

Bayesian modeling of multivariate spatial binary data with applications to dental caries

Dipankar Bandyopadhyay^{1,*}, Brian J. Reich² and Elizabeth H. Slate¹

¹*Department of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina, 135 Cannon Street, Charleston, SC 29425, U.S.A.*

²*Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A.*

SUMMARY

Dental research gives rise to data with potentially complex correlation structure. Assessments of dental caries yield a binary outcome indicating the presence or absence of caries experience for each surface of each tooth in a subject's mouth. In addition to this nesting, caries outcome exhibit spatial structure among neighboring teeth. We develop a Bayesian multivariate model for spatial binary data using random effects autologistic regression that controls for the correlation within tooth surfaces and spatial correlation among neighboring teeth. Using a sample from a clinical study conducted at the Medical University of South Carolina, we compare this autologistic model with covariates to alternative models to demonstrate the improvement in predictions and also to assess the effects of covariates on caries experience. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: autologistic; binary; caries; MCMC; spatial; WinBUGS

1. INTRODUCTION

Dental caries is an infectious disease damaging any surface of a tooth that is exposed to the oral cavity, but not structures retained within the bone [1]. Four main criteria contribute to caries formation: tooth surface (enamel or dentin); cariogenic (or potentially caries-causing) bacteria; fermentable carbohydrates (like sucrose); and time [2]. Dental caries develops by the localized dissolution of the tooth's hard tissues, caused by acids produced by the bacteria in the biofilms (dental plaque) on the teeth and eventually leading to cavities. The biofilm consists of highly

*Correspondence to: Dipankar Bandyopadhyay, Department of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina, 135 Cannon Street, Charleston, SC 29425, U.S.A.

†E-mail: bandyopd@musc.edu

Contract/grant sponsor: South Carolina COBRE for Oral Health Research; contract/grant number: NIH/NCRR P20 RR017696-06

Contract/grant sponsor: NIH; contract/grant number: R01 LM009153

Received 7 October 2008

Accepted 22 May 2009

cariogenic agents and a matrix made up mainly of extracellular polysaccharides. The destructive acids are formed when fermentable carbohydrates (sugars) reach these biofilms resulting in tooth damage. If left untreated, caries can give rise to tooth pulp infection, which can spread to supporting tissues and jaws, resulting in advanced disease conditions, which are often painful [3]. The caries process does not have an inevitable outcome; different individuals will be susceptible to different degrees depending on the shape of the teeth, oral hygiene habits and the buffering capacity of the saliva.

In clinical studies that examine dental caries status, each participant produces multiple outcome data [4] with tooth surfaces clustered within tooth that is also clustered within the oral cavity of that subject. In this paper, we analyze data from a clinical study [5] conducted at the Medical University of South Carolina (MUSC) to determine the dental caries status of Type-2 diabetic Gullah-speaking African-Americans. The total count of decayed, missing and filled surface, which we abbreviate as DMFS, was recorded for each tooth. Note that the DMFS counts in our data pertain to permanent tooth. Although the popular DMFS measure [6] is the indicative of the cumulative severity of caries status for that particular tooth, it does not explore the caries experience for a particular tooth surface nor does it explain the rate of how a diseased surface within a particular tooth influences the decay of a neighboring surface. Here, we develop a model for caries status of tooth surfaces that accommodates the subject and tooth level clustering and facilitates borrowing of strength across all teeth when assessing the effects of covariates such as age, gender, smoking habits, brush-floss habits, poverty status, etc.

Presence of clustering (teeth within subject and surfaces within a tooth) calls for binary multi-level modeling [4] accommodating significant heterogeneity through random effects. Additionally, different surfaces within a tooth and between neighboring teeth have spatial association with each other. In a recent dental study examining presence/absence of caries experience on the eight deciduous molars of seven-year old children [7], a high association was found between symmetrically opponent, vertically opponent as well as diagonally opponent molars. However, the study focussed on caries experience of a particular molar and not the binary caries status of a tooth surface in terms of DMFS. A particular tooth surface experiencing caries lesions may affect a set of neighboring surfaces of the same tooth and also surfaces of neighboring teeth with caries. To analyze these correlated binary data with latent spatial autocorrelation, we use an extension of the autologistic model of Besag [8, 9]. Autologistic models are suitable for establishing a correspondence between a binary response (presence/absence of caries) and the potential explanatory covariates by a logistic regression, while accounting for spatial dependence by an autoregression [10]. They also provide estimates of the probability of caries occurrence at a particular surface and predict the outcome at some unsampled surface, thereby improving understanding of a space-referenced binary outcome.

Autologistic regression models have been used as an attractive tool in many disciplines like epidemiology, image analysis and environmental studies [11–14]. However, application of a multivariate autologistic model in analyzing caries data has not been explored and it comes with unique challenges. For example, there are layers of clustering, e.g. by subject and tooth. Also, unlike most application where the spatial structure is clearly defined by a grid or geopolitical boundaries, the spatial structure within a mouth must be carefully chosen. Very recently, multiple neighbor relationships within the mouth in the context of periodontal disease modeling has been explored in [15]. Building on [16] for Gaussian data, we identify several types of spatial associations, such as surfaces on the same or neighboring tooth or contacting surfaces on different teeth. We show that accounting for these different types of spatial associations dramatically improves the predictive performance of our model. However, statistical inference with autologistic models is

associated with complications that arise due to the intractable normalizing factor and consequential problems in writing the data likelihood. We circumvent this problem by utilizing a pseudolikelihood estimation (PLE) technique [17]. Under suitable regularity conditions, maximum PLE's are consistent and asymptotically normal [18]. However, the efficiency of the PLE's depends on the strength of spatial dependence and is comparable to MLE's provided the spatial association is not too strong [13, 19].

In this present study, we adopt a Bayesian approach to address two broad questions arising from our available dental data: (a) How do subject level covariates influence caries status at a surface level after accounting for spatial association? and (b) How does spatial association influence development of caries lesions in a group of neighboring teeth? The key advantage of relying on Bayesian inference to address these questions is the ability to incorporate background (prior) information about the unknown parameters in the model, both for subject level covariates and those determining spatial dependencies. To implement the Bayesian modeling, we adopt the pseudolikelihood as an approximation to the full data likelihood along with relevant Markov chain Monte Carlo (MCMC) steps [20]. We compare our random effects autologistic model accommodating spatial components to alternative models that do not distinguish between the effect of different spatial associations among the tooth surfaces through cross-validators techniques [21] and posterior predictive model choice criterion [22].

The remainder of the paper unfolds as follows. Section 2 describes the aforementioned data on dental caries that motivates this research. In Section 3, we propose our autologistic model that incorporates spatial dependence among different tooth surfaces. Section 4 proposes Bayesian inference using pseudolikelihood to estimate model parameters and prediction. Section 5 applies the spatial autologistic regression model to the dental caries data and uses Bayesian model selection tools to determine the best model. It also summarizes and discusses the estimation of spatial association and fixed effects parameters for the best model. Conclusions and future developments are in Section 6.

2. MOTIVATING DATA

The motivating data were collected from a clinical study conducted by the Center for Oral Health Research (COHR) of the Medical University of South Carolina (MUSC) as part of the South Carolina Center for Biomedical Research Excellence (COBRE) Program for Oral Health. The study was primarily aimed to assess caries status in Type-2 diabetic Gullah-speaking (or simply Gullah) African-American population (13 years or older) residing in the coastal sea-islands of South Carolina. All subjects answered a detailed questionnaire that focussed on their social, medical and dental history and underwent an oral exam. Since this is a part of an ongoing study, we selected a random sample of 100 subjects with complete covariate information. The primary response (outcome) variable is a binary (1/0) indicator that denotes whether a particular surface within a tooth of the individual is either carious, i.e. either decayed (D), missing (M) or filled (F), or not. Note that if the whole tooth is missing due to caries, then we consider all the surfaces to be missing, whereas if the loss is due to a cause other than caries, the tooth does not contribute to the data analysis. The reason for a missing tooth was determined from the questionnaire administered to the participants in our study, which distinguished caries from gum disease, orthodontics, injury and other causes for each missing tooth. We acknowledge that this self report may be inaccurate, but it is the best information available. Each quadrant of teeth

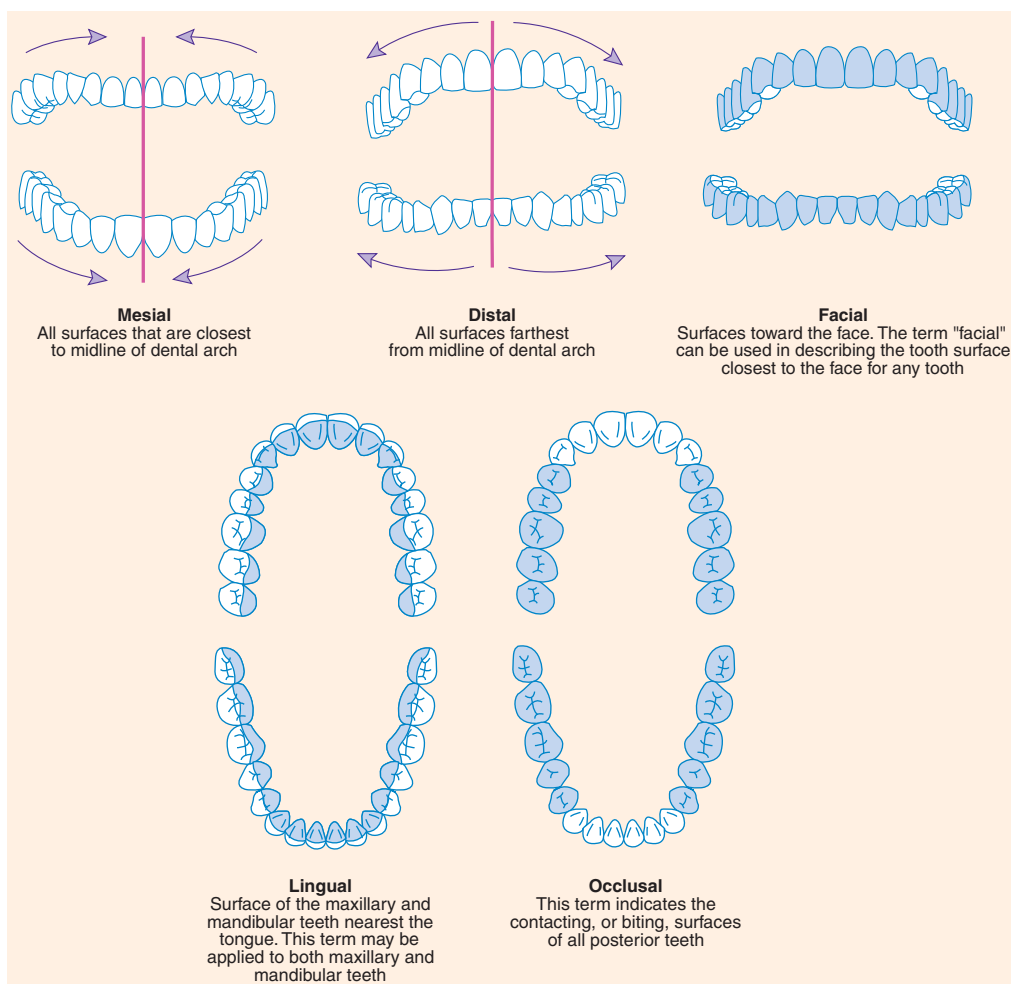


Figure 1. Different surfaces of permanent dentition within a mouth. Adapted from *Dental Hygiene Theory and Practice*, 2nd Edition, Elsevier by Michele Darby and Margaret Walsh.

(consisting of a cluster of eight teeth, two in each jaw) in a human mouth [6] is made up of (a) the non-anterior teeth (three molars and two pre-molars) and (b) the anterior teeth (one incisor and two canines). For measuring caries, each non-anterior tooth contribute five surfaces (occlusal, mesial, distal, facial and lingual) while an anterior tooth contributes four surfaces (the occlusal not being recorded) to the response. Figure 1 illustrates the different surfaces of teeth within a mouth. Additionally, several subject level covariates were also collected including Age (in years), Gender (0=Male, 1=Female), Smoking status (0=Never, 1=Smoker), Brush-Floss (1=Brushed twice and flossed once every day, 0=otherwise) and Poverty (1=Below poverty line, 0=Above poverty line). The mean age of the subjects in the sample is 53 years with a range of 27–73 years. Although study recruitment was gender blind, females participated at a higher rate (75 per cent) than males and this is reflected in the sample drawn for our analyses. Higher rates of enrollment among

females than males is not unusual for studies among Gullah African Americans, unfortunately [23] and our clinical collaborators are working with the community (using patient navigator techniques, for example) to achieve better gender balance. There are only about 8 per cent ‘former smokers’ in our sample. Thus, in order to avoid a separate category for ‘smokers’ and keep our autologistic model relatively simple, we collapsed the groups of ‘former smokers’ and ‘present smokers’ into ‘smokers’, which makes about 36 per cent of the study sample. The poverty status was determined according to family per capita income level, i.e. total income generated by the family divided by the number of family members. If this value was less than \$5000, the subject was classified as below the poverty line. About 34 per cent of the subjects in the sample live below the poverty line and 24 per cent reported to have brushed twice and flossed once every day. In this paper, we explore the effects of spatial associations among tooth surfaces with respect to surface level dental caries while controlling for the effects of these subject-level covariates.

3. STATISTICAL MODEL

Let $\{Y_{ijs} : (i, j, s) \in D\}$ be the binary (1/0) indicator of DMFS collected at the s th surface of the j th tooth of individual i , where D denotes the set of indices for the observed data, $i = 1, \dots, 100$, $j = 1, \dots, 32$ and $s = 1, \dots, 4$ (for incisors and canines) and $s = 1, \dots, 5$ (for pre-molars and molars). Define $|D|$ to be the total number of observations. We assume that the random variable Y_{ijs} is distributed as Bernoulli (p_{ijs}). Under a generalized linear mixed model (GLMM) framework [24], a basic logistic regression framework incorporating heterogeneity at the subject, tooth and surface level is defined by

$$\text{logit}(p_{ijs}) = \beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + G_{ijs} \quad (1)$$

where β_0 is the intercept term, \mathbf{X}_i denotes the vector of subject-level covariates, such as age, gender, smoking status, brush-floss habits and poverty with the corresponding fixed-effects parameter vector $\boldsymbol{\beta}$, and G_{ijs} are the random effects corresponding to the binary response of the (i, j, s) th surface.

3.1. Random effects autologistic model

Since the responses are spatially correlated (through associations among different surfaces within the same tooth or neighboring teeth), the independence assumption among the response is questionable and hence there is need to introduce a formal spatial structure. The spatially correlated heterogeneity accounts for any unmeasured risk to caries progression that is common for adjacent/proximal tooth surfaces [25]. We now define the neighborhood in our spatial set-up. We consider three types of spatial neighbors for tooth surfaces. They are:

1. Surfaces on the same tooth (first-order)
2. Surfaces on adjacent teeth on the same jaw (second-order), and
3. Contacting surfaces on opposite jaws (between-jaw).

We also assume a Markov random field (MRF) model such that the full conditional distribution for a surface depends only on the outcome of its neighboring surfaces. Thus, accommodating subject-specific (spatially un-correlated) random-effects U_i along with spatially associated random

effects E_{ijs} at the surface level, the random effects G_{ijs} corresponding to the (i, j, s) th surface can be decomposed as

$$G_{ijs} = U_i + E_{ijs} \tag{2}$$

where

$$E_{ijs} = \sum_{l \neq s} b_{ls} Y_{ijl} + \sum_{m \sim j} \sum_l c_{ls} Y_{iml} + \sum_{p \leftrightarrow j} d_s Y_{ips} \tag{3}$$

In the above specification (3), the (spatial) regression parameters $\mathbf{b} = \{b_{ls} : b_{ls} = b_{sl}, l < s; l, s = 1, \dots, 5\}$ control first-order spatial association, i.e. spatial association of different surfaces within the same tooth; $m \sim j$ denotes that teeth m and j are adjacent on the same jaw, $\mathbf{c} = \{c_{ls} : c_{ls} = c_{sl}, l, s = 1, \dots, 5\}$ controls the second-order spatial association, i.e. association among surfaces of adjacent teeth on the same jaw; $p \leftrightarrow j$ denotes that teeth p and j are contacting on opposite jaws and $\mathbf{d} = \{d_k : k = 1, \dots, 5\}$ controls between jaw spatial association. The logistic model is thus extended to a spatial autologistic model accounting for the simultaneous presence of unknown sources of (tooth and site-specific) spatial variation through spatially smoothed random effects and unknown subject level heterogeneity through subject-specific random effects. The independent random effects U_i influence the estimates of the unknown spatial parameters that seems to control the spatial association among both the site-specific responses and random effects [26]. Thus, under the autologistic specification given by (1–3), we have

$$\Pr(Y_{ijs} = y_{ijs} | Y_{i'j's'}, (i, j, s) \neq (i', j', s'), \mathbf{U}, \mathbf{X}) = \frac{\exp\{y_{ijs}[\beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + U_i + E_{ijs}]\}}{1 + \exp\{\beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + U_i + E_{ijs}\}} \tag{4}$$

where E_{ijs} is given in (3). Note that the conditional distribution of Y_{ijs} depends on all other observations only through its neighbors.

3.2. Data likelihood

Let $\boldsymbol{\Omega} = (\beta_0, \boldsymbol{\beta}, \mathbf{b}, \mathbf{c}, \mathbf{d})$ denote the parameter vector in this spatial autologistic regression model. The primary goal is to estimate $\boldsymbol{\Omega}$ and draw inference on these parameters to determine the degree of spatial association. Then, the joint likelihood without integrating out the random-effects U_i is given by

$$L(\boldsymbol{\Omega}; \mathbf{U}, \mathbf{X}, \mathbf{y}) = c(\boldsymbol{\Omega})^{-1} \exp \left\{ \sum_{(i,j,s) \in D} y_{ijs} (\beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + U_i + \frac{1}{2} E_{ijs}) \right\} \tag{5}$$

where

$$c(\boldsymbol{\Omega}) = \sum_{\mathbf{y}} \exp \left\{ \sum_{(i,j,s) \in D} y_{ijs} (\beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + U_i + \frac{1}{2} E_{ijs}) \right\} \tag{6}$$

and $\sum_{\mathbf{y}}$ denotes the sum over all possible $2^{|D|}$ realizations of Y . The normalizing constant in (6) does not have a closed form, hence direct maximization of the likelihood in (5) becomes analytically intractable. This motivated the pseudolikelihood approximation in which the autocovariate term is retained in the regression, but the surfaces are assumed to be independent for constructing the likelihood [27, 28]. The Hammersley–Clifford Theorem [29] posits that the joint distribution of the

spatial process as determined by the pseudolikelihood function specified by the full conditionals as in (4) is well defined, and is given as the product of full conditionals written as:

$$\text{PL}(\boldsymbol{\Omega}; \mathbf{U}, \mathbf{X}, \mathbf{y}) = \prod_{(i,j,s) \in D} \Pr(Y_{ijs} = y_{ijs} | \mathbf{U}, \mathbf{X}, Y_{i'j's'}, (i, j, s) \neq (i', j', s')) \quad (7)$$

3.3. Estimation schemes

Under a frequentist set-up, the estimate of the model parameters obtained by maximizing (7) is termed the maximum pseudolikelihood estimate (MPLE). The estimation technique using pseudolikelihood approximation of the usual likelihood function is ‘no free lunch’ [17] and is actually a trade-off between computational simplicity and statistical efficiency [11]. Estimation of the autologistic model using pseudolikelihood may be affected by high levels of intrinsic spatial correlation [13]. However, very recently it was shown through extensive simulation studies [27] that the autologistic model was robust to high levels of intrinsic spatial auto-correlation, providing good predictive performance at the highest levels of simulated underlying spatial dependence. Several authors have considered alternative techniques of computing the MLE in this spatial autologistic set-up by approximating the constant through MCMC-based techniques, like, path sampling [30]. Huffer and Wu [31] considered that Markov chain Monte Carlo (MCMC) approximation of the normalizing constant and proposed MCMC MLE method of estimation. Berthelsen and Møller [32] used perfect simulation and path sampling techniques to obtain likelihood and nonparametric Bayesian MCMC inference for spatial point process. Very recently, Zheng and Zhu [33] considered combining a Metropolis–Hastings algorithm within a Gibbs sampler to obtain posterior distribution of model parameters as well as the posterior predictive distributions in a spatio-temporal set-up. However, all these methods involve intense computations, which become very complicated in our set-up because of the three types of spatial association and the large number of observed data points.

In this paper, we use the pseudolikelihood approximation (bypassing the complications involved in our data structure to estimate the normalizing constant) to the full likelihood under a Bayesian framework. Most of the recent work on Bayesian autologistic models [14, 34] has been implemented using a non-normalized distribution employing MCMC methodology [20]. Our method is particularly straightforward and simple to implement in available freeware WinBUGS [35] and is in contrast to the method of [14, 34] as we do not rely on the normal approximation of the pseudolikelihood or the nonlinear maximization steps involved in those approximations.

4. BAYESIAN INFERENCE

Our Bayesian approach has several important advantages over the likelihood-based frequentist methods. Besides the ability to incorporate pertinent background (prior) information as described in the Introduction, the Bayesian paradigm utilizes the Gibbs sampler and MCMC algorithms [36] to determine the estimates of exact posterior distributions and hence does not depend on large-sample methodology other than the number of MCMC iterations, which can be increased easily [37]. The Bayesian method provides the entire posterior distribution of the parameters and any arbitrary parameter functionals considering both the data likelihood and the priors assigned to the parameters. Next, we investigate the choice of priors on model parameters to conduct our Bayesian inference.

4.1. Prior and hyperprior distributions

We specify practical weakly informative prior opinion on the fixed effects regression parameters β , the spatial autologistic regression parameters $\theta_1 = (\mathbf{b}, \mathbf{c}, \mathbf{d})$ and non-informative opinion on the intercept term β_0 and the random effects U . We assume the elements of the parameter space Ω are independently distributed. Since we have no prior information from historical data or experiment, we assign weakly informative priors/hyperpriors to obtain well-defined posteriors. The prior distributions were centered at zero as we were uncertain about parameter values and thereby providing shrinkage toward zero as a tool to ascertain computational stability for moderately diffuse priors in mixed effects model [38].

Specifically, we assign weakly informative i.i.d. Normal (0, Precision=0.25) priors on the elements of β . This implies that the density of the associated odds-ratio centered at 1 with 95 per cent intervals (e^{-4}, e^4) , sufficiently will include a wide range of prior guesses [39]. For the intercept term β_0 , we use an improper uniform prior (using the dflat() option in WinBUGS). For the elements in θ_1 , we put double exponential (DE) priors (also called Laplace priors) centered at 0 to reflect a skeptical view about spatial effects and requiring strong evidence from data to suggest otherwise *aposteriori*. The DE prior (also called the Bayesian Lasso) is widely used to induce sparsity as it is strongly peaked at zero with a heavy tail [40]. This is particularly ideal for situations in Bayesian regression frameworks similar to our random-effects autologistic model where we have a highly parameterized θ_1 (due to three different neighborhood structures for the five different tooth surfaces) and thereby strive to encourage covariate removal [41] unless strongly supported by the data. The zero centered DE $\beta \sim DE(0, \tau)$, with p.d.f $(\tau/2) \exp(-\tau|\beta|)$ for $-\infty < \beta < \infty$, has variance $2/\tau^2$. For the choice of the hyperprior on the Lasso parameter τ^2 , we choose a Gamma(1,1) density (also exponential with mean 1) to reflect a diffuse hyperprior far enough from zero [42]. For a reasonably non-informative prior on the random effects U , we use $U \sim \text{Normal}(0, \text{Precision}=(1/\sigma_u^2))$ with $\sigma_u \sim \text{Uniform}(0, 100)$ following [43].

4.2. Posterior distribution and estimation

The posterior conclusions from our Bayesian analysis will be based on the joint posterior distribution of all the model parameters conditional on the data which is obtained by combining the pseudolikelihood given in (14) and the joint prior densities using Bayes' theorem:

$$p(\Omega, \mathbf{U}, \tau, \sigma_u^2 | \mathbf{X}, \mathbf{y}) \propto \text{PL}(\Omega; \mathbf{U}, \mathbf{X}, \mathbf{y}) \times \pi_0(\beta_0) \times \pi_1(\beta) \times \pi_2(\theta_1 | \tau) \times \pi_3(\mathbf{U} | \sigma_u^2) \times \pi_4(\tau^2) \times \pi_5(\sigma_u^2) \tag{8}$$

where $\pi_j(\cdot), j=0, \dots, 5$ denote the prior/hyperprior distributions on the model parameters as described above.

The full Bayesian approach utilizing posterior densities is particularly attractive over the MCMC MLE method of [13] which exhibits some instability in searching the solution of the MLE score equation. Our autologistic Bayesian framework carefully avoids the whole issue of estimating the normalizing constant, which often turns out to be computationally expensive (as mentioned earlier) with the increase in the number of spatial association terms and data points. Posterior computations were performed via Monte Carlo approximations with the help of MCMC techniques [20]. The forms of the conditional distributions of model parameters are proportional to the joint posterior density as in (8) and can be obtained by retaining relevant parameter associated quantities from the joint posterior. The computations are done via WinBUGS software which automatically samples

from the pseudolikelihood and generates the conditional distributions needed for the MCMC sampling. The recent monograph [41] underscores the versatility of WinBUGS software to sample directly from the autologistic (pseudolikelihood) framework in the context of Bayesian disease mapping and relative risk estimation. The samples from the posterior distribution in (8) thus obtained by MCMC allow to compute summary measures and credible intervals of any arbitrary functional of the parameters.

4.3. Model selection

In the midst of a variety of model selection procedures the Bayesian toolbox has to offer, our model comparisons are based solely on predictive performances of competing models. Currently, there is not any unanimous ‘correct’ approach for model selection [44]. We present two popular approaches to Bayesian model selection based solely on the predictive performance of the model using cross-validation techniques and posterior predictive L-measure. If $\Theta = (\Omega, \tau, \sigma_u)$ denotes the entire parameter space in our model and \mathbf{y}_{pr} denotes the predictive data vector, then the posterior predictive distribution is given by:

$$p(\mathbf{y}_{pr}|\mathbf{y}) = \int p(\mathbf{y}_{pr}|\Theta)p(\Theta|\mathbf{y})d\Theta \quad (9)$$

One can obtain predictive data easily from a converged posterior sample and samples from the posterior predictive distribution are replicates of the observed model generated data.

Cross-validation techniques [21] involve fitting the model to a randomly selected subset of the data and then predicting the remaining data. The estimated error in this prediction (misclassification rates for our binary response) can be compared across alternative models. Clearly, one chooses the model with the lowest estimated predictive error. In our example, our cross-validated measure randomly removes 10 per cent of the data and monitors the proportion of incorrectly classified deleted binary responses. This is given as:

$$\begin{aligned} CV_1 &= N_1^{-1} \sum_{(i,j,s)} \{y_{ijs} * (1 - y_{ijs,pr})\} \\ CV_0 &= N_0^{-1} \sum_{(i,j,s)} \{(1 - y_{ijs}) * y_{ijs,pr}\} \end{aligned} \quad (10)$$

where y_{ijs} is the observed response, $y_{ijs,pr}$ is the predicted response, N_0 is the total number of deleted 0’s and N_1 is the total number of deleted 1’s from our study and the summations are taken over the respected deleted observations. Under a Bayesian paradigm, this measure is obtained as a posterior mean of the proportions over all parameters and the data computed as ECV_0 and ECV_1 where the expectation is over all other parameters and data. Thus, ECV_0 and ECV_1 are the average posterior misclassification probabilities for deleted zeros and ones, respectively.

We also evaluate the predictive performance of our models using posterior predictive L-measure as proposed by [22] with modifications [45] to clustered correlated data. The L-measure quantifies model fit through comparisons of model-based posterior predictive distribution to equivalent features from observed data [45] utilizing an user-defined loss-function. The key idea is to minimize the expectation of the posterior predictive loss, i.e. the loss with respect to the posterior predictive distribution. We summarize our observed multivariate responses y_{ijs} to a single univariate

subject-specific summary value $M_i = \sum_{j,s} w_{js} y_{ijs}$ where $w_{js} = 1$ for all j and s as described in [45]. Then, the L-measure using the squared-error loss-function [22] is defined as:

$$L = \sum_{i=1}^n \sigma_{i(M)}^2 + \frac{k}{k+1} \sum_{i=1}^n (\mu_{i(M)} - M_i)^2 = P + \frac{k}{k+1} G \tag{11}$$

Here, $\mu_{i(M)} = E(M_{i,pr} | \mathbf{y})$ and $\sigma_{i(M)}^2 = \text{Var}(M_{i,pr} | \mathbf{y})$ where $M_{i,pr} = \sum_{j,s} w_{js} y_{ijs,pr}$. The first term P is a penalty term where model overfitting might result in large predictive variances and hence larger P , whereas the second term $G = \sum_{i=1}^n (\mu_{i(M)} - M_i)^2$ is a goodness-of-fit term. The choice of k is determined by the weight placed on G relative to P . As $k \rightarrow \infty, k/(k+1) \rightarrow 1$ and hence, L reduces to $G + P$, which is the overall summary statistic used here. The smaller the value of L , the better is the model's predictive performance.

We did not apply the popular likelihood-based deviance information criterion (DIC) [46] measure for Bayesian model choice in our problem as we were approximating the true likelihood by the pseudolikelihood. Also, ECV and L-measures are invariant to model parametrization (we employed several parameters to determine spatial associations in our highly parameterized model) as it is based solely on predictive distributions.

5. DATA ANALYSIS AND FINDINGS

In this section, we apply our method to the real data-set on caries progression described in Section 2. Including all the subject-specific covariates (age, gender, smoking status, brush-floss habits and poverty), we posit four competing models to fit our data by varying the spatial parameters, viz. (1) whether the spatial model is independence ($b=c=d=0$), (2) isotropic ($b=c=d$), (3) first-ordered ($c=d=0$) or (4) non-stationary (the full model as in Section 3). To select our best model, we compare between the four models using the model selection criterion described in the previous section.

1. *Model 1*: Simple random effects logistic regression with no spatial association, i.e. **b=c=d=0**
2. *Model 2*: Random effects autologistic regression with same autologistic parameters, i.e. **b=c=d**
3. *Model 3*: Random effects autologistic regression with only the first-order autologistic parameters, i.e. **c=d=0**
4. *Model 4*: Random effects autologistic regression with varying autologistic parameters, i.e. **b, c and d** unconstrained.

For all of the models above, we did not include any between-jaw spatial association in the model between surfaces 2–5. Thinking pragmatically, there is minimal spatial interference among tooth surfaces located on two opposite jaws other than the occlusal surfaces of the non-anterior teeth (molars and pre-molars), which often come in contact during mastication or other mechanical action. Considering this fact and also to avoid an overparameterized model, the parameters $d_i, i = 2, \dots, 5$ were set to zero. Hence, all subsequent model comparison and data analysis were performed considering **d=d₁**. Within WinBUGS, we used the Gibbs sampler [20] to obtain samples from the posterior distributions powered by MCMC approximations. For each of the models, we ran two chains with widely dispersed initial values. Initial values of precision parameters were selected arbitrarily. Our pseudolikelihood based autologistic model induces substantial model complexity

due to two layers of clustering (subject \rightarrow teeth \rightarrow surface) as well as the first-order, second-order and between-jaw spatial associations among teeth surfaces. Thus, parameter convergence was achieved at the cost of a high burn-in period of 45 000 with a total iteration size of 50 000. For each model parameter, the posterior convergence was assessed using trace plots, auto-correlation plots, density estimates as well as the Gelman–Rubin potential scale-reduction factor R [44]. To reduce the auto-correlation among successive Markov draws, we used a spacing of 2. After discarding the initial burn-in of 45 000 samples, we used the remaining 2500 to compute the posterior estimates of our parameters of interest. The associated WinBUGS code is available in the webpage of the first author (<http://people.musc.edu/~bandyopd>).

Table I presents the comparison among the four competing models using the Bayesian model choice criterion. With the increase in model complexity by inclusion of spatial autologistic parameters, both the cross-validation values (ECV_0 and ECV_1) and the L-measure (determining goodness-of-fit and penalty) decrease substantially. Both G and P are substantially lower in Model 4 as compared with Models 2 and 3, though Model 4 contains more estimable parameters than both Models 2 and 3. However, ECV_1 (misclassification index for deleted 1's) were not substantially different for Models 3 and 4. Performance of Model 1 (the simple random effects logistic regression model) is worst among all the four models, indicating the presence of significantly strong spatial association in the caries data. It is clear that Model 4 reigns supreme among all the models in terms of predictive performance. We thus select Model 4 as our best fitting model and henceforth discuss parameter estimates based on this model.

5.1. Spatial association parameters

Tables II and III report the posterior estimates of the mean, standard deviation and 95 per cent credible interval (CI) of first-order, second-order and between-jaw spatial association parameters (as described in Section 3) under Model 4. The surfaces within a tooth are denoted as 1 = Occlusal, 2 = Mesial, 3 = Distal, 4 = Facial, 5 = Lingual. For example, the parameter b_{12} controls (in log-odds terms) the ‘first-order’ spatial association between surfaces 1 and 2, i.e. the occlusal and the mesial surface; the parameter c_{13} controls (in log-odds terms) the ‘second-order’ spatial association between surfaces 1 and 3, i.e. the occlusal and the distal surface and similarly, the parameter d_1 controls (in log-odds terms) the ‘between-jaw’ spatial association, i.e. the association between the occlusal surfaces of the molars/pre-molars on the opposing jaws. From Table II (1st-order associations), it is immediate that all the spatial association parameters are positive. The strongest association observed is between surfaces 2 and 4, i.e. the mesial and facial surfaces while the weakest association was found between surfaces 1 and 4, i.e. between occlusal and facial surface. In terms of odds ratio, we can interpret that after controlling for all other fixed-effect and spatial covariates, the odds of having a decayed occlusal surface is $e^{0.631} = 1.93$ times more (considering

Table I. Model comparison using cross-validation and L-measure.

Model	ECV_0	ECV_1	G	P	L
Model 1	0.338	0.515	3577.0	2364.0	5941.0
Model 2	0.223	0.259	2474.0	500.5	2974.5
Model 3	0.111	0.108	823.1	813.7	1636.8
Model 4	0.067	0.096	557.0	493.7	1050.7

Table II. Posterior estimates of 1st-order autologistic parameters with 95 per cent credible intervals for Model 4.

Parameter	Mean	Std. dev.	95 per cent CI
b_{12}	1.922	0.195	(1.572, 2.225)
b_{13}	2.806	0.104	(2.593, 2.969)
b_{14}	0.631	0.251	(0.238, 1.049)
b_{15}	1.551	0.204	(1.144, 1.874)
b_{23}	1.636	0.157	(1.28, 1.897)
b_{24}	3.386	0.142	(3.052, 3.592)
b_{25}	2.152	0.155	(1.869, 2.51)
b_{34}	1.961	0.164	(1.625, 2.259)
b_{35}	2.697	0.214	(2.21, 3.017)
b_{45}	2.713	0.138	(2.418, 2.926)

Subscripts (i, j) denote association between i th and j th surface, $(i, j) \in 1, 2, 3, 4, 5$ where 1=Occlusal, 2=Mesial, 3=Distal, 4=Facial and 5=Lingual.

Table III. Posterior estimates of 2nd-order and between-jaw autologistic parameters with 95 per cent credible intervals for Model 4.

Parameter	Mean	Std. dev.	95 per cent CI
c_{11}	1.95	0.0931	(1.781, 2.124)
c_{12}	-0.708	0.0941	(-0.881, -0.534)
c_{13}	-0.512	0.1056	(-0.6968, -0.275)
c_{14}	0.016	0.1392	(-0.292, 0.245)
c_{15}	-0.218	0.114	(-0.422, -0.023)
c_{22}	1.219	0.253	(0.816, 1.775)
c_{23}	2.048	0.193	(1.724, 2.501)
c_{24}	-1.538	0.123	(-1.758, -1.323)
c_{25}	-0.822	0.116	(-1.044, -0.603)
c_{33}	-0.091	0.262	(-0.59, 0.278)
c_{34}	-0.869	0.183	(-1.186, -0.381)
c_{35}	-0.274	0.197	(-0.582, 0.063)
c_{44}	3.934	0.162	(3.665, 4.249)
c_{45}	-0.908	0.196	(-1.371, -0.636)
c_{55}	2.29	0.205	(1.937, 2.667)
d_{11}	2.41	0.211	(2.123, 2.88)

Subscripts (i, j) denote association between i th and j th surface, $(i, j) \in 1, 2, 3, 4, 5$ where 1=Occlusal, 2=Mesial, 3=Distal, 4=Facial and 5=Lingual.

first-order association) if the facial surface is decayed with the 95 per cent credible interval to be (1.269, 2.856) and so on.

Among the second-order spatial associations (Table III), the strongest association (log-odds=3.934) is between the two facial surfaces on neighboring teeth. The odds ratio interpretation is the same as above. Some of the second-order association parameters, i.e. between surfaces (1,4), (3,3), (3,5) do not significantly (95 per cent credible interval includes 0) influence the decay process. Intuitively, we would expect the spatial association parameters to be positive, or non-significant at all. However, several significantly negative associations are apparent in Table III, i.e. surfaces (1,2),

Table IV. Posterior estimates of fixed/random-effects parameters with 95 per cent credible intervals for Models 1 and 4.

Parameter	Model 1			Model 4		
	Mean	Std. dev.	95 per cent CI	Mean	Std. dev.	95 per cent CI
Intercept	-1.305	0.4473	(-2.2, -0.5126)	-5.076	0.081	(-5.21, -4.962)
Age	2.072	0.6853	(0.9745, 3.258)	-0.0198	0.023	(-0.069, 0.024)
Gender	0.439	1.743	(-1.878, 2.666)	-0.019	0.035	(-0.078, 0.051)
Smoking status	-1.72	3.314	(-5.441, 1.905)	-0.026	0.064	(-0.112, 0.067)
Brush-floss	0.722	0.854	(-0.556, 2.122)	-0.019	0.044	(-0.068, 0.098)
Poverty	-0.5003	1.238	(-2.315, 1.191)	0.024	0.056	(-0.067, 0.109)
σ_u	3.232	0.410	(2.563, 4.023)	0.075	0.055	(0.015, 0.189)
τ				0.532	0.133	(0.319, 0.818)

(1,3), (1,5), (2,4), (2,5), (3,4) and (4,5). We speculate that there is some possible collinear effect among the first- and second-order parameters and that the negative values of c are serving primarily to fine tune the model fit as determined by the significantly positive first-order parameters b . For the between-jaw association parameters (Table III), there is significant association (log-odds=2.41) between the occlusal surfaces of the molar/pre-molars. As suggested by a reviewer, we fitted a relatively simple and less parameterized model (labeled Model 3) with $\mathbf{c}=\mathbf{d}=\mathbf{0}$, i.e. containing only the first-order associations, to explore parameter estimates as well as model fit. Although model comparison results (Table I) reveal Model 4 to perform better than the simpler Model 3, the estimated standard deviations of parameters \mathbf{b} for Model 3 are lower than those in Model 4.

5.2. Fixed effects parameters and random effects

The posterior mean, standard deviation and 95 per cent credible interval for each fixed effects parameters (for both Models 1 and 4) are given in Table IV. It should be noted that the estimates of the fixed effects parameters (for gender, age, etc) are in fact conditional (site-specific) estimates. Here, β_j represents the log-odds of a diseased surface (in terms of DMFS index) at the same spatial location for subjects with different values of the covariate X_j , conditioned on the same values at the neighboring locations for each subject.

However, we warn against the direct comparison of fixed-effects estimates of Models 1 and 4 (as presented in Table IV) especially for these data with strong spatial associations. The intercept term is significant for both models, though it is much smaller for Model 4. After controlling for all other model covariates, we find 'age' to be positively associated (log-odds=2.072) with tooth decay in Model 1, the simple random effects logistic regression model. None of the other fixed effects appear to be significantly influencing dental caries. For Model 4 (model with spatial autologistic parameters as well as the fixed-effects), none of the covariates appear to be significant though the fixed-effects estimates has considerably smaller standard deviations than what was obtained in Model 1.

Using a normal density to model the random effects, we obtain (conditional) log-odds interpretation of the fixed effects but this information is not preserved marginally. In order to provide both conditional (site-level) and marginal (subject-level) fixed-effects parameter interpretation in terms of log-odds, one might choose to use the Bridge distribution as proposed in [47] assuming the pseudolikelihood to be a reasonable approximation to the true likelihood. Although we are

not aware of a closed-form expression using a full likelihood, simulation studies (not reported here) show that the mouth-level average response using the pseudolikelihood framework increases monotonically with $X^T\boldsymbol{\beta}$. Hence, inference on $\boldsymbol{\beta}$ in the autologistic model still informs about the population mean number of DMFS and hence the risk factors (the fixed effects) may not always be considered a nuisance in the full autologistic model. This phenomenon may not be unique to the autologistic model and has been discussed in other settings [48, 49] in the context of exploring the effect of spatial correlation on the fixed effect parameters in a model.

Using our Bayesian inference, we conclude that first-order spatial association (surfaces in the same tooth) is significantly strong and sufficiently explains the spatial variation in these caries data. Note that the estimates of both the random effects standard deviation, σ_u , and the posterior standard deviation of the fixed effects are substantially smaller for Model 4 than Model 1. The posterior estimate of the Lasso parameter (τ) (which was assumed *a priori* to be exponential with mean = 1) is 0.532 with 95 per cent credible interval to be (0.319, 0.818).

We also conducted a sensitivity analysis for the prior distribution of the random effects precision parameter σ_u and the prior normal precision of the fixed effects parameters for all our models. In particular, we allowed $\sigma_u \sim \text{Uniform}(0, k)$ where $k \in \{50, 150, 200\}$ and the normal precision on the fixed effects to be 0.1, 0.01. We checked the sensitivity analysis by changing one parameter at a time and refitting models 1–4. Although slight changes were observed in the parameter estimates (fixed-effects and spatial parameters) as well as the model comparison measures by changing the priors on model parameters, results were quite robust overall and did not change any conclusions regarding our best model, the strength of spatial association parameters or inference on the fixed-effects.

6. DISCUSSION

In determining dental caries status, spatial dependence of one tooth surface on another neighboring surface cannot be denied. Our study is motivated to address this issue and to determine the strength and effect of this spatial association. Using randomly chosen data from a caries assessment study, this study examines the effects of spatial association as well as subject-specific fixed-effects in determining caries progression as described by the binary indicator that a tooth surface is either decayed or missing or filled (DMFS). The tooth surfaces are clustered within a tooth, which in turn is clustered within a subject. We develop a random effects autologistic regression model under a Bayesian paradigm where the ‘random effects’ term explains the heterogeneity due to subject-specific effects and the ‘autologistic parameters’ describe the spatial effects induced on one particular type of tooth surface by a neighboring tooth surface in determining caries status. Our model choice mechanism (based on posterior predictive distribution) compares competing models based solely on their predictive performance and selects the autologistic regression model to be a far better candidate than the traditional random effects logistic regression model.

Autologistic regression models has its applications in medical imaging, remote sensing, disease mapping, species-habitat relationships [41]. Our model is unique in a sense that we have considered a multivariate version (incorporating random effects) of the autologistic model along with three different types of spatial dependencies, determined by the three different spatial association parameters explaining first-order, second-order and between-jaw associations. To the best of our knowledge, we do not find any other application of the spatial autologistic model in dental epidemiology. Because of the intractable normalizing constant, we have considered a pseudolikelihood-based

inference under a Bayesian framework, an attractive option to an applied user because it can be readily programmed in freeware like WinBUGS.

In this paper we have included a tooth determined to be missing due to caries as contributing all surfaces (4 or 5 depending on the tooth position) to the binary response. Given the potential inaccuracy of the self-reported cause for missingness, however, missing teeth could alternatively be handled as 'missing data', for which a battery of Bayesian imputation techniques (involving different assumptions, viz. missing completely at random, missing at random, etc.) can be applied [45]. Additionally, subjects with more severe caries may have fewer teeth than subjects with good caries status, a situation that could be handled using informative cluster size models, where the size of the cluster (number of teeth present for each subject) influences the outcome among cluster subunits [50].

Our spatial autologistic model can be extended incorporating temporal patterns when we have DMFS measures for subjects brought in and subjected to randomized treatments and subsequent longitudinal followups. This calls for a more complex 'spatio-temporal' Bayesian estimation scheme under the pseudolikelihood framework with the addition of more parameters and is a consideration for future research. Also easier said than done, we also plan to explore computation schemes to estimate the untractable normalizing constant in our random effects autologistic regression framework of order 3 for a valid 'full-likelihood' inference. However, inference using those techniques like path sampling can be extremely complex in our set-up involving three types of spatial associations along with random effects. Whether using such techniques will lead to a substantial improvement in model predictive performance for our data is a subject of future research. However, drifting away from the pseudolikelihood domain will always be encompassed with computational complexities involving writing one's own 'situation specific' computer code in a relatively low-level language like Fortran or C/C++, which might not always be very attractive for an applied user. Despite its relatively lesser appeal to a full-likelihood minded audience, this research underscores the potential of the pseudolikelihood autologistic framework in analyzing multivariate spatial binary dental data.

ACKNOWLEDGEMENTS

This research was supported in part by NIH/NCRR Grant P20 RR017696-06 and NIH Grant 1R01LM009153. The authors thank the Center for Oral Health Research (COHR) at the Medical University of South Carolina for providing the data and context for this work. In particular, we are grateful to the following COHR personnel: Drs S. London, J. Fernandes, C. Salinas, W. Zhao, Ms L. Summerlin, Ms P. Hudson. The first author would also like to thank Dr Andrew Lawson for interesting discussions. We also thank the Associate Editor and two referees for their helpful comments and suggestions, which have led to an improved presentation of this article. We are grateful to Elsevier for granting us permission to use Figure 1.

REFERENCES

1. Kidd EAM, Smith BGN, Watson TF. *Pickard's Manual of Operative Dentistry* (8th edn). Oxford University Press: Oxford, 2003.
2. Soames JV, Southam JC. *Oral Pathology* (2nd edn). Oxford University Press: Oxford, 1993.
3. Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, Jha P, Mills A, Musgrove P (eds). *Disease Control Priorities in Developing Countries* (2nd edn). Oxford University Press: New York, 2006.
4. Burnside G, Pine CM, Williamson PR. The application of multilevel modelling to dental caries data. *Statistics in Medicine* 2007; **26**:4139–4149.

5. Fernandes J, Slate EH, Wiegand RE, London SD, Grewal JS, Werner P, Sanders JJ, Lopes-Virella M, Salinas CF. Dental caries in Type 2 Gullah diabetics. *Journal of Dental Research* 2007; **86**(Special Issue A):1054. Available from: www.dentalresearch.org.
6. Darby ML, Walsh MM. *Dental Hygiene: Theory and Practice* (1st edn). W. B. Saunders Company: U.S.A., 1995.
7. García-Zattera MJ, Jara A, Lessafre E, Declerck D. Conditional independence of multivariate binary data with an application in caries research. *Computational Statistics and Data Analysis* 2007; **51**:3223–3234.
8. Besag J. Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society, Series B* 1972; **34**:75–83.
9. Besag J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* 1974; **36**:192–236.
10. Zhu J, Huang H-C, Wu J. Modeling spatial-temporal binary data using Markov random fields. *Journal of Agricultural, Biological and Environmental Statistics* 2005; **10**:212–225.
11. Griffith DA. A spatial filtering specification for the autologistic model. *Environment and Planning A* 2004; **36**:1791–1811.
12. Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Mathematical Statistics* 1991; **43**:1–59.
13. Wu H, Huffer FW. Modeling the distribution of plant species using autologistic regression model. *Environmental and Ecological Statistics* 1997; **4**:49–64.
14. Hoeting JA, Leecaster M, Bowden D. An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological and Environmental Statistics* 2000; **5**:102–114.
15. Reich BJ. Neighbor relation modelling and variance component identification in hierarchical areal data models with application to Periodontology, Cancer Epidemiology, and Sports. *Unpublished Ph.D. Dissertation*, University of Minnesota, 2005.
16. Reich BJ, Hodges JS, Carlin BP. Spatial analysis of periodontal data using conditionally autoregressive priors having two types of neighbor relations. *Journal of the American Statistical Association* 2007; **102**:44–55.
17. Besag J, Tantrum J. Likelihood analysis of binary data in space and time. In *Highly Structured Stochastic Systems*, Chapter 9A, Green P, Hort N, Richardson S (eds). Oxford University Press: New York, 2003; 289–295.
18. Guyon X. *Random Fields on a Network*. Springer: Berlin, 1995.
19. Gumpertz ML, Graham JM, Ristiano JB. Autologistic model of spatial pattern of phytophthora epidemic in bell pepper: effects of soil variables on disease presence. *Journal of Agricultural, Biological and Environmental Statistics* 1997; **2**:131–156.
20. Gelfand A, Smith AFM. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* 1990; **85**:398–409.
21. Vehtari A, Lampinen J. Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Computation* 2002; **14**:2439–2468.
22. Gelfand AE, Ghosh SK. Model choice: a minimum posterior predictive loss approach. *Biometrika* 1998; **85**:1–11.
23. Johnson-Spruill I, Hammond P, Davis B, McGee Z, Loudon D. Health of Gullah families in South Carolina with Type-2 diabetes. *The Diabeted Educator* 2009; **35**:117–123.
24. Jiang J. *Linear and Generalized Linear Mixed Models and their Applications*. Springer: New York, 2007.
25. Zhou H, Lawson AB, Hebert JR, Slate EH, Hill EG. A Bayesian hierarchical modeling approach for studying the factors affecting the stage at diagnosis of prostate cancer. *Statistics in Medicine* 2008; **27**:3612–3628.
26. Alfò M, Postiglione P. Semiparametric modelling of spatial binary observations. *Statistical Modelling* 2002; **2**:123–137.
27. Wintle BA. Dealing with uncertainty in wildlife habitat model. *Unpublished Ph.D. Dissertation*, University of Melbourne, 2003.
28. Wintle BA, Bardos DC. Modeling species-habitat relationships with spatially autocorrelated observation data. *Ecological Applications* 2006; **16**:1945–1958.
29. Cressie N. *Statistics for Spatial Data*. Wiley: New York, 1993.
30. Gelman A, Meng XL. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science* 1998; **13**:163–185.
31. Huffer FW, Wu H. Markov Chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics* 1998; **54**:509–524.
32. Berthelsen KK, Møller J. Likelihood and nonparametric-Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics* 2003; **30**:549–564.
33. Zheng Y, Zhu J. Markov Chain Monte Carlo for a spatial-temporal autologistic regression model. *Journal of Computational and Graphical Statistics* 2008; **17**:123–137.

34. Heikkinen J, Hogmander H. Fully Bayesian approach to image restoration with an application in biogeography. *Applied Statistics* 1994; **43**:569–582.
35. Spiegelhalter D, Thomas A, Best N, Lunn D. 'WinBUGS User Manual, Version 1.4.2', MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, 2005. Available from: <http://www.mrc-bsu.cam.ac.uk/bugs>.
36. Gamerman D, Lopes HF. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC: New York, 2006.
37. O'Brien SM, Dunson DB. Bayesian multivariate logistic regression. *Biometrics* 2004; **60**:739–746.
38. Natarajan R, McCulloch CE. Gibbs sampling with diffuse priors: a valid approach to data-driven inference? *Journal of Computational and Graphical Statistics* 1988; **7**:267–277.
39. Dunson DB, Chen Z, Harry J. A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics* 2003; **59**:521–530.
40. Park T, Casella G. The Bayesian Lasso. *Journal of the American Statistical Association* 2008; **103**:681–686.
41. Lawson A. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. CRC Press/Chapman & Hall: New York, 2008.
42. Kaban A. On Bayesian classification with Laplace priors. *Pattern Recognition Letters* 2007; **28**:1271–1282.
43. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2006; **1**:515–533.
44. Gelman A, Carlin JB, Stern H, Rubin D. *Bayesian Data Analysis* (2nd edn). Chapman & Hall/CRC: London, Boca Raton, 2004.
45. Daniels MJ, Hogan JW. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman & Hall/CRC: Boca Raton, FL, 2008.
46. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with Discussion). *Journal of the Royal Statistical Society, Series B* 2002; **64**:583–639.
47. Wang Z, Louis TA. Matching conditional and marginal shapes in binary mixed-effects models using a bridge distribution function. *Biometrika* 2003; **90**:765–775.
48. Paciorek CJ. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Harvard University Biostatistics Working Paper 98*.
49. Reich BJ, Hodges JS, Zadnik V. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* 2006; **62**:1197–1206.
50. Williamson JM, Datta S, Satten GA. Marginal analysis of clustered data when the cluster size is informative. *Biometrics* 2003; **59**:36–42.