Bayesian Nonparametric Spatio-Temporal Models for Disease Incidence Data

Athanasios Kottas[1], Jason A. Duan[2], Alan E. Gelfand[3]
University of California, Santa Cruz[1]
Yale School of Management[2]
Duke University[3]

Abstract

Typically, disease incidence or mortality data are 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 with efficient Gibbs samplers through strategically chosen latent variables. 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.

Keywords: 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 $i = 1, ..., n$ indexes the regions B_i, and $t = 1, ..., T$ 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. Such information can be accommodated in our modeling framework as discussed in Kottas *et al.* (2006). However, the focus here is on flexible modeling of areal unit spatial random effects and so we do not consider covariates.

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 (parametric) models; see, e.g., Elliott *et al.* (2000), and Banerjee *et al.* (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.) E_{it} is obtained as $R^* n_{it}$, with R^* an overall disease rate, using either *external* or *internal* standardization, the former developing R^* from reference tables (available for certain types of cancer), the latter computed from the given data set, e.g., $R^* = \sum_{it} y_{it} / \sum_{it} n_{it}$. Next, the relative risks R_{it} 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., $\mu_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta}$ for regression coefficients $\boldsymbol{\beta}$), u_i are regional random effects (typically, the u_i are assumed i.i.d. $\text{N}(0, \sigma_u^2)$), v_i are spatial random effects, and δ_t are temporal effects (say, with an autoregressive prior).

The most commonly used prior model for the v_i is based on some form of a conditional autoregressive (CAR) structure (see, e.g., Clayton and Kaldor, 1987; Cressie and Chan, 1989; Besag *et al.*, 1991; Bernardinelli *et al.*, 1995; Besag *et al.*, 1995; Waller *et al.*, 1997; Pascutto *et al.*, 2000). For instance, the widely-used specification suggested by 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. Alternatively, a multivariate normal distribution for the v_i, with correlations driven by the distances between region centroids, has been used (see, e.g., Wakefield and Morris, 1999; Banerjee *et al.*, 2003).

A different hierarchical formulation, discussed in Böhning *et al.* (2000), involves replacing the normal mixing distribution with a discrete distribution taking values φ_j, $j = 1, ..., k$ (that represent the relative risks for k underlying time-space clusters) with corresponding probabilities p_j, $j = 1, ..., k$. 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, Schlattmann and Böhning (1993) and Militino *et al.* (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) and Giudici *et al.* (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. Further related Bayesian work is that of Gangnon and Clayton (2000), Green and Richardson (2002), and Lawson and Clark (2002).

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 Banerjee *et al.*, 2004).

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 *et al.*, 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 $\boldsymbol{z}_{t,D} = \{z_t(s) : s \in D\}$ 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 $\boldsymbol{z}_t = (z_{1t}, ..., z_{nt})$, which we define by block averaging the surfaces $\boldsymbol{z}_{t,D}$ 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; Antoniak, 1974) 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 *et al.*, 2005). To model the $\boldsymbol{z}_t$, we can specify them to be independent replications under the DP or we can add a further dynamic level to the model with $\boldsymbol{z}_t$ evolving from $\boldsymbol{z}_{t-1}$ 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 mixture representation, which is more general than that of Böhning *et al.* (2000) as it incorporates spatial dependence and it allows, through the structure of 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. 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 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 prior model is discussed in Section 2.1. Section 2.2 briefly reviews spatial DPs and demonstrates how their use provides foundation for the modeling approach presented in Section 2.1. Section 2.3 develops a nonparametric spatio-temporal modeling framework.

2.1 The spatial prior probability model

Here, we treat the log-rate surfaces $\boldsymbol{z}_{t,D}$ 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. Thus, the y_{it} are assumed conditionally independent, given $z_{it} = \log p_{it}$, from $\mathrm{Po}(y_{it} \mid n_{it}\exp(z_{it}))$.

This specification can be derived through aggregation of an underlying Poisson process under assumptions and approximations as follow. For the time period t, assume that the disease incidence cases, over region D, are distributed according to a non-homogeneous Poisson 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 Poisson 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 $\int_{B_i} n_t(s)p_t(s)ds = n_{it}p^*_{it}$, where $p^*_{it} = |B_i|^{-1}\int_{B_i} p_t(s)ds$. If we approximate the distribution of the p^*_{it} with the distribution of the $\exp(z_{it})$, we can write $y_{it} \mid z_{it} \overset{ind.}{\sim} \mathrm{Po}(y_{it} \mid n_{it}\exp(z_{it}))$ 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 *et al.*, 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. 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 $\mu_t(s) = \mu$, for all t, and use a normal prior for μ. Alternatively, we could set $\mu_t(s) = \mu_t$, where the μ_t are i.i.d. $\mathrm{N}(0, \sigma^2_\mu)$ with random hyperparameter σ^2_μ. In what follows for the spatial prior model, we illustrate with the common μ specification.

To develop the model for the spatial random effects, first, let the $\boldsymbol{\theta}_{t,D}$, $t = 1, ..., T$, 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 $z_{it} = \mu + \theta_{it}$, where $\theta_{it} = |B_i|^{-1}\int_{B_i}\theta_t(s)ds$ is the block average of the surface $\boldsymbol{\theta}_{t,D}$ over region B_i. The induced distribution for $\boldsymbol{\theta}_t = (\theta_{1t}, ..., \theta_{nt})$ 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

$$|B_i|^{-1}|B_j|^{-1}\int_{B_i}\int_{B_j}\rho(||s - s'||; \phi)\, dsds'.$$

Next, consider a DP prior for the spatial random effects $\boldsymbol{\theta}_t$ with precision parameter $\alpha > 0$ and centering (base) distribution $\mathrm{N}_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi))$ (we will write $N_p(\cdot \mid \boldsymbol{\lambda}, \Sigma)$ for the p-variate normal density/distribution with mean vector $\boldsymbol{\lambda}$ and covariance matrix Σ). We denote this DP prior by $\mathrm{DP}(\alpha, \mathrm{N}_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi)))$. The choice of the DP in this context yields data-driven deviations from the normality assumption for the spatial random effects; at the same time, it allows relatively simple implementation of simulation-based model fitting.

Note that the above structure implies for the vector of counts $\boldsymbol{y}_t = (y_{1t}, ..., y_{nt})$ a nonparametric Poisson mixture model given by $\int\prod_{i=1}^n \mathrm{Po}(y_{it} \mid n_{it}\exp(\mu + \theta_{it}))dG(\boldsymbol{\theta}_t)$, where the mixing distribution $G \sim \mathrm{DP}(\alpha, \mathrm{N}_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi)))$. Under this mixture specification, the distribution for the vectors of log-rates, $\boldsymbol{z}_t = \mu 1_n + \boldsymbol{\theta}_t$, 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 (e.g., the methods suggested by MacEachern and Müller, 1998, and Neal, 2000) than the standard Gibbs sampler for DP-based hierarchical models (e.g., West *et al.*, 1994; Bush and MacEachern, 1996). Thus, to overcome both concerns above, we replace the DP prior for the $\boldsymbol{z}_t$ with a DP mixture prior,

$$\boldsymbol{z}_t \mid \mu, \tau^2, G \overset{ind.}{\sim} \int \mathrm{N}_n(\boldsymbol{z}_t \mid \mu 1_n + \boldsymbol{\theta}_t, \tau^2 I_n)dG(\boldsymbol{\theta}_t),$$

where, again, $G \sim \mathrm{DP}(\alpha, \mathrm{N}_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi)))$. That is, we now write $z_{it} = \mu + \theta_{it} + u_{it}$, with u_{it} i.i.d. $\mathrm{N}(0, \tau^2)$. 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) and Bernardinelli *et al.* (1995), 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.

Hence, the mixture model for the $\boldsymbol{y}_t$ now assumes the form

$$f(\boldsymbol{y}_t \mid \mu, \tau^2, G) = \int\prod_{i=1}^n p(y_{it} \mid \mu, \tau^2, \theta_{it})dG(\boldsymbol{\theta}_t),$$

where $p(y_{it} \mid \mu, \tau^2, \theta_{it}) = \int \mathrm{Po}(y_{it} \mid n_{it}\exp(z_{it}))\mathrm{N}(z_{it} \mid \mu + \theta_{it}, \tau^2)dz_{it}$ is a Poisson-lognormal mixture. Equiv-

alently, the model can be written in the following semiparametric hierarchical form

$$\begin{array}{rcl} y_{it} \mid z_{it} & \overset{ind.}{\sim} & \mathrm{Po}(y_{it}|n_{it}\exp(z_{it})) \\ z_{it} \mid \mu, \theta_{it}, \tau^2 & \overset{ind.}{\sim} & \mathrm{N}(z_{it} \mid \mu + \theta_{it}, \tau^2) \\ \boldsymbol{\theta}_t \mid G & \overset{i.i.d.}{\sim} & G \\ G \mid \sigma^2, \phi & \sim & \mathrm{DP}(\alpha, \mathrm{N}_n(\cdot \mid \mathbf{0}, \sigma^2 R_n(\phi))). \end{array} \tag{1}$$

The model is completed with independent priors $p(\mu)$, $p(\tau^2)$ and $p(\sigma^2)$, $p(\phi)$ for μ, τ^2, and for the hyperparameters σ^2, ϕ of the DP prior. In particular, we use a normal prior for μ, inverse gamma priors for τ^2 and σ^2, and a discrete uniform prior for ϕ. Although not implemented for the examples of Section 4, a prior for α can be added, without increasing the complexity of the posterior simulation method (Escobar and West, 1995).

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

$$\begin{array}{c} p(\mu)p(\tau^2)p(\sigma^2)p(\phi)p(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_T \mid \sigma^2, \phi) \\ \prod_{i=1}^{n}\prod_{t=1}^{T} \mathrm{Po}(y_{it}|n_{it}\exp(z_{it}))\mathrm{N}(z_{it} \mid \mu + \theta_{it}, \tau^2), \end{array} \tag{2}$$

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 $\boldsymbol{\theta}_t$, $p(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_T \mid \sigma^2, \phi)$, is given by

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

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 $t = 2, ..., T$, $\boldsymbol{\theta}_t$ is either set equal to $\boldsymbol{\theta}_j$, $j = 1, ..., t-1$, 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 $\mathrm{N}_n(\mathbf{0}, \sigma^2 R_n(\phi))$ 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 briefly review SDPs, which provide nonparametric prior models for random fields $\boldsymbol{W}_D = \{W(s) : s \in D\}$ over a region $D \subseteq R^d$, and thus yield suitable nonparametric priors for the analysis of spatial or spatio-temporal geostatistical data. Central to their development is the constructive definition of the DP (Sethuraman, 1994). According to this definition, a random distribution arising from $\mathrm{DP}(\alpha, G_0)$, 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 $\omega_1 = z_1$, $\omega_\ell = z_\ell\prod_{r=1}^{\ell-1}(1-z_r)$, ℓ = 2,3,..., with $\{z_r,\ r = 1,2,...\}$ i.i.d. from $\mathrm{Beta}(1, \alpha)$, and, independently, $\{\varphi_\ell,\ \ell = 1,2,...\}$ 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, $\boldsymbol{\varphi}_{\ell,D} = \{\varphi_\ell(s) : s \in D\}$, 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 $\sum_{\ell=1}^{\infty}\omega_\ell\delta_{\boldsymbol{\varphi}_{\ell,D}}$. The interpretation is that for any collection of spatial locations in D, say, $(s_1, ..., s_M)$, G_D induces a random probability measure $G^{(M)}$ on the space of distribution functions for $(W(s_1), ..., W(s_M))$. In fact, $G^{(M)} \sim \mathrm{DP}(\alpha, G_0^{(M)})$, where $G_0^{(M)}$ is the M-variate normal distribution for $(W(s_1), ..., W(s_M))$ 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}.

SDPs provide an illustration of dependent Dirichlet processes (MacEachern, 1999) in that they yield a stochastic process of random distributions, one at each location in D. These distributions are dependent but such that, at each index value, the distribution is a univariate DP. See De Iorio *et al.* (2004) for an illustration in the ANOVA setting; Teh *et al.* (2006) for related work on hierarchical DPs; and Griffin and Steel (2006) and Duan, Guindani and Gelfand (2005) for recent extensions and alternative constructions.

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 $\boldsymbol{y}_t = (y_t(s_1), ..., y_t(s_n))'$, t = 1,...,T, where $(s_1, ..., s_n)$ are the locations where the observations are obtained. Working with continuous real-valued measurements, the SDP is used as a prior for the spatial random effects surfaces, say, $\boldsymbol{\zeta}_{t,D} = \{\zeta_t(s) : s \in D\}$, 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. $\mathrm{N}(0, \tau^2)$, 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 $\boldsymbol{\beta}$ a $p\times 1$ vector of regression coefficients, we could write $X_t'\boldsymbol{\beta}$ for the mean structure associated with $\boldsymbol{y}_t$. Hence, the $\boldsymbol{y}_t$, given $\boldsymbol{\beta}$, τ^2, and $G^{(n)}$, are independent from the DP mixture model $\int \mathrm{N}_n\left(\boldsymbol{y}_t \mid X_t'\boldsymbol{\beta} + \boldsymbol{\zeta}_t, \tau^2 I_n\right) dG^{(n)}(\boldsymbol{\zeta}_t)$, where $\boldsymbol{\zeta}_t = (\zeta_t(s_1), ..., \zeta_t(s_n))$, $G^{(n)} \sim \mathrm{DP}(\alpha, G_0^{(n)})$ (induced by the SDP prior for the $\boldsymbol{\zeta}_{t,D}$), with $G_0^{(n)}$ an n-variate normal (induced at $(s_1, ..., s_n)$ by the base GP of the SDP prior). Details for prior specification, simulation-based model fitting, 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. For instance, the $y_t(s_i)$ could arise from an exponential-dispersion family. Hence, we can formulate nonparametric spatial generalized linear models, extending the work in Diggle *et al.* (1998) where a stationary GP was used for the spatial random effects (see also, e.g., Heagerty and Lele, 1998, Diggle *et al.*, 2002, and Christensen and Waagepetersen, 2002).

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 $\boldsymbol{\theta}_{t,D}$, $t = 1, ..., T$, 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 $G_{0D} = \mathrm{GP}(\mathbf{0}, \sigma^2\rho(||s - s'||; \phi))$ (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, $r = 1, ..., M$, 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 $t = 2, ..., T$, $\boldsymbol{\theta}_{t,D}$ is identical to $\boldsymbol{\theta}_{j,D}$, $j = 1, ..., t-1$, 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 $\mathrm{N}_n(\mathbf{0}, \sigma^2 R_n(\phi))$ 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 $\mathrm{N}_n(\mathbf{0}, \sigma^2 R_n(\phi))$ with probability $\alpha(\alpha + 1)^{-1}$ or $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1$ with probability $(\alpha + 1)^{-1}$. Analogously, for any $t = 2, ..., T$, the induced conditional prior $p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{t-1}, \sigma^2, \phi)$ is a mixed distribution with point masses at $\boldsymbol{\theta}_j$, $j = 1, ..., t-1$, and continuous piece given by the $\mathrm{N}_n(\mathbf{0}, \sigma^2 R_n(\phi))$ distribution; the corresponding weights are $(\alpha+t-1)^{-1}$, $j = 1, ..., t-1$, 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 $\boldsymbol{z}_t = \mu 1_n + \boldsymbol{\theta}_t$ to $\boldsymbol{z}_t = \mu 1_n + \boldsymbol{\theta}_t + \boldsymbol{u}_t$, where the $\boldsymbol{u}_t$ are independent $\mathrm{N}_n(\mathbf{0}, \tau^2 I_n)$. Hence, model (2) is equivalent to the marginal version of the model above, i.e., with G_D marginalized over its SDP prior.

The argument above, based on SDPs, 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 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 Banerjee *et al.*, 2004, 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, as in Section 2.1, by writing $z_t(s) = \mu_t + \theta_t(s)$ and add temporal structure to the model through *transition equations* for the $\theta_t(s)$, say,

$$\theta_t(s) = \nu\theta_{t-1}(s) + \eta_t(s), \tag{4}$$

where, in general, $|\nu| < 1$, and the innovations $\boldsymbol{\eta}_{t,D} = \{\eta_t(s) : s \in D\}$ are independent realizations from a spatial stochastic process. We can now define the nonparametric prior for the block averages $\eta_{it} = |B_i|^{-1}\int_{B_i}\eta_t(s)ds$ of the $\boldsymbol{\eta}_{t,D}$ surfaces following the approach of Section 2.1 or, equivalently, of Section 2.2. Proceeding with the latter, we assume that the $\boldsymbol{\eta}_{t,D}$, given G_D, are independent from G_D, and assign a SDP prior to G_D with parameters α and $G_{0D} = \mathrm{GP}(\mathbf{0}, \sigma^2\rho(||s - s'||; \phi))$. Marginalizing G_D over its prior, the induced prior, $p(\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T \mid \sigma^2, \phi)$, for the $\boldsymbol{\eta}_t = (\eta_{1t}, ..., \eta_{nt})$ is given by (3) (with $\boldsymbol{\eta}_t$ replacing $\boldsymbol{\theta}_t$). Block averaging the surfaces in the transition equations (4), we obtain $\boldsymbol{\theta}_t = \nu\boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t$, where $\boldsymbol{\theta}_{t-1} = (\theta_{1,t-1}, ..., \theta_{n,t-1})$. Adding, as before, the i.i.d. $\mathrm{N}(0, \tau^2)$ terms to the z_{it}, we obtain the following general form for the spatio-temporal hierarchical model

$$\begin{array}{rcl} y_{it} \mid z_{it} & \overset{ind.}{\sim} & \mathrm{Po}(y_{it} | n_{it}\exp(z_{it})) \\ z_{it} \mid \mu_t, \theta_{it}, \tau^2 & \overset{ind.}{\sim} & \mathrm{N}(z_{it} \mid \mu_t + \theta_{it}, \tau^2) \\ \boldsymbol{\theta}_t & = & \nu\boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t \\ \boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T \mid \sigma^2, \phi & \sim & p(\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T \mid \sigma^2, \phi). \end{array} \tag{5}$$

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 $\mathrm{N}(0, \sigma_\mu^2)$ distribution (with random σ_μ^2), or they could be explained through a parametric function $h(t; \boldsymbol{\beta})$, say, a polynomial trend, $h(t; \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^m \beta_j t^j$, or the autoregressive structure could be extended to include the μ_t, say, $\mu_t = \nu_\mu\mu_{t-1} + \gamma_t$, with $|\nu_\mu| < 1$, and γ_t i.i.d. $\mathrm{N}(0, \sigma_\mu^2)$. For the Ohio state lung cancer data (discussed in Section 4.2), we work with a linear trend function $\mu_t = \beta_0 + \beta_1 t$. We set $\boldsymbol{\theta}_1 = \boldsymbol{\eta}_1$, i.e., $\boldsymbol{\theta}_0 = \mathbf{0}$ (alternatively, an informative prior for $\boldsymbol{\theta}_0$ can be used). We choose priors for τ^2, σ^2 and ϕ as in model (2); we take independent normal priors for the components of $\boldsymbol{\beta}$; and a discrete uniform prior for ν.

3 Posterior inference and prediction

We discuss here the types of posterior inference that can be obtained based on the models of Section 2. In particular, Section 3.1 comments on the (smoothed) inference for the disease rates while, under the dynamic model, Section 3.2 discusses forecasting of disease rates using the extension of Section 2.3.

3.1 Spatial model

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 $(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_T)$ and denote by $\boldsymbol{\theta}^* = \{\boldsymbol{\theta}_j^* : j = 1, ..., T^*\}$ the vector of distinct values. Defining the vector of configuration indicators, $\boldsymbol{w} = (w_1, ..., w_T)$, 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 $(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_T)$. 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 data $= \{(y_{it}, n_{it}) : i = 1, ..., n, t = 1, ..., T\}$, can be obtained using the Gibbs sampler discussed briefly below.

The full conditional for each z_{it} can be expressed as $p(z_{it} \mid ..., \text{data}) \propto \exp(-n_{it} \exp(z_{it}))\text{N}(z_{it} \mid \mu + \theta_{it} + \tau^2 y_{it}, \tau^2)$. We can sample from this full conditional introducing an auxiliary variable u_{it}, with positive values, such that $p(z_{it}, u_{it} \mid ..., \text{data}) \propto \text{N}(z_{it} \mid \mu + \theta_{it} + \tau^2 y_{it}, \tau^2)\ 1_{(0<u_{it}<\exp(-n_{it}\exp(z_{it})))}$. Now the Gibbs sampler is extended to draw from $p(u_{it} \mid z_{it}, \text{data})$ and $p(z_{it} \mid u_{it}, ..., \text{data})$. The former is a uniform distribution over $(0, \exp(-n_{it} \exp(z_{it})))$. The latter is a $\text{N}(\mu + \theta_{it} + \tau^2 y_{it}, \tau^2)$ 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 $\boldsymbol{\theta}_t$, $t = 1, ..., T$, and hyperparameters μ, τ^2, σ^2, ϕ, can be updated as in the spatial DP mixture model (reviewed briefly 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 *et al.*, 2005, for details.) All these updates require computations involving the matrix $R_n(\phi)$. 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, $i = 1, ..., n$. Note that, with the discrete uniform prior for ϕ, these calculations need only be performed once at the beginning of the MCMC algorithm.

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* $\boldsymbol{z}_0 = (z_{10}, ..., z_{n0})$,

$$p(\boldsymbol{z}_0 \mid \text{data}) = \int\int p(\boldsymbol{z}_0 \mid \boldsymbol{\theta}_0, \mu, \tau^2) p(\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}^*, \boldsymbol{w}, T^*, \sigma^2, \phi)\, p(\boldsymbol{\psi} \mid \text{data}). \tag{6}$$

Here, $p(\boldsymbol{z}_0 \mid \boldsymbol{\theta}_0, \mu, \tau^2)$ is a $\text{N}_n(\mu 1_n + \boldsymbol{\theta}_0, \tau^2 I_n)$ density, $\boldsymbol{\theta}_0 = (\theta_{10}, ..., \theta_{n0})$ is the vector of spatial random effects corresponding to $\boldsymbol{z}_0$, and

$$p(\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}^*, \boldsymbol{w}, T^*, \sigma^2, \phi) = \tfrac{\alpha}{\alpha+T}\text{N}_n(\boldsymbol{\theta}_0 \mid \boldsymbol{0}, \sigma^2 R_n(\phi)) + \tfrac{1}{\alpha+T}\sum_{j=1}^{T^*} T_j \delta_{\boldsymbol{\theta}_j^*}(\boldsymbol{\theta}_0), \tag{7}$$

where T_j is the size of the j-th cluster $\boldsymbol{\theta}_j^*$. Therefore, $p(\boldsymbol{z}_0 \mid \text{data})$ arises by averaging the mixture

$$\tfrac{\alpha}{\alpha+T}\text{N}_n(\boldsymbol{z}_0 \mid \mu 1_n, \tau^2 I_n + \sigma^2 R_n(\phi)) + \tfrac{1}{\alpha+T}\sum_{j=1}^{T^*} T_j \text{N}_n(\boldsymbol{z}_0 \mid \mu 1_n + \boldsymbol{\theta}_j^*, \tau^2 I_n)$$

with respect to the posterior of $\boldsymbol{\psi}$. 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.

3.2 Spatio-temporal model

Turning to the spatio-temporal model of Section 2.3, let $\mu_t = \beta_0 + \beta_1 t$ (as in the example of Section 4.2). Denoting by $\boldsymbol{\psi} = (\beta_0, \beta_1, \tau^2, \nu, \sigma^2, \phi, \{(\boldsymbol{z}_t, \boldsymbol{\eta}_t) : t = 1, ..., T\})$ the parameter vector corresponding to model (5), the posterior $p(\boldsymbol{\psi}|\text{data})$ is proportional to

$$p(\beta_0)p(\beta_1)p(\nu)p(\tau^2)p(\sigma^2)p(\phi)p(\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T \mid \sigma^2, \phi) \prod_{t=1}^{T} \text{N}_n(\boldsymbol{z}_t|\boldsymbol{\lambda}_t, \tau^2 I_n) \prod_{i=1}^{n}\prod_{t=1}^{T} \text{Po}(y_{it}|n_{it}\exp(z_{it})), \tag{8}$$

where $\boldsymbol{\lambda}_t = (\beta_0 + \beta_1 t)1_n + \sum_{\ell=1}^{t} \nu^{t-\ell}\boldsymbol{\eta}_\ell$. The Gibbs sampler described below can be used to obtain draws from $p(\boldsymbol{\psi}|\text{data})$.

The form of the full conditionals for the z_{it} is similar to the one for the spatial model, and, thus, either auxiliary variables or adaptive rejection sampling can be used to update these parameters.

For each $t = 1, ..., T$, the full conditional for $\boldsymbol{\eta}_t$ is proportional to

$$p(\boldsymbol{\eta}_t|\{\boldsymbol{\eta}_j : j \neq t\}, \sigma^2, \phi) \prod_{\ell=t}^{T} \text{N}_n(\boldsymbol{z}_\ell|\boldsymbol{d}_\ell + \nu^{\ell-t}\boldsymbol{\eta}_t, \tau^2 I_n)$$

where $\boldsymbol{d}_\ell = (\beta_0 + \beta_1 \ell)1_n + \sum_{m=1, m\neq t}^{\ell} \nu^{\ell-m}\boldsymbol{\eta}_m$, $\ell = t, ..., T$. The product term above is proportional to a $\text{N}_n(\boldsymbol{\eta}_t|\boldsymbol{\mu}_t, \Sigma_t)$ density, with $\boldsymbol{\mu}_t = (\sum_{\ell=t}^{T}\nu^{2(\ell-t)})^{-1}\sum_{\ell=t}^{T}\nu^{\ell-t}(\boldsymbol{z}_\ell - \boldsymbol{d}_\ell)$ and $\Sigma_t = \tau^2(\sum_{\ell=t}^{T}\nu^{2(\ell-t)})^{-1} I_n$. Let T^{*-} be the number of distinct $\boldsymbol{\eta}_j$ in $\{\boldsymbol{\eta}_j : j \neq t\}$, $\boldsymbol{\eta}_j^{*-}$, $j = 1, ..., T^{*-}$, be the distinct values, and T_j^- be the size of the cluster corresponding to $\boldsymbol{\eta}_j^{*-}$. The prior full conditional $p(\boldsymbol{\eta}_t|\{\boldsymbol{\eta}_j : j \neq t\}, \sigma^2, \phi)$ 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 $\text{N}_n(\boldsymbol{0}, \sigma^2 R_n(\phi))$ distribution. Hence, $p(\boldsymbol{\eta}_t|..., \text{data})$ 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 distribution with covariance matrix $H_t = (\Sigma_t^{-1} + \sigma^{-2} R_n^{-1}(\phi))^{-1}$ and mean vector $H_t \Sigma_t^{-1} \boldsymbol{\mu}_t$. Here, q_j is the value of the $\mathrm{N}_n(\boldsymbol{\mu}_t, \Sigma_t)$ density at $\boldsymbol{\eta}_j^{*-}$, and $q_0 = \int \mathrm{N}_n(\boldsymbol{u}|\mathbf{0}, \sigma^2 R_n(\phi)) \mathrm{N}_n(\boldsymbol{u}|\boldsymbol{\mu}_t, \Sigma_t) d\boldsymbol{u}$, an integral that is available analytically.

Updating σ^2 and ϕ proceeds as in the spatial model. The full conditional for τ^2 is an inverse gamma distribution, and β_0 and β_1 have normal full conditionals. Finally, working with a discrete uniform prior for ν, we sample directly from its discretized full conditional.

Having sampled from $p(\boldsymbol{\psi}|\text{data})$, we can obtain the posterior for $\boldsymbol{z}_t$, the vector of log-rates corresponding to specific time periods t. Moreover, given the temporal structure of model (5), of interest is temporal forecasting for disease rates at future time points. In particular, the posterior forecast distribution for the vector of log-rates $\boldsymbol{z}_{T+1}$ at time $T+1$,

$$p(\boldsymbol{z}_{T+1}|\text{data}) = \int\int p(\boldsymbol{z}_{T+1}|\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T, \boldsymbol{\eta}_{T+1}, \beta_0, \beta_1, \nu, \tau^2) \\ p(\boldsymbol{\eta}_{T+1}|\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T, \sigma^2, \phi) p(\boldsymbol{\psi}|\text{data}),$$

where $p(\boldsymbol{z}_{T+1}|\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T, \boldsymbol{\eta}_{T+1}, \beta_0, \beta_1, \nu, \tau^2)$ is an n-variate normal with mean vector $(\beta_0 + \beta_1(T+1))1_n + \sum_{\ell=1}^{T+1} \nu^{T+1-\ell} \boldsymbol{\eta}_\ell$ and covariance matrix $\tau^2 I_n$, and $p(\boldsymbol{\eta}_{T+1}|\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T, \sigma^2, \phi)$ can be expressed as in (7) by replacing $\boldsymbol{\theta}_0$ with $\boldsymbol{\eta}_{T+1}$ and using the, analogous to $(\boldsymbol{\theta}^*, \boldsymbol{w}, T^*)$, clustering structure in the $(\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T)$.

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. The county location, area, and polygons are obtained from the "map" package in R.

Regarding prior specification, for both models (1) and (5) we work with an exponential correlation function, $\rho(||s - s'||; \phi) = \exp(-\phi||s - s'||)$. For both data examples, the discrete uniform prior for ϕ takes values in $[0.001, 1]$, corresponding to the range from 3 to 3000 miles; σ^{-2} and τ^{-2} have gamma$(0.1, 0.1)$ priors (with mean 1); and α is set equal to 1 (results were practically identical under $\alpha = 5$ and $\alpha = 10$). Finally, the normal priors for μ (Section 4.1) and for β_0 and β_1 (Section 4.2) have mean 0 and large variance (there was very little sensitivity to choices between 10^2 and 10^8 for the prior variance).

We observed very good mixing and fast convergence in the implementation of the Gibbs samplers discussed in the Appendix. In both of our simulation and Ohio lung cancer example below, we obtain 15,000 samples from the Gibbs sampler, and discard the first 3,000 samples as burn-in. We use 3,000 subsamples from the remaining 12,000 samples, with thinning equal to 4, for our posterior inference.

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.

In particular, we proceed as follows. For $i = 1, ..., 88$ and $t = 1, ..., T$ (with $T = 40$), we first generate z_{it} independent $\mathrm{N}(\mu + \theta_{it}, \tau^2)$ and, then, y_{it} independent $\mathrm{Po}(n_i \exp(z_{it}))$, where n_i is the population of county i in 1988. 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 $\boldsymbol{\theta}^{(\ell)} = \left(\theta_1^{(\ell)}, \ldots, \theta_n^{(\ell)}\right) \sim \mathrm{N}_n((-1)^\ell \mu_\theta 1_n, \sigma_\ell^2 R)$, with the (i,j)-th element of the correlation matrix R given by $|B_i|^{-1}|B_j|^{-1} \int_{B_i}\int_{B_j} \exp(-\phi||s - s'||)\, ds ds'$. 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 $\phi = 0.6$. Under these choices, marginally, each θ_{it} has a bimodal distribution of the form $0.5\mathrm{N}\left(-\mu_\theta, \sigma_1^2\right) + 0.5\mathrm{N}\left(\mu_\theta, \sigma_2^2\right)$.

We fit model (1) to this data set. The Bayesian goodness of fit is illustrated with univariate and bivariate posterior predictive densities for the log-rates, which are estimated using (6). In Figure 2 we compare the true densities of the model from which we simulated the data with the SDP model posterior predictive densities for four selected counties. They are "Delaware" and "Franklin" in central Ohio, "Hamilton" in southwest, and "Stark" in northeast. "Franklin" includes Columbus and "Hamilton" includes Cincinnati so these are highly populated counties. "Delaware" is more suburban and "Stark" is very rural. The "+" mark the values of the 40 observed log-rates $\log(y_{it}/n_i)$ in each of these four counties. In addition, Figure 2 includes posterior predictive densities from a parametric model based on a $\mathrm{GP}(\mathbf{0}, \sigma^2 \exp(-\phi||s - s'||))$ 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. $\mathrm{N}_n(\mathbf{0}, \sigma^2 R_n(\phi))$. The SDP model clearly outperforms the GP model with regard to posterior predictive inference.

Next, we pair the four counties above to show in Figure 3 the predictive joint densities, based on the SDP model, and, again, to compare with the true joint densities (using samples in both cases). The first pair "Delaware" and "Franklin" are next to each other. The second pair "Hamilton" and "Stark" are distant. We note that, with only 40 replications, our model captures quite well both marginal and joint densities for the log-rates.

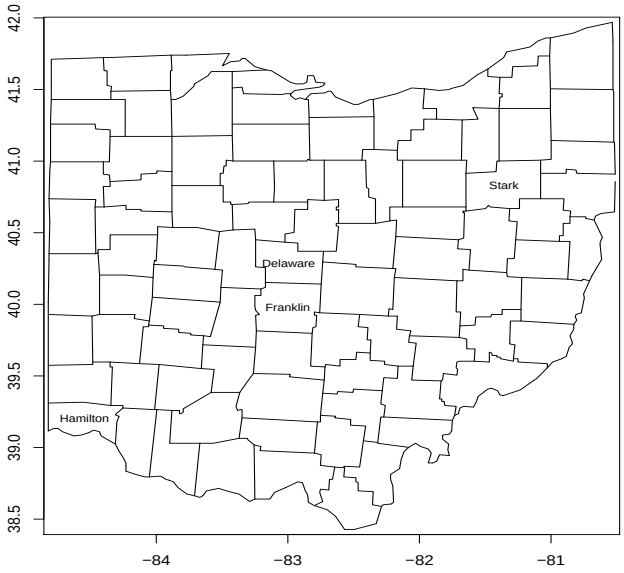

Figure 1: Map of the 88 counties in the state of Ohio.

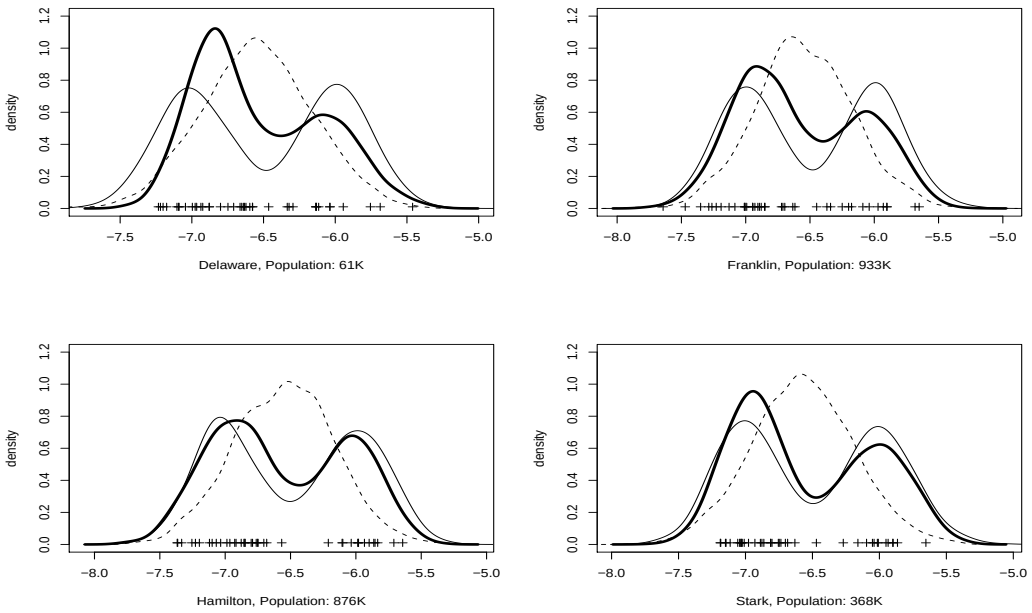

Figure 2: Simulation example. Posterior predictive densities for the log-rates, corresponding to four counties, based on the SDP model (thick curves) and the GP model (dashed curves). The true densities are denoted by the thin curves, and the observed log-rates by "+".

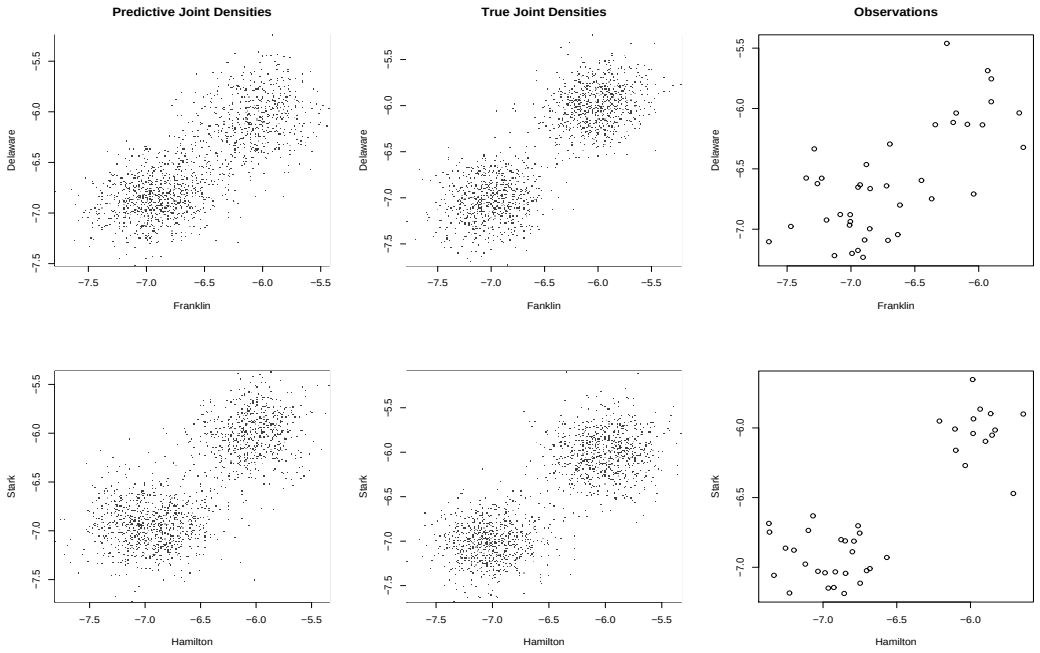

Figure 3: Simulation example. Posterior predictive densities (left column) and true bivariate densities (middle column) for log-rates associated with two pairs of counties. The right column includes plots of the corresponding observed log-rates.

4.2 Ohio lung cancer data

The exploratory study of the Ohio lung cancer mortality data reveals a spatio-temporal varying structure in the incidence rates. We display the observed log-rates $\log(y_{it}/n_{it})$ for the aforementioned four counties in Figure 4. This plot shows clear evidence of an increasing, roughly linear, trend in the log-rate. Therefore we apply the dynamic SDP model (5) with a linear trend over time, setting $\mu_t = \beta_0 + \beta_1 t$. Moreover, because negative values for ν do not appear plausible, we use a discrete uniform prior on $[0, 1)$ for ν.

The time t is normalized to be from year $t = 1$ to 21. In order to validate our model, we leave year 21 (year 1988) out in our model fitting and predict the log-rates for all 88 counties in that year, using the posterior forecast distribution developed in Section 3.2. Posterior point (posterior medians) and 95% equal-tail interval estimates for β_0, β_1 and for ν are given by -8.208 $(-8.319, -8.100)$, 0.0367 $(0.0292, 0.0448)$ and 0.7 $(0.6, 0.8)$, respectively. There was also prior to posterior learning for the other hyperparameters, in particular, point and interval estimates were 0.0586 $(0.0552, 0.0656)$ for ϕ; 0.104 $(0.0855, 0.113)$ for τ^2; and 0.133 $(0.101, 0.152)$ for σ^2.

In Figure 5 we display the marginal posterior forecast density of the log-rate for the earlier four counties in the hold-out year 1988. We also calculated 95% marginal predictive intervals for all 88 counties in 1988 and found that 83 out of 88 observed log-rates (94.3 %) are within their 95% interval; we do not seem to be overfitting or underfitting. In Figure 6 we provide the contour plot of the predictive log-rate surface for 1988, using medians from the posterior forecast distribution for each county.

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. We applied the modeling to both real and simulated data. With the simulated data we clearly demonstrated the advantage of such flexibility.

Extensions in several directions may be envisioned. Three examples are the following. In treating the specification for the μ_t we could provide a nonparametric model as well through i.i.d. realizations obtained under DP mixing or the associated dynamic version with independent innovations under such a model. Next, we often study concurrent disease maps to try to understand the pattern of joint incidence of diseases. In our setting, for a

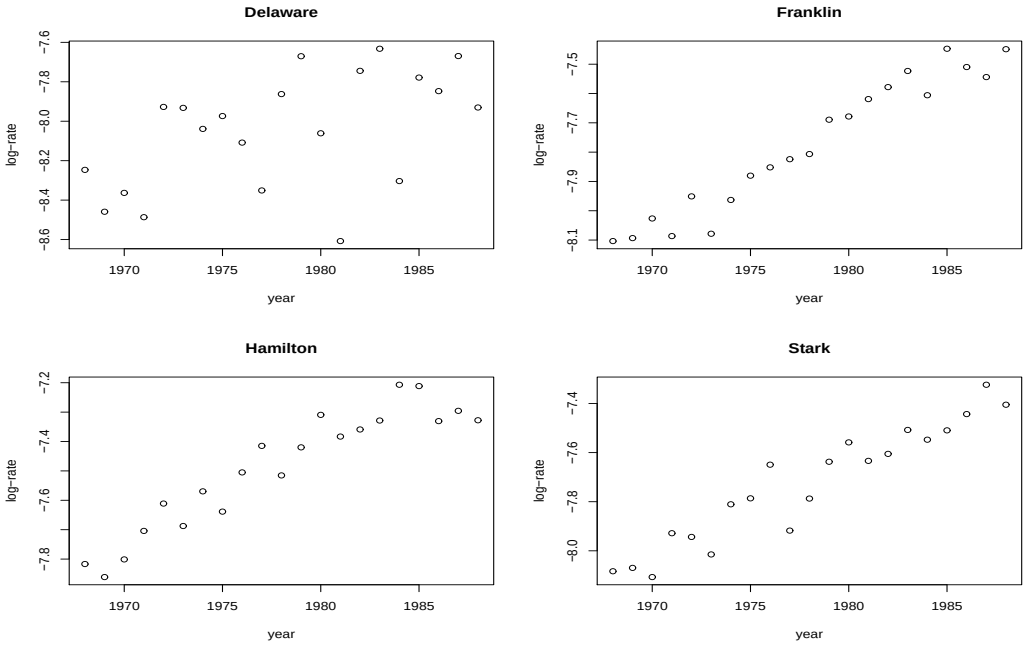

Figure 4: Observed log-rates for four counties from 1968 to 1988 for the Ohio data example.

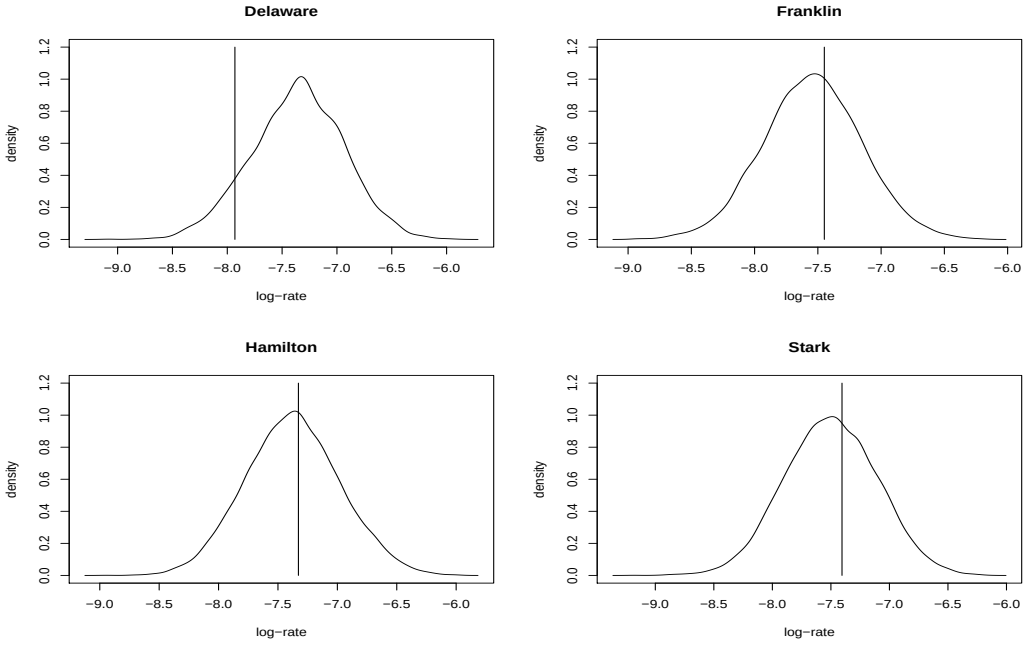

Figure 5: Posterior forecast densities for the log-rate of four counties in the hold-out year (year 1988) for the Ohio data example. The vertical line in each plot is the observed log-rate.

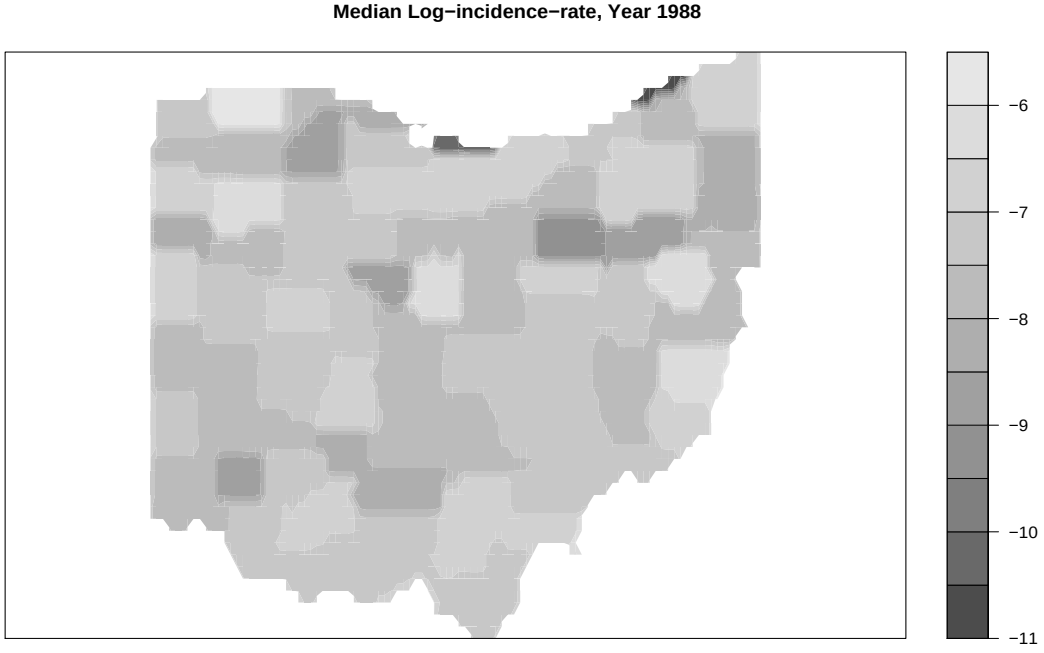

Figure 6: For the Ohio data example of Section 4.2, medians of the posterior forecast distribution for the log-rate in each county for year 1988.

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. Finally, how would we handle misalignment issues in this nonparametric setting? That is, what should we do if disease counts are observed for one set of areal units while covariate information is supplied for a different set of units? Banerjee *et al.* (2004) suggest strategies for treating misalignment but exclusively in the context of GPs. Extensions to our SDP setting would be useful.

Acknowledgments

The work of the first and the third author was supported in part by NSF grants DMS-0505085 and DMS-0504953, respectively.

References

Antoniak, C.E. (1974), "Mixtures of Dirichlet processes with applications to nonparametric problems," *The Annals of Statistics*, **2**, 1152-1174.

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

Banerjee, S., Wall, M.M., and Carlin, B.P. (2003), "Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota," *Biostatistics*, **4**, 123-142.

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.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), "Bayesian computation and stochastic systems (with discussion)," *Statistical Science*, **10**, 3-66.

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

Bush, C.A., and MacEachern, S.N. (1996), "A semiparametric Bayesian model for randomised block designs," *Biometrika*, **83**, 275-285.

Christensen, O.F., and Waagepetersen, R. (2002), "Bayesian prediction of spatial count data using generalized linear mixed models," *Biometrics*, **58**, 280-286.

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

Cressie, N.A.C., and Chan, N.H. (1989), "Spatial modeling of regional variables," *Journal of the American Statistical Association*, **84**, 393-401.

De Iorio, M., Müller, P., Rosner, G.L., and MacEachern, S.N. (2004), "An ANOVA model for dependent random measures," *Journal of the American Statistical Association*, **99**, 205-215.

Diggle, P.J., Tawn, J.A., and Moyeed, R.A. (1998), "Model-based geostatistics (with discussion)," *Applied Statistics*, **47**, 299-350.

Diggle, P., Moyeed, R., Rowlingson, B., and Thomson, M. (2002), "Childhood malaria in the Gambia: A case-study in model-based geostatistics," *Applied Statistics*, **51**, 493-506.

Duan, J.A., Guindani, M., and Gelfand, A.E. (2005), "Generalized spatial DIrichlet process models," ISDS Discussion Paper 2005-23, Duke University.

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

Escobar, M.D., and West, M. (1995), "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, **90**, 577-588.

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

Gangnon, R.E., and Clayton, M.K. (2000), "Bayesian detection and modeling of spatial disease clustering," *Biometrics*, **56**, 922-935.

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.

Giudici, P., Knorr-Held, L., and Rasser, G. (2000), "Modelling categorical covariates in Bayesian disease mapping by partition structures," *Statistics in Medicine*, **19**, 2579-2593.

Green, P.J., and Richardson, S. (2002), "Hidden Markov models and disease mapping," *Journal of the American statistical Association*, **97**, 1055-1070.

Griffin, J.E., and Steel, M.F.J. (2006), "Order-based dependent Dirichlet processes," *Journal of the American statistical Association*, **101**, 179-194.

Heagerty, P.J., and Lele, S.R. (1998), "A composite likelihood approach to binary spatial data," *Journal of the American Statistical Association*, **93**, 1099-1111.

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., 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), "Modeling Disease Incidence Data with Spatial and Spatio-temporal Dirichlet Process Mixtures," Technical Report AMS 2006-04, Department of Applied Mathematics and Statistics, University of California, Santa Cruz.

Lawson, A.B., and Clark, A. (2002), "Spatial mixture relative risk models applied to disease mapping," *Statistics in Medicine*, **21**, 359-370.

MacEachern, S.N. (1999), "Dependent nonparametric processes," *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association, pp. 50-55.

MacEachern, S.N., and Müller, P. (1998), "Estimating mixture of Dirichlet process models," *Journal of Computatioanl and Graphical Statistics*, **7**, 223-238.

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.

Neal, R.M. (2000), "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, **9**, 249-265.

Pascutto, C., Wakefield, J.C., Best, N.G., Richardson, S., Bernardinelli, L., Staines, A., and Elliott, P. (2000), "Statistical issues in the analysis of disease mapping data," *Statistics in Medicine*, **19**, 2493-2519.

Schlattmann, P., and Böhning, D. (1993), "Mixture models and disease mapping," *Statistics in Medicine*, **12**, 1943-1950.

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.

Teh, Y.W., Jordan, M.I., Beal, M.J., and Blei, D.M. (2006), "Hierarchical Dirichlet processes," To appear in the *Journal of the American Statistical Association*.

Wakefield, J., and Morris, S. (1999), "Spatial dependence and errors-in-variables in environmental epidemiology," in Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (eds), *Bayesian Statistics 6*. Oxford University Press, pp. 657-684.

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.

West, M., Müller, P., and Escobar, M.D. (1994), "Hierarchical priors and mixture models, with application in regression and density estimation," in Smith, A.F.M. and Freeman, P. (eds), *Aspects of Uncertainty: A Tribute to D.V. Lindley*. New York: Wiley, pp. 363-386.