Published in final edited form as:

J Biopharm Stat. 2009 July ; 19(4): 700–720. doi:10.1080/10543400902964159.

A Novel Gibbs Maximum A Posteriori (GMAP) Approach on Bayesian Nonlinear Mixed-Effects Population Pharmacokinetics (PK) Models

Seongho Kim¹, Stephen D. Hall², and Lang Li^{1,*}

¹Division of Biostatistics, Department of Medicine, School of Medicine, Indiana University, Indianapolis, IN, 46032

²Division of Clinical Pharmacology, Department of Medicine, School of Medicine, Indiana University, Indianapolis, IN, 46032

Abstract

In this paper, various Bayesian Monte Carlo Markov Chain (MCMC) methods and the proposed algorithm, Gibbs maximum a posteriori (GMAP) algorithm, are compared for implementing the nonlinear mixed-effects model in pharmacokinetics (PK) studies. An intravenous two-compartmental PK model is adopted to fit the PK data from the midazolam (MDZ) studies, which recruited 24 individuals with 9 different time points per subject. The three-stage hierarchical nonlinear mixed model is constructed. Data analysis and model performance comparisons show that GMAP converges the fastest, and provides reliable results. At the mean time, data augmentation (DA) methods are used for the Random-walk Metropolis method. Data analysis shows that the speed of the convergence of Random-walk Metropolis can be improved by DA, but all of them are not as fast as GMAP. The performance of GMAP and various MCMC algorithms are compared through Midazolam data analysis and simulation.

Keywords

Bayesian model; Monte Carlo Markov Chain (MCMC); Pharmacokinetics (PK); Gibbs Maximum a Posteriori (GMAP)

1 Introduction

The nonlinear mixed-effect model (NLMM) is, in general, defined by two levels of hierarchy. At the first level, drug concentrations from an individual are considered as arising from an individual probability model, whose mean is given by a PK model with subject-specific parameters. The subject parameters, at the second level, are considered as random variables and their probability distributions are modeled as a function of subject level covariates (Beal and Sheiner, 1982; Ette et al., 1997).

The maximum likelihood (ML) approach is the most popular method in estimating the PK parameters from the NLMM. For the ML approach, however, it is difficult to calculate the likelihood of the data for NLMM due to the nonlinear function of the observations on the random parameters. For this reason, several approximate methods have been proposed to overcome this problem (Sheiner and Beal, 1982; Mallet, 1986; Lindstrom and Bates, 1990;

*Corresponding author lali@iupui.edu.

Davidian and Gallant, 1992; Vonesh and Carter, 1992; Wolfinger, 1993). Nevertheless, it is still hard to evaluate the adequacy of the approximation and the appropriateness of the asymptotic theory. Alternatively, Markov Chain Monte Carlo (MCMC) (Smith and Roberts, 1993) was used in NLMM (Wakefield et al., 1994), and in theory it provided an exact solution without approximation. MCMC is Monte Carlo integration using Markov chains. There are many strategies of constructing these chains (Geman and Geman, 1984; Smith and Gelfand, 1992; Tierney, 1994; Wakefield et al., 1994; Gilks, 1996; Gilks and Roberts, 1996; Roberts, 1996), but all of them are special cases of the general framework of Metropolis et al. (1953) and Hastings (1970).

Unfortunately, the more complex the models become, the slower the MCMC converges or mixes, i.e., the Markov chain will not move rapidly throughout the support of the target distribution. Various strategies, such as data augmentation (DA) (or auxiliary variable (AV)) algorithms (Tanner and Wong, 1987; Damien, Wakefield, and Walker, 1999; van Dyk and Meng, 2001), were developed to improve the speed of MCMC. DA algorithms are iterative methods for constructing optimization or sampling algorithms with augmenting the observed data by unobserved data or latent variables. They have been employed to make simulation feasible and simple as well as to improve the speed of iterative simulation.

In a previous work, Bennett, Racine-Poon, and Wakefield (1996) compared various MCMC strategies in a NLMM, such as Rejection Gibbs, Ratio Gibbs, Random-walk Metropolis, Independence Metropolis-Hastings, and MLE/prior Metropolis-Hastings. They demonstrated that Ratio Gibbs approach could be a daunting process for the uninitiated case because it involves several maximizations per random variable generation. In addition, computing time per iteration for Rejection Gibbs approach is far more than that for Metropolis-Hastings methods due to the large number of rejection sampling iterations per acceptance. Bennett et al. (1996) finally concluded that Random-walk Metropolis is the best MCMC strategy based on their studies. In this paper, besides what has been discussed in Bennett et al. (1996), we add data augmentation strategies into the MCMC and developed Gibbs maximum a posteriori (GMAP) algorithms. GMAP consists of Maximum a Posteriori (MAP) and Gibbs sampler, and MAP is implemented for the parameters not having a closed form of the posterior distribution and Gibbs sampler is adapted for the parameters with a closed form. We compare all of these approaches in PK data analysis and simulation studies.

In Section 2, a brief description of the midazolam (MDZ) PK model is given; and MCMC and GMAP approaches are defined. In Section 3 MCMC methods and GMAP are compared through MDZ PK data analysis as well as simulation experiment; and in section 4, conclusions are reached.

2. Methods

2.1. Midazolam (MDZ) PK Model

The PK of MDZ is assumed to follow a two-compartment model with intravenous infusion (IV). Then the PK is described by the system of the ordinary differential equations (ODEs) as follows:

$$\begin{aligned}\frac{dA_1}{dt} &= -CL \cdot \frac{A_1}{V_1} + CL_{12} \cdot \left(-\frac{A_1}{V_1} + \frac{A_2}{V_2}\right) \\ \frac{dA_2}{dt} &= CL_{12} \cdot \left(\frac{A_1}{V_1} - \frac{A_2}{V_2}\right) \\ (A_1, A_2)|_{t=0} &= (Dose, 0)\end{aligned}$$

where $(V1, V2)$ are volumes of distribution in systemic and peripheral compartments respectively; $CL12$ is the inter-compartment rate constant; CL is the systemic clearance. When solving the ODEs, the predicted concentration at time t is given by

$$\log f(\theta, t) = \log \frac{A1(t)}{V1}$$

where

$$\begin{aligned} A1(t) &= Dose \cdot \left[\frac{\left(\frac{CL12}{V2} - \lambda_1\right)e^{-\lambda_1 t}}{(\lambda_2 - \lambda_1)} + \frac{\left(\frac{CL12}{V2} - \lambda_2\right)e^{-\lambda_2 t}}{(\lambda_1 - \lambda_2)} \right] \\ \lambda_1 &= \frac{1}{2} \left[\frac{CL}{V1} + \frac{CL12}{V1} + \frac{CL12}{V2} + \sqrt{\left(\frac{CL}{V1} + \frac{CL12}{V1} + \frac{CL12}{V2}\right)^2 - 4 \cdot \frac{CL12}{V2} \cdot \frac{CL}{V1}} \right] \\ \lambda_2 &= \frac{1}{2} \left[\frac{CL}{V1} + \frac{CL12}{V1} + \frac{CL12}{V2} - \sqrt{\left(\frac{CL}{V1} + \frac{CL12}{V1} + \frac{CL12}{V2}\right)^2 - 4 \cdot \frac{CL12}{V2} \cdot \frac{CL}{V1}} \right] \\ \theta &= \left(\log V1, \log V2, \log CL12, \log CL \right)' \end{aligned}$$

2.2 MDZ PK Study

In an MDZ PK study approved by Indiana University IRB, twenty-four volunteers, 18 to 55 years of age, received single dose of 2.74–4.80 mg MDZ intravenously (IV). No study volunteers were taking any prescription medication including hormonal contraceptives throughout the course of the study. All volunteers were non-smokers and had body mass indexes (BMI) within 19 to 30 kg/m². Other exclusion criteria were metabolic, cardiac, pulmonary, renal, hepatic dysfunction, or women who were pregnant or breast feeding. Blood samples for MDZ assays were collected in non-heparinized evacuated blood collection tubes at 0.5, 0.75, 1, 1.5, 2, 4, 6, and 9 hours after IV MDZ dosing.

2.3 A Nonlinear Mixed-Effect Model Specification

The three-stage hierarchical NLMM specification follows the notation given by Bennett et al. (1996). The first-stage model has the form of

$$\left[\log y_{ij} \mid \theta_i, \tau^{-1}, t_{ij} \right] \mathcal{N} \left(\log f(\theta_i, t_{ij}), \tau^{-1} \right), \quad i=1, \dots, M; j=1, \dots, n_i,$$

where drug concentration y_{ij} from subject i at time t_{ij} is modeled as the nonlinear function $f(\theta_i, t_{ij})$ with variance τ^{-1} , and θ_i is a p -dimensional subject-specific PK parameter vector. The second-stage structure is given by

$$\left[\theta_i \mid \theta, \Sigma_p \right] \mathbf{MVN}(\theta, \Sigma_p),$$

where θ is the population-average PK parameter vector, and Σ_p is the between-subject covariance matrix. Note that **MVN** is shorthand for multivariate normal distribution. The third stage of the model describes the prior distributions

$$\left[\tau \right] \mathcal{Ga} \left(\frac{\nu_0}{2}, \frac{\tau_0 \nu_0}{2} \right), \left[\theta \right] \mathbf{MVN}(\mathbf{c}, \mathbf{C}_p), \left[\Sigma_p^{-1} \right] \mathcal{W} \left(\rho, \left[\rho \mathbf{R}_p \right]^{-1} \right)$$

for given ρ , $\mathbf{R}_p$, v_0 , τ_0 , $\mathbf{c}$, and $\mathbf{C}_p$, where Ga is a gamma distribution and W a Wishart distribution. Then the posterior distribution for $(\tau, \theta_1, \dots, \theta_M, \theta, \Sigma_p)$ is proportional to

$$\prod_{i=1}^M \prod_{j=1}^{n_i} P(\log y_{ij} | \log f(\theta_i, t_{ij}), \tau) P(\theta_i | \theta, \Sigma_p) \times \text{Prior}(\tau, \theta, \Sigma_p);$$

$$\text{Prior}(\tau, \theta, \Sigma_p) = Ga\left(\frac{v_0}{2}, \frac{\tau_0 v_0}{2}\right) \cdot \text{MVN}(\mathbf{c}, \mathbf{C}_p) \cdot W(\rho, [\rho \mathbf{R}]^{-1}).$$

Let $l_i(\theta_i, \tau)$ denote the conditional log-likelihood for the i th subject, while ignoring terms which do not depend on θ_i :

$$l_i(\theta_i, \tau) = -\frac{\tau}{2} \sum_{j=1}^{n_i} (\log y_{ij} - \log f(\theta_i, t_{ij}))^2.$$

For setting of the hyperparameters (prior distributions), we employed the previous results from the meta-analysis for MDZ studies as depicted in Table II in Li et al. (2007). Li et al. (2007) used a nonlinear two-compartmental model, while we use a linear two-compartmental model. Therefore, we use the following equation (*Well-stirred* and *Michaelis-Menten* model) to recover CL (clearance) from V_{max} and K_m in their Table II:

$$CL = \frac{80 \times \frac{V_{max}}{K_m}}{80 + \frac{V_{max}}{K_m}}.$$

Overall, we assign the previous results below to the prior distributions in our linear two-compartmental model:

$$\rho=5; c=\begin{bmatrix} \log V1 \\ \log V2 \\ \log CL \\ \log CL12 \end{bmatrix} = \begin{bmatrix} \log 67.05 \\ \log 45.08 \\ \log 76.97 \\ \log 42.9 \end{bmatrix}; \tau_0=0.3; v_0=\frac{2 \cdot \tau_0^2}{0.2^2}; C_p=\begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix}; R_p=C_p^{-1}.$$

2.4 Markov Chain Monte Carlo Algorithms

The Gibbs sampler is defined by the following full conditional distributions except for random effect θ_i :

$$\begin{aligned}
& [\tau | y_1, \dots, y_M, \theta_1, \dots \\
& \quad , \theta_M \propto \prod_{i=1}^M \prod_{j=1}^{n_i} P(\log y_{ij} | \log f(\theta_i) \\
& \quad , t_{ij}, \tau \times \text{Prior}(\tau) \text{Ga}\left(\frac{1}{2} \{v_0 \right. \\
& \quad + \sum_{i=1}^M n_i \\
& \quad , \frac{1}{2} \left\{ \sum_{i=1}^M \sum_{j=1}^{n_i} [\log y_{ij} - \log f(\theta_i) \right. \\
& \quad , t_{ij}^2 + v_0 \tau_0
\end{aligned} \tag{1}$$

$$[\theta | \theta_1, \dots, \theta_M, \Sigma_p] \propto \prod_{i=1}^M P(\theta_i | \theta, \Sigma_p) \times \text{Prior}(\theta) \text{MVN}\left(V_p \left\{ M \Sigma_p^{-1} \bar{\theta} + C_p^{-1} c \right\}, V_p\right) \tag{2}$$

$$[\Sigma_p^{-1} | \theta_1, \dots, \theta_M, \theta] \propto \prod_{i=1}^M P(\theta_i | \theta, \Sigma_p) \times \text{Prior}(\Sigma_p) \mathcal{W}\left(M + \rho, \left\{ \sum_{i=1}^M (\theta_i - \theta)(\theta_i - \theta)' + \rho \mathbf{R} \right\}^{-1}\right), \tag{3}$$

where $\bar{\theta} = \frac{1}{M} \sum_{i=1}^M \theta_i$ and $V_p^{-1} = M E_p^{-1} + C_p^{-1}$.

Because of the nonlinear function $f(\theta_i, t_{ij})$, the conditional distribution of θ_i is non-standard and known up to a normalizing constant,

$$\begin{aligned}
[\theta_i | y_i, \theta, \Sigma_p, \tau] & \propto \prod_{j=1}^{n_i} P(\log y_{ij} | \log f(\theta_i, t_{ij}), \tau) \cdot P(\theta_i | \theta, \Sigma_p) \\
& \propto \exp\{l_i(\theta_i, \tau)\} \cdot P(\theta_i | \theta, \Sigma_p) \\
& \propto \exp\left\{-\frac{\tau}{2} \sum_{j=1}^{n_i} (\log y_{ij} - \log f(\theta_i, t_{ij}))^2 - \frac{1}{2} (\theta_i - \theta)' \Sigma_p^{-1} (\theta_i - \theta)\right\}
\end{aligned}$$

Now, we introduce data augmentation for drawing samples from the full conditional distribution of θ_i . Other methods are described in Bennett et al. (1996).

Let y be the observed data and θ the parameter of interest. The DA algorithm (Tanner and Wong, 1987) starts with the construction of the missing data z in the original model. A DA scheme is a model for $y, p(y, z | \theta)$, that satisfies the following constraint

$$\int p(y, z | \theta) dz = p(y | \theta) \tag{4}$$

Our nonlinear model can be written in the DA algorithm by introducing a missing data z_{ij} for each observation. This DA model is obtained by specifying

$$\left[\log y_{ij} | \theta_i, \tau^{-1}, z_{ij}, t_{ij} \right] \mathbf{N} \left(\log f(\theta, t_{ij}) + z_{ij}, \tau^{-1} \left(1 - \frac{\xi^{-1}}{\tau^{-1}} \right) \right) \quad (5)$$

and

$$\left[z_{ij} | \theta_i, \tau^{-1} \right] \mathcal{N} \left(0, \xi^{-1} \right) \quad (6)$$

where $i = 1, \dots, M; j = 1, \dots, n_i$. It is not difficult to check that integrating over the z_{ij} is satisfied with Equation (4). Then the DA algorithm is as follows:

Step 1: Draw z_{ij} conditional on $(\theta_i, \tau^{-1}, y_{ij}, t_{ij})$, where $i = 1, \dots, M; j = 1, \dots, n_i$; that is, draw

$$\left[z_{ij} | \theta_i, \tau^{-1}, y_{ij}, t_{ij} \right] \mathcal{N} \left(\frac{\xi^{-1}}{\tau^{-1}} (\log y_{ij} - \log f(\theta_i, t_{ij})), \xi^{-1} \left(1 - \frac{\xi^{-1}}{\tau^{-1}} \right) \right)$$

Step 2: Draw θ_i conditional on $(\tau^{-1}, z_{ij}, y_{ij}, t_{ij}, j = 1, \dots, n_i)$ with the prior $\mathbf{N}(\theta, \sigma^2)$, where $i = 1, \dots, M$; that is, draw

$$\left[\theta_i | \tau^{-1}, z_{ij}, y_{ij}, t_{ij}, j = 1, \dots, n_i \right] \propto \exp \left\{ -\frac{1}{2(\tau^{-1} - \xi^{-1})} \sum_{j=1}^{n_i} (\log y_{ij} - \log f(\theta_i - t_{ij}) - z_{ij})^2 - \frac{1}{2} (\theta_i - \theta)' \Sigma_p^{-1} (\theta_i - \theta) - \frac{1}{\xi^{-1}} \sum_{j=1}^{n_i} z_{ij}^2 \right\}.$$

Note that we must use the Metropolis-Hastings algorithm for *Step 2* for our nonlinear model because there is no closed form.

In linear models, Liu et al (1994, 1995) have shown that the convergence rate of any DA scheme is equal to the square of the maximal correlation between the missing variable and the parameter (as for our linear model, see Appendix I). In order to investigate the rate of the convergence of our nonlinear model, we approximate f to $f(0) + f'(0)\theta_i$ by the first-order Taylor expansion since it is not available in the closed form to calculate the posterior distribution. Then the rate of convergence for this nonlinear model becomes

$$\begin{aligned} r_0 &= \frac{n_i f'(0) \frac{\xi^{-1}}{\tau^{-1}} \sigma^{-1}}{\tau^{-1} \left(\frac{n_i f'(0) \sigma^{-1}}{\tau^{-1}} + 1 - \frac{\xi^{-1}}{\tau^{-1}} \right)} \\ &= \frac{n_i f'(0) r \sigma^{-1}}{\tau^{-1} \left(\frac{n_i f'(0) \sigma^{-1}}{\tau^{-1}} + 1 - r \right)}, \text{ where } r = \frac{\xi^{-1}}{\tau^{-1}} \end{aligned} \quad (7)$$

Note that the convergence rate of (7), r_0 , decreases with $r \left(= \frac{\xi^{-1}}{\tau^{-1}} \right)$, which is the ratio of the variances, going to 0.

2.5 Gibbs Maximum a Posteriori (GMAP) Algorithm

We develop a hybrid algorithm with Gibbs sampling and Maximum a Posteriori (MAP), which is called Gibbs maximum a posteriori (GMAP) algorithm. To implement GMAP, the parameters are first divided into two parts, which are tractable and intractable. If a parameter

has a closed form to its posterior distribution, it is categorized into a tractable parameter, and, if not, it is intractable. As for tractable parameters, Gibbs sampling is implemented, and intractable parameters are estimated by MAP. In our NLMM case, because there is a closed form available for the fixed effects, it is implemented using Gibbs Sampling according to Equations (1), (2), and (3), while the random effects are estimated by MAP, instead of MCMC methods, as follows: As the likelihood of θ_i is proportional to

$$-\frac{\tau}{2} \sum_{j=1}^{n_i} (\log y_{ij} - \log f(\theta_i, t_{ij}))^2 - \frac{1}{2} (\theta_i - \theta)' \Sigma_p^{-1} (\theta_i - \theta),$$

MAP estimator of θ_i , θ_i^{MAP} , is estimated by

$$\operatorname{argmax}_{\theta_i} \left[-\frac{\tau}{2} \sum_{j=1}^{n_i} (\log y_{ij} - \log f(\theta_i, t_{ij}))^2 - \frac{1}{2} (\theta_i - \theta)' \Sigma_p^{-1} (\theta_i - \theta) \right]$$

We use the quasi-Newton method (Broyden, 1970), to perform maximization of the posterior distribution.

GMAP is interested in estimating a distribution of the mode of the posterior distribution and estimates the standard deviation for the estimate of the mode of each parameter, while other MCMC methods give us the standard deviation of the estimate of each parameter. Therefore we used the Fisher information matrix of random-effects for each subject in order to recover the standard deviations of random-effects for each subject (Appendix II).

2.6 The Comparison Criterion

To compare various MCMC methods and GMAP, we formally consider the required number of iterations needed to achieve adequate precision in ergodic averages, using the method of Raftery and Lewis (1992), and comment informally the time involved in implementing the algorithm and the computer time required per iteration. In addition, we compare the performance of the estimation of each model

The method of Raftery and Lewis (1992) estimates, on the basis of a pilot run, the number of iterations required to estimate the q -th posterior quantile of some function of the model parameters to within $\pm\delta$ with probability s . The length of burn-in is controlled by another variable, denoted by ε (see, Raftery and Lewis, 1996).

Note that all the statistical analyses and simulations are performed using the statistical package R (R Development Core Team). To generate the random samples of each distribution (Normal distribution, Multivariate Normal distribution, Whishart distribution, Gamma distribution, and Uniform distribution), we use the packages, which are *MCMCpack* and *mnormt*. The R function *optim* is used for estimating the random-effects of our proposed method. To solve the differential equations for PK models, we use the package *nlmeODE* along with *nlme* and *odesolve*. The package *coda* is used for calculating the number of iterations required to converge of Raftery and Lewis (1992). It is available from the authors on request.

3 Results

3.1 Convergence Speed Comparisons

The required number of iterations is compared by the method of Raftery and Lewis (1992). We first simulated a pilot run of 1,000 iterations, and performed the Raftery and Lewis

calculations with $q = 0.025$, $\delta = 0.005$, $s = 0.95$, and $\varepsilon = 0.01$ for all the parameters of each method.

Figure 1 gives the average and maximum of the total number of iterations (in ten thousands) recommended for the fixed- and random-effects of each method by the method of Raftery and Lewis (1992). We see that there is no large difference for each method in fixed-effects. On the other hand, there are huge differences among random-effects.

The rejection Gibbs method (RG) has the smallest recommended number of iterations out of all of the strategies for random-effects; the average recommended number is 975 and the maximum 1,242 as depicted in Figure 1(b). Its average number of rejection-sampling iterations needed per acceptance, however, is between 63 of 14-th subject and 23,610 of 18-th subject, with an overall average of 5,303 in Table 1. This causes RG to have 135 times expensive computer time per iteration more than others.

For RWM, we use a multivariate normal proposal distribution with covariance matrix given by a constant c times the inverse information matrix, where $c \in \{0.25, 0.5, 1, 2, 4\}$. As demonstrated in Bennett et al. (1996), Table 2 shows that the acceptance rate increases as c decreases. Overall, a best of RWM is the $c = 0.5$ version with 33.3% as an acceptance rate.

To improve the speed of the mixing, we implement the data augmentation method for RWM. In order for RWM to satisfy the constraint which is Equation (4), the $c = 1$ version is used to implement the data augmentation method. As mentioned earlier, the convergence rate of any DA scheme is equal to the square of the maximal correlation between the missing variable and the parameter in linear models. Unfortunately, the rate of the convergence is not available for our nonlinear model in the closed form, and hereby its performance can not be inferred properly. For this reason, we assume that the convergence rate of our nonlinear model follows that of the nonlinear model in the previous section, which is Equation (7), and implement the data augmentation method with several versions, which are related to the ratio of the variances r , where $r \in \{10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 0.01, 0.1, 0.5, 0.9\}$. The $r = 10^{-5}$ and $r = 0.01$ versions give smaller recommended numbers of iterations for random-effects, which are respectively 15,552 and 14,453 in average, than other versions as depicted in Figure 1(b), while the $r = 10^{-5}$ version has smaller maximum recommended numbers of iterations than the $r = 0.01$ version, which are 93,177 and 271,362, respectively. As we anticipated, one will not be able to see a clear relationship between the different r 's and the speed of the convergence due to the nonlinear function f , even though it is surely that the speed of the convergence of RWM can be improved by using the DA for the nonlinear model.

We used GMAP to estimate the random-effects for each subject in order to avoid the computation burden due to large dimensions. Each subject has four random-effects, which are $\log V1$, $\log V2$, $\log CL$, and $\log CLI2$. The average and maximum recommended numbers of iterations for random-effects of GMAP are 1,274 and 2,336, respectively, while the average of the average and maximum recommended numbers of iterations for random-effects of other MCMC methods (without RG) are 47,557 and 1,336,186, respectively. Overall, Figure 1 shows that GMAP has the smallest recommended numbers of iteration for both of fixed-effects and random-effects except for RG causing the expensive computation burden. Indeed, GMAP keeps the performance better than others, resulting in converging up to 31 (average) and 536 (maximum) times faster.

3.2 The performance of the estimation and fitting

We conduct a run of 100,000 iterations, with burn-in = 10,000 and thin = 10, and compare the performance of the estimation among RWM with $c = 0.5$, DA with $r = 10^{-5}$, IMH, and GMAP. We adapt the $r = 10^{-5}$ version for DA based on the recommended numbers of iterations in

random-effects, since there is no large difference in fixed-effects. Note that GMAP for random-effects is focused on the distribution of the estimation for the mode, not the distribution of the posterior.

We calculated the mean squared error (MSE) between the logarithms of predicted and observed concentrations as depicted in Figure 2. The MSE of the methods, RWM, DA, IMH, and GMAP, is 0.0257, 0.0163, 0.011, and 0.009, respectively. In other words, GMAP has 2.9, 1.8, 1.2 times smaller MSE than other methods, suggesting that GMAP has a better fitting. It is highly likely due to its fastest MCMC convergence as shown earlier. Note that all the points are near the line in case of IMH and GMAP, while there are several points far from the line for RWM and DA in Figure 2.

The estimates of means and their standard deviations (SD) for fixed- and random-effects are shown by Tables 3 and 4. Table 3 gives the means and SDs for fixed-effects, while Tables 4 shows us the means and SDs for random-effects for each subject. The SDs of random-effects for GMAP in Table 4 are estimated using the Fisher information matrix as mentioned in Section 2.5. As all of the results are overlapped within 2SDs, they suggest that all of the estimates are comparable.

3.3. Simulation Validation

To establish validity of the proposed method GMAP, this section presents a computer simulation study based on the MDZ PK model whose structures are presented in the two-compartmental model as described in Section 2. For the simulation experiment, a dataset was generated from the MDZ PK model, with the number of subjects (individuals) $M = 5$ and true values of both of fixed- and random-effect as depicted in Table 6. Note the blood sampling time points are 0.25, 0.5, 0.75, 0.9, 1, 2, 5, 7, 9 h after dose and MDZ is dosed at 4.7 mg. We then applied the proposed algorithm GMAP, as well as RWM with $c = 0.5$, DA with $r = 10^{-5}$, and IMH, to the MDZ PK model. The initial values of parameters were taken as the true values. The procedures were repeated 100 times. The average (maximum) of acceptance rates for RWM, DA, and IMH were 26.95% (52.54%), 18.37% (37.61%), and 11.53% (32.05%), respectively.

Table 5 gives the average and maximum of the total number of iterations (in ten thousands) recommended for the fixed- and random-effects of each method by the method of Raftery and Lewis (1992) for a run of 1,000 iterations with $q = 0.025$, $\delta = 0.005$, $s = 0.95$, and $\varepsilon = 0.01$. As shown in the previous experiment with the real dataset, simulation studies demonstrate the superiority of GMAP in random-effects, resulting in converging up to 33 (average) and 137 (maximum) times faster than RWM. Not surprisingly, there is a little difference among the methods in fixed-effects as depicted in Table 5.

The coverage probabilities based on the 90% interval estimates (90% CP) are presented, along with average bias (true value – estimate) and standard errors of each estimate, in Table 6. The classical 90% C.I.'s are computed based on the normal distribution and then 90% CP's are inferred. Note that the standard errors of GMAP for 90% CP's are computed by Fisher's information matrix as mentioned earlier in Section 2.5. The average biases of all the parameters are less than 1.0 except for $\log CL$, w_{33} , and $\log CL_5$ for which 90% CP's are less than 10%, suggesting that it might be necessary for those to run more iteration to converge than other parameters. After not considering three parameters $\log CL$, w_{33} , and $\log CL_5$, the averages of 90% CP's of RWM, DA, IMH, and GMAP are 99.86%, 99.71%, 100%, and 98.14% in fixed-effects, respectively. As for random-effects, 87.42%, 86.63%, 86.42%, and 87.63% are the average 90% CP's for each method. Overall, the results given in Table 6 indicate that the estimates of the parameters of all the methods are very similar, while GMAP converges faster than other methods in Table 5.

4. Discussion and Conclusion

In this paper, we compare the various MCMC methods, including DA, with GMAP, using a PK dataset, which not only confirms the results of Bennett et al. (1996) but also suggests that both DA and GMAP are capable of speeding up the convergence.

Our computation experiment shows that DA can certainly improve the speed of the convergence of RWM in terms of the number of iterations to converge in our nonlinear model. The difficulty to implement DA for the nonlinear model, however, is that it is not simple and clear to find the good candidate for DA since there is no closed form in the convergence rate as mentioned earlier. As DA is always implemented to a specific MCMC algorithm, a more relevant question that can be addressed is whether DA can improve the speed of IMH algorithm. To implement DA into IMH, the only thing to do is to use IMH instead of RWM in *Step 2* in Section 2.4. Because DA can improve the mixing of within-subject variance component and subject-specific fitting, we expect that DA will be able to improve the speed of convergence of IMH. The improvement will be very similar to RWM, as their DA implementations share the same mechanism.

When prior data or information is not available, their prior distributions can be specified as uniform distributions or normal/Wishart distributions with large variations. As a matter of fact, in Li et al. (2007), uniform distributions with pre-specified bounds were employed as prior distributions. Their argument was that based on *in-vitro* data and clinical experiences, the PK parameters and their between-subject variances have to be positive and won't exceed some up bounds. This uniform distribution assumption led to a conjugate prior-posterior, hence, it wouldn't affect the formulations of either our proposed GMAP or other MCMC methods. On the other hand, if normal distribution and Wishart distribution with large variations are specified as priors for PK parameters and their variance components, they won't change the formulation of our model at all.

Estimation approaches, such as EM, stochastic approximation EM (SAEM; Delyon et al., 1999), and algorithms implemented in NONMEM are all approximate MLE. MLE is a point-wise estimator, i.e. the mode of a likelihood function. The convergence relies on assessing the mode estimate. In particular, in SAEM, a Monte-Carlo step is implemented to approximate the integral of the E-step of the point estimate. On the other hand, GMAP, like every MCMC approach, is a stochastic approach that estimates the distributions PK parameters and their variance components. The convergence relies on assessing the percentiles of posterior distributions. Both GMAP and MCMCs estimate posterior distributions of fixed effects. Unlike traditional MCMCs, GMAP only estimates the mode distribution of subject-specific PK parameters, while the other MCMCs estimate the whole distribution of subject-specific PK parameters. This compromise made by GMAP gains us the speed of convergence.

The experiment and simulation illustrate how GMAP can lead to very efficient algorithms when used in conjunction with Gibbs sampling in terms of the speed of the convergence. GMAP out performs all the other MCMC approaches. It results in converging up to **31** (average) and **536** (maximum) times faster. If all of the MCMC and GMAP are run with the same number of iterations, GMAP shows the best fitting, i.e. smallest MSE. In other words, GMAP has 2.9, 1.8, 1.2 times smaller MSE than RWM, DA, and IMH, respectively.

We think that the key of this speed improvement is GMAP focuses on the posterior distribution of a random effect' *mode*, rather than a random effect's posterior distribution. As long as the inference is primarily for population average parameters, GMAP is an optimal Bayesian approach, especially for its speed.

Acknowledgments

Drs. Lang Li and Stephen D. Hall's researches are supported by NIH grants, R01 GM74217 (LL), R01 GM67308 (SH), FD-T-001756 (SH).

Appendix I. The rate of the convergence of DA in a linear model

We consider the linear hierarchical model to describe basic ideas of the DA algorithm. The conditional probability model for the observations is given by

$$[y_{ij}|\theta_i, \tau^{-1}] \mathbf{N}(\theta_i, \tau^{-1})$$

where $i = 1, \dots, M$; $j = 1, \dots, n_i$. The θ_i is the random-effect associated with the i th subject, and it is assumed to be normally distributed with mean θ and variance σ^{-1} . Here τ , θ , and σ^{-1} are the fixed-effects.

This linear model can be written in the DA algorithm by introducing a missing data z_{ij} for each observation. The DA algorithm is obtained by specifying

$$[y_{ij}|\theta_i, \tau^{-1}, z_{ij}] \mathbf{N}(\theta_i + z_{ij}, \tau^{-1} - \xi^{-1})$$

and

$$[z_{ij}|\theta_i, \tau^{-1}] \mathbf{N}(0, \xi^{-1})$$

where $i = 1, \dots, M$; $j = 1, \dots, n_i$. Integrating over the z_{ij} is definitely satisfied with Equation (4). Then the DA algorithm is as follows:

Step 1: Draw z_{ij} conditional on $(\theta_i, \tau^{-1}, y_{ij})$, where $i = 1, \dots, M$; $j = 1, \dots, n_i$; that is, draw

$$[z_{ij}|\theta_i, \tau^{-1}, y_{ij}] \mathbf{N}\left(\frac{\xi^{-1}}{\tau^{-1}}(y_{ij} - \theta_i), \xi^{-1}\left(1 - \frac{\xi^{-1}}{\tau^{-1}}\right)\right)$$

Step 2: Draw θ_i conditional on $(\tau^{-1}, z_{ij}, y_{ij}, j = 1, \dots, n_i)$ with the prior $\mathbf{N}(\theta, \sigma^{-1})$, where $i = 1, \dots, M$; that is, draw

$$\left[\theta_i|\tau^{-1}, z_{ij}, y_{ij}, j=1, \dots, n_i\right] \mathbf{N}\left(\left(\frac{n_i}{\tau^{-1} - \xi^{-1}} + \frac{1}{\sigma^{-1}}\right)^{-1} \left(\frac{\sum_j (y_{ij} - z_{ij})}{\tau^{-1} - \xi^{-1}} + \frac{\theta}{\sigma^{-1}}\right), \left(\frac{n_i}{\tau^{-1} - \xi^{-1}} + \frac{1}{\sigma^{-1}}\right)^{-1}\right).$$

The rate of convergence for this linear model then is

$$\begin{aligned} r_0 &= \frac{n_i \frac{\xi^{-1}}{\tau^{-1}} \sigma^{-1}}{\tau^{-1} \left(\frac{n_i \sigma^{-1}}{\tau^{-1}} + 1 - \frac{\xi^{-1}}{\tau^{-1}}\right)} \\ &= \frac{n_i r \sigma^{-1}}{\tau^{-1} \left(\frac{n_i \sigma^{-1}}{\tau^{-1}} + 1 - r\right)}, \text{ where } r = \frac{\xi^{-1}}{\tau^{-1}}, \end{aligned}$$

using

$$\left[\theta_i | \tau^{-1}, y_{ij}, j=1, \dots, n_i \right] \\ \mathbf{N} \left(\left(\frac{n_i}{\tau^{-1}} + \frac{1}{\sigma^{-1}} \right)^{-1} \left(\frac{\sum y_{ij}}{\tau^{-1}} + \frac{\theta}{\sigma^{-1}} \right), \left(\frac{n_i}{\tau^{-1}} + \frac{1}{\sigma^{-1}} \right)^{-1} \right).$$

It is clear that the convergence rates of this linear model, r_0 , decreases as $r \left(= \frac{\xi^{-1}}{\tau^{-1}} \right)$, which is the ratio of the variances, goes to 0.

Appendix II. Standard deviations for GMAP

The standard deviation for GMAP estimates is obtained using the Fisher information matrix. In other words, the standard deviation of estimates $\widehat{\theta}_i$ of random-effects for each subject i can be estimated by the inverse of the observed information matrix.

Let $\ell_i(\theta_i | \mathbf{g})$ denote the log-posterior function of a random-effect θ_i for the i th subject given $\mathbf{y}_i$, $\widehat{\theta}$, $\widehat{\Sigma}_p$, $\widehat{\tau}$, while ignoring terms which do not depend on θ_i :

$$\ell_i(\theta_i | \mathbf{g}) = -\frac{\widehat{\tau}}{2} \sum_{j=1}^{n_i} \left(\log y_{ij} - \log f(\theta_i, t_{ij}) \right)^2 - \frac{1}{2} (\theta_i - \widehat{\theta})^T \widehat{\Sigma}_p^{-1} (\theta_i - \widehat{\theta}).$$

In order to compute the Fisher information matrix $-E \left[\frac{\partial^2}{(\partial \theta_i)^2} \ell_i(\theta_i | \mathbf{g}) \right]$, we need the second derivatives of the log-posterior function:

$$\frac{\partial^2}{(\partial \theta_i)^2} \ell_i(\theta_i | \mathbf{g}) = \left[\cdot \ell_{obs_i}^{ab} \right]_{a,b} + \frac{1}{2} (\widehat{\Sigma}_p^{-1})^T + \frac{1}{2} \widehat{\Sigma}_p^{-1},$$

where

$$\cdot \ell_{obs_i}^{ab} = \frac{\partial^2 \ell_i(\theta_i | \mathbf{g})}{\partial \theta_i^b \partial \theta_i^a} = \widehat{\tau} \sum_{j=1}^{n_i} \left(\log y_{ij} - \log f(\theta_i, t_{ij}) \right) \frac{\partial \log f(\theta_i, t_{ij})}{\partial \theta_i^a} - \widehat{\tau} \sum_{j=1}^{n_i} \frac{\partial \log f(\theta_i, t_{ij})}{\partial \theta_i^b} \frac{\partial \log f(\theta_i, t_{ij})}{\partial \theta_i^a},$$

$\widehat{\theta}_i^k$ is the k th entry of the vector θ_i , and $\left[\cdot \ell_{obs_i}^{ab} \right]_{a,b}$ is a $p \times p$ matrix,

Then the Fisher information matrix becomes

$$\left[\widehat{\tau} \sum_{j=1}^{n_i} \frac{\partial \log f(\widehat{\theta}_i, t_{ij})}{\partial \theta_i^b} \frac{\partial \log f(\widehat{\theta}_i, t_{ij})}{\partial \theta_i^a} \right]_{a,b} + \frac{1}{2} \left[(\widehat{\Sigma}_p^{-1})^T + \widehat{\Sigma}_p^{-1} \right],$$

resulting that the approximate variances and covariances are found by taking the inverse of the Fisher information matrix above and so are the standard deviations.

References

- Aarons L. Population Pharmacokinetics: Theory and Practice. *British Journal of Clinical Pharmacology* 1991;32:669–670. [PubMed: 1768557]
- Beal SL, Sheiner LB. Estimating population kinetics. *Crit Rev Biomed Eng* 1982;8:195–222. [PubMed: 6754254]
- Bennett, JE.; Racine-Poon, A.; Wakefield, JC. MCMC for Nonlinear Hierarchical Models. In: Gilks, WR.; Richardson, S.; Spiegelhalter, DJ., editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall; London: 1996. p. 339-337.
- Booth J, Hobert JP. Maximizing generalized linear model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society B* 1999;61:265–285.
- Broniatowski M, Celeux G, Diebolt J. Reconnaissance De Melanges De Densites Par Un Algorithme D'apprentissage Probabiliste. *Data Analysis and Informatics* 1983;3:359–374.
- Broyden CG. The Convergence of a Class of Double-Rank Minimization Algorithms: 2. The New Algorithm. *IMA Journal of Applied Mathematics* 1970;6:222–231.
- Celeux G, Diebolt J. The SEM Algorithm: A Probabilistic Teacher Algorithm Derived from the EM Algorithm for the Mixture Problem. *Computational Statistics Quarterly* 1985;2:73–82.
- Celeux G, Diebolt J. A Stochastic Approximation Type EM Algorithm for the Mixture Problem. *Stochastics: An International Journal of Probability and Stochastic Processes* 1992;41:119–134.
- Celeux G, Chauveau D, Diebolt J. Stochastic Versions of the EM Algorithm: An Experimental Study in the Mixture Case. *Journal of Statistical Computation and Simulation* 1996;55:287–314.
- Chauveau D. A Stochastic EM Algorithm for Mixtures with Censored Data. *Journal of Statistical Planning and Inference* 1995;46:1–25.
- Damien P, Wakefield J, Walker S. Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variables. *Journal of the Royal Statistical Society B* 1999;61:331–344.
- Davidian M, Gallant AR. Smooth Nonparametric Maximum Likelihood Estimation for Population Pharmacokinetics, with Application to Quinidine. *Journal of Pharmacokinetics and Pharmacodynamics* 1992;20:529–556.
- Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics* 1999;27:94–128.
- Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society B* 1977;39:1–38.
- Diebolt J, Celeux G. Asymptotic Properties of A Stochastic EM Algorithm for Estimating Mixing Proportions. *Stochastic Models* 1993;9:599–613.
- Diebolt, J.; Ip, EHS. Stochastic EM: Method and Application. In: Gilks, WR.; Richardson, S.; Spiegelhalter, DJ., editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall; London: 1996. p. 259-273.
- van Dyk DA, Meng XL. The Art of Data Augmentation. *Journal of Computational and Graphical Statistics* 2001;10:1–50.
- Ette, EI.; Miller, R.; Gillespie, WR.; Huang, SM.; Lesko, L.; Williams, R. The Population Approach: FDA Experience. In: Concentration and Dose. Balant, LP.; Aarons, L., editors. *The Population Approach: Measuring and Managing Variability in Response*. Commission of the European Communities, European Cooperation in the field of Scientific and Technical Research; Brussels: 1997. p. 271-275.
- Fox JP. Stochastic EM for Estimating the Parameters of a Multilevel IRT Model. *British Journal of Mathematical and Statistical Psychology* 2003;56:65–81. [PubMed: 12803822]
- Geman S, Geman D. Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984;6:721–741.
- Gilks, WR. Full Conditional Distributions. In: Gilks, WR.; Richardson, S.; Spiegelhalter, DJ., editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall; London: 1996. p. 75-88.
- Gilks, WR.; Roberts, GO. Strategies for Improving MCMC. In: Gilks, WR.; Richardson, S.; Spiegelhalter, DJ., editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall; London: 1996. p. 89-114.

- Hamon A, Iovleff S. Three Stochastic Versions of the EM Algorithm for Estimating Longitudinal Rasch Model. *Communications in Statistics* 2003;32:275–295.
- Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 1970;57:97–109.
- Li L, Yu M, Chin R, Lucksiri A, Flockhart DA, Hall SD. Drug-drug interaction prediction: A Bayesian meta-analysis approach. *Statistics in Medicine* 2007;26:3700–3721. [PubMed: 17357990]
- Lindstrom MJ, Bates DM. Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association* 1988;83:1014–1022.
- Lindstrom MJ, Bates DM. Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics* 1990;46:673–687. [PubMed: 2242409]
- Liu C, Rubin DB. The ECME Algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence. *Biometrika* 1994;81:633–648.
- Liu J, Wong WH, Kong A. Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes. *Biometrika* 1994;81:27–40.
- Liu JS, Wong WH, Kong A. Covariance Structure and Convergence Rate of the Gibbs Sampler with Various Scans. *Journal of the Royal Statistical Society B* 1995;57:157–169.
- Mallet A. A Maximum Likelihood Estimation Method for Random Coefficient Regression Models. *Biometrika* 1986;73:645–656.
- McCulloch CE. Maximum likelihood algorithms for generalized linear mixed Models. *Journal of the American Statistical Association* 1997;92:162–170.
- Meng XLI, Rubin DB. Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika* 1993;80:267–278.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computer Machines. *Journal of Chemical Physics* 1953;21:1087–1092.
- Nielsen SF. The Stochastic EM Algorithm: Estimation and Asymptotic Results. *Bernoulli* 2000;6:457–489.
- Pinheiro JC, Bates DM. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics* 1995;4:12–35.
- Racine-Poon, A.; Wakefield, J. Bayesian Analysis of Population Pharmacokinetic and Instantaneous Pharmacodynamic Relationships. Berry, D.; Stangl, D., editors. *Bayesian Biostatistics*; Marcel-Dekker: 1996. p. 355–388.
- Raftery, AE.; Lewis, S. How Many Iterations in the Gibbs Sampler?. In: Bernardo, M.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. *Bayesian Biostatistics 4*. Oxford: Oxford University Press; Oxford: 1992. p. 763–773.
- Raftery, AE.; Lewis, S. Implementing MCMC. Markov Chain Monte Carlo in Practice. In: Gilks, WR.; Richardson, S.; Spiegelhalter, DJ., editors. *Chapman and Hall*; London: 1996. p. 115–130.
- Roberts, GO. Markov Chain Concepts Related to Sampling Algorithms. Markov Chain Monte Carlo in Practice. In: Gilks, WR.; Richardson, S.; Spiegelhalter, DJ., editors. *Chapman and Hall*; London: 1996. p. 45–57.
- Sheiner LB, Beal SL. Evaluation of Methods for Estimating Population Pharmacokinetics Parameters. I. Michaelis-Menten Model: Routine Clinical Pharmacokinetic Data. *Journal of Pharmacokinetics and Biopharmaceutics* 1980;8:553–571. [PubMed: 7229908]
- Sheiner LB, Beal SL. Bayesian Individualization of Pharmacokinetics: Simple Implementation and Comparison with Non-Bayesian Methods. *Journal of Pharmaceutical Science* 1982;71:1344–1348.
- Sheiner LB, Beal SL. Evaluation of Methods for Estimating Population Pharmacokinetic Parameters. III: Monoexponential Model; Routine Clinical Pharmacokinetic Data. *Journal of Pharmacokinetics and Biopharmaceutics* 1983;11:303–319. [PubMed: 6644555]
- Sheiner LB, Grasela TH. An Introduction to Mixed Effect Modeling: Concepts, Definitions, and Justification. *Journal of Pharmacokinetics and Pharmacodynamics* 1991;19:11S–24S.
- Sheiner LB, Rosenberg B, Melmon KL. Modelling of Individual Pharmacokinetics for Computer-Aided Drug Dosage. *Computers and Biomedical Research* 1972;5:411–459. [PubMed: 4634367]
- Smith AFM, Gelfand AE. Bayesian Statistics without Tears: A Sampling-Resampling Perspective. *The American Statistician* 1992;46:84–88.

- Smith AFM, Roberts GO. Bayesian Computation via the Gibbs Sampler and Relater Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society B* 1993;55:3–23.
- Tanner, MA. Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions. Third Edition. Springer; New York: 1991.
- Tanner MA, Wong WH. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* 1987;82:528–550.
- Tierney L. Markov Chains for Exploring Posterior Distributions (with Discussion). *Annals of Statistics* 1994;22:1701–1762.
- Tregouet DA, Tired L. Cox Proportional Hazards Survival Regression in Haplotype-based Association Analysis using the Stochastic-EM Algorithm. *European Journal of Human Genetics* 2004;12:971–974. [PubMed: 15241485]
- Tregouet DA, Escolano S, Tired L, Mallet A, Golmard JL. A New Algorithm for Haplotype-based Association Analysis: the Stochastic-EM Algorithm. *Annals of Human Genetics* 2004;68:165–177. [PubMed: 15008795]
- Vonesh EF, Carter RL. Mixed-Effects Nonlinear Regression for Unbalanced Repeated Measures. *Biometrics* 1992;48:1–17. [PubMed: 1581479]
- Wakefield JC, Smith AFM, Racine-Poon A, Gelfand AE. Bayesian Analysis of Linear and Non-Linear Population Models by Using the Gibbs Sampler. *Applied Statistics* 1994;43:201–221.
- Wei GCG, Tanner MA. A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association* 1990;85:699–704.
- Wolfinger R. Laplace's Approximation for Nonlinear Mixed Models. *Biometrika* 1993;80:791–795.

(a) Fixed-Effects

(b) Random-Effects **Figure 1.**

Recommended numbers of iterations (in ten thousands) for each method in (a) fixed-effects and (b) random-effects. The left ordinate denotes the number of iteration in average and the right ordinate denotes the number of iteration in maximum. RWM is implemented using several constants, $(c1, c2, c3, c4, c5) = (0.25, 0.5, 1, 2, 4)$, and the various ratios of variance, $(r1, r2, r3, r4, r5, r6, r7, r8, r9) = (10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 0.01, 0.1, 0.5, 0.9)$, are used to implement DA. MaxIter is the maximum of recommended numbers of iterations for each method.

Figure 2.

Predicted versus observed concentrations for each method. $\log(\text{pred.conc})$ is the logarithm of the predicted concentrations; $\log(\text{obs.conc})$ is the logarithm of the observed concentrations; MSE is the mean squared error.

Table 1

The average number of rejection-sampling iterations needed per acceptance for RG

Subject	1	2	3	4	5	6	7	8
Iteration	1,707	3,572	528	14,813	3,563	12,200	4,560	4,292
Subject	9	10	11	12	13	14	15	16
Iteration	1,604	5,945	1,721	10,345	1,601	63	243	4,247
Subject	17	18	19	20	21	22	23	24
Iteration	9,060	<u>23,610</u>	6,180	3,598	5,008	132	611	8,058

Table 2

Acceptance rates (in percentages) for each method.

Subject	RWM					DA					IMH				
	c=0.25	c=0.5	c=1	c=2	c=4	r=10 ⁻⁸	r=10 ⁻⁷	r=10 ⁻⁶	r=10 ⁻⁵	r=10 ⁻⁴	r=0.01	r=0.1	r=0.5	r=0.9	
1	36.7	20.9	16.9	9.4	3.2	12.3	17.7	11.9	15.4	15.7	13.8	12.1	14.0	1.9	11.0
2	47.3	34.5	20.1	14.4	6.7	21.7	22.7	18.6	19.0	22.2	22.9	22.0	13.3	4.0	1.0
3	22.2	18.3	7.6	5.2	2.8	10.9	9.3	9.9	8.8	12.3	11.1	6.4	9.0	2.0	0.9
4	42.0	28.4	19.6	10.9	5.6	20.9	19.8	18.1	18.9	19.2	18.0	19.1	14.8	2.1	17.5
5	58.8	46.1	28.0	19.4	6.9	29.3	34.0	32.6	32.5	30.1	31.3	28.1	19.4	3.1	52.7
6	57.8	44.6	33.5	19.6	9.4	32.0	32.6	32.5	29.6	36.0	31.8	30.7	17.0	3.9	63.3
7	40.1	26.1	11.2	9.8	3.8	13.6	16.9	13.9	16.2	16.6	14.2	14.9	11.4	1.9	19.6
8	61.8	51.8	32.8	20.4	10.4	33.9	35.3	35.6	31.4	34.6	31.1	30.8	22.1	3.2	16.6
9	47.3	44.3	25.9	14.7	10.5	25.7	17.4	27.4	25.0	28.0	29.2	26.5	14.3	3.1	44.9
10	53.8	39.6	27.1	17.5	10.4	30.2	28.3	22.2	25.6	24.5	28.7	22.9	17.2	2.2	45.8
11	48.7	27.9	22.7	8.6	6.4	16.9	18.1	24.8	19.2	22.6	22.3	23.9	14.2	2.5	38.2
12	41.0	33.3	20.8	12.2	5.8	17.0	23.7	19.7	18.0	23.3	20.9	16.2	13.0	2.7	36.3
13	57.5	38.2	33.8	16.3	10.5	30.0	30.0	28.8	28.8	29.9	28.6	29.5	17.2	3.3	25.1
14	2.3	1.2	0.4	0.9	0.7	0.9	0.9	0.9	1.5	0.9	0.7	0.5	0.7	0.3	0.8
15	18.7	15.6	8.1	6.4	2.8	11.3	10.2	8.0	7.2	12.1	5.7	5.8	2.0	1.7	2.4
16	63.3	50.3	36.2	20.2	10.9	36.3	34.0	33.9	34.0	34.8	32.8	35.5	23.7	3.0	49.9
17	46.1	34.7	17.3	10.3	3.8	19.4	21.1	20.5	19.2	20.1	21.8	18.2	13.9	2.6	8.0
18	42.2	26.0	20.3	9.7	5.7	15.5	19.6	15.2	17.7	19.7	17.4	15.9	16.4	1.9	19.5
19	51.6	38.3	25.9	15.9	7.3	25.1	24.7	24.6	24.3	23.0	25.6	21.4	12.7	2.7	23.6
20	62.8	48.2	33.3	20.9	9.5	34.2	33.2	34.3	31.4	33.0	32.9	31.9	20.7	3.5	35.8
21	42.0	25.4	19.4	8.8	4.0	15.1	22.6	14.6	7.9	19.5	13.6	17.7	11.7	3.4	23.2
22	38.1	31.9	15.0	12.5	2.9	24.5	22.6	15.4	21.1	20.3	24.3	17.6	11.7	2.8	0.7
23	45.4	28.7	19.9	11.1	5.5	19.2	21.2	17.1	18.8	19.8	19.0	16.4	10.6	2.4	4.2
24	53.2	44.2	28.2	16.0	8.7	34.2	30.0	29.7	32.8	32.3	32.0	30.0	18.6	3.2	47.4
Average	45.0	33.3	21.8	13.0	6.4	22.1	22.7	21.3	21.0	22.9	22.1	20.6	14.2	2.6	24.5

Table 3

The estimates of means and standard deviations of the fixed-effects. w_{ij} is an (i,j) -th entry of $\Sigma_p = [w_{ij}]_{i,j}$. $E(X)$ is the mean of a parameter; SD is the standard deviation of a parameter.

Parameters	RWM ($c=0.5$)		DA ($r=10^{-5}$)		IMH		GMAP	
	E(X)	SD	E(X)	SD	E(X)	SD	E(X)	SD
$\log V1$	3.135	0.236	3.140	0.230	3.027	0.217	2.887	0.198
$\log V2$	3.794	0.284	3.709	0.281	3.748	0.285	3.623	0.260
$\log CL$	3.180	0.287	3.184	0.287	3.175	0.286	3.164	0.287
$\log CL12$	3.197	0.276	3.197	0.286	3.076	0.282	3.177	0.263
w_{11}	1.245	0.409	1.229	0.404	1.315	0.431	1.186	0.373
w_{12}	0.113	0.372	0.022	0.368	0.201	0.377	0.081	0.322
w_{13}	-1.728	0.748	-1.710	0.738	-1.674	0.776	-1.683	0.734
w_{14}	0.574	0.388	0.576	0.391	0.656	0.403	0.712	0.376
w_{22}	2.121	0.691	2.041	0.662	2.205	0.730	1.848	0.572
w_{23}	0.385	0.861	0.430	0.839	0.457	0.888	0.490	0.800
w_{24}	-1.408	0.574	-1.437	0.561	-1.400	0.584	-1.177	0.491
w_{33}	7.730	2.306	7.714	2.344	7.752	2.339	7.713	2.335
w_{34}	-1.465	0.919	-1.465	0.915	-1.412	0.951	-1.446	0.893
w_{44}	2.145	0.681	2.186	0.692	2.300	0.719	2.072	0.630
τ^{-1}	0.018	0.003	0.018	0.003	0.017	0.002	0.009	0.001

The estimates of means and standard deviations of the random-effects for each subject. The Fisher information matrix is used to estimate the standard deviations for GMAP. E(X) is the mean of a parameter; SD is the standard deviation of a parameter.

Table 4

subject	logV1						logV2					
	RWM ($c=0.5$)			IMH			RWM ($c=0.5$)			IMH		
	E(X)	SD	E(X)	E(X)	SD	SD	E(X)	SD	E(X)	E(X)	SD	SD
1	3.22	0.61	3.25	0.60	3.10	0.39	2.88	0.48	3.81	0.18	3.79	0.18
2	1.67	0.50	1.68	0.48	1.09	0.26	0.98	0.44	2.64	0.35	2.65	0.34
3	3.91	0.41	3.65	0.57	3.03	0.54	2.72	0.68	5.14	1.14	4.09	0.85
4	2.53	0.48	2.54	0.47	2.27	0.41	2.13	0.42	3.33	0.23	3.34	0.23
5	3.04	0.41	3.05	0.40	3.03	0.38	2.96	0.34	3.90	0.17	3.91	0.17
6	2.62	0.40	2.62	0.39	2.55	0.40	2.41	0.33	3.50	0.19	3.49	0.18
7	2.80	0.61	2.74	0.58	2.52	0.53	2.27	0.60	3.66	0.19	3.66	0.19
8	3.14	0.21	3.14	0.21	3.11	0.20	3.09	0.15	4.03	0.32	4.01	0.30
9	3.67	0.46	3.59	0.44	3.45	0.38	3.51	0.30	4.07	1.06	3.71	0.68
10	3.36	0.65	3.30	0.70	3.37	0.59	3.25	0.45	4.15	0.20	4.13	0.24
11	3.85	0.53	3.82	0.52	3.73	0.29	3.65	0.31	4.15	0.94	4.03	0.99
12	2.74	0.48	2.73	0.47	2.50	0.45	2.41	0.47	3.19	0.16	3.19	0.15
13	3.30	0.36	3.31	0.37	3.25	0.36	3.28	0.26	3.93	0.21	3.93	0.21
14	3.46	0.37	3.80	0.19	3.97	0.11	3.05	0.81	3.60	1.48	2.77	1.80
15	3.21	0.53	3.21	0.54	3.10	0.43	3.16	0.42	3.23	0.18	3.23	0.18
16	3.44	0.40	3.43	0.41	3.45	0.39	3.45	0.25	4.17	0.21	4.16	0.20
17	4.07	0.17	4.08	0.17	4.06	0.19	4.10	0.13	5.25	0.41	5.21	0.39
18	2.48	0.50	2.50	0.50	2.24	0.44	2.04	0.49	3.33	0.30	3.34	0.30
19	2.81	0.44	2.83	0.45	2.82	0.44	2.67	0.40	3.67	0.21	3.67	0.21
20	3.39	0.29	3.39	0.29	3.50	0.31	3.39	0.22	4.10	0.20	4.09	0.19
21	2.91	0.59	2.98	0.60	2.73	0.55	2.50	0.61	3.39	0.20	3.39	0.19
22	3.26	0.76	3.23	0.66	3.47	0.08	2.97	0.68	3.21	0.58	3.36	0.25
23	3.10	0.55	3.08	0.53	2.84	0.48	2.79	0.38	3.86	0.25	3.84	0.25
24	3.58	0.58	3.52	0.62	3.70	0.43	3.66	0.33	4.35	0.21	4.33	0.23

Subject	logCL						logCL/2											
	RWM ($c=0.5$)			IMH			GMAP			RWM ($c=0.5$)			IMH			GMAP		
	E(X)	SD	E(X)	SD	E(X)	SD	E(X)	SD	E(X)	SD	E(X)	SD	E(X)	SD	E(X)	SD	E(X)	SD
1	3.17	0.13	3.18	0.12	3.11	0.08	3.13	0.10	3.40	0.33	3.42	0.30	3.39	0.20	3.54	0.22		
2	2.67	0.21	2.67	0.20	2.43	0.14	2.36	0.21	2.20	0.48	2.22	0.47	1.47	0.28	1.50	0.46		
3	2.83	0.36	3.06	0.15	3.03	0.17	2.94	0.22	2.67	0.36	2.60	0.42	2.83	0.39	2.71	0.42		
4	3.13	0.15	3.14	0.15	3.05	0.15	3.01	0.15	2.87	0.31	2.88	0.30	2.73	0.34	2.65	0.33		
5	2.96	0.09	2.96	0.09	2.96	0.08	2.96	0.07	3.29	0.20	3.30	0.20	3.30	0.19	3.33	0.14		
6	2.86	0.11	2.86	0.11	2.84	0.11	2.81	0.09	2.85	0.23	2.85	0.22	2.82	0.24	2.79	0.20		
7	3.04	0.14	3.04	0.14	3.00	0.15	2.94	0.16	3.45	0.36	3.44	0.36	3.36	0.40	3.24	0.43		
8	3.13	0.08	3.13	0.08	3.14	0.07	3.14	0.05	2.64	0.15	2.63	0.15	2.62	0.14	2.60	0.11		
9	3.01	0.22	3.09	0.12	3.12	0.08	3.14	0.07	2.82	0.50	2.81	0.48	2.94	0.39	2.93	0.24		
10	3.55	0.15	3.53	0.17	3.55	0.14	3.54	0.10	3.75	0.32	3.71	0.36	3.77	0.31	3.86	0.22		
11	3.33	0.13	3.34	0.14	3.42	0.06	3.41	0.06	2.99	0.72	3.05	0.77	3.45	0.48	3.56	0.29		
12	3.05	0.13	3.04	0.13	2.98	0.13	2.97	0.13	3.01	0.28	3.01	0.27	2.95	0.30	2.95	0.26		
13	3.03	0.08	3.03	0.08	3.03	0.08	3.05	0.05	3.20	0.19	3.19	0.19	3.21	0.18	3.26	0.12		
14	3.03	0.09	3.09	0.06	3.00	0.07	3.02	0.16	3.60	1.44	3.92	1.88	1.19	1.22	4.07	1.01		
15	3.32	0.13	3.32	0.13	3.32	0.12	3.32	0.09	3.27	0.42	3.23	0.44	3.36	0.27	3.40	0.18		
16	3.46	0.10	3.46	0.10	3.46	0.09	3.48	0.06	3.39	0.19	3.38	0.19	3.39	0.18	3.45	0.12		
17	3.65	0.15	3.67	0.13	3.66	0.12	3.68	0.10	3.54	0.16	3.54	0.15	3.53	0.14	3.56	0.11		
18	3.41	0.19	3.42	0.19	3.32	0.18	3.25	0.21	3.02	0.41	3.04	0.42	2.80	0.40	2.67	0.46		
19	3.37	0.14	3.38	0.14	3.36	0.14	3.34	0.13	3.18	0.29	3.18	0.29	3.14	0.28	3.14	0.27		
20	3.37	0.08	3.38	0.08	3.36	0.07	3.39	0.05	3.23	0.17	3.23	0.16	3.20	0.16	3.26	0.11		
21	3.43	0.17	3.45	0.16	3.39	0.16	3.33	0.19	3.41	0.36	3.42	0.35	3.36	0.39	3.28	0.42		
22	3.60	0.18	3.60	0.16	3.62	0.03	3.56	0.17	3.91	0.77	3.74	0.41	3.84	0.07	3.74	0.33		
23	3.53	0.16	3.53	0.16	3.48	0.16	3.46	0.13	3.24	0.29	3.24	0.29	3.23	0.35	3.19	0.27		
24	3.59	0.13	3.57	0.14	3.61	0.09	3.62	0.07	3.81	0.27	3.79	0.30	3.88	0.20	3.95	0.13		

Simulation Studies: recommended numbers of iterations (in ten thousands) for each method of fixed-effects and random-effects in average and maximum.

Table 5

	RWM	DA	IMH	GMAP
Fixed-Effects	Average	0.1021	0.1020	0.1020
	Maximum	0.1304	0.1358	0.1343
Random-Effects	Average	3.751	7.417	20.95
	Maximum	29.04	67.32	163.1

Table 6

Simulation studies: the average biases (true value - estimate), standard errors of the estimates, and coverage probability based on 90% C.I.

	TRUE	RWM			DA			IMH			GMAP		
		Bias	SE	90% CP	Bias	SE	90% CP	Bias	SE	90% CP	Bias	SE	90% CP
Fixed-Effects	$\log V_I$	3.1350											
		-0.0012	0.0181	100	-0.0042	0.0192	100	0.0009	0.0186	100	0.0086	0.0170	100
	$\log V_2$	3.7940											
		0.8945	0.0425	99	0.9012	0.0391	98	0.8805	0.0353	100	0.9455	0.0556	89
	$\log CL$	3.1800											
		1.2103	0.0700	0	1.1972	0.0659	0	1.2686	0.0697	0	1.1125	0.0693	0
	$\log CL_{I2}$	3.1970											
		-0.4923	0.1029	100	-0.4945	0.1045	100	-0.4658	0.0945	100	-0.4301	0.0798	99
	W_{II}	1.2450											
		-0.6399	0.0550	100	-0.6419	0.0498	100	-0.6387	0.0546	100	-0.6368	0.0572	100
Random-Effects	w_{22}	2.1210											
		-0.6023	0.1156	100	-0.5930	0.1205	100	-0.5900	0.1046	100	-0.6290	0.1903	100
	w_{33}	7.7300											
		5.8992	0.3813	2	5.9713	0.3358	1	5.5315	0.4611	10	6.3648	0.2910	0
	w_{44}	2.1450											
		0.2027	0.5243	100	0.2523	0.4248	100	0.3653	0.3372	100	0.7139	0.2316	99
	$\tau\tau^{-1}$	0.0003											
		-1.2E-10	5.5E-10	100	0.1E-10	7.1E-10	100	-1.2E-10	5.5E-10	100	-1.2E-10	6.6E-10	100
	$\log V_{I_1}$	2.2971											
		0.0017	0.0180	79	0.0017	0.0189	79	0.0010	0.0159	88	0.0009	0.0213	84
Random-Effects	$\log V_{2_1}$	0.7637											
		-0.0260	0.0896	84	-0.0247	0.0927	81	-0.0275	0.0755	87	-0.0157	0.1003	82
	$\log CL_1$	1.9225											
		-0.0023	0.0070	86	-0.0022	0.0070	81	-0.0026	0.0053	91	-0.0016	0.0078	88
	$\log CL_{I2_1}$	3.7673											
		-0.8378	0.3671	99	-0.8203	0.3224	98	-0.7410	0.4030	88	-0.5138	0.1952	99
	$\log V_{1_2}$	2.9228											
		-0.0018	0.0204	85	-0.0010	0.0199	85	-0.0004	0.0171	83	0.0017	0.0223	88
	$\log V_{2_2}$	2.0639											
		0.0140	0.0578	91	0.0136	0.0560	90	0.0115	0.0439	90	0.0271	0.0609	90
Random-Effects	$\log CL_2$	3.1210											
		0.0012	0.0071	97	0.0016	0.0070	97	0.0016	0.0062	89	0.0045	0.0079	88
	$\log CL_{I2_2}$	4.8801											
		-0.0728	0.2846	100	-0.0698	0.2636	100	0.0091	0.2176	96	0.2325	0.2622	87

	TRUE	RWM			DA			IMH			GMAP		
		Bias	SE	90% CP	Bias	SE	90% CP	Bias	SE	90% CP	Bias	SE	90% CP
	$\log VI_3$	3.2407	0.0187	87	-0.0010	0.0186	89	-0.0016	0.0177	88	0.0013	0.0166	89
	$\log V2_3$	2.7884	0.0335	91	0.0020	0.0325	92	0.0027	0.0303	92	-0.0034	0.0318	93
	$\log CL_3$	2.3392	0.0063	86	0.0000	0.0060	92	0.0000	0.0056	88	0.0010	0.0064	91
	$\log CLI2_3$	4.0591	0.0903	91	-0.0076	0.0917	90	-0.0034	0.0823	90	0.0133	0.0830	93
	$\log VI_4$	2.4127	0.0165	82	0.0014	0.0161	84	0.0008	0.0158	80	0.0033	0.0172	85
	$\log V2_4$	4.9140	0.0560	94	0.0032	0.0560	95	0.0037	0.0556	88	0.0118	0.0524	93
	$\log CL_4$	2.5675	0.0272	93	0.0005	0.0274	92	-0.0005	0.0287	88	-0.0032	0.0262	90
	$\log CLI2_4$	2.9374	0.0234	89	0.0015	0.0232	91	0.0018	0.0227	83	0.0053	0.0214	89
	$\log VI_5$	4.8253	0.0143	80	0.0068	0.0150	78	0.0039	0.0108	82	0.0085	0.0134	86
	$\log V2_5$	3.0177	0.1617	61	0.2811	0.1596	57	0.1349	0.0777	69	0.5054	0.3293	63
	$\log CL_5$	-1.6572	0.3490	16	-1.4854	0.3199	13	-1.0898	0.3619	36	-1.8602	0.3098	4
	$\log CLI2_5$	2.4360	0.4359	86	-0.0998	0.4399	75	-0.0370	0.2757	82	-0.2641	0.2922	87