Spatial Analyses of Periodontal Data Using Conditionally Autoregressive Priors Having Two Classes of Neighbor Relations

By BRIAN J. REICH[a], JAMES S. HODGES[b] and BRADLEY P. CARLIN[b] [1]

[a] *Department of Statistics, North Carolina State University, 2501 Founders Drive, Box 8203, Raleigh, NC 27695, U.S.A.*

[b] *Division of Biostatistics, School of Public Health, University of Minnesota, 2221 University Ave SE, Suite 200, Minneapolis, MN 55414, U.S.A.*

reich@stat.ncsu.edu hodges@ccbr.umn.edu brad@biostat.umn.edu

Correspondence Author: Brian J. Reich
Telephone: (919) 513-7686
Fax: (919) 515-7591

April 25, 2006

[1] The third author was supported in part by NIH grants 5–R01–ES07750-07 and 1–R01–CA95955–01. The authors thank Prof. David Madigan of Rutgers University for showing us the no-free-terms structure.

Spatial Analyses of Periodontal Data Using Conditionally Autoregressive Priors Having Two Classes of Neighbor Relations

Abstract

Attachment loss, the extent of a tooth's root (in millimeters) that is no longer attached to surrounding bone by periodontal ligament, is often used to measure the current state of a patient's periodontal disease and monitor disease progression. Attachment loss data can be analyzed using a conditionally autoregressive (CAR) prior distribution, which smooths fitted values toward neighboring values. However, it may be desirable 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, we may wish to allow smoothing of neighbor pairs bridging the gap between teeth to be different from smoothing of pairs that do not bridge such gaps. Adequately modelling the spatial structure may improve monitoring of periodontal disease progression. This paper develops a two-neighbor-relation CAR model to handle this situation and presents associated theory to help explain the sometimes unusual posterior distributions of the parameters controlling the different types of smoothing. The posterior of these smoothing parameters often has long upper tails, and its shape can change dramatically depending on the spatial structure. Like previous authors, we show that the prior distribution on these parameters has little effect on the posterior of the fixed effects, but has a marked influence on the posterior of both the random effects and the smoothing parameters. Our analysis of attachment loss data also suggests that the spatial structure itself varies between individuals.

Key Words: Conditional autoregressive prior; Gaussian Markov random field; identification; neighbor relations; periodontal data.

1 Introduction

In periodontics, attachment loss (AL), the extent of a tooth's root (in millimeters) that is no longer attached to surrounding bone by periodontal ligament, is used to assess the cumulative damage to a patient's periodontium and to see if treatment stops disease progression. Many texts (e.g., Darby and Walsh 1995) describe periodontal measurement. Figure 1 shows AL for a particular patient (Patient 1 in our study). One patient's mouth has up to 168 measurements (six per tooth) with at least one "island" (disconnected group of regions) per jaw; missing teeth can create more islands. The patient shown in Figure 1 is missing tooth number 2 on the left side of the maxilla (upper jaw), resulting in three islands.

This paper presents the first analyses of periodontal data using spatial statistical methods. The data are from a clinical trial conducted at the University of Minnesota's Dental School comparing three active treatments to placebo and to no treatment. The 50 patients presented here were at least 35 years old, had moderate to severe periodontal disease, and were not undergoing endodontic or surgical periodontal therapy. Each patient was examined once at baseline and four times after administration of treatment, at three month intervals. The original analysis used whole-mouth averages of clinical measures (including attachment loss) or averages of subsets of sites defined by baseline disease status. These standard, non-spatial analyses found no treatment effect (Shievitz 1997).

A natural spatial model for analyzing attachment loss is the conditionally autoregressive (CAR) model, popularized for Bayesian disease mapping by Besag et al. (1991). In an exam with n measurement sites, assume $\boldsymbol{x}_i'\boldsymbol{b} + \theta_i$ is the true attachment loss at site i, $i = 1, ..., n$, where $\boldsymbol{x}_i$ is a column vector containing covariates having coefficients $\boldsymbol{b}$, and θ_i captures spatial variation in true attachment loss that is not explained by the covariates. We include seven population-level

covariates: six tooth number indicators (Figure 1 defines the tooth numbers) and an indicator for direct sites (sites not in a gap between two teeth). Let y_i be site i's observed attachment loss, and assume the likelihood $y_i|\theta_i, \boldsymbol{b}, \tau_0$ is normal with mean $\boldsymbol{x}_i\boldsymbol{b} + \theta_i$ and precision τ_0, conditionally independent across i. The spatial structure governing the θ_i is described by a lattice of neighbor relations among sites. A CAR model for $\boldsymbol{\theta}$ with L_2 norm (also called a Gaussian Markov random field model) has improper density

$$p(\boldsymbol{\theta}|\tau) \propto \tau^{(n-G)/2} \exp\left(-\frac{\tau}{2}\boldsymbol{\theta}' Q \boldsymbol{\theta}\right), \tag{1}$$

where the parameter $\tau \geq 0$ controls smoothing induced by this prior, larger values smoothing more than smaller; G is the number of islands in the spatial structure; $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$; and Q is $n \times n$ with non-diagonal entries $q_{ij} = -1$ if i and j are neighbors and 0 otherwise, and diagonal entries q_{ii} equal to the number of region i's neighbors. This is an n-variate normal kernel specified by its precision matrix τQ instead of its covariance.

Attachment loss measurements are spatially correlated, but their correlation may not be simply a function of distance. Figure 2 identifies four types of neighbor pairs, labelled I-IV. The four neighbor types may have different correlations as suggested by previous studies (e.g., Sterne et al., 1988; Gunsolley et al., 1994; Roberts, 1999) and by the empirical correlations in Table 1a for the 50 subjects analyzed here. Thus, modelling these data may require two or more classes of neighbor relations in the spatial structure, with the lth class having its own τ_l so the different neighbor-relation classes can induce different degrees of smoothing. Besag and Higdon (1999) introduced CAR priors with two classes of neighbor relations, modelling a rectangular grid of plots with different smoothing parameters for row and column neighbors. They also extended the model to adjust for edge effects and to analyze data from multiple experiments.

Table 1: Empirical correlations of each type of neighbor pair and the neighbor pairs controlled by each smoothing parameter for each grid. Figure 2 defines the types. "Number of pairs" is the number of pairs of each type of neighbor relation for the 50 patients combined. "Empirical correlation" is the correlation of each type of neighbor pair using the residuals from the regression (non-spatial) of the 50 patients' AL onto population-level covariates.

(a) Empirical correlations

	Type I	Type II	Type III	Type IV
Number of pairs	5233	2341	2345	2618
Empirical correlation	0.47	0.57	0.52	0.60

(b) Neighbor pairs controlled by each smoothing parameter for each grid

	Grid	Type I	Type II	Type III	Type IV
One class of neighbor pairs	1NR	τ_1	τ_1	τ_1	τ_1
Sides vs interproximal	A	τ_1	τ_1	τ_2	τ_2
Interproximal vs direct only	B	τ_1	τ_2	τ_2	τ_2
Type II vs others	C	τ_2	τ_1	τ_2	τ_2

As possible models for attachment loss, we consider four neighborhood structures ("grids") with one or two classes of neighbor relations, defined in Table 1b. The first grid (1NR) allows only one class of neighbor relations and has just one smoothing parameter, τ_1. Grid A distinguishes neighbor pairs entirely on either the buccal (cheek) or lingual (tongue) sides of the teeth (types I and II) from other neighbor pairs. Grid B distinguishes neighbors bridging the gap between teeth, the "interproximal region" (types II, III, and IV) from type I neighbor pairs. Finally, Grid C distinguishes type II neighbor pairs from the other types.

Spatial analysis of periodontal data can potentially serve several purposes. In research, it can be desirable to take periodontal measurements at only a subset of sites. For example, the National Health and Nutrition Examination Survey III (NHANES III) measured only two sites per tooth on a randomly-selected half-mouth (Drury et al., 1996). Different spatial structures may imply different sampling schemes. Also, different spatial structures are consistent with different etiologies of attachment loss. Compared to the 1NR model, grids A and B imply a special role for interproximal regions; compared to each other, they imply different effects for different interproximal

sites. Clinically, measurement error is relatively large. Calibration studies commonly show that a single attachment loss measurement has an error with a standard deviation of roughly 0.4 to 1 mm (Osborn et. al, 1990; Osborn et. al, 1992). Figure 1 shows a severe case of periodontal disease, so measurement error with a 1 mm standard deviation is substantial. Practitioners in effect do t-tests at each site to determine if an apparent change is real, and commonly a site's measured attachment loss must change by at least 2 mm to be deemed a true change. It should be possible to exploit the spatial correlation of attachment loss measurements to mitigate the effects of measurement error and improve sensitivity.

Section 2 develops a spatial model for periodontal data using a CAR prior with two neighbor relations (2NRCAR). Our analysis was initially hampered by technical problems, such as MCMC autocorrelations near one for the precision parameters, requiring us to more carefully consider identification in these models. Sections 3 and 4 derive the marginal posterior density of (z_1, z_2), for $z_l = log(\tau_l/\tau_0)$, $l = 1, 2$, and examine identification of the z_l. Section 5 then applies Section 2's model our 50-patient data set. While the spatial structure of these AL data appears to vary considerably among patients, grids with two neighbor relations are superior to the 1NR grid for some patients. The choice of grid has noteworthy effects on fitted values and the posterior of the fixed effects. Section 6 considers the effect of (z_1, z_2)'s prior, and Section 7 concludes. Technical results are relegated to the Appendix.

2 A spatial model for periodontal data

Let $\boldsymbol{y}_p$, the n_p-vector of patient p's observed AL, follow a normal distribution with mean $X_p\boldsymbol{b}_p + \boldsymbol{\theta}_p$ and precision $\tau_{0_p} I_{n_p}$, where X_p is a $n_p \times k_p$ matrix of known covariates, $\boldsymbol{b}_p$ is a k_p-vector of fixed effects, and $\boldsymbol{\theta}_p$ is patient p's n_p-vector of spatial random effects, $p = 1, ..., N$. Correlation between a

patient's AL at contiguous sites is modelled in the prior for the spatial random effects $\boldsymbol{\theta}_p$. Although it may be possible to specify a model with spatial correlation in measurement error and in the true mean attachment loss, it would be difficult for the data to differentiate between these competing sources of spatial correlation. We have resolved this by assuming all the spatial correlation is accounted for by $\boldsymbol{\theta}_p$'s prior. $\boldsymbol{\theta}_p$ has a CAR prior with two neighbor relations, written, extending (1), as

$$p(\boldsymbol{\theta}_p|\tau_{1_p}, \tau_{2_p}) \propto c(\tau_{1_p}, \tau_{2_p})^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{\theta}_p'\{\tau_{1_p}Q_{1_p} + \tau_{2_p}Q_{2_p}\}\boldsymbol{\theta}_p\right), \tag{2}$$

where Q_{l_p} and τ_{l_p} respectively describe and control smoothing of class l neighbor pairs for patient p, and $c(\tau_{1_p}, \tau_{2_p})$ is the product of the positive eigenvalues of $\tau_{1_p}Q_{1_p} + \tau_{2_p}Q_{2_p}$ (see below). We assume a pair of regions are neighbors of at most one type. Q_{l_p} has rank $n_p - G_{l_p}$, G_{l_p} being the number of islands in neighbor class l's spatial structure for patient p; assume $G_{l_p} < n_p$, i.e., the lth neighborhood structure is not null. If G_p is the number of islands in patient p's combined spatial structure, $G_{l_p} \geq G_p$. We assume all patients have the same grid, e.g., Grid A, (although patients may have different adjacency matrices due to missing teeth), but it is possible to consider models where the grid can vary between subjects.

Appendix A.1 derives the following results. For any Q_{1_p} and Q_{2_p}, there is a nonsingular B_p such that $Q_{1_p} = B_p'D_{1_p}B_p$ and $Q_{2_p} = B_p'D_{2_p}B_p$, where D_{l_p} is diagonal with $n_p - G_{l_p}$ positive diagonal entries and G_{l_p} zero entries (Newcomb, 1961). Call D_{l_p}'s diagonal elements $d_{l_{pj}}$ and without loss of generality assume the last G_p diagonal elements of both D_{l_p} are zero. Then the product of $\tau_{1_p}Q_{1_p} + \tau_{2_p}Q_{2_p}$'s positive eigenvalues is proportional to

$$c(\tau_{1_p}, \tau_{2_p}) = \prod_{j=1}^{n_p - G_p} \left(\tau_{1_p}d_{1_{pj}} + \tau_{2_p}d_{2_{pj}}\right).$$

We assume the fixed effects are shared across patients, i.e., $\boldsymbol{b}_p \equiv \boldsymbol{b}$ and $k_p \equiv k$, to capture known

patterns in attachment loss. There are $k = 7$ fixed effects: six tooth number indicators (with Tooth 1 serving as the reference) and an indicator for direct sites (sites not in the gap between teeth). Previous studies have shown these to be the only substantial and consistent fixed effects (e.g., Sterne et al., 1988; Gunsolley et al., 1994; Shievitz, 1997; Roberts, 1999). Since the rows and columns of Q_{1_p} and Q_{2_p} sum to zero, the two-neighbor relation CAR model necessarily implies a flat prior on $\boldsymbol{\theta}_p$'s average on each island. To ensure $\boldsymbol{b}$'s identifiability, we do not include a column for the intercept in X_p, so that the intercept is implicit in $\boldsymbol{\theta}_p$.

The precision parameters $\{\tau_{0_p}, \tau_{1_p}, \tau_{1_p}\}$ are allowed to vary between patients. The transformation from $\{\tau_{0_p}, \tau_{1_p}, \tau_{1_p}\}$ to $\{z_{0_p} = log(\tau_{0_p}),\ z_{1_p} = log(\tau_{1_p}/\tau_{0_p}),\ z_{2_p} = log(\tau_{2_p}/\tau_{0_p})\}$ allows for Gaussian priors which more naturally capture vague prior information and allow the $\{z_{l_p}\}$ to be correlated *a priori*. Under this parameterization, z_{0_p} sets the scale and z_{l_p} controls the amount of smoothing of class l neighbors for patient p. Our analysis in Section 5 considers two priors for the $(z_{0_p}, z_{1_p}, z_{2_p})$: (1) the patient's z_{l_p} are independent *a priori* with $z_{l_p} \sim Uniform(-10, 10)$, $l \in \{0, 1, 2\}, s = 1, ..., 50$; and (2) the patient's $(z_{0_p}, z_{1_p}, z_{2_p})$ are drawn from (and thus smoothed by) the Gaussian prior $(z_{0_p}, z_{1_p}, z_{2_p})' \sim N(\mu, \Sigma)$ where $\mu = (\mu_0, \mu_1, \mu_2)'$ and Σ is a diagonal precision matrix with diagonal elements (η_0, η_1, η_2). While it may be reasonable to expect a patient with large z_{1_p} will also have large z_{2_p}, in the absence of any pre-existing data to this effect, we prefer to let z_{1_p} and z_{2_p} be independent *a priori* and let the data induce any posterior correlation. In many cases, the data overcome this prior independence; the posterior correlation of z_{1_p} and z_{2_p} is often very high. To complete the hierarchical model, the μ_l and η_l are given independent $\mathrm{N}(m_l, p_l)$ and $\mathrm{Gamma}(a_l, b_l)$ priors respectively, $l \in \{0, 1, 2\}$. Assuming the Gaussian prior on the patients' precisions, the full posterior, $p(\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2|\boldsymbol{y})$, is

$$\prod_{p=1}^{N} \exp\left(\frac{n_p z_{0_p}}{2} - \frac{e^{z_{0_p}}}{2}\{(\boldsymbol{y}_p - X_p\boldsymbol{b} - \boldsymbol{\theta}_p)'(\boldsymbol{y}_p - X_p\boldsymbol{b} - \boldsymbol{\theta}_p)\}\right) \tag{3}$$

$$\times \prod_{p=1}^{N} \left[\prod_{j=1}^{n_p - G_p} (e^{z_{1_p} + z_{0_p}} d_{1_{pj}} + e^{z_{2_p} + z_{0_p}} d_{2_{pj}})^{1/2} \exp\left(-\frac{e^{z_{0_p}}}{2} \{ \boldsymbol{\theta}_p' (e^{z_{1_p}} Q_{1_p} + e^{z_{2_p}} Q_{2_p}) \boldsymbol{\theta}_p \} \right) \right]$$

$$\times \left[\prod_{p=1}^{N} \eta_l^{1/2} \exp\left(-\frac{\eta_l}{2} (z_{l_p} - \mu_l)^2 \right) \right] \times \left[\prod_{l \in \{0,1,2\}} \eta_l^{a_l - 1} \exp(-\frac{p_l}{2} (\mu_l - m_l)^2 - b_l \eta_l) \right]$$

where $\boldsymbol{y} = (\boldsymbol{y}_1', ..., \boldsymbol{y}_N')'$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_N')'$, and $z_l = (z_{l_1}, ..., z_{l_N})$, $l \in \{0, 1, 2\}$. Section 5's analysis assumes $m_l = 0$, $p_l = 0.001$, and $a_l = b_l = 0.01$, $l \in \{0, 1, 2\}$.

3 Exploring identification of (τ_1, τ_2) by inspecting $p(\tau_1, \tau_2 | \boldsymbol{\theta})$

Our analysis of AL data was initially hampered by poor identification of the smoothing parameters $\{z_{1_p}, z_{2_p}\}$, whose posteriors often have long tails and MCMC draws with autocorrelations near one. Identification of the smoothing parameters warrants further consideration because, as our data analysis shows, they play a key role in estimating attachment loss. This section explores (τ_1, τ_2)'s identification through their conditional posterior p$(\tau_1, \tau_2 | \boldsymbol{\theta})$, provides sufficient conditions that ensure both are τ_1 and τ_2 are identified, and gives characteristics of spatial grids that lead to well-identified smoothing parameters. Section 4 explores (z_1, z_2)'s full marginal posterior using one patient's data. To simplify notation, both sections drop the subscripts that identifies patients.

The conditional density of (τ_1, τ_2) can be re-expressed to highlight identification issues. As in Section 2, assume a non-singular B such that $Q_1 = B' D_1 B$ and $Q_2 = B' D_2 B$, D_1 and D_2 being diagonal. For $\boldsymbol{\theta}^* = B\boldsymbol{\theta}$,

$$p(\tau_1, \tau_2 | \boldsymbol{\theta}^*) \propto \prod_{j=1}^{n-G} \left[(d_{1j}\tau_1 + d_{2j}\tau_2)^{1/2} \exp(-\frac{1}{2} \theta_j^{*2} \{ d_{1j}\tau_1 + d_{2j}\tau_2 \}) \right] p(\tau_1, \tau_2). \tag{4}$$

Denoting $\gamma_j = d_{1j}\tau_1 + d_{2j}\tau_2$, (τ_1, τ_2) has conditional density

$$p(\tau_1, \tau_2|\boldsymbol{\theta}^*) \propto \left[\prod_{j=1}^{n-G} \gamma_j^{1/2} \exp\left(-\frac{\theta_j^{*2}\gamma_j}{2}\right)\right] p(\tau_1, \tau_2), \tag{5}$$

Thus τ_1 and τ_2 enter $p(\tau_1, \tau_2|\boldsymbol{\theta})$, and hence $p(\tau_1, \tau_2|\mathbf{y})$, only through the prior $p(\tau_1, \tau_2)$ and the $N - G$ linear combinations $\{\gamma_j\}$. The conditional density $p(z_1, z_2|\boldsymbol{\theta}^*)$ is also a function of (z_1, z_2) only through the prior and $n - G$ linear functions of (e^{z_1}, e^{z_2}); we omit details.

The jth term of the product in (5) is constant for (τ_1, τ_2) satisfying $d_{1j}\tau_1 + d_{2j}\tau_2 = c$ for $c > 0$, so individual terms in (5)'s product do not identify τ_1 and τ_2. Rather, identification arises from multiplying terms with different ratios d_{1j}/d_{2j}. If there are two or more distinct ratios d_{1j}/d_{2j}, τ_1 and τ_2 are identified. This holds provided each pair of regions are neighbors of at most one type and neither neighborhood structure is null, as assumed; see Appendix A.2.

Each term $\gamma_j^{1/2} \exp(-\frac{\gamma_i \theta_j^{*2}}{2})$ has the form of a gamma density with variate $\gamma_j = d_{1j}\tau_1 + d_{2j}\tau_2$, mode θ_j^{*-2}, and an infinite set of modal (τ_1,τ_2) satisfying $d_{1j}\tau_1 + d_{2j}\tau_2 = \theta_j^{*-2}$. Terms with $d_{1j} \neq 0$ and $d_{2j} \neq 0$ give non-identified modal lines $\tau_2 = -\tau_1 d_{1j}/d_{2j} + \theta_k^{*-2}/d_{2j}$. Only the intercepts of these lines depend on $\boldsymbol{\theta}$; the slopes, $-d_{1j}/d_{2j}$, do not.

Each term in (5) can be deemed a *free term* or a *mixed term* depending on (d_{1j}, d_{2j}). We define the jth term to be a *free term for* τ_1 if $d_{2j} = 0$ and $d_{1j} \neq 0$, and vice versa for τ_2. A free term for τ_1 is a function of τ_1 only, taking the form of a gamma density with variate τ_1. *Mixed terms* have both $d_{1j} \neq 0$ and $d_{2j} \neq 0$.

As Section 4 shows, grids with free terms give better identification than grids with no free terms. As noted, G_1 d_{1j} are zero, G_2 d_{2j} are zero, and G pairs (d_{1j}, d_{2j}) are (0,0), so τ_1 has $G_2 - G$ free terms and τ_2 has $G_1 - G$ free terms. The θ_j^{*2} corresponding to, say, τ_2's free terms are functions of the differences between averages of the $\boldsymbol{\theta}$'s on the G_1 islands defined by class 1 neighbors. For

example, under Grid A, if there are no missing teeth, class 1 neighbors define two islands on each jaw: the long strips on measurements on the jaw's lingual (tongue) and buccal (cheek) sides. For each jaw, the θ_j^{*2} for τ_2's free term is proportional to the difference between the average of lingual $\boldsymbol{\theta}$ and buccal $\boldsymbol{\theta}$, which depends on only τ_2, not τ_1. Similarly, the difference between $\boldsymbol{\theta}$'s average on the four sites in the gap between Teeth 1 and 2 and $\boldsymbol{\theta}$'s average on the four sites in the gap between Teeth 2 and 3 only depends on τ_1, not τ_2, resulting in a free term for τ_1.

Grid C gives no free terms for τ_1 because neighbor pairs controlled by τ_2 (Types I, III, and IV) form a connected graph (Figure 2). Considering spatial maps outside of periodontal analysis, certain grids give no free terms for *either* smoothing parameter. For example, both class 1 and class 2 neighbors in Figure 3's grid form connected graphs, leaving no free terms for either τ_1 or τ_2. Both τ_1 and τ_2 are still identified under this grid because there are mixed terms with different ratios d_{1j}/d_{2j}, but identification is likely to be poor.

If all terms are free terms, τ_1 and τ_2 are conditionally independent *a posteriori* if they are independent *a priori*. This happens if, for example, the data consist of two islands, each with its own τ_l. Mixed terms induce negative correlation between τ_1 and τ_2 conditional on $\boldsymbol{\theta}$. Specifically, a quadratic approximation to $\log p(\tau_1, \tau_2|\boldsymbol{\theta})$ gives $corr(\tau_1, \tau_2|\boldsymbol{\theta}) \approx -\Delta_{12}/\sqrt{\Delta_{11}\Delta_{22}}$, where $\Delta_{ab} = \sum_{j=1}^{n-G} \frac{d_{aj}d_{bj}}{(d_{1j}\tau_1+d_{2j}\tau_2)^2}$. This approximate *conditional* correlation is never positive, but the *marginal* posterior correlation of τ_1 and τ_2 can be.

4 Exploring $p(z_1, z_2|\boldsymbol{y})$

This section derives and explores the marginal posterior of the smoothing parameters using Patient 1's data (Figure 1). No tidy expression like (5) is available for $p(z_1, z_2|\mathbf{y})$ except in special cases. Of course, an MCMC algorithm draws from $p(z_1, z_2|\tau_0, \boldsymbol{\theta}, \boldsymbol{y}) = p(z_1, z_2|\boldsymbol{\theta})$, so the free/mixed

terms ideas developed in Section 3 for $p(\tau_1, \tau_2|\boldsymbol{\theta})$ may still help explain $p(z_1, z_2|\mathbf{y})$.

To compute the marginal posterior of (z_1, z_2), we temporarily ignore the fixed effects and give τ_0 a Gamma(a_0, b_0) prior, giving the full posterior

$$\begin{aligned} p(\boldsymbol{\theta}, \tau_0, \tau_1, \tau_2|\boldsymbol{y}) &\propto p(\tau_1, \tau_2)\tau_0^{n/2+a_0} \prod_{j=1}^{n-G} (\tau_1 d_{1j} + \tau_2 d_{2j})^{1/2} \\ &\propto \exp\left(-\frac{1}{2}\{\tau_0 \left[(\boldsymbol{y}-\boldsymbol{\theta})'(\boldsymbol{y}-\boldsymbol{\theta}) + 2b_0\right] + \boldsymbol{\theta}(\tau_1 Q_1 + \tau_2 Q_2)\boldsymbol{\theta}\}\right), \end{aligned} \tag{6}$$

where $p(\tau_1, \tau_2)$ is (τ_1, τ_2)'s prior. Next, reparameterize to $z_l = log(\tau_l/\tau_0)$ and integrate $\boldsymbol{\theta}$ and τ_0 out of (6), leaving the marginal posterior of the smoothing parameters (z_1, z_2),

$$p(z_1, z_2|\mathbf{y}) \propto p(z_1, z_2) \prod_{j=1}^{n-G} (e^{z_1} d_{1j} + e^{z_2} d_{2j})^{1/2} |I_n + e^{z_1} Q_1 + e^{z_2} Q_2|^{-1/2} R_*^{-a}, \tag{7}$$

where $R_* = b_0 + \frac{1}{2}\{\mathbf{y}' \left(I_n - (I_n + e^{z_1} Q_1 + e^{z_2} Q_2)^{-1}\right) \mathbf{y}\}$ and $a = (n-G)/2 + a_0$.

Figure 4 contains contour plots of the log marginal posterior of (z_1, z_2) under each grid using Subject 1's data (Figure 1). With each contour plot we present a graph of the $n - G = 159$ unidentified lines evaluated at the marginal posterior median of (z_1, z_2), e.g., the set of (x_1, x_2) satisfying $d_{1j}e^{x_1} + d_{2j}e^{x_2} = d_{1j}e^{\tilde{z}_1} + d_{2j}e^{\tilde{z}_2}, i = 1, ..., N - G$, where $\tilde{z}_l$ is the posterior median of z_l. In these plots, the straight lines represent the unidentified modal lines arising from free terms for z_1 (vertical) or z_2 (horizontal), and the curved lines represent the unidentified modal curves arising from mixed terms.

Table 2 gives counts of free and mixed terms for the three grids with two neighbor relations. The shape of (z_1, z_2)'s posterior is largely determined by the free terms. Grids A and B (Figures 4b and 4d) have long upper tails at specific z_1 and z_2 arising from the free terms, e.g., at $z_1 \approx 1$ and $z_2 \approx -2.5$ for Grid A. The long upper tails (as opposed to lower tails) may reflect positive

Table 2: Free-term counts for each periodontal grid for Patient 1

Grid	n	G	G_1	G_2	Free Terms for z_1	Free Terms for z_2	Mixed Terms
A	162	3	6	84	81	3	75
B	162	3	54	84	81	51	27
C	162	3	114	3	0	111	48

skewness in the distribution of the free terms' intercepts. Grid C has no free terms for z_1, which explains the absence of a long upper tail for z_2 (Figure 4f).

Grid A has many free terms for z_1, few for z_2, and many mixed terms. The disparity in free terms implies better identification of z_1 than z_2. The ridge along the vertical line $z_1 \approx 1$ in Figure 4b is more narrow than the ridge along the horizontal line $z_2 \approx -2$. The many mixed terms (Figure 4a) induce some curvature of the L-shaped contours near the intersection of the lines of mass along $z_1 \approx 1$ and $z_2 \approx -2$.

Grid B has many free terms for both z_1 and z_2 and far fewer mixed terms than Grid A. Again, z_1 has more free terms than z_2, so the density is more peaked along $z_1 \approx 1$ than along $z_2 \approx 1$ (Figure 4d). The relative paucity of mixed terms (Figure 4c) means little curvature in the L-shape of (z_1, z_2)'s posterior.

The absence of free terms for z_1 under Grid C explains the poor identification of z_1 in Figure 4f. For large z_2, class 2 neighbors are highly smoothed; because they form a connected graph, class 1 neighbors are forced to be similar regardless of z_1, that is, the data provide little information for z_1. Thus, for large z_2, say $z_2 \geq 4$, z_1's posterior is nearly flat (Figure 4f). For small z_2, say $z_2 = -1$, smoothing of class 2 neighbors does not obscure smoothing of class 1 neighbors, allowing the data to rule out $z_1 < 0$. For $z_2 = -1$, z_1's posterior is still flat for large z_1 because all large z_1 correspond to almost complete smoothing of class 1 neighbors.

5 Analysis of periodontal data for many patients

This section uses Section 2's model to analyze baseline attachment loss from $N = 50$ patients in the clinical trial described in Section 1, originally described and analyzed by Shievitz (1997). MCMC convergence was improved by integrating out the CAR random effects and sampling from the marginal posterior of $(\boldsymbol{b}, z_{0_p}, z_{1_p}, z_{2_p})$. For each model, structured MCMC (Sargent et al., 2000) with blocks $\boldsymbol{b}$ and $(z_{0_p}, z_{1_p}, z_{2_p})$ was used to make 20,000 draws from $\mathrm{p}(\boldsymbol{b}, z_{0_p}, z_{1_p}, z_{2_p}|\boldsymbol{y})$. Draws of $\boldsymbol{\theta}_p$ were then generated from the conditional distribution of $\boldsymbol{\theta}_p$ given each iteration's $(\boldsymbol{b}, z_{0_p}, z_{1_p}, z_{2_p})$. Convergence was assessed by comparing summaries of the (z_{1_p}, z_{2_p}) draws to contour plots of the exact posterior of (z_{1_p}, z_{2_p}) given by Equation (7).

Figure 5a is a scatterplot of $(\tilde{z}_{0_p}, \tilde{z}_{1_p})$ from the 1NR grid assuming the patients' z_{l_p} have independent Uniform(-10,10) priors, where $\tilde{z}_{l_p}$ is z_{l_p}'s posterior median. Most of the $\tilde{z}_{0_p} = log(\tilde{\tau}_{0_p})$ are near zero, which corresponds to measurement error with standard deviation roughly 1.0, as mentioned in Section 1. The most striking feature of this plot is the patient-to-patient variation in the smoothing parameters z_{1_p}. For example, $\tilde{z}_{1_{15}} = -6.48$, i.e., smoothing of $\boldsymbol{\theta}_{15}$ is negligible (Figure 6) and $\tilde{z}_{1_1} = 5.97$, i.e., $\boldsymbol{\theta}_1$ is substantially smoothed (Figure 7). The different amounts of smoothing for these two patients may be driven by the steady increase in AL from the front to the back of the lower jaw for Patient 15. In contract with the random scatter for Patient 1, this distinct spatial pattern indicates that the site-to-site variation in Patient 15's AL is real and should not be smoothed over. By contrast, the model with $(z_{0_p}, z_{1_p}, z_{2_p})' \sim N(\mu, \Sigma)$ *a priori* smooths the $(\tilde{z}_{0_p}, \tilde{z}_{1_p})$ towards (0,0) (Figure 5b).

Figure 5c also shows considerable patient-to-patient variation in $(\tilde{z}_{1_p}, \tilde{z}_{2_p})$ for Grid A assuming the z_{l_p} have independent Uniform(-10,10) priors. For the most part, $\tilde{z}_{1_p} \in (0, 4)$, but the $\tilde{z}_{2_p}$ vary almost uniformly from -3 to 5. The plot of posterior medians resembles the shape of the marginal

posterior distribution of Patient 1's (z_{1_1}, z_{2_1}) under Grid A, in Figure 4b. This may be explained by the counts of free terms in Table 2. Since Grid A gives many free terms for z_1 and few free terms for z_2, we expect the data to provide less information about z_2 than z_1. Therefore, the sampling distribution of $\tilde{z}_{2_p}$ should have more variation than the sampling distribution of $\tilde{z}_{1_p}$, as in Figure 5c.

The counts of free terms may also explain $(\tilde{z}_{1_p}, \tilde{z}_{2_p})$'s shrinkage under the model where the $(z_{0_p}, z_{1_p}, z_{2_p})$ have a multivariate normal prior. Figure 5d shows that the z_{1_p} are moderately shrunk compared to Figure 5c, but the z_{2_p} are shrunk almost completely. Since z_1 has more free terms than z_2, the data are more informative for z_1 than z_2 and the prior has less influence on z_1's posterior than z_2's posterior.

We compare the models using the deviance information criterion (DIC) of Speigelhalter et al. (2002). Defining $D(\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2) = -2 \log f(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2)$, $DIC = \bar{D} + P_D$ where $P_D = \bar{D} - \hat{D}$, $\bar{D} = E(D(\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2)|\boldsymbol{y})$, and $\hat{D} = D(E(\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2|\boldsymbol{y}))$, the expectations being taken with respect to the full posterior. The model's fit is measured by the posterior mean of the deviance, $\bar{D}$, while the model's complexity is captured by P_D, the effective number of parameters in the model. Models with smaller DIC and $\bar{D}$ are favored. DIC can also be computed individually for each patient. Defining $D_p(\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2) = -2 \log f(\boldsymbol{y}_p|\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2)$, $DIC_p = \bar{D}_p + P_{D_p}$ where $P_{D_p} = \bar{D}_p - \hat{D}_p$, $\bar{D}_p = E(D_p(\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2)|\boldsymbol{y})$, and $\hat{D} = D_p(E(\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2|\boldsymbol{y}))$. Since $\sum_{p=1}^{N} D_p(\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2) = D(\boldsymbol{\theta}, \boldsymbol{b}, z_0, z_1, z_2)$, $\sum_{p=1}^{N} DIC_p = DIC$ and $\sum_{p=1}^{N} P_{D_p} = P_D$.

Table 3 gives DIC for each grid and prior choice for the z_{l_p}. Although the spatial models are more complex (i.e., have larger P_D) than the non-spatial models, DIC strongly favors spatial modelling of these periodontal data because of the substantial improvement in fit (i.e., reduction in $\bar{D}$); the two models with only fixed effects and without spatial CAR random effects ("FE only") have the largest DIC statistics by far. No patient's DIC_p favors these non-spatial models. Also,

Table 3: Summary of DIC for various models. "DIC_1" is the DIC specific to only Patient 1. "# pats min DIC_p" is the number of patients that have the smallest DIC_p for the given model. "FE only" is the non-spatial model with only fixed effects.

(a) z_{l_p} ∼Uniform(-10,10) *a priori*, independent across l and p

	FE only	1NR Grid	Grid A	Grid B	Grid C
DIC	16449	7863	8669	8720	8793
$\bar{D}$	16391	6244	7213	7361	7152
P_D	57.8	1619	1456	1359	1640
DIC_1	358.1	283.8	272.4	283.4	281.7
# pats min DIC_p	0	15	16	9	10

(b) $(z_{0_p}, z_{1_p}, z_{2_p})' \sim N(\mu, \Sigma)$ *a priori*

	FE only	1NR Grid	Grid A	Grid B	Grid C
DIC	16451	9953	9968	10199	9899
$\bar{D}$	16394	8140	8065	8597	8091
P_D	57.3	1813	1903	1602	1808

for each grid, DIC and $\bar{D}$ overwhelmingly favor independent z_{l_p} over shrunken z_{l_p}. This is not surprising considering the large patient-to-patient variation and non-Gaussian scatters in Figures 5a and 5c.

Assuming independent z_{l_p} *a priori*, the 1NR grid has smaller DIC and $\bar{D}$ than each two-neighbor-relation grid (Table 3a). However, patients are far from unanimous in favoring the 1NR grid. A two-neighbor-relation grid minimizes 35 of the 50 patients' DIC_p ("# pats min DIC_p"). Grid A has the smallest DIC of the 2NR grids and minimizes DIC_p for more patients than the 1NR grid. Even Grid C, which has the largest DIC of the spatial grids, minimizes DIC_p for 10 of the 50 patients.

Table 4 shows posterior summaries for the fixed effects. Under the non-spatial model ("FE only") each fixed-effect posterior interval except Tooth 4 excludes zero. Under this model, mean AL is higher at direct sites than sites in the gap between teeth, lower on teeth 2 and 3 than tooth 1, and higher on teeth 5-7 than tooth 1. As expected, the 95% interval of each fixed effect is wider under each spatial model than under the FE-only model. Also, the posterior medians of the tooth-number fixed effects are generally closer to zero under the spatial models, especially Grid

Table 4: Posterior median and 95% interval of the fixed effects under different grids assuming $z_{l_p} \sim$ Uniform(-10,10) independent across l and p. "FE only" refers to the non-spatial model with only fixed effect.

	FE only	1NR	Grid A	Grid B	Grid C
Direct	0.18 (0.13, 0.22)	0.21 (0.16, 0.26)	0.20 (0.15, 0.25)	0.21 (0.16, 0.26)	0.21 (0.16, 0.26)
Tooth 2	-0.14 (-0.21, -0.06)	-0.09 (-0.19, 0.01)	-0.10 (-0.20, 0.00)	-0.10 (-0.20, 0.00)	-0.09 (-0.18, 0.02)
Tooth 3	-0.26 (-0.33, -0.18)	-0.22 (-0.34, -0.10)	-0.24 (-0.36, -0.13)	-0.25 (-0.38, -0.13)	-0.22 (-0.34, -0.10)
Tooth 4	-0.02 (-0.10, 0.06)	0.01 (-0.14, 0.14)	-0.01 (-0.14, 0.12)	-0.03 (-0.17, 0.11)	0.01 (-0.13, 0.15)
Tooth 5	0.19 (0.11, 0.27)	0.19 (0.03, 0.35)	0.18 (0.02, 0.33)	0.13 (0.04, 0.29)	0.20 (0.03, 0.35)
Tooth 6	0.80 (0.72, 0.88)	0.76 (0.56, 0.94)	0.76 (0.58, 0.92)	0.69 (0.50, 0.88)	0.74 (0.54, 0.92)
Tooth 7	0.91 (0.83, 0.99)	0.82 (0.61, 1.03)	0.83 (0.63, 1.01)	0.74 (0.53, 0.96)	0.83 (0.62, 1.03)

B, compared to the FE-only model. Under the spatial models, variation in $\boldsymbol{\theta}$ absorbs some of the tooth-number effects and nudges the tooth-number fixed effects towards zero. Reich et al. (2006) explore the effect of adding CAR parameters on the fixed effects in spatial regression. These authors show that spatial smoothing has the largest effect on covariates that vary smoothly in space, which explains why the changes in medians from the non-spatial model to the spatial models are larger for the tooth-number effects than for the direct-site effect.

Grid A minimizes DIC_1, the DIC specific to Patient 1. Figure 7 plots Patient 1's data (symbols) and fitted values, i.e., $\boldsymbol{\theta}_1 + X_1\boldsymbol{b}$'s posterior mean, under the 1NR grid (solid lines) and Grid A (dashed lines). The fitted values under the 1NR grid are similar to the fitted values under Grids B and C (not shown), but often differ from Grid A's fitted values by over 0.5 mm. For Grid A, draws of z_1 are generally larger than draws of z_2 and $(\tilde{z}_{1_1}, \tilde{z}_{2_1}) = (7.36, -2.70)$; here z_1 controls smoothing of Type I and II neighbor pairs, which form long strips of sites along the buccal and lingual sides of each jaw. Large z_1 and small z_2 smooth substantially within these long strips but not between them. Figure 7 shows that this is preferable for the upper jaw, where attachment loss is similar across tooth number but is larger for lingual sites (mean AL = 3.49) than buccal sites (mean AL = 2.20). Grids B, C, and 1NR smooth more between these strips.

6 Effect of the prior distribution

Choosing parameterizations and priors for the scale parameters in hierarchical models is an important and unresolved issue (Daniels and Kass, 1999,2001; Natarajan and Kass, 2000; Kelsall and Wakefield, 2002) that we cannot settle here. However, the generally poor identification of the smoothing parameters calls for an investigation of the sensitivity of Section 5's results to changes in the priors for the scale parameters. This section considers five putatively vague parameterizations/priors, each independent across patients: Uniform(-10,10) and LogGamma(0.01,0.01) priors for the z_{l_p}, Uniform(0,10) priors on the standard deviations $\sigma_{l_p} = \tau_{l_p}^{-1/2}$, Gamma(0.01,0.01) priors on the variances $\sigma_{l_p}^2$, and a prior motivated by Besag and Higdon (1999) having $\tau_{0_p} \sim$ G(0.01,0.01), $\lambda_p = \tau_{1_p} + \tau_{2_p} \sim$ G(0.01,0.01), and $\beta_p = \tau_{1_p}/(\tau_{1_p} + \tau_{2_p}) \sim$ U(0,1). For each parameterization/prior, we consider Grid A and discuss the types of fit it encourages and examine the influence on fixed-effect estimates, fitted values, and the posteriors of the smoothing parameters.

Figure 8 summarizes the induced prior (contour lines) and posterior samples (boxes and whiskers) of the (z_{1_p}, z_{2_p}) for each parameterization/prior. For several patients, the posterior of the smoothing parameters is affected by the prior, especially for patients with large z_{1_p} or z_{2_p}. For example, Figure 8b's LogGamma(0.01,0.01) prior for z_{l_p} favors small z_{l_p} and precludes extremely smooth fits with $z_{1_p} > 5$ or $z_{2_p} > 5$. The priors in Figures 8c – 8e all have mode (0,0) and encourage moderate levels of smoothing. Of these three priors, the uniform prior on the standard deviations (Figure 8c) shrinks the (z_{1_p}, z_{2_p}) most towards (0, 0) and the inverse gamma prior for the variances (Figure 8d) is most like the uniform prior on z_{l_p}. The Uniform(0,1) prior on the β_p, which control the relative amount of smoothing of the two classes of neighbor pairs, discourages fits with large z_{l_p} and small z_{2_p} (i.e., $\beta_p \approx 1$), or small z_{l_p} and large z_{2_p} (i.e., $\beta_p \approx 0$), and shrinks the smoothing parameters towards the line $z_1 = z_2$ (Figure 8e).

Table 5: Posterior median and 95% interval of the fixed effects under Grid A and different priors/parameteriztions for the scale parameters. Each column corresponds to a different parameteriztion/prior. The scale parameters are independent across and within patients *a priori* under each prior. The transformations used in this table are $\sigma^2_{l_p} = 1/\tau_{l_p}$, $\lambda_p = \tau_{1_p}+\tau_{2_p}$, and $\beta_p = \tau_{1_p}/(\tau_{1_p}+\tau_{2_p})$.

	$z_{0_p} \sim$ U(-15,15) $z_{1_p} \sim$ U(-15,15) $z_{2_p} \sim$ U(-15,15)	$z_{0_p} \sim$ LG(.01,.01) $z_{1_p} \sim$ LG(.01,.01) $z_{2_p} \sim$ LG(.01,.01)	$\sigma_{0_p} \sim$ U(0,10) $\sigma_{1_p} \sim$ U(0,10) $\sigma_{2_p} \sim$ U(0,10)	$\sigma^2_{0_p} \sim$ IG(.01,.01) $\sigma^2_{1_p} \sim$ IG(.01,.01) $\sigma^2_{2_p} \sim$ IG(.01,.01)	$\tau_{0_p} \sim$ G(.01,.01) $\lambda_p \sim$ G(.01,.01) $\beta_p \sim$ U(0,1)
Direct	0.20 (0.15, 0.25)	0.20 (0.15, 0.26)	0.21 (0.16, 0.26)	0.21 (0.15, 0.26)	0.20 (0.15, 0.25)
Tooth 2	-0.10 (-0.20, 0.00)	-0.10 (-0.20, 0.00)	-0.09 (-0.20, 0.01)	-0.10 (-0.20, 0.00)	-0.09 (-0.19, 0.01)
Tooth 3	-0.24 (-0.36, -0.13)	-0.24 (-0.36, -0.12)	-0.23 (-0.36, -0.11)	-0.24 (-0.36, -0.13)	-0.24 (-0.36, -0.12)
Tooth 4	-0.01 (-0.14, 0.12)	-0.02 (-0.16, 0.12)	-0.01 (-0.16, 0.13)	-0.02 (-0.15, 0.12)	-0.01 (-0.16, 0.13)
Tooth 5	0.18 (0.02, 0.33)	0.16 (0.01, 0.32)	0.17 (0.00, 0.33)	0.17 (0.02, 0.32)	0.17 (0.01, 0.33)
Tooth 6	0.75 (0.58, 0.92)	0.73 (0.56, 0.90)	0.73 (0.54, 0.92)	0.75 (0.58, 0.92)	0.74 (0.55, 0.92)
Tooth 7	0.82 (0.63, 1.01)	0.80 (0.61, 0.99)	0.81 (0.60, 1.02)	0.82 (0.63, 1.01)	0.80 (0.60, 1.00)

The prior also has a noticeable effect on Patient 1's fitted values. Figure 9 shows $X_1\beta + \boldsymbol{\theta}_1$'s posterior mean for Patient 1's left maxillary island under two priors for the scale parameters: $z_{l_p} \sim$ Uniform(-10,10) (solid lines) and the prior motivated by Besag and Higdon (dashed lines). Under Grid A, z_{1_1} controls smoothing of Type I and II neighbor pairs, which form long strips of sites along the buccal and lingual sides of each jaw, that is, z_{1_1} controls smoothing of adjacent sites within Figure 9's two rows. With $z_{l_p} \sim$ Uniform(-10,10), β_1's posterior median is 0.998, i.e., z_{1_1} is generally larger than z_{2_1}, and the fitted values within Figure 9's rows are very similar. Recall that the uniform prior on β_p discourages fits with large z_{1_1} and small z_{2_1} (Figure 8e). Under this prior, β_1's posterior median is 0.867 and the fitted values within Figure 9's rows are less smooth, differing from the fitted values under the uniform prior on the z_{l_p} by as much as 1.47 mm (Tooth 1).

The fixed-effect estimates and intervals in Table 5 are nearly identical under the five priors. This agrees with previous work in this area, e.g., Daniels and Kass (1999). These authors show that the prior on the scale parameter may lead to substantial changes in the posterior of the scale parameters, but such changes typically do not affect the fixed effects' posteriors.

7 Discussion

In our periodontal data analysis, our model choice statistic sometimes favored models with two neighbor relations despite their increased complexity. Although the 1NR grid had a smaller overall DIC than any of the two-neighbor-relation grids, Grid A minimized the patient-specific DIC for more patients than the 1NR grid. The spatial structure appeared to vary considerably among these 50 patients, who were haphazardly selected from the study population.

The empirical correlations in Table 1a suggest a fourth two-neighbor-relation grid, Grid D, that allows for differential smoothing of neighbor pairs on the same tooth (Types I and III) and neighbor pairs on different teeth (Types II and IV). Assuming the z_{l_s} have independent Unif(-10,10) priors, Grid D has a smaller overall DIC than Grid A. However, as Figure 10 shows, this result is largely driven by Patients 3, 15, 29, 36 and 38. These patients all have severe periodontal disease (e.g., Patient 15 in Figure 6) and may not represent the study population. Although these patients do not seem to be influential on the fixed effects, they do influence the choice of spatial grid; Grid A has smaller patient-specific DIC than Grid D for 28 of the 45 remaining patients. Stratification into more homogeneous groups of patients, such as "moderate" and "advanced" groups, may be needed for one grid to prevail as "best" for all patients. Stratification may also allow the degree of smoothing to be the same for all patients in a stratum, although our model still enables different smoothing for different patients when this is appropriate (Figures 6 and 7). Modelling with random grids may also be possible (Lu, 2005).

The smoothing parameters $z_l = log(\tau_l/\tau_0)$ are identified except in trivial cases, but identification can be poor depending on the spatial structure. Free terms greatly enhance identification. Free terms arise from prior contrasts in $\boldsymbol{\theta}$ with precision depending on only one z_l. Generally, z_l with no free terms are poorly identified, especially if neighbor pairs of the other class are highly

smoothed. This may cause computing problems such as poor MCMC convergence. MCMC algorithms exploiting the free-term structure may give better performance.

For CAR models with two neighbor relations, the prior on the smoothing parameters is very important. The posterior of several patients' smoothing parameters were affected by the priors considered in Section 6. The prior also had a large effect on Patient 1's fitted values. However, similar to other work in this area (e.g., Daniels and Kass, 1999), for these data, the smoothing parameters' prior did not affect the fixed-effect estimates.

Finally, this paper presents a method for analyzing baseline periodontal data. In practice, longitudinal data may also be of interest. The 2NRCAR prior could be applied in this spatiotemporal setting by defining the two neighbor types to be "spatial neighbors" at the same visit and "temporal neighbors" at the same spatial location.

References

Besag J, Higdon D (1999). Bayesian analysis of agricultural field experiments (with discussion). *J. Roy. Stat. Soc., Series B*, **61**: 691–746.

Besag J, York JC, Mollié A (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Stat. Math.*, **43**:1–59.

Best NG, Waller LA, Thomas A, Conlon EM, Arnold RA (1999). Bayesian models for spatially correlated disease and exposure data (with discussion). *Bayesian Statistics 6*:131–156.

Daniels MJ, Kass RE (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J. Amer. Statist. Assoc.*, **94**:1254–1263.

Daniels MJ, Kass RE (2001) Shrinkage estimators for covariance matrices. *Biometrics*, **57**:1173–1184.

Darby ML, Walsh MM (1995). *Periodontal and Oral Hygiene Assessment. Dental Hygiene Theory and Practice.* Philadelphia: W. B. Saunders.

Drury TF, Winn DM, Snowden CB, Kingman A, Kleinman DV, Lewis B (1996). An overview of the oral health component of the 1988-1991 National Health and Nutrition Examination Survey (NHANES III-Phase 1). *J. Dent. Research* **75**:620–630.

Graybill FA (1983). *Matrices with Applications in Statistics.* Pacific Grove, CA: Wadsworth & Brooks.

Gunsolley JC, Williams DA, Schenkein HA (1994). Variance component modeling of attachment level measurements. *J. Clin. Perio.*, **21**:289–295.

Hodges JS, Carlin BP, Fan Q (2003). On the precision of the conditionally autoregressive prior in spatial models. *Biometrics*, **59**:317–322.

Kelsall JE, Wakefield JC (2002). Modeling spatial variation in disease risk: A geostatistical approach. *J. Amer. Statist. Assoc.*, **97**:692–701.

Lu H (2005). Measuring the complexity of generalized linear hierarchical models and Bayesian areal Wombling for boundary analysis. PhD Thesis, University of Minnesota.

Natarajan R, Kass RE (2000). Reference bayesian methods for generalized linear models. *J. Amer. Statist. Assoc.*, **95**:227–237.

Newcomb RW (1961). On the simultaneous diagonalization of two semi-definite matrices. *Quart. Appl. Math.*, **19**: 144–146.

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

Osborn JB, Stoltenberg JL, Huso BA, Aeppli DM, Pihlstrom BL. (1992). Comparison of measurement variability in subjects with moderate periodontitis using a conventional and constant force periodontal probe. *J. Periodontology*, **63**:283–289.

Reich BJ, Hodges JS, Zadnik V (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, In press.

Roberts T (1999). Examining mean structure, covariance structure and correlation of interproximal sites in a large periodontal data set. MS Thesis, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN.

Sargent DJ, Hodges JS, Carlin BP (2000). Structured Markov chain Monte Carlo. *J. Comp. Graph. Stat.* **9**:217–234.

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

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002). Bayesian measures of model complexity and fit (with discussion and rejoinder) *J. Roy. Statist. Soc., Ser. B*, **64**:583–639.

Sterne JAC, Johnson NW, Wilton JMA, Joyston-Bechel S, Smales FC (1988). Variance components analysis of data from periodontal research. *J. Perio. Res.*, **23**:148–153.

Appendix

A.1 The CAR Prior with Two Neighbor Relations

Newcomb (1961) showed how to construct a nonsingular B such that $Q_1 = B'D_1B$ and $Q_2 = B'D_2B$, where D_l is diagonal with $n - G_l$ positive diagonal entries and G_l zero entries. Thus (2)'s exponent can be written as $-\frac{1}{2}\boldsymbol{\theta}'B'\{\tau_1 D_1 + \tau_2 D_2\}B\boldsymbol{\theta}$. B is orthogonal only if Q_1Q_2 is symmetric (Graybill 1983, Theorem 12.2.12). Also, B is not unique, but apart from permuting rows or

columns, any B can be obtained from any other B by pre-multiplying by a diagonal matrix with positive diagonal entries. As will become clear, any such change, or any permutation of B's rows or columns, has no noteworthy effect.

For a given B, define D_l's diagonal elements as $d_{lj} \geq 0, l = 1, 2$ and $j = 1, \ldots, n$. (The d_{lj} depend on B; we suppress this for simplicity.) For exactly G values of j, $d_{1j} = d_{2j} = 0$. To see this, set $\tau_1 = \tau_2$, turning the problem back into a CAR prior with one class of neighbor relation; $D_1 + D_2$ has exactly G zero diagonal entries, and the result follows. Without loss of generality, define B so D_l's last G diagonal entries are zero and $d_{1j} + d_{2j} > 0$ for $j = 1, \ldots, n - G$.

Following Hodges et al. (2003), define $\boldsymbol{\theta}^* = B\boldsymbol{\theta}$ and partition $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}^{*'} = (\boldsymbol{\theta}_1^{*'}, \boldsymbol{\theta}_2^{*'})$, where $\boldsymbol{\theta}_1^*$ has length $n-G$ and $\boldsymbol{\theta}_2^*$ has length G. Then (2)'s exponent is $-\frac{1}{2}\boldsymbol{\theta}_1^{*'} diag\{\tau_1 d_{1j} + \tau_2 d_{2j}\}\boldsymbol{\theta}_1^*$, $diag\{\nu_j\}$ being a diagonal matrix with $\{\nu_j\}$ on the diagonal in the order $j = 1, \ldots, n - G$. This exponent is the kernel of a proper multivariate normal density for $\boldsymbol{\theta}_1^*$, which has multiplier

$$\prod_{j=1}^{n-G} (\tau_1 d_{1j} + \tau_2 d_{2j})^{1/2}. \tag{8}$$

For j with both d_{lj} positive, the jth term's contribution to the multiplier is determined by the ratio d_{1j}/d_{2j}, because $d_{2j} > 0$ can be factored out and disappears in the proportionality constant. For j with only one d_{lj} positive, that d_{lj} can likewise be factored out. Thus the proper version of this CAR prior is unique even though B is not.

A.2 Proof that z_1 and z_2 are identified in non-trivial cases

If $d_{1j}/d_{2j} = c$ for $j = 1, ..., n - G$, then $D_1 = cD_2$, which implies $Q_1 = cQ_2$. Since off-diagonal elements of Q_l are 0 or -1, either $c = 1$ and $Q_1 = Q_2$, so each neighbor pair is a pair of both classes, or $c = 0$ and Q_1 is the zero matrix, i.e., its neighborhood structure is null. Both possibilities were ruled out by assumption, so there are at least two distinct d_{1j}/d_{2j}.

Figure 1: Patient 1's attachment loss. The shaded boxes represent teeth and the circles represent measurement sites. "Maxillary" and "Mandibular" refer to upper and lower jaws respectively. The maxilla's second tooth on the left is missing.

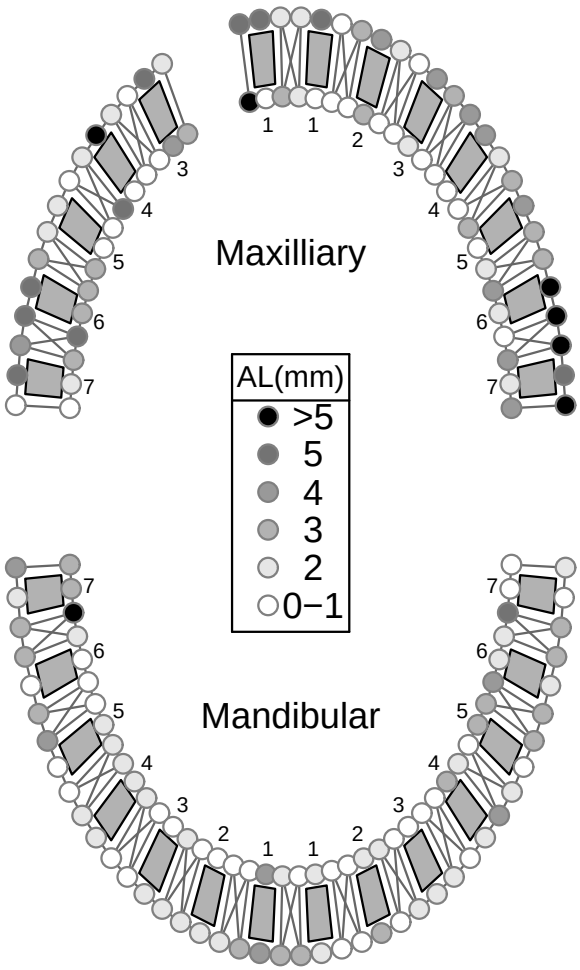


Figure 2: Neighbor pairs in a three-tooth periodontal grid. Rectangles represent teeth, circles represent sites where AL is measured, and different line types represent the types of neighbor pairs.

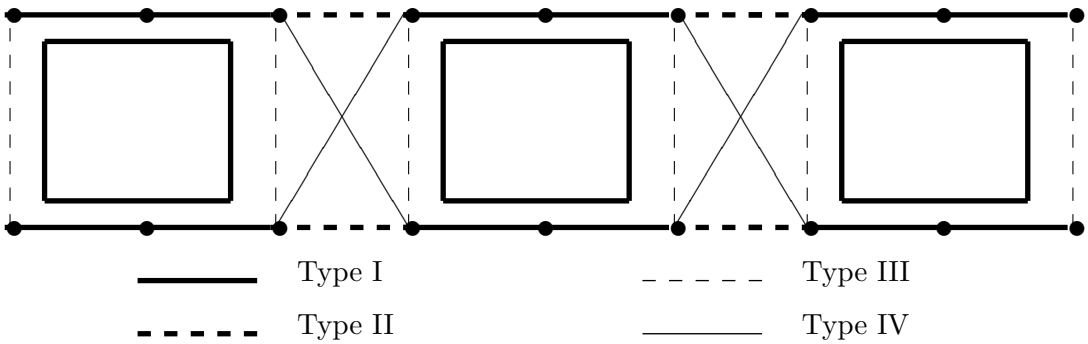


Figure 3: Non-periodontal grid with no free terms for either τ_1 or τ_2. The solid lines are class 1 neighbors and the dashed lines are class 2 neighbors.

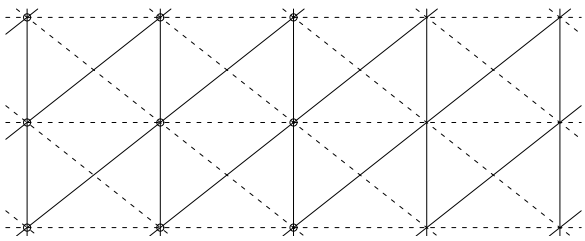
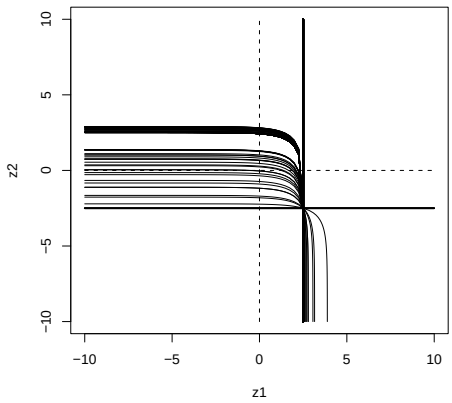
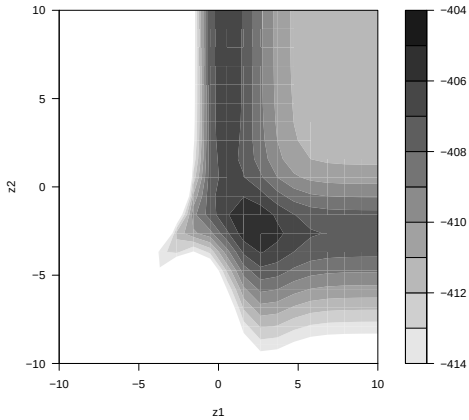


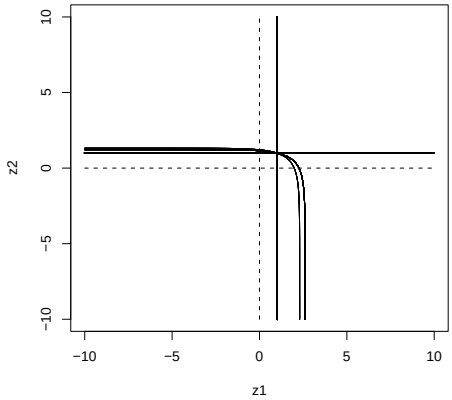
Figure 4: Posterior of (z_1, z_2) for Patient 1 under various spatial grids



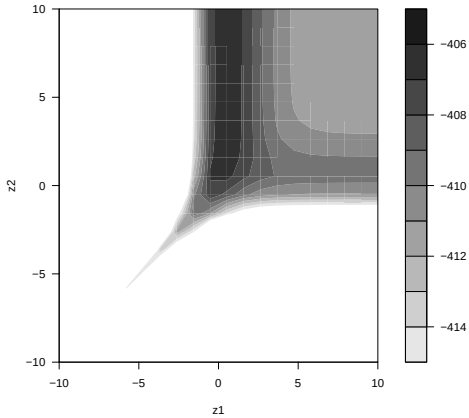
(a) Non-identified curves based on (z_1, z_2)'s posterior median under Grid A



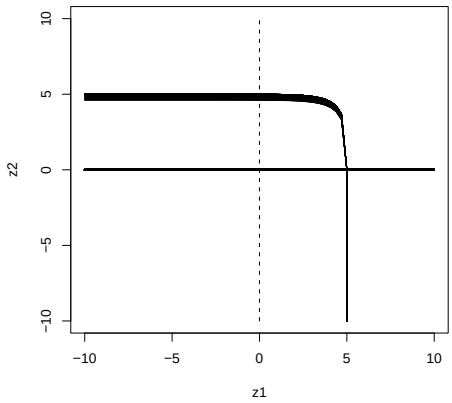
(b) Contour plot of (z_1, z_2)'s log marginal posterior under Grid A



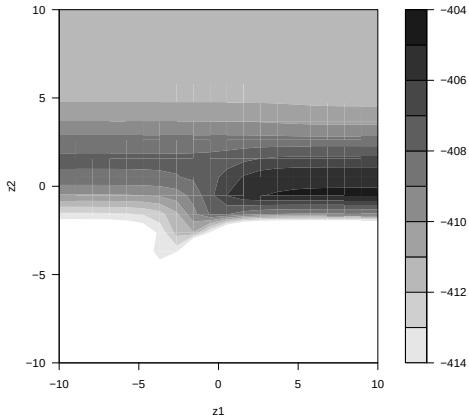
(c) Non-identified curves based on (z_1, z_2)'s posterior median under Grid B



(d) Contour plot of (z_1, z_2)'s log marginal posterior under Grid B

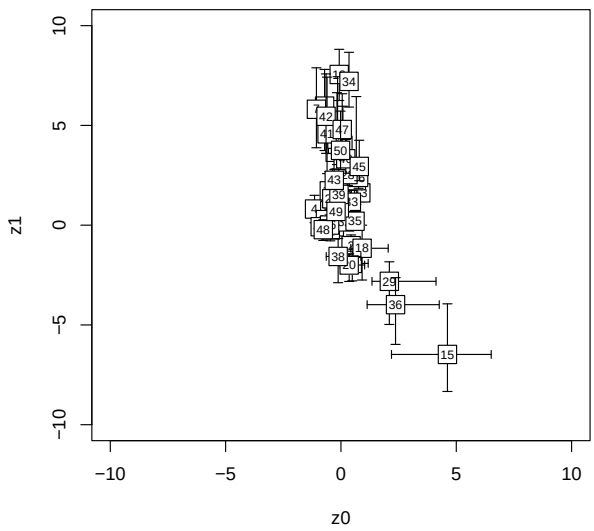


(e) Non-identified curves based on (z_1, z_2)'s posterior median under Grid C

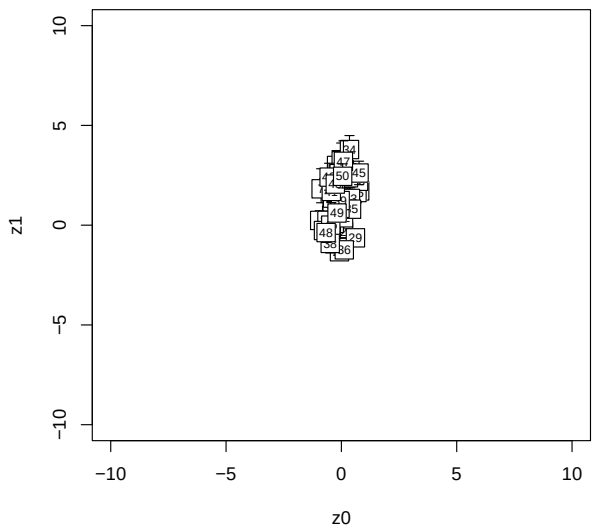


(f) Contour plot of (z_1, z_2)'s log marginal posterior under Grid C

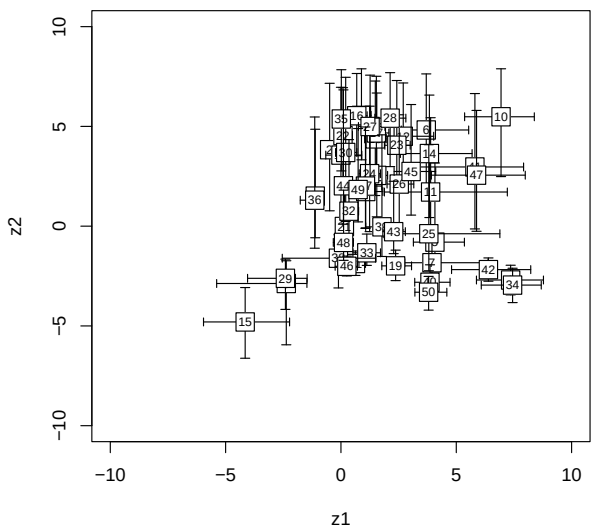
Figure 5: Summary of the posterior of each patient's smoothing parameters. The boxes represent the posterior medians and the whiskers represent the interquartile ranges of each patient's (z_{0_p}, z_{1_p}) or (z_{1_p}, z_{2_p}), $p = 1, ..., 50$.



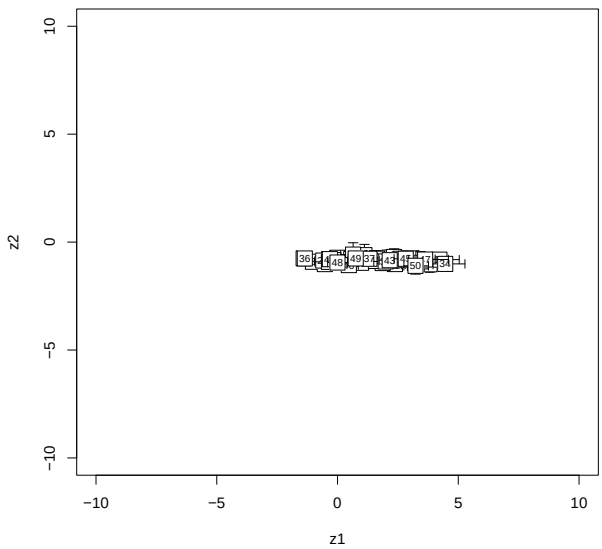
(a) (z_{0_p}, z_{1_p}) under the 1NR grid assuming the z_{l_p} are independent *a priori*



(b) (z_{0_p}, z_{1_p}) under the 1NR grid assuming the (z_{0_p}, z_{l_p}) are shrunk with a normal prior



(c) (z_{1_p}, z_{2_p}) under Grid A assuming the precisions are independent *a priori*



(d) (z_{1_p}, z_{2_p}) under Grid A assuming the $(z_{0_p}, z_{l_p}, z_{2_p})$ are shrunk with a normal prior

Figure 6: Patient 15's data (symbols) and posterior mean of $X_{15}\beta + \boldsymbol{\theta}_{15}$ (solid line) for the 1NR grid assuming the z_{l_s} are independent *a priori*. "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.

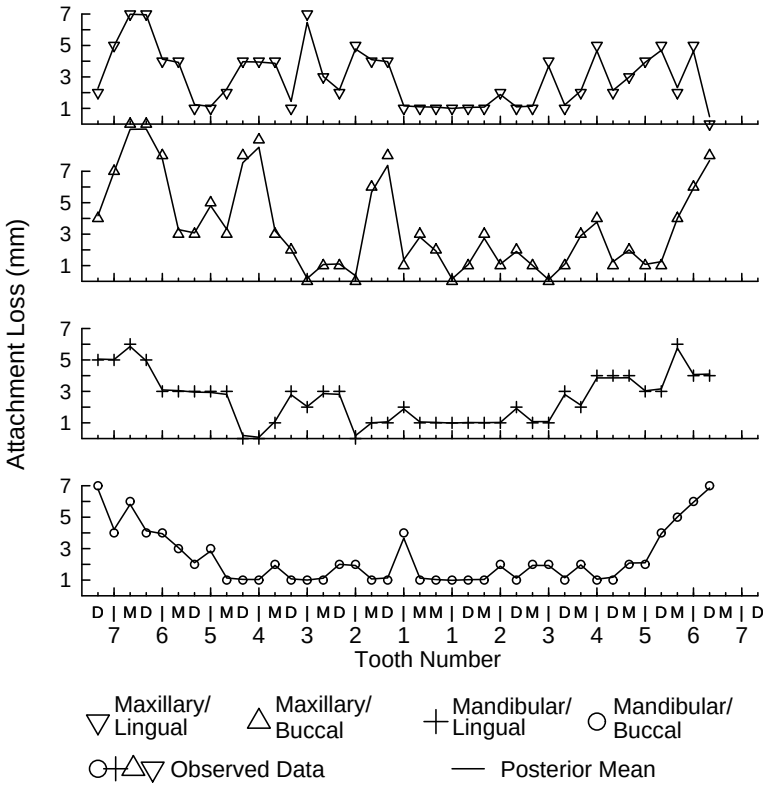


Figure 7: Patient 1's data (symbols) and posterior mean of $X_1\beta + \boldsymbol{\theta}_1$ assuming the z_{l_p} are independent *a priori* for the 1NR grid (solid lines) and Grid A (dashed lines). "Maxillary" and "Mandibular" refer to upper and lower jaws respectively, while "buccal" and "lingual" refer to the cheek and the tongue sides of the teeth, respectively.

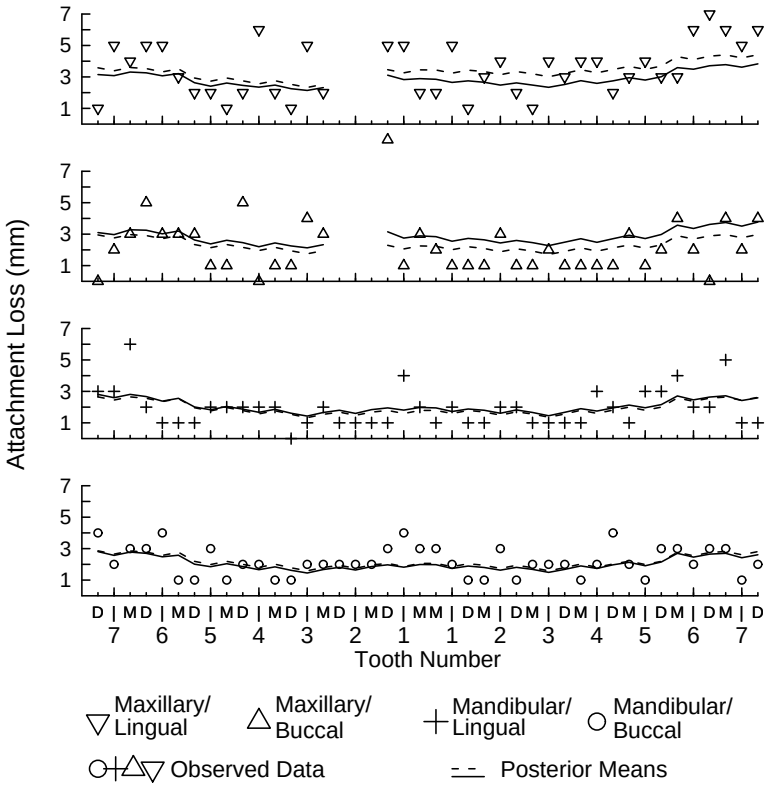
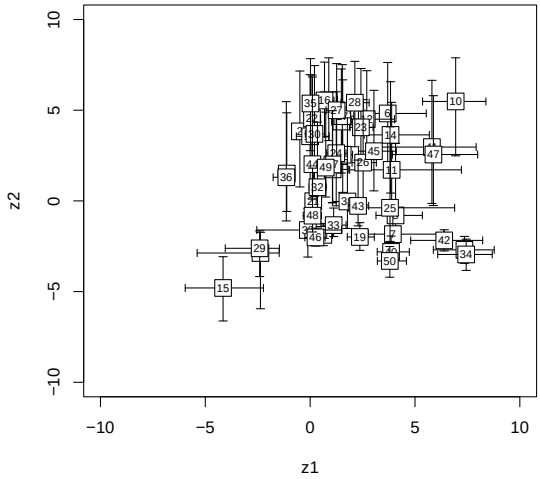
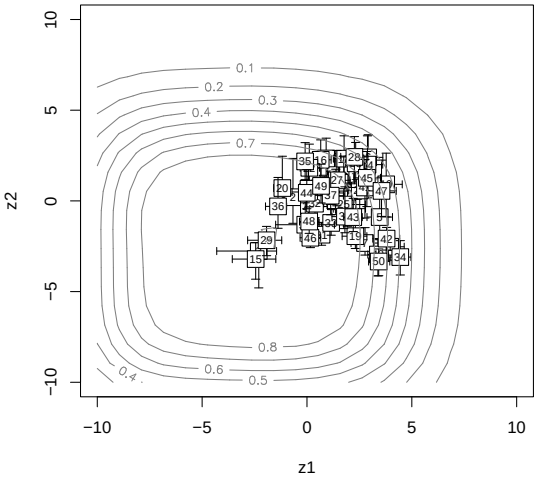


Figure 8: Summary each patient's smoothing parameters under different priors. The boxes represent the posterior medians and the whiskers represent the interquartile ranges of each patient's (z_{1_p}, z_{2_p}), $p = 1, ..., 50$. The shaded lines are contours of (z_{1_p}, z_{2_p})'s induced prior density. The transformations used are $\sigma^2_{l_p} = 1/\tau_{l_p}$, $\lambda_p = \tau_{1_p} + \tau_{2_p}$, and $\beta_p = \tau_{1_p}/(\tau_{1_p} + \tau_{2_p})$.

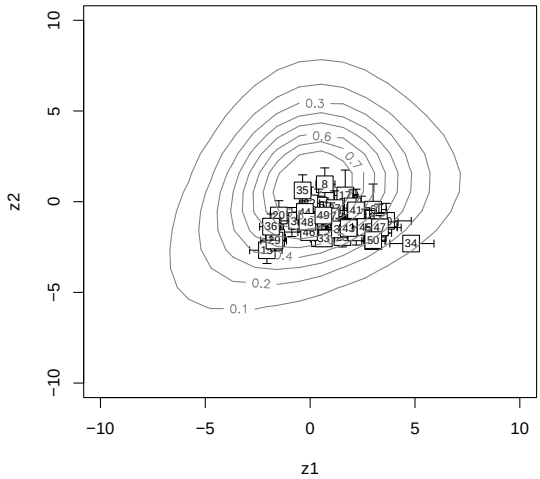
(a) $z_{l_p} \sim$ Unif(-10,10), $l \in \{0, 1, 2\}$



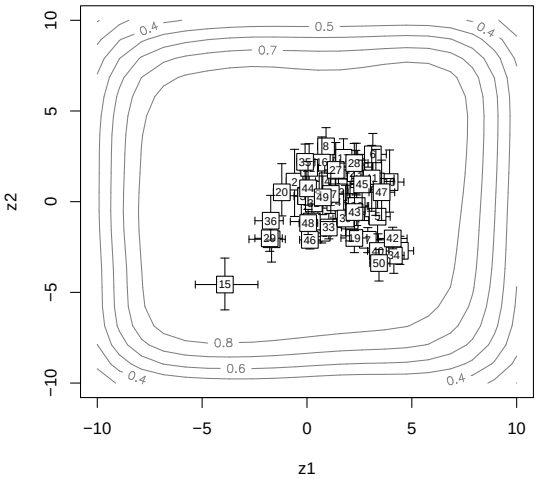
(b) $z_{l_p} \sim$ LG(.01,.01), $l \in \{0, 1, 2\}$



(c) $\sigma_{l_p} \sim$ U(0,10), $l \in \{0, 1, 2\}$



(d) $\sigma^2_{l_p} \sim$ IG(0.01,0.01), $l \in \{0, 1, 2\}$



(e) $\tau_{0_p} \sim$ G(.01,.01), $\lambda_p \sim$ G(.01,.01), $\beta_p \sim$ U(0,1)

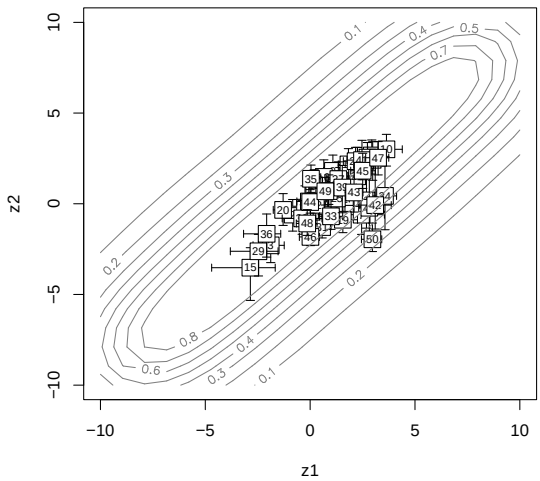


Figure 9: Data (symbols) and posterior mean of $X_1\beta+\boldsymbol{\theta}_1$ for Patient 1's left maxillary island under Grid A assuming $z_{l_p} \sim$ Uniform(-10,10) (solid lines) and $\tau_{0_p} \sim$ G(.01,.01), $\lambda_p \sim$ G(.01,.01), and $\beta_p \sim$ U(0,1) (dashed lines), each independent across $l \in \{0,1,2\}$ and $p \in \{1,...,50\}$. "Buccal" and "lingual" refer to the cheek and the tongue sides of the teeth, respectively.

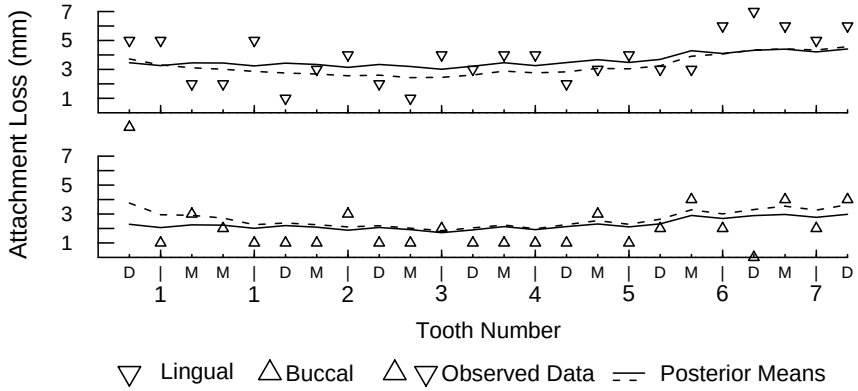


Figure 10: Plot of each patient's DIC_p under grids A and D.