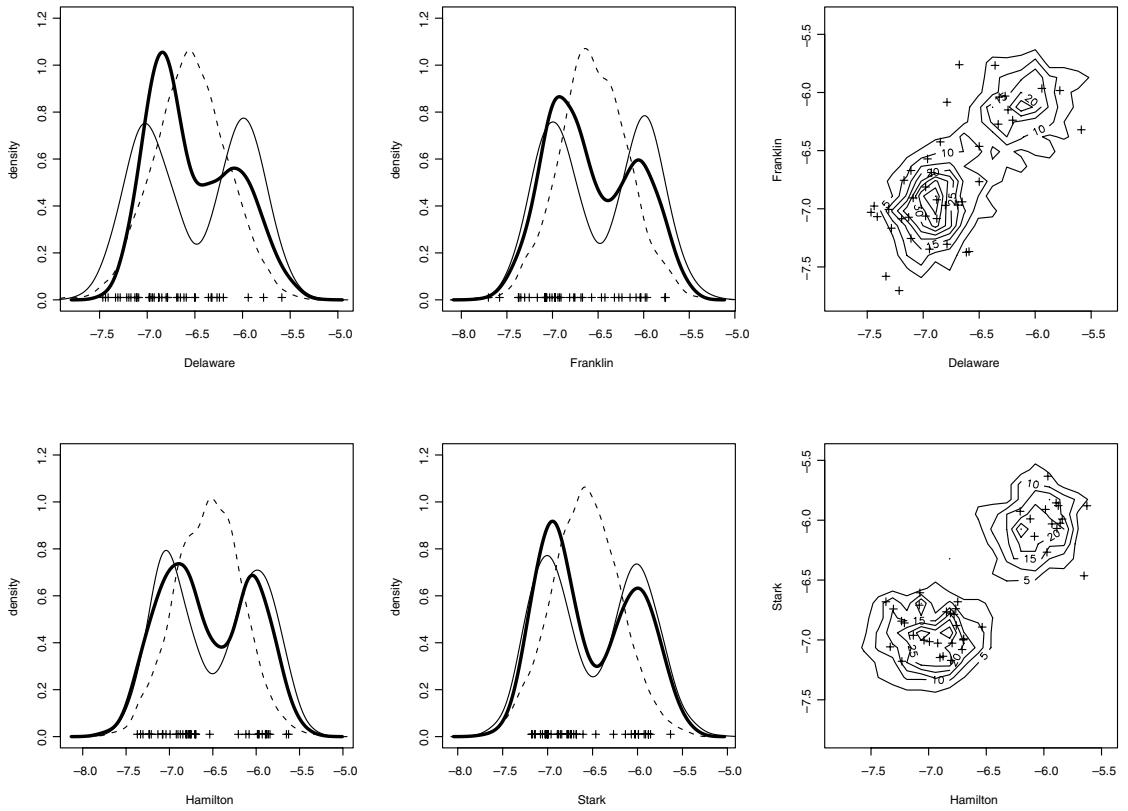

Figure 2 For the simulation example of Section 4.1, the left and middle columns include posterior predictive densities for the log-rates corresponding to four counties, based on the SDP model (thick curves) and the GP model (dashed curves). The true densities are denoted by the thin curves. The right column includes contour plots of bivariate posterior predictive densities for log-rates associated with two pairs of counties, overlaid on the corresponding observed log-rates. In all panels, the 40 observed log-rates, $\log(y_{it}/n_i)$, are denoted by "+".

Next, we pair the four counties above to show the posterior predictive joint densities, based on the SDP model (see the right column in Figure 2). We note that, with only 40 replications, our model captures quite well both marginal and joint densities for the log-rates.

4.2 Ohio lung cancer data

For our illustrative analysis of the Ohio lung cancer mortality data we consider the subset consisting of observations for white males between 55 and 64 years; the same subset was analyzed in Knorr-Held (2000). (See Kottas, Duan and Gelfand, 2006, for an analysis of the full data set, using a version of model (4) with a linear trend βt replacing the μ_t.)

Time t is normalized to be from $t = 1$ to 20 for years 1968 to 1987. To validate our model, we leave year 1988 out in our model fitting. There was prior to posterior learning for all model hyper-parameters. In particular, posterior point (posterior medians) and 90% equal-tail interval estimates for μ_0, τ^2 and ν are given by -6.281 $(-6.323, -6.246)$, 0.501 (0.479, 0.523) and 0.259 (0.127, 0.401), respectively.

In Figure 3 we display posterior point (medians) and 95% interval estimates for the log-rates over time for the four counties considered in Section 4.1, as well as counties "Clermont" and "Marion" (see Figure 1). Results for years 1968 to 1987 are based on the posterior samples for the correspond-

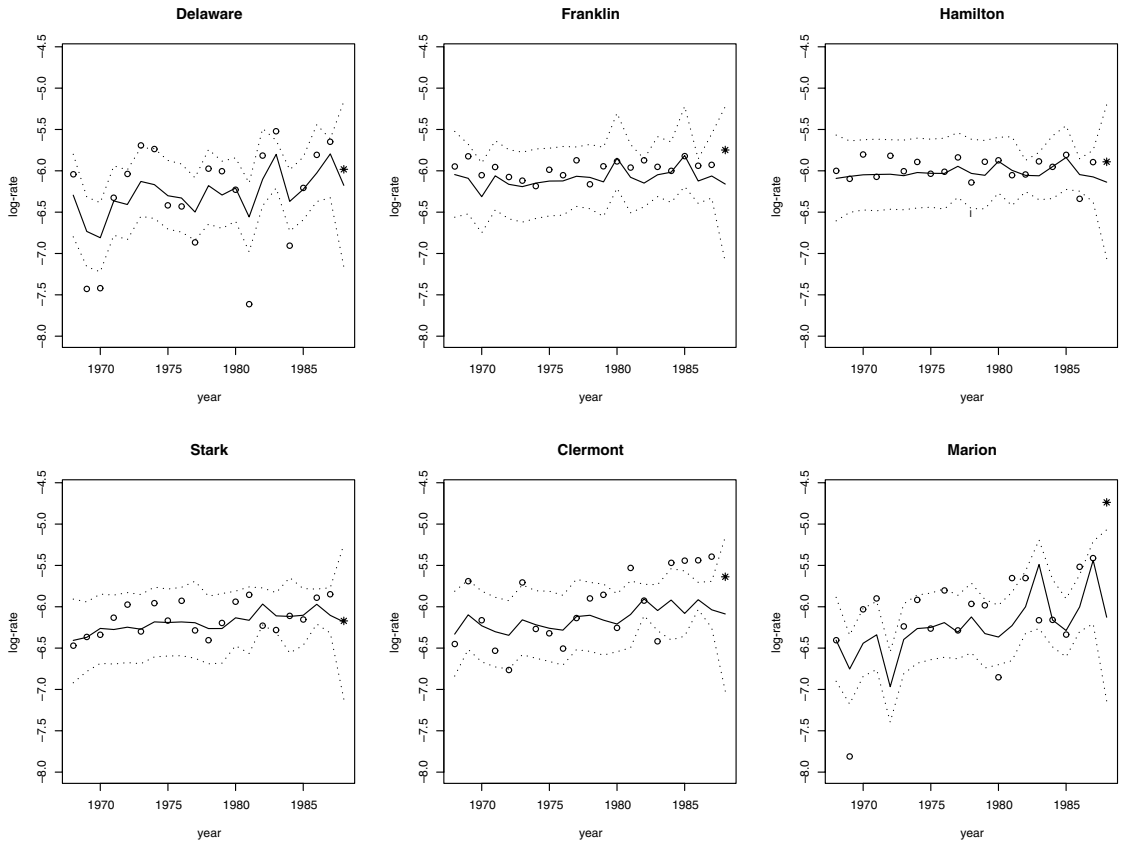

Figure 3 For the Ohio data of Section 4.2, posterior point estimates (solid lines) and 95% interval estimates (dotted lines) of the log-rates over time for six counties. In each panel, the circles denote the observed log-rates for years 1968 to 1987, the data of which were used in the model fitting, and the "$*$" indicates the observed log-rate for year 1988, which was used to validate the model.

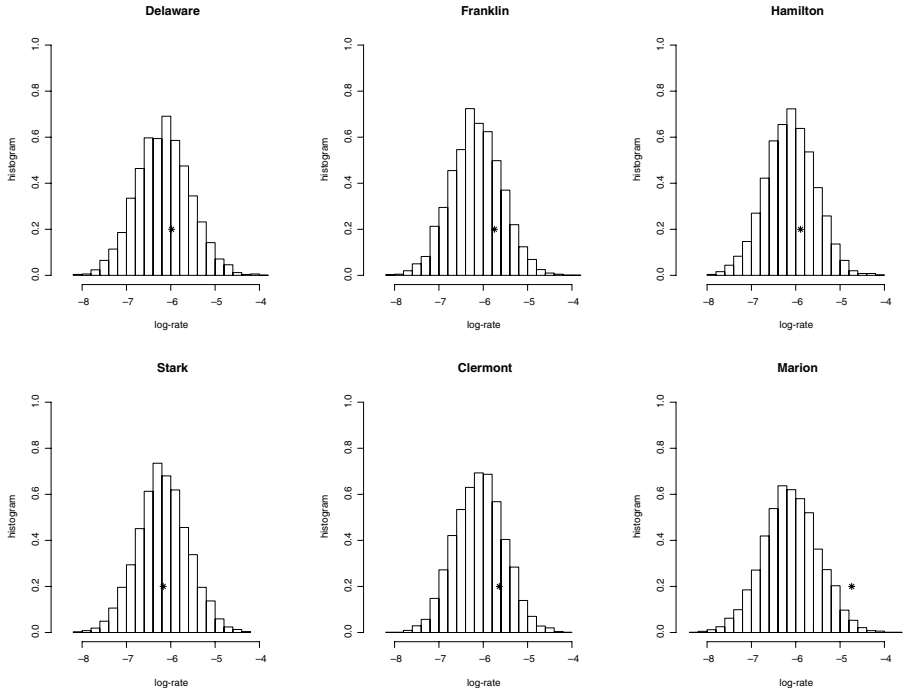

Figure 4 For the Ohio data example, histograms of draws from the posterior forecast distribution for the log-rate of six counties in the hold-out year (year 1988). The "$*$" in each panel indicates the value of the observed log-rate.

ing z_{it}, $t = 1, \ldots, 20$, whereas the estimates for year 1988 are based on draws from the posterior forecast distribution developed in Section 3. Histograms of these draws are plotted in Figure 4. Finally, we also calculated 95% marginal posterior forecast intervals for the log-rates for all 88 counties in 1988 and found that 84 out of 88 observed log-rates (95.45%) are within their 95% interval; we do not seem to be overfitting or underfitting.

5 Discussion

We have argued that, with regard to disease mapping, it may be advantageous to conceptualize the model as a spatial point process rather than through more customary areal unit spatial dependence specifications. Aggregation of the point process to suitable spatial units enables us to use it for the observable data. Specifying a non-homogeneous point process requires a model for the latent risk surface. Here, we have argued that there are advantages to viewing this surface as a process realization rather than through parametric modeling. But then, the flexibility of a nonparametric process model as opposed to the limitations of a stationary GP model becomes attractive. The choice of a spatial DP finally yields our proposed approach.

Extensions in several directions may be envisioned. Two examples are the following. We often study concurrent disease maps to try to understand the pattern of joint incidence of diseases. In our setting, for a pair of diseases, this would take us to a pair of dependent surfaces from a bivariate spatial process. We could envision modeling based upon a bivariate SDP centered around a bivariate GP. Moreover, it would be of interest to extend our nonparametric modeling framework to handle misalignment issues, i.e., data settings where the disease counts are observed for one set of areal units while covariate information is supplied for a different set of units.

Appendix: Posterior Simulation Methods

Spatial model: Under model (2), the posterior full conditional for each z_{it} is proportional to $\exp(-n_{it}\exp(z_{it}))\,N(z_{it} \mid \mu_0 + \theta_{it} + \tau^2 y_{it}, \tau^2)$, and can be sampled by introducing an auxiliary variable u_{it} (> 0) such that $p(z_{it}, u_{it} \mid \ldots, \text{data}) \propto N(z_{it} \mid \mu_0 + \theta_{it} + \tau^2 y_{it}, \tau^2)\ 1_{(0<u_{it}<\exp(-n_{it}\exp(z_{it})))}$. Now the Gibbs sampler is extended to draw from $p(u_{it} \mid z_{it}, \text{data})$ and $p(z_{it} \mid u_{it}, \ldots, \text{data})$. The former is a uniform distribution over $(0, \exp(-n_{it}\exp(z_{it})))$. The latter is a $N(\mu_0 + \theta_{it} + \tau^2 y_{it}, \tau^2)$ distribution truncated over the interval $(-\infty, \log(-n_{it}^{-1}\log u_{it}))$. Alternatively, adaptive rejection sampling can be used to draw from the full conditional for z_{it} noting that its density is log-concave.

Having updated all the z_{it}, the mixing parameters $\boldsymbol{\theta}_t$, $t = 1, \ldots, T$, and hyperparameters μ_0, τ^2, σ^2, ϕ, can be updated as in the spatial DP mixture model (reviewed in Section 2.2), with $\boldsymbol{z}_t$ playing the role of the data vector $\boldsymbol{y}_t$. (We refer to the Appendix in Gelfand, Kottas and MacEachern, 2005, for details.) All these updates require computations involving $R_n(\phi)$. To approximate the entries of this matrix, we use Monte Carlo integrations based on sets of locations distributed independently and uniformly over each region B_i.

Spatio-temporal model: Parameters z_{it}, σ^2 and ϕ of model (4) are updated in a similar fashion as in the spatial model. A random-walk Metropolis-Hastings step was used to update ν. The full conditional for μ_0 and for each of the μ_t is normal, while τ^2 and σ_μ^2 have inverse gamma full conditionals.

For $t = 1, \ldots, T$, $p(\boldsymbol{\eta}_t \mid \ldots, \text{data}) \propto p(\boldsymbol{\eta}_t \mid \{\boldsymbol{\eta}_j : j \neq t\}, \sigma^2, \phi) \prod_{\ell=t}^{T} N_n(\boldsymbol{z}_\ell \mid \boldsymbol{d}_\ell + \nu^{\ell-t}\boldsymbol{\eta}_t, \tau^2 I_n)$, where $\boldsymbol{d}_\ell = (\mu_0 + \mu_\ell)1_n + \sum_{m=1, m\neq t}^{\ell} \nu^{\ell-m}\boldsymbol{\eta}_m$, $\ell = t, \ldots, T$. The product above is proportional to a $N_n(\boldsymbol{\eta}_t \mid \boldsymbol{m}_t, \Sigma_t)$ density, with $\boldsymbol{m}_t = (\sum_{\ell=t}^{T} \nu^{2(\ell-t)})^{-1} \sum_{\ell=t}^{T} \nu^{\ell-t}(\boldsymbol{z}_\ell - \boldsymbol{d}_\ell)$, $\Sigma_t = \tau^2(\sum_{\ell=t}^{T} \nu^{2(\ell-t)})^{-1} I_n$. Let T^{*-} be the number of distinct $\boldsymbol{\eta}_j$ in $\{\boldsymbol{\eta}_j : j \neq t\}$, $\boldsymbol{\eta}_j^{*-}$, $j = 1, \ldots, T^{*-}$, the distinct values, and T_j^- the size of the cluster corresponding to $\boldsymbol{\eta}_j^{*-}$. The prior full conditional $p(\boldsymbol{\eta}_t \mid \{\boldsymbol{\eta}_j : j \neq t\}, \sigma^2, \phi)$ is a mixed distribution with point masses $T_j^-(\alpha + T - 1)^{-1}$ at the $\boldsymbol{\eta}_j^{*-}$ and continuous mass $\alpha(\alpha + T - 1)^{-1}$ on the $N_n(\boldsymbol{0}, \sigma^2 R_n(\phi))$ distribution. Hence, $p(\boldsymbol{\eta}_t \mid \ldots, \text{data})$ is also a mixed distribution with point masses, proportional to $T_j^- q_j$, at the $\boldsymbol{\eta}_j^{*-}$ and continuous mass, proportional to αq_0, on an n-variate normal with covariance matrix $H_t = (\Sigma_t^{-1} + \sigma^{-2}R_n^{-1}(\phi))^{-1}$ and mean vector $H_t\Sigma_t^{-1}\boldsymbol{m}_t$. Here, q_j is the value of the $N_n(\boldsymbol{m}_t, \Sigma_t)$ density at $\boldsymbol{\eta}_j^{*-}$, and $q_0 = \int N_n(\boldsymbol{u} \mid \boldsymbol{0}, \sigma^2 R_n(\phi))\, N_n(\boldsymbol{u} \mid \boldsymbol{m}_t, \Sigma_t)\, d\boldsymbol{u}$, an integral that is available analytically.

Acknowledgement The work of the first and the third author was supported in part by NSF grants DMS-0505085 and DMS-0504953, respectively. The authors thank a referee and the Associate Editor for helpful comments.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall, Boca Raton.

Bernardinelli, L., Clayton, D. G., and Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine* **14**, 2411–2432.

Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000). Spatial Poisson regression for health and exposure data measured at disparate resolutions. *Journal of the American Statistical Association* **95**, 1076–1088.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* **1**, 353–355.

Böhning, D., Dietz, E., and Schlattmann, P. (2000). Space-time mixture modelling of public health data. *Statistics in Medicine* **19**, 2333–2344.

Brix, A. and Diggle, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society Series B* **63**, 823–841.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–681.

Elliott, P., Wakefield, J., Best, N., and Briggs, D. (Eds.) (2000). *Spatial Epidemiology: Methods and Applications*. University Press, Oxford.

Escobar, M. D. and West, M. (1998). Computing nonparametric hierarchical models. In Dey, D., Müller, P., and Sinha, D. (Eds), *Practical Nonparametric and Semiparametric Bayesian Statistics*. New York: Springer, pp. 1–22.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.

Gelfand, A. E., Banerjee, S., and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics* **16**, 465–479.

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* **100**, 1021–1035.

Kelsall, J. and Wakefield, J. (2002). Modeling spatial variation in disease risk: a geostatistical approach. *Journal of the American Statistical Association* **97**, 692–701.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* **19**, 2555–2567.

Knorr-Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine* **17**, 2045–2060.

Knorr-Held, L. and Rasser, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* **56**, 13–21.

Kottas, A., Duan, J., and Gelfand, A. E. (2006). Bayesian nonparametric spatio-temporal models for disease incidence data. *Proceedings of the 2006 Joint Statistical Meetings, ASA Section on Bayesian Statistical Science*, pp. 60–71.

Militino, A. F., Ugarte, M. D., and Dean, C. B. (2001). The use of mixture models for identifying high risks in disease mapping. *Statistics in Medicine* **20**, 2035–2049.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

Short, M., Carlin, B. P., and Gelfand, A. E. (2005). Covariate-adjusted spatial CDF's for air pollutant data. *Journal of Agricultural, Biological, and Environmental Statistics* **10**, 259–275.

Wakefield, J. and Shaddick, G. (2006). Health-exposure modeling and the ecological fallacy. *Biostatistics* **7**, 438–455.

Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**, 607–617.

