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

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

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

brianr@biostat.umn.edu hodges@ccbr.umn.edu brad@biostat.umn.edu

Correspondence Author: Bradley P. Carlin
telephone: (612) 624-6646
fax: (612) 626-0660

April 28, 2004

[1]The second author was supported by the University of Minnesota School of Dentistry. 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 Types of Neighbor Relations

Abstract

Attachment loss data are used to measure the current state of a patient's periodontal disease and can be analyzed using a conditionally autoregressive (CAR) prior distribution, which smooths fitted values toward neighboring values. However, in this setting it is desirable to have more than one type of neighbor relation in the spatial structure, so the different types 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 modeling the spatial structure of these data may improve monitoring of periodontal disease progrssion. 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 the smoothing parameters often has long upper tails or multiple modes, and its shape can change dramatically depending on the spatial structure and true values of the parameters. We show that the prior distribution on these parameters has a marked effect on the posterior of both the means and the smoothing parameters; in some cases, the conventional Gamma(0.01,0.01) prior almost completely overrides the data. Our analysis of attachment loss data also suggests that the spatial structure itself may vary between individuals.

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

1 Introduction

Bayesian analyses of areal data often use a conditionally autoregressive (CAR) distribution, popularized for Bayesian disease mapping by Besag et al. (1991) and defined as follows. In a map with n regions, suppose each region has an unknown quantity of interest θ_i, $i = 1, 2, \ldots, n$, and that the map's spatial structure is described by a lattice of neighbor relations among regions. The CAR model 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 positive parameter τ controls smoothing induced by this prior, larger values smoothing more than smaller; G is the number of "islands" (disconnected groups of regions) in the spatial structure (Hodges et al., 2003); $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$; 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.

It can be desirable to have two or more types of neighbor relations in the spatial structure, with the l^{th} type having its own τ_l so the different neighbor-relation types can induce different degrees of smoothing. For example, in an agricultural field trial, the fertilities of a grid of plots might be smoother in one grid direction than in the other; in spatio-temporal data, τ_1 might smooth spatially while τ_2 smooths over time. Besag and Higdon (1999) introduced CAR priors with two types of neighbor relations, modeling a rectangular grid of plots with different smoothing parameters for row and column neighbors, parameterized as $\beta\lambda$ and $(1-\beta)\lambda$. They also extended the model to adjust for edge effects and to analyze data from multiple experiments.

Such models are also natural for analyzing a common measure of periodontal disease status,

attachment loss, which is the extent of a tooth's root (in millimeters) that is no longer attached to surrounding bone by periodontal ligament. Attachment loss is used to assess the cumulative damage to a patient's periodontium and to see if treatment halts disease progression. (Many texts, e.g., Darby and Walsh 1995, describe periodontal measurement.) Figure 1 shows the six sites on each tooth where attachment loss is usually measured. In this spatial lattice, θ_i is the true attachment loss at site i. A full-mouth set of measurements has up to 168 measurements per subject with at least two islands (one per jaw); missing teeth can create more islands. The subject whose data are plotted in Figure 9 is missing tooth number 2 on the left side of the maxilla (upper jaw), resulting in three islands. Calibration studies commonly show that a single attachment loss measurement has an error with a standard deviation of roughly 0.75 to 1 mm. Figure 9 shows a severe case of periodontal disease, so measurement error with a 1 mm standard deviation is substantial.

Attachment loss measurements are spatially correlated, but their correlation may not be simply a function of distance. Figure 1 identifies four types of neighbor pairs, labeled I-IV. Previous studies (e.g., Sterne et al. 1988, Gunsolley et al. 1994, Roberts 1999) suggest that the four neighbor types may have different correlations. Thus, appropriate statistical modeling of these data may require different degrees of smoothing for different types of neighbor pairs.

This paper considers models with two types. As possible models for attachment loss, we consider four neighborhood structures ("grids") defined in Table 1. The first grid (1NR) allows only one type of neighbor pair 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 neighbors. Finally, Grid C distinguishes type II neighbor pairs from the other types.

Spatial analyses of periodontal data can potentially serve several purposes. In research, it can

Table 1: Neighbor pairs controlled by each smoothing parameter for each grid. Figure 1 defines the types.

	Grid	Type I	Type II	Type III	Type IV
One type of neighbors	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

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. 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.

This paper presents the first analyses of periodontal data using spatial methods. The data come from an industry-sponsored randomized clinical trial of an anti-inflammatory gel (with three different drug concentrations versus placebo versus no gel) conducted at the University of Minnesota's Dental School. The original analysis used whole-mouth average of clinical measures (including attachment loss) or averages of subsets of sites defined by baseline disease status. These classical, non-spatial analyses found no treatment effect (Shievitz 1997).

We explore the case of observables with additive Gaussian errors having mean zero and precision

τ_e, a reasonable model of attachment loss. Our analyses were initially hampered by substantial technical problems, such as MCMC autocorrelations near one for the precision parameters, requiring us to more carefully consider identification in these models. Section 2 derives the marginal posterior density of (r_1, r_2), for $r_l = \tau_l/\tau_e$, $l = 1, 2$, while Sections 3 and 4 examine identification of the r_l. Section 5 applies these insights to periodontal data from three patients. The spatial structure of the data appears to vary considerably among patients, though grids with two neighbor relations are clearly superior to the 1NR grid for some patients. The choice of grid can have noteworthy effect on $\boldsymbol{\theta}$'s posterior. Section 5 also shows that for some spatial structures, increasing sample size has little effect on identification. Section 6 considers the effect of (r_1, r_2)'s prior, and Section 7 concludes. Overly technical results are relegated to an appendix.

2 CAR with two types of neighbor relations

A CAR prior with two types of neighbor relation can be written, extending (1), as

$$p(\boldsymbol{\theta}|\tau_1, \tau_2) \propto c(\tau_1, \tau_2)^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{\theta}'\{\tau_1 Q_1 + \tau_2 Q_2\}\boldsymbol{\theta}\right), \tag{2}$$

where Q_l and τ_l respectively describe and control smoothing of type l neighbor pairs, and $c(\tau_1, \tau_2)$ is the product of the positive eigenvalues of $\tau_1 Q_1 + \tau_2 Q_2$ (see below). We assume a pair of regions are neighbors of at most one type. Q_l has rank $n - G_l$, G_l being the number of islands in neighbor type l's spatial structure; assume $G_l < n$, i.e., the l^{th} neighborhood structure is not null. If G is the number of islands in the combined spatial structure, $G_l \geq G$.

Appendix Section A.1 derives the following results. For any Q_1 and Q_2, there is 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 (Newcomb, 1961). Call D_l's diagonal elements d_{li} and without loss of

generality assume the last G diagonal elements of both D_l are zero. Then

$$p(\boldsymbol{\theta}|\tau_1,\tau_2) \propto \prod_{j=1}^{n-G} (\tau_1 d_{1j} + \tau_2 d_{2j})^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{\theta}'\{\tau_1 Q_1 + \tau_2 Q_2\}\boldsymbol{\theta}\right). \tag{3}$$

Let y_i be region i's observation, and let $y_i|\theta_i,\tau_e$ be independent normal with mean θ_i and precision τ_e. Then given $\mathbf{y} = (y_1, ..., y_n)'$, the unknowns have joint posterior

$$\begin{aligned} p(\boldsymbol{\theta},\tau_e,\tau_1,\tau_2|\mathbf{y}) &\propto p(\tau_e,\tau_1,\tau_2)\tau_e^{n/2} \prod_{j=1}^{n-G} (\tau_1 d_{1j} + \tau_2 d_{2j})^{1/2} \\ &\times \exp\left(-\frac{1}{2}\{\tau_e(\mathbf{y}-\boldsymbol{\theta})'(\mathbf{y}-\boldsymbol{\theta}) + \boldsymbol{\theta}'(\tau_1 Q_1 + \tau_2 Q_2)\boldsymbol{\theta}\}\right), \end{aligned} \tag{4}$$

where $p(\tau_e,\tau_1,\tau_2)$ is the prior for (τ_e,τ_1,τ_2).

Next, reparameterize to $r_l = \tau_l/\tau_e$ and integrate $\boldsymbol{\theta}$ and τ_e out of (4), leaving the marginal posterior of the smoothing parameters (r_1, r_2). If the τs have independent gamma priors, **i.e.**, $p(\tau_e,\tau_1,\tau_2) \propto \prod_{k\in\{e,1,2\}} \tau_k^{a_k-1}\exp(-\tau_k b_k)$ for $a_k, b_k > 0$, the marginal posterior is

$$p(r_1,r_2|\mathbf{y}) \propto \prod_{j=1}^{n-G} (r_1 d_{1j} + r_2 d_{2j})^{1/2} |I_n + r_1 Q_1 + r_2 Q_2|^{-1/2} r_1^{a_1-1} r_2^{a_2-1} R_*^{-a}, \tag{5}$$

where $R_* = b_e + r_1 b_1 + r_2 b_2 + \{\mathbf{y}'\mathbf{y} - \mathbf{y}'(I_n + r_1 Q_1 + r_2 Q_2)^{-1}\mathbf{y}\}/2$ and $a = (n-G)/2 + a_e + a_1 + a_2$. If instead the prior is $p(\tau_e, r_1, r_2) \propto \tau_e^{a_e-1}\exp(-\tau_e b_e)p(r_1, r_2)$, the marginal posterior takes the more general form

$$p(r_1,r_2|\mathbf{y}) \propto p(r_1,r_2) \prod_{j=1}^{n-G} (r_1 d_{1j} + r_2 d_{2j})^{1/2} |I_n + r_1 Q_1 + r_2 Q_2|^{-1/2} R_*^{-a}, \tag{6}$$

where $R_* = b_e + \frac{1}{2}\{\mathbf{y}'\mathbf{y} - \mathbf{y}'(I_n + r_1 Q_1 + r_2 Q_2)^{-1}\mathbf{y}\}$ and $a = (n-I)/2 + a_e$. We might also consider a further transformation to (z_1, z_2), for $z_l = log(r_l)$, to spread out $p(r_1, r_2|\mathbf{y})$'s mass, which tends

to concentrate on small r_1 and r_2. This parameterization allows Gaussian or t priors, which more naturally capture vague prior information and allow z_1 and z_2 to be correlated *a priori*.

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

The conditional density of the smoothing parameters (τ_1, τ_2) can be re-expressed to highlight identification issues. As in (3), assume a non-singular B such that $Q_1 = B'D_1B$ and $Q_2 = B'D_2B$, D_1 and D_2 being diagonal. For $\boldsymbol{\theta}^* = B\boldsymbol{\theta}$, $p(\tau_1, \tau_2|\boldsymbol{\theta}^*) \propto$

$$p(\boldsymbol{\theta}^*|\tau_1, \tau_2)p(\tau_1, \tau_2) \propto \prod_{j=1}^{n-G} \left[(d_{1j}\tau_1 + d_{2j}\tau_2)^{1/2} \exp(-\frac{1}{2}\theta_j^{*2}\{d_{1j}\tau_1 + d_{2j}\tau_2\})\right] p(\tau_1, \tau_2). \tag{7}$$

Combine terms with identical (d_{1j}, d_{2j}), label the M distinct pairs (d_{1j}, d_{2j}) as $\mathbf{s}_1, ..., \mathbf{s}_M$ for $\mathbf{s}_i = (s_{1i}, s_{2i})$, and let j belong to the set S_i if $(d_{1j}, d_{2j}) = \mathbf{s}_i$. Denoting $\gamma_i = s_{1i}\tau_1 + s_{2i}\tau_2$, (τ_1, τ_2) has conditional density

$$p(\tau_1, \tau_2|\boldsymbol{\theta}^*) \propto \left[\prod_{i=1}^{M} \gamma_i^{m_i/2} \exp\left(-\frac{\gamma_i}{2} \sum_{k \in S_i} \theta_k^{*2}\right)\right] p(\tau_1, \tau_2), \tag{8}$$

where m_i is S_i's cardinality. 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 M linear combinations $\{\gamma_i\}$. The conditional densities $p(r_1, r_2|\boldsymbol{\theta}^*)$ and $p(z_1, z_2|\boldsymbol{\theta}^*)$ are also functions of (r_1, r_2) or (z_1, z_2) respectively only through the prior and M linear functions of (r_1, r_2) or (e^{z_1}, e^{z_2}); we omit details.

The i^{th} term of the product in (8) is constant for (τ_1, τ_2) satisfying $s_{1i}\tau_1 + s_{2i}\tau_2 = c$ for $c > 0$, so individual terms in the product do not identify τ_1 and τ_2. Rather, identification arises from multiplying terms with different ratios s_{1i}/s_{2i}. If there are two or more distinct ratios s_{1i}/s_{2i}, 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 $\gamma_i^{m_i/2}\exp(-\gamma_i\frac{1}{2}\sum_{k\in S_i}\theta_k^{*2})$ has the form of a gamma density with variate $\gamma_i = s_{1i}\tau_1 + s_{2i}\tau_2$, mode $m_i/\sum_{k\in S_i}\theta_k^{*2}$, and an infinite set of modal (τ_1,τ_2) satisfying $s_{1i}\tau_1 + s_{2i}\tau_2 = m_i/\sum_{k\in S_i}\theta_k^{*2}$. Terms with $s_{1i}\neq 0$ and $s_{2i}\neq 0$ give non-identified modal lines $\tau_2 = -\tau_1 s_{1i}/s_{2i} + m_i/s_{2i}\sum_{k\in S_i}\theta_k^{*2}$. Only the intercepts of these lines depend on $\boldsymbol{\theta}$; the slopes, $-s_{1i}/s_{2i}$, do not.

Each term in (8) can be deemed a *free term* or a *mixed term* depending on (s_{1i}, s_{2i}). We define the i^{th} term to be a *free term for* τ_1 if $s_{2i}=0$ and $s_{1i}\neq 0$, 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 $s_{1i}\neq 0$ and $s_{2i}\neq 0$. Free terms are important because, as Sections 4 and 5 show, they greatly enhance identification. 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.

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 tend to induce negative correlation between τ_1 and τ_2 or between r_1 and r_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 $\Delta_{ab} = \sum_{j=1}^{N-G}\frac{d_{aj}d_{bj}}{(d_{1j}\tau_1+d_{2j}\tau_2)^2}$. Similarly, $corr(r_1,r_2|\boldsymbol{\theta}) \approx -\Gamma_{12}/\sqrt{\Gamma_{11}\Gamma_{22}}$ where $\Gamma_{ab} = \sum_{j=1}^{N-G}\frac{d_{aj}d_{bj}}{(d_{1j}r_1+d_{2j}r_2)^2}$. These approximate correlations are never positive, but the marginal posterior correlation $corr(r_1,r_2|\mathbf{y})$ can be (Section 4.1 gives an example).

4 Exploring $p(r_1, r_2|\mathbf{y})$ and $p(z_1, z_2|\mathbf{y})$

No tidy expression like (8) is available for $p(\tau_1,\tau_2|\mathbf{y})$ except in special cases. Of course, as each m_i increases, $p(\tau_1,\tau_2|\mathbf{y})$ should more closely resemble $p(\tau_1,\tau_2|\boldsymbol{\theta})$ for the true $\boldsymbol{\theta}$. An MCMC algorithm draws from $p(\tau_1,\tau_2|\tau_e,\boldsymbol{\theta},\mathbf{y}) = p(\tau_1,\tau_2|\boldsymbol{\theta})$; as $\boldsymbol{\theta}$ varies between iterations, the intercepts

of the non-identified lines, and hence $p(\tau_1, \tau_2|\boldsymbol{\theta})$'s shape, change with $\boldsymbol{\theta}$. Still, the (s_{1i}, s_{2i}) help explain $p(\tau_1, \tau_2|\mathbf{y})$.

To explore (τ_1, τ_2)'s identifiability, we consider three regular neighborhood structures that shed light on more complex structures like the periodontal grids. The first is a rectangular grid with column and row neighbor relations, which we call a *BH grid* (Figure 2a), referring to Besag and Higdon (1999). The grid can be spatial; it can also be a simple space-time smoother, each column representing the system at one time, so smoothing of row neighbors corresponds to smoothing over time. Multiple islands of this type might arise from different experiments or locations. This model has many free terms for τ_1 and τ_2, which are well-identified.

In the second regular structure, each neighbor relation forms a connected graph (Figure 2b). This leaves no free terms for τ_1 or τ_2, even with multiple islands of this structure, and τ_1 and τ_2 are only weakly identified. We call this the *no-free-terms* grid.

The third regular structure has two offset rows with one neighbor relation for horizontal neighbors in the same row and a second neighbor relation for diagonal neighbors in different rows (Figure 2c). This gives one free term per island for the diagonal smoothing parameter τ_1 but none for τ_2, indicating a single free term's effect. We call this the *horizontal/diagonal grid.*

Equation (8) and simulated data from these structures give insight into the effects of free and mixed terms on identification of (τ_1, τ_2). For each structure, we consider datasets with about 200 observations, generated using $\tau_e = 1$. A Gamma(0.01,0.01) prior (mean 1, variance 100) was used for τ_e, r_1, and r_2, as suggested by Best et al. (1999), and the corresponding log Gamma(0.01,0.01) for z_1 and z_2. As will become clear, the prior on (r_1, r_2) or (z_1, z_2) is very important. With each contour plot of $p(r_1, r_2|\mathbf{y})$ or $p(z_1, z_2|\mathbf{y})$, we present a graph of the M lines of unidentifiability, e.g., the set of (x_1, x_2) satisfying $s_{1i}e^{x_1} + s_{2i}e^{x_2} = s_{1i}e^{z_1} + s_{2i}e^{z_2}, i = 1, ..., M$ for the true (z_1, z_2).

4.1 BH Grid

Let τ_1 and τ_2 control smoothing of row and column neighbors respectively. Consider the grid in Figure 2a with two rows and two columns. This CAR prior has, *a priori*, precision

$$\Sigma^{-1}(\tau_1,\tau_2) = \tau_1 Q_1 + \tau_2 Q_2 = \tau_1 \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} + \tau_2 \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}. \quad (9)$$

A convenient diagonalizing matrix B is

$$B = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad \text{giving} \quad \boldsymbol{\theta}^* = B\boldsymbol{\theta} = \begin{pmatrix} (\theta_1+\theta_3)-(\theta_2+\theta_4) \\ (\theta_1+\theta_2)-(\theta_3+\theta_4) \\ (\theta_1-\theta_2)-(\theta_3-\theta_4) \\ \theta_1+\theta_2+\theta_3+\theta_4 \end{pmatrix}. \quad (10)$$

Thus $B\Sigma^{-1}B'$, $\boldsymbol{\theta}^*$'s precision matrix, is proportional to a diagonal matrix with diagonal entries τ_1, τ_2, $\tau_1+\tau_2$, and 0.

The first element of $\boldsymbol{\theta}^*$ is the difference between column totals, proportional to the column main effect. Its precision depends only on τ_1, giving a free term for τ_1. Similarly, $\boldsymbol{\theta}^*$'s second element is proportional to the row main effect, giving a free term for τ_2. The third element is like the interaction in a two-way analysis of variance (ANOVA) and gives a mixed term. Finally, $\sum\theta_j$ has precision zero, which is always true for a CAR prior.

For general BH grids, the counts of free and mixed terms equal the counts of degrees of freedom in the analogous two-way ANOVA: c columns and r rows give $c-1$ free terms for τ_1, $r-1$ free terms for τ_2, and $(c-1)(r-1)$ mixed terms. Multi-island structures combine by adding numbers of free and mixed terms across islands. Here, as for all spatial structures considered in this paper, a free term in τ_l arises from the difference between two groups of regions such that all neighbor

relations between a region in one group and a region in the other group are of type l.

Figure 3b shows the contour plot of $\log p(r_1, r_2|\mathbf{y})$ for a simulated data set on a BH grid, generated with $r_1 = r_2 = 1$ and twelve 4×4 islands, for 192 total observations. Figure 3a shows the non-identified lines evaluated at the true (r_1, r_2). Each line has $m_i = 12$ (one per island) except the vertical and horizontal lines, corresponding to the free terms in r_1 and r_2 respectively, which have $m_i = 36$ (three per island).

The smoothing parameters are reasonably well-identified, as each of the 12 islands gives three free terms for each r_l. Figure 3b suggests a positive correlation of $(r_1, r_2)|\mathbf{y}$, near the mode at least. As noted, r_1 and r_2 are generally negatively correlated given $(\boldsymbol{\theta}, \tau_e)$, but as $(\boldsymbol{\theta}, \tau_e)$ varies, the intercepts of the non-identified lines vary, giving in this case a marginal correlation of 0.02.

To illustrate further, we generated 20,000 samples from (r_1, r_2)'s marginal posterior (Figure 4). Conditional on each MCMC iteration's $\boldsymbol{\theta}$, we computed the centroid of all intersections of pairs of non-identified lines, to indicate the center of (r_1, r_2)'s conditional density. A plot of the centroids is visually indistinguishable from Figure 4. That is, (r_1, r_2) has high precision given $(\boldsymbol{\theta}, \tau_e)$, but its conditional location varies greatly between iterations (recall that the intercepts of the non-identified lines depend on $\boldsymbol{\theta}$ while the slopes do not). Figure 4's horizontal axis shows a boxplot of the 20,000 x-intercepts of the non-identified line corresponding to r_1's free terms; the vertical axis has an analogous boxplot for r_2. Although $m_i = 36$ for r_1's and r_2's free terms, these lines vary a great deal between iterations. In Figure 3b, the conditional distribution of r_1 given any r_2 resembles this boxplot, suggesting that the free terms provide most of the information about r_1, and similarly for r_2.

Figure 3 shows that $p(r_1, r_2|\mathbf{y})$ is highly skewed with most of its mass near zero. Transforming to $z_l = log(r_l)$ distributes this mass over the real line, giving a more symmetric posterior with "bicycle seat" contours (Figure 5).

4.2 No-Free-Terms Neighborhood Structure

This structure (Figure 2b) has no free terms for τ_1 or τ_2; that is, in a given no-free-terms island, both neighbor relations define connected graphs. It is like the BH grid but with diagonal neighbor pairs added. The data for Figure 6 were generated with $\tau_e = 1$, $r_1 = r_2 = 10$ ($z_1 = z_2 = 2.3$) and two 10×10 islands, giving 200 total observations. Figure 6b's most striking feature is the surface's flatness along the elbow-shaped curve $e^{z_1} + e^{z_2} = e^{2.3} + e^{2.3}$ (i.e., $r_1 + r_2 = 10 + 10$, the true $r_1 + r_2$). This is the non-identified curve for the term in $p(z_1, z_2|\boldsymbol{\theta})$ with $s_{1i} = 1$ and $s_{2i} = 1$, evaluated at the true (z_1, z_2). In Figure 6a, this curve is roughly in the middle of the M curves, which are distinct but have similar shapes. Thus, the data provide some information about $e^{z_1} + e^{z_2}$ (i.e., $r_1 + r_2$), but very little about z_1 and z_2 individually.

For $z_1 < 0$ (little smoothing of type 1 neighbors) the data are informative about z_2 and the posterior has a well-defined peak near $z_2 = 3$. For these small z_1, smoothing of type 1 neighbors is negligible, so pairs of type 2 neighbors are only affected by z_2, allowing the data to rule out extreme z_2. But for large z_1 (much smoothing of type 1 neighbors), the data say little about z_2 because large z_1 forces all type 2 neighbors to have similar θ_i. Thus, for large z_1, the posterior rules out only large z_2, which would smooth *more* than indicated by z_1, while the posterior is quite flat for $z_2 < 2$. That is, a large z_1 precludes differentiating between smaller z_2.

4.3 Horizontal/Diagonal Grid

To examine the effect of a few free terms, consider the structure in Figure 2c. Neighbor relation 1 (τ_1, r_1, z_1) smooths diagonal neighbor pairs with one neighbor in each row, while neighbor relation 2 (τ_2, r_2, z_2) smooths horizontal neighbor pairs in the same row. The difference between the two row sums has precision depending only on τ_1 (i.e., r_1, z_1), giving one free term per island for τ_1. The horizontal smoothing parameter τ_2 has no free terms. Figures 7 and 8 give curves of

non-identifiability and contour plots of $\log p(z_1, z_2|\mathbf{y})$ for two datasets generated with different true (z_1, z_2). Each dataset had two islands, each with two rows of length 50, for 200 total observations and two free terms total for z_1.

Figure 7's data had true $z_1 = 2.3$ and $z_2 = -4.6$, i.e., smoothing substantially for diagonal neighbors (z_1) and little for horizontal neighbors (z_2). As for the no-free-terms data (Figure 6), $p(z_1, z_2|\mathbf{y})$ has most of its mass along $e^{z_1} + e^{z_2} = e^{2.3} + e^{-4.6}$. For large z_1, diagonal neighbors are very similar; because they form a connected graph, horizontal neighbors are forced to be similar regardless of z_2. Thus, for large z_1, the data cannot differentiate small z_2 so the posterior is flat for small z_2. The data *are* informative about z_1 through the difference between the two row sums, that is, through z_1's free term. Hence z_1 near the true value have high posterior density, giving golf-club-shaped contours.

By contrast, Figure 8's data had $z_1 = -4.6$ and $z_2 = 2.3$, i.e., little smoothing of diagonal neighbors (z_1). The posterior now has a very different shape and the z_l are well-identified. Figure 8 shows posterior mass on the curve for which $e^{z_1} + e^{z_2}$ equals its true value, including an unusual "arm" of probability to the right of the main peak. Unlike the no-free-terms case with the same true (z_1, z_2) (Figure 6), this arm is over 30 log units below the peak. For the no-free-terms model z_2 was well-identified for small z_1 but not large z_1. For the diagonal/horizontal grid, z_1's two free terms center the posterior mass around the true z_1 and because z_1 is small, the data rule out both extremes for z_2.

5 Analysis of Periodontal Data

This section analyzes attachment losses from three representative subjects in a recent industry-sponsered periodontal clinical trial (initially described and analyzed by Shievitz 1997). The subjects

Table 2: Posterior summaries of θ_{43} and θ_i for site #43's neighbors, for subject 1. "Type" is the site's neighbor type relative to site #43.

			Posterior mean (SD)				Posterior correlation with θ_{43}			
Site	Type	y_i	Grid A	Grid B	Grid C	1NR	Grid A	Grid B	Grid C	1NR
3	III	4	3.34(.55)	3.46(.51)	3.58(.61)	3.39(.61)	0.35	0.99	0.65	0.67
4	IV	5	3.38(.56)	3.46(.51)	3.63(.60)	3.52(.65)	0.37	0.99	0.66	0.69
42	I	3	2.76(.56)	3.09(.62)	3.10(.63)	3.10(.60)	0.73	0.37	0.38	0.47
43		5	2.88(.61)	3.46(.51)	3.48(.56)	3.45(.63)	-	-	-	-
44	II	3	2.74(.60)	3.45(.52)	3.38(.56)	3.19(.54)	0.82	0.98	0.81	0.58

are first analyzed separately, then together assuming common τ_e, τ_1, and τ_2. For τ_e, we used a gamma prior with $a_e = b_e = 0.01$; for z_1 and z_2, independent Unif(−15,15) priors. For each subject and candidate grid, structured MCMC (Sargent et al., 2000) with blocks $\boldsymbol{\theta}$ and (τ_e, z_1, z_2) was used to make 30,000 posterior draws. Convergence was assessed by comparing summaries of the (z_1, z_2) draws to contour plots of the exact posterior of (z_1, z_2) given by Equation (6).

5.1 Data Analysis

Figure 9 plots Subject 1's data $\mathbf{y}$ (symbols) and posterior mean for $\boldsymbol{\theta}$ (solid lines) for each grid. The grid noticeably affects $\boldsymbol{\theta}$'s posterior mean: Grids B, C and 1NR give similar posterior means, but often differ from Grid A by over 0.5 mm. For Grid A, draws of z_1 are generally larger than draws of z_2 (Figure 10a); 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 the islands. Large z_1 and small z_2 smooth substantially within these long strips but not between them. Grids B, C and 1NR smooth more between these strips, and z_1 tends to be less than z_2 (Figures 10b,c,d).

The grid also affects $\boldsymbol{\theta}$'s posterior covariance. Table 2 gives posterior summaries for subject 1 for site #43 and its four neighbors. Only for Grid A is θ_{43} highly correlated with θ_{42}, its type I neighbor, which has smaller posterior standard deviation for Grid A than for the other grids.

For Grid B, neighbor pairs controlled by z_2 (types II, III, and IV) have correlation near 1 and nearly identical means and standard deviations, consistent with Figure 10b, where posterior mass concentrates on large z_2 and moderate z_1. The large correlations give these θ_i relatively small standard deviations. The covariances under Grid C and 1NR are similar except for type II pairs, which have larger correlation under Grid C because they have their own smoothing parameter z_1. Otherwise, however, z_2 for Grid C has marginal posterior much like z for 1NR (Figures 10c,d).

We compared the four grids using the deviance information criterion (DIC) of Speigelhalter et al. (2002). Defining $D(\boldsymbol{\theta}, \tau_e, z_1, z_2) = -2 \log f(y|\boldsymbol{\theta}, \tau_e, z_1, z_2)$, $DIC = \bar{D} + p_D$ where $\bar{D} = E(D(\boldsymbol{\theta}, \tau_e, z_1, z_2)|\mathbf{y})$ and $p_D = \bar{D} - D(E(\boldsymbol{\theta}, \tau_e, z_1, z_2|\mathbf{y}))$, the expectations being taken with respect to the joint posterior. $\bar{D}$ and p_D measure fit and complexity, respectively; models with smaller DIC are favored.

Table 3 gives DIC and p_D for each subject and grid. A different grid minimizes DIC for each subject. Each model has 165 or 171 parameters (3 precisions plus 162 θ_i for subject 1 or 168 θ_i for the others) but the effective model size p_D is much smaller. Grid C minimizes DIC for subject 1, but the one-neighbor-relation model (1NR) has smaller DIC and larger p_D than grids A and B: the second smoothing parameter helps only if it makes the model fit the data better, as for Grid C. For subject 2, p_D is very small: $\boldsymbol{\theta}$ is smoothed almost to a constant vector. Each grid permits this fit, so the DICs are very similar. Finally for subject 3, Grids A, B, and C fit the data similarly, each having DIC markedly superior to 1NR's despite larger p_D. Comparing subjects 1 and 3, it is clear that adding the second smoothing parameter can increase *or* decrease effective model complexity (p_D) relative to the 1NR model.

If the three subjects' precisions are independent *a priori*, DIC for the fit including all three subjects is the sum of their individual DICs. Table 3 gives these DICs ("3 subjects combined; different precisions"). Grid C fits best in this average sense and all three grids with two neighbor

Table 3: DIC for the various models

	Subject 1		Subject 2		Subject 3		3 subjects combined different precisions		3 subjects combined common precision	
Grid	DIC	p_D	DIC	p_D	DIC	p_D	DIC	p_D	DIC	p_D
1NR	269.4	32.8	156.3	5.0	231.3	27.4	657.0	65.2	689.8	67.9
A	278.5	27.5	155.3	7.4	212.9	36.2	646.7	71.1	684.6	73.1
B	272.4	28.9	156.2	4.0	211.3	37.3	639.9	70.2	678.1	69.5
C	264.6	36.8	155.8	3.2	214.5	36.9	634.9	76.9	659.7	95.8

relations outperform 1NR, mainly because of subject 3.

Many of these fits have flat, unusually shaped posteriors for (z_1, z_2), shown in Figure 10 for subject 1. For Grids A and B, $p(z_1, z_2|y)$'s contours (Figure 10a,b) are L-shaped with long tails extending to large z_1 or z_2. For Grid C (Figure 10c), z_2 is well-identified but there is little information about z_1. These odd posteriors are considered further in Section 5.2.

To see whether spatial structure varies between subjects, the three subjects' data were combined in one fit assuming common precision parameters (τ_e, z_1, z_2). Table 3's final two columns ("3 subjects combined; common precisions") give DIC and p_D for this analysis. For each grid, DIC strongly favors different precisions for each subject, not surprising considering the p_D for the single-subject fits. As in the analysis with different precisions, this analysis favors Grid C and 1NR fits worst.

Figure 11 gives contour plots of the posteriors from analyzing all three subjects assuming common precisions. Compared to Figure 10, the contours are steeper, but the general shapes are unchanged. Figure 11a's most striking feature is the two modes. The three subjects' individual posterior medians for z_2 are -1.75, 0.08 and 6.09, conflicting information which, in this case, gives a bimodal posterior. More modes may be possible with a larger group of subjects.

Even when all three subjects are combined to give a fairly large dataset (498 observations),

Table 4: Free-term counts for each periodontal grid, with 3 subjects and common precisions

Grid	n	I	I_1	I_2	Free Terms for z_1	Free Terms for z_2	Mixed Terms
A	498	7	14	256	249	7	235
B	498	7	266	256	249	159	83
C	498	7	346	7	0	339	152

z_1 and z_2 are poorly identified. However, the primary focus of this type of analysis is usually $\boldsymbol{\theta}$, and even when z_1 and z_2 are not identified, MCMC analysis of $\boldsymbol{\theta}$ can be valid. Analysis of free and mixed terms helps understand and anticipate poor identification and computing methods can adapted accordingly. For example, informative priors could be given to parameters that are expected to be poorly-identified, or MCMC methods could be developed that more efficiently sample from L-shaped joint distributions.

5.2 Analysis of Periodontal Grids

This section uses the insights from Section 4 to interpret Figures 10 and 11. The shape of (z_1, z_2)'s posterior seems largely determined by free terms. For each grid with two neighbor relations, there are long upper tails at specific z_1 and z_2 arising from the free terms, e.g., at $z_1 \approx 2$ and $z_2 \approx -2$ for Grid A. The long upper tails may reflect positive skewness in the distribution of the free terms' intercepts, as for the BH grid (Figure 4). Table 4 gives counts of free and mixed terms for the three grids with two neighbor relations. Because subject 1 is missing a tooth and subjects 2 and 3 are not, their spatial structures differ and so do their $\{d_{ij}\}$ and contributions of free and mixed terms.

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. Figure 11a shows that only z_1 between 0 and 5 have log density within 5 log units of the peak, while z_2 ranging from -4 to 15 satisfy this criterion.

The many mixed terms induce some curvature of the L-shaped contours near the intersection of the lines of mass along $z_1 \approx 2$ and $z_2 \approx -2$.

Grid B has many free terms for both z_1 and z_2, more than a comparably sized BH grid, and far fewer mixed terms than Grid A. Again, z_1 has more free terms than z_2, so the density is larger along $z_1 \approx 1$ than along $z_2 \approx 1$ (Figure 11b). The relative paucity of mixed terms means little curvature in the L-shape of (z_1, z_2)'s posterior (Figure 11b).

As for the horizontal/diagonal grid (Section 4.3), Grid C's neighbor pairs controlled by z_2 (I, III and IV) form a connected graph on each island, so z_1 has no free terms. Figure 11c shows that z_2's 339 free terms make it well-identified with mean near zero, while z_1 is not well-identified, and the golf-club shape from the horizontal/diagonal grid is faintly recognizable for $z_1 < 5$ (although oriented differently).

6 Effect of the Prior Distribution

Now consider Figure 12, showing the prior's effect on $p(r_1, r_2|\mathbf{y})$ for the no-free-terms grid and the same simulated data used in Figure 6 (Section 4.2). Figure 12a shows (r_1, r_2)'s log marginal likelihood, which is extremely flat, ranging only 0.5 log units over most of the graph's domain. This is not unexpected: with no free terms, identification is poor. The likelihood is quite small near the origin and appears to have a single maximum favoring considerable smoothing for neighbor relation 1 (τ_1, r_1, z_1).

With such a flat likelihood, the prior has great influence. Figure 12b is the contour plot of $p(r_1, r_2|\mathbf{y})$ using Figure 12a's likelihood and independent Gamma(0.01,0.01) priors for r_1 and r_2. This prior puts almost all its mass near the axes and is fairly flat for large r_1 and r_2. The posterior thus has large mass near each axis but the likelihood forces the posterior to be small near the origin,

giving a mode near each axis.

Figure 6 shows the contours of $p(z_1, z_2|\mathbf{y})$ for the same data and priors (i.e., log gamma priors for z_1 and z_2). This posterior appears unimodal, though still quite flat. For many examples in this paper, $p(r_1, r_2|\mathbf{y})$ had two modes while $p(z_1, z_2|\mathbf{y})$ appeared to have one.

We have used two priors for z_1 and z_2: log Gamma(0.01,0.01) for the regular structures in Section 4 and Uniform(-15,15) for the attachment loss data in Section 5. For the latter, (z_1, z_2)'s marginal posteriors have long upper tails along lines corresponding to the free terms. The BH grid (Section 4.1) also has many free terms for z_1 and z_2 but $p(z_1, z_2|\mathbf{y})$ had no upper tails (Figure 5). This is the prior's effect. The log Gamma(0.01,0.01) prior used in Section 4.1 places little weight on z_1 or z_2 larger than 5 and eliminates the likelihood's tails. A log gamma prior has the same effect on the attachment loss analyses. Figure 13b is the contour plot for subject 1 for Grid B and log gamma priors for z_1 and z_2; compare it to Figure 10b, which used uniform priors. The log gamma priors largely override the likelihood and the density is small for z_1 or z_2 above 5, preventing fits with extremely large amounts of smoothing.

The choice of prior can also noticeably affect the fit with respect to $\boldsymbol{\theta}$. Figure 13a shows the differences between posterior means of the lingual and buccal sides of subject 1's small maxillary island under Grid B. With the uniform prior, z_2 is large (smoothing a lot), ranging from 0 to 15 *a posteriori*, so neighbors at the mesial or distal sites (D or M) of the same tooth (i.e., type III neighbor pairs controlled by z_2) are smoothed together and their difference is nearly zero. But with the log gamma prior, z_2 is forced to be smaller and these neighbor pairs differ by as much as 0.31 mm. These results seem to indicate that the Gamma(0.01,0.01) and corresponding log gamma prior are anything but non-informative or diffuse, casting doubt on their utility as reference priors for problems like this.

7 Discussion

In our periodontal data analysis, DIC sometimes favored models with two neighbor relations despite their increased complexity. Specifically, adding the second neighbor relation had a noteworthy effect on the marginal posteriors of individual θ_i for two of the patients (subjects 1 and 3). However, the spatial structure appeared to vary considerably even among these three subjects, who were haphazardly selected from the study population.

The smoothing parameters $r_l = \tau_l/\tau_e$ and $z_l = log(r_l)$ are identified except in trivial cases, but identification can be poor depending on the spatial structure and the true r_l or z_l. Free terms greatly enhance identification. For the models considered here, free terms arose from prior contrasts in $\boldsymbol{\theta}$ with precision depending on only one r_l or z_l. Generally, r_l or z_l with no free terms are poorly identified even by moderately large samples, especially if neighbor pairs of the other type are highly smoothed. This may cause computing problems such as poor MCMC convergence, which is obvious from contour plots like Figure 10a. 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. Bimodal posteriors for (r_1, r_2) arose from independent Gamma(0.01,0.01) priors, while log gamma priors for (z_1, z_2) precluded large values of these parameters and forced less smoothing on $\boldsymbol{\theta}$ than uniform priors. Among other things, this consideration weighs against using the conventional Gamma(ϵ, ϵ) as a reference prior.

The connection between free terms and identification can be extended to a variety of models. Work in progress is examining other richly-parameterized models with three unknown precisions, including crossed random-effect models and dynamic linear models. A general treatment of three-precision models may be possible.

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.

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.

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

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, 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, Univerisity 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 type 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

$$\prod_{j=1}^{n-G} (\tau_1 d_{1j} + \tau_2 d_{2j})^{1/2}. \tag{11}$$

For j with both d_{lj} positive, the j^{th} 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 types, 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: Neighbor pairs in a three-tooth periodontal grid. Bold-face squares represent teeth, small circles represent sites where attachment loss is measured, Roman numerals represent types of neighbor pairs, and Arabic numerals index measurement sites.

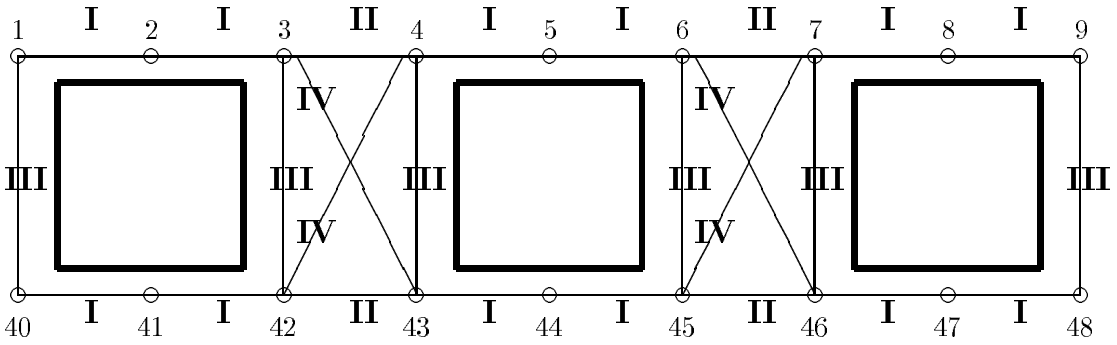
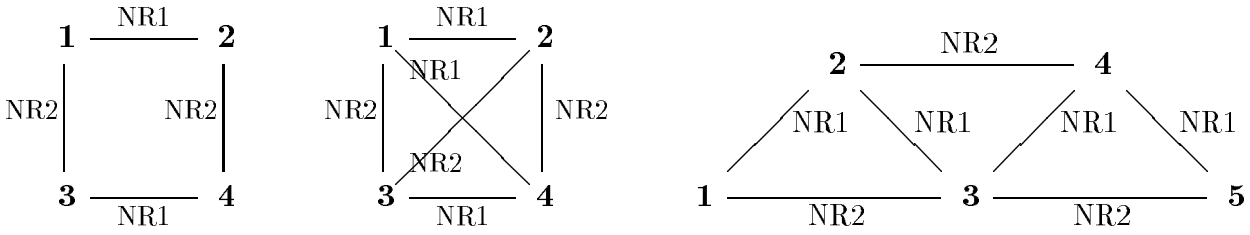
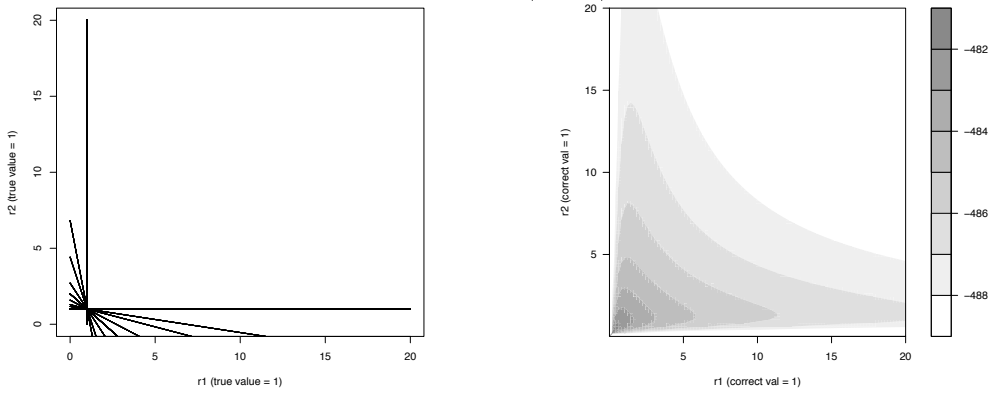


Figure 2: Neighbor relations (NR) for the regular grids in Section 4.



(a) BH Grid (b) No-Free-Terms Grid (c) Horizontal/Diagonal Grid

Figure 3: Posterior of (r_1, r_2) for BH grid



(a) Non-Identified Lines Based on True (r_1, r_2) (b) Contour Plot, Log Posterior of (r_1, r_2)

Figure 4: Scatterplot of draws from (r_1, r_2)'s posterior. Box plots show the modal non-identified lines for the free terms in τ_1 (horizontal axis) and τ_2 (vertical axis). The true $(r_1, r_2) = (1, 1)$.

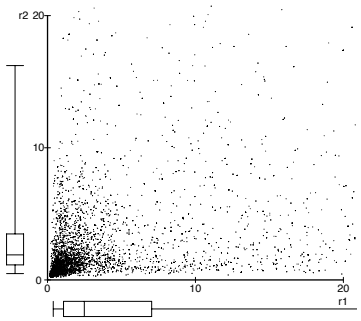
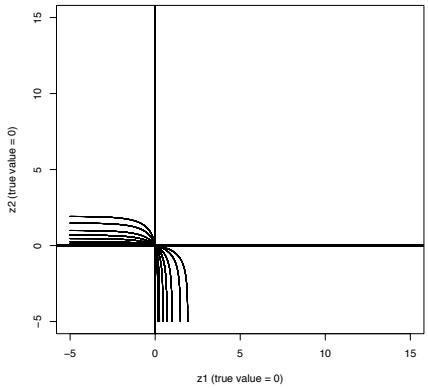
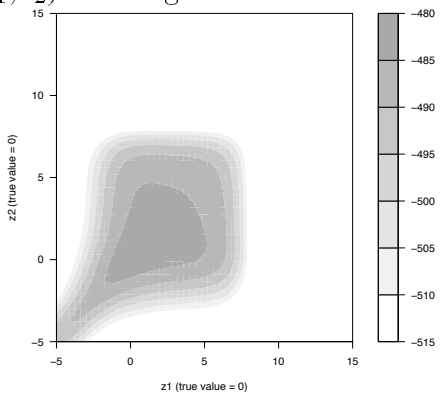


Figure 5: Posterior of (z_1, z_2) for BH grid

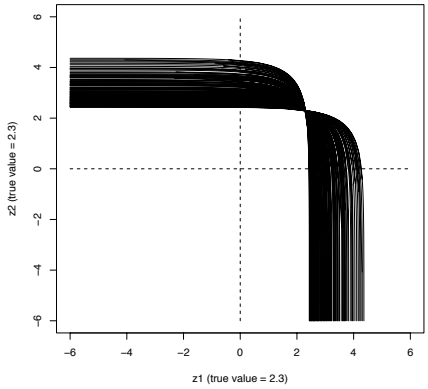


(a) Non-Identified Lines Based on True (z_1, z_2)

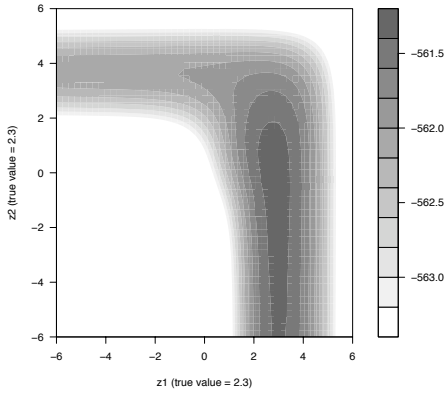


(b) Contour Plot, Log Posterior of (z_1, z_2)

Figure 6: Posterior of (z_1, z_2) for the no-free-terms grid

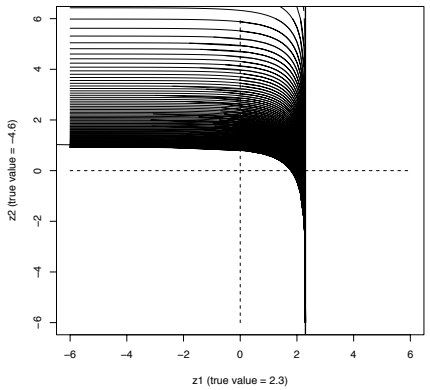


(a) Non-Identified Curves Based on True (z_1, z_2)

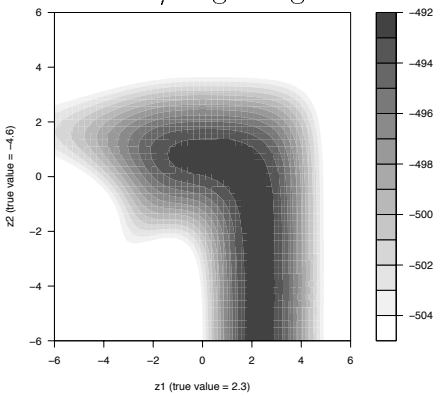


(b) Contour Plot, Log Posterior of (z_1, z_2)

Figure 7: Posterior of (z_1, z_2) for the horizontal/diagonal grid

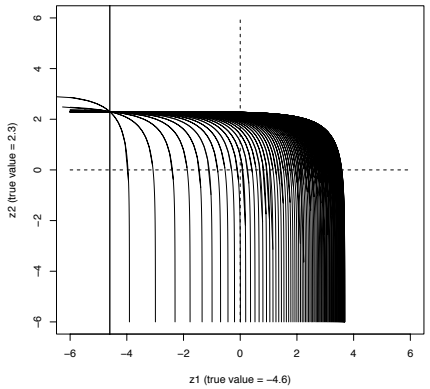


(a) Non-Identified Curves Based on True (z_1, z_2)

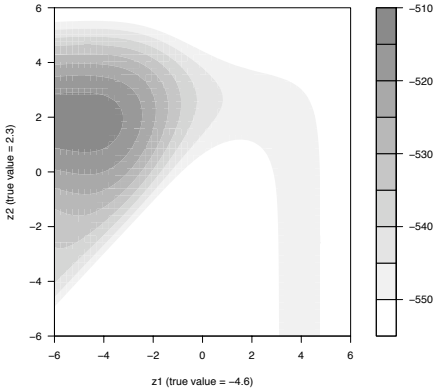


(b) Contour Plot, Log Posterior of (z_1, z_2)

Figure 8: Posterior of (z_1, z_2) for the horizontal/diagonal grid



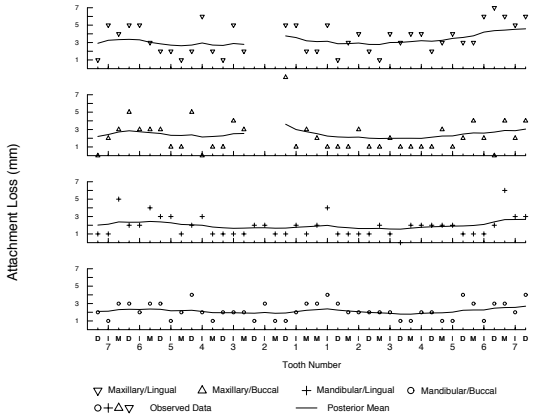
(a) Non-Identified Curves Based on True (z_1, z_2)



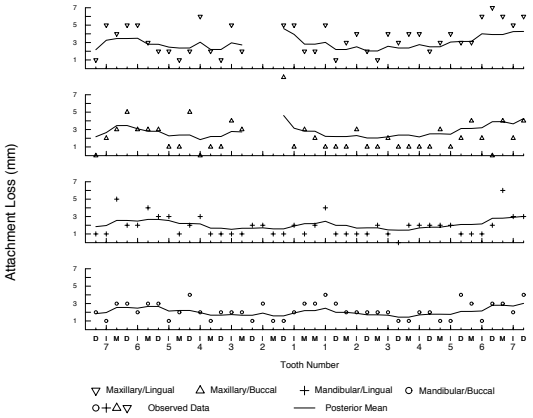
(b) Contour Plot, Log Posterior of (z_1, z_2)

Figure 9: Data and posterior means of θ_i for subject 1. "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.

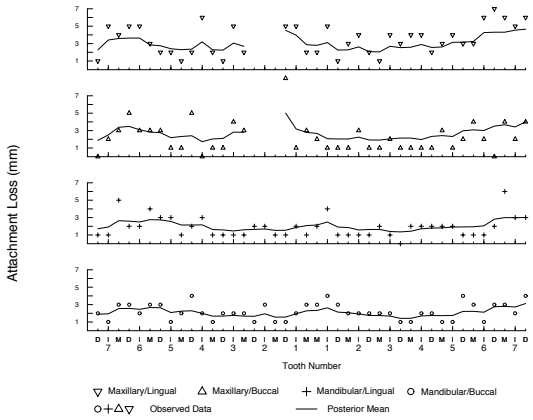
(a) Grid A



(b) Grid B



(c) Grid C



(d) 1 Neighbor Relation

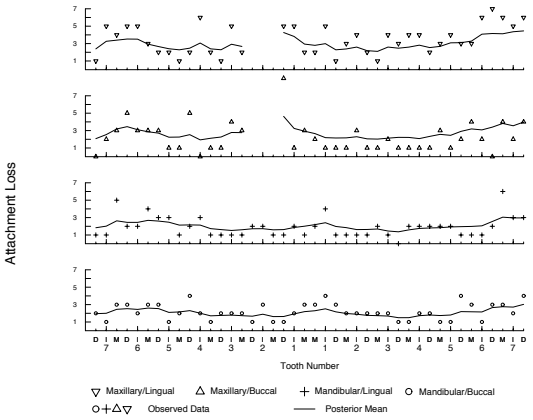
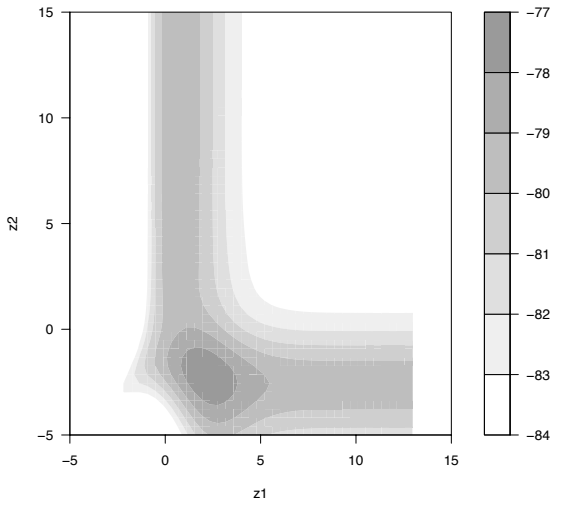
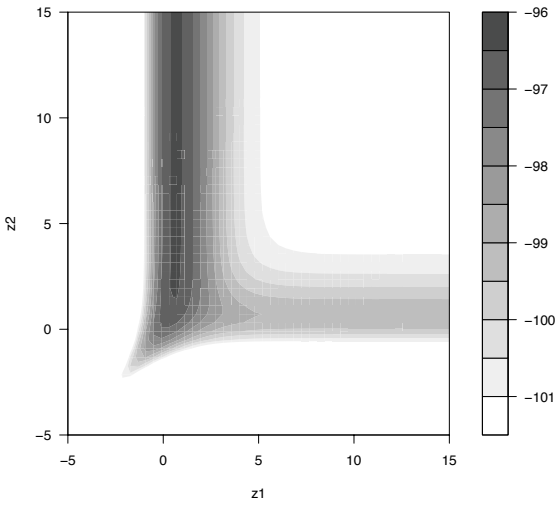


Figure 10: Posterior contour plots for subject 1. Panels (a), (b), and (c): Log marginal posterior of (z_1, z_2) for grids A, B, and C. Panel (d): Log marginal posterior of z for 1NR grid

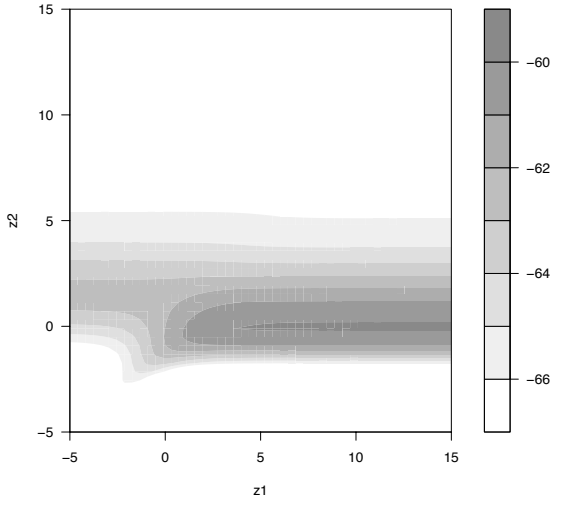
(a) Grid A



(b) Grid B



(c) Grid C



(d) 1 Neighbor Relation

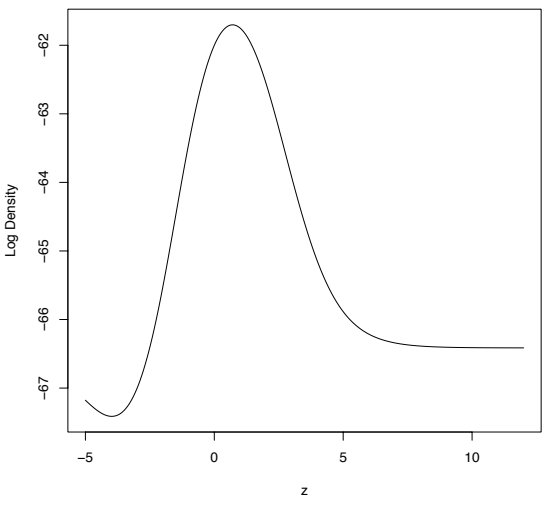
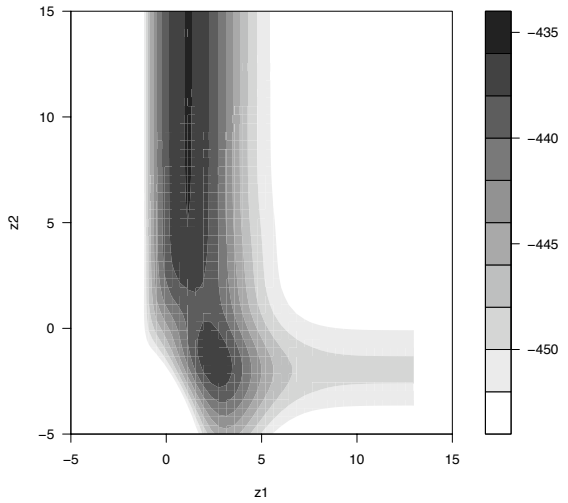
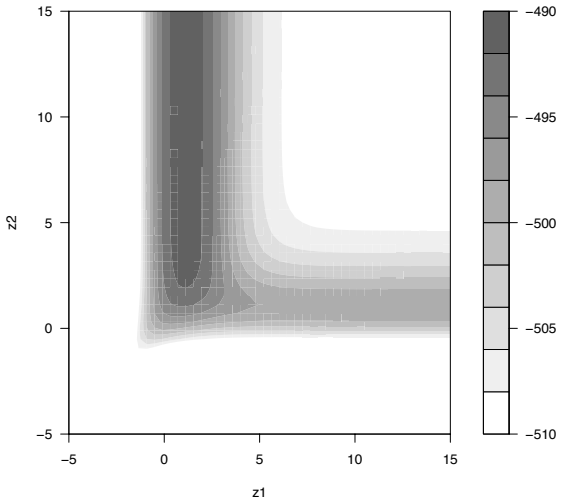


Figure 11: Log posterior of (z_1, z_2) combining all 3 subjects assuming common (τ_e, z_1, z_2)

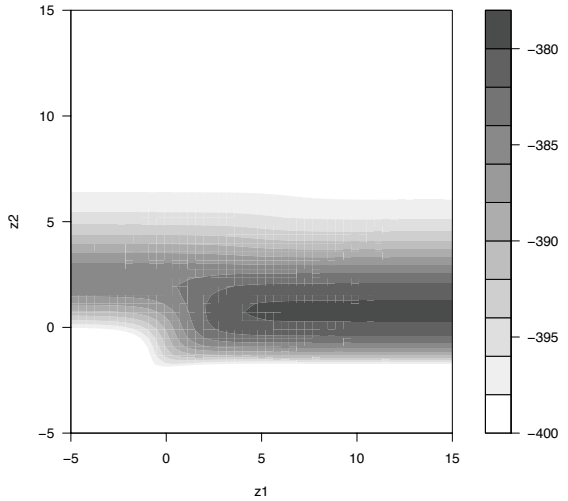
(a) Grid A



(b) Grid B



(c) Grid C



(d) 1 Neighbor Relation

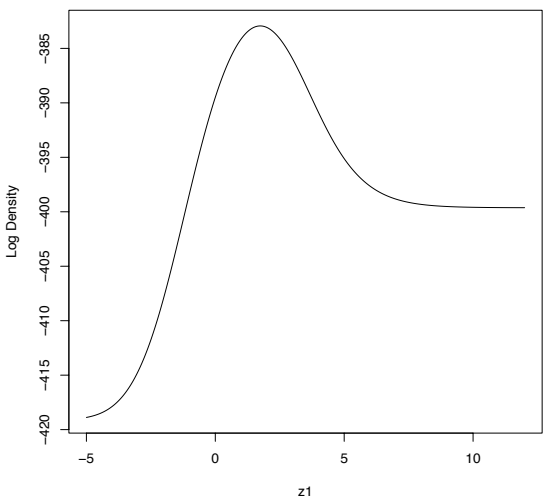


Figure 12: Marginal likelihood and posterior of (r_1, r_2) for the no-free-terms grid.

(a) Contour Plot, Log Likelihood of (r_1, r_2) (b) Contour Plot, Log Posterior of (r_1, r_2)

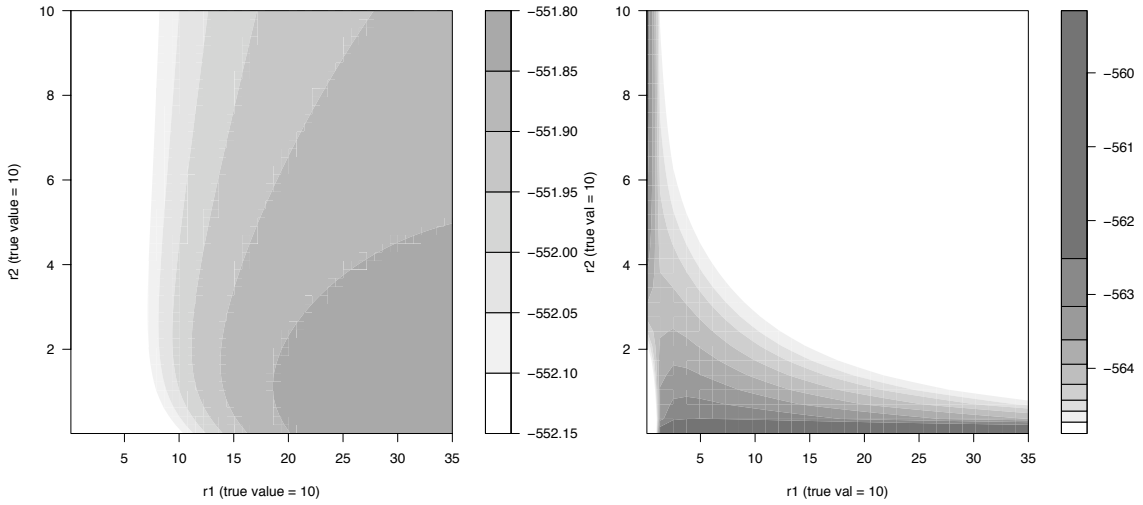
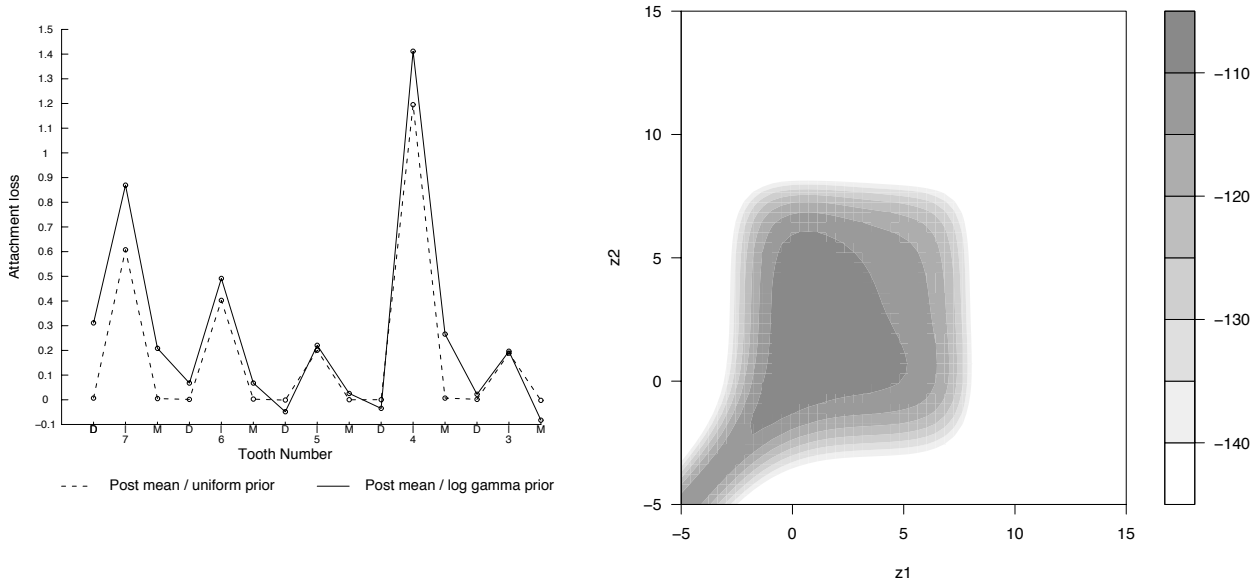


Figure 13: Subject 1, Grid B, comparing uniform and log gamma priors for z_1 and z_2.



(a) Difference between posterior mean of the buccal and lingual sides of the small left maxillary island

(b) Contour plot, log posterior using a log gamma prior