Bayesian nonparametric modeling for disease incidence data

Athanasios Kottas *

1 Introduction

Disease incidence or mortality data are routinely recorded as summary counts for contiguous geographical regions (e.g., census tracts, zip codes, districts, or counties) and collected over discrete time periods. The count responses are typically accompanied by covariate information associated with the region (e.g., median family income, or percent with a specific type of education), and occasionally, by covariate information associated with each incidence case (e.g., sex, race, age), even though we only know the region into which the case falls. A key inferential objective in the analysis of disease incidence data is identification and explanation of spatial and spatio-temporal patterns of disease risk (disease mapping). Also of interest is forecasting of disease risk.

The statistical literature of the past twenty five years or so has witnessed a growing emphasis on fairly sophisticated methods to model heterogeneity in disease event rates. Most of the methodology has been developed within a hierarchical framework through introduction of spatial and spatio-temporal models tailored to the disease mapping inference goals. In this context, the Bayesian approach to modeling and inference is naturally attractive.

In this chapter, we review Bayesian nonparametric spatial and spatio-temporal modeling approaches for disease incidence data. Section 2 provides the necessary background on nonparametric priors, mainly the Dirichlet process prior, and its extension to spatial Dirichlet process models. Bayesian nonparametric work has focused on modeling methods for the stochastic mechanism that generates the region-specific count responses, and this is where we place the emphasis in this review, providing only brief discussion on modeling the covariate information. We thus focus on distributional specifications for the disease incidence counts (number of observed cases of the disease). These count responses are denoted by y_{it}, where $i = 1, ..., m$ indexes the geographic regions A_i, and $t = 1, ..., T$ indexes the (discrete) time periods. Note that although cases occur at specific spatial point locations, the available responses are associated with entire

*Athanasios Kottas (thanos@soe.ucsc.edu) is Professor, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA, 95064, USA. The author wishes to thank Andrew Lawson and Sudipto Banerjee for the invitation to write this book chapter, as well as a reviewer for useful feedback. This work was supported in part by the National Science Foundation under award DMS 1310438.

subregions, $A_1, ..., A_m$, that form a partition of the study region. As a consequence, there exist two distinct perspectives to model formulation. The more straightforward approach is to develop the hierarchical spatio-temporal model building the spatial dependence through a finite set of spatial random effects, one for each region. An alternative prior specification approach emerges by modeling the underlying continuous-space disease risk surface, which yields an implied prior for the finite dimensional distribution of the spatial random effects through aggregation of the continuous surface. Sections 3 and 4 discuss Bayesian nonparametric methods under these two modeling frameworks. In the former case, nonparametric mixtures of Poisson distributions have been used to model directly the distribution for the disease incidence counts (e.g., Hossain et al., 2013). In the latter setting, nonparametric spatial or spatio-temporal prior models have been considered for the disease risk (or rate) surface (e.g., Kelsall and Wakefield, 2002; Kottas et al., 2008). Finally, Section 5 provides concluding remarks.

2 Background on Bayesian nonparametrics

Bayesian nonparametric methods enable flexible modeling and inference for a wide range of problems, since they built from prior probability models for entire spaces of distributions or functions instead of scalar or vector parameters as in traditional parametric Bayesian modeling. Such methods have led to substantive applications in several fields, since they free the data analyst from customary parametric modeling restrictions yielding more accurate inference and more reliable predictions. Bayesian nonparametrics is at this point a burgeoning area of Bayesian statistics; we refer to Hjort et al. (2010) and Müller and Mitra (2013) for general reviews of related theory, methods, and applications.

Most of the Bayesian nonparametric methods for disease mapping discussed here are based on (extensions of) the Dirichlet process (DP) prior (Ferguson, 1973), the earliest example of a nonparametric prior for spaces of distributions. The DP can be defined in terms of two parameters: a parametric baseline distribution G_0, which defines the expectation of the process; and a scalar parameter $\alpha > 0$, which can be interpreted as a precision parameter, since larger α values result in DP realizations that are *closer* to G_0. We use $G \sim \text{DP}(\alpha, G_0)$ to denote that a DP prior, with parameters α and G_0, is placed on random distribution G. The most widely used definition of the DP is the constructive definition given by Sethuraman (1994). According to this definition, a distribution G generated from a $\text{DP}(\alpha, G_0)$ prior is (almost surely) of the form $G = \sum_{i=1}^{\infty} w_i \, \delta_{\vartheta_i}$, where δ_x denotes a point mass at x. Here, the ϑ_i are i.i.d. from G_0, and the weights are constructed through a *stick-breaking* procedure, specifically, $w_1 = \zeta_1$, $w_i = \zeta_i \prod_{k=1}^{i-1}(1-\zeta_k), i = 2, 3, \ldots$, with the ζ_k i.i.d. $\text{Beta}(1, \alpha)$; moreover, the sequences $\{\zeta_k : k = 1, 2, \ldots\}$ and $\{\vartheta_i : i = 1, 2, \ldots\}$ are independent. Hence, the DP generates discrete

distributions that can be represented as countable mixtures of point masses, with locations drawn independently from G_0 and weights generated according to a stick-breaking mechanism based on i.i.d. draws from a Beta$(1, \alpha)$ distribution.

A natural way to increase the applicability of DP-based modeling is by using the DP as a prior for the mixing distribution in a mixture model with a parametric kernel density function (or probability mass function) $k(\cdot \mid \boldsymbol{\theta})$. This approach yields the class of DP mixture models, which can be generically expressed as $f(\cdot \mid G) = \int k(\cdot \mid \boldsymbol{\theta})\, \mathrm{d}G(\boldsymbol{\theta})$, with $G \sim \mathrm{DP}(\alpha, G_0)$. The model is typically extended by adding hyperpriors to the DP precision parameter α and/or the parameters of G_0. Semiparametric versions are also possible by mixing on only a portion of the kernel parameters. The kernel can be chosen to be a (possibly multivariate) continuous distribution, thus overcoming the discreteness of the DP. In fact, the discreteness of G is an asset in this context, since, given the data, it enables ties among the corresponding mixing parameters. Thus, the class of DP mixture models offers an appealing choice for applications where clustering is anticipated, as in, e.g., density estimation, classification, and regression, and is indeed the most widely used Bayesian nonparametric method in applications.

The DP constructive definition has motivated extensions in several directions. One such extension is the spatial DP (Gelfand et al., 2005), a nonparametric prior for the distribution of random fields $\boldsymbol{W}_D = \{W(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ over a region $D \subseteq R^d$. To model the distribution of $\boldsymbol{W}_D$, the atoms in the DP stick-breaking representation, ϑ_i, are extended to realizations from a random field, $\boldsymbol{\vartheta}_{i,D} = \{\vartheta_i(\boldsymbol{s}) : \boldsymbol{s} \in D\}$. Thus, G_0 is extended to a spatial stochastic process G_{0D} over D. For instance, a Gaussian process (GP) can be used for G_{0D}. The resulting spatial DP provides a random distribution for $\boldsymbol{W}_D$, with realizations G_D given by $\sum_{i=1}^{\infty} w_i \delta_{\boldsymbol{\vartheta}_{i,D}}$. Consequently, for any (finite) set of spatial locations in D, $\boldsymbol{s} = (\boldsymbol{s}_1, ..., \boldsymbol{s}_M)$, G_D induces a random distribution $G_{\boldsymbol{s}}^{(M)}$ for $(W(\boldsymbol{s}_1), ..., W(\boldsymbol{s}_M))$. In fact, $G_{\boldsymbol{s}}^{(M)} \sim \mathrm{DP}(\alpha, G_0^{(M)})$, where $G_0^{(M)}$ is the M-variate normal distribution induced by the GP for G_{0D} at $(\boldsymbol{s}_1, ..., \boldsymbol{s}_M)$. It can be shown that the random process G_D yields non-Gaussian finite dimensional distributions, has nonconstant variance, and is nonstationary, even if it is centered around a stationary GP G_{0D}. Moreover, if G_{0D} has continuous sample paths, then as the distance between two spatial locations $\boldsymbol{s}$ and $\boldsymbol{s}'$ becomes smaller, the difference between distributions $G_{\boldsymbol{s}}$ and $G_{\boldsymbol{s}'}$ becomes *smaller*; formal details can be found in Gelfand et al. (2005) and Guindani and Gelfand (2006). For alternative constructions of nonparametric prior models for spatial random surfaces, we refer to Griffin and Steel (2006), Duan et al. (2007), Reich and Fuentes (2007), and Rodriguez and Dunson (2011). Bayesian nonparametric mixture modeling has also been explored for spatial (marked) point processes, including the work of Wolpert and Ickstadt (1998), Ishwaran and James (2004), Kottas and Sansó (2007), and Taddy and Kottas (2012).

3 Nonparametric mixture modeling for incidence counts

In order to model heterogeneity or discontinuity in disease event rates one is naturally led to mixture models. Indeed, various types of parametric mixture model specifications have been explored in the disease mapping literature. Chapter 20 in this volume reviews several of the existing methods based on both spatial and spatio-temporal mixture models. As discussed in Section 2, a significant portion of methodological and applied work in Bayesian nonparametrics has built from (countable) nonparametric mixtures. It is therefore natural to consider nonparametric mixture prior models for the disease incidence count distribution in order to expand on the inferential power of corresponding parametric finite mixtures.

To arrive at a generic mixture model formulation, and to fix notation, recall the typical assumption for the disease incidence counts: $y_{it} \mid \theta_{it} \overset{ind.}{\sim} \text{Poisson}(y_{it} \mid e_{it}\theta_{it})$, that is, conditionally on parameters θ_{it}, the y_{it} are independent Poisson distributed with mean $e_{it}\theta_{it}$. Here, e_{it} is the expected disease count for region i at time period t, and θ_{it} is the associated relative risk. The expected counts are typically computed through $e_{it} = R\, n_{it}$, where n_{it} is the specified number of individuals at risk in region i at time t, and R is an overall disease rate. The given data set can be used to obtain R, for example, $R = \sum_{i,t} y_{it} / \sum_{i,t} n_{it}$ (internal standardization), or R can be developed from reference tables (external standardization). Standard parametric hierarchical models explain the relative risk parameters through different types of random effects. For instance, a specification with random effects additive in space and time is

$$\log \theta_{it} = \mu_{it} + u_i + v_i + \delta_t \tag{1}$$

where μ_{it} is a component for the regional covariates, u_i are regional random effects (typically, assumed i.i.d. from a zero-mean normal distribution), δ_t are temporal effects (say, with an autoregressive prior), and v_i are spatial random effects with prior typically built from a conditional autoregressive (CAR) structure. For further details, we refer to Banerjee et al. (2015), as well as to chapters 7 and 15 in this volume. When spatio-temporal interaction is sought, $v_i + \delta_t$ in (1) may be replaced by space-time random effects v_{it}, which have been modeled using independent CAR structures over time, dynamically with independent CAR innovations, or as a CAR in space and time; see chapter 19 in this volume for related references.

A general (finite) mixture model formulation arises by replacing the continuous mixing distribution for the v_i (or the v_{it}), implied by the CAR prior, with a discrete distribution taking K possible values. These values represent the relative risks for K underlying space or space-time clusters, and have corresponding mixing weights that form probability vector(s) on the $(K-1)$-dimensional simplex. The simplest form for the discrete mixing distribution involves values ϕ_j with corresponding probabilities ω_j, for $j = 1, ..., K$, which upon marginalization over the

random effects, results in the mixture $\sum_{j=1}^{K} \omega_j \text{Poisson}(y_{it} \mid e_{it}\phi_j)$ (e.g., Böhning et al., 2000). In the setting without a temporal component, related is the work of Knorr-Held and Rasser (2000), Denison and Holmes (2001), and Hegarty and Barry (2008) based on spatial partition structures, which divide the study region into a number of distinct clusters (sets of contiguous regions) with constant relative risk, assuming a priori random number, size, and location for the clusters. More flexible mixture model specifications arise through use of spatially dependent vectors of weights, ω_{ij}, such that $\sum_{j=1}^{K} \omega_{ij} = 1$, for all regions i. For instance, for spatial only incidence counts, Fernández and Green (2002) model ω_{ij} through a logistic transformation, $\omega_{ij} = \exp(\eta_{ij}/\psi)/\sum_{\ell=1}^{K} \exp(\eta_{i\ell}/\psi)$, where the vectors $(\eta_{1j}, ..., \eta_{mj})$, for $j = 1, ..., K$, arise conditionally independent from a Markov random field prior model, inducing spatial dependence to the collection of weights $(\omega_{1j}, ..., \omega_{mj})$ for each mixture component. Finally, as discussed in chapter 20, for spatio-temporal disease incidence data one can envision the further extension to space-time dependent weights, ω_{itj}, where now $\sum_{j=1}^{K} \omega_{itj} = 1$, for any region i and any time period t.

As a nonparametric version of this last modeling scenario, Hossain et al. (2013) developed mixtures of Poisson distributions for the count responses with weights that depend on both space and time. In particular, Hossain et al. (2013) propose the model

$$y_{it} \mid \boldsymbol{\omega}, \boldsymbol{\phi} \sim \sum_{j=1}^{K} \omega_{itj} \text{Poisson}(y_{it} \mid e_{it}\phi_j) \tag{2}$$

where $\boldsymbol{\phi} = (\phi_1, ..., \phi_K)$ is the vector of mixing relative risk parameters that are common across both space and time, and $\boldsymbol{\omega} = \{(\omega_{it1}, ..., \omega_{itK}) : i = 1, ..., m; t = 1, ..., T\}$ collects the space-time dependent mixing weights, which satisfy $\sum_{j=1}^{K} \omega_{itj} = 1$, for any i and t. The weights are defined through $\omega_{itj} = B_{itj} \prod_{k=1}^{j-1}(1 - B_{itk})$. Here, $B_{itj} = q_{itj}\zeta_j$, where q_{itj} is a space-time dependent kernel function (that may also depend on regional covariates), and the ζ_j are i.i.d. Beta$(1, \alpha)$. This construction can be recognized as an extension of the stick-breaking representation for the DP weights, and is an example of a kernel stick-breaking process (Dunson and Park, 2008). The kernel function is specified by an extension of the logistic form from Fernández and Green (2002). More specifically, $q_{itj} = \exp(\eta_{itj}/\psi)/\sum_{\ell=1}^{K} \exp(\eta_{it\ell}/\psi)$, where $\psi > 0$ plays the role of a smoothness parameter, and the η_{itj} may be modeled through different types of spatial and temporal random effects, and possibly also as a function of time-varying regional covariates. Hossain et al. (2013) implemented the model in (2) with a specified number of components K, which is a finite truncation approximation to the general kernel stick-breaking mixture model that involves a countable number of components in the prior. This approximation allows ready Markov chain Monte Carlo (MCMC) posterior simulation for the model through use of routine techniques for discrete finite mixture models.

Structured nonparametric mixtures of Poisson distributions are also explored in Li et al. (2015) for spatial only count data. In particular, the development of their "areal referenced spatial stick breaking process" prior model is along the lines of the finite mixture models discussed above. Using the previous notation, but excluding the time component, that model involves the first-stage specification: $y_i \mid \mu_i, v_i \overset{ind.}{\sim} \text{Poisson}(y_i \mid e_i \exp(\mu_i + v_i))$, for $i = 1, ..., m$, where the μ_i are defined through a linear regression on regional covariates. The prior for the spatial random effects is built from $v_i \mid G^{(i)} \overset{ind.}{\sim} G^{(i)}$, where $G^{(i)} = \sum_{j=1}^{K} \omega_{ij} \delta_{\varphi_j}$, with a normal prior for the mixing parameters φ_j, and a kernel stick-breaking structure for the weights. More specifically, $\omega_{ij} = q_{ij}\zeta_j \prod_{k=1}^{j-1}(1 - q_{ik}\zeta_k)$, where the ζ_j are i.i.d. Beta$(1, \alpha)$, and a CAR prior on the logit scale is used for $(q_{1j}, ..., q_{mj})$. Evidently, collapsing the two stages of the model by marginalizing over the v_i, yields a finite mixture of Poisson distributions with spatially dependent vectors of weights, in the spirit of the model specification in Fernández and Green (2002). The focus in Li et al. (2015) is on detection of boundaries between disparate neighboring regions. This is formulated as a multiple hypothesis testing problem, based on the posterior probabilities of the events $\{v_i = v_{i'}\}$ for each pair of adjacent regions A_i and $A_{i'}$, for which discreteness is a necessary property of the prior probability model for the spatial random effects.

Finally, we note that a different semiparametric extension of the hierarchical structure for the relative risks in (1) can be developed by replacing the normal distribution for the regional random effects, u_i, with a nonparametric prior. Malec and Müller (2008) provide an example in the context of small area estimation with binary responses under a setting that involves multivariate regional random effects, modeled with a DP mixture of multivariate normal distributions.

4 Nonparametric prior models for continuous-space risk surfaces

Here, we review methods that build the hierarchical model for disease incidence counts (and related covariates) from prior models for the underlying continuous-space relative risk (or rate) surface, which is aggregated to provide the induced prior for the relative risk (or rate) spatial random effects. Although less commonly used in disease mapping, this approach offers a more coherent modeling framework, since it avoids the dependence of the prior model on the data collection procedure (e.g., the number, shapes, and sizes of the regions in the particular study), and it can more naturally accommodate data sources available at different levels of spatial aggregation. Focusing in all cases on the modeling for the spatial or space-time random effects, Section 4.1 provides an overview of two methods for spatial only data (Best et al., 2000; Kelsall and Wakefield, 2002), and Section 4.2 discusses a spatio-temporal approach (Kottas et al., 2008).

4.1 Spatial models

For disease counts recorded over space only, Kelsall and Wakefield (2002) construct a hierarchical model the first stage of which assumes $y_i \mid \theta_i \overset{ind.}{\sim} \text{Poisson}(y_i \mid e_i\theta_i)$, for $i = 1, ..., m$ (again, the notation is similar to the one in Section 3, excluding the temporal component). This familiar specification is derived through Poisson process continuous-space models for the area/stratum population at risk and corresponding cases. In particular, the population at risk within stratum k is assumed to follow a non-homogeneous Poisson process (NHPP) with intensity $\lambda_k(\boldsymbol{s})$, and the cases are viewed as a spatial point pattern from a NHPP with intensity $\lambda_k(\boldsymbol{s})p_k(\boldsymbol{s})$, where $p_k(\boldsymbol{s})$ denotes the probability of disease for stratum k at location $\boldsymbol{s}$. It is further assumed that $p_k(\boldsymbol{s}) = p_k\theta_k(\boldsymbol{s})$, where $\theta_k(\boldsymbol{s})$ is the relative risk for stratum k at location $\boldsymbol{s}$, and p_k is the reference disease probability in stratum k. For locations $\boldsymbol{s}$ in region A_i, the intensity $\lambda_k(\boldsymbol{s})$ is conceptualized as $\lambda_k(\boldsymbol{s}) = N_{ik}f_{ik}(\boldsymbol{s})$, where N_{ik} is the area/stratum population count, and $f_{ik}(\boldsymbol{s})$ is the density function for the spatial distribution of the population in stratum k and region A_i. Then, using (for rare diseases) the Poisson approximation to the Binomial distribution for the number of cases in stratum k and region A_i, and summing over k, the distribution for y_i arises as Poisson with mean $\sum_k N_{ik}p_k \int_{A_i} \theta_k(\boldsymbol{s})f_{ik}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$. The mean can be re-written as $e_i \sum_k w_{ik}\theta_{ik}$, where $e_i = \sum_k N_{ik}p_k$ and $\theta_{ik} = \int_{A_i} \theta_k(\boldsymbol{s})f_{ik}(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$, such that the relative risk for region A_i is given by $\theta_i = \sum_k w_{ik}\theta_{ik}$, a weighted average of stratum-specific relative risks θ_{ik}, with weights $w_{ik} = N_{ik}p_k/e_i$, the expected proportions of cases in region A_i that are in stratum k. Kelsall and Wakefield (2002) make two simplifying assumptions/approximations to arrive at the final form for the θ_i. First, the relative risk surface is assumed constant across strata, that is, $\theta_k(\boldsymbol{s}) = \theta(\boldsymbol{s})$, for all k and $\boldsymbol{s}$. This results in $\theta_i = \int_{A_i} \theta(\boldsymbol{s})f_i(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$, where $f_i(\boldsymbol{s}) = \sum_k w_{ik}f_{ik}(\boldsymbol{s})$ is a weighted average of the stratum-specific population densities over region A_i. Finally, the population density is assumed uniform across regions (an assumption implicitly made in most disease mapping modeling approaches), such that $f_i(\boldsymbol{s}) = |A_i|^{-1}$, for all $\boldsymbol{s} \in A_i$.

The model is completed with a prior for the vector of log-relative risk random effects, $(\log(\theta_1), ..., \log(\theta_m))$, where, based on the argument above, $\theta_i = |A_i|^{-1} \int_{A_i} \theta(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$. This prior is given by an m-dimensional normal distribution with a structured covariance matrix H, induced by a GP prior for the underlying log-relative risk surface, $\{\log(\theta(\boldsymbol{s})) : \boldsymbol{s} \in D\}$, where D is the region under study. Kelsall and Wakefield (2002) use an isotropic GP with cubic correlation function defined in terms of a single range of dependence parameter. For computational feasibility, in particular, to simplify the form of the elements for the covariance matrix H, an additional approximation is applied. More specifically, the distribution of $\log(\theta_i) = \log(|A_i|^{-1} \int_{A_i} \theta(\boldsymbol{s})\mathrm{d}\boldsymbol{s})$ is approximated by the distribution of $|A_i|^{-1} \int_{A_i} \log(\theta(\boldsymbol{s}))\mathrm{d}\boldsymbol{s}$. Posterior inference under the model is implemented using standard MCMC methods for GP-based models. A benefit of the continuous-space modeling approach is that, in addition to estimation for the relative risks,

predictive inference for the underlying relative risk surface is also possible.

Similar in spirit is also the contribution of Best et al. (2000), although their data structure does not exactly fit within the standard disease mapping setting. This work develops a spatial regression model to study the effect of traffic pollution on respiratory disorders in children. The particular study encompasses data on: the residential postcode for "severe wheezing" cases; case-specific individual attributes (e.g., age, gender) and home environment covariates (e.g., home dampness, maternal smoking); population density available at a district level; and traffic pollution levels available on a spatial grid. The modeling approach incorporates the incidence cases essentially as spatially referenced data, using the centroid locations of the home postcode for all 191 cases; five postcodes contain 2 cases each, with the remaining 181 cases corresponding to a unique postcode. Best et al. (2000) model these locations of severe wheezing cases along with the individual attributes as realizations from a marked NHPP process, using a semiparametric formulation for its intensity measure. The parametric component of the intensity incorporates the information from the covariates, population density, and the risk factor (traffic pollution). The nonparametric component models the (continuous-space) spatial random effects through a kernel mixture with a gamma process for the mixing measure, using the approach in Wolpert and Ickstadt (1998). The specific model for the spatial random effects can also be represented as a DP mixture, using the direct connection between the gamma process and the Dirichlet process.

4.2 Space-time modeling

The model of Kelsall and Wakefield (2002) is extended in Kottas et al. (2008), where a spatial DP prior is used for the underlying disease rate surface under a dynamic setting that handles disease incidence data collected over both space and time. We first discuss the spatial component of the modeling approach, followed by the dynamic spatial extension.

The first stage of the hierarchical model involves Poisson$(y_{it} \mid n_{it} \exp(\gamma_{it}))$ distributions, where n_{it} is the number of individuals at risk for region A_i and time t, and $p_{it} = \exp(\gamma_{it})$ is the corresponding disease rate (with rare diseases, the logarithmic and logit transformations are practically equivalent). Kottas et al. (2008) argue for this form for the Poisson mean instead of $e_{it}\theta_{it}$, since it avoids the need to develop the e_{it} through standardization. The γ_{it} are viewed as log-rate spatial effects arising by aggregating log-rate surfaces $\boldsymbol{\gamma}_{t,D} = \{\gamma_t(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ over the regions A_i. That is, $\gamma_{it} = |A_i|^{-1} \int_{A_i} \gamma_t(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$ is the block average of the surface $\boldsymbol{\gamma}_{t,D}$ over region A_i. Similar to Kelsall and Wakefield (2002), the first-stage Poisson specification is derived through aggregation of an underlying NHPP under certain assumptions and approximations. For time period t, the disease incidence cases are assumed to follow a NHPP with intensity function $n_t(\boldsymbol{s})p_t(\boldsymbol{s})$, where $\{n_t(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ is the population density surface, and $p_t(\boldsymbol{s}) = \exp(\gamma_t(\boldsymbol{s}))$ is the disease rate at time t and location $\boldsymbol{s}$. Assuming a uniform population density over each

region at each time period, $n_t(\boldsymbol{s}) = n_{it}|A_i|^{-1}$, for $\boldsymbol{s} \in A_i$. Hence, aggregating the NHPP over the regions A_i, each y_{it} follows a Poisson distribution with mean $\int_{A_i} n_t(\boldsymbol{s})p_t(\boldsymbol{s})\mathrm{d}\boldsymbol{s} = n_{it}p^*_{it}$, where $p^*_{it} = |A_i|^{-1}\int_{A_i} p_t(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$. If the distribution of the p^*_{it} is approximated by the distribution of the $\exp(\gamma_{it})$, one obtains $y_{it} \mid \gamma_{it} \overset{ind.}{\sim} \text{Poisson}(y_{it} \mid n_{it}\exp(\gamma_{it}))$ for the first stage distribution.

To develop the prior model for the spatial log-rate random effects, first, the log-rate surfaces $\boldsymbol{\gamma}_{t,D}$ are taken as realizations from a mean-zero isotropic GP with variance σ^2 and exponential correlation function $\exp(-\varphi||\boldsymbol{s} - \boldsymbol{s}'||)$. The induced distribution for $\boldsymbol{\gamma}_t = (\gamma_{1t}, ..., \gamma_{mt})$ is an m-dimensional normal with covariance matrix $\sigma^2 H(\phi)$, where the (i,j)-th element of $H(\phi)$ is given by $|A_i|^{-1}|A_j|^{-1}\int_{A_i}\int_{A_j}\exp(-\varphi||\boldsymbol{s} - \boldsymbol{s}'||)\mathrm{d}\boldsymbol{s}\mathrm{d}\boldsymbol{s}'$. Next, a DP prior is assumed for the distribution of the $\boldsymbol{\gamma}_t$ with centering distribution given by the m-dimensional normal above, $\text{N}_m(\boldsymbol{0}, \sigma^2 H(\phi))$. The choice of the DP in this context allows for data-driven deviations from the normality assumption for the spatial random effects.

Note that this structure implies for the vector of counts $\boldsymbol{y}_t = (y_{1t}, ..., y_{mt})$ a Poisson DP mixture model: $\int\prod_{i=1}^m \text{Poisson}(y_{it} \mid n_{it}\exp(\gamma_{it}))\mathrm{d}G(\boldsymbol{\gamma}_t)$, where $G \sim \text{DP}(\alpha, \text{N}_m(\boldsymbol{0}, \sigma^2 H(\phi)))$. To overcome the discreteness of the distribution for the log-rate vectors (induced by the discreteness of DP realizations), the DP prior for the $\boldsymbol{\gamma}_t$ can be replaced with a DP mixture prior,

$$\boldsymbol{\gamma}_t \mid \tau^2, G \sim \int \text{N}_m(\boldsymbol{\gamma}_t \mid \boldsymbol{\gamma}^*_t, \tau^2 I_m)\mathrm{d}G(\boldsymbol{\gamma}^*_t), \quad G \sim \text{DP}(\alpha, \text{N}_m(\boldsymbol{0}, \sigma^2 H(\phi))),$$

where $\boldsymbol{\gamma}^*_t = (\gamma^*_{1t}, ..., \gamma^*_{mt})$. This extension essentially involves the introduction of a heterogeneity effect, writing $\gamma_{it} = \gamma^*_{it} + u_{it}$, with u_{it} i.i.d. $\text{N}(0, \tau^2)$. The mixture model for the $\boldsymbol{y}_t$ now becomes $f(\boldsymbol{y}_t \mid \tau^2, G) = \int\prod_{i=1}^m p(y_{it} \mid \tau^2, \gamma^*_{it})\mathrm{d}G(\boldsymbol{\gamma}^*_t)$, where $p(y_{it} \mid \tau^2, \gamma^*_{it}) = \int \text{Poisson}(y_{it} \mid n_{it}\exp(\gamma_{it}))\text{N}(\gamma_{it} \mid \gamma^*_{it}, \tau^2)\mathrm{d}\gamma_{it}$ is a Poisson-lognormal mixture. In full hierarchical form, the model is given by

$$\begin{array}{rcl} y_{it} \mid \gamma_{it} & \overset{ind.}{\sim} & \text{Poisson}(y_{it} \mid n_{it}\exp(\gamma_{it})), \;\; i = 1, ..., m, \;\; t = 1, ..., T \\ \gamma_{it} \mid \gamma^*_{it}, \tau^2 & \overset{ind.}{\sim} & \text{N}(\gamma_{it} \mid \gamma^*_{it}, \tau^2), \;\; i = 1, ..., m, \;\; t = 1, ..., T \\ \boldsymbol{\gamma}^*_t \mid G & \overset{i.i.d.}{\sim} & G, \;\; t = 1, ..., T \\ G \mid \alpha, \sigma^2, \phi & \sim & \text{DP}(\alpha, \text{N}_m(\boldsymbol{0}, \sigma^2 H(\phi))). \end{array} \tag{3}$$

The model is completed with independent hyperpriors for τ^2 and for the DP prior parameters.

Both to establish the connection with a spatial DP prior, and for MCMC posterior simulation, it is useful to marginalize the random mixing distribution G in (3) over its DP prior (Blackwell and MacQueen, 1973). The resulting joint prior distribution for the $\boldsymbol{\gamma}^*_t$ is given by

$$\text{N}_m(\boldsymbol{\gamma}^*_1 \mid \boldsymbol{0}, \sigma^2 H(\phi))\prod_{t=2}^T \left\{ \frac{\alpha}{\alpha + t - 1}\text{N}_m(\boldsymbol{\gamma}^*_t \mid \boldsymbol{0}, \sigma^2 H(\phi)) + \frac{1}{\alpha + t - 1}\sum\nolimits_{j=1}^{t-1} \delta_{\boldsymbol{\gamma}^*_j}(\boldsymbol{\gamma}^*_t) \right\}. \tag{4}$$

Hence, the $\boldsymbol{\gamma}_t^*$ arise according to a Pólya urn scheme, which highlights the DP-induced clustering: $\boldsymbol{\gamma}_1^*$ is drawn from the centering distribution, and then for each $t = 2, ..., T$, $\boldsymbol{\gamma}_t^*$ is either set equal to $\boldsymbol{\gamma}_j^*$, $j = 1, ..., t-1$, with probability $(\alpha + t - 1)^{-1}$ or is drawn from the centering distribution with the remaining probability.

The prior model for the spatial random effects $\boldsymbol{\gamma}_t^*$ discussed above is defined starting with a GP prior for the corresponding surfaces $\{\gamma_t^*(\boldsymbol{s}) : \boldsymbol{s} \in D\}$, block averaging the associated GP realizations over the regions to obtain the $\mathrm{N}_m(\mathbf{0}, \sigma^2 H(\phi))$ distribution, and finally centering a DP prior for the $\boldsymbol{\gamma}_t^*$ around this m-dimensional normal distribution. Kottas et al. (2008) show that the joint prior distribution for the $\boldsymbol{\gamma}_t^*$ is exactly as in (4) if one starts instead with a spatial DP prior for the distribution G_D of the $\{\gamma_t^*(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ (centered around the same isotropic GP used above), marginalizes G_D over its spatial DP prior, and then block averages the (marginal) realizations from the spatial DP prior over the regions. Hence, the marginal version of model (3) (which is the one used for posterior predictive inference) is consistent with the marginal version of the corresponding (continuous-space) spatial DP mixture model, regardless of the number and geometry of the subregions chosen to partition the region under study.

Finally, to extend the spatial model described above to a spatio-temporal setting, Kottas et al. (2008) use a dynamic spatial modeling framework, viewing the log-rate process $\boldsymbol{\gamma}_{t,D}$ as a temporally evolving spatial process. In particular, the log-rate surface is modeled as $\gamma_t(\boldsymbol{s}) = \xi_t + \gamma_t^*(\boldsymbol{s})$, adding temporal structure through transition equations for the $\gamma_t^*(\boldsymbol{s})$. (Note that, both here as well as in the spatial model discussed above, a mean structure $\mu_t(\boldsymbol{s})$ would typically be added to $\gamma_t(\boldsymbol{s})$ to incorporate covariate information.) For instance, $\gamma_t^*(\boldsymbol{s}) = \nu\gamma_{t-1}^*(\boldsymbol{s}) + \eta_t(\boldsymbol{s})$, where the innovations $\boldsymbol{\eta}_{t,D} = \{\eta_t(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ are independent realizations from a spatial stochastic process with distribution G_D. A spatial DP prior is assigned to G_D, with parameters α and $G_{0D} = \mathrm{GP}(\mathbf{0}, \sigma^2 \exp(-\varphi||\boldsymbol{s} - \boldsymbol{s}'||))$. Marginalizing G_D over its prior, the induced joint prior $p(\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T \mid \alpha, \sigma^2, \phi)$ for the block averaged $\boldsymbol{\eta}_t = (\eta_{1t}, ..., \eta_{mt})$, where $\eta_{it} = |A_i|^{-1} \int_{A_i} \eta_t(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$, is given by (4) (with $\boldsymbol{\eta}_t$ replacing $\boldsymbol{\gamma}_t^*$). Block averaging the surfaces in the transition equations, results in $\boldsymbol{\gamma}_t^* = \nu\boldsymbol{\gamma}_{t-1}^* + \boldsymbol{\eta}_t$. And, adding again the i.i.d. $\mathrm{N}(0, \tau^2)$ terms to the γ_{it}, the following general form for the spatio-temporal model emerges

$$
\begin{aligned}
y_{it} \mid \gamma_{it} &\overset{ind.}{\sim} \mathrm{Poisson}(y_{it} \mid n_{it} \exp(\gamma_{it})), \;\; i = 1, ..., m, \;\; t = 1, ..., T \\
\gamma_{it} \mid \xi_t, \gamma_{it}^*, \tau^2 &\overset{ind.}{\sim} \mathrm{N}(\gamma_{it} \mid \xi_t + \gamma_{it}^*, \tau^2), \;\; i = 1, ..., m, \;\; t = 1, ..., T \\
\boldsymbol{\gamma}_t^* &= \nu\boldsymbol{\gamma}_{t-1}^* + \boldsymbol{\eta}_t \\
\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T \mid \alpha, \sigma^2, \phi &\sim p(\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T \mid \alpha, \sigma^2, \phi).
\end{aligned}
$$

The ξ_t could be i.i.d. $\mathrm{N}(0, \sigma_\xi^2)$, modeled with a parametric autoregressive structure, or explained through a parametric trend.

5 Conclusions

Although Bayesian nonparametric methodology is now an integral component of applied Bayesian modeling, applications in spatial epidemiology, and more specifically in disease mapping, are relatively limited. We have reviewed Bayesian nonparametric spatial and spatio-temporal models for disease incidence data, categorizing the modeling approaches according to whether the nonparametric prior is placed on the finite dimensional distribution of the region-specific spatial effects or, more generally, on the latent disease risk or rate surface.

From a practical point of view, more work is needed on empirical comparison between the existing methods as well as with more standard parametric hierarchical models. In terms of new methodological developments, it is arguably of interest to expand the scope of existing methods that build from modeling the underlying temporally evolving continuous-space disease risk (or rate) surfaces. In this context, it is important to elaborate on the modeling framework to handle spatial misalignment issues for 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. Finally, it would be of practical and methodological interest to explore flexible nonparametric methodology for describing and forecasting patterns of joint incidence of multiple diseases.

References

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2015). *Hierarchical Modeling and Analysis for Spatial Data (Second Edition)*. Chapman & Hall/CRC.

Best, N. G., K. Ickstadt, and R. L. Wolpert (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 J. MacQueen (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* **1**, 353–355.

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

Denison, D. G. T. and C. C. Holmes (2001). Bayesian partitioning for estimating disease risk. *Biometrics* **57**, 143–149.

Duan, J. A., M. Guindani, and A. E. Gelfand (2007). Generalized spatial Dirichlet process models. *Biometrika* **94**, 809–825.

Dunson, D. and J. Park (2008). Kernel stick-breaking processes. *Biometrika* **95**, 307–323.

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

Fernández, C. and P. J. Green (2002). Modelling spatially correlated data via mixtures: a Bayesian approach. *Journal of the Royal Statistical Society, Series B* **64**, 805–826.

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

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

Guindani, M. and A. E. Gelfand (2006). Smoothness properties and gradient analysis under spatial Dirichlet process models. *Methodology and Computing in Applied Probability* **8**, 159–189.

Hegarty, A. and D. Barry (2008). Bayesian disease mapping using product partition models. *Statistics in Medicine* **27**, 3868–3893.

Hjort, N. L., C. Holmes, P. Müller, and S. G. Walker (2010). *Bayesian Nonparametrics*. Cambridge University Press.

Hossain, M., A. B. Lawson, B. Cai, J. Choi, J. Liu, and R. S. Kirby (2013). Space-time stick-breaking processes for small area disease cluster estimation. *Environmental and Ecological Statistics* **20**, 91–107.

Ishwaran, H. and L. F. James (2004). Computational methods for multiplicative intensity models using weighted gamma processes: Proportional hazards, marked point processes, and panel count data. *Journal of the American Statistical Association* **99**, 175–190.

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

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

Kottas, A., J. Duan, and A. E. Gelfand (2008). Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures. *Biometrical Journal* **50**, 29–42.

Kottas, A. and B. Sansó (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference* **137**, 3151–3163.

Li, P., S. Banerjee, T. A. Hanson, and A. M. McBean (2015). Bayesian models for detecting difference boundaries in areal data. *Statistica Sinica* **25**, 385–402.

Malec, D. and P. Müller (2008). A Bayesian semi-parametric model for small area estimation. In B. Clarke and S. Ghosal (Eds.), *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, pp. 223–236. Beachwood, Ohio, USA: Institute of Mathematical Statistics.

Müller, P. and R. Mitra (2013). Bayesian nonparametric inference – why and how (with discussion). *Bayesian Analysis* **8**, 269–360.

Reich, B. and M. Fuentes (2007). A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics* **1**, 249–264.

Rodriguez, A. and D. B. Dunson (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis* **6**, 145–178.

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

Taddy, M. A. and A. Kottas (2012). Mixture modeling for marked Poisson processes. *Bayesian Analysis* **7**, 335–362.

Wolpert, R. L. and K. Ickstadt (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.