The International Journal of Biostatistics

Volume 8, *Issue* 1 2012 *Article* 28

A Reproducing Kernel-Based Spatial Model in Poisson Regressions

Hongmei Zhang, *University of South Carolina - Columbia*
Jianjun Gan, *GlaxoSmithKline*

Recommended Citation:
Zhang, Hongmei and Gan, Jianjun (2012) "A Reproducing Kernel-Based Spatial Model in Poisson Regressions," *The International Journal of Biostatistics*: Vol. 8: Iss. 1, Article 28.
DOI: 10.1515/1557-4679.1360

A Reproducing Kernel-Based Spatial Model in Poisson Regressions

Hongmei Zhang and Jianjun Gan

Abstract

A semi-parametric spatial model for spatial dependence is proposed in Poisson regressions to study the effects of risk factors on incidence outcomes. The spatial model is constructed through an application of reproducing kernels. A Bayesian framework is proposed to infer the unknown parameters. Simulations are performed to compare the reproducing kernel-based method with several commonly used approaches in spatial modeling, including independent Gaussian and CAR models. Compared with these models, the reproducing kernel-based method is easy to implement and more flexible in terms of the ability to model various spatial dependence patterns. To further demonstrate the proposed method, two real data applications are discussed: Scottish lip cancer data and Florida smoke-related cancer data.

KEYWORDS: semi-parametric, reproducing kernel, Gaussian kernel, CAR models, poisson regression

Author Notes: The authors thank Professor William Browne at the University of Bristol for his insights offered at the beginning of the project, Hal Stern at the University of California, Irvine, for his precious comments and suggestions, and Professor Raid Amin at the University of West Florida and Mr. Daniel Stadel for their kind help with data collection. The authors also thank Andrew Ortaglia and Meredith Ray for their help on the editing. The authors are grateful to the three referees for their constructive and insightful comments.

1 Introduction

The work presented in this article was motivated by a study evaluating the effects of risk factors on incidence outcomes through a Poisson regression model, in which spatial dependence needs to be explained. In spatial data analysis, non-specific random spatial effects are usually modeled parametrically (Lawson, 2001, Yan and Clayton, 2006). For this purpose, some studies consider identical and independent Gaussian distributions (Lawson, 2001, 2008). This assumption is acceptable if the spatial effect is unstructured and may be approximated by independent random effects. However, in many situations, this assumption does not hold and spatial dependence needs to be appropriately addressed. Conditional autoregressive (CAR) models are a popular choice to characterize spatial dependence. CAR models are usually constructed in the Gaussian framework (Cressie and Chan, 1989, Besag, Mollie, York, and Mollié, 1991), although double exponential (Laplace) distributions are utilized as well (Best, Arnold, Thomas, Waller, and Conlon, 1999). Despite the popularity of CAR models, they have some critical limitations. First, constructing a positive definite covariance matrix in a joint distribution of random spatial effects is not a trivial task. This limits the flexibility of CAR models in describing various patterns of spatial effects (Cressie and Chan, 1989, Stern and Cressie, 1999, Banerjee, Carlin, and Gelfand, 2004). Second, the spatial structure indicated by a CAR model may not fit well for geographical entities of different sizes and arranged in an irregular pattern (Richardson, 1992, Kelsall and Eld, 2002). Third, in the autoregressive process of fitting a CAR model of order one, information from non-neighboring regions may not be fully incorporated. This is similar to the phenomenon in a first order autoregressive process.

Inspired by these limitations, in this article, we present an alternative method to model spatial dependence and incorporate the method into the Poisson regression model. It is a Bayesian semi-parametric method built on reproducing kernels. This method utilizes information from all event locations when evaluating conditional spatial effects of each specific region. Due to the implementation of a reproducing kernel, the method is able to describe various forms of spatial dependence and to provide informative evaluation on the strength of spatial effects (Liu, Lin, and Ghosh, 2007, Liu, Ghosh, and Lin, 2008). To our knowledge, reproducing kernels have not been considered in the area of spatial modeling.

Assuming the data are continuous, reproducing kernels are related to variogram models in geostatistics in that both methods involve an evaluation of distance-based correlations between different regions (Cressie, 1985, Ecker and Gelfand, 1997, Gorsich and Genton, 2000). However, these two methods have fundamental differences. Reproducing kernels evaluate spatial random effects of a specific region using a weighted average over all regions' contributions; correlations between

regions only partially control the region's spatial effects. In contrast, a variogram model focuses on the evaluation of distance-based correlations between regions and utilizes correlations to assess the contribution of a region.

The remainder of the article is organized into four sections. In Section 2, we introduce the reproducing kernel-based (RKB) method for spatial dependence modeling. In this section, a fully Bayesian approach is presented for data analysis. Markov Chain Monte Carlo (MCMC) related computing issues are also discussed. Simulation studies based on various criteria are included in Section 3 to examine the performance of the method. A comparison between different methods is also discussed in this section. In Section 4, we apply the RKB method to two data sets, the classical Scottish lip cancer data (Clayton and Kaldor, 1987, Breslow and Clayton, 1993) and the Florida smoke-related cancer incidence data. We summarize our findings and propose possible future work in Section 5.

2 The spatial model in Poisson regressions

In epidemiological studies, to examine risk factor effects on rare incidence rates, Poisson regression models are usually chosen (Frome and Checkoway, 1985, Zou, 2004, Feldens, Kramer, Ferreira, Spiguel, and Marquezan, 2010). In the following, we propose a spatial semi-parametric Poisson regression model taking random spatial effects into account.

2.1 The Poisson regression model

Let $Y_i, i = 1, \cdots, N$, be the event (incidence) counts in region R_i and $\mathbf{X}_i$ be a vector of risk factors. We assume that:

$$\begin{aligned} P(Y_i = y_i|\lambda_i) &= \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \\ \log(\lambda_i|\alpha, \boldsymbol{\beta}, \delta_i) &= \log(n_i) + \alpha + \boldsymbol{\beta}'\mathbf{X}_i + \delta_i, \end{aligned} \tag{1}$$

where $P(\cdot)$ denotes the probability mass function of a discrete random variable and $\boldsymbol{\beta}$ is a vector of parameters evaluating the effects of risk factors. The parameter n_i is an offset to adjust for background effect. The inclusion of offset enables us to model rates of events. In our real data applications, n_i is the population size in region R_i. The term δ_i reflects the spatial effect of region i on the average number of incidences in R_i. Under δ_i and parameters α and $\boldsymbol{\beta}$, the counts Y_i's are independent.

2.2 The reproducing kernel-based method for spatial effects

With all surrounding regions considered, implicitly the random spatial effect δ_i is a function of region locations. We assume that this function lies in a space of functions generated by a positive definite reproducing kernel, which is, under some regularity conditions, equivalent to a space of function defined by a particular set of orthogonal basis functions (Cristianini and Shawe-Taylor, 2000). The reproducing kernel-based (RKB) method has been popularized in the area of machine learning (Vapnik, 1998, Scholkopf and Smola, 2002); the support vector machine is one common example. In our application, the RKB method is used to describe distance-based random spatial effects of any form. Kindermann and Snell (1980) and Cressie and Chan (1989) also used the RKB method to evaluate non-distance based spatial effects.

To define the random spatial effect δ_i using reproducing kernels, we apply the representer theorem (Kimeldorf and Wahba, 1970, O'Sullivan, Yandell, and Raynor, 1986). Based on this theorem, δ_i can be defined as a linear combination of kernels,

$$\delta_i \;=\; \mathbf{k}_i'\boldsymbol{\eta} = \sum_{j=1}^{N} \eta_j k(s_i, s_j), \tag{2}$$

where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_N)'$ is a vector of unknown parameters, $(s_i, s_j) \mapsto k(s_i, s_j)$ is a kernel, and $\mathbf{k}_i = (k(s_i, s_1), \cdots, k(s_i, s_N))'$. The kernel is a function of region locations, and $s_i = (x_i, y_i)$ denotes the location of region R_i defined by the (x, y) coordinates (latitude, longitude) of R_i's centroid. Using centroids to represent regions may cause information loss, but usually they are good representations of region locations especially when the regions are relatively small and homogeneous. Intuitively, the parameter vector $\boldsymbol{\eta}$ can be regarded as a vector of importance indices of surrounding regions of R_i. As a consequence, the spatial effect δ_i is a combination of the effect of R_i and the effects of R_j's ($j \neq i$).

Kernel $(s_i, s_j) \mapsto k(s_i, s_j)$ determines the space of functions used to approximate a function of interest. For instance, $(s_i, s_j) \mapsto k(s_i, s_j) = (s_i s_j + h)^q = (x_i x_j + y_i y_j + h)^q$ generates a space of functions spanned by all possible q-th order monomials of region locations with h being a smoothing parameter. In our application, a Gaussian radial basis function kernel will be applied. As discussed in Liu et al. (2007), where reproducing kernels are utilized to estimate gene-gene interactions, a Gaussian kernel is preferred compared to first and second order polynomial kernels if the underlying function is complex. Since our study is in a different framework compared to Liu et al. (2007), simulations are performed to assist the choice of a Gaussian kernel, where we also consider other radial basis function kernels besides the first and second order polynomial kernels.

A Gaussian kernel for regions R_i and R_j is defined as

$$(s_i, s_j) \mapsto k(s_i, s_j) = \exp\left\{\frac{-[(x_i - x_j)^2 + (y_i - y_j)^2]}{h}\right\}, \tag{3}$$

where $(s_i, s_j) \mapsto k(s_i, s_j) = 1$ if $i = j$ and $(s_i, s_j) \mapsto k(s_i, s_j)$ approaches zero if two regions are far apart. In the Gaussian kernel, equal weights are given to the latitude and longitude as done in other types of spatial modeling such as in the CAR models and the powered exponential family models (Cressie and Chan, 1989, Richardson, 1992). Clearly $k(\cdot, \cdot)$ takes its values in $[0, 1]$. Applying the kernel to all regions, a kernel matrix $\mathbf{K}$ is then defined with row vectors of $\mathbf{k}_i'$. Built on Euclidean distances between every two locations, the kernel matrix $\mathbf{K}$ is positive definite and essentially measures the correlations among different locations. The parameter h in the Gaussian kernel is an unknown smoothing parameter critical in function approximations. It controls the rate of decay of correlations. Given the locations of regions, the smaller the value of h, the faster the correlations between regions decrease. In the area of machine learning, the parameter h is usually pre-specified based on some ad-hoc methods. Brewer (2000) discusses the estimation of h through frequentist approaches. In this article, a Bayesian method will be used to infer h. Combining (2) and (3), the spatial effect of region R_i on the outcome is a weighted sum of correlations between region R_i and R_j.

In addition to the RKB method, other approaches are also available to approximate unknown functions, for instance, the spline-based methods. The geoadditive models proposed by Kammann and Wand (2003) was built upon P-splines containing a certain number of knots that need to be pre-specified (Eilers and Marx, 1996). P-splines were also implemented in the additive regression model proposed by Fahrmeir, Kneib, and Lang (2004). Ruppert, Wang, and Carroll (2003) has an in depth discussion on splines. In terms of function approximation, the RKB method and spline-based methods have a similar theoretical foundation, but the logic of their model-fitting is different. Spline-based methods start with the smoothness conditions of an unknown function in order to achieve quick convergence (Wahba, 1985), and a corresponding kernel function can usually be derived from these conditions. Instead, the RKB method starts from a kernel function that potentially determines the smoothness property of the unknown function. When comparing these two approaches, the RKB method gives a direct start of function approximation and substantially simplify the approximation process (Liu et al., 2007).

2.3 Bayesian inferences

Following the probability mass function (1), under $\boldsymbol{\theta} = \{\alpha, \boldsymbol{\beta}, \boldsymbol{\eta}, h\}$, the joint density of $\mathbf{Y}$ satisfies,

$$\begin{aligned} P(\mathbf{Y} = \mathbf{y}|\boldsymbol{\theta}) &= \prod_i P(Y_i = y_i|\boldsymbol{\theta}) \\ &= \prod_i \frac{\exp\{-n_i \exp(\alpha + \boldsymbol{\beta}'\mathbf{X}_i + \delta_i)\}(n_i \exp(\alpha + \boldsymbol{\beta}'\mathbf{X}_i + \delta_i))^{y_i}}{y_i!}, \end{aligned} \quad (4)$$

with δ_i defined in (2). We apply a fully Bayesian approach to infer the parameters.

2.3.1 Prior and hyper-prior distributions

To be fully Bayesian, we start from specifying the prior distributions of the parameters $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\eta}, h)$. Independence among these parameters is assumed.

Prior distributions of α and $\boldsymbol{\beta}$: Vague prior distributions are assigned to these coefficient parameters. Specifically we choose normal distributions with mean zero and large variances.

Prior distribution of $\boldsymbol{\eta}$: The prior distribution of $\boldsymbol{\eta}$ is assumed to be $N(\mathbf{0}, \tau\mathbf{K}^{-1})$, where $\mathbf{K}$ is the kernel matrix. This prior distribution is selected based on a connection between linear mixed models and semi-parametric models with reproducing kernels included. Liu et al. (2008) and Gonzalez-Recio, Gianola, Long, Weigel, Rosa, and Avendano (2008) discuss in detail this choice. This prior distribution assumes that the variations of $\boldsymbol{\eta}$ are related to the distances between regions. The farther apart two regions are, the more uncertainty over one region's influence on the other. Hyper parameter τ is a regularization parameter. Under $\mathbf{K}^{-1}$, small values of τ indicate small contributions of spatial dependence. In the extreme case, the spatial effect disappears when τ approaches to zero. The prior distribution of $\boldsymbol{\eta}$ is chosen for computational convenience. Other less-informative prior distributions can be constructed too – for instance, independent normal distributions a priori assuming unstructured spatial dependence between regions. This is discussed in detail in the simulation studies in Section 3.

Prior distribution of h and hyper prior distribution of τ in $\boldsymbol{\eta}$: The prior distributions of h and τ are selected as inverse gamma with both shape and scale parameters being 0.5. According to Gustafson (2003) and Kass and Wasserman (1995),

this choice of prior distribution can be viewed as giving a "unit-information" prior such that our prior guess for each parameter is one and the prior gets little weight compared to the data. Note that with this choice of prior distribution for h, the Gaussian kernel is scale dependent and h itself is consequently scale variant. However, the functionality of h on the control of smoothness remains the same. The estimates of h will be comparable between different data sets as long as the same unit in distance measures is used.

2.3.2 The posterior distribution:

With all prior and hyper-prior distributions specified, the joint posterior distribution of the expanded $\boldsymbol{\theta}$, $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\eta}, h, \tau)$, satisfies (up to a normalizing constant),

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto P(\mathbf{Y} = \mathbf{y}|\boldsymbol{\theta}) p(\alpha) p(\boldsymbol{\beta}) p(\boldsymbol{\eta}|h, \tau) p(h) p(\tau), \tag{5}$$

where

$$\begin{aligned}
P(\mathbf{Y} = \mathbf{y}|\theta) &= \prod_i \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}, \\
\log(\lambda_i|\boldsymbol{\theta}) &= \log(n_i) + \alpha + \boldsymbol{\beta}'\mathbf{X}_i + \delta_i, \\
\delta_i &= \mathbf{k}_i'\boldsymbol{\eta} = \sum_{j=1}^{N} \eta_j k(s_i, s_j), \\
\boldsymbol{\eta}|h, \tau &\sim N(\mathbf{0}, \tau\mathbf{K}^{-1}), \\
\alpha &\sim N(0, \sigma_\alpha^2),\ \boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_{\boldsymbol{\beta}}^2\mathbf{I}), \\
h &\sim \text{Inv-Gam}(0.5, 0.5),\ \tau \sim \text{Inv-Gam}(0.5, 0.5).
\end{aligned} \tag{6}$$

In (5), we use $p(\cdot)$ to denote a probability density function of a continuous random variable, and $\mathbf{I}$ in (6) denotes an identity matrix. Spatial dependence δ_i is evaluated with updated information on $\boldsymbol{\eta}$ obtained from its posterior inferences. Note that we will conclude the same model (6) by assigning the following prior distribution to $\boldsymbol{\delta}$,

$$\boldsymbol{\delta}|h, \tau \sim N(\mathbf{0}, \tau\mathbf{K}), \tag{7}$$

which is linked to a Gaussian random field. A Gaussian random field in general induces a multivariate normal distribution for spatial dependence (Siegmund and Worsley, 1995, Emmerich, Giannakoglou, and Naujoks, 2006). Hyper-parameters σ_α^2 and $\sigma_{\boldsymbol{\beta}}^2$ are selected to be large in order to form diffused normal distributions. Since all the prior distributions are proper, the joint posterior distribution is proper as well.

2.3.3 Computation of the posterior distribution:

The posterior distribution of parameter vector $\boldsymbol{\theta}$ given by (5) is difficult to study analytically. Instead we use a Markov chain Monte Carlo (MCMC) simulation method, specifically, a Gibbs sampler coupled with Metropolis-Hastings (M-H) steps to generate samples from the joint posterior distribution. Gibbs sampler utilizes full conditional posterior distributions of each parameter. The full conditional posterior distribution of τ is inverse gamma,

$$\tau|\boldsymbol{\eta},h \sim \text{Inv-Gam}((N+1)/2,(\boldsymbol{\eta}^T\mathbf{K}\boldsymbol{\eta}+1)/2).$$

Given (7), it is straightforward to show

$$\tau|\boldsymbol{\delta},h \sim \text{Inv-Gam}((N+1)/2,(\boldsymbol{\delta}^T\mathbf{K}^{-1}\boldsymbol{\delta}^T+1)/2). \tag{8}$$

Thus, we can sample τ without knowing $\boldsymbol{\eta}$. Equations (7) and (8) are derived based on the prior distribution $\boldsymbol{\eta}|h,\tau \sim N(\mathbf{0},\tau\mathbf{K}^{-1})$. However, if we assign a prior distribution to $\boldsymbol{\eta}$ that is not normal, then the prior distribution of $\boldsymbol{\delta}$ will be different from those in a classical Gaussian random field, and the conditional posterior distribution of τ will be different from (8).

The full conditional posterior distributions of the remaining parameters can be easily derived from (5) and (6). They are not standard and will be sampled using the M-H algorithm, in which proposal distribution functions are needed to generate proposal samples for each parameter. The proposal distributions for $\alpha,\boldsymbol{\beta},\boldsymbol{\eta}$ and $\log(h)$ are normal distributions centered at the previous posterior draw of the parameters. The variance components in the proposal distributions are selected to achieve efficient convergence. Specifically, they are adaptively tuned by controlling acceptance rates during the sampling process (Gelman, Carlin, Stern, and Rubin, 2003, Browne and Draper, 2006). Let v denote the variance and assume the initial value of v is $v=v_0$. Usually v_0 takes a relatively small value in order to initiate the MCMC simulation process with a relatively large acceptance rate. If after T iterations, the acceptance rate r_a is close to the target rate r_t, that is, $r_a \in (r_t-r_0,r_t+r_0)$ with r_0 being a positive small number pre-determined, then we set $v=v_0$; otherwise, update v_0 as

$$v_m \quad = \quad \begin{cases} v_0[2-(1-r_a)/(1-r_t)], & \text{if } r_a > r_t+r_0 \\ v_0/(2-r_a/r_t), & \text{if } r_a < r_t-r_0, \end{cases} \tag{9}$$

for $m=1,\cdots,$ until $r_a \in (r_t-r_0,r_t+r_0)$.

Once the proposal variance components are selected, we continue to run a certain number of iterations for burn-in and then draw posterior samples to infer the parameters. The convergence of the sampled sequences is evaluated using the

method developed by Gelman and Rubin (1992a,b) and discussed further in Gelman et al. (2003). We programmed the sampling process in C++; for users unfamiliar with C++, we also coded it in R2WinBUGS. Both codes are available upon request.

3 Simulations

In this section, through simulated data generated from various simulation scenarios, we demonstrate and evaluate the RKB method discussed in the previous sections. We compare the method with three commonly used methods for spatial dependence modeling: the independent Gaussian (IG) model and two CAR models including the intrinsic CAR (ICAR) model (Besag et al., 1991) and a distance-based CAR (DCAR) model (Cressie and Chan, 1989, Stern and Cressie, 1999). A brief description of these competing models is included in the Appendix.

3.1 Simulated data

In each of 800 Monte Carlo (MC) replicates, 30 regions are considered and are arranged in a spatial structure of a rectangle formed by 5 columns and 6 rows in a map view. The coordinate of each region is determined by the row and column indices; the coordinate of a region in row r and column c is (r,c).

We assume the expected effect of region i on the number of incidences (in log-scale) is $\log(\lambda_i|\alpha,\beta,\delta_i) = \log(n_i) - 5 + x_i + \delta_i, i = 1,\cdots,30$, which gives $\alpha = -5, \beta = 1$. The covariate x_i is generated from $UNIF(0.01, 0.8)$. The offset n_i is generated from $N(5000, 100^2)$, which implies the average offset in log-scale is around 8.51. The parameters of interest are α and β along with the component parameters in the reproducing kernel, h and τ. To simulate spatial effects δ_i, we consider the following two data-generating scenarios:

SC1. The 30 regions are simulated assuming an unstructured spatial dependence such that $\delta_i \overset{iid}{\sim} N(0, \sigma_\delta^2)$ with $\sigma_\delta = 0.5$.

SC2. Data are simulated assuming the existence of two clusters of spatially dependent regions; cluster 1 is formed by regions 1 to 3 and 6 to 8, and cluster 2 is formed by regions 19, 20, 24, 25, 29, and 30. The correlation is set at $\rho = 1$ among the clustered regions, indicating a strong spatial dependence in a cluster.

In the second scenario, we expect a stronger spatial effect. On the other hand, since the dependence suddenly disappears outside the two small clusters, the value of h can be small.

3.2 Empirical evidence

The posterior inferences for each data set are based on 20,000 MCMC iterations after 40,000 burn-in iterations. To assess the performance of the methods in estimating Poisson regression coefficients, we record the average bias, coverage rate of true β, and average length of credible intervals (Table 1). Summaries of Moran's I, the deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, and van der Linde, 2002), and mean squared prediction error (MSPE) (Lawson, 2008) are used to assess the model fit and prediction quality, respectively (Table 2). Moran's I evaluates global spatial autocorrelation of the residuals and is estimated as $\frac{\mathbf{e}^T\mathbf{A}\mathbf{e}}{\mathbf{e}^T\mathbf{e}}$, with $\mathbf{e}$ being the vector of standardized residuals and $\mathbf{A}$ being an adjacency matrix. Moran's I being zero indicates no global autocorrelation. DIC is a Bayesian measure of model fit which penalizes complex models while MSPE measures the quality of prediction of the fitted model; small values of DIC and MSPE are preferred.

Overall, regardless of the pattern of spatial dependence, the RKB method performs slightly better than the other three methods in terms of accuracy and precision, although it does not do the best on every individual criterion. Some noteworthy observations from the simulation results are highlighted below and may deserve a further systematic study:

- E1 With respect to prediction (MSPE) and model fitting (DIC), the RKB method outperforms other methods in both examples (unstructured and correlated spatial dependence) (Table 2).
- E2 In terms of accuracy of parameter β estimates, when we examine average bias and coverage rates together, the RKB method overall outperforms the other three methods (Table 1). However, the RKB method tends to produce wider credible intervals.
- E3 Largest bias was observed in results from the IG model. Its model fitting and prediction quality is better or comparable to DCAR.
- E4 ICAR provides the overall worst performance with respect to model fitting, prediction, and accuracy of parameter estimates. The relatively bigger bias and small coverage rate might be due to the impropriety of the model, which deserves further investigation.

A close inspection at the measures on all the criteria between different methods reveals that among the competing methods, DCAR performs most closely to RKB. However, as noted in Appendix, there is a parameter γ in DCAR that evaluates the strength of spatial dependence. The lower and upper bounds of γ in DCAR are the same across all levels of spatial dependence. These constraints potentially

limit DCAR to properly reflect the strength of spatial dependence. For the simulated spatial structure, the upper bound of γ is $\gamma_{max} = 0.065$. In all the MC replicates, the upper limits of the 95% credible intervals are close to γ_{max} (results not shown). Table 3 lists the inferences of h, τ, and γ. The estimate of h in the second scenario is smaller, which is likely due to the sharp diminishment of spatial dependence outside the two clusters. The larger estimate of τ in the second scenario indicates stronger spatial effects.

To examine the consistency of the pattern described above, two additional data-generating scenarios were considered with each representing a specific level of spatial dependence. One scenario assumes $\boldsymbol{\delta} \sim N(0,\Sigma)$ with $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_N)$ and

$$\Sigma = \begin{pmatrix} 0.5^2 & \rho 0.5^2 & \cdots & \rho/[(1-r)^2+(1-c)^2]0.5^2 & \cdots & \frac{\rho}{41}0.5^2 \\ \rho 0.5^2 & 0.5^2 & \cdots & \rho/[(1-r)^2+(2-c)^2]0.5^2 & \cdots & \frac{\rho}{34}0.5^2 \\ \cdots & & \cdots & \cdots & \cdots & \cdots \\ \frac{\rho}{41}0.5^2 & \frac{\rho}{34}0.5^2 & \cdots & \rho/[(6-r)^2+(5-c)^2]0.5^2 & \cdots & 0.5^2 \end{pmatrix}_{30\times 30,} \quad (10)$$

where $\rho = 0.48$. This definition of Σ yields a moderate spatial dependence among three contiguous regions and the dependence then decreases with the increase of Euclidean distances between regions. The other scenario considers one cluster of spatially dependent regions composed of 6 contiguous regions with spatial correlation being 1 in the cluster. The findings in general were consistent with those given in Tables 1 to 3 (results not shown).

3.3 Further simulation studies

Further simulation studies are carried out with focus on sensitivity analysis: the impact of a different prior distribution for $\boldsymbol{\eta}$, the choice of reproducing kernel, and the number of regions.

3.3.1 Prior distribution of $\boldsymbol{\eta}$

As noted earlier, we assumed $\boldsymbol{\eta} \sim N(\mathbf{0}, \tau\mathbf{K}^{-1})$. To examine if the RKB method is limited to this prior, we simplify the covariance matrix of $\boldsymbol{\eta}$ and assume $\boldsymbol{\eta} \sim N(\mathbf{0}, \tau\mathbf{I})$. This distribution does not enforce any prior structured spatial dependence on the variation of η_i. We simulated 100 MC replicates following the second scenario: the population has two clusters of regions with strong spatial dependence.

Table 1: Average bias, coverage rate, and average length of credible intervals (CI) of β across 800 MC replicates. Empirical 95% intervals are in parentheses.

	Average Bias	Coverage Rate (%)	Average CI Length
		Unstructured $\boldsymbol{\delta}$	
IG	-0.0067 (-0.89, 0.91)	92.63	1.71 (1.22, 2.37)
ICAR	-0.0069 (-0.95, 0.94)	91.75	1.70 (1.16, 2.49)
DCAR	-0.0064 (-1.00, 0.73)	92.74	1.69 (1.17, 2.33)
RKB	-0.0062 (-0.65, 1.14)	93.55	1.86 (1.40, 2.54)
		Correlated $\boldsymbol{\delta}$	
IG	0.0042 (-0.80, 0.83)	95.42	1.75 (1.43, 2.24)
ICAR	0.016 (-0.58, 0.64)	96.78	1.35 (1.09, 1.73)
DCAR	0.0016 (-0.73, 0.76)	96.16	1.66 (1.37, 2.12)
RKB	0.0019 (-1.00, 0.78)	95.97	1.84 (1.50, 2.34)

Table 2: Averaged MSPE, DIC, and Moran's I across 800 MC replicates. Empirical 95% intervals are in parentheses. All 95% credible intervals of Moran's I cover zero indicating overall spatial independence.

	MSPE	DIC	Moran's I
		Unstructured $\boldsymbol{\delta}$	
IG	113.00 (95.09, 133.74)	230.49 (223.98, 236.61)	0.010
ICAR	116.02 (97.52, 136.66)	231.90 (225.52, 237.48)	-0.045
DCAR	113.18 (95.66, 134.04)	230.34 (223.95, 236.07)	-0.058
RKB	112.38 (95.01, 133.28)	228.33 (222.54, 234.18)	-0.056
		Correlated $\boldsymbol{\delta}$	
IG	170.09 (153.37, 185.28)	243.03 (240.09, 245.86)	0.018
ICAR	173.19 (155.90, 189.25)	243.04 (239.84, 245.99)	-0.045
DCAR	170.03 (153.27, 185.03)	242.85 (239.86, 245.67)	-0.034
RKB	169.68 (153.12, 184.75)	242.43 (239.49, 245.23)	-0.039

The results from the new prior are comparable to those based on the prior discussed earlier, $\boldsymbol{\eta} \sim N(\mathbf{0}, \tau\mathbf{K}^{-1})$. In terms of the estimate of coefficient β, the

Table 3: Statistics summary of parameters measuring spatial dependence based on 800 MC replicates. Empirical 95% intervals are in parentheses.

	Unstructured $\boldsymbol{\delta}$	Correlated $\boldsymbol{\delta}$
γ	0.031 (0.025, 0.043)	0.042 (0.038, 0.047)
τ	0.24 (0.14, 0.42)	0.28 (0.25, 0.33)
h	0.087 (0.067, 0.14)	0.062 (0.059, 0.065)

average bias is -0.001 with 95% empirical interval (-0.84, 0.81), coverage rate of true β is 95.97%, and the average length of 95% credible intervals is 1.85 with 95% credible interval (1.50, 2.34). Turning our attention to model assessment, the value of MSPE is 169.56 with 95% empirical interval (154.24, 187.65) and DIC is 240.82 with (238.20, 243.87) being a 95% empirical interval. The averaged Moran's I is -0.039 and all 95% credible intervals cover zero. The results presented here and in the previous section demonstrate that the posterior inference from the RKB method is not sensitive to these two prior distributions. However, as typical in Bayesian inference, the posterior inference can be dominated by strong prior beliefs.

3.3.2 Kernel selection

To assess the performance of the selected Gaussian kernel, we consider several other reproducing kernels and compare their performances. Besides the first order ($(s_i,s_j) \mapsto k(s_i,s_j) = s_i s_j$) and second order ($(s_i,s_j) \mapsto k(s_i,s_j) = (s_i s_j + h)^2$) polynomial kernels discussed earlier, we also considered an analysis of variance (ANOVA) kernel and an Laplacian kernel. The ANOVA kernel is a radial basis function kernel, $(s_i,s_j) \mapsto k_d(s_i,s_j) = \left\{\exp(-\frac{(x_i-x_j)^2}{h}) + \exp(-\frac{(y_i-y_j)^2}{h})\right\}^d$ (Karatzoglou, Smola, and Hornik, 2012). The parameter d is the order of the kernel and represents the degree of interactions (Shawe-Taylor and Cristianini, 2004). This kernel has been used in multivariate regressions to model the dependencies between response variables in addition to the dependence on covariates (Stitson, Gammerman, Vapnik, Vovk, Watkins, and Weston, 1999, Hofmann, Scholköpf, and Smola, 2008). Laplacian kernel and exponential kernel belong to the family of radial basis function kernels as well. They are closely related to the Gaussian kernel (with squared norm in the Gaussian kernel left out). These two kernels are almost equivalent to each other, and thus only the Laplacian kernel is implemented. We expect this kernel will perform similarly to that of the Gaussian kernel. Comparing the

selection of a reproducing kernel is equivalent to comparing models. Thus DIC is used for this purpose. We record DIC values from each of the above 100 MC replicates for each reproducing kernel and plot them in Figure 1.

Because of a large DIC value from the second polynomial kernel, the differences between the Gaussian, Laplacian, and ANOVA kernels are hard to see from the figure. We here present the summary statistics of DIC. The sample mean of DIC values when using the Gaussian kernel is 226.55 with a 95% empirical interval of $(159.43, 244.38)$, and for the Laplacian kernel the mean is 232.00 and a 95% empirical interval is $(225.25, 243.54)$. These two kernels give similar estimates. The RKB method based on the ANOVA kernel gives the mean of DIC 242.89 and a 95% empirical interval is $(238.84, 246.48)$. The result is slightly inferior to those from the Gaussian and Laplacian kernels. The ANOVA kernel has been shown to work well in multi-dimensional support vector regression models (Stitson et al., 1999, Hofmann et al., 2008). However, in our application, this kernel evaluates spatial dependence based on a separate assessment on longitudinal and latitudinal distances, which can possibly bring in noise and reduce the quality of model fitting. For the first order polynomial kernel the mean is 814.83 and a 95% empirical interval is $(679.42, 926.55)$, and for the second order polynomial kernel the mean is 1787.91 with a 95% empirical interval of $(950.19, 3585.78)$. The poor performances of the first and second order polynomials are not surprising. To model spatial dependence, we prefer kernels with the ability of detecting stronger dependence for two geographically closer regions and weaker dependence if they are far apart. However, the first and second order polynomial kernels do not satisfy this property, which is the possible cause for large DIC values. Overall, the DIC values based on the Gaussian and those on Laplacian kernels agree with each other and are the smallest compared to the DICs from the others.

Through simulations, we also examined the effect of the number of regions to see if the proposed method is in favor of small numbers of study regions. We expand the number of simulated regions from 30 to 64. The 64 regions are arranged in 8 rows and 8 columns. Other than the number of regions, the remaining settings are the same as in the scenario with 30 regions. Overall, the RKB method can reasonably handle a large number of regions and its performances are at least comparable to those of the CAR models. The performance of CAR models, on the other hand, seems to vary depending on the choice of covariance matrix Σ and the feature of spatial dependence (results not shown).

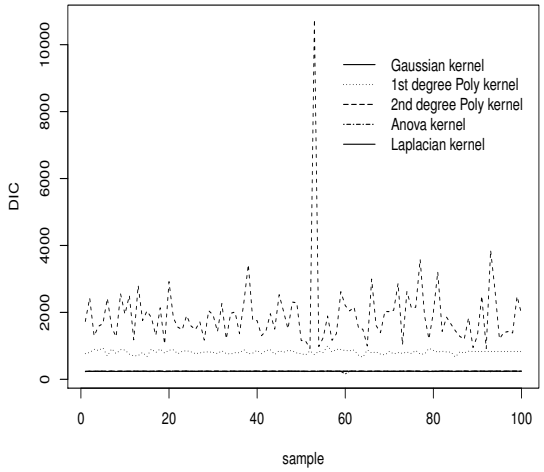


Figure 1: DICs from 100 samples with different kernels applied in the RKB method.

4 Applications

We apply the RKB method to analyze two data sets. One is the classic Scottish lip cancer data examined in previous studies (Clayton and Kaldor, 1987, Breslow and Clayton, 1993) and further discussed in Waller and Gotway (2004). The other data set includes incidence rates and measures of possible risk factors for smoke related cancers in the respiratory system, oral cavity, and pharynx (hereafter, ROP cancers) in Florida.

The Scottish lip cancer: Data include the number of lip cancer cases, expected number of cases, and percent of population engaged in agriculture, fishery, or forestry (AFF) in 56 counties in Scotland. In the model of DCAR, we combine the ideas given by Cressie and Chan (1989) and Besag et al. (1991) to define neighboring regions such that two regions are treated as neighbors if they are adjacent or if their centroid distance is within a pre-specified value. For other parameter settings in DCAR, we follow the parameterization discussed in the simulation studies. Table 4 and 5 list the parameter estimates together with DIC, MSPE, and Moran's I. They are based on 20,000 iterations after 20,000 burn-in iterations. As before, α is the overall intercept, and β is the slope measuring the effect of AFF on lip cancer incidence rates.

As indicated in Table 4, although the estimates of β from different methods consistently imply a significant effect of AFF on lip cancer (indicated by credible

intervals), the estimate of β from the ICAR model tends to be lower than that from other methods. As seen in Table 5, the estimate of h is very small indicating the possibility of unstructured spatial effects. Thus the inconsistent estimate of β is likely due to the relatively larger bias and lower coverage rate of ICAR for unstructured spatial dependence as seen in simulations (Table 1). In addition, comparing DICs and MSPEs, the RKB method gives the smallest DIC and MSPE while the largest DIC and MSPE are observed in both CAR models. This result is also consistent with that listed in Table 2. In summary, the RKB method performs slightly better than the other methods in model fitting and prediction, especially better than ICAR and DCAR.

Table 4: Posterior estimates of the parameters from the Scottish lip cancer data. Included in parentheses are the 95% credible intervals.

	Posterior estimates				
	α	β	DIC	MSPE	Moran's I
RKB	-0.50	0.069	309.32	19.13	0.041
	(-0.83, -0.18)	(0.040, 0.097)			(-0.16, 0.17)
IG	-0.49	0.068	310.33	19.29	0.046
	(-0.82, -0.16)	(0.038, 0.097)			(-0.15, 0.18)
ICAR	-0.29	0.044	312.57	20.91	-0.036
	(-0.53, -0.056)	(0.018, 0.068)			(-0.24, 0.14)
DCAR	-0.54	0.065	328.37	20.73	0.0087
	(-0.87, -0.20)	(0.038, 0.093)			(-0.17, 0.14)

Florida ROP cancer: The Florida ROP cancer data include age adjusted ROP cancer incidence rates in 2006, information of behavioral risk factors collected in 2002, and gender information for each of the 67 counties in Florida. The data are from the Florida Cancer Data System and the Behavioral Risk Factor Surveillance System (BRFSS). Due to the latency period of cancer, information of behavioral risk factors collected in 2002 were used (which were the only data prior to 2006 available to the public upon data collection). The following three behavioral risk factors are considered and measured as random-sample-based sample proportions for each county: no regular moderate physical activity, current smokers, and heavy or binge drinkers. Gender is included as a risk factor due to the significant differences of mortality rates between gender in Florida (CDC, 2007).

Table 5: Statistics summary of parameters unique to each method (Scottish lip cancer). Included in parentheses are the 95% credible intervals.

	h	τ	V	γ
RKB	0.0032	0.37	—	—
	$(8.18\times10^{-6}, 0.012)$	(0.23, 0.68)	—	—
ICAR	—	—	0.52	—
	—	—	(0.26, 0.95)	—
DCAR	—	—	1.60	0.15
	—	—	(1.07, 3.68)	(0.028, 0.18)

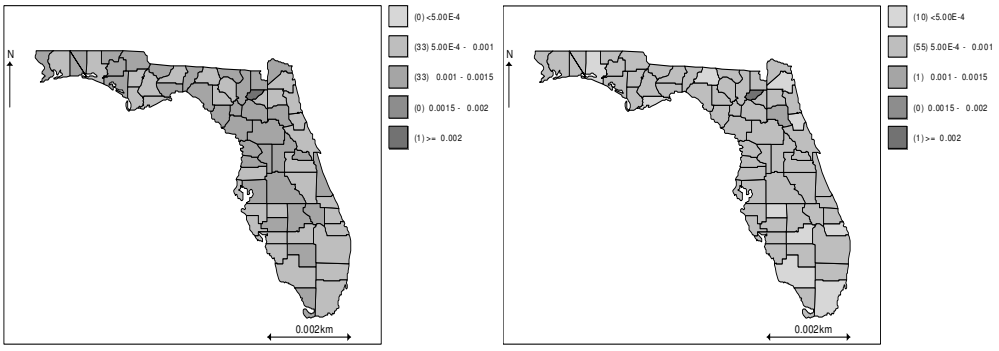
To set up the Poisson regression model, we consider gender specific risk factor effects and an overall gender effect. Thus, $\log(\lambda_i|\boldsymbol{\theta})$ in model (1) becomes

$$\log(\lambda_{ig}|\boldsymbol{\theta}) = \log(n_{ig}) + \alpha_0 + \alpha_g + \boldsymbol{\beta}_g'\mathbf{X}_{ig} + \delta_i, \tag{11}$$

where $g = 1,2$ indexes gender ($g = 1$ for female) and n_{ig} is the population size of county i gender g. Additionally, the constraint $\alpha_1 + \alpha_2 = 0$ is applied.

Table 6 lists the posterior estimates of the parameters. Due to the relatively large number of parameters, the estimates are based on 100,000 iterations after 200,000 burn-in iterations. Posterior means and 95% credible intervals of each parameter are given in the table. Positive estimates of the parameters imply that the incidence rate in a county increases with the increased level of risk factors. As expected, smoking is associated with increased incidence, which supports findings from other studies (Blot and Fraumeni, 1982, CDC, 1998). The effects of the other two factors are less susceptible, which might be due to ecological influences (Blakely and Woodward, 2000). This is not our focus but an interesting direction to explore. The relative risks of males and females across the state are given in Figure 2, which indicates that statewide males have higher risk of cancer. This is consistent with the pattern described elsewhere (CDC, 2007). Additionally, the posterior mean of h is 0.036 with a 95% credible interval of $(0.0013, 0.064)$ and for τ the corresponding inferences are 0.066 $(0.041, 0.10)$. Compared to the estimates of h and τ from the Scottish lip cancer data, the estimate of h in this example is higher suggesting a stronger spatial dependence. The smaller value of τ, on the other hand, implies a weaker effect of spatial dependence on cancer incidence.

We also inferred parameters using the other three methods, IG, ICAR, and DCAR (results not shown). Overall the posterior estimates of parameters agree with those from the RKB method in terms of the direction of factor effects, except that



a) b)

Figure 2: Estimated relative risks of ROP cancer incidence for (a) males and (b) females using the RKB method (the darker the color, the higher the relative risk).

the DCAR model did not detect significant physical activity effect. This might be due to the limitation of $\gamma_{\max}$. The posterior mean of γ in DCAR is 0.11 with a 95% credible interval of $(0.043, 0.14)$. The upper bound is about its maximum value, $\gamma_{\max} = 0.14$, which implies a possible strong spatial dependence, but the estimate of its strength is limited by $\gamma_{\max}$, as seen in our simulations. For model fit and prediction errors, we summarized DIC and MSPE values together with Moran's I in Table 7. The ICAR model gives the largest DIC and MSPE. The DCAR model and the RKB method are comparable in that DCAR gives the smallest MSPE while the RKB method produces the smallest DIC.

Table 6: Posterior estimates of the parameters from the ROP cancer data using the RKB method. Included in parentheses are the 95% credible intervals.

Gender (α_1) effect	Factor ($\boldsymbol{\beta}$) effects	Physical activity	Smoke	Drinking
0.17	Male	-0.013	0.018	-0.021
(-0.19, 0.49)		(-0.017, -0.0088)	(0.013, 0.023)	(-0.028, -0.013)
	Female	-0.0080	0.011	-0.0051
		(-0.013, -0.0039)	(0.0072, 0.016)	(-0.010, -0.000067)

Table 7: Evaluation of model fit and prediction for the ROP cancer data. Included in parentheses are the 95% credible intervals.

	Moran's I^F	Moran's I^M	DIC	MSPE
RKB	0.038	0.039	1425.80	2608.00
	(-0.020, 0.085)	(-0.049, 0.10)	—	—
IG	0.0038	0.014	1428.46	2628.00
	(-0.064, 0.058)	(-0.079, 0.081)	—	—
ICAR	0.016	-0.012	1445.51	2707.00
	(-0.060, 0.073)	(-0.11, 0.077)	—	—
DCAR	0.013	-0.0015	1430.43	2563.00
	(-0.052, 0.065)	(-0.091, 0.088)	—	—

5 Summary

A semi-parametric model built upon a reproducing kernel to describe spatial dependence is proposed to study risk factor effects through Poisson regressions. Simulation studies are performed to demonstrate and evaluate the method, and to compare the method with several other methods including an independent Gaussian (IG) model for unstructured spatial dependence, a CAR model based on the number of contiguous neighbors (ICAR), and a CAR model based on distances between regions (DCAR). Results from two real data applications, Scottish lip cancer data and Florida ROP cancer data, agree with previous findings.

Based on our simulations, the RKB method performs better than the other three methods do on model fitting and prediction. In terms of bias and coverage rates, the RKB method overall does better as well, although this method has the tendency to give wider credible intervals. These findings indicate that the proposed method has the potential to serve as a reasonable alternative to the commonly used CAR models. Compared to the other three methods, the RKB method has three promising features that make it attractive. First, it is flexible and has the ability to appropriately account for different formulations of spatial dependence. Second, unlike the CAR models, the RKB method is easy to implement and program. The kernel matrix $\mathbf{K}$ in the prior distributions in general is positive definite and we only require parameters h and τ be positive. Thus the RKB method does not have the

theoretical (e.g., the identifiability issue of spatial effects induced by impropriety of the prior in ICAR) and computational (e.g., the limit on the range of precision parameters to ensure identifiable spatial effects using ICAR or DCAR) problems inherent in CAR models as noted in Banerjee et al. (2004). Lastly, among the three methods, DCAR performed most closely to the RKB method, which is not surprising because both methods are distance-based and both include parameters evaluating spatial dependence. However, the upper and lower bounds of parameter γ in the DCAR model are controlled by the spatial structure but not the level of spatial dependence. This can limit its potential to properly infer the strength of spatial dependence. Such limitation does not exist in the RKB method. In terms of computing performance, the speed of the RKB method written in C++ is similar to DCAR but a little slower than the speed of the other two methods. When written in R2WinBUGS, due to the limitation in flexibility of the package, the kernel has to be calculated in every iteration of MCMC and the RKB method is the least efficient. It will be beneficial to develop a distribution function for the RKB method in WinBUGS.

Other possible future work can be pursued in different directions. The RKB method is designed to utilize information from all surrounding regions. It may be interesting to incorporate variable selection techniques into this reproducing kernel framework to select regions. Additionally, although it is not the focus of this article, ecological effects on health outcomes have been studied intensively (Blakely and Woodward, 2000, Jones, Patel, Levy, Storeygard, Balk, Gittleman, and Daszak, 2008, Hu and Rao, 2009). To this end, joint modeling that links ecological and spatial models together may be an informative way for a thorough evaluation of the risk factors.

Appendix

5.1 Competing methods

An independent Gaussian model assumes $\delta_i \overset{iid}{\sim} N(0, \sigma_\delta^2)$. The formulation of δ_i in a CAR model under a Gaussian framework is given as (Besag, 1974),

$$\begin{aligned} \delta_i | \{\delta_k : k \in A_i\} &\sim N(E_i, V_i) \\ E_i &= \mu_i + \sum_{k \in A_i} \gamma w_{ik} (\delta_k - \mu_k) \\ V_i &= V \xi_i^2, \quad \xi_i^2 > 0 \\ w_{ik} \xi_k^2 &= w_{ki} \xi_i^2, \end{aligned} \tag{12}$$

where A_i defines a collection of neighbors (contiguous or non-contiguous) of region i, and V is the overall variance parameter. The last equation in (12) is a constraint to ensure a symmetric covariance matrix, denoted by Σ, in the joint distribution of δ_i's. Parameter γ measures the strength of spatial dependence ranging from $\gamma_{\min}$ to $\gamma_{\max}$. The lower and upper bounds of γ, respectively, are the inverse of the smallest and largest eigenvalues of the matrix $\text{diag}(\xi_i)^{-1/2} \mathbf{W} \text{diag}(\xi_i)^{1/2}$, where $\text{diag}(\xi_i)$ is a diagonal matrix composed of ξ_i's and $\mathbf{W}$ is an $N \times N$ matrix composed of w_{ik}'s. The pre-determined bounds of γ ensure a positive definite covariance matrix Σ.

In the article, we consider two CAR models, the intrinsic CAR model (Besag et al., 1991), ICAR, and a distance-based CAR model (Cressie and Chan, 1989, Stern and Cressie, 1999), DCAR. The ICAR model assumes $\mu_i = 0, \gamma = 1$, and $w_{ik} = 1/b_i$ if regions i and k are adjacent and $w_{ik} = 0$ otherwise. The quantity b_i is the number of neighboring regions of region i. The DCAR model defines neighbors based on distances. If two regions are within a pre-specified distance, then they are treated as "neighbors". In our simulations, we set the cut-off distance large enough so that the effects of all surrounding regions are counted. Following Stern and Cressie (1999), w_{ik} and ξ_i are defined as,

$$\begin{aligned} w_{ik} &= \begin{cases} (E_k/E_i)^{0.5}, & \text{if } k \in A_i \\ 0, & \text{elsewhere,} \end{cases} \\ \xi_i &= E_i^{-1}, \end{aligned}$$

where the value of E_i is taken as the population size in region i. Clearly, $\gamma_{\min}$ and $\gamma_{\max}$ noted above are constant across all levels of spatial dependence. This potentially imposes a limitation on the ability of γ to evaluate the strength of spatial dependence.

Prior distributions for the parameters, including σ_δ^2 in the independent Gaussian model and the overall variance parameter V in the CAR models, are selected to be inverse gamma distributions with shape parameter 0.05 and scale parameter 0.0005, as suggested by Kelsall and Wakefield (1999). Prior distribution of γ in the distance-based CAR is chosen to be non-informative and assumed to be a uniform distribution ranged from $\gamma_{\min}$ to $\gamma_{\max}$.

References

Banerjee, S., B. Carlin, and A. Gelfand (2004): *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC.

Besag, J. (1974): "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, Series B: Methodological*, 36, 192–236.

Besag, J., A. Mollie, J. York, and A. Mollié (1991): "Bayesian image restoration, with two applications in spatial statistics (Disc: P21-59)," *Annals of the Institute of Statistical Mathematics*, 43, 1–20.

Best, N. G., R. A. Arnold, A. Thomas, L. A. Waller, and E. M. Conlon (1999): "Bayesian models for spatially correlated disease and exposure data," in J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. Smith, eds., *Bayesian Statistics 6 – Proceedings of the Sixth Valencia International Meeting*, Clarendon Press [Oxford University Press], 131–156.

Blakely, T. and A. Woodward (2000): "Ecological effects in multi-level studies," *Journal of Epidemiology and Community Health*, 54, 367–374.

Blot, W. J. and J. Fraumeni (1982): "Changing patterns of lung cancer in the United States," *American Journel of Epidemiology*, 115, 664–673.

Breslow, N. and D. Clayton (1993): "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, 88, 9–25.

Brewer, M. J. (2000): "A Bayesian model for local smoothing in kernel density estimation," *Statistics and Computing*, 10, 299–309.

Browne, W. J. and D. Draper (2006): "A comparison of Bayesian and likehood-based methods for fitting mutilevel models (pkg: P473-550)," *Bayesian Analysis*, 1, 473–514.

CDC (1998): "Preventing and controlling oral and pharyngeal cancer: recommendations from a national strategic planning conference," *Morb. Mortal. Wkly. Rep.*, 47.

CDC (2007): "U.S. Cancer statistics working group. United States cancer statistics: 1999-2004 incidence and mortality web-based report. Atlanta: U.S. Department of Health and Human Services; available at: www.cdc.gov/uscs." .

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

Cressie, N. (1985): "Fitting variogram models by weighted least squares," *Journal of the International Association for Mathematical Geology*, 17, 563–586.

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

Cristianini, N. and J. Shawe-Taylor (2000): *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press.

Ecker, M. D. and A. E. Gelfand (1997): "Bayesian variogram modeling for an isotropic spatial process," *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 347–369.

Eilers, P. H. C. and B. D. Marx (1996): "Flexible smoothing with B-splines and penalties," *Statistical Science*, 11, 89–121.

Emmerich, M., K. Giannakoglou, and B. Naujoks (2006): "Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels," *IEEE Transactions on Evolutionary Computation*, 10, 421–439.

Fahrmeir, L., T. Kneib, and S. Lang (2004): "Penalized structured additive regression for space-time data: A Bayesian perspective," *Statistica Sinica*, 14, 731–761.

Feldens, C., P. Kramer, S. Ferreira, M. Spiguel, and M. Marquezan (2010): "Exploring factors associated with traumatic dental injuries in preschool children: a Poisson regression analysis," *Dental Traumatology*, 26, 143–148.

Frome, E. and H. Checkoway (1985): "Use of Poisson regression models in estimating incidence rates and ratios," *American Journal of Epidemiology*, 121, 309–323.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003): *Bayesian Data Analysis*, Chapman & Hall/CRC.

Gelman, A. and B. D. Rubin (1992a): "Inference from iterative simulation using multiple sequences," *Statistical Science*, 7, 457–511.

Gelman, A. and D. B. Rubin (1992b): "A single series from the Gibbs sampler provides a false sense of security," in J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*, Clarendon Press [Oxford University Press], 625–631.

Gonzalez-Recio, O., D. Gianola, N. Long, K. Weigel, G. Rosa, and S. Avendano (2008): "Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers," *Genetics*, 178, 2305–2313.

Gorsich, D. and M. Genton (2000): "Variogram model selection via nonparametric derivative estimation," *Mathematical Geology*, 32, 249–270.

Gustafson, P. (2003): *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*, Chapman and Hall/CRC Press.

Hofmann, T., B. Scholköpf, and A. Smola (2008): "Kernel methods in machine learning," *The Annals of Statistics*, 36, 1171–1220.

Hu, Z. and K. Rao (2009): "Particulate air pollution and chronic ischemic heart disease in the eastern United States: a county level ecological study using satellite aerosol data," *Environmental Health*, 8, 26.

Jones, K., N. Patel, M. Levy, A. Storeygard, D. Balk, J. Gittleman, and P. Daszak (2008): "Global trends in emerging infectious diseases," *Nature*, 451.

Kammann, E. E. and M. P. Wand (2003): "Geoadditive models," *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 52, 1–18.

Karatzoglou, A., A. Smola, and K. Hornik (2012): "Kernel-based machine learning lab (version 0.9-14)," .

Kass, R. E. and L. Wasserman (1995): "A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion," *Journal of the American Statistical Association*, 90, 928–934.

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

Kelsall, J. and J. Wakefield (1999): "Discussion of 'Bayesian models for spatially correlated disease and exposure data' by Best et al." *Bayesian Statistics*, 6, 151.

Kimeldorf, G. and G. Wahba (1970): "Some results on Tchebycheffian spline functions," *Journal of Mathematical Analysis and Applications*, 33, 82–95.

Kindermann, R. and J. L. Snell (1980): *Markov Random Fields and Their Applications*, American Mathematical Society, Providence, R.I.

Lawson, A. (2001): *Statistical Methods in Spatial Epidemiology*, Wiley: London.

Lawson, A. (2008): *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology (Interdisciplinary Statistics)*, Chapman and Hall/CRC.

Liu, D., D. Ghosh, and X. Lin (2008): "Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models," *Bioinformatics*, 9, 292–302.

Liu, D., X. Lin, and D. Ghosh (2007): "Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models," *Biometrics*, 63, 1079–1088.

O'Sullivan, F., B. S. Yandell, and J. Raynor, William J. (1986): "Automatic smoothing of regression functions in generalized linear models," *Journal of the American Statistical Association*, 81, 96–103.

Richardson, S. (1992): "Statistical methods for geographical correlation studies," in P. Elliott, J. Cuzick, D. English, and R. Stem, eds., *Geographical and Environmental Epidemiology: Methods for Small Area Studies*.

Ruppert, D., M. Wang, and R. Carroll (2003): *Semiparametric Regression*, Cambridge: Cambridge University Press.

Scholkopf, B. and A. Smola (2002): *Learning with Kernels*, Cambridge, Massachusetts: MIT press.

Shawe-Taylor, J. and N. Cristianini (2004): *Kernel methods for pattern analysis*, Cambridge, UK ; New York : Cambridge University Press.

Siegmund, D. O. and K. J. Worsley (1995): "Testing for a signal with unknown location and scale in a stationary Gaussian random field," *The Annals of Statistics*, 23, 608–639.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002): "Bayesian measures of model complexity and fit (Pkg: P583-639)," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 64, 583–616.

Stern, H. and N. Cressie (1999): "Inference for extremes in disease mapping," in A. Lawson, A. Biggeri, D. Bohning, E. LesaKre, J. Viel, and R. Bertollini, eds., *Disease Mapping and Risk Assessment for Public Health*.

Stitson, M., A. Gammerman, V. Vapnik, V. Vovk, C. Watkins, and J. Weston (1999): "Support vector regression with ANOVA decomposition kernels," in B. Schölkopf, C. J. C. Burges, and A. J. Smola, eds., *Advances in Kernel MethodsSupport Vector Learning*, MIT Press, Cambridge, MA, 285–292.

Vapnik, V. N. (1998): *Statistical Learning Theory*, John Wiley & Sons.

Wahba, G. (1985): "A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem," *The Annals of Statistics*, 13, 1378–1402.

Waller, L. A. and C. A. Gotway (2004): *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons.

Yan, P. and M. Clayton (2006): "A cluster model for space-time disease counts," *Statistics in Medicine*, 25, 867–881.

Zou, G. (2004): "A modified Poisson regression approach to prospective studies with binary data," *American Journal of Epidemiology*, 159, 702–706.