Modeling Longitudinal Spatial Periodontal Data: A Spatially Adaptive Model with Tools for Specifying Priors and Checking Fit

Brian J. Reich[1],[*] and James S. Hodges[2]

[1]Department of Statistics, North Carolina State University,
2501 Founders Drive, Box 8203, Raleigh, North Carolina 27695, U.S.A.
[2]Division of Biostatistics, School of Public Health, University of Minnesota,
2221 University Avenue SE, Suite 200, Minneapolis, Minnesota 55414, U.S.A.
[*]*email:* reich@stat.ncsu.edu

SUMMARY. Attachment loss (AL), the distance down a tooth's root that is no longer attached to surrounding bone by periodontal ligament, is a common measure of periodontal disease. In this article, we develop a spatiotemporal model to monitor the progression of AL. Our model is an extension of the conditionally autoregressive (CAR) prior, which spatially smooths estimates toward their neighbors. However, because AL often exhibits a burst of large values in space and time, we develop a nonstationary spatiotemporal CAR model that allows the degree of spatial and temporal smoothing to vary in different regions of the mouth. To do this, we assign each AL measurement site its own set of variance parameters and spatially smooth the variances with spatial priors. We propose a heuristic to measure the complexity of the site-specific variances, and use it to select priors that ensure parameters in the model are well identified. In data from a clinical trial, this model improves the fit compared to the usual dynamic CAR model for 90 of 99 patients' AL measurements.

KEY WORDS: Conditional autoregressive prior; Disease monitoring; Nonstationarity; Periodontal data; Spatiotemporal data.

1. Introduction

Periodontal disease is the primary cause of adult tooth loss. It has been estimated that over half of adults over age 35 years are already in the early stages of periodontal disease (Oliver, Brown, and Loe, 1998). A common measure of periodontal disease is attachment loss (AL), the distance down a tooth's root that is no longer attached to surrounding bone by the periodontal ligament. During a full periodontal exam, AL is measured at six locations on each tooth as shown in Figure 1. For a patient with no missing teeth, there are $n_s = 168$ observations. The patient whose data are plotted in Figure 1 is missing the second tooth on the left side of the upper jaw, so there are 162 observations.

Calibration studies commonly show that a single AL measurement has an error with a standard deviation of roughly 0.4 to 1 mm (Osborn et al., 1990, 1992). Figure 1 shows a moderate to severe case of periodontal disease, so measurement error with a 1-mm standard deviation is substantial. In clinical practice, clinicians in effect do a *t*-test at each site to determine if an apparent change from one office visit to the next is real, and commonly a site's measured AL must change by at least 2 mm to be deemed a true change.

The aim of this article is to develop a statistical model to improve periodontal disease monitoring. Several treatments for periodontal disease are available and early detection is important. The methods developed here could be implemented in some form in the increasingly popular software used to record measurements in periodontal offices. This would allow periodontists to identify changes in a patient's condition more quickly and reliably so corrective measures could be taken.

Reich, Hodges, and Carlin (2007) analyzed AL data from a single visit using a conditionally autoregressive (CAR) distribution, popularized for Bayesian disease mapping by Besag, York, and Mollié (1991). This spatial model borrows strength across neighboring sites to smooth away measurement error and improve estimates of true AL at each site. Reich et al. (2007) show that it can be advantageous to have more than one class of neighbor relation in the spatial structure, so the different classes of neighbor relations can induce different degrees of smoothing. For example, they allow smoothing of neighbor pairs bridging the gap between teeth to be different from smoothing of pairs that do not bridge such gaps. In this article, we propose a more flexible spatial model by allowing each spatial location, rather than each neighbor type, to have a different amount of spatial smoothing. Also, we extend this model to analyze repeated periodontal measurements using a spatiotemporal model that smooths values both toward neighbors in space and toward consecutive measurements at the same location.

The nature of periodontal disease progression has been debated for some years. The "linear theory" posits that AL changes gradually at a steady rate (Loe et al., 1978). A competing theory, the "burst theory" (Socransky et al., 1984), contends that a given location alternates between bursts of
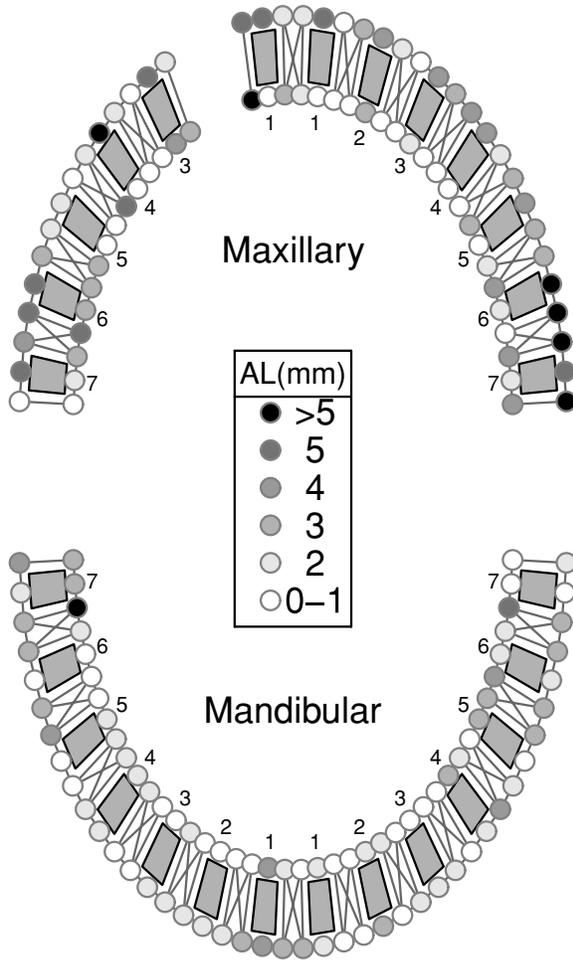
Figure 1. Observed AL for a typical clinical trial subject. The shaded boxes represent teeth, the circles represent measurement sites, and the gray lines represent neighbor pairs connecting adjacent sites on the same tooth and sites that share a gap between teeth. "Maxillary" and "Mandibular" refer to upper and lower jaws, respectively. The small numbers beside each tooth are the "tooth numbers." The maxilla's second tooth on the left is missing; third molars ("wisdom teeth") are excluded.

progression and periods of remission or repair. Gilthorpe et al. (2003) proposed a nonspatial hierarchical model that unified these theories by allowing the time trajectories at individual sites to have a polynomial form, and found evidence that linear trajectories were insufficient for modeling longitudinal AL data.

To accommodate both the linear and the burst theories of disease progression, a spatiotemporal model should allow the amount of spatial and temporal smoothing to vary in different regions of the mouth. Models that allow the amount of variation to change over time are common in time series analysis (Engle, 1982; Kitagawa and Gersh, 1985; Bollerslev, 1986). Several spatial models have been proposed that allow the amount of spatial smoothing to vary spatially. For example, Lawson and Clark (2002) used a mixture of L_1 and

L_2 spatial processes to handle discontinuities. More recently, Brewer and Nolan (2007) allowed the amount of smoothing to vary across the spatial domain by assigning each site its own spatial smoothing parameter. They used an empirical Bayes approach, using results from the usual CAR model to provide prior information about the region-specific smoothing parameters.

Our method follows Brewer and Nolan in assigning each site its own smoothing parameter for the change in AL at each visit. Also, because measurement error for AL data is often not constant throughout the mouth, we assign each site its own error variance as well. Both types of site-specific variances display spatial pattern, so we smooth the area-specific variances using CAR priors. This fully Bayesian model allows parts of the mouth to have different amounts of spatial smoothing.

A highly parameterized model with different variances for each measurement location is susceptible to poor identification and slow Markov chain Monte Carlo (MCMC) convergence. Also, due to the complicated nature of longitudinal periodontal data, it is difficult to visually inspect model output to determine if the model fits well. Therefore, we propose two new methods to accompany this model. First, to guide prior specification and avoid identification problems, we derive a measure of complexity for the site-specific smoothing parameters and use that to choose an intuitive informative CAR prior for the smoothing parameters. Second, we develop a diagnostic to search for features of the data that a given model fits poorly and to suggest elaboration, e.g., additional fixed effects or spatially varying smoothing parameters.

The article proceeds as follows. Section 2 proposes a spatially adaptive CAR model to analyze AL data from a single visit, which allows both the error variance and the smoothing variance to vary spatially. Section 2.4 develops a measure of complexity for the site-specific variances, which is used to choose priors that ensure identification. This model is extended to the spatiotemporal setting in Section 3. The spatiotemporal model has three types of variances: error variance, variances controlling spatial smoothing of the baseline mean, and variances controlling spatial smoothing of the changes between visits. Each type of variance is allowed to vary spatially and is smoothed with a CAR model. These models are illustrated in Sections 4 and 5 using periodontal data from a recent clinical trial. Section 6 concludes.

2. A Spatially Adaptive CAR Model for Cross-Sectional AL Data

This section extends the usual CAR model of Besag et al. (1991) to allow both the error and smoothing variances to vary spatially.

2.1 *The Usual CAR Model*

Let y_s be the observed AL at site s. Assume $y_s = \mu_s + \mathbf{x}'_s \boldsymbol{\beta} + \epsilon_s$, where μ_s is the spatially varying intercept, $\mathbf{x}_s$ is the vector of covariates for site s, $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients, and measurement error $\epsilon_s \sim N(0, \sigma^2)$ independently across s. The covariates used throughout the article are six tooth-number indicators (referring to the tooth numbers in Figure 1) and an indicator that a site is adjacent to the gap between teeth. Based on previous work (Hao,

1998; Roberts, 1999; Zhao, 1999), we do not include interactions (e.g., between tooth and quadrant) and no intercept term is included in $\mathbf{x}_s$ because the intercept is implicit in the CAR model for $\boldsymbol{\mu}$.

Spatial dependence is introduced through the prior (or model) on $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{n_s})'$, which smooths μ_s toward its neighbors defined in Figure 1. The CAR model with L_2 norm (also called a Gaussian Markov random field) can be defined by priors for each μ_s conditional on the value of $\boldsymbol{\mu}$ at all other sites, $\mu_{(s)}$. Under the $\mathrm{CAR}(\tau^2)$ model, $\mu_s \mid \mu_{(s)}$ is normal with mean $\bar{\mu}_s$ and variance τ^2/m_s, where $\bar{\mu}_s$ is the mean of $\boldsymbol{\mu}$ at site s's m_s neighbors and τ^2 is the variance parameter that controls prior smoothing. The resulting joint prior for $\boldsymbol{\mu}$ can be written

$$p(\boldsymbol{\mu} \mid \tau^2) \propto (\tau^2)^{-(n_s-G)/2} \exp\left(-\frac{1}{2}\boldsymbol{\mu}'Q\boldsymbol{\mu}\right)$$

$$\propto (\tau^2)^{-(n_s-G)/2} \exp\left(-\frac{1}{2}\sum_{s \sim j}\frac{(\mu_s - \mu_j)^2}{\tau^2}\right), \quad (1)$$

where G is the number of islands (disconnected groups of regions) in the spatial grid (Hodges, Carlin, and Fan, 2003), $Q_{ss} = m_s/\tau^2$, $Q_{sj} = -I(s \sim j)/\tau^2$, and $I(s \sim j)$ is the binary indicator of whether sites s and j are neighbors. In (1), τ^2 can be loosely interpreted as the variance of the difference between the values of $\boldsymbol{\mu}$ for a pair of neighboring sites.

2.2 *Spatially Varying Error Variances*
One difficulty in modeling AL data is that measurement error is often not constant throughout the mouth (Hao, 1998; Roberts, 1999; Zhao, 1999). For example, measurement error is often higher in the gap between teeth because it is more difficult to make measurements there. Yan (2007) proposed a heteroskedastic CAR model that allows the error variance to vary spatially. To allow for heteroskedasticity in the error variances, we assign each site its own error variance, σ_s^2. The model then becomes

$$y_s = \mu_s + x_s'\boldsymbol{\beta} + \epsilon_s, \text{where } \epsilon_s \sim N\left(0, \sigma_s^2\right). \quad (2)$$

We use a prior on the σ_s^2 that spatially smooths them using a CAR prior that allows regions to have high or low measurement-error variances. Let $\sigma_s^2 = \exp(u_0 + u_s + x_s'\boldsymbol{\alpha}_1)$, where u_0 represents the typical log error variance across the entire spatial domain, u_s is the change of log measurement variance for region s compared to the average measurement variance, $\mathbf{x}_s$ is a vector of explanatory variables that predict the magnitude of measurement error, and $\boldsymbol{\alpha}_1$ contains the corresponding parameters. The average error variance u_0 is given a N(0,10) prior and $\boldsymbol{u} = (u_1, \ldots, u_n)'$ is given a $\mathrm{CAR}(\gamma_u)$ prior, subject to the constraint $\sum_{s=1}^{n_s} u_s = 0$. For notational convenience, we have written the model for σ_s^2 in terms of the same covariates x_s as in (2), although this is not necessary and the model is trivially extended to have different covariates for the error variances.

2.3 *Spatially Varying Smoothing Variances*
To allow spatial smoothing to vary in space, we follow the lead of Brewer and Nolan (2007) and extend $\boldsymbol{\mu}$'s joint prior

(1) to have a different variance for each neighbor pair,

$$p\left(\boldsymbol{\mu} \mid \tau_{sj}^2\right) \propto |Q|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{\mu}'Q\boldsymbol{\mu}\right)$$

$$\propto |Q|^{1/2} \exp\left(-\frac{1}{2}\sum_{s \sim j}\frac{(\mu_s - \mu_j)^2}{\tau_{sj}^2}\right), \quad (3)$$

where $|Q|$ is the product of Q's positive eigenvalues, $Q_{ss} = \sum_{s \sim j} 1/\tau_{sj}^2$ and $Q_{sj} = -I(s \sim j)/\tau_{sj}^2$. Rather than treating the τ_{sj}^2 as separate unknown parameters, each location is assigned its own smoothing variance, τ_s^2. We assume that $\tau_{sj}^2 = \tau_s \tau_j$. (We part ways with Brewer and Nolan at this point; see further comments.) Using multiplicative variances, the conditional prior for μ_s is normal with

$$E\left(\mu_s \mid \mu_{(s)}\right) = \sum_{s \sim j}\left[\frac{\tau_j^{-1}}{\sum_{s \sim k}\tau_k^{-1}}\right]\mu_j,$$

$$\mathrm{Var}\left(\mu_s \mid \mu_{(s)}\right) = \tau_s\left(\sum_{s \sim j}\tau_j^{-1}\right)^{-1}. \quad (4)$$

We denote this spatially adaptive CAR (SACAR) model for $\boldsymbol{\mu}$ as $\boldsymbol{\mu} \sim \mathrm{SACAR}(\tau_1^2, \ldots, \tau_{n_s}^2)$. Under this model, the conditional prior mean of μ_s is a weighted average of the neighboring μ_j, with neighboring sites having small τ_j^2 receiving more weight than those having large τ_j^2. As τ_s^2 increases, site s is effectively removed from the spatial grid: its conditional prior variance increases, reducing the influence of the spatial prior, i.e., of its neighbors, and it is given little weight in the conditional priors of its neighbors.

Brewer and Nolan (2007) used the additive model $\tau_{sj}^2 = (\tau_s^2 + \tau_j^2)/2$. The additive and multiplicative models for the smoothing variances give similar results for the analyses in Sections 4 and 5. Because our multiplicative model gives somewhat more intuitive expressions for the conditional mean and variance of μ_s in (4) and enables the prior distribution heuristic in Section 2.4, we present the results for the multiplicative model.

Brewer and Nolan took an empirical Bayes approach to build priors for the τ_s^2. First, they fit the usual CAR model with constant spatial smoothing described in Section 2.1. The differences between the fitted value at site s and its neighbors were then used to specify a prior for τ_s^2.

We take a fully Bayesian approach to this model and use a spatial prior for the smoothing variances, similar to the spatial prior for the error variances in Section 2.2. Define $\tau_s^2 = \exp(v_0 + v_s + \mathbf{x}_s'\boldsymbol{\alpha}_2)$, where v_0 represents the typical log smoothing variance across the entire spatial domain, v_s is the change of log smoothing variance for region s compared to the average smoothing variance, x_s is a vector of explanatory variables, and $\boldsymbol{\alpha}_2$ are the corresponding parameters. The average log smoothing variance v_0 is given a N(0,10) prior and $\mathbf{v} = (v_1, \ldots, v_n)'$ is given a $\mathrm{CAR}(\gamma_v)$ prior, subject to the constraint $\sum_{s=1}^{n_s} v_s = 0$.

2.4 *Measuring the Complexity of the Smoothing Variances to Specify Their Prior*

Proper priors at each stage of the hierarchy ensure a proper posterior for the spatially adaptive CAR model described above. However, given that the model has far more parameters than observations, identification may be poor without substantial spatial smoothing of the τ_s^2, the smoothing variances. We now develop a heuristic for specifying a prior for γ_v, which controls the smoothness of the smoothing variances τ_s^2. Specifically, we define a rough measure of the complexity of $\mathbf{v}$ based on Hodges and Sargent's (2001) degrees of freedom for hierarchical linear models, and use that to facilitate prior specification of γ_v, the parameter that controls spatial smoothing of $\mathbf{v}$.

From (3), under the multiplicative model for the smoothing variances $\tau_{sj}^2 = \tau_s \tau_j$, each term in (3) has the form of a standard normal density with variate $(\mu_j - \mu_s)/\sqrt{\tau_s \tau_j}$. To roughly measure the complexity of $\mathbf{v}$, we set aside the covariates $\mathbf{x}_s$ and approximate the distribution of $\log[(\mu_j - \mu_s)^2/(\tau_s \tau_j)]$ by a normal distribution, i.e.,

$$\log(\mu_j - \mu_s)^2 = \log(\tau_s) + \log(\tau_j) + e_{js}$$
$$= (\tilde{v}_j + \tilde{v}_s)/2 + e_{js}, \qquad (5)$$

where $\tilde{v}_j = v_0 + v_j$. Because $(\mu_j - \mu_s)/\sqrt{\tau_s \tau_j}$ is approximately standard normal, the error $e_{js} = \log[(\mu_j - \mu_s)^2/(\tau_s \tau_j)]$ is roughly the log of a χ_1^2 random variable, which has variance $\xi_1^2 = 4.98$. To account for the correlation between the terms of (3) and preserve the appropriate total variance $\sum \text{Var}(e_{js}) = (n_s - G)\xi_1^2$, we inflate the approximate variance of e_{js} to $\omega^2 = m\xi_1^2/(n_s - G)$, where m is the total number of neighbor pairs in the spatial grid.

With these simplifying assumptions, the model for L, the m-vector of log-squared differences $\log(\mu_j - \mu_s)^2$, can be written

$$L \sim N\left(Z\tilde{v}, \omega^2 I_m\right) \qquad (6)$$

$$\tilde{v} \sim CAR(\gamma_v), \qquad (7)$$

where Z is the $m \times n$ design matrix implied by (5) and I_m is the $m \times m$ identity matrix. Applying the definition of degrees of freedom in Hodges and Sargent (2001), the degrees of freedom describing the complexity of the smoothing variances for a given γ_v^2 is

$$\rho\left(\gamma_v^2\right) = \text{trace}\left[\left(Z'Z + \frac{\omega^2}{\gamma_v^2}D\right)^{-1}Z'Z\right], \qquad (8)$$

where D is the $n_s \times n_s$ adjacency matrix with $D_{ss} = m_s$ and $D_{sj} = -I(s \sim j)$. The degrees of freedom for $\mathbf{v}$, $\rho(\gamma_v^2)$, is strictly increasing in γ_v^2 and $0 \leq \rho(\gamma_v^2) \leq n_s$ for any $\gamma_v^2 > 0$ and any spatial grid. To ensure that $\mathbf{v}$ is well identified, we give γ_v^2 the prior InvGamma(5.0, 1.0), parameterized to have mode 1/6, so the prior 95% percentile for $\rho(\gamma_v^2)$ is 20.6 out of a possible 168, i.e., $\mathbf{v}$ is forced to be fairly smooth.

3. A Spatially Adaptive Dynamic CAR Model for Longitudinal Periodontal Data

3.1 *Statistical Model*

Section 2 proposed a model for AL data from a single periodontal exam, which smooths AL estimates toward their spatial neighbors. This section extends the model to the spatiotemporal setting. Let $y_{ts} = \mu_{ts} + \mathbf{x}_{ts}'\boldsymbol{\beta} + \epsilon_{ts}$ be the observed AL at site s and visit t, where μ_{ts} is the random intercept at site s and visit t, $\mathbf{x}_{ts}$ is a vector of covariates, $\boldsymbol{\beta}$ is the corresponding vector of regression parameters, and $\epsilon_{ts} \sim N(0, \sigma_s^2)$ is the measurement error. Following Section 2.2, we allow the error variance to vary spatially, i.e., $\sigma_s^2 = \exp(u_0 + u_s + x_s'\boldsymbol{\alpha}_1)$, where u_0 is given a N(0,10) prior and $\mathbf{u} = (u_1, \dots, u_n)'$ is given a CAR(γ_u) prior, subject to the constraint $\sum_{s=1}^{n_s} u_s = 0$. However, we hold the error variance constant across visits because in practice the measurements are usually taken by the same periodontist and because measurement error depends on site-specific factors, e.g., ease of access.

The intercepts for the baseline visit, $\boldsymbol{\mu}_0 = (\mu_{01}, \dots, \mu_{0n_s})'$, are given Section 2.3's spatially adaptive CAR prior. That is, $\boldsymbol{\mu}_0 \sim \text{SACAR}(\tau_1^2, \dots, \tau_{n_s}^2), \tau_s^2 = \exp(v_0 + v_s + x_{0s}'\boldsymbol{\alpha}_2)$, where v_0 is given a N(0,10) prior and $v = (v_1, \dots, v_{n_s})'$ is given a CAR(γ_v) prior, subject to the constraint $\sum_{s=1}^{n_s} v_s = 0$.

The subsequent n_t visits are modeled as a dynamic spatiotemporal process (Gelfand, Banerjee, and Gammerman, 2005),

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \Delta_t, \qquad (9)$$

where $\Delta_t = (\Delta_{t1}, \dots, \Delta_{tn_s})$ is the vector of changes in mean AL between visit $t-1$ and visit t. As with the baseline mean $\boldsymbol{\mu}_0$, each vector of changes Δ_t is given a spatially adaptive CAR prior to borrow strength from neighboring sites. However, because the spatial patterns in AL changes are likely to be different from the spatial pattern in baseline AL, the changes are smoothed with a different set of variances parameters, i.e., $\Delta_t \sim SACAR(\delta_1^2, \dots, \delta_{n_s}^2)$. The smoothing variances for the changes in AL $\delta_1^2, \dots, \delta_{n_s}^2$ are modeled as constant across time and smoothed in the same fashion as the baseline smoothing variances, i.e., $\delta_s^2 = \exp(w_0 + w_s + x_{ts}'\boldsymbol{\alpha}_3)$, where w_0 is given a N(0,10) prior and $w = (w_1, \dots, w_{n_s})'$ is given a CAR(γ_w) prior, subject to the constraint $\sum_{s=1}^{n_s} w_s = 0$. Although this is the same prior as for the baseline smoothing variances, because we have three vectors of observations on Δ_t, it is less restrictive, with prior the 95% percentile for $\rho(\gamma_w^2)$ being 45.5 out of a possible 168.

For a visit other than the first or last visit, the prior for μ_{ts} conditional on every component of $\boldsymbol{\mu}$ except μ_{ts} is normal with

$$E\left(\mu_{ts} \mid \mu_{(ts)}\right) = \sum_{s \sim j}\left(\left[\frac{\delta_j^{-1}}{\sum_{s \sim k}\delta_k^{-1}}\right]\mu_{tj} + \left[\frac{\delta_j^{-1}}{\sum_{s \sim k}\delta_k^{-1}}\right]\right.$$
$$\left. \times \left[(\mu_{t-1,s} - \mu_{t-1,j}) + (\mu_{t+1,s} - \mu_{t+1,j})\right]\right),$$

$$\text{Var}\left(\mu_{ts} \mid \mu_{(ts)}\right) = \frac{1}{2}\left(\sum_{s \sim j}\delta_j^{-1}\right)^{-1}. \qquad (10)$$

The conditional prior mean is the sum of two components. The first component is a weighted average of the μ_{ts}'s spatial neighbors at the same visit, similar to the conditional prior mean for baseline AL in (4). The first component, based only on μ_t, is adjusted by the second component using the differences between the true AL at site s and its spatial neighbors at the previous and subsequent visits. For example, if the true AL at site s is larger than its neighbors at times $t-1$ and $t+1$, the second component of $E(\mu_{ts} \mid \mu_{(ts)})$ will be positive and the prior mean of μ_{ts} will be larger than the weighted average of the true AL of sites s's neighbors at time t.

All told, the fit is controlled by three types of variance parameters: error variance (σ_s^2), baseline spatial smoothing variance (τ_s^2), and the smoothing variance for the change in mean AL (δ_s^2). Each type of variance is smoothed spatially using a CAR prior with neighborhood structure defined as in Figure 1, allowing the nature of the time series to be different in different regions of the mouth. The full model is

$$y_{ts} = x'_{ts}\boldsymbol{\beta} + \mu_{0s} + \sum_{j=1}^{t} \Delta_{js} + \epsilon_{ts},$$

$$\text{where } \sum_{j=1}^{t} \Delta_{js} \text{ is null for } t=0;$$

$$\epsilon_{ts} \sim N\left(0, \sigma_s^2\right), \text{ where } \sigma_s^2 = \exp(u_0 + u_s + x_{0s}\boldsymbol{\alpha}_1) \text{ and}$$

$$u \sim CAR\left(\gamma_u^2\right), \text{ subject to } \sum_s u_s = 0;$$

$$\boldsymbol{\mu}_0 \sim SACAR\left(\tau_1^2, \ldots, \tau_{n_s}^2\right),$$
$$\text{where } \tau_s^2 = \exp(v_0 + v_s + x_{0s}\boldsymbol{\alpha}_2)$$

$$\mathbf{v} \sim CAR\left(\gamma_v^2\right), \text{ subject to } \sum_s v_s = 0;$$

$$\Delta_t \sim SACAR\left(\delta_1^2, \ldots, \delta_{n_s}^2\right),$$
$$\text{where } \delta_s^2 = \exp(w_0 + w_s + x_{ts}\boldsymbol{\alpha}_3)$$

$$w \sim CAR\left(\gamma_w^2\right), \text{ subject to } \sum_s w_s = 0. \tag{11}$$

We evaluate the posterior using MCMC sampling using R (http://www.R-project.org). As far as we know the model cannot be fit in WinBUGS. The "car.normal" distribution does not allow for random weights and the "dmnorm" distribution requires a proper covariance matrix. Gibbs sampling is used to update $\boldsymbol{\mu}_0$, Δ_t, and $\boldsymbol{\beta}$. The elements of $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, and $\boldsymbol{\alpha}_3$ are updated using Metropolis–Hastings sampling, tuned so the acceptance ratio is near 40%. The variances u, v, and w are sampled using blocked-Metropolis updates, with teeth as blocking units. We make 50,000 draws using MCMC sampling and discard the first 10,000 as burn-in. Based on trace plots of the deviance and several individual parameters, the algorithm seems to converge fairly quickly.

When fitting this model and simplifications of it, we compare models using the deviance information criterion (DIC) of Spiegelhalter et al. (2002), defined as $\text{DIC} = \bar{D} + p_D$ where $\bar{D}$ is the posterior mean of the deviance, $p_D = \bar{D} - \hat{D}$ is the effective number of parameters, and $\hat{D}$ is the deviance evaluated at the posterior mean of the parameters in the likelihood. The model's fit is measured by $\bar{D}$, while the model's complexity is captured by p_D. Models with smaller DIC are preferred.

3.2 *A Goodness-of-Fit Diagnostic*

Visually inspecting a fit to determine whether the model captures the data's key features is an important step in model building for time series and spatial data. However, it can be difficult to visually inspect complicated spatiotemporal data such as longitudinal AL data, because of its dimensionality. Therefore, this section proposes a new diagnostic to search for inadequacies in a proposed model.

The spatiotemporal model in (11) can be written

$$y \mid \boldsymbol{\mu}, \boldsymbol{\beta}, \Sigma_1 \sim N\left(\boldsymbol{\mu} + X\boldsymbol{\beta}, \Sigma_1\right) \text{ and } \boldsymbol{\mu} \,\big|\, \Sigma_2^{-1} \sim N\left(0, \Sigma_2\right),$$

where $y = (y'_0, \ldots, y'_{n_t})'$ is the vector of observed AL for all sites and visits, $\boldsymbol{\mu} = (\boldsymbol{\mu}'_0, \ldots, \boldsymbol{\mu}'_{n_t})'$, and Σ_1^{-1} and Σ_2^{-1} are precision (inverse covariance) matrices. The conditional precision matrix of y, Σ_1^{-1}, is diagonal with diagonal elements $1/\sigma_s^2$. The prior precision matrix of $\boldsymbol{\mu}$ is $\Sigma_2^{-1} = Z^{-1'}\tilde{Q}Z^{-1}$, where Z is defined by $\boldsymbol{\mu} = Z(\boldsymbol{\mu}'_0, \Delta'_1, \ldots, \Delta'_{n_t})'$ and $\tilde{Q}$ is the block diagonal prior precision matrix of $(\mu'_0, \Delta'_1, \ldots, \Delta'_{n_t})'$ which depends on τ_s and δ_s.

To judge whether the spatiotemporal covariance implied by this model fits the data well, we compute the marginal distribution of y after integrating out the random effects $\boldsymbol{\mu}$. The marginal distribution of y is normal with mean $X\boldsymbol{\beta}$ and variance

$$\Sigma_y = \left(\Sigma_1^{-1} - \Sigma_1^{-1}\left[\Sigma_1^{-1} + \Sigma_2^{-1}\right]\Sigma_1^{-1}\right)^{-1}, \tag{12}$$

where Σ_y depends on $(\sigma_s, \tau_s, \delta_s)$. We can evaluate the model's fit by comparing Σ_y evaluated at the posterior medians of $(\sigma_s, \tau_s, \delta_s)$, denoted $\hat{\Sigma}_y$, with the residuals $\hat{r} = y - X\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the posterior median of $\boldsymbol{\beta}$.

It is common in multivariate analysis to analyze a covariance matrix by transforming to orthogonal coordinates. Let $z = (z_1, \ldots, z_{n_s n_t})' = \hat{\Sigma}_y^{-1/2}\hat{r}$. If the model is correct, the z_j^2 approximately follow independent χ_1^2 distributions. A qq-plot of the z_j^2 against the quantiles of the χ_1^2 distribution can be used to check for deviations from the spatiotemporal model. Large z_j^2 result from linear combinations of the data with more variation than can be explained by the spatiotemporal covariance structure. Outlying z_j^2 may be caused by any of several deficiencies, including missing fixed effects, nonstationarity, or misspecification of the spatial or temporal association structure. The sources of the deficiencies can be examined by plotting the weights of the linear combinations of $\hat{r}$ corresponding to the outlying z_j^2 (i.e., the rows of $\hat{\Sigma}_y^{-1/2}$). Although this method of searching for outlying linear combinations of the data is not exhaustive, we find it a useful tool.

4. Analysis of Cross-Sectional Periodontal Data

Using the model of Section 2, we first analyze one subject's baseline AL data, displayed in Figure 2 (the baseline observations are plotted using a "1"). Six tooth number indicators and an indicator of a site adjacent to the gap between teeth are used as x_s, i.e., as predictors of the AL and as predictors for any spatially varying variance parameters. The usual CAR model has DIC = 463 and $p_D = 31.6$. Its fitted values are smooth throughout the mouth, resulting in a poor fit in areas with large AL such as the left side of the lower jaw (mandible).
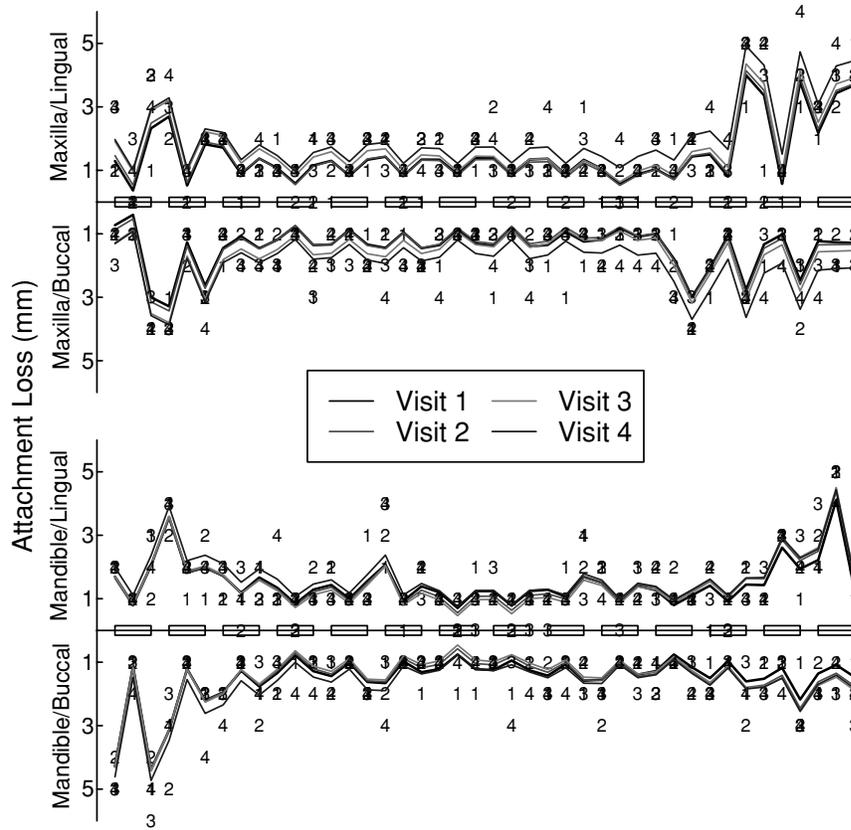
Figure 2. Observed AL and posterior mean AL for one patient over four visits. The numerals are observed AL with the character representing the visit number. The colored lines are the fitted values (posterior means) under the model with spatially varying baseline smoothing parameters, and the rectangles along the horizontal axes represent teeth. "Maxillary" and "mandibular" refer to upper and lower jaws, respectively, while "buccal" and "lingual" refer to the cheek and the tongue sides of the teeth, respectively. This figure appears in color in the electronic version of the journal.

Allowing the error variances to vary in different areas of the mouth reduces the DIC from 463 to 456. Allowing the smoothing variances to vary in different areas of the mouth gives a larger reduction in DIC from 463 to 436. Figure 2 shows that there is more volatility in the back of the mouth than the front of the mouth. The models with spatially varying smoothing parameters are preferred because they allow for differential smoothing in the back and front of the mouth. Among models with spatially varying smoothing parameters, the model for the error variances has little effect on DIC.

5. Analysis of Longitudinal Periodontal Data

This section uses Section 3's model to analyze AL for the patient in Section 4 over the course of four visits 3 months apart. The whole-mouth average AL increases over time: the average AL is 1.55, 1.57, 1.66, and 1.92 for visits 1, 2, 3, and 4, respectively. The data are plotted in Figure 2 along with posterior means from the "best" model described further below. In Figure 2, the first two rows represent the upper jaw. The first row shows the results for the strip of sites along the tongue (lingual) side of the upper jaw.

DIC is smaller for the full model with all three spatially varying variances (DIC = 1415) described in Section 3 than

for the model with the three variances constant across space (DIC = 1457). The effective number of parameters is also smaller for the full model ($p_D = 148.0$) than for the model with the three variances constant across space ($p_D = 158.6$). This may be due to the differential smoothing of the spatiotemporal random effects, shown in Figure 3. Figure 3b shows more spatial variability in the mean baseline AL in the back of the mouth, while the spatial random effects are smoothed considerably in the front of the mouth. The error variances (Figure 3a) and spatial smoothing of the change in AL from one visit to the next (Figure 3c) are fairly constant throughout the mouth, except that the error variances are generally larger for sites in the gap between teeth.

The fixed effects for mean AL under the spatially adaptive CAR model are given in Table 1a. Mean AL is larger for sites in the gap between teeth than for direct sites. This observation has been made previously (Shievitz, 1997; Roberts, 1999; Reich, Hodges, and Zadnik, 2006). Although the plot of observed AL in Figure 2 suggests that AL is larger in the back of the mouth (left and right extremes of Figure 2), the 95% intervals for all the tooth number effects cover zero (Table 1a). This is probably because the spatially varying intercepts μ_{0s} absorb some of the trend from the front to the back of the mouth. That is, this spatial trend can be explained

(a) Error standard deviations, σ_s.

(b) Baseline smoothing standard deviations, τ_s.

(c) Change smoothing standard deviations, δ_s.
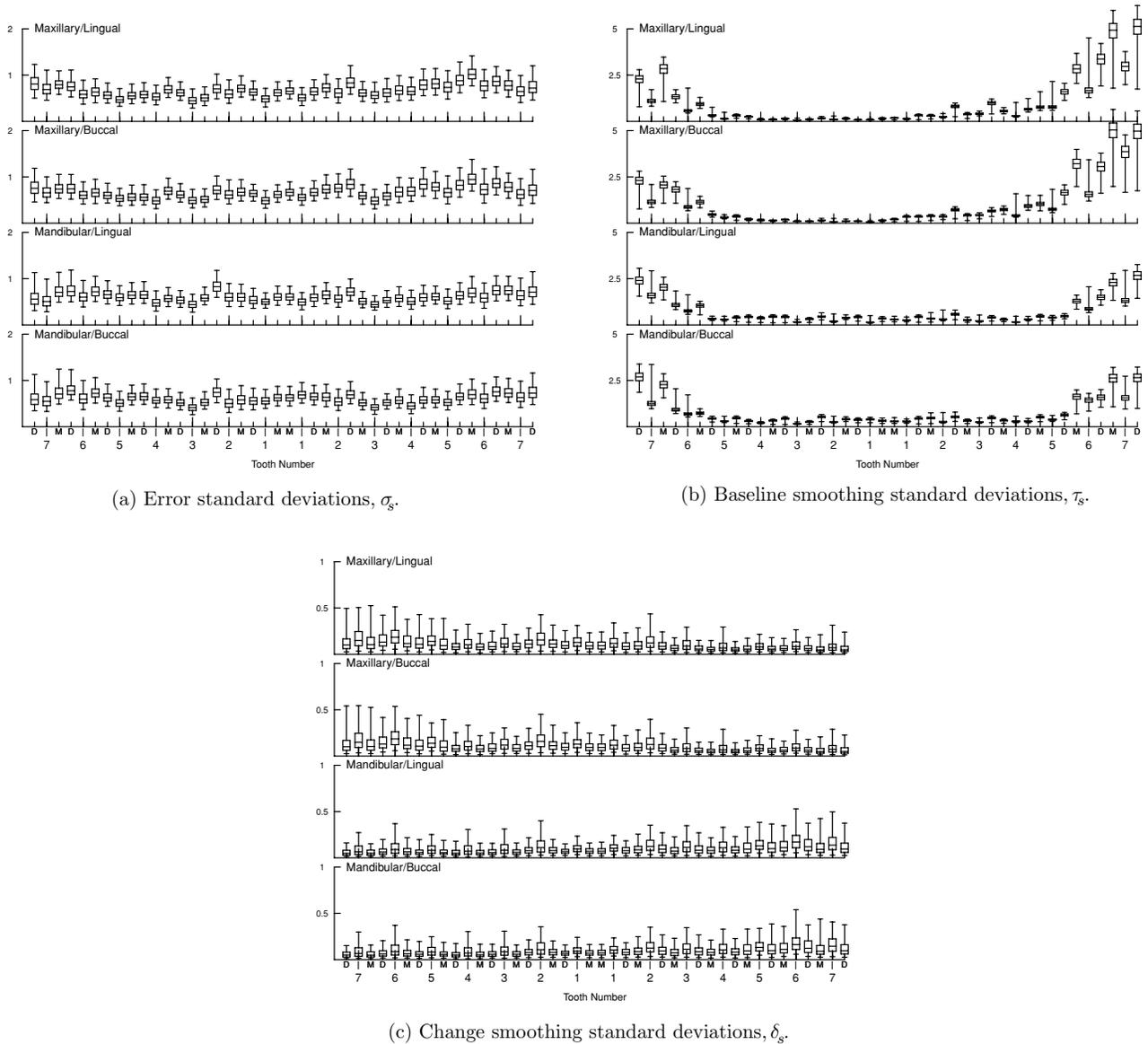
Figure 3. Boxplots with whiskers representing 95% intervals of the posteriors of the three variance components. "Maxillary" and "mandibular" refer to upper and lower jaws, respectively, while "buccal" and "lingual" refer to the cheek and the tongue sides of the teeth, respectively.

either by the tooth number effects or by the spatial random effects, and this uncertainty reduces the tooth number effects and increases their posterior standard deviations (for further discussion, see Reich et al., 2006).

The posteriors of the variances suggest some simplifications. First, based on Table 1 and Figure 3c, we assume that spatial smoothing of the change in AL from one visit to the next is constant throughout the mouth (i.e., $\delta_s^2 \equiv \delta^2$). Also, for the error variances, we remove the CAR random effects and tooth number fixed effects, so the error variance depends only on whether a site is in a gap between teeth. This reduced model has a smaller DIC and p_D (DIC = 1375, $p_D = 82.9$) than the full model described above (DIC = 1415, $p_D = 148.0$).

For the model with all three types of variance parameters held constant throughout the mouth, several of the z_j^2 lie substantially above the 45-degree line in the qq-plot. The row of $\hat{\Sigma}_y^{-1/2}$ for the largest z_j^2 is constant across the four visits and only nonzero for the back left tooth of the lower jaw. This provides further evidence that the model with constant baseline smoothing is not appropriate because it provides an especially poor fit around the back left tooth of the lower jaw.

For the model with spatially varying baseline smoothing variances, the weights for the largest single z_j^2 are a function of the changes in AL between the third and fourth visits around the second tooth from the right on the upper jaw. Some of the measured increases between visits 3 and 4 around this tooth are quite large (Figure 2). However, the qq-plot of the z_j^2 is

Table 1

Summary of the fixed effects for the full model in Section 5. Table (a) gives the posterior medians and 95% intervals for the mean AL regression parameters ($\boldsymbol{\beta}$). Table (b) gives the posterior medians and 95% intervals for the exponentials of parameters for the error variances ($\exp(\boldsymbol{\alpha}_1)$), the baseline smoothing variances ($\exp(\boldsymbol{\alpha}_2)$), and the change in AL smoothing variances ($\exp(\boldsymbol{\alpha}_3)$).

(a)

	Mean AL
Gap	0.59 (0.43, 0.73)
Tooth 2	0.05 (−0.24, 0.36)
Tooth 3	−0.18 (−0.58, 0.21)
Tooth 4	−0.12 (−0.57, 0.31)
Tooth 5	−0.11 (−0.60, 0.40)
Tooth 6	0.23 (−0.40, 0.83)
Tooth 7	−0.01 (−0.94, 0.85)

(b)

	Error variance	Baseline smoothing	Change smoothing
Gap	1.42 (1.10, 1.81)	1.14 (0.25, 4.09)	0.89 (0.20, 3.91)
Tooth 2	1.10 (0.74, 1.78)	1.40 (0.41, 4.17)	1.25 (0.09, 7.47)
Tooth 3	0.65 (0.43, 1.02)	0.45 (0.13, 1.63)	0.95 (0.21, 4.53)
Tooth 4	0.74 (0.48, 1.11)	0.56 (0.10, 2.94)	1.02 (0.25, 3.56)
Tooth 5	0.68 (0.41, 1.05)	0.56 (0.12, 2.34)	1.20 (0.21, 4.82)
Tooth 6	0.71 (0.42, 1.14)	5.45 (1.82, 16.51)	0.99 (0.14, 4.02)
Tooth 7	0.70 (0.40, 1.10)	8.18 (1.87, 28.72)	0.92 (0.19, 3.98)

nearly linear, suggesting further modifications of the model may not be necessary.

Figure 2 shows the fitted values—posterior means for true AL—for the model with spatially varying baseline smoothing variances. For nearly all sites, the first visit (black) has the smallest fitted value and the final visit (blue) has the largest fitted value. However, the increase does not appear to be linear in time or consistent across space. In the lower right quadrant the mean AL increases between the first and second visits, but is then fairly constant. Between the second and third visits, the mean AL increases in the upper left quadrant, and between the third and fourth visits the mean AL increases in the upper right and lower left quadrants.

Figure 4a shows the posterior probability under the reduced model of an increase from baseline in true AL for each visit, that is, the posterior probability of $\mu_{ts} > \mu_{0s}$. To simulate how disease monitoring would go in practice, Figure 4 uses only data from the first t visits to compute the probability of an increase from baseline at visit t. The probability of an increase from baseline is less than 0.90 throughout the mouth for the second visit. The probabilities do not change substantially between the second and third visits except for an increase in the upper left quadrant. In this region, the probability of an increase is greater than 0.99 for sites in the back of the mouth. For the final visit, the probability of an increase from baseline is greater than 0.50 throughout the mouth and greater than 0.90 in most sites in the back of the mouth.

A periodontist may also be interested in the probability of an increase of more than a certain amount, which is also
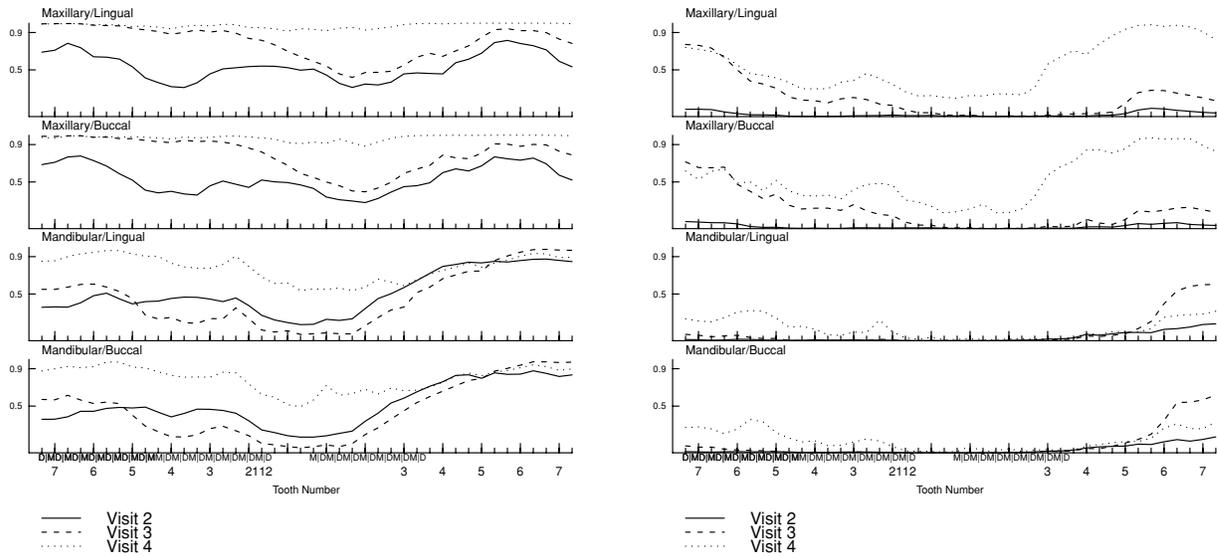
(a) Probability of an increase from baseline.

(b) Probability of an increase from baseline of more than 0.5 mm.

Figure 4. Posterior probability of an increase from baseline (panel (a)) and the probability of an increase of at least 0.5 mm (panel (b)). "Maxillary" and "mandibular" refer to upper and lower jaws, respectively, while "buccal" and "lingual" refer to the cheek and the tongue sides of the teeth, respectively.

immediately available from the MCMC output. For example, Figure 4b maps the posterior probability of an increase from baseline of more than 0.5 mm. Although there is large probability of an increase throughout the mouth, only around teeth 6 and 7 in the upper right quadrant for visit 4 is there substantial probability of an increase of at least 0.5 mm.

In addition to analyzing this particular subject, we applied the spatially adaptive CAR model to all 99 patients in this clinical trial with fewer than 20% missing teeth. Only the first and fourth visits were used in this analysis and only the baseline smoothing parameters in the SACAR model were allowed to vary spatially. We fit the usual CAR model and spatially adaptive CAR model separately for each patient and recorded DIC for each model. DIC is smaller for the spatially adaptive CAR model for 90 of the 99 patients; for 71 of the 90, DIC was smaller by at least 10 for the spatially adaptive CAR.

These 99 subjects may also be used to specify informative priors for future analyses. For example, we analyzed the baseline AL of each patient using the model of Section 2 with constant error variance, spatially varying smoothing variances, and informative InvGamma(5,1) priors for γ_v^2. For each patient we compute the posterior median of γ_v^2. The median of the 99 posterior medians for γ_v^2 was 2.58 (IQR = 0.30). A prior that captures this spread could be an alternative to the informative inverse gamma prior based on degrees of freedom in future analyses.

6. Discussion

This article presented a nonstationary spatiotemporal model for longitudinal periodontal data that allows the variance components to be different in different regions of the mouth. For the patient analyzed in detail in Section 5, it did not help to allow spatial variation in the error or smoothing variances for the change in AL between visits. However, the model with spatially varying baseline smoothing produced a smaller DIC than the usual dynamic CAR model with constant variances. This model also improved the fit for the vast majority of patients from a recent study. The final model using four visits' data can be run in an hour or so on an ordinary PC. Therefore, given the computing power needed to run the imaging software currently used in some dental offices, it should be possible to implement a real-time version of the present analysis.

Given the many variance parameters in the model, Section 2.4 proposed a measure of complexity for spatially smoothed variance parameters and proposed a way to use this measure to place informative priors on the site-specific variances. This led to a well-identified model that was still flexible enough to allow considerably less spatial smoothing in volatile areas of the mouth. Also, Section 3.2's z_j^2 diagnostic revealed inadequacies in the usual CAR model and suggested improvements.

Modeling AL as Gaussian is the norm in periodontal research. This may seem inappropriate for integer-valued outcomes. In the late 1980s, new periodontal probes ("Florida probes") were introduced, which recorded measurements electronically to one decimal place. However, reproducibility studies (e.g., Osborn et al., 1990, 1992) showed that the extra dec-

imal place did not reduce the standard error of individual AL. Therefore, we assume normality, although models for ordered categorical data are also possible.

More sophisticated spatiotemporal models are also possible. For example, the change variances δ_s^2 could vary with space *and* time and be given a second spatiotemporal prior. This model is intuitively appealing because it may avoid oversmoothing bursts in time, just as the spatially adaptive CAR model is designed to prevent oversmoothing bursts in space. Future work might examine prior specifications that impose enough constraint to allow identification, without going so far as to fix the variances across time as we have. Also, periodontal disease monitoring may be improved by simultaneously analyzing other measures of periodontal disease, such as pocket depth and bleeding on probing. This suggests a multivariate version of our spatially adaptive dynamic CAR model, perhaps using a multivariate conditionally autoregressive model (Gelfand and Vounatsou, 2003; Jin, Carlin, and Banerjee, 2005).

References

Besag, J., York, J. C., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43,** 1–59.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31,** 307–327.

Brewer, M. J. and Nolan, A. J. (2007). Variable smoothing in Bayesian intrinsic autoregressions. *Envirometrics* **18,** 841–857.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of variance of UK inflation. *Econometrica* **50,** 987–1007.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4,** 11–25.

Gelfand, A. E., Banerjee, S., and Gammerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics* **16,** 465–479.

Gilthorpe, M. S., Zamzuri, A. T., Griffiths, G. S., Maddick, I. H., Eaton, K. A., and Johnson, N. W. (2003). Unification of the "burst" and "linear" theories of periodontal disease progression: A multilevel manifestation of the same phenomenon. *Journal of Dental Research* **82,** 200–205.

Hao, W. (1998). Examining mean, variance, and correlation structure in a large periodontal dataset. Master of Science Plan B Paper, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota.

Hodges, J. S. and Sargent, D. J. (2001). Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika* **88,** 367–379.

Hodges, J. S., Carlin, B. P., and Fan, Q. (2003). On the precision of the conditionally autoregressive prior in spatial models. *Biometrics* **59,** 317–322.

Jin, X., Carlin, B. P., and Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61,** 950–961.

Kitagawa, G. and Gersch, W. (1985). A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series. *IEEE Transactions on Automatic Control* **30,** 48–56.

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

Loe, H., Anerud, A., Boysen, H., and Smith, M. (1978). The natural history of periodontal disease in man. The rate of periodontal disease destruction before 40 years of age. *Journal of Periodontal Research* **64,** 713–718.

Oliver, R. C., Brown, L. J., and Loe, H. (1998). Periodontal diseases in the United States population. *Journal of Periodontology* **69,** 269–278.

Osborn, J., Stoltenberg, J., Huso, B., Aeppli, D., and Pihlstrom, B. (1990). Comparison of measurement variability using a standard and constant force periodontal probe. *Journal of Periodontology* **61,** 497–503.

Osborn, J. B., Stoltenberg, J. L., Huso, B. A., Aeppli, D. M., and Pihlstrom, B. L. (1992). Comparison of measurement variability in subjects with moderate periodontitis using a conventional and constant force periodontal probe. *Journal of Periodontology* **63,** 283–289.

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* **62,** 1197–1206.

Reich, B. J., Hodges, J. S., and Carlin, B. P. (2007). Spatial analyses of periodontal data using conditionally autoregressive priors having two types of neighbor relations. *Journal of the American Statistical Association* **102,** 44–55.

Roberts, T. (1999). Examining mean structure, covariance structure and correlation of interproximal sites in a large periodontal data set. Master of Science Plan B Paper, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota.

Shievitz, P. (1997). The effect of a non-steroidal anti-inflammatory drug on periodontal clinical parameters after scaling. Master of Science Thesis, School of Dentistry, University of Minnesota, Minneapolis, Minnesota.

Socransky, S. S., Haffajee, A. D., Goodson, J. M., and Lindhe, J. (1984). New concepts of destructive periodontal disease. *Journal of Clinical Periodontology* **11,** 21–32.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion and rejoinder). *Journal of the Royal Statistical Society, Series B* **64,** 583–639.

Yan, J. (2007). Spatial stochastic volatility for lattice data. *Journal of Agricultural, Biological, and Environmental Statistics* **12,** 25–40.

Zhao, Y. (1999). Examining mean structure, correlation structure, and differences between examiners in a large periodontal dataset. Master of Science Plan B Paper, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota.