

# A BAYESIAN HIERARCHICAL SPATIAL MODEL FOR DENTAL CARIES ASSESSMENT USING NON-GAUSSIAN MARKOV RANDOM FIELDS<sup>1</sup>

BY ICK HOON JIN, YING YUAN AND DIPANKAR BANDYOPADHYAY

*University of Notre Dame, University of Texas and  
Virginia Commonwealth University*

Research in dental caries generates data with two levels of hierarchy: that of a tooth overall and that of the different surfaces of the tooth. The outcomes often exhibit spatial referencing among neighboring teeth and surfaces, that is, the disease status of a tooth or surface might be influenced by the status of a set of proximal teeth/surfaces. Assessments of dental caries (tooth decay) at the tooth level yield binary outcomes indicating the presence/absence of teeth, and trinary outcomes at the surface level indicating healthy, decayed or filled surfaces. The presence of these mixed discrete responses complicates the data analysis under a unified framework. To mitigate complications, we develop a Bayesian two-level hierarchical model under suitable (spatial) Markov random field assumptions that accommodates the natural hierarchy within the mixed responses. At the first level, we utilize an autologistic model to accommodate the spatial dependence for the tooth-level binary outcomes. For the second level and conditioned on a tooth being nonmissing, we utilize a Potts model to accommodate the spatial referencing for the surface-level trinary outcomes. The regression models at both levels were controlled for plausible covariates (risk factors) of caries and remain connected through shared parameters. To tackle the computational challenges in our Bayesian estimation scheme caused due to the doubly-intractable normalizing constant, we employ a double Metropolis–Hastings sampler. We compare and contrast our model performances to the standard nonspatial (naive) model using a small simulation study, and illustrate via an application to a clinical dataset on dental caries.

**1. Introduction.** Dental caries, also known as tooth decay, is one of the most prevalent chronic diseases worldwide [Selwitz, Ismail and Pitts (2007)]. Although preventable, people remain susceptible to the disease throughout their lifetime [Featherstone (2000), Pitts (2004)]; hence it remains a major global oral health burden and is prevalent in the United States. Caries is triggered by acids produced during bacterial fermentation of food debris that accumulates on the tooth surface.

---

Received May 2014; revised November 2015.

<sup>1</sup>Supported in part by the National Institutes of Health (NIH) Grants R01CA154591, R03DE020114 and R03DE021762.

*Key words and phrases.* Autologistic model, Bayesian inference, dental caries, Markov chain Monte Carlo, Potts model, spatial data analysis.

This causes localized dissolution of the tooth's hard tissues and leads to the development of cavities or holes in the teeth [Kidd, Smith and Watson (2003)]. The four main factors influencing the formation of dental caries are the person's age, the health of the tooth surface, the presence of cariogenic bacteria and the presence of fermentable carbohydrates [Soames and Southam (1993)]. The degree of caries progression varies with the individual, depending on the shape of the teeth, oral hygiene habits and the buffering capacity of saliva. If left untreated, caries can spread to supporting tissues and the jaws, and result in advanced conditions that are often painful [Jamison et al. (2006)] and which may lead to tooth loss.

Clinical Studies on dental caries produce clustered multivariate data, consisting of information collected at the tooth surfaces, which are in turn clustered within the oral cavity of each participant [Burnside, Pine and Williamson (2007)]. The surface-level data measure whether they are decayed (D), filled (F) or missing (M) tooth surfaces (S), these are often aggregated over the whole mouth/individual to yield the popular DMFS measure [Darby and Walsh (1995)]. However, the use of this measure provides unrealistic equal weighting to a tooth surface described as D, M or F, even though these indices might represent some latent ordinal levels of caries progression. While the "M" measure is indicative of a tooth-level feature (because a missing tooth contributes to all surfaces as missing for that tooth), the "D" and "F" measures categorize surface-level outcomes. That means the assignment of the "M" component in computing the DMFS measure might lead to possible overestimation or underestimation of the individuals true carious status. In addition, caries status might be spatially referenced [García-Zattera et al. (2007)], that is, a diseased surface (within a particular tooth) might be influenced by the disease status of a set of proximal surfaces, or teeth.

There is a rich body of statistical literature on analyzing spatially referenced discrete data in many disciplines, such as epidemiology, image analysis and environmental studies, starting with the autologistic propositions for binary data [Besag (1974), Hoeting, Leecaster and Bowden (2000), Preisler (1993), Sherman, Apanasovich and Carroll (2006), Wu and Huffer (1997)] and the Potts model for multi-category outcomes [Alfó, Nieddu and Vicari (2009), Green and Richardson (2002), Potts (1952)]. These are based on the popular non-Gaussian Markov random field (MRF) assumptions, where the full conditional distribution for a response (say, a tooth surface in the analysis of dental caries) depends on only a set of prespecified neighbors. Autologistic models, popularized by Besag (1974), are often used to establish the correspondence between a binary response (say, the presence/absence of caries) and the potential explanatory covariates via logistic regression while accounting for spatial dependence via an autoregression. One can also predict the outcome at some unsampled surface, thereby improving the understanding of a spatially referenced binary outcome. There have been some applications of the autologistic specification for modeling caries outcomes in the dental literature [Afroughi et al. (2010), Bandyopadhyay, Reich and Slate

(2009), García-Zattera et al. (2007), Mustvari et al. (2013)]. Other spatial applications in the context of modeling (continuous) periodontal disease responses utilize a parametric conditionally autoregressive (CAR) framework [Reich and Bandyopadhyay (2010), Reich and Hodges (2008), Reich, Hodges and Carlin (2007)] or a nonstationary nonparametric framework [Reich, Bandyopadhyay and Bondell (2013)] mostly under a Bayesian paradigm. The Potts model, a generalization of the Ising model that is popular in statistical mechanics [Winkler (2003)], fits multi-category [Zhu, He and Zhou (2008)] regression models to multiple discrete responses. The Potts model includes the autologistic model as a special case and has been widely applied in image segmentation [Brémaud (1999), Higdon (1998), Johnson and Piert (2009)] and functional magnetic resonance imaging data analysis [Johnson et al. (2012)].

In this paper, we develop a Bayesian hierarchical two-level framework that closely resembles the caries evolution process in humans. At the first level of hierarchy, the binary probability of a tooth being present or absent is modeled via an autologistic model. Conditioned on the tooth being present, we next model the probabilities of a D, F or health (H)/sound surface via a Potts model. At the surface level, we identify various types of spatial association quantified by between-teeth and within-tooth interactions. A Gaussian copula is employed to link the spatial interaction parameters in the autologistic model to those in the Potts model. For illustration, we apply our proposed methodology to a dataset generated from a clinical study on dental caries [Fernandes et al. (2007)]. Our framework handles both individual- and tooth-level (spatial) clustering, and facilitates borrowing strength across all teeth and surfaces for caries risk assessment of important covariables such as age, gender, smoking status, oral brushing/flossing habits and glycemic level.

The remainder of the paper is structured as follows. In Section 2, we describe the aforementioned motivating data on dental caries. In Section 3, we propose our Bayesian hierarchical spatial model utilizing the autologistic and Potts specifications. We present the Bayesian computational framework using the double MH sampler in Section 4. In Section 5, we apply our model to the caries dataset, assess model fit and summarize the inference on fixed effects and spatial associations. In Section 6, we present a small simulation study to study the effect of excluding spatial associations parameters on the regression parameters. We provide our conclusions and future developments in Section 7.

**2. Motivating data.** The motivating data were collected in a clinical study by the Center for Oral Health Research at the Medical University of South Carolina. The study aimed to assess caries status among the Gullah-speaking (or simply Gullah) African Americans who had type 2 diabetes, were 13 years of age or older, and resided in the coastal islands of South Carolina. All study participants underwent an oral examination and answered a detailed questionnaire that focused on

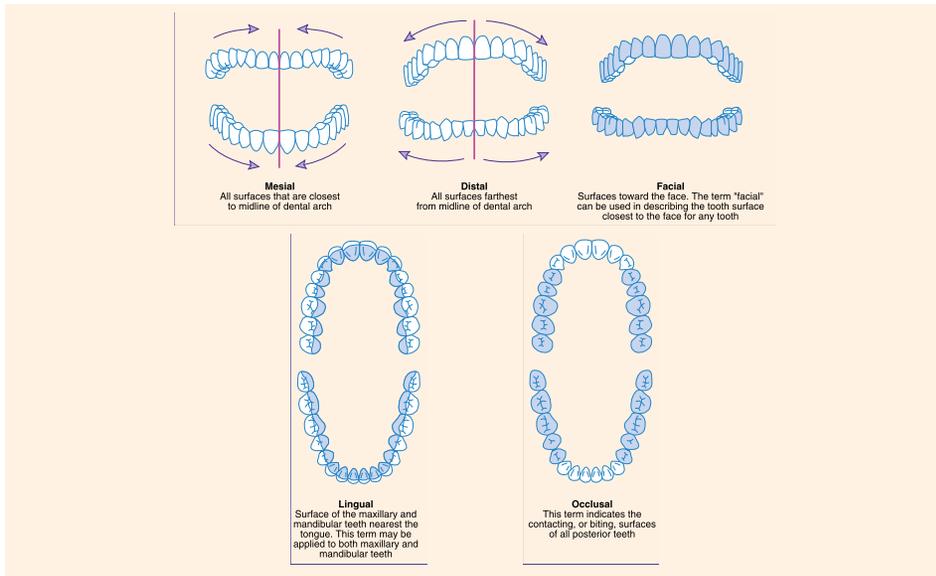


FIG. 1. Permanent dentition within a mouth showing various tooth surfaces. Adapted from Darby and Walsh (1995).

their social, medical and dental histories. There were 281 subjects in the dataset with complete covariate information.

Figure 1 illustrates the different surfaces of permanent dentition within a human mouth. Following Darby and Walsh (1995), the entire dentition can be divided into four quadrants, two on each jaw bone, the mandible (lower jaw) and maxilla (upper jaw). Each quadrant consists of a cluster of 8 teeth: the nonanterior teeth (three molars and two premolars) and the anterior teeth (one incisor and two canine). Furthermore, each nonanterior tooth contributes five surfaces (occlusal, mesial, distal, facial and lingual), while each anterior tooth contributes four of these surfaces (no occlusal surface for these teeth). Caries data has two levels of hierarchy, a tooth level and a surface level, and hence the primary response is recorded according to the level of hierarchy. An assessment of the caries progression at the tooth level yields a binary indicator for the presence or absence of a particular tooth. For that indicator, we denote a nonmissing and missing tooth as N and M, respectively. Note that, for a missing tooth, all its surfaces are considered missing, whereas if the tooth is lost due to a cause other than caries, it does not contribute to the data analysis. The cause of missingness was determined from the questionnaire administered to the study participants, in which the participants distinguished the cause from among the choices of caries, gum disease, orthodontics, injury and other factors. We acknowledge that this self-reported information may be inaccurate, but it is the best information available. Next, conditional on the tooth being present, an

assessment of the caries progression status at the tooth-surface level is a trinary indicator that the surface is either healthy (H), decayed (D) or filled (F).

Several individual-level covariates were also collected, including gender (0 = male, 1 = female), brushing and flossing habits (1 = brushed twice and flossed once every day; 0 = otherwise), smoking status (0 = never, 1 = smoker), age (in years) and glycemic level (determined by HbA1c). Although study recruitment was blind to gender, females participated at a higher rate (77%) than males, and this is reflected in the sample drawn for our analysis. Higher rates of enrollment among females than males are typical for studies among Gullah African Americans. About 15% of the individuals in our sample are smokers and 19% brush twice and floss once every day. The mean age of the participants in the sample is approximately 55 years, with a range of 26–78 years. The mean glycemic level of the individuals in the sample is approximately 7.8, with a range of 5.0–16.4. One primary objective of this study is the risk assessment of these covariates of interest on the caries response, controlling for the effects of spatial association.

### 3. Bayesian hierarchical spatial model for dental caries.

3.1. *Types of spatial associations.* Because caries in a tooth is commonly related to the health of the adjacent/proximal tooth (or surfaces), a spatial structure may be introduced into the models. At the tooth level, we consider only one spatial association, that is, the interaction with neighboring teeth. We denote the corresponding parameter for this association as  $\psi_A \in [0, \infty)$ . At the surface level, three types of spatial associations can be conjectured. These are related to the following:

1. surfaces on the same tooth (type-A association),
2. surfaces on adjacent teeth on the same jaw (type-B association), and
3. contact surfaces on the opposite jaw (type-C association).

For the sake of simplicity and ease of interpretation, we eliminated the associations between non-neighboring surfaces in adjacent teeth on the same jaw. Type-A associations are illustrated in Figure 2(a) and can be divided into three categories, as characterized by three spatial parameters,  $\psi_{P,1}$ ,  $\psi_{P,2}$  and  $\psi_{P,3}$ . Specifically,  $\psi_{P,1}$  denotes associations between the occlusal surface and the other four surfaces on the same tooth, while  $\psi_{P,2}$  denotes associations between adjacent nonocclusal surfaces on the same tooth, that is, between mesial and facial, mesial and lingual, distal and facial, and distal and lingual surfaces.  $\psi_{P,3}$  denotes associations between nonadjacent opposite-site surfaces on the same tooth, that is, between mesial and distal, and facial and lingual surfaces. Type-B associations, as illustrated in Figure 2(b), also consist of two categories characterized by two spatial parameters,  $\psi_{P,4}$  and  $\psi_{P,5}$ . While  $\psi_{P,4}$  denotes associations between the contacting mesial and distal surfaces of adjacent teeth on the same jaw,  $\psi_{P,5}$  quantifies the associations between adjacent occlusal surfaces, facial surfaces and lingual surfaces

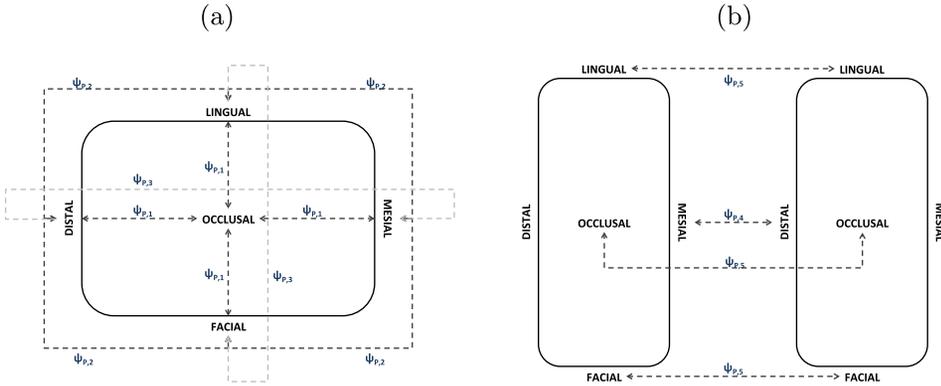


FIG. 2. Illustrations of spatial interactions at the surface level. Panel (a) denotes the type-A (within-tooth) associations, while panel (b) represents the type-B (between-teeth) associations.

of adjacent teeth on the same jaw. Finally,  $\psi_{P,6}$  is the parameter that captures the spatial correlation between the contacting occlusal surfaces on opposite jaws (Type-C association). We denote the vector of the spatial association parameters by  $\psi_P = \{\psi_{P,1}, \dots, \psi_{P,6}\}$ , where  $\psi_{P,1}, \dots, \psi_{P,6} \in [0, \infty)$ . Using these definitions of spatial associations  $\psi_A$  and  $\psi_P$ , we develop a model for the presence/absence of teeth and a model for the status of the tooth surfaces (given that the tooth is present) using autologistic and Potts models, respectively.

3.2. Model for the presence/absence of teeth. Let  $x_{ij}$  be the binary variable indicating whether the  $j$ th tooth of individual  $i$  is missing ( $x_{ij} = 1$ ) or nonmissing ( $x_{ij} = 0$ ), with  $i = 1, \dots, 281$  and  $j = 1, \dots, 32$ . Assuming the autologistic model, the joint distribution of  $\mathbf{x} = \{x_{ij}\}$  is given by

$$\begin{aligned}
 f(\mathbf{x}|\theta_A) &= \frac{1}{\kappa(\theta_A)} \exp \left[ \sum_{(i,j)} x_{ij} \left\{ \alpha_M + \sum_{l=1}^7 \beta_{M,l} Z_{i,l} + R_{A,i} \right\} \right. \\
 (3.1) \quad &\left. + \frac{1}{2} \psi_A \sum_{(i,j)} \sum_{(ij) \sim (ij)'} \{x_{ij} x_{(ij)'} + (1 - x_{ij})(1 - x_{(ij)'})\} \right],
 \end{aligned}$$

where  $\psi_A$  determines the intensity of association between  $x_{ij}$  and its neighbor, represented by  $(ij) \sim (ij)'$ ;  $Z_{i,l}$  is the  $l$ th individual-level covariate with  $l = 1, \dots, 7$  denoting gender, brushing-flossing habit, smoking status, age,  $\text{age}^2$ , HbA1c and  $\text{HbA1c}^2$ ;  $\beta_{M,l}$  measures the effect of covariate  $l$  on the missingness of teeth;  $\alpha_M$  is an intercept; and  $R_{A,i}$  is the random effect that accounts for the within-individual correlation. We assume  $R_{A,i} \sim N(0, 1/\sigma_A^2)$ , where  $N(\cdot)$  denotes a normal distribution. To accommodate potential nonlinear effects of age and HbA1c, we include quadratic terms of age and HbA1c in our model. In the autologistic model,  $\kappa(\theta_A)$ , with  $\theta_A = (\psi_A, \beta_M = \{\beta_{M,l}\}, \alpha_M)$ , is the doubly-intractable normalizing constant

which involves the sum over all possible realizations of  $\mathbf{x}$  [see equation (A.1) in Appendix A]. In order to simplify a complex spatial structure of dental caries, we here assume pairwise-only dependencies and the free boundary condition that all nodes in the external field have a zero value [Lim, Wang and Sherman (2007)].

3.3. *Model for surface-level caries data.* Let  $y_{ijs}$  be the trinary indicator of the surface condition on a nonmissing tooth ( $x_{ij} = 0$ ), representing a healthy ( $y_{ijs} = 1$ ), decayed ( $y_{ijs} = 2$ ) or filled ( $y_{ijs} = 3$ ) surface. Here,  $y_{ijs}$  is observed at the  $s$ th surface of the  $j$ th tooth for individual  $i$ , where the 5 surface types are represented as  $s = 1$ : occlusal;  $s = 2$ : mesial;  $s = 3$ : distal;  $s = 4$ : facial; and  $s = 5$ : lingual. Note that this corresponds to the posterior premolars and molars. For the anterior incisors and canines, there are only four surfaces, corresponding to  $s = 2, \dots, 5$  (see Figure 1). We assume  $\mathbf{y} = \{y_{ijs}\}$ , arising from a multinomial distribution with the following Potts model,

$$\begin{aligned}
 p(\mathbf{y}|\boldsymbol{\theta}_P, x_{ij} = 0) &= \frac{1}{\kappa(\boldsymbol{\theta}_P)} \exp \left[ \sum_{(i,j,s)} I(y_{ijs} = 2) \left\{ \alpha_D + \sum_{l=1}^7 \beta_{D,l} Z_{i,l} + R_{P,i} \right\} \right. \\
 &\quad \left. + \sum_{(i,j,s)} I(y_{ijs} = 3) \left\{ \alpha_F + \sum_{l=1}^7 \beta_{F,l} Z_{i,l} + R_{P,i} \right\} + \sum_{q=1}^6 \psi_{P,q} S_{P,q} \right],
 \end{aligned}
 \tag{3.2}$$

where  $Z_{i,l}$  denotes the individual-level covariates (as described in Section 3.2);  $\beta_{D,l}$  and  $\beta_{F,l}$  measure the effects of covariate  $l$  for the decayed and filled surfaces, with  $\alpha_D$  and  $\alpha_F$  as the intercepts, respectively;  $R_{P,i}$  is the individual-specific random effect following  $N(0, 1/\sigma_P^2)$ ; and  $\psi_{P,q}$ , where  $q = 1, \dots, 6$ , are the six spatial association parameters defined in Section 3.1. With a one-to-one correspondence to  $\psi_{P,q}$ , the six spatial terms  $S_{P,q}$  are defined as follows:

- $S_{P,1} = \sum_{(i,j)} \sum_{s \neq 1} I(y_{ijs} = y_{ij1}) I(x_{ij} = 0)$ , corresponding to  $\psi_{P,1}$ , which represents the associations between the occlusal surface and the other surfaces on the same tooth;
- $S_{P,2} = \sum_{(i,j)} \sum_{s=4,5} \{I(y_{ijs} = y_{ij2}) + I(y_{ijs} = y_{ij3})\} I(x_{ij} = 0)$ , corresponding to  $\psi_{P,2}$ , which represents the associations between adjacent nonocclusal surfaces on the same tooth;
- $S_{P,3} = \sum_{(i,j)} \{I(y_{ij2} = y_{ij3}) + I(y_{ij4} = y_{ij5})\} I(x_{ij} = 0)$ , corresponding to  $\psi_{P,3}$ , which represents the associations between nonadjacent opposite-site surfaces on the same tooth;
- $S_{P,4} = \sum_{(i,j)} \sum_{m \sim j} I(y_{ij2} = y_{im3}) I(x_{ij} = 0) I(x_{im} = 0)$ , where  $m \sim j$  represents  $m = j - 1$  for  $j = 2, \dots, 8, 18, \dots, 24$ ;  $m = j + 1$  with  $j = 9, \dots, 15, 25, \dots, 31$ ; and  $m$  does not exist with  $j = 1, 16, 17, 32$ . This corresponds to  $\psi_{P,4}$ , representing the association between the mesial and distal surfaces of adjacent teeth on the same jaw;

- $S_{P,5} = \sum_{(i,j)} \sum_{m \sim j} \sum_{s=1,4,5} I(y_{ijs} = y_{ims}) I(x_{ij} = 0) I(x_{im} = 0)$ , corresponding to  $\psi_{P,5}$ , which represents the associations between the adjacent occlusal, facial and lingual surfaces of teeth on the same jaw;
- $S_{P,6} = \sum_{(i,j)} \sum_{o \leftrightarrow j} I(y_{io1} = y_{io1}) I(x_{ij} = 0) I(x_{io} = 0)$ , where  $o \leftrightarrow j$  denotes the contacting teeth  $o$  and  $j$  on opposite jaws, corresponding to  $\psi_{P,6}$ , which represents the association between the occlusal surfaces of these teeth.

Similar to the autologistic model, the normalizing constant of the Potts model  $\kappa(\theta_P)$ , where  $\theta_P = (\psi_{P,1}, \dots, \psi_{P,6}, \beta_D = \{\beta_{D,l}\}, \beta_F = \{\beta_{F,l}\}, \alpha_D, \alpha_F)$ , is intractable [see equation (A.2) in Appendix A]. We also assume pairwise-only dependencies and the free boundary condition.

In spatial models, such as the autologistic and Potts models, the spatial parameters are typically the parameters that are most difficult to identify due to limited data information. To alleviate this issue, we link spatial parameters  $\psi_A$  (in the autologistic model) and  $\psi_P$  (in the Potts model) using a Gaussian copula [Nelsen (2006)] to borrow information across the autologistic and Potts models. We first set boundaries  $[\psi_{A,\min}, \psi_{A,\max}]$  and  $[\psi_{P,\min}, \psi_{P,\max}]$  for  $\psi_A$  and  $\psi_P$ , respectively, to map these parameters into  $[0, 1]$ . In our analysis, we choose the lower bound,  $\psi_{A,\min} = \psi_{P,\min} = 0$ , to represent the independent surfaces, and the upper bound,  $\psi_{A,\max} = 1$  and  $\psi_{P,\max} = 3$ , to represent highly dependent surfaces, similarly to the methods of Liang, Liu and Carroll (2007) and Zhang et al. (2010), respectively. Let  $W_A = \Phi^{-1}(\psi_A)$ ,  $W_{P,1} = \Phi^{-1}(\psi_{P,4}/3)$ ,  $W_{P,2} = \Phi^{-1}(\psi_{P,5}/3)$  and  $W_{P,3} = \Phi^{-1}(\psi_{P,6}/3)$ , where  $\Phi^{-1}$  is the inverse cumulative distribution of a standard normal distribution. The Gaussian copula assumes that  $\mathbf{W} = (W_A, W_{P,1}, W_{P,2}, W_{P,3})$  follows a multivariate normal distribution,  $\mathbf{W} \sim N_4(\mathbf{0}, \Sigma)$ , where  $\Sigma$  is a  $4 \times 4$  covariance (or correlation) matrix with the elements on the diagonal equal to 1. As  $\psi_{P,1}$ ,  $\psi_{P,2}$  and  $\psi_{P,3}$  represent the associations between surfaces on the same tooth, these parameters are presumably less likely to be correlated with  $\psi_A$ , which characterizes the spatial association at the tooth level. Therefore, we exclude  $\psi_{P,1}$ ,  $\psi_{P,2}$  and  $\psi_{P,3}$  in the copula model. We assign  $\psi_{P,1}$ ,  $\psi_{P,2}$  and  $\psi_{P,3}$  independent uniform priors with support  $[0, 3]$ . In our data analysis, for simplicity and also to be consistent with the common practice of using the Potts model, we do not impose correlations among the spatial parameters within the Potts model, that is,  $\psi_{P,1}, \dots, \psi_{P,6}$  are pairwise independent a priori. We only specify and estimate the correlations between  $\psi_A$  and each of  $\psi_{P,4}, \dots, \psi_{P,6}$  (i.e., across the autologistic and Potts models). We denote these correlation parameters as  $\rho_1 = \text{corr}(W_A, W_{P,1})$ ,  $\rho_2 = \text{corr}(W_A, W_{P,2})$ ,  $\rho_3 = \text{corr}(W_A, W_{P,3})$  and  $\boldsymbol{\rho} = (\rho_1, \dots, \rho_3)$ , where  $\text{corr}(\cdot)$  denotes the correlation coefficient. In some applications, more complicated, unstructured covariance structures can be entertained in the copula model if desired. In practice, it is possible that the strength of the dependency between the two Markov random fields (but also within them) depends on covariates. One way to accommodate that is to model the dependency parameters as the functions of covariates. However, as data

generally contain limited information on the dependency parameters, incorporating covariates and introducing extra parameters may lead to unstable estimates and identification problems.

**4. Estimation.**

4.1. *Prior specification.* Denote  $\Theta = (\theta_A, \theta_P, \sigma_A^2, \sigma_P^2, \rho)$ . To conduct our Bayesian analysis, we assign noninformative or weakly informative priors to the parameters of the autologistic model and Potts model. A uniform prior on  $\alpha = (\alpha_M, \alpha_D, \alpha_F) \in [-20, 20]$  is assumed. To assign prior distributions for  $\beta = (\beta_M = \{\beta_{M,l}\}, \beta_D = \{\beta_{D,l}\}, \beta_F = \{\beta_{F,l}\})$ , we adopt the weakly informative prior for the logistic regression proposed by Gelman et al. (2008). We assign each component of  $\beta$  an independent Cauchy prior, that is, Cauchy (0, 2.5), after standardizing the covariates to have a mean of 0 and a standard deviation of 0.5. For the variances of random effects  $R_A$  and  $R_P$ , we assume a vague prior  $\sigma_A^2, \sigma_P^2 \sim \text{Uniform}(0, 100)$ . In the copula model, we take  $\rho \sim \text{Uniform}(-1, 1)$ . Letting  $\pi(\Theta)$  denote the prior distribution of  $\Theta$ , the posterior distribution of our Bayesian hierarchical spatial model is given by

$$(4.1) \quad \begin{aligned} &\pi(\Theta|\mathbf{x}, \mathbf{y}) \\ &\propto f(\mathbf{x}|R_A, \theta_A) f(\mathbf{y}|R_P, \theta_P, \mathbf{x}) f(R_A, R_P|\Theta) f(\psi_A, \psi_P|\Theta) \pi(\Theta). \end{aligned}$$

4.2. *Markov chain Monte Carlo (MCMC).* The general MH algorithm cannot be applied to simulate from  $\pi(\Theta|\mathbf{x}, \mathbf{y})$  because the acceptance probability would involve an unknown intractable normalizing constant ratio  $\{\kappa(\theta_A)\kappa(\theta_P)\} / \{\kappa(\theta'_A)\kappa(\theta'_P)\}$ , where  $\theta'_A$  and  $\theta'_P$  denote the proposed values. To circumvent this, we use a double MH (DMH) sampler [Liang (2010)] as an approximate version of the exchange algorithm [Murray, Ghahramani and MacKay (2006)]. By replacing the perfect sampler in the exchange algorithm with MH sampling steps, the DMH sampler is easier to implement and computationally more efficient. One iteration of the DMH sampler can be described as follows:

1. Simulate a new sample  $\Theta'$  from  $\pi(\Theta)$  using the MH algorithm starting with  $\Theta_t$ , where  $t$  denotes the iteration index and  $\Theta_t$  denotes the current state of the Markov chain.
2. Generate an auxiliary variable  $(x', y') \sim P_{\Theta'}^{(m)}(x', y'|x, y)$ , where  $P_{\Theta'}^{(m)}(x, y|x', y')$  denotes the transition probability from the current state  $(x, y)$  to a sample  $(x', y')$ , through  $m$  MH updates, with the detailed balance condition  $r(\Theta_t, \Theta', x', y'|x, y) = \frac{\pi(\Theta_t|x', y') \pi(\Theta'|x, y)}{\pi(\Theta_t|x, y) \pi(\Theta'|x', y')}$ .
3. Set  $\Theta_{t+1} = \Theta'$  if the auxiliary variable is accepted in step (2), and set  $\Theta_{t+1} = \Theta_t$  otherwise.

In our analysis, updating all parameters at one time is not feasible. This is because the parameter space  $\Theta$  consists of 30 components, in addition to the individual-level random effects, leading to slow mixing of the MCMC. Instead, we sequentially update the parameters as follows:

1. Update  $\psi_A$ ,  $\alpha_M$  and  $\beta_M$  simultaneously in the autologistic model;
2. Update  $\sigma_A$  in the autologistic model;
3. Update  $R_{A,i}$  sequentially in the autologistic model;
4. Update  $\psi_P$ ,  $\alpha_D$  and  $\alpha_F$  simultaneously in the Potts model;
5. Update  $\beta_D$  and  $\beta_F$  simultaneously in the Potts model;
6. Update  $\sigma_P$  in the Potts model;
7. Update  $R_{P,i}$  sequentially in the Potts model;
8. Update  $\rho$  in the Gaussian copula model.

The details of the MH ratios in each update are presented in Appendix B.

**5. Application: Dental caries dataset.** In this section, we apply our method to the motivating dataset on caries progression described in Section 2. We employed the DMH sampler to run five independent chains with random starting values. Each run consisted of 60,000 iterations, with 20 cycles of Gibbs updates to generate the auxiliary variables. To assess the convergence of the chains, we used trace plots, autocorrelation plots and the Gelman–Rubin convergence diagnostic  $\hat{R}$ . We discarded the first 10,000 iterations of each run for the burn-in process and collected 5000 samples from the remaining iterations, with a thinning of 10 iterations.

5.1. *Covariate-effect parameters.* Table 1 summarizes the posterior mean and 95% highest posterior density (HPD) interval for the parameter vectors  $(\beta_M, \beta_D, \beta_F)$ , quantifying the effect of various covariates on the carious conditions, that is, Missing, Decayed and Filled surfaces, respectively. Each element of these parameter vectors can be interpreted in terms of the increase/decrease in the log odds of having a missing (M) tooth (for the first level), and decayed (D) or filled (F) surface (with healthy surface as the baseline for the second level) at the same spatial location, when the value of a covariate  $Z_{i,l}$  increases one unit (or changes in category, e.g., from a nonsmoker to smoker, for discrete covariates), conditioned on the other covariates and spatial referencing for that spatial location remaining unchanged.

Covariates corresponding to parameters whose 95% HPD intervals do not include 0 are considered to have substantial effects on the caries outcomes. Intuitively, a proper habit of brushing and flossing leads to reduced odds (log odds =  $-0.325$ ), while smoking increases the odds (log odds =  $0.196$ ) of experiencing a decayed tooth surface. Also, a higher glycemic level leads to increased odds (log odds =  $0.139$ ) of a decayed surface. Glycemic level demonstrates some trend of nonlinear effect on the missing tooth, as the HPD of  $\text{HaA1c}^2$  is largely positive.

TABLE 1  
*Posterior mean estimates and 95% HPD intervals of the covariate-effect parameters obtained after fitting the copula model to the caries dataset*

| Covariates         | Condition | Copula    |                  | Noncopula |                  |
|--------------------|-----------|-----------|------------------|-----------|------------------|
|                    |           | Estimates | 95% HPD          | Estimates | 95% HPD          |
| Gender             | Missing   | 0.004     | (−0.113, 0.123)  | 0.000     | (−0.111, 0.118)  |
|                    | Decayed   | −0.068    | (−0.191, 0.032)  | −0.036    | (−0.159, 0.086)  |
|                    | Filled    | 0.091     | (0.007, 0.175)   | 0.098     | (−0.024, 0.212)  |
| Brush-Floss        | Missing   | −0.079    | (−0.195, 0.028)  | −0.074    | (−0.192, 0.037)  |
|                    | Decayed   | −0.325    | (−0.450, −0.174) | −0.255    | (−0.366, −0.132) |
|                    | Filled    | 0.041     | (−0.042, 0.115)  | 0.060     | (−0.043, 0.156)  |
| Smoking            | Missing   | 0.096     | (−0.038, 0.224)  | 0.090     | (−0.046, 0.223)  |
|                    | Decayed   | 0.196     | (0.061, 0.324)   | 0.225     | (0.099, 0.414)   |
|                    | Filled    | −0.123    | (−0.239, −0.005) | −0.124    | (−0.276, 0.027)  |
| Age                | Missing   | 0.348     | (0.245, 0.458)   | 0.344     | (0.242, 0.451)   |
|                    | Decayed   | 0.001     | (−0.102, 0.106)  | −0.004    | (−0.122, 0.185)  |
|                    | Filled    | 0.025     | (−0.065, 0.103)  | 0.027     | (−0.066, 0.167)  |
| Age <sup>2</sup>   | Missing   | −0.005    | (−0.140, 0.130)  | −0.009    | (−0.153, 0.142)  |
|                    | Decayed   | 0.065     | (−0.098, 0.238)  | 0.034     | (−0.131, 0.200)  |
|                    | Filled    | −0.184    | (−0.378, −0.036) | −0.200    | (−0.351, −0.091) |
| HaA1c              | Missing   | −0.023    | (−0.155, 0.094)  | −0.029    | (−0.149, 0.094)  |
|                    | Decayed   | 0.139     | (0.013, 0.343)   | 0.039     | (−0.068, 0.145)  |
|                    | Filled    | 0.023     | (−0.090, 0.138)  | −0.012    | (−0.099, 0.099)  |
| HaA1c <sup>2</sup> | Missing   | 0.110     | (−0.012, 0.236)  | 0.108     | (−0.014, 0.238)  |
|                    | Decayed   | 0.081     | (−0.077, 0.178)  | 0.167     | (0.024, 0.276)   |
|                    | Filled    | −0.081    | (−0.092, 0.034)  | −0.059    | (−0.172, 0.040)  |

The upper HPD value of the Brush-Floss covariate for a missing tooth is a very small positive number; therefore, we can infer that maintaining good dental hygiene (via brushing and flossing) also leads to reduced odds (log odds = −0.079) of a missing tooth. Similarly, we infer that smoking increases odds (log odds = 0.096) of a missing tooth because the lower HPD value of the missing tooth for smokers is a very small negative number. Older age is related to increased odds of a missing tooth (log odds = 0.348), while reduced odds (log odds = −0.184) are observed for a filled surface for a quadratic term of age. This supports the finding that missing teeth are largely associated with elderly subjects, while more filled tooth surfaces are associated with younger subjects.

In our approach, two spatial models (i.e., the model for the presence/absence of teeth and the model for surface-level caries) are linked through the copula. We also considered an alternative approach that fits the two spatial models independently without imposing the copula structure. We note that although the separate modeling approach seems simpler to fit than the proposed joint modeling approach, in

TABLE 2

Posterior mean estimates and 95% HPD intervals of spatial association parameters, variance components and correlation parameters, obtained after fitting the copula and noncopula models to the caries dataset

| Parameter    | Copula    |                  | Noncopula |                  |
|--------------|-----------|------------------|-----------|------------------|
|              | Estimates | 95% HPD          | Estimates | 95% HPD          |
| $\psi_A$     | 0.377     | (0.352, 0.401)   | 0.378     | (0.353, 0.400)   |
| $\psi_{P,1}$ | 0.156     | (0.127, 0.183)   | 0.153     | (0.119, 0.183)   |
| $\psi_{P,2}$ | 0.905     | (0.837, 0.986)   | 0.897     | (0.846, 0.950)   |
| $\psi_{P,3}$ | 0.984     | (0.857, 1.110)   | 0.995     | (0.888, 1.116)   |
| $\psi_{P,4}$ | 1.113     | (1.004, 1.227)   | 1.106     | (0.985, 1.196)   |
| $\psi_{P,5}$ | 0.647     | (0.595, 0.706)   | 0.647     | (0.587, 0.708)   |
| $\psi_{P,6}$ | 0.444     | (0.309, 0.624)   | 0.486     | (0.317, 0.650)   |
| $\sigma_A^2$ | 0.273     | (0.201, 0.360)   | 0.256     | (0.184, 0.360)   |
| $\sigma_P^2$ | 0.249     | (0.171, 0.379)   | 0.286     | (0.158, 0.441)   |
| $\alpha_M$   | -0.352    | (-0.432, -0.283) | -0.351    | (-0.446, -0.269) |
| $\alpha_D$   | -0.751    | (-0.865, -0.627) | -0.787    | (-0.948, -0.672) |
| $\alpha_F$   | -0.261    | (-0.340, -0.165) | -0.288    | (-0.407, -0.201) |
| $\rho_1$     | 0.988     | (0.960, 0.996)   | -         | -                |
| $\rho_2$     | 0.876     | (0.856, 0.892)   | -         | -                |
| $\rho_3$     | 0.635     | (0.475, 0.744)   | -         | -                |

our case, these two approaches are actually comparable in terms of computational complexity. This is because the majority of computation time and complexity is related to handling a doubly-intractable normalizing constant that appears in both joint and separate models. For comparison purposes, the estimates from the separate model approach are also presented in Tables 1 and 2.

Although the estimates are generally similar between the separate and joint approaches, slight differences are noted in the mean estimates for intercepts, variance parameters for random effects and covariates for decayed and filled surfaces. In addition, the separate model approach yields wider HPD intervals for intercepts, variance parameters for random effects and covariate effects for decayed and filled surfaces.

5.2. *Spatial association parameters.* Table 2 presents the posterior mean and 95% HPD interval of the spatial association parameters and other remaining parameters (e.g., intercepts, variance components and correlations induced via the Gaussian copula) from our full copula model, and compares these to those obtained from the noncopula model. Usually, in these autologistic and Potts specifications, a value of 1.0 for the spatial association parameters ( $\psi$ ) amounts to a very high degree of association [Green and Richardson (2002)]. The estimate of the tooth-level  $\psi_A$  is 0.38, which provides evidence of a moderate level of association. At the surface level, the posterior estimates of  $\psi_{P,1} - \psi_{P,6}$  can be ordered to determine the

strength of associations between various teeth surfaces. The strongest association is observed for  $\psi_{P,4}$ , representing the effect between the mesial and distal surfaces of adjacent teeth on the same jaw. Given the proximity of these surfaces within the same tooth, this was expected and, intuitively, the gaps between teeth (the area between the mesial and distal surfaces of adjacent teeth in the same jaw) might serve as ideal pockets for trapping food during mastication, and can trigger vigorous caries progression. The next one following (in terms of the magnitude of association) is  $\psi_{P,3}$ , representing the associations between nonadjacent opposite surfaces on the same tooth. This is followed by  $\psi_{P,2}$ , that is, between adjacent nonocclusal surfaces on the same tooth. Associations between adjacent occlusal surfaces, facial surfaces and lingual surfaces of teeth on the same jaw (represented by  $\psi_{P,5}$ ) and of contacting occlusal surfaces on opposite jaws (represented by  $\psi_{P,6}$ ) are of moderate strength, while that between the occlusal surface and the other surfaces on the same tooth (represented by  $\psi_{P,1}$ ) is negligible. Furthermore, the estimates of  $\rho_1$ ,  $\rho_2$  and  $\rho_3$  explain the correlation between spatial association parameters  $\psi_A$  (from the first level) and  $\psi_{P,4}$ ,  $\psi_{P,5}$  and  $\psi_{P,6}$ , respectively, from the second level, as specified by the Gaussian copula structure. The strength is maximum for  $\rho_1$ , which might explain the fact that the presence/absence of a tooth is highly related with the association between the mesial and distal surfaces of adjacent teeth on the same jaw. Also, as in Table 1, the corresponding estimates of the spatial association parameters and variance components for the noncopula model are presented in Table 2. Once again, these estimates are very close to the ones from the full (copula) model.

To summarize, the associations with occlusal surfaces tend to be much lower than those with nonocclusal surfaces, that is, the carious state of the occlusal surfaces exerts minimum influence on the carious status of neighboring surfaces. Note that only the nonanterior teeth (the molars and premolars) have occlusal surfaces; therefore, occlusal surfaces are outnumbered by the other types of surfaces.

*5.3. Model assessment.* We assessed the goodness of fit for our model using the posterior predictive checking method [Gelman et al. (2013)]. Using this approach, the first step is to choose a discrepancy measure  $T(\mathbf{y})$ , a function of data  $\mathbf{y}$ , to target a certain model specification. Then we simulate the data replicate  $\mathbf{y}^{\text{rep},k}$ ,  $k = 1, \dots, K$ , from its posterior predictive distribution based on the posited model and observed data. The model assessment is based on the posterior predictive probability of  $T(\mathbf{y})$ , which is estimated as the proportion of simulations for which  $T(\mathbf{y}^{\text{rep},k})$  equals or exceeds its realized value  $T(\mathbf{y})$ . A large (i.e.,  $>0.95$ ) or small (i.e.,  $<0.05$ ) value of the posterior predictive probability signals a lack of fit of the model, that is, the observed value  $T(\mathbf{y})$  is very unlikely under the posited model, whereas a posterior predictive probability close to 0.5 indicates that the model adequately fits the data.

To assess the adequacy of the model we proposed, we chose 9 discrepancy measures that target different components of the autologistic model (3.1),

$$T_1 = \sum_{(i,j)} I(x_{ij} = 1), \quad T_{2,l} = \sum_{(i,j)} z_{i,l} I(x_{ij} = 1), \quad l = 1, \dots, 7,$$

$$T_3 = \sum_{(i,j)} \sum_{(ij) \sim (ij)'} x_{ij} x_{(ij)'} + (1 - x_{ij})(1 - x_{(ij)'}),$$

where  $T_1$  diagnoses the overall (or marginal effect) adequacy of the model, and  $T_{2,l}$  and  $T_3$  diagnose the adequacy of covariate effects and spatial associations, respectively. Along similar lines, for the Potts model (3.2), we examined 22 discrepancy measures

$$\tilde{T}_1 = \sum_{(i,j,s)} I(y_{ijs} = 2), \quad \tilde{T}_2 = \sum_{(i,j,s)} I(y_{ijs} = 3),$$

$$\tilde{T}_{3,l} = \sum_{(i,j,s)} z_{i,l} I(y_{ijs} = 2), \quad l = 1, \dots, 7,$$

$$\tilde{T}_{4,l} = \sum_{(i,j,s)} z_{i,l} I(y_{ijs} = 3), \quad l = 1, \dots, 7,$$

$$\tilde{T}_{5,q} = S_{P,q}, \quad q = 1, \dots, 6,$$

where  $\tilde{T}_1$  and  $\tilde{T}_2$  diagnose the overall (or marginal effects) adequacy of the model,  $\tilde{T}_{3,l}$  and  $\tilde{T}_{4,l}$  diagnose the adequacy of covariate effects, and  $\tilde{T}_{5,q}$  diagnoses the adequacy of spatial associations. We generated 25,000 data replicates using posterior predictive simulations, and calculated the posterior predictive probability for each of the aforementioned 31 discrepancy measures. The resulting posterior predictive probabilities were all between 0.40 and 0.60, suggesting that the proposed model adequately fits the data. Figure 3 shows the posterior predictive distributions of selected discrepancy measures ( $T_3$  and  $\tilde{T}_{5,2}$ ).

**6. Simulation study.** We conducted a small simulation study in this section to explore the finite sample frequentist/classical properties of the covariate-effect parameter estimates, and to quantify the effect of excluding spatial associations on these estimates when the underlying data generation mechanism is spatially referenced. Here, we used a much simpler spatial model (as compared to the model in Section 3) for data generation by reducing the number of spatial associations and covariates. We considered three spatial association parameters  $\psi_A, \psi_{P,2}$  and  $\psi_{P,4}$ , and two covariate-effect parameters  $\beta_2$  and  $\beta_4$ , corresponding respectively to a binary and a continuous covariate. We fixed  $\psi_{P,1} = 0.1$ . We generated the binary covariate from a Bernoulli(0.19) distribution, depicting the original dataset with 19% subjects with proper brushing and flossing habits. We generated the continuous covariate from  $N(0, 0.5)$ . We considered

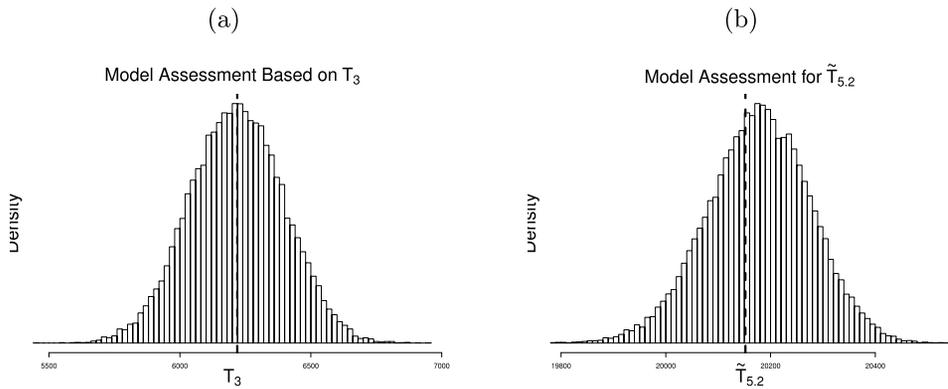


FIG. 3. Posterior predictive distribution of  $T_3$  (panel a) and  $\tilde{T}_{5,2}$  (panel b) based on 25,000 posterior predictive simulations. The vertical lines indicate the observed values of  $T_3$  and  $\tilde{T}_{5,2}$ .

four sets of values for the association parameters:  $(\psi_A, \psi_{P,2}, \psi_{P,4}) = (0.3, 1.0, 0.7)$ ,  $(0.3, 1.5, 1.0)$ ,  $(0.4, 1.0, 0.7)$  and  $(0.4, 1.5, 1.0)$ , representing a sequence of spatial associations increasing from moderate to high levels. We fixed all the other parameter values as follows:  $(\alpha_M, \alpha_D, \alpha_F, \beta_{M,2}, \beta_{M,4}, \beta_{D,2}, \beta_{D,4}, \beta_{F,2}, \beta_{F,4}) = (0.0, -0.2, 0.0, 0.1, 0.15, 0.35, 0.25, -0.15, -0.15, 0.15)$ . For each scenario, we generated 500 datasets. We compared this model to a naive model that ignores the spatial associations, that is, fitting the missing/nonmissing tooth data using the standard logistic regression and fitting the surface-level data using a multinomial regression.

Table 3 summarizes the simulation results, including the empirical bias, root mean squared error (RMSE) and the coverage probability of the 95% HPD interval, based on 500 simulations. Compared to the naive method, the proposed approach has smaller bias and RMSE in all four scenarios, especially for the covariate-effect parameters, including  $\beta_{M,2}$ ,  $\beta_{M,4}$ ,  $\beta_{D,2}$ ,  $\beta_{D,4}$ ,  $\beta_{F,2}$  and  $\beta_{F,4}$ . In addition, under the proposed method, the coverage probabilities of the 95% HPD interval are generally close to the nominal value of 0.95, whereas those of the naive method are substantially lower than the nominal value and often lower than 50% for some parameters (e.g.,  $\beta_{M,4}$ ,  $\beta_{D,2}$  and  $\beta_{F,4}$ ). Based on the simulation results, we conclude that the exclusion of spatial associations can lead to serious issues in the estimation of the covariate-effect parameters.

**7. Discussion.** In this paper, we set out to alleviate some of the methodological issues not previously addressed in the context of modeling dental caries. Specifically, risk assessment of “missing” tooth/surfaces separately from the “decayed” and “filled” tooth surfaces via a two-level model in the context of DMFS/DMFT for dental caries has never been conducted. If we are not mistaken, this is the first paper exploring such a model. In particular, our Bayesian hierarchical model not

TABLE 3  
*Empirical bias, root mean squared error (RMSE) and coverage probability (CP) of 95% HPD interval for parameter estimates under the proposed method and the naive method (which ignores spatial correlations)*

| Setting                                  | Method   |        | $\alpha_M$ | $\beta_{M,2}$ | $\beta_{M,4}$ | $\alpha_D$ | $\alpha_F$ | $\beta_{D,2}$ | $\beta_{D,4}$ | $\beta_{F,2}$ | $\beta_{F,4}$ |
|--|----------|--------|------------|---------------|---------------|------------|------------|---------------|---------------|---------------|---------------|
| $\psi_A = 0.3$                           | Proposed | Bias   | 0.001      | 0.006         | 0.014         | -0.002     | -0.001     | 0.007         | 0.002         | -0.005        | 0.005         |
|  |          | RMSE   | 0.001      | 0.007         | 0.014         | 0.002      | 0.001      | 0.008         | 0.004         | 0.006         | 0.005         |
|  |          | CP (%) | 94.4       | 95.0          | 94.8          | 95.4       | 94.6       | 90.4          | 93.8          | 91.8          | 94.2          |
| $\psi_{P,2} = 1.0$<br>$\psi_{P,3} = 0.7$ | Naive    | Bias   | 0.003      | 0.251         | 0.587         | -0.221     | 0.001      | 0.266         | -0.066        | -0.212        | 0.246         |
|  |          | RMSE   | 0.004      | 0.251         | 0.587         | 0.221      | 0.001      | 0.266         | 0.066         | 0.212         | 0.246         |
|  |          | CP (%) | 74.4       | 49.4          | 0.2           | 0.0        | 77.0       | 36.2          | 76.0          | 51.6          | 17.2          |
| $\psi_A = 0.3$                           | Proposed | Bias   | 0.000      | 0.008         | 0.015         | -0.003     | 0.000      | 0.002         | 0.000         | -0.003        | 0.000         |
|  |          | RMSE   | 0.001      | 0.009         | 0.015         | 0.003      | 0.000      | 0.004         | 0.003         | 0.004         | 0.002         |
|  |          | CP (%) | 96.8       | 94.0          | 92.4          | 93.6       | 95.8       | 90.2          | 91.8          | 93.4          | 94.2          |
| $\psi_{P,2} = 1.5$<br>$\psi_{P,3} = 1.0$ | Naive    | Bias   | 0.000      | 0.251         | 0.587         | -0.406     | 0.003      | 0.463         | -0.078        | -0.391        | 0.468         |
|  |          | RMSE   | 0.002      | 0.251         | 0.587         | 0.406      | 0.004      | 0.463         | 0.078         | 0.391         | 0.468         |
|  |          | CP (%) | 80.8       | 51.8          | 0.2           | 0.0        | 68.0       | 14.2          | 67.2          | 18.0          | 0.8           |
| $\psi_A = 0.4$                           | Proposed | Bias   | 0.001      | 0.004         | 0.016         | -0.002     | 0.000      | 0.002         | 0.006         | -0.005        | 0.002         |
|  |          | RMSE   | 0.001      | 0.005         | 0.016         | 0.002      | 0.000      | 0.004         | 0.007         | 0.006         | 0.003         |
|  |          | CP (%) | 94.8       | 93.6          | 93.8          | 94.4       | 95.8       | 93.4          | 94.0          | 92.8          | 91.2          |
| $\psi_{P,2} = 1.0$<br>$\psi_{P,3} = 0.7$ | Naive    | Bias   | 0.004      | 0.421         | 1.009         | -0.224     | 0.004      | 0.264         | -0.054        | -0.214        | 0.249         |
|  |          | RMSE   | 0.004      | 0.421         | 1.009         | 0.224      | 0.004      | 0.264         | 0.054         | 0.214         | 0.249         |
|  |          | CP (%) | 69.4       | 26.6          | 0.0           | 0.0        | 79.4       | 37.2          | 77.4          | 48.0          | 18.8          |
| $\psi_A = 0.4$                           | Proposed | Bias   | -0.001     | 0.015         | 0.021         | -0.004     | 0.000      | 0.001         | 0.000         | -0.008        | 0.004         |
|  |          | RMSE   | 0.001      | 0.016         | 0.021         | 0.004      | 0.000      | 0.003         | 0.003         | 0.009         | 0.004         |
|  |          | CP (%) | 95.4       | 94.0          | 92.6          | 92.2       | 96.4       | 90.0          | 92.4          | 91.4          | 95.0          |
| $\psi_{P,2} = 1.5$<br>$\psi_{P,3} = 1.0$ | Naive    | Bias   | -0.001     | 0.447         | 1.017         | -0.415     | 0.003      | 0.474         | -0.064        | -0.410        | 0.502         |
|  |          | RMSE   | 0.002      | 0.447         | 1.017         | 0.415      | 0.004      | 0.474         | 0.064         | 0.410         | 0.502         |
|  |          | CP (%) | 69.8       | 21.4          | 0.0           | 0.0        | 67.4       | 15.0          | 69.2          | 16.8          | 0.6           |

only quantifies various kinds of (spatial) between-tooth and within tooth-surfaces associations, but also efficiently explore the response-covariate relationships in caries progression. We tackled the problem of doubly-intractable normalizing constants (a standard byproduct of autologistic or Potts models) by utilizing an efficient double MH sampler.

A multinomial/logistic framework under the umbrella of spatial generalized linear mixed models (SGLMM) can be a useful alternative to our two-level (autologistic and Potts) formulation. We chose to use the autologistic and Potts models because of ease of interpretation for spatial dependence. Unlike the spatial generalized linear mixed model (SGLMM), which uses a latent Gaussian Markov random field to model spatial dependence, the autologistic and Potts models explain dependence directly using a function of the observations, which makes spatial interactions to be easy to interpret. Our approach provides an extension to the seminal work of Besag (1972) for analyzing discrete (spatial) data by considering inference using the intractable full likelihood, and avoiding ad hoc (and often inefficient) pseudo-likelihood approximations.

Our motivating dataset does not contain missing data. To handle missing data, a natural and general approach is to impute the missing data using the multiple imputation method, and then apply the proposed method. Due to the difficulty and computational burden of calculating doubly-intractable normalizing constants in the autologistic and Potts models, in practice, we may conduct the multiple imputation based on a simple imputation model that ignores the spatial interactions, and then apply the proposed models that account for the spatial interactions. In general, handling missing data for the models with doubly-intractable normalizing constants remains a very challenging problem due to the complexity of posterior calculation. It is a topic worthy of further research.

When there are many covariates, a certain variable selection procedure may be needed to obtain a parsimonious model. In general, Bayesian model selection for the models with doubly-intractable normalizing constant is a very challenging problem because of the difficulty of calculating posteriors. A simple and practical approach is to use the traditional forward and backward variable selection procedure based on whether the 90% or 95% credible interval of the parameter estimates includes 0.

Although our modeling approach has been developed for the analysis of dental caries assessment data, it can be readily extended to handle general bivariate spatial data with mixed binary and multinomial outcomes. Our proposal can also be extended into various directions. Longitudinal clinical trials on dental caries, such as the X-ACT trial [Bader et al. (2013)], lead to a spatio-temporal setup in which our methods can be conveniently explored. In this paper, we focused entirely on the efficient risk assessment to understand the complex covariate-response relationships for progression of dental caries. Certainly, prediction (determining responses at future time points) can be an important direction of research, primarily under a spatio-temporal setting, where the objective is to understand whether the D

and F surfaces might progress to more severe carious states (say, M). Our current framework is entirely parametric and can be conveniently applied to the evaluation of other dental caries datasets. Under model misspecification of the parametric framework (say for the Potts model), one can investigate various nonparametric specifications as in Johnson et al. (2012). These suggestions are also viable areas for future research.

APPENDIX A: THE FORM OF NORMALIZING CONSTANTS

The intractable normalizing constant of the autologistic model is

$$(A.1) \quad \kappa(\theta_A) = \exp \left[ \sum_{(i,j)} x_{ij} \left\{ \alpha_M + \sum_{l=1}^7 \beta_{M,l} Z_{i,l} + R_{A,i} \right\} + \frac{1}{2} \psi_A \sum_{(i,j)} \sum_{(ij) \sim (ij)'} \{ x_{ij} x_{(ij)'} + (1 - x_{ij})(1 - x_{(ij)'}) \} \right],$$

where  $\mathcal{X}$  is the sample space of  $x_{ij}$ . The intractable normalizing constant of the Potts model is

$$(A.2) \quad \kappa(\theta_P) = \sum_{\forall y \in \mathcal{Y}} \exp \left[ \sum_{(i,j,s)} I(y_{ijs} = 2) \left\{ \alpha_D + \sum_{l=1}^7 \beta_{D,l} Z_{i,l} + R_{P,i} \right\} + \sum_{(i,j,s)} I(y_{ijs} = 3) \left\{ \alpha_F + \sum_{l=1}^5 \beta_{F,l} Z_{i,l} + R_{P,i} \right\} + \sum_{q=1}^6 \psi_{P,q} S_{P,q} \right],$$

where  $\mathcal{Y}$  is the sample space of  $y_{ijs}$ .

APPENDIX B: MCMC UPDATES

When updating  $\psi_A, \alpha_M, \beta_M$  simultaneously in the autologistic model using the approximate exchange algorithm, the MH ratio can be derived from equation (4.1) as follows:

$$(B.1) \quad r(\theta_A, \theta'_A, \mathbf{x}'|\mathbf{x}) = \frac{\pi(\theta'_A)}{\pi(\theta_A)} \cdot \frac{f(x|R_A, \theta'_A)}{f(x|R_A, \theta_A)} \cdot \frac{f(x'|R_A, \theta_A)}{f(x'|R_A, \theta'_A)} \cdot \frac{f(\psi'_A, \psi_P|\rho)}{f(\psi_A, \psi_P|\rho)}.$$

When updating  $\sigma_A^2$  in the autologistic model, the MH ratio can be derived from equation (4.1) as follows:

$$(B.2) \quad r(\sigma_A^2, \sigma_A'^2 | R_P) = \frac{\pi(R_A | \sigma_A'^2) \pi(\sigma_A'^2)}{\pi(R_A | \sigma_A^2) \pi(\sigma_A^2)}.$$

When updating  $R_{A,i}$  sequentially in the autologistic model, the MH ratio can be derived from equation (4.1) as follows:

$$(B.3) \quad \log \{ r(R'_{A,i}, R_{A,i} | \sigma_A^2) \} = \frac{\sum_{(i,j)} x_{ij} R'_{A,i}}{\sum_{(i,j)} x_{ij} R_{A,i}} \times \frac{\sum_{(i,j)} x'_{ij} R_{A,i}}{\sum_{(i,j)} x'_{ij} R'_{A,i}}.$$

When updating  $\psi_P, \alpha_D, \beta_D, \alpha_F, \beta_F$  simultaneously in the Potts model using the approximate exchange algorithm, the MH ratio can be derived from equation (4.1) as follows:

$$(B.4) \quad r(\theta_P, \theta'_P, \mathbf{y}'|\mathbf{y}) = \frac{\pi(\theta'_P)}{\pi(\theta_P)} \cdot \frac{f(y|R_P, \theta'_P)}{f(y|R_P, \theta_P)} \cdot \frac{f(y'|R_P, \theta_P)}{f(y'|R_P, \theta'_P)} \cdot \frac{f(\psi_A, \psi'_P|\rho)}{f(\psi_A, \psi_P|\rho)}.$$

When updating  $\sigma_P$  in the Potts model, the MH ratio can be derived from equation (4.1) as follows:

$$(B.5) \quad r(\sigma_P^2, \sigma_P'^2|R_P) = \frac{\pi(R_P|\sigma_P'^2)\pi(\sigma_P'^2)}{\pi(R_P|\sigma_P^2)\pi(\sigma_P^2)}.$$

When updating  $R_{P,i}$  sequentially in the Potts model, the MH ratio can be derived from equation (4.1) as follows:

$$(B.6) \quad \log\{r(R'_{P,i}, R_{P,i}|\sigma_P^2)\} = \frac{\sum_{(i,j,s)}\{I(y_{ijs} = 2) + I(y_{ijs} = 3)\}R'_{P,i}}{\sum_{(i,j,s)}\{I(y_{ijs} = 2) + I(y_{ijs} = 3)\}R_{P,i}} \times \frac{\sum_{(i,j,s)}\{I(y'_{ijs} = 2) + I(y'_{ijs} = 3)\}R_{P,i}}{\sum_{(i,j,s)}\{I(y'_{ijs} = 2) + I(y'_{ijs} = 3)\}R'_{P,i}}.$$

The MH ratio for updating  $\rho$  is

$$(B.7) \quad r(\rho|\psi_A, \psi_P) = \frac{\pi(\psi_A, \psi_P|\rho') f(\rho')}{\pi(\psi_A, \psi_P|\rho) f(\rho)}.$$

For the separate model, equation (B.1) and equation (B.4) will be changed as follows:

$$(B.8) \quad r(\theta_A, \theta'_A, \mathbf{x}'|\mathbf{x}) = \frac{\pi(\theta'_A)}{\pi(\theta_A)} \cdot \frac{f(x|R_A, \theta'_A)}{f(x|R_A, \theta_A)} \cdot \frac{f(x'|R_A, \theta_A)}{f(x'|R_A, \theta'_A)},$$

$$r(\theta_P, \theta'_P, \mathbf{y}'|\mathbf{y}) = \frac{\pi(\theta'_P)}{\pi(\theta_P)} \cdot \frac{f(y|R_P, \theta'_P)}{f(y|R_P, \theta_P)} \cdot \frac{f(y'|R_P, \theta_P)}{f(y'|R_P, \theta'_P)}.$$

**Acknowledgements.** The authors thank the Associate Editor and two anonymous referees for their thoughtful comments that led to a much improved presentation. They are grateful to the Center for Oral Health Research at the Medical University of South Carolina for providing the motivating data and context of this work. They also acknowledge various interesting insights from Laurie Barker (CDC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## SUPPLEMENTARY MATERIAL

**Supplement to “A Bayesian hierarchical spatial model for dental caries assessment using non-Gaussian Markov random fields”** (DOI: [10.1214/16-AOAS917SUPP](https://doi.org/10.1214/16-AOAS917SUPP); .zip). We provide the C code and associated instructions for implementing our Bayesian two-level hierarchical model [Jin and Yuan (2016)].

## REFERENCES

- AFROUGHI, S., FAGHIHZADEH, S., KHALEDI, M. J. and MOTLAGH, M. G. (2010). Dental caries analysis in 3–5-years-old children: A spatial modelling. *Arch. Oral Biol.* **55** 374–378.
- ALFÓ, M., NIEDDU, L. and VICARI, D. (2009). Finite mixture models for mapping spatially dependent disease counts. *Biom. J.* **51** 84–97. [MR2667513](#)
- BADER, J. D., VOLLMER, W. M., SHUGARS, D. A., GILBERT, G. H., AMAECHI, B. T., BROWN, J. P., LAWS, R. L., FUNKHOUSER, K. A., MAKHIJA, S. K., RITTER, A. V. et al. (2013). Results from the Xylitol for Adult Caries Trial (X-ACT). *J. Am. Dent. Assoc.* **144** 21–30.
- BANDYOPADHYAY, D., REICH, B. J. and SLATE, E. H. (2009). Bayesian modeling of multivariate spatial binary data with applications to dental caries. *Stat. Med.* **28** 3492–3508. [MR2744378](#)
- BESAG, J. E. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *J. Roy. Statist. Soc. Ser. B* **34** 75–83. [MR0323276](#)
- BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. [MR0373208](#)
- BRÉMAUD, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues. Texts in Applied Mathematics* **31**. Springer, New York. [MR1689633](#)
- BURNSIDE, G., PINE, C. M. and WILLIAMSON, P. R. (2007). The application of multilevel modelling to dental caries data. *Stat. Med.* **26** 4139–4149. [MR2405797](#)
- DARBY, M. L. and WALSH, M. M. (1995). *Dental Hygiene: Theory and Practice*. W. B. Saunders Company.
- FEATHERSTONE, J. D. (2000). The science and practice of caries prevention. *J. Am. Dent. Assoc.* **131** 887–899.
- FERNANDES, J., SLATE, E. H., WIEGAND, R. E., LONDON, S. D., GREWAL, J. S., WERNER, P., SANDERS, J. J., LOPES-VIRELLA, M. and SALINAS, C. F. (2007). Dental caries in type 2 Gullah diabetics. *J. Dent. Res.* **86** 1054.
- GARCÍA-ZATTERA, M. J., JARA, A., LESAFFRE, E. and DECLERCK, D. (2007). Conditional independence of multivariate binary data with an application in caries research. *Comput. Statist. Data Anal.* **51** 3223–3234. [MR2345637](#)
- GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2013). *Bayesian Data Analysis*, 3rd ed. Chapman & Hall, New York, NY.
- GELMAN, A., JAKULIN, A., PITTAU, M. G. and SU, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2** 1360–1383. [MR2655663](#)
- GREEN, P. J. and RICHARDSON, S. (2002). Hidden Markov models and disease mapping. *J. Amer. Statist. Assoc.* **97** 1055–1070. [MR1951259](#)
- HIGDON, D. M. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *J. Amer. Statist. Assoc.* **93** 585–595.
- HOETING, J. A., LEECASTER, M. and BOWDEN, D. (2000). An improved model for spatially correlated binary responses. *J. Agric. Biol. Environ. Stat.* **5** 102–114. [MR1817027](#)
- JAMISON, D. T., BREMAN, J. G., MEASHAM, A. R., ALLEYNE, G., GLAESON, M., EVANS, D. B., JHA, P., MILLS, A. and MUSGROVE, P. (2006). *Disease Control Priorities in Developing Countries*, 2nd ed. Oxford Univ. Press, London.

- JIN, I. H., YUAN, Y. and BANDYOPADHYAY, D. (2016). Supplement to “A Bayesian hierarchical spatial model for dental caries assessment using non-Gaussian Markov random fields.” DOI:10.1214/16-AOAS917SUPP.
- JOHNSON, T. D. and PIERT, M. (2009). A Bayesian analysis of dual autoradiographic images. *Comput. Statist. Data Anal.* **53** 4570–4583. MR2744348
- JOHNSON, T. D., LIU, Z., BARTSCH, A. J. and NICHOLS, T. E. (2012). A Bayesian non-parametric Potts model with application to pre-surgical FMRI data. *Stat. Methods Med. Res.* **22** 364–381. DOI:10.1177/0962280212448970.
- KIDD, E. A. M., SMITH, B. G. N. and WATSON, T. F. (2003). *Pickard’s Manual of Operative Dentistry*, 8th ed. Oxford Univ. Press, London.
- LIANG, F. (2010). A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants. *J. Stat. Comput. Simul.* **80** 1007–1022. MR2742519
- LIANG, F., LIU, C. and CARROLL, R. J. (2007). Stochastic approximation in Monte Carlo computation. *J. Amer. Statist. Assoc.* **102** 305–320. MR2345544
- LIM, J., WANG, X. and SHERMAN, M. (2007). An adjustment for edge effects using an augmented neighborhood model in the spatial auto-logistic model. *Comput. Statist. Data Anal.* **51** 3679–3688. MR2364483
- MURRAY, I., GHAHRAMANI, Z. and MACKAY, D. J. C. (2006). MCMC for doubly-intractable distributions. In *Proceedings of 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*.
- MUSTVARI, T., BANDYOPADHYAY, D., LESAFFRE, E. and DECLERCK, D. (2013). A multilevel model for spatially correlated binary data in the presence of misclassification: An application in oral health research. *Stat. Med.* DOI:10.1002/sim.5944.
- NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. Springer, New York. MR2197664
- PITTS, N. B. (2004). Are we ready to move from operative to non-operative/preventive treatment of dental caries in clinical practice? *Caries Res.* **38** 294–304.
- POTTS, R. B. (1952). Some generalized order-disorder transformations. *Proc. Cambridge Philos. Soc.* **48** 106–109. MR0047571
- PREISLER, H. K. (1993). Modeling spatial patterns of trees attacked by Bark-beetles. *Appl. Stat.* **42** 501–514.
- REICH, B. J. and BANDYOPADHYAY, D. (2010). A latent factor model for spatial data with informative missingness. *Ann. Appl. Stat.* **4** 439–459. MR2758179
- REICH, B. J., BANDYOPADHYAY, D. and BONDELL, H. D. (2013). A nonparametric spatial model for periodontal data with nonrandom missingness. *J. Amer. Statist. Assoc.* **108** 820–831. MR3174665
- REICH, B. J. and HODGES, J. S. (2008). Modeling longitudinal spatial periodontal data: A spatially adaptive model with tools for specifying priors and checking fit. *Biometrics* **64** 790–799. MR2526629
- REICH, B. J., HODGES, J. S. and CARLIN, B. P. (2007). Spatial analyses of periodontal data using conditionally autoregressive priors having two classes of neighbor relations. *J. Amer. Statist. Assoc.* **102** 44–55. MR2345531
- SELWITZ, R. H., ISMAIL, A. I. and PITTS, N. B. (2007). Dental caries. *Lancet* **369** 51–59.
- SHERMAN, M., APANASOVICH, T. V. and CARROLL, R. J. (2006). On estimation in binary autologistic spatial models. *J. Stat. Comput. Simul.* **76** 167–179. MR2223145
- SOAMES, J. V. and SOUTHAM, J. C. (1993). *Oral Pathology*, 2nd ed. Oxford Univ. Press, London.
- WINKLER, G. (2003). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*, 2nd ed. *Applications of Mathematics (New York)* **27**. Springer, Berlin. MR1950762
- WU, H. and HUFFER, F. W. (1997). Modeling the distribution of plant species using the autologistic regression model. *Ecological Statistics* **4** 49–64.

ZHANG, X., JOHNSON, T. D., LITTLE, R. J. A. and CAO, Y. (2010). A Bayesian image analysis of radiation induced changes in tumor vascular permeability. *Bayesian Anal.* **5** 189–212. [MR2596441](#)

ZHU, H., HE, F. and ZHOU, J. (2008). Auto-multicategorical regression model for the distribution of vegetation. *Stat. Interface* **1** 63–73. [MR2425345](#)

I. H. JIN

DEPARTMENT OF APPLIED AND  
COMPUTATIONAL MATHEMATICS AND STATISTICS

UNIVERSITY OF NOTRE DAME  
NOTRE DAME, INDIANA, 46556  
USA

E-MAIL: [ijin@nd.edu](mailto:ijin@nd.edu)

Y. YUAN

DEPARTMENT OF BIostatISTICS  
UNIVERSITY OF TEXAS  
MD ANDERSON CANCER CENTER  
HOUSTON, TEXAS, 77030  
USA

E-MAIL: [yyuan@mdanderson.org](mailto:yyuan@mdanderson.org)

D. BANDYOPADHYAY

DEPARTMENT OF BIostatISTICS  
SCHOOL OF MEDICINE

VIRGINIA COMMONWEALTH UNIVERSITY  
RICHMOND, VIRGINIA, 23298

USA

E-MAIL: [dbandyop@vcu.edu](mailto:dbandyop@vcu.edu)