Aust. N. Z. J. Stat. **48**(2), 2006, 225–236 doi: 10.1111/j.1467-842X.2006.00436.x

BAYESIAN ANALYSIS OF THE ADDITIVE MIXED MODEL FOR RANDOMIZED BLOCK DESIGNS

JENTING WANG[1] AND JOHN S.J. HSU[2*]

State University of New York and University of California

Summary

This paper deals with the Bayesian analysis of the additive mixed model experiments. Consider b randomly chosen subjects who respond once to each of t treatments. The subjects are treated as random effects and the treatment effects are fixed. Suppose that some prior information is available, thus motivating a Bayesian analysis. The Bayesian computation, however, can be difficult in this situation, especially when a large number of treatments is involved. Three computational methods are suggested to perform the analysis. The exact posterior density of any parameter of interest can be simulated based on random realizations taken from a restricted multivariate t distribution. The density can also be simulated using Markov chain Monte Carlo methods. The simulated density is accurate when a large number of random realizations is taken. However, it may take substantial amount of computer time when many treatments are involved. An alternative Laplacian approximation is discussed. The Laplacian method produces smooth and very accurate approximates to posterior densities, and takes only seconds of computer time. An example of a pipeline cracks experiment is used to illustrate the Bayesian approaches and the computational methods.

Key words: Laplacian approximation; Monte Carlo simulation.

1. Introduction

Suppose we have t treatments that are to be compared and b blocks. The randomized complete block design is used to control and reduce experimental error. Following Box & Tiao (1973), we consider the additive mixed model, with one observation per cell. The response of the unit with the ith treatment in the jth block is

$$y_{ij} = \theta_i + b_j + \epsilon_{ij} \qquad i = 1, 2, \ldots, t; \qquad j = 1, 2, \ldots, b, \tag{1}$$

where θ_i is the mean of the ith treatment, b_j is the effect of the jth block, and ϵ_{ij} is the experimental error. The experimental errors are assumed independent and normally distributed with zero means and common variance σ_e^2. We further assume that the block effects are independent and normally distributed with zero means and common variance σ_b^2, and are

Received May 2001; revised July 2005; accepted August 2005.

*Author to whom correspondence should be addressed.

[1]Department of Mathematical Sciences, State University of New York College at Oneonta, Oneonta, NY 13820, USA.

[2]Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA. e-mail: hsu@pstat.ucsb.edu

Acknowledgments. The authors wish to thank an anonymous associate editor for valuable comments and suggestions.

independent of experimental errors. Define the sums of squares $SSB = t\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2$ and $SSE = \sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$, where $\bar{y}_{i.}$, $\bar{y}_{.j}$ and $\bar{y}_{..}$ represent sample means for the ith treatment, for the jth block, and for all observations, respectively. The likelihood function of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_t)^\top$, σ_e^2, and $\sigma_{be}^2 = \sigma_e^2 + t\sigma_b^2 > \sigma_e^2$ given the data $\mathbf{y} = (y_{11}, y_{12}, \ldots, y_{tb})^\top$ is

$$L\left(\boldsymbol{\theta}, \sigma_{be}^2, \sigma_e^2 \middle| \mathbf{y}\right) \propto \left(\sigma_{be}^2\right)^{-b/2}\left(\sigma_e^2\right)^{-b(t-1)/2} \exp\left(-\frac{1}{2}\left(\frac{SSB}{\sigma_{be}^2} + \frac{SSE}{\sigma_e^2} + (\bar{\mathbf{y}}_t - \boldsymbol{\theta})^\top \mathbf{V}^{-1}(\bar{\mathbf{y}}_t - \boldsymbol{\theta}) \right)\right), \quad \text{for } \sigma_{be}^2 > \sigma_e^2 \tag{2}$$

where $\bar{\mathbf{y}}_t = (\bar{y}_{1.}, \bar{y}_{2.}, \ldots, \bar{y}_{t.})$ and $\mathbf{V} = b^{-1}(\sigma_e^2\,\mathbf{I} + \sigma_b^2\,\mathbf{1}\,\mathbf{1}^\top)$ with $\mathbf{I}$ denoting the identity matrix and $\mathbf{1}$ denoting the unit vector.

Let $\bar{\theta} = \frac{1}{t}(\theta_1 + \theta_2 + \cdots + \theta_t)$ represent the overall mean, and $\tau_i = \theta_i - \bar{\theta}$ represent the ith treatment effect, for $i = 1, 2, \ldots, t$. Note that we may express the tth treatment effect τ_t by $\tau_t = -\sum_{i=1}^{t-1}\tau_i$. We reparametrize model (1) and rearrange the last term in (2). The likelihood function of $\bar{\theta}$, $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_{t-1})$, σ_{be}^2, and σ_e^2 is

$$L\left(\bar{\theta}, \boldsymbol{\tau}, \sigma_{be}^2, \sigma_e^2 \middle| \mathbf{y}\right) \propto \left(\sigma_{be}^2\right)^{-b/2} \exp\left(-\frac{1}{2\sigma_{be}^2}\left(SSB + tb(\bar{\theta} - \bar{y}_{..})^2\right)\right) \times \left(\sigma_e^2\right)^{-b(t-1)/2} \exp\left(-\frac{1}{2\sigma_e^2}\left(SSE + (\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})^\top \mathbf{D}^{-1}(\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})\right)\right), \tag{3}$$

for $\bar{\theta} \in \mathbb{R}$, $\boldsymbol{\tau} \in \mathbb{R}^{t-1}$, and $\sigma_{be}^2 > \sigma_e^2 > 0$, where $\mathbf{D} = b^{-1}(\mathbf{I} - t^{-1}\mathbf{1}\mathbf{1}^\top)$, $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_{t-1})$, and $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$, for $i = 1, 2, \ldots, t-1$. Note that the above likelihood function consists of two components: one for $(\bar{\theta}, \sigma_{be}^2)$ and the other for $(\boldsymbol{\tau}, \sigma_e^2)$.

Suppose that some prior information about the experiment is available. In this case, a Bayesian analysis is desirable. The prior distributions may be assigned subjectively or obtained from previously performed experiments. For example, suppose that similar experiments have been performed in the past. Summary statistics such as treatment means $\bar{\mathbf{y}}_t$, and sums of squares SSB and SSE can be calculated for each experiment. Those quantities may be used to formulate the prior distributions for the overall mean $\bar{\theta}$, treatment effects $\boldsymbol{\tau}$, and variance components σ_e^2 and σ_{be}^2. The Bayesian analysis is more informative in general. The computation, however, is sometimes not easy. The posterior distribution of $\bar{\theta}$, $\boldsymbol{\tau}$, σ_{be}^2, σ_e^2 given the data $\mathbf{y}$ is a $(t+2)$-dimensional distribution. The posterior distribution of any parameter of interest, $\eta = g(\bar{\theta}, \boldsymbol{\tau}, \sigma_{be}, \sigma_e)$, will require a $(t+2)$-dimensional integration. The distribution cannot be obtained analytically and numerical methods such as the Monte Carlo methods (model sampling, Rubinstein, 1981; Gibbs sampling, Gelfand & Smith, 1990) and approximations (Laplacian approximations, Leonard *et al.*, 1989) are needed to overcome the difficulties in computation. The detailed prior structure and the computational methods are discussed in Section 2. In Section 3, a numerical example will be discussed to illustrate the model as well as the computational methods. A detailed Bayesian analysis for the experiment will also be discussed.

2. Bayesian Analysis

We consider a conjugate prior for $\bar{\theta}$, $\boldsymbol{\tau}$, σ_{be}^2, and σ_e^2. The prior can be specified in the following three stages.

(i) Stage 1: Given σ_{be}^2, with specified parameters $\mu_{\bar{\theta}}$ and c, the overall mean $\bar{\theta}$ possesses a normal distribution with mean $\mu_{\bar{\theta}}$ and variance $c\sigma_{be}^2$. Given also σ_e^2, with specified vector $\boldsymbol{\mu}$, and matrix $\mathbf{C}$, the vector of $(t-1)$ treatment effects $\boldsymbol{\tau}$ follows a multivariate normal distribution, with a mean $\boldsymbol{\mu}$, and a covariance matrix $\sigma_e^2\,\mathbf{C}$.

(ii) Stage 2: Given σ_e^2, with specified parameters ν_1 and λ_1, the quantity $\nu_1\lambda_1/\sigma_{be}^2$ follows a truncated chi-squared distribution with ν_1 degrees of freedom, with the density

$$\pi\left(\sigma_{be}^2|\sigma_e^2\right) \propto \begin{cases} \left(\sigma_{be}^2\right)^{-\frac{\nu_1}{2}-1} \exp\left(-\dfrac{\nu_1\lambda_1}{2\sigma_{be}^2}\right) & \text{for } \sigma_{be}^2 > \sigma_e^2 \\ 0 & \text{otherwise.} \end{cases}$$

(iii) Stage 3: With specified parameters ν_2 and λ_2, the quantity $\nu_2\lambda_2/\sigma_e^2$ possesses a chi-squared distribution with ν_2 degrees of freedom. That is,

$$\pi\left(\sigma_e^2\right) \propto \begin{cases} \left(\sigma_e^2\right)^{-\frac{\nu_2}{2}-1} \exp\left(-\dfrac{\nu_2\lambda_2}{2\sigma_e^2}\right) & \text{for } \sigma_e^2 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

It is well known that the conjugate priors are very easy to work with because the posterior and prior have the same distributional form and the effect of the data is just to update the parameters from the prior to the posterior. Therefore, the resulting posterior distribution of $\bar{\theta}$, $\boldsymbol{\tau}$, σ_{be}^2 and σ_e^2 has the same form as the prior but with prior parameters $\mu_{\bar{\theta}}$, c, $\boldsymbol{\mu}$, $\mathbf{C}$, ν_1, λ_1, ν_2 and λ_2 replaced by $\mu_{\bar{\theta}}^*$, c^*, $\boldsymbol{\mu}^*$, $\mathbf{C}^*$, ν_1^*, λ_1^*, ν_2^* and λ_2^* where $\mu_{\bar{\theta}}^* = \bar{\theta}^* = (c^{-1}+tb)^{-1}(c^{-1}\mu_{\bar{\theta}} + tb\bar{y}_{..})$, $c^{*-1} = c^{-1} + tb$, $\boldsymbol{\mu}^* = \boldsymbol{\tau}^* = (\mathbf{D}^{-1}+\mathbf{C}^{-1})^{-1}(\mathbf{D}^{-1}\hat{\boldsymbol{\tau}} + \mathbf{C}^{-1}\boldsymbol{\mu})$, $\mathbf{C}^{*-1} = \mathbf{C}^{-1} + \mathbf{D}^{-1}$, $\nu_1^* = \nu_1 + b$, $\nu_1^*\lambda_1^* = \nu_1\lambda_1 + SSB + (c + t^{-1}b^{-1})^{-1}(\mu_{\bar{\theta}} - \bar{y}_{..})^2$, $\nu_2^* = \nu_2 + b(t-1)$ and $\nu_2^*\lambda_2^* = \nu_2\lambda_2 + SSE + (\hat{\boldsymbol{\tau}} - \boldsymbol{\mu})^\top \mathbf{H}(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu})$, with $\mathbf{H} = \mathbf{D}^{-1}(\mathbf{D}^{-1}+\mathbf{C}^{-1})^{-1}\,\mathbf{C}^{-1}$. Such updating procedures, in general, have been widely discussed in the Bayesian literature. For details see, for example, Berger (1985), Gelman *et al.* (1995), Leonard & Hsu (1999) and Congdon (2003).

The prior parameters $\mu_{\bar{\theta}}$, c, $\boldsymbol{\mu}$, $\mathbf{C}$, ν_1, λ_1, ν_2 and λ_2 may be specified subjectively according to researchers' prior knowledge. They may also be specified based on previously performed experiments. Alternatively, vague, improper, or objective priors might be used. For example, to determine the prior parameters we may first consider uniform priors for $\bar{\theta}$, $\boldsymbol{\tau}$, $\log\sigma_{be}^2$ and $\log\ \sigma_e^2$, where $\sigma_{be}^2 > \sigma_e^2$. The resulting posteriors are:

(i) Given σ_{be}^2, the posterior of $\bar{\theta}$ follows a normal distribution with mean $\bar{y}_{..}$ and variance σ_{be}^2/tb. Given σ_e^2, the posterior of $\boldsymbol{\tau}$ follows a normal distribution with mean vector $\hat{\boldsymbol{\tau}}$ and covariance matrix $\sigma_e^2\,\mathbf{D}$.

(ii) Given σ_e^2, the posterior of SSB/σ_{be}^2 follows a truncated chi-squared distribution with $b-1$ degrees of freedom, where $\sigma_{be}^2 > \sigma_e^2$.

(iii) The posterior of SSE/σ_e^2 follows a chi-squared distribution with $b(t-1)-1$ degrees of freedom.

This should be the choice of prior if $\bar{y}_{..}$, $\hat{\boldsymbol{\tau}}$, SSB and SSE are replaced by $\bar{y}_{..}^*$, $\hat{\boldsymbol{\tau}}^*$, SSB^* and SSE^*, where $\bar{y}_{..}^*$, $\hat{\boldsymbol{\tau}}^*$, SSB^* and SSE^* are summary statistics based upon previous

experiments. Therefore, prior parameters $\mu_{\bar{\theta}}, c, \boldsymbol{\mu}, \mathbf{C}, \nu_1, \lambda_1, \nu_2$ and λ_2 can be determined accordingly.

The primary goal for the experiment is to compare treatments. The following parameters may be considered for the analysis.

(i) The parameter of interest is a linear contrast of $\theta_1, \theta_2, \ldots, \theta_t$. That is, $\eta = \sum_{i=1}^{t} a_i\theta_i = \sum_{i=1}^{t} a_i\tau_i$, where $\sum_{i=1}^{t} a_i = 0$. Such parameters include the difference between two groups of treatments. For example, $\eta = \theta_i - \theta_j$ represents the difference between treatments i and j. We might be interested in the probability $P(\theta_i > \theta_j) = P(\tau_i > \tau_j) = P(\eta > 0)$. Such probability can be obtained by using Monte Carlo simulations from the posterior distribution of $\boldsymbol{\tau}$ given $\mathbf{y}$ or by performing a simple one-dimensional numerical integration of the approximated posterior density of η given $\mathbf{y}$ over the range $\eta > 0$. The Monte Carlo simulations and the approximations to the posterior densities for the parameters of interest will be discussed later in the section.

(ii) The parameter of interest is a quadratic function of $\theta_1, \theta_2, \ldots, \theta_t$. For example, $\eta = \sum_{i=1}^{t}(\theta_i - \bar{\theta})^2 = \sum_{i=1}^{t} \tau_i^2$. The concentration of the density of η in the neighborhood of zero indicates that the treatment means are similar.

In order to perform a detailed analysis, it is desirable to calculate the posterior density of the parameter of interest. To derive the density, we first consider the posterior density without the constraint C: $\sigma_{be}^2 > \sigma_e^2$, and denote the posterior densities and posterior probabilities by π^* and P^*, when the constraint C is ignored. Ignoring the constraint C, the posterior density of $\boldsymbol{\tau}$, given $\mathbf{y}$, can be obtained by integrating $\bar{\theta}$, σ_{be}^2 and σ_e^2 out from the joint posterior density. The resulting posterior density $\pi^*(\boldsymbol{\tau} \mid \mathbf{y})$ is the $(t-1)$-dimensional multivariate t distribution, with ν degrees of freedom, location vector $\boldsymbol{\tau}^*$ and scale matrix $\mathbf{T}$, denoted by $t_{t-1}(\nu, \boldsymbol{\tau}^*, \mathbf{T})$, and has the form

$$\pi^*(\boldsymbol{\tau} \mid \mathbf{y}) \propto \left(1 + \frac{1}{\nu}(\boldsymbol{\tau} - \boldsymbol{\tau}^*)^\top \mathbf{T}(\boldsymbol{\tau} - \boldsymbol{\tau}^*)\right)^{-[\nu_2 + b(t-1)+1]/2}, \tag{4}$$

where $\nu = \nu_2 + (t-1)(b-1) + 1$ and

$$\mathbf{T} = \frac{\nu}{\nu_2\lambda_2 + SSE + (\hat{\boldsymbol{\tau}} - \boldsymbol{\mu})^\top \mathbf{H}(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu})}(\mathbf{D}^{-1} + \mathbf{C}^{-1}).$$

Note that the actual posterior density of $\boldsymbol{\tau}$, given $\mathbf{y}$, is not $\pi^*(\boldsymbol{\tau} \mid \mathbf{y})$, the multivariate t density described in (4). We denote that the actual posterior density of $\boldsymbol{\tau}$ given $\mathbf{y}$, is the one with the constraint C: $\sigma_{be}^2 > \sigma_e^2$, by $\pi(\boldsymbol{\tau} \mid \mathbf{y}) = \pi^*(\boldsymbol{\tau} \mid C, \mathbf{y})$. Note the equality

$$\pi^*(\boldsymbol{\tau} \mid C, \mathbf{y}) \times P^*(C \mid \mathbf{y}) = P^*(C \mid \boldsymbol{\tau}, \mathbf{y}) \times \pi^*(\boldsymbol{\tau} \mid \mathbf{y}),$$

and the term $P^*(C \mid \mathbf{y})$ is constant in $\boldsymbol{\tau}$. Therefore, the desired posterior density $\pi(\boldsymbol{\tau} \mid \mathbf{y}) = \pi^*(\boldsymbol{\tau} \mid C, \mathbf{y})$ can be represented as

$$\pi(\boldsymbol{\tau} \mid \mathbf{y}) \propto \pi^*(\boldsymbol{\tau} \mid \mathbf{y}) \times P^*(C \mid \boldsymbol{\tau}, \mathbf{y}). \tag{5}$$

To calculate $P^*(C \mid \boldsymbol{\tau}, \mathbf{y}) = P^*(\sigma_{be}^2 > \sigma_e^2 \mid \boldsymbol{\tau}, \mathbf{y})$, we need to study the conditional distributions of σ_{be}^2 and σ_e^2, given $\boldsymbol{\tau}$ and $\mathbf{y}$. Ignoring the constraint C, the quantities σ_{be}^2 and σ_e^2 are independent, and the quantity U_{be} follows a chi-squared distribution with degrees of freedom

$\nu_{be} = \nu_1 + b + 1$, where

$$U_{be} = \frac{1}{\sigma^2_{be}} \left(\nu_1 \lambda_1 + SSB + (c + t^{-1} b^{-1})^{-1} (\mu_\theta - \bar{y}_{..})^2 \right).$$

Furthermore, given $\boldsymbol{\tau}$, the quantity U_e follows a chi-squared distribution with $\nu_e = \nu_2 + b\,(t-1) + 1$ degrees of freedom, where

$$U_e = \frac{1}{\sigma^2_e} \left(\nu_2 \lambda_2 + SSE + (\hat{\boldsymbol{\tau}} - \boldsymbol{\mu})^\top \mathbf{H} (\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}) + (\boldsymbol{\tau} - \boldsymbol{\tau}^*)^\top (\mathbf{D}^{-1} + \mathbf{C}^{-1}) (\boldsymbol{\tau} - \boldsymbol{\tau}^*) \right). \quad (6)$$

Therefore, given $\boldsymbol{\tau}$, the quantity

$$F = \frac{U_{be}/\nu_{be}}{U_e/\nu_e} = K(\boldsymbol{\tau}) \frac{\sigma^2_e}{\sigma^2_{be}}$$

follows an F distribution with ν_{be} and ν_e degrees of freedom, where $K(\boldsymbol{\tau})$ is a function of $\boldsymbol{\tau}$ and is defined as

$$K(\boldsymbol{\tau}) = \frac{\left(\nu_1 \lambda_1 + SSB + (c + t^{-1} b^{-1})^{-1} (\mu_\theta - \bar{y}_{..})^2 \right) / \nu_{be}}{\left(\nu_2 \lambda_2 + SSE + (\hat{\boldsymbol{\tau}} - \boldsymbol{\mu})^\top \mathbf{H} (\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}) + (\boldsymbol{\tau} - \boldsymbol{\tau}^*)^\top (\mathbf{D}^{-1} + \mathbf{C}^{-1}) (\boldsymbol{\tau} - \boldsymbol{\tau}^*) \right) / \nu_e}.$$

Following (5), the actual posterior density of $\boldsymbol{\tau}$ given $\mathbf{y}$ is

$$\pi(\boldsymbol{\tau} \mid \mathbf{y}) \propto \pi^*(\boldsymbol{\tau} \mid \mathbf{y}) \times P^*\left(\sigma^2_e < \sigma^2_{be} \middle| \boldsymbol{\tau}, \mathbf{y}\right),$$

where $\pi^*(\boldsymbol{\tau} \mid \mathbf{y})$ is the multivariate t density defined in (4), and the probability $P^*(\sigma^2_e < \sigma^2_{be} \mid \boldsymbol{\tau}, \mathbf{y})$ can be calculated via an F distribution. That is,

$$\begin{aligned} P^*\left(\sigma^2_e < \sigma^2_{be} \middle| \boldsymbol{\tau}, \mathbf{y}\right) &= P^*\left(\frac{\sigma^2_e}{\sigma^2_{be}} < 1 \mid \boldsymbol{\tau}, \mathbf{y} \right) \\ &= P\left(F_{\nu_{be}, \nu_e} < K(\boldsymbol{\tau}) | \boldsymbol{\tau}, \mathbf{y}\right), \end{aligned}$$

where the variable F_{ν_{be}, ν_e} denotes an F random variable with ν_{be} and ν_e degrees of freedom. Therefore, the posterior density of $\boldsymbol{\tau}$, given $\mathbf{y}$, is proportional to the product of a multivariate t density and the cumulative distribution function of an F random variable, evaluated at $K(\boldsymbol{\tau})$. That is,

$$\pi(\boldsymbol{\tau} \mid \mathbf{y}) \propto \left(1 + \frac{1}{\nu} (\boldsymbol{\tau} - \boldsymbol{\tau}^*)^\top \mathbf{T} (\boldsymbol{\tau} - \boldsymbol{\tau}^*) \right)^{-[\nu_2 + b(t-1) + 1]/2} \times P\left(F_{\nu_{be}, \nu_e} < K(\boldsymbol{\tau}) \middle| \boldsymbol{\tau}, \mathbf{y} \right). \quad (7)$$

The posterior densities of the parameters of interest, such as the linear contrasts and quadratic functions of $\boldsymbol{\tau}$ cannot be written in analytic forms. Therefore, approximations or simulation methods are needed to calculate the desired density. Three calculation methods are discussed below: (i) direct simulations; (ii) Markov chain Monte Carlo (MCMC) simulations; and (iii) Laplacian approximations.

(i) *Direct simulations according to the posterior distribution of $\boldsymbol{\tau}$ in equation (7).* To simulate the parameter of interest $\eta = g(\boldsymbol{\tau})$, we first simulate a $\boldsymbol{\tau}$ from the multivariate t-distribution with ν degrees of freedom, mean vector $\boldsymbol{\tau}^*$ and precision matrix $\mathbf{T}$, then keep the simulated $\boldsymbol{\tau}$ with probability $P(F_{\nu_{be}, \nu_e} < K(\boldsymbol{\tau}) \mid \boldsymbol{\tau}, \mathbf{y})$. The quantity η is then calculated according to the function $\eta = g(\boldsymbol{\tau})$.

(ii) *Markov chain Monte Carlo simulations.* Markov chain Monte Carlo (Metropolis *et al.*, 1953; Hastings, 1970; Geman & Geman, 1984; Besag & Higdon, 1999; Gelfand & Smith, 1990; Geyer, 1992; Brooks, 1998; Gelman *et al.*, 1995) has been widely used in Bayesian computations. The Gibbs sampler (Gelfand & Smith, 1990) is a particular MCMC method and can be used here to simulate $\boldsymbol{\tau}$ and hence $\eta = g(\boldsymbol{\tau})$. The Gibbs sampler requires that it is possible to simulate directly from each conditional distribution of the parameter given all the remaining parameters. Following the posterior distributions of $\bar{\theta}$, $\boldsymbol{\tau}$, σ_{be}^2 and σ_e^2, the conditional distributions are reported as follows:

(D1) Given $\boldsymbol{\tau}$, σ_{be}^2, and σ_e^2, the posterior of $\bar{\theta}$ follows a normal distribution with mean $\bar{\theta}^*$ and variance $(c^{-1} + tb)^{-1}\,\sigma_{be}^2$.

(D2) Given $\bar{\theta}$, σ_{be}^2, σ_e^2, the posterior of $\boldsymbol{\tau}$ is a multivariate normal with mean vector $\boldsymbol{\tau}^*$ and covariance matrix $(\mathbf{D}^{-1} + \mathbf{C}^{-1})^{-1}\,\sigma_e^2$.

(D3) Given $\bar{\theta}$, $\boldsymbol{\tau}$ and σ_e^2, the quantity

$$W_{be} = \frac{\nu_1\lambda_1 + SSB + (c + t^{-1}b^{-1})^{-1}(\mu_\theta - \bar{y}_{\cdot\cdot})^2 + (c^{-1} + tb)(\bar{\theta} - \theta^*)^2}{\sigma_{be}^2}$$

follows a truncated chi-squared distribution with $\nu_1 + b + 1$ degrees of freedom, where $\sigma_{be}^2 > \sigma_e^2$.

(D4) Given $\bar{\theta}$, $\boldsymbol{\tau}$, and σ_{be}^2, the quantity U_e, defined in (6), follows a truncated chi-squared distribution with $\nu_2 + b(t-1) + 1$ degrees of freedom, where $\sigma_e^2 < \sigma_{be}^2$.

The quantities $\bar{\theta}$, $\boldsymbol{\tau}$, σ_{be}^2 and σ_e^2 can be successively simulated according to the conditional distributions in (D1) – (D4). Hence, the parameter of interest $\eta = g(\boldsymbol{\tau})$ can be calculated accordingly.

(iii) The Laplacian approximations for calculating marginal densities have been used successfully in many applications. See for example, Leonard (1982), Leonard, Hsu & Tsui (1989), Tierney, Kass & Kadane (1989), Hsu (1995) and Leonard & Hsu (1999), for details. Let $\eta = g(\boldsymbol{\tau})$ be the parameter of interest. The quantity $\boldsymbol{\tau}_\eta$ conditionally maximizes $\pi(\boldsymbol{\tau} \mid \mathbf{y})$ given that $\eta = g(\boldsymbol{\tau})$. Let $\boldsymbol{\ell}_\eta$ and $\mathbf{R}_\eta$ be defined as

$$\boldsymbol{\ell}_\eta = \frac{\partial \log \pi(\boldsymbol{\tau} \mid \mathbf{y})}{\partial \boldsymbol{\tau}}\Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_\eta}$$

and

$$\mathbf{R}_\eta = -\frac{\partial^2 \log \pi(\boldsymbol{\tau} \mid \mathbf{y})}{\partial \boldsymbol{\tau}\boldsymbol{\tau}^\top}\Big|_{\boldsymbol{\tau}=\boldsymbol{\tau}_\eta}.$$

Note that the matrix $\mathbf{R}_\eta$ is the posterior information matrix of $\pi(\boldsymbol{\tau} \mid \mathbf{y})$ evaluated at $\boldsymbol{\tau} = \boldsymbol{\tau}_\eta$. Following Leonard, Hsu & Tsui (1989), the posterior density of $\pi(\eta \mid \mathbf{y})$ can be approximated by

$$\bar{\pi}(\eta \mid \mathbf{y}) \propto \pi(\boldsymbol{\tau}_\eta \mid \mathbf{y})|\mathbf{R}_\eta|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\boldsymbol{\ell}_\eta^\top \mathbf{R}_\eta^{-1}\boldsymbol{\ell}_\eta\right) f(\eta \mid \boldsymbol{\tau}_\eta^*, \mathbf{R}_\eta^{-1}), \tag{8}$$

where $\boldsymbol{\tau}_\eta^* = \boldsymbol{\tau}_\eta + \mathbf{R}_\eta^{-1}\,\boldsymbol{\ell}_\eta$, and the function $f(\eta \mid \boldsymbol{\tau}_\eta^*, \mathbf{R}_\eta^{-1})$ denotes the density of $\eta = g\,(\boldsymbol{\tau})$ when $\boldsymbol{\tau}$ possesses a multivariate normal with mean vector $\boldsymbol{\tau}_\eta^*$ and covariance matrix $\mathbf{R}_\eta^{-1}$. Note that when g is a linear function in $\boldsymbol{\tau}$, f is a normal density.

When g is a non-linear function of $\boldsymbol{\tau}$, we will need to replace f by an appropriate approximation f^*, and the accuracy of $\bar{\pi}(\eta \mid \mathbf{y})$ will depend on the precision of the approximated density f^*. For example, when g is a quadratic function in $\boldsymbol{\tau}$, the density f is well approximated by a Gamma density f^* by matching the first two moments. It has been reported in many numerical examples with many different functions $g(\boldsymbol{\tau})$, by Leonard, Hsu & Tsui (1989), Hsu, Leonard & Tsui (1991) and Leonard & Hsu (1999) that the approximation (8) possesses excellent numerical accuracy when compared with the exact result. Other approximations such as Edgeworth expansions (Johnson & Ladalla, 1979; Zellner & Rossi, 1984), and variational methods (Jordan *et al.*, 1999; Jaakkola & Jordan, 2000; Beal & Ghahramani, 2003) can also be considered.

Other Bayesian approaches to ANOVA may also be considered for the analysis. In estimating variance components, the maximum likelihood (ML) estimation takes no account of the degrees of freedom that are involved in estimating fixed effects. Thus, the resulting ML estimators are biased. This feature is overcome by restricted maximum likelihood (REML) estimation which estimates variance components based on residuals calculated after fitting just the fixed effects part of the model. See for example, Thompson (1962) and Harville (1977). Our method improves upon REML by expressing uncertainty regarding the variance components. Besag & Higdon (1993, 1999) and Besag *et al.* (1995) discussed Bayesian approaches for analysing agricultural field experiments. They proposed complex formulations for situations when spatial effects were considered, while our approach is for the standard additive mixed model. An advantage of our approach, when compared with other Bayesian approaches, is that we are easily able to access the marginals of complex functions of the first stage parameters.

3. An Example

Construction of a land pipeline for transporting domestic waste water from a primary treatment plant was completed by the Perth Metropolitan Water Authority (Western Australia) in the 1980s. Tests for cracking of the cement mortar lining of the pipeline were performed during the course of construction to determine if autogenous healing would seal the cracks for, if not, costly repairs using epoxy would be necessary. After cracks were observed, the pipeline was kept filled with water for 14 weeks, and Cox & Kelsall (1986) reported that periodic measurements (in millimetres) of crack widths were taken at 12 randomly chosen crack locations initially before filling the pipeline, and then after the 2nd, 6th, and 14th weeks during the wet period. Those measurements are reproduced in Table 1 and displayed in Figure 1 for the purpose of illustration. The intervals, 0, 2, 6 and 14 weeks, represent the four treatments of the experiment, and the corresponding treatment means are denoted by θ_1, θ_2, θ_3 and θ_4 respectively. The primary interest of this experiment is to learn whether the four predetermined times of measuring after the initial cracks appeared affected the crack widths.

Sincich (1995) suggested the additive mixed model (1) for analysis, with the four predetermined treatments as fixed effects and the twelve crack locations as random blocks. The analysis of variance is summarized in Table 2. The treatment F ratio has a p-value of nearly zero indicating significant different treatment means. The block F ratio has a p-value of 0.1255, which indicates that the location faction is not very significant. Furthermore, the interaction between time periods and location does not seem to be significant. Although this cannot be tested in this case because of the lack of degrees of freedom, it can be seen from the scatter plot of crack width against wet period in Figure 1 that the interaction is not

TABLE 1

The Pipeline Cracks Data

Crack location	Crack width (in millimetres) after being wet for			
	0 weeks	2 weeks	6 weeks	14 weeks
1	0.50	0.20	0.10	0.10
2	0.40	0.20	0.10	0.10
3	0.60	0.30	0.15	0.10
4	0.80	0.40	0.10	0.10
5	0.80	0.30	0.05	0.05
6	1.00	0.40	0.05	0.05
7	0.90	0.25	0.05	0.05
8	1.00	0.30	0.05	0.10
9	0.70	0.25	0.10	0.10
10	0.60	0.25	0.10	0.05
11	0.30	0.15	0.10	0.05
12	0.30	0.14	0.05	0.05

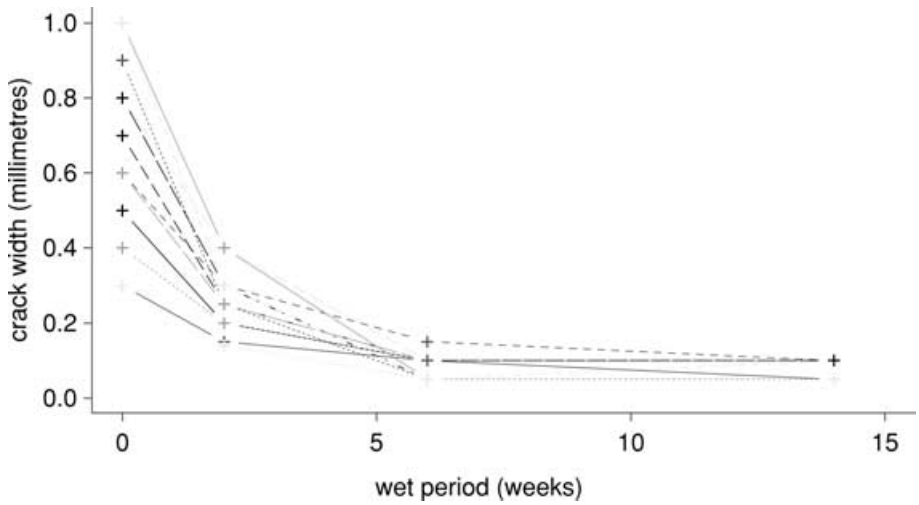

Figure 1. Scatter plot for the pipeline cracks data.

significant because similar curve patterns are presented for different locations over four measuring times. For the Bayesian analysis, we consider the three-stage conjugate prior, discussed in Section 2, with vague prior information about the overall mean and the treatment effects, by letting $c \to \infty$ and $|\mathbf{C}| \to \infty$. Furthermore, we choose prior parameters $\lambda_1 = \nu_1 = \lambda_2 = \nu_2 = 1$, as an example for illustration. The procedure remains the same for any other prior parameters supplied. We study whether the measuring time affected the crack widths by evaluating $\eta = \sum_{i=1}^{4} \tau_i^2 = \sum_{i=1}^{4} (\theta_i - \bar{\theta})^2$. The idea is that if the measuring time does not affect the crack widths then the treatment means θ_1, θ_2, θ_3 and θ_4 are all identical, hence $\eta = 0$. The posterior distributions of η are presented in Figure 2. Curve (a) in the figure is the simulated exact histogram of η, which was simulated from the exact posterior density $\pi(\boldsymbol{\tau} \mid \mathbf{y})$ of $\boldsymbol{\tau}$ given $\mathbf{y}$ in (7). The MCMC approach produced a similar histogram to (a). Curve (b) in the figure represents the approximate posterior density of η which is obtained using the Laplacian approximation (8). Curve (b) agrees very well with the histogram and shows the

TABLE 2

Analysis of variance of the pipeline cracks data

Source of variation	Sum of squares	Degrees of freedom	Mean square	F_0	p-value
Treatments	2.685	3	0.895	58.117	<0.0001
Blocks	0.277	11	0.025	1.667	0.1255
Error	0.509	33	0.015		
Total	3.471	47			

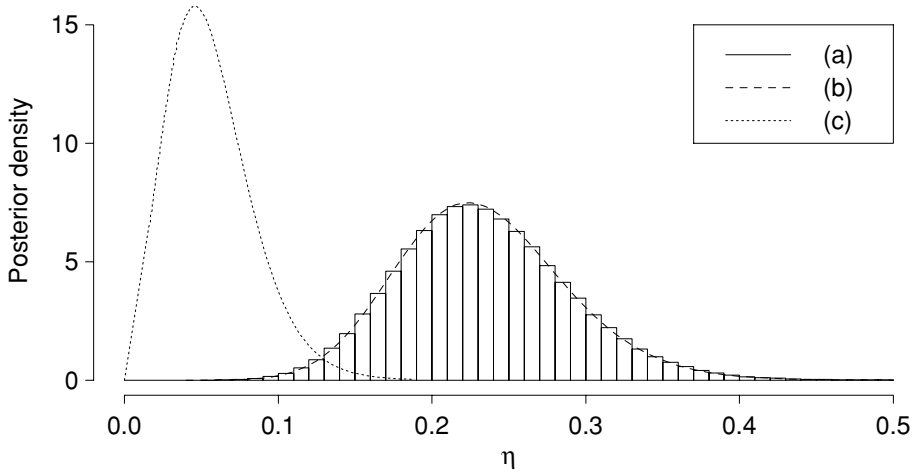

Figure 2. Marginal posterior density of $\eta = \sum_{i=1}^{4}(\theta_i - \bar{\theta})^2$ for the pipeline cracks data. (a) Histogram, based on 100 000 simulations from exact posterior density with vague prior; (b) Laplacian Approximation with vague prior; (c) Laplacian Approximation with $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{C} = \mathbf{D}$.

accuracy of the approximation (8). The posterior density curve is centered about 0.22 and is visible within the range (0.05, 0.50). The curve being focused away from 0 suggests that the four treatments have significantly different effects. This conclusion of significant differences between treatments with the vague prior information, coincides with the standard ANOVA approach. With a more informative prior, however, the results might be quite different. For example, suppose that the researchers believe that the four measuring times do not affect the crack widths. Curve (c) represents the approximate posterior density of η when the prior parameters are $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{C} = \mathbf{D}$ instead. Those parameters may be obtained according to researchers' knowledge or by using the Bayesian's updating procedure discussed in Section 2, based on summary information from previous experiments. Curve (c) is much more peaked, centered about 0.045, and ranged between 0.0 and 0.2. The curve is shifted towards 0 compared with curve (b) using a vague prior. It is not obvious if the conclusion of indifference between the treatments can be drawn from curve (c).

The primary interest for the experiment is to determine whether the four measuring times affected the crack widths. Supposing that the vague prior is used, we have now concluded that the four predetermined times of measuring after the initial cracks affected the crack widths. It may be of interest to investigate further the structure of the differences. We can study the linear, quadratic and cubic relationships between the changes in size of the cracks

TABLE 3

Table of orthogonal polynomial coefficients

Wet period (weeks)	Normalized coefficients for		
	Linear contrast	Quadratic contrast	Cubic contrast
0	−0.5129	0.5296	−0.4544
2	−0.3264	−0.1059	0.7952
6	0.0466	−0.7680	−0.3976
12	0.7926	0.3443	0.0568

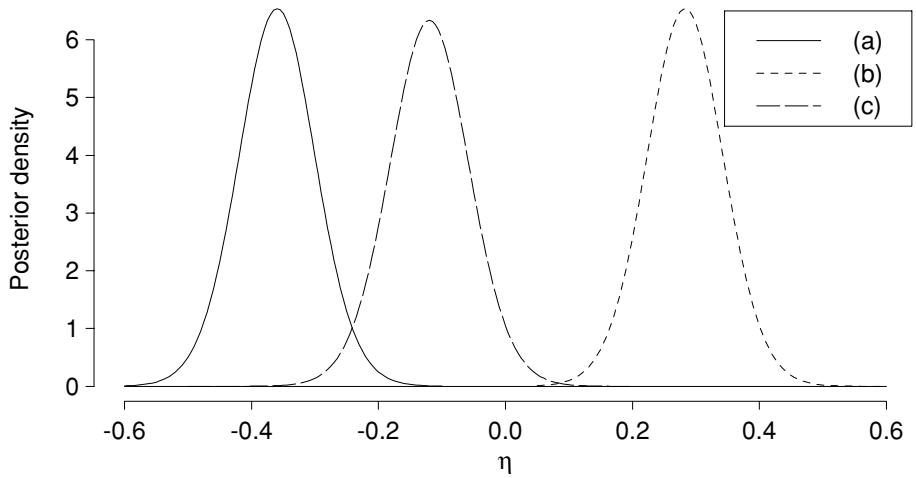

Figure 3. Marginal posterior densities of linear, quadratic and cubic contrasts for the pipeline cracks data. (a) Linear contrast; (b) Quadratic contrast; (c) Cubic contrast.

over time by constructing orthogonal contrasts in the treatment effects τ_1, τ_2, τ_3 and τ_4. The four predetermined times of measuring were not equally spaced, so the standard tables for constructing orthogonal contrasts are not applicable. However, the contrasts can still be constructed (see for example Wishart & Metakides, 1953 and Robson, 1959). Following the procedure provided by Robson (1959) the normalized coefficients for the orthogonal contrasts are constructed and reported in Table 3.

The normalized linear contrast

$$L = -0.5129\tau_1 - 0.3264\tau_2 + 0.0466\tau_3 + 0.7926\tau_4$$

represents a measure of linear trend of the measuring time. The case $L = 0$ represents no linear trend, and the linear trend is significant while L is away from zero. Curve (a) in Figure 3 represents the posterior density of L. Curve (a) is centered about -0.36 and is visible within the range $(-0.60, -0.10)$. The posterior probability $P(L > 0) \approx 0$ indicates that the linear trend is strong. Similarly, the normalized quadratic and cubic contrasts

$$Q = 0.5296\tau_1 - 0.1059\tau_2 - 0.7680\tau_3 + 0.3443\tau_4 \text{ and}$$
$$C = -0.4544\tau_1 + 0.7952\tau_2 - 0.3976\tau_3 + 0.0568\tau_4$$

represent measures of quadratic and cubic trends of the measuring times, respectively. Curves (b) and (c) in Figure 3 represent the posterior densities for Q and C, respectively. Curve (b) is centered about 0.28 and is visible within the range (0.05, 0.50) and curve (c) is centered about -0.12 and is visible within range $(-0.40, 0.18)$. The posterior probabilities $P(Q > 0) \approx 1$ and $P(C > 0) = 0.0320$ indicate that the quadratic trend is strong but not as much as the linear trend, and the cubic trend is insubstantial. The arguments here coincide with the scatter plot in Figure 1: while the linear and quadratic trends are strong, the cubic trend is not. In conclusion, the crack sizes do differ depending on the time of measuring. The shrinkage in the cracks varies both linearly and quadratically over time. For more discussion on pipeline cracks please see, for example, Kannappan (1986), Lu (1998) and Zhang *et al.* (1999).

References

BEAL, M.J. & GHAHRAMANI, Z. (2003). The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures, in *Bayesian Statistics*, Vol. **7**, eds J.M. Mernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith & M. West, pp. 453–464. New York: Oxford University Press.

BERGER, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.

BESAG, J. & HIGDON, D. (1993). Bayesian inference for agricultural field experiments. *Bull. Int. Statist. Inst.* **55**, 121–136.

BESAG, J. & HIGDON, D. (1999). Bayesian analysis of agricultural field experiments (with discussion). *J. R. Statist. Soc. B* **61**, 691–746.

BESAG, J., GREEN, P.J., HIGDON, D. & MENGERSEN, K.L. (1995). Bayesian computation and stochastic systems (with discussion). *Statist. Sci.* **10**, 3–66.

BOX, G.E.P. & TIAO, G.C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Wiley.

BROOKS, S.P. (1998). Markov chain Monte Carlo method and its application. *The Statistician* **47**, 69–100.

CONGDON, P. (2003). *Applied Bayesian Modeling*. Chichester: Wiley.

COX, B.G. & KELSALL, K.J. (1986). Construction of Cape Peron ocean outlet, Perth, Western Australia. *Proceedings of the Institute of Civil Engineers, Part 1* **80**, 465–491.

GELFAND, A.E. & SMITH, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409.

GELMAN, A., CARLIN, J.B., STERN, H.S. & RUBIN, D.B. (1995). *Bayesian Inference in Statistical Analysis*. New York: Chapman and Hall.

GEMAN, S. & GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

GEYER, C.J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.* **7**, 473–511.

HARVILLE, D.A. (1977). Maximum-likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72**, 320–340.

HASTINGS, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

HSU, J.S.J. (1995). Generalized Laplacian approximations in Bayesian inference. *Canadian J. Statist.* **23**, 399–410.

HSU, J.S.J., LEONARD, T. & TSUI, K. (1991). Statistical inference for multiple choice tests. *Psychometrika* **56**, 327–348.

JAAKKOLA, T. & JORDAN, M. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing* **10**, 25–37.

JOHNSON, R.A. & LADALLA, J.N. (1979). The large sample behavior of posterior distributions which sample from multiparameter exponential models, and allied results. *Sankyha* **41**, 196–215.

JORDAN, M.I., GHAHRAMANI, A., JAAKKOLA, T.S. & SAUL, L.K. (1999). An introduction to variational methods for graphical models. *Machine Learning* **37**, 183–233.

KANNAPPAN, S. (1986). *Introduction to Pipe Stress Analysis*. New York: John Wiley and Sons.

LEONARD, T. (1982). Comment on "A simple predictive density function." *J. Amer. Statist. Assoc.* **77**, 657–658.

LEONARD, T. & HSU, J.S.J. (1999). *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. New York: Cambridge University Press.

LEONARD, T., HSU, J.S.J. & TSUI, K. (1989). Bayesian marginal inference. *J. Amer. Statist. Assoc.* **84**, 1051–1057.

LU, Y. (1998). Frequency-mode selection for ultrasonic detection and characterization of circumferential cracks in pipelines. *Proceedings of SPIE* **3398**, 18–27.

METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H. & TELLER, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

ROBSON, D.S. (1959). A simple method for constructing orthogonal polynomials when the independent variable is unequally spaced. *Biometrics* **15**, 197–191.

RUBINSTEIN, R.Y. (1981). *Simulation and the Monte Carlo Method*. New York: Wiley.

SINCICH, T. (1995). *Business Statistics by Example*. Upper Saddle River, NJ: Prentice-Hall.

THOMPSON, W.A., Jr (1962). The problem of negative estimates of variance components. *Ann. Math. Stat.* **33**, 273–289.

TIERNEY, L., KASS, R.E. & KADANE, J. (1989). Approximate marginal densities of non-linear functions. *Biometrika* **76**, 425–433.

WISHART, J. & METAKIDES, T. (1953). Orthogonal polynomial fitting. *Biometrika* **40**, 361–369.

ZELLNER, A. & ROSSI, P.E. (1984). Bayesian analysis of dichotomous quantal response models. *J. Econometrics* **25**, 365–393.

ZHANG, X.Y., LAMBERT, S.B., PLUMTREE, A. & SUTHERBY, R. (1999). Transgranular stress corrosion cracking of X-60 pipeline steel in simulated ground water. *Corrosion* **55**, 297–305.