

## ORIGINAL ARTICLE

# Bayesian adaptive Markov chain Monte Carlo estimation of genetic parameters

B Mathew<sup>1</sup>, AM Bauer<sup>1</sup>, P Koistinen<sup>2</sup>, TC Reetz<sup>1</sup>, J Léon<sup>1</sup> and MJ Sillanpää<sup>2,3,4</sup>

Accurate and fast estimation of genetic parameters that underlie quantitative traits using mixed linear models with additive and dominance effects is of great importance in both natural and breeding populations. Here, we propose a new fast adaptive Markov chain Monte Carlo (MCMC) sampling algorithm for the estimation of genetic parameters in the linear mixed model with several random effects. In the learning phase of our algorithm, we use the hybrid Gibbs sampler to learn the covariance structure of the variance components. In the second phase of the algorithm, we use this covariance structure to formulate an effective proposal distribution for a Metropolis-Hastings algorithm, which uses a likelihood function in which the random effects have been integrated out. Compared with the hybrid Gibbs sampler, the new algorithm had better mixing properties and was approximately twice as fast to run. Our new algorithm was able to detect different modes in the posterior distribution. In addition, the posterior mode estimates from the adaptive MCMC method were close to the REML (residual maximum likelihood) estimates. Moreover, our exponential prior for inverse variance components was vague and enabled the estimated mode of the posterior variance to be practically zero, which was in agreement with the support from the likelihood (in the case of no dominance). The method performance is illustrated using simulated data sets with replicates and field data in barley. *Heredity* (2012) **109**, 235–245; doi:10.1038/hdy.2012.35; published online 18 July 2012

**Keywords:** adaptive MCMC; identifiability problem; Bayesian analysis; Gibbs sampling; estimation of genetic parameters

## INTRODUCTION

Estimations of heritability (proportion of phenotypic variance attributable to genetic factors) and breeding values are of great interest because they are necessary to plan an efficient breeding program for the trait of interest. Therefore, considerable effort has been devoted to develop new statistical methods that estimate the breeding values and heritability in the linear mixed model context (Lynch and Walsh, 1998). However, when the random effects or variance components fitted to the model have multiple 'solutions' among their parameter spaces given the observed data, such parameters are said to be unidentifiable. Even if it is often well justified to include multiple random effects into the model (for example, Wall *et al.*, 2005), in practice, identifiability problems due to small data size and the familial structure of the data complicate the estimation of random genetic effects and their variances (Misztal, 1997; Waldmann *et al.*, 2008; Norris *et al.*, 2010). Evolutionary studies of natural populations can be more prone to this problem because data sets that are used for mixed model analyses of natural populations are typically much smaller than those that are used in plant or animal breeding (Kruuk, 2004; O'Hara *et al.*, 2008). In addition, the pedigree information for natural populations may be inaccurate and/or incomplete.

Bayesian methods can be applied to estimate genetic parameters (for example, Wang *et al.*, 1993; Blasco, 2001; Sorensen and Gianola, 2002; Hadfield, 2010), and such methods can, unlike REML (residual maximum likelihood), be helpful in diagnosing identifiability problems. In the Bayesian approach, one combines what is known about

the parameter (represented as a prior distribution) with the information that comes from the data to obtain the posterior distribution. This probability distribution represents the uncertainty about the parameter after the data has been taken into account (for example, Blasco, 2001). In the Bayesian approach, one is not limited to calculating only point estimates of variance components (as is traditionally done in a REML analysis) or only their confidence intervals, but one may explore any aspect of the remaining uncertainty, such as the uncertainty in the estimation of heritability. The standard computational approach is to use Markov chain Monte Carlo (MCMC) methods to draw samples from posterior distributions. The Gibbs sampler and the Metropolis-Hastings (M-H) algorithm are the two commonly used MCMC methods. The Gibbs sampler (Casella and George, 1992) uses draws from the fully conditional posteriors, and is a special case of M-H sampling (Chib and Greenberg, 1995).

Recently, Bauer *et al.* (2009) and Waldmann *et al.* (2008) applied Gibbs sampling to quantitative genetics research studies in plants, and the latter article developed a fast hybrid Gibbs sampler that accounted for additive and dominance variances in the mixed model. In a study carried out by De Boer and Hoeschele (1993), it was shown that the presence of inbreeding induces nonzero covariances between additive and dominance effects. However, Bauer *et al.* (2006), Oakey *et al.* (2006), Bauer and Léon (2008) predicted the breeding values (assuming no dominance) for self-pollinating crops by accounting for inbreeding among the lines. When nonzero covariance exists due to inbreeding, computational procedures for the estimation of the

<sup>1</sup>Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, Germany; <sup>2</sup>Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; <sup>3</sup>Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland and <sup>4</sup>Department of Mathematical Sciences and Department of Biology, University of Oulu, Oulu, Finland

Correspondence: Dr B Mathew, Institute of Crop Science and Resource Conservation, University of Bonn, Katzenburgweg 5, 53115 Bonn, Germany.  
E-mail: boby.mathew@hotmail.com

Received 31 January 2012; revised 15 May 2012; accepted 25 May 2012; published online 18 July 2012

variance components are further complicated. In our current study, we do not consider inbreeding together with dominance.

Since the 1980s, the use of MCMC methods has revolutionized the Bayesian analysis of complex statistical models (Robert and Casella, 2004). Even today, much focus is given to improving the efficiency and convergence of MCMC samplers. The efficiency of an MCMC algorithm critically depends on the transition kernel of the MC (Hastings, 1970; Roberts and Rosenthal, 2001), but the choice of an efficient kernel, which produces a rapidly mixing chain, is often difficult. Therefore, adaptive MCMC algorithms have been proposed that can use the previous history of the chain to learn the proposal distribution parameters. These adaptive algorithms are efficient for exploring the posterior distribution of the model using the data at hand. Recent developments (Haario *et al.*, 2001; Rosenthal, 2011) have increased the interest in applying adaptive MCMC methods in research studies. Although these methods would allow one to keep adapting the proposal distribution during the whole time the algorithm runs, we selected the proposal distribution on the basis of a learning phase, whose output we otherwise omitted from the posterior analysis. After the learning phase, we fixed the proposal distribution so that we were able to justify the use of analysis tools (such as the effective sample size (ESS)) that have been developed for analyzing the output from non-adaptive MCMC algorithms. As we stopped adapting after the learning phase, one could argue that our algorithm was not truly adaptive. However, because we manage to reduce analytically the number of unknowns down to three, the benefits of adaptation can be obtained in this simplified manner. Moreover, our method of selecting the proposal distribution is strongly motivated by the work on adaptive MCMC methods.

The hybrid Gibbs sampler is a combination of both a single-site Gibbs sampling algorithm (for example, Sorensen and Gianola, 2002) and a blocked Gibbs sampling algorithm (Garcia-Cortes and Sorensen, 1996). The convergence of the single-site Gibbs sampler can be slow due to posterior dependencies. In our new approach, the adaptive MCMC runs in two phases. First, in the learning phase, we ran the MCMC algorithm to obtain an estimate of the posterior covariance structure for log-transformed variance components. In the learning phase, we used a hybrid Gibbs sampler to sample random (additive genetic and dominance) effects. The dependencies among breeding values and dominance effects slow the convergence of the MCMC chain. So the effect of breeding values and dominance effects were marginalized away (integrated out analytically) before computing the posterior for the second phase, which we call the adapted phase. In the adapted phase, we utilized the estimated covariance structure from the learning phase, to generate multivariate correlated proposals for (log-transformed) variance components in a random walk M-H algorithm. The acceptance of these proposals was checked jointly as a single block. Block updates of the variance components after marginalizing the random effects helped the MC to converge to its equilibrium distribution reasonably fast.

In this study, we developed a fast adaptive MCMC algorithm for the estimation of additive and dominance variance in the traditional infinitesimal model without inbreeding. We compared the efficiency of the two estimation algorithms (conventional and adaptive MCMC samplers) with simulated and real data sets. In this assessment, it was important to ensure that differences in the analysis results were not due to reasons other than real differences in the sampling efficiencies between the two algorithms. Weak identifiability of variance components would make a fair comparison difficult. To alleviate this problem, we decided to simulate moderately high heritability. At the same time, we wanted to keep the number of individuals in

our sample relatively small to maintain reasonable computation times in the examples. However, results from these example analyses of data with high heritability and small sample size arguably correspond to the results obtained from more realistic data with smaller heritability and a larger sample size.

## MODELS AND METHODS

### Model 1

We consider the mixed linear model (Henderson, 1985a,b):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{d} + \mathbf{e}, \quad (1)$$

where  $\mathbf{y}$  is an  $n \times 1$  vector of phenotypic observations,  $\boldsymbol{\beta}$  is a  $k \times 1$  vector of fixed (environmental) effects,  $\mathbf{a}$  is a  $q \times 1$  vector of random additive genetic effects,  $\mathbf{d}$  is a  $q \times 1$  vector of random dominance genetic effects and  $\mathbf{e}$  is a  $n \times 1$  vector of error terms, which are independently normally distributed with mean zero and variance  $\sigma_e^2$ . Moreover,  $\mathbf{X}$ ,  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  are known incidence matrices, where  $\mathbf{X}$  associates  $\boldsymbol{\beta}$  to the phenotypic observations  $\mathbf{y}$ . For the simulated data sets,  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  associate genetic effects,  $\mathbf{a}$  and  $\mathbf{d}$ , respectively, to the observation vector  $\mathbf{y}$ . For the field data,  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  associate random additive genetic effects  $\mathbf{a}$  and  $\mathbf{G} \times \mathbf{E}$  (genotype-by-environment interaction) to  $\mathbf{y}$ . The numerator relationship matrix  $\mathbf{A}$ , which describes additive genetic relationships between lines, can be calculated from pedigree information using well-known methods (for example, see p 763 in Lynch and Walsh, 1998) or alternatively, its inverse  $\mathbf{A}^{-1}$  can be calculated directly (Henderson, 1976; Quaas, 1976). Similarly, the dominance matrix  $\mathbf{D}$ , which describes dominance genetic relationships among lines, can be calculated from pedigree information (for example, see p 768 in Lynch and Walsh, 1998; Waldmann *et al.*, 2008) or alternatively; its inverse can be calculated directly (Hoeschele and VanRaden, 1991). We want to emphasize that for studies of genotype-by-environment interactions, the methodology presented here works simply by considering the  $\mathbf{G} \times \mathbf{E}$  covariance structure in place of  $\mathbf{D}$ . For details of the  $\mathbf{G} \times \mathbf{E}$  covariance structure, see Bauer *et al.* (2009).

In the following sections, we present two different hierarchical models; the former to be used in the learning phase and the latter in the adapted phase of the estimation algorithm. If all the priors are chosen to be the same, then these two hierarchical models are identical, except that most parameters have been integrated out analytically from the latter.

### Hierarchical model 1

Let the precision parameters  $\psi_a$ ,  $\psi_d$  and  $\psi_e$  be the inverses of the variances  $\sigma_a^2$ ,  $\sigma_d^2$  and  $\sigma_e^2$ , respectively. Then using model 1, the phenotypic observation for a given trait is modeled as a linear combination of explanatory variables. For given  $\boldsymbol{\beta}$ ,  $\mathbf{a}$ ,  $\mathbf{d}$  and  $\psi_e$ , vector  $\mathbf{y}$  follows a multivariate normal distribution

$$\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{a}, \mathbf{d}, \psi_e \sim \text{MVN}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{d}, \mathbf{I}/\psi_e), \quad (2)$$

where  $1/\psi_e$  is the residual variance of the model. Let  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{a}, \mathbf{d})$  be the unknown location parameters and  $\boldsymbol{\psi} = (\psi_a, \psi_d, \psi_e)$  be the precision parameters. By Bayes theorem, the joint posterior density of unknown parameters is proportional to

$$p(\boldsymbol{\theta}, \boldsymbol{\psi} \mid \mathbf{y}) \propto p(\boldsymbol{\psi})p(\boldsymbol{\theta} \mid \boldsymbol{\psi})p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\psi}), \quad (3)$$

where  $p(\boldsymbol{\psi}) = p(\psi_a)p(\psi_d)p(\psi_e)$  and  $p(\boldsymbol{\theta} \mid \boldsymbol{\psi}) = p(\boldsymbol{\beta})p(\mathbf{a} \mid \psi_a)p(\mathbf{d} \mid \psi_d)$  are the prior distributions and  $p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\psi})$  is the likelihood from Equation (2). For the Bayesian analysis, one must assign a prior distribution for the unknown model parameters. Therefore,  $\boldsymbol{\beta}$  was assigned an improper uniform prior distribution.

$$p(\boldsymbol{\beta}) \propto \text{constant}. \quad (4)$$

Conditionally on the precision parameters, the genetic effects were assigned multivariate normal prior distributions with a zero mean vector  $\mathbf{0}$  (of size  $q$ ),

$$\mathbf{a} \mid \psi_a \sim \text{MVN}(\mathbf{0}, \mathbf{A}/\psi_a), \quad \mathbf{d} \mid \psi_d \sim \text{MVN}(\mathbf{0}, \mathbf{D}/\psi_d) \quad (5)$$

Before assigning a prior distribution for the precision parameters, we standardized the phenotypic observation vector  $\mathbf{y}$  to use the same prior for different data sets (which may originally have had very different phenotypic scales). After the standardization, the precision parameters  $\psi_a$ ,  $\psi_d$  and  $\psi_e$  were assumed to follow a gamma prior distribution with parameters

$k_i$  and  $\lambda_i$  and mean  $k_i/\lambda_i$ ,

$$\psi_i \sim \text{Gamma}(k_i, \lambda_i), \quad i = a, d, e. \quad (6)$$

We chose  $k_i=1$  and  $\lambda_i=0.001$  (that is, the exponential distribution with mean  $1/\lambda_i$ ) to obtain flat priors. This choice allows the variance components to be shrunk to very nearly zero, if this is warranted by the data. This follows because the prior Equation (6) implies an inverse gamma prior with parameters  $(k_i, \lambda_i)$  for the variance component  $\sigma_i^2$ . The inverse gamma density increases from a value of zero to its maximum at the mode  $\lambda_i/(k_i+1)$  and then decays slowly. Shrinkage-type priors have been used before, for example, in variable selection (O'Hara and Sillanpää, 2009) and in haplotype estimation (Gasbarra *et al.*, 2011), as well as in the penalized likelihood estimation of genetic covariance matrices (Meyer and Kirkpatrick, 2010).

### Hierarchical model 2

In the adapted phase of the algorithm, we use a model in which all the unknown location parameters  $\theta$  are integrated out from model (1). The joint posterior density of parameters  $\psi$  is

$$p(\psi | \mathbf{y}) \propto p(\psi)p(\mathbf{y} | \psi). \quad (7)$$

To mimic the improper uniform prior Equation (4), the fixed effects  $\beta$  were assigned a normal prior distribution with a zero mean vector  $\mathbf{0}$  and a large covariance matrix  $\mathbf{B}\sigma_\beta^2$ , where  $\sigma_\beta^2=10^6$ ,

$$\beta \sim \text{MVN}(\mathbf{0}, \mathbf{B}\sigma_\beta^2).$$

Here,  $\mathbf{B}$  is the unscaled prior covariance matrix between fixed effects. The genetic effects  $\mathbf{a}$  and  $\mathbf{d}$  were assigned the multivariate normal priors Equation (5), and the variance components, the gamma priors Equation (6). After these choices, it is a simple matter to integrate out the location parameters from the model (cf. pp 313–314 in Sorensen and Gianola, 2002), namely

$$\mathbf{y} | \psi \sim \text{MVN}(\mathbf{0}, \Sigma), \quad (8)$$

where  $\Sigma = \mathbf{X}\mathbf{B}\mathbf{X}'\sigma_\beta^2 + \mathbf{Z}_1\mathbf{A}\mathbf{Z}_1'/\psi_a + \mathbf{Z}_2\mathbf{D}\mathbf{Z}_2'/\psi_d + \mathbf{I}/\psi_e$ .

### Estimation in the learning phase

To implement the Gibbs sampler, one needs the fully conditional posterior distributions of all unknown parameters ( $\theta$  and  $\psi$ ) of hierarchical model 1. These can be found, for example, from Waldmann *et al.* (2008). To update  $\theta$ , samples can be drawn either element-wise or block-wise from the fully conditional posterior distribution

$$\theta | \psi, \mathbf{y} \sim \text{MVN}(\hat{\theta}, \mathbf{C}^{-1}/\psi_e), \quad (9)$$

where  $\hat{\theta}$  is the solution to the linear system  $\mathbf{C}\theta = \mathbf{W}'\mathbf{y}$ . Here,

$$\mathbf{C} = \mathbf{W}'\mathbf{W} + \mathbf{V}, \quad \mathbf{W} = [\mathbf{X}, \mathbf{Z}_1, \mathbf{Z}_2], \quad \mathbf{V} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{A}^{-1}\alpha_a & 0 \\ 0 & 0 & \mathbf{D}^{-1}\alpha_d \end{bmatrix} \quad (10)$$

with  $\alpha_a = \psi_a/\psi_e$ ,  $\alpha_d = \psi_d/\psi_e$ . The precision parameters are sampled from their fully conditional posterior distributions,

$$\psi_i | \theta, \mathbf{y} \sim \text{Gamma}(k_i^*, \lambda_i^*), \quad i = a, d, e$$

where  $k_a^* = k_a + q/2$ ,  $\lambda_a^* = \lambda_a + (\mathbf{a}'\mathbf{A}^{-1}\mathbf{a})/2$ ,  $k_d^* = k_d + q/2$ ,  $\lambda_d^* = \lambda_d + (\mathbf{d}'\mathbf{D}^{-1}\mathbf{d})/2$ ,  $k_e^* = k_e + n/2$ , and  $\lambda_e^* = \lambda_e + 1/2\|\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}_1\mathbf{a} - \mathbf{Z}_2\mathbf{d}\|^2$ . During the learning phase of the algorithm, we use a hybrid Gibbs sampler with a block update every 50th iteration to sample the random additive and dominance effects. See the appendix for details of the sampling algorithm.

### Estimation in the adapted phase

We use the history of the chain during the learning phase to form the proposal distribution for the parameters of hierarchical model 2. In the second, adapted phase of the algorithm, we use an M–H algorithm to update the log-variance components block-wise using putative samples generated from the learned proposal distribution.

Our M–H algorithm uses random-walk proposals; the proposed parameter vector is generated by adding to the current parameter vector an increment from a multivariate normal distribution with a zero mean and covariance matrix  $\mathbf{S}_p$ . We base our selection of the proposal covariance matrix on the theoretical results of Roberts *et al.* (1997) and Roberts and Rosenthal (2001).

These authors showed that if the posterior distribution is approximately multivariate normal with covariance matrix  $\mathbf{S}$ , then the optimal choice for the proposal covariance matrix  $\mathbf{S}_p$  is approximately  $(2.38)^2/d\mathbf{S}$ , where  $d$  is the number of unknown parameters in the posterior distribution. To improve our ability to use this result, our algorithm works on the logarithmic scale, that is, we use the vector  $\tau = (\tau_a, \tau_d, \tau_e)$  as the new parameter vector, where the  $\tau$ 's are the logarithms of the variance components,  $\tau_i = \log(\sigma_i^2) = -\log(\psi_i)$ ,  $i = a, d, e$ . This reparameterization eliminates the positivity constraints that are present for the variance components or their inverses. At the same time, it makes the posterior distribution resemble a multivariate normal distribution more closely. As the posterior covariance matrix  $\mathbf{S}$  of the vector  $\tau$  is unknown, we estimate it with the sample covariance matrix  $\hat{\mathbf{S}}$ , which is calculated from the log-transformed variance components that are simulated during the learning phase.

After the proposed parameter vector  $\tau^*$  has been generated by adding a noise vector to the current parameter vector  $\tau$ , the proposed  $\tau^*$  is either accepted or rejected as the new state of the MC based on the value of the M–H acceptance ratio  $r$ , which is now given by

$$r = \frac{p(\tau^*)p(\mathbf{y} | \tau^*)}{p(\tau)p(\mathbf{y} | \tau)} \quad (11)$$

Here, the likelihood ratio can be evaluated based on Equation (8), after the log-transformed variance components  $\tau = (\tau_a, \tau_d, \tau_e)$  and  $\tau^* = (\tau_a^*, \tau_d^*, \tau_e^*)$  have been transformed to precision parameters, using the formulas

$$\psi_i = e^{-\tau_i}, \quad \psi_i^* = e^{-\tau_i^*}, \quad i = a, d, e.$$

For  $\tau$ , the likelihood is

$$p(\mathbf{y} | \tau) = (2\pi)^{-n/2} \frac{1}{\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}\mathbf{y}'\Sigma^{-1}\mathbf{y}\right\} \quad (12)$$

where  $\Sigma$  is the covariance matrix of  $\mathbf{y}$ , conditionally on the current values of the parameters,

$$\Sigma = \mathbf{X}\mathbf{B}\mathbf{X}'\sigma_\beta^2 + \mathbf{Z}_1\mathbf{A}\mathbf{Z}_1'/\psi_a + \mathbf{Z}_2\mathbf{D}\mathbf{Z}_2'/\psi_d + \mathbf{I}/\psi_e.$$

For  $\tau^*$ , the likelihood  $p(\mathbf{y} | \tau^*)$  is obtained from a similar formula in which  $\Sigma$  is replaced by the covariance matrix of  $\mathbf{y}$  conditional on the proposed values of the parameters,

$$\Sigma^* = \mathbf{X}\mathbf{B}\mathbf{X}'\sigma_\beta^2 + \mathbf{Z}_1\mathbf{A}\mathbf{Z}_1'/\psi_a^* + \mathbf{Z}_2\mathbf{D}\mathbf{Z}_2'/\psi_d^* + \mathbf{I}/\psi_e^*.$$

See the appendix for the details of how the likelihood ratio is calculated. To evaluate the prior ratio  $p(\tau^*)/p(\tau)$  in Equation (11), we must take into account that we have formulated the prior for the vector of precision parameters  $\psi$ . Using the change-of-variables formula for probability densities, the prior ratio can be calculated as

$$\frac{p(\tau^*)}{p(\tau)} = \frac{p(\psi^*) | J^* |}{p(\psi) | J |}. \quad (13)$$

Here,  $p(\psi) = p(\psi_a|k_a, \lambda_a)p(\psi_d|k_d, \lambda_d)p(\psi_e|k_e, \lambda_e)$  is the product of the three gamma densities, Equation (6); and similarly,  $p(\psi^*)$  is the product of the same gamma densities evaluated at the proposed precision parameters. Furthermore,  $J = -\exp(-\tau_a - \tau_d - \tau_e)$  is the Jacobian (determinant) arising from expressing  $\psi$  in terms of  $\tau$ , and  $J^* = -\exp(-\tau_a^* - \tau_d^* - \tau_e^*)$  is the Jacobian from expressing  $\psi^*$  in terms of  $\tau^*$ . In the actual M–H algorithm, we first calculate the logarithm of the M–H ratio  $r$ , and then we calculate the logarithm of the ratio of the absolute Jacobians,

$$\log \frac{|J^*|}{|J|} = -(\tau_a^* - \tau_a + \tau_d^* - \tau_d + \tau_e^* - \tau_e). \quad (14)$$

The sampling algorithm during the adapted phase is as follows. First, we estimate the posterior covariance matrix  $\mathbf{S}$  of the log-transformed variance components from the output of the learning phase and calculate the proposal covariance matrix as  $\mathbf{S}_p = (2.38)^2\hat{\mathbf{S}}/d$ . Then, we iterate the following steps:

1. Let  $\tau$  be the current values in logarithmic scale. Generate new values  $\tau^* = \tau + \mathbf{w}$ , where  $\mathbf{w}$  is simulated from  $\text{MVN}(\mathbf{0}, \mathbf{S}_p)$ . Transform  $\tau$  and  $\tau^*$  to precision parameter vectors  $\psi$  and  $\psi^*$ .
2. Calculate the logarithm of the M–H acceptance ratio  $\log(r)$  using Equations (11)–(14).
3. Accept the proposed value  $\tau^*$ , if a random number drawn from the uniform distribution over  $[0,1]$  is less than  $r$ . If the proposal is accepted, then the

proposed parameter vector is taken as the current vector  $\tau = \tau^*$ , otherwise, the current value is retained.

As the breeding values and dominance effects have been integrated out from the likelihood, this sampling algorithm reduces the problems of the Gibbs sampler that arise due to posterior dependences between the random effects and variance components.

The whole adaptive algorithm, consisting of the learning phase and adapted phase, is described more fully in the appendix. It has been implemented in the Matlab (2007) environment, where most of our analyses have also been performed.

### Effective sample size

ESS (Geyer, 1992; Waagepetersen *et al.*, 2008) is a popular diagnostic tool for MCMC methods. ESS determines the approximate number of independent samples that would provide the same estimation accuracy as the dependent MCMC samples. The ESS values were calculated with the R package coda (Plummer *et al.*, 2006).

## EXAMPLE ANALYSES

### Simulated data

We developed a C program that simulates ‘virtual’ populations for the variance component estimation. Owing to the identifiability problems faced during the analysis, we decided to consider two different data sets, one of which resulted in an unimodal posterior distribution of dominance variance and the other in a bimodal posterior. To develop the bimodal data set, we considered a base population of 50 unrelated lines, wherein each of the 25 females were mated with 25 males, and each crossing resulted in five offspring (in total, 3175 individuals, including the base population). For the unimodal data set, we considered a base population of 40 lines, 20 females and 20 males, and each crossing resulted in nine offspring (in total 3640 individuals, including the base population).

Additive genetic relationship matrix  $A$  and dominance relationship matrix  $D$  were calculated from the pedigree information as described in the model section. To simulate a quantitative trait  $y$ , we generated three factors; the additive effect  $a$ , the dominance effect  $d$  and noise  $e$ , and the vector of phenotypic observations was calculated as their sum,

$$y = a + d + e.$$

Here, vectors  $a$ ,  $d$  and  $e$  were drawn from  $MVN(\mathbf{0}, A\sigma_a^2)$ ,  $MVN(\mathbf{0}, D\sigma_d^2)$  and  $MVN(\mathbf{0}, I\sigma_e^2)$ , respectively. We used the Cholesky decomposition of the covariance matrices  $A\sigma_a^2$  and  $D\sigma_d^2$  to draw samples from these distributions. Hence, the random genetic effects  $a$  and  $d$  were calculated as  $a = Pz_a$  and  $d = Tz_d$ , where  $z_i \sim MVN(\mathbf{0}, I)$  and  $P$  and  $T$  are the Cholesky factors  $PP^T = A\sigma_a^2$  and  $TT^T = D\sigma_d^2$ . To validate our estimation methods, we generated the two data sets using a heritability of 0.31 ( $\sigma_a^2 = 800$ ,  $\sigma_d^2 = 600$ ,  $\sigma_e^2 = 3025$ ). Using the same set of parameters, we generated 10 simulation replicates of the unimodal data set (an existing unimodal data set and 9 new replicates) by sampling new residuals  $e$  from  $MVN(\mathbf{0}, I\sigma_e^2)$ , each time, keeping the original pedigree. However, one realization was removed and simulated again because it resulted in a bimodal posterior.

### QTLMAS XII workshop data

This is the simulated data set obtained from the QTLMAS XII workshop web page, <http://www.computationalgenetics.se/QTLMAS08/QTLMAS/DATA.html>.

The data set was generated following an animal breeding protocol, consisting of 5865 individuals from seven generations. For the first four generations (a total of 4665 individuals), both pedigree and

phenotype information are available, and we considered this subset of the data for our analysis. The additive genetic relationship matrix  $A$  and dominance relationship matrix  $D$  were calculated from the pedigree information.

### Field data

Real data from 82 spring barley (*Hordeum vulgare* L.) lines originating from the German North Rhine Westphalia (Bauer *et al.*, 2006, 2009; Bauer and Léon, 2008) core collection were analyzed. These lines were cultivated in a randomized complete-block design with three replications over 3 different years (2001, 2002 and 2003) at the Research Station ‘Dikopshof’ of University of the Bonn, Germany. For the real data, a few replications were missing, and the missing values were imputed by the average value of non-missing replications for the corresponding year. There are a number of alternative ways of dealing with missing data. However, as the number of missing values was very low, we expected that method used would not make a significant difference. Pedigree information was available for all the lines, and the phenotypic observations of the trait ‘thousand kernel mass’ were measured for all the lines. For the field data, we considered genotype-by-environment interaction instead of the dominance relationship in the linear mixed model (1) and accounted for the inbreeding among lines. Following Bauer *et al.* (2009), two different covariance structures were applied to model the genotype-by-environment interaction. In the first approach, called Bayes\_ID, the genotype-by-environment interaction was assumed to be independently and identically normally distributed. In the second approach (Bayes\_A<sup>ext</sup>), an extended relationship matrix  $A^{\text{ext}} = A \otimes I$  (here ‘ $\otimes$ ’ is the Kronecker product of two matrices) was used to model the genotype-by-environment interaction. Moreover, the fixed year effect was considered in the  $X$  matrix, along with the overall mean for the analysis. Note that in model (1), the simplifying assumption of independent errors with the constant variance was again made (cf. Burgueno *et al.*, 2012; Piepho *et al.*, 2012).

### Analyses and results, simulated data

To validate our new algorithm, we analyzed the two simulated data sets; the unimodal data set with 3640 individuals and the bimodal data set with 3175 individuals. The estimates for variance components based on all the individuals of the two simulated data sets and 10 simulation replicates were calculated using our adaptive MCMC method and the REML method (Tables 1 and 2). The REML estimates of the variance components were calculated using ASReml software (Gilmour *et al.*, 2006). True values, given in Table 1, are the values used in the simulations. The implemented MCMC had a total chain length of 50 000, consisting of a burn-in period of 2000 iterations, a learning phase from iteration number 2000 to 5000, and finished with the adapted phase from iterations 5000 to 50 000. Acceptance ratios for the bimodal and unimodal data sets were 28% and 26%, respectively. The point estimates, mean and median of the posterior distribution of the variance components were calculated from the MCMC samples. To calculate the mode of the posterior distribution, a kernel smoothing approach following Hoti *et al.* (2002) was used.

In Table 2, we demonstrate that Bayesian point estimates averaged over 10 replicates for variance components were always more close to the true simulated values than the averaged REML estimates. The same was often true for the extreme values (minimum and maximum) of the estimates over 10 replicates. This suggests that the Bayesian adaptive MCMC method can give variance component estimates that are competitive with the REML estimates.

A properly implemented MCMC sampler should be able to cover all the areas supported by the target distribution, but the existence of multiple modes makes this difficult (Geyer and Thompson, 1995). A conventional MCMC algorithm may fail to jump between the different modes and therefore may visit only a single mode. Although running the chain for a very long time may remedy this problem, this is computationally highly demanding. In our approach, the posterior covariance structure estimated from the learning phase helps the sampler to move freely between the different modes of the target distribution. Our MCMC algorithm was able to detect two different modes in the posterior for dominance and residual variances in the bimodal data set, whereas REML always returns a single mode (and the identified mode may depend on its starting values). Table 3 summarizes the rough estimates for two different modes (estimated using the kernel smoothing approach; see Hoti *et al.*, 2002.) The posterior mode 1 values of dominance and residual variance are close to the true simulated values and are somewhat better than the corresponding REML estimates. The posterior mode 2 values of dominance and residual variance are poor, and their existence indicates that there is an identifiability problem. From Figures 1

and 2, it is clear that the adaptive MCMC algorithm is able to move between the different modes of the posterior of the bimodal data set. It can be seen from Figures 1 and 3 that our new adaptive MCMC algorithm was able to detect different modes in the distribution with a relatively low number of iterations, whereas the conventional MCMC method had problems visiting the different modes. To visualize the different modes in the posterior, a histogram with hexagonal bins was drawn for the log-transformed dominance and error variance components (Figure 3) with the aid of the hexbin package of R.

**Table 3 The two different modes of the variance components for the simulated bimodal data set**

	Mode 1	Mode 2	REML	True
$\sigma_a^2$	675.40	2.67	716.00	600
$\sigma_e^2$	3020.20	3505.2	2882.60	3025

Abbreviation: REML, residual maximum likelihood. The posterior mode estimates are obtained from the adapted phase of the Markov chain Monte Carlo analysis. REML estimates and true simulated values are also shown.

**Table 1 The estimates of variance components and broad-sense heritabilities for the learning and adapted phases from the MCMC analyses of the two simulated data sets**

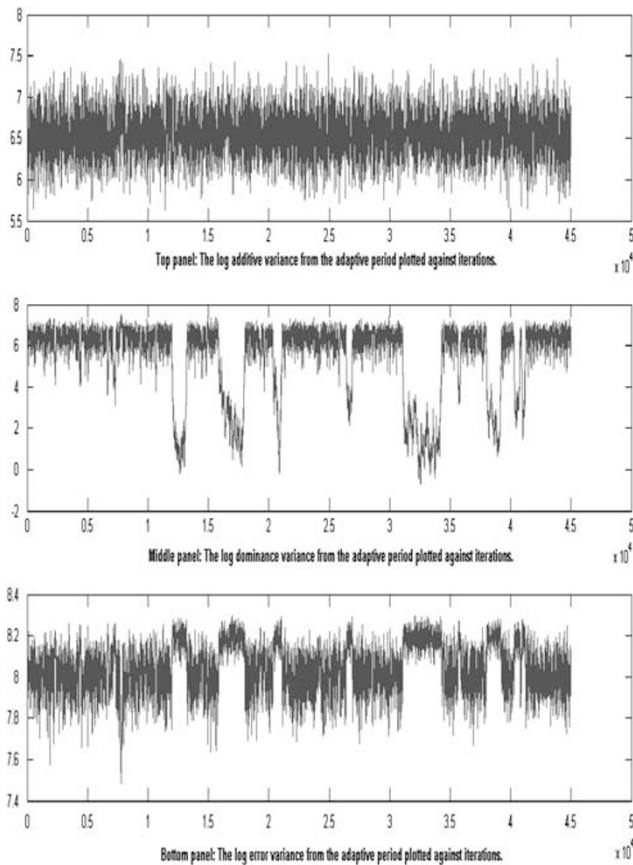
	Learning phase			Adapted phase			REML	True
	Mean	Median	Mode	Mean	Median	Mode		
<i>Bimodal data</i>								
$\sigma_a^2$	672.57	607.73	573.67	721.49	695.87	679.37	752.99	800
$\sigma_d^2$	545.93	510.21	453.20	493.10	522.30	675.40	716.00	600
$\sigma_e^2$	3107.70	3143.70	3013.70	3132.10	3105.00	3020.20	2882.60	3025
$h^2$	0.28	0.26	0.25	0.27	0.28	0.30	0.33	0.31
<i>Unimodal data</i>								
$\sigma_a^2$	873.36	820.90	879.80	751.20	744.53	779.80	781.28	800
$\sigma_d^2$	619.36	642.23	658.70	591.50	585.77	579.80	571.68	600
$\sigma_e^2$	2845.40	2865.00	2894.70	2965.00	2971.90	2960.90	2928.79	3025
$h^2$	0.34	0.33	0.34	0.33	0.31	0.31	0.31	0.31

Abbreviations: MCMC, Markov chain Monte Carlo; REML, residual maximum likelihood. REML estimates and true simulated values are also shown. The names 'unimodal data' and 'bimodal data' are based on the characteristics that these data sets exhibited during the MCMC analysis.

**Table 2 Estimated variance components of two Markov chain Monte Carlo algorithms and REML based on 10 simulation replicates**

	$\sigma_a^2$			$\sigma_d^2$			$\sigma_e^2$		
	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
<i>Learning phase</i>									
Mean	778.04	659.95	873.01	593.67	512.86	698.77	3031.13	2845.09	3184.00
Median	767.32	627.37	831.31	584.36	503.76	690.93	3057.02	2894.05	3191.20
Mode	728.63	518.87	879.81	568.09	515.13	658.08	3008.63	2737.08	3282.07
<i>Adapted phase</i>									
Mean	774.70	732.99	839.36	577.07	545.11	620.89	3065.10	2901.00	3184.05
Median	764.48	718.31	827.60	574.23	529.86	615.08	3069.23	2907.08	3191.20
Mode	779.02	747.16	838.08	580.98	531.31	634.40	3103.76	2928.00	3284.50
REML estimates	731.27	588.17	881.47	555.99	527.02	629.06	3064.94	2928.00	3174.91
True values	800.00			600.00			3025.00		

Abbreviation: REML, residual maximum likelihood. True simulated values are also shown. Average, minimum and maximum values of the posterior (mean, median and mode) and REML estimates are calculated for variance components from the analyses of 10 simulation replicates.



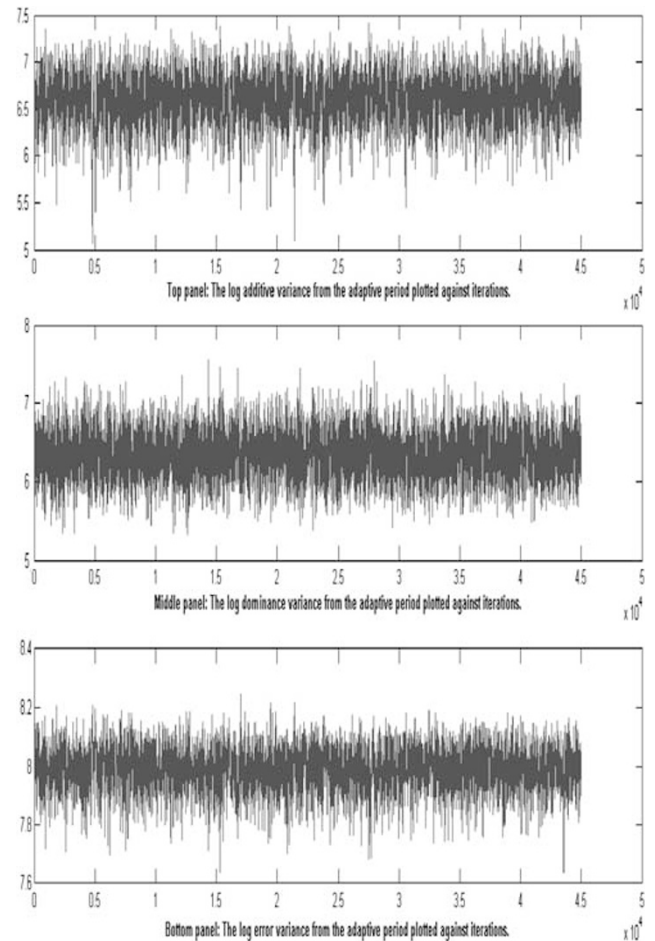
**Figure 1** The logarithm of the variance components for the bimodal data set plotted against the MCMC iteration number. The trace plots show 45 000 iterations from the adapted phase.

#### Analyses and results, QTLMAS XII workshop data

We considered a subset of 4665 individuals (the first four generations) from the QTLMAS XII workshop data for the analysis. The pedigree information for the first four generations was available, and hence **A** and **D** matrices were calculated from the pedigree. The heritability of the QTLMAS XII workshop data was around 0.30 with the zero dominance effect. For further details of the data, see Lund *et al.* (2009). Our main motivation in analyzing the QTLMAS data set was to test how our method behaves in the absence of a dominance effect. The variance components were estimated using the adaptive MCMC and the REML methods (Table 4). The implemented MCMC had a total chain length of 50 000 with a burn-in period of 2000 iterations, a learning phase from iteration number 2000 to 5000, and the adapted phase from iterations 5000 to 50 000. The acceptance ratio for the data set was 35%. The point estimates were calculated as before. In our analysis, we obtained a heritability of approximately 0.30. Hallander *et al.* (2010) used a different prior and obtained a heritability point estimate of 0.34 from a smaller subset of data, using a Bayesian model containing additive polygenic effects only. They used uniform distributions as non-informative choice of priors to the standard deviations.

#### Analyses and results, field data

The trait 'thousand kernel mass' for 82 spring barley lines from 3 different years with three replications were considered for analysis with our adaptive MCMC method and the REML method using ASReML software (by assuming the same covariance structure for the



**Figure 2** The logarithm of the variance components for the unimodal data set plotted against the MCMC iteration number. The trace plots show 45 000 iterations from the adapted phase.

genotype-by-environment interaction as in Bayes\_ID). The implemented MCMC method had a total chain length of 50 000, consisting of a burn-in period of 2000 iterations, a learning phase from iteration number 2000 to 5000, and finishing with the adapted phase from iterations 5000 to 50 000. For analysis, each year was considered as a different location. Therefore, to account for the number of locations, heritability was calculated using the formula (Hanson, 1963):

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \left(\sigma_{g \times e}^2 / j\right) + \left(\sigma_e^2 / j * k\right)}$$

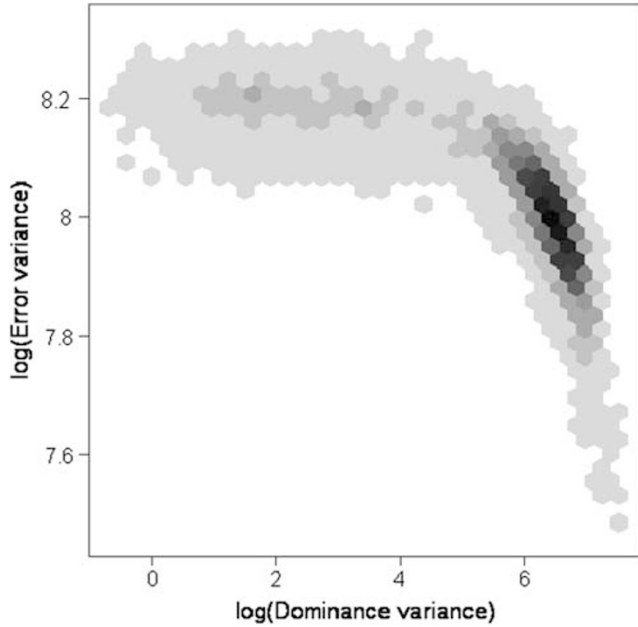
where  $\sigma_a^2$  is the additive genetic variance,  $\sigma_{g \times e}^2$  is the variance due to genotype-by-environment interactions,  $\sigma_e^2$  is the error variance,  $j$  is the number of years and  $k = 3$  is the number of replications. We also calculated the point estimates and 95% highest posterior density intervals for the posterior distribution from the adapted phase of the algorithm using the Bayes\_ID and Bayes\_A<sup>ext</sup> methods (Table 5). Bauer *et al.* (2009) considered data from two different years (2002 and 2003) for the analysis; and in our current study, we considered data from 3 different years (2001, 2002 and 2003). Hence, our analysis provided higher heritabilities than in Bauer *et al.* (2009). Both studies showed that the Bayes\_A<sup>ext</sup> estimates were closer to the REML estimates. Moreover, results from both studies indicated that it is important to consider the relationship information between lines in a Bayesian model when estimating the genotype-by-environment interactions.

**Effective sample size**

ESS values were calculated for different data sets from both the learning phase and the adapted phase. ESS is a measure of the mixing properties of the MCMC chain. High values of ESS imply that the

autocorrelation is low and are an indication that the mixing of the MCMC chain is good.

Adequate mixing of the MCMC sampler over different parts of the parameter space is essential for the convergence of MCMC algorithms, but conventional MCMC algorithms may suffer from slow mixing. From the trace plots (Figure 4), for the learning phase and adapted phase, it seems clear that the adapted MCMC was mixing well compared with the general hybrid Gibbs sampler (used in the learning phase). Thus, the adaptation significantly improved the mixing property of the algorithm, by learning an appropriate covariance structure for the proposal distribution. This visual impression is confirmed by Table 6, which summarizes the ESS for the unimodal, bimodal and QTLMAS data sets. In addition, the ESS values and the ESS ratios were calculated for the 10 simulation replicates (see Table 7). The ESS ratios were calculated by dividing the ESS from the adapted phase by the ESS from the learning phase. The ESS ratio gives how many more learning phase iterations one needs to run to obtain the same estimation accuracy as with the adapted phase. Based on the ESS values, one needs to run the standard MCMC chain (learning phase) at least six times longer to obtain comparable estimates as with from the adapted phase MCMC sampler. To calculate the ESS, an MCMC chain with a length of 3000 from the learning phase and a chain of the same length from the beginning of the adapted phase were considered after a burn-in period of 2000 iterations. For the field data, the ESS was calculated using Bayes\_ID

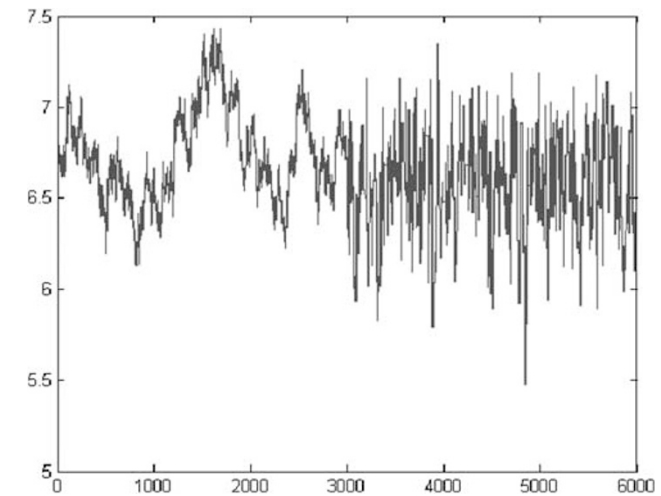


**Figure 3** Histogram of the log-transformed dominance and error variance components using hexagonal bins.

**Table 4** The estimates of the variance components and broad-sense heritabilities for the learning and adapted phases from the Markov chain Monte Carlo analysis of the QTLMAS XII data set

	Learning phase			Adapted phase			REML	True
	Mean	Median	Mode	Mean	Median	Mode		
$\sigma_a^2$	1.33	1.32	1.10	1.34	1.33	1.31	1.35	1.36
$\sigma_d^2$	0.09	0.10	0.10	0.01	0.00	0.00	0.00	0.00
$\sigma_e^2$	3.06	3.06	2.84	3.13	3.13	3.15	3.12	3.20
$h^2$	0.46	0.46	0.29	0.30	0.29	0.29	0.30	0.30

Abbreviation: REML, residual maximum likelihood. REML estimates and true simulated values for the entire pedigree are also shown. The true value for the additive variance was calculated as the variance of true genomic breeding values omitting relationships between individuals, and the residual variance was calculated accordingly to obtain a heritability of approximately 0.30.



**Figure 4** Trace plot of the log-transformed additive variance component for the unimodal simulated data set. The first 3000 samples are taken from the learning phase and the remaining samples are from the adapted phase. A full color version of this figure is available at the *Heredity* journal online.

**Table 5** The estimates of variance components, heritabilities and the 95% HPD intervals for the field data from the adapted phases of the algorithm using Bayes\_ID and Bayes\_A<sup>ext</sup> covariances

	Bayes_ID					Bayes_A <sup>ext</sup>					REML
	Mean	Median	Mode	2.5	97.5	Mean	Median	Mode	2.5	97.5	
$\sigma_a^2$	9.27	9.13	9.08	5.15	13.54	9.10	9.15	9.05	5.20	13.69	9.21
$\sigma_h^2$	2.45	2.49	2.60	0.00	4.61	2.98	2.70	2.84	0.00	4.77	3.18
$\sigma_e^2$	17.67	17.58	17.50	15.33	20.30	17.18	17.23	17.29	15.36	20.25	17.08
$h^2$	0.76	0.76	0.76			0.76	0.76	0.76			0.75

Abbreviation: REML, residual maximum likelihood. REML estimates are also shown.

**Table 6 ESS for 3000 iterations of the two Markov chain Monte Carlo algorithms with the unimodal, bimodal, QTLMAS and field data sets**

	Learning phase			Adapted phase		
	$\sigma_a^2$	$\sigma_d^2$	$\sigma_e^2$	$\sigma_a^2$	$\sigma_d^2$	$\sigma_e^2$
<i>Unimodal data set</i>						
ESS	10.36	8.76	29.56	176.71	318.15	191.81
<i>Bimodal data set</i>						
ESS	12.82	12.10	11.27	240.56	159.68	176.44
<i>QTLMAS data set</i>						
ESS	41.24	3.58	116.98	242.90	93.37	204.64
	Learning phase			Adapted phase		
	$\sigma_a^2$	$\sigma_{g \times e}^2$	$\sigma_e^2$	$\sigma_a^2$	$\sigma_{g \times e}^2$	$\sigma_e^2$
<i>Field data (Bayes_ID)</i>						
ESS	178.28	21.35	27.85	125.72	196.51	133.22

Abbreviation: ESS, effective sample size.

**Table 7 Efficiencies of two MCMC algorithms based on 10 simulation replicates**

	Learning phase			Adapted phase			ESS ratio		
	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
$\sigma_a^2$	17.96	10.01	22.66	185.34	162.67	236.52	10.31	16.25	10.46
$\sigma_d^2$	17.08	8.09	24.64	231.32	136.18	318.08	13.54	16.83	12.90
$\sigma_e^2$	34.17	20.29	44.88	231.02	175.47	276.06	6.76	8.64	6.15

Abbreviations: ESS, effective sample size; MCMC, Markov chain Monte Carlo.

Average, minimum and maximum values of ESS are calculated for variance components from the analyses of 10 simulated data sets. Each MCMC analysis was run for 3000 iterations. In addition, the ESS ratio (adapted phase/learning phase) for the 10 simulation replicates is also given.

covariances. The ESS values from Tables 6 and 7 clearly support better mixing properties of variance components in the adapted phase for all the data sets. Our prior allows the chain to mix well, and at the same time, it allows a realistic estimate of the dominance variance in the case of no dominance, because in such a case the prior shrinks the posterior towards zero.

When the target distribution is multimodal, the conventional MCMC algorithm may have difficulties moving between modes. Additionally, the REML method fails to identify different modes of the distribution. Our new adaptive MCMC algorithm was able to visit the different modes even after a low number of iterations and exhibited good mixing properties.

## DISCUSSION

One of the main problems associated with Bayesian analysis of mixed models with several random effects is that the analysis is computationally demanding. Waldmann *et al.* (2008) have shown that the hybrid Gibbs sampler is much faster than the normal blocked Gibbs sampler for estimating additive and dominance genetic variances in the traditional infinitesimal model. In our current study, we

compared the performance of the hybrid Gibbs sampler with an adaptive MCMC method using simulated pedigree data sets with non-zero additive and dominance genetic variances but no inbreeding, showing that the new adaptive MCMC algorithm was almost two times faster than the hybrid Gibbs sampler. To compare the running times, we compared an adaptive MCMC chain of total length 50 000 (a burn-in period of 2000 iterations, 3000 iterations in the learning phase and 45 000 iterations in the adapted phase) with a hybrid Gibbs sampling chain of the same total length (a burn-in period of 2000 iterations and 48 000 iterations from the normal hybrid Gibbs sampling). What is more, the adaptive algorithm has superior mixing properties, as is shown by the ESS in Tables 6 and 7. The increase in speed is partly due to the fact that, unlike the algorithm of Waldmann *et al.* (2008), our adaptive MCMC method does not sample additive and dominance genetic values for individuals. In the adaptive MCMC algorithm (see appendix for details), the determinants and quadratic forms associated with the covariance matrices at the proposed and current points are needed to calculate the likelihood ratio. Once the proposed value is accepted, the determinant and quadratic form at the current point can be replaced by the determinant and quadratic form corresponding to the accepted variance components. This makes the calculation of the likelihood ratio computationally lighter than the block update of the Gibbs sampler. A MATLAB implementation of the adaptive method (described in the appendix) is available in the Supplementary Materials.

During the adapted phase of the algorithm, our sampler generates values only from the marginal posterior of the variance components. Even if our method is primarily intended for the estimation of the genetic variances, it is possible to generate MCMC samples for the additive and dominance genetic values afterwards, by sampling them block-wise from their fully conditional posterior distribution conditionally on each of the values of the variance components in the MCMC sample generated by the adaptive MCMC sampler. In contrast, in the normal hybrid Gibbs sampler, the genetic values are sampled conditionally on each of the values of the variance components. We tested this procedure by calculating the genetic values for the QTLMAS workshop data with the blocked Gibbs sampler conditionally on every 10th realization (of three variance components) out of 45 000 samples from the adapted phase. The linear correlation between the true genetic values (that is, the sum of the additive and dominance values) and the estimated genetic values was around 0.71 for the QTLMAS workshop data. In addition, the genetic values given by ASReML showed a correlation of approximately 0.71 with the true genetic values for the same data set. Our adaptive MCMC genetic values showed a strong correlation of approximately 0.99 with the genetic values from ASReML, demonstrating that our posterior mean estimates were close to the classical point estimates.

De Boer and Hoeschele (1993) showed that the presence of inbreeding changes the mean and complicates the genetic covariance structure of a population. Although the mixed model was not considered here, it can, in principle, account for inbreeding by including a complex covariance matrix among the additive and dominance effects. To accomplish this analysis, our method would require adjustments. However, another type of a model that would suit our estimation framework well is a Gaussian process model (Crossa *et al.*, 2010) or an extension of a ridge regression model (Piepho, 2009; Schulz-Streeck and Piepho, 2010). Then, the dominance relationship matrix would be replaced by a marker similarity matrix or by the covariance function that was proportional to the evaluations of a reproducing kernel evaluated in the marker genotypes.



Identifiability problems can arise, especially when the dominance relationship matrix  $D$  is close to a multiple of the identity matrix. This occurs when the pedigree is incomplete and/or lacks full sibs or double cousins. Then, certain features of the phenotypic observations be attributed to dominance effects can almost as well as to noise. In such a case the joint marginal posterior of the dominance variance and the error variance should be bimodal, and then a conventional MCMC sampler may have difficulties moving between the modes. Gibbs samplers are especially vulnerable, but M–H sampling schemes may behave better. Adding more full-sibs to the pedigree file can improve the multimodality problem to some extent. In our simulation experiments, our adaptive MCMC algorithm was able to explore the entire parameter space with good mixing properties, and therefore was able to detect different modes in the posterior distribution.

The proposal covariance matrix  $(2.38)^2/d S$  from Roberts *et al.* (1997) and Roberts and Rosenthal (2001) is optimal in a large-dimensional context when the posterior is approximately Gaussian (Roberts and Rosenthal, 2007; Rosenthal, 2011). We also experimented with other scalings of the posterior covariance matrix, but the theoretical formula turned out to work well enough in our applications. This scaling factor  $(2.38)^2/d$  was also employed in the MCMC sampler of Fang *et al.* (2011), who introduced a new method for QTL mapping. In their sampling scheme, they utilized REML estimates in the construction of the proposal covariance matrix. If the target distribution is multimodal, this approach may fail to move between different modes. In contrast, our new adaptive MCMC method uses the history of the chain to learn the proposal covariance matrix, which enables the algorithm to move between different modes. The success of adaptive MCMC methods generally depends on how well the proposal covariance structure is learned from the previous history of the chain. Therefore, it is important to use a sufficient number of samples in the learning period. The required sample depends first on the dimensionality and on the other characteristics of the posterior distribution, and second on the mixing properties of the MCMC sampler. Therefore, it is impossible to give general prescriptions for it.

In our study, we also tested adaptation in a version of the model where the random effects were not marginalized away. However, this formulation suffered from poor mixing and slow convergence because of posterior dependencies among the random effects and the variance components (results not shown). The marginalized model (that is, hierarchical model 2) that we used in our study was able to explore the entire parameter space with good mixing properties. The ESS of a parameter is the number of independent samples from the posterior distribution that our correlated MCMC sample is worth. If the ESSs are low, then the autocorrelations will be high, and that may be an indication of poor mixing of the chain. The adaptive scheme was able to decrease the autocorrelation of the chain to yield much larger ESS.

The choice of the prior is one of the important steps in any Bayesian analysis. Generally, the influence of the prior distribution on the posterior is related to the sample size of the data. We carried out a sensitivity analysis using different priors, and most priors seemed to lead to non-zero estimates of dominance variance for the QTLMAS data (results not shown). However, the gamma prior for the precision parameters ( $k_i = 1$  and  $\lambda_i = 0.001$ ) was able to provide good mixing, while still resulting in a realistic estimate of dominance variance in the case of no dominance. This follows because the prior can then shrink the posterior towards zero.

## DATA ARCHIVING

Data have been deposited at Dryad:doi:10.5061/dryad.0p88f.

## ACKNOWLEDGEMENTS

This work was supported by funding from the Theodor-Brinkman-Graduate School of the Faculty of Agriculture at the University of Bonn, research grants from the Academy of Finland and the University of Helsinki's research funds. We are grateful to two anonymous referees for their constructive comments on the manuscript. Also, we gratefully thank Patrik Waldmann and Fabian Hoti for a Gibbs sampling Matlab code that provided a starting point for the development of our method.

- Bauer AM, Hoti F, Reetz TC, Schuh W-D, Léon J, Sillanpää MJ (2009). Bayesian prediction of breeding values by accounting for genotype-by-environment interaction in self-pollinating crops. *Genet Res* **91**: 193–207.
- Bauer AM, Léon J (2008). Multiple-trait breeding values for parental selection in self-pollinating crops. *Theor Appl Genet* **116**: 235–242.
- Bauer AM, Reetz TC, Léon J (2006). Estimation of breeding values of inbred lines using best linear unbiased prediction (BLUP) and genetic similarities. *Crop Sci* **46**: 2685–2691.
- Blasco A (2001). The Bayesian controversy in animal breeding. *J Anim Sci* **79**: 2023–2046.
- Burgueno J, de Los Campos G, Weigel K, Crossa J (2012). Genomic prediction of breeding values when modeling genotype environment interaction using pedigree and dense molecular markers. *Crop Sci* **52**: 707–719.
- Casella G, George EI (1992). Explaining the Gibbs sampler. *Am Stat* **46**: 167–174.
- Chib S, Greenberg E (1995). Understanding the Metropolis-Hastings algorithm. *Am Stat* **49**: 327–335.
- Crossa J, de Los Campos G, Pérez P, Gianola D, Burguño J, Luis Araus J *et al.* (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* **186**: 713–724.
- De Boer IJM, Hoeschele I (1993). Genetic evaluation methods for populations with dominance and inbreeding. *Theor Appl Genet* **86**: 245–258.
- Fang M, Liu J, Sun D, Zhang Y, Zhang Q, Zhang Y *et al.* (2011). QTL mapping in outbred half-sib families using Bayesian model selection. *Heredity* **107**: 265–276.
- García-Cortés LA, Sørensen D (1996). On a multivariate implementation of the Gibbs sampler. *Genet Sel Evol* **28**: 121–126.
- Gasbarra D, Pirinen M, Kulathinal S, Sillanpää MJ (2011). Estimating haplotype frequencies by combining data from large DNA pools with database information. *IEEE/ACM Trans Comput Biol Bioinf* **8**: 36–44.
- Geyer CJ (1992). Practical Markov chain Monte Carlo (with discussion). *Stat Sci* **7**: 473–511.
- Geyer CJ, Thompson A (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *J Am Stat Assoc* **90**: 909–920.
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006). ASReml User Guide, Release 2.0. VSN International: Hemel Hempstead, UK.
- Haario H, Saksman E, Tamminen J (2001). An adaptive Metropolis algorithm. *Bernoulli* **7**: 223–242.
- Hadfield JD (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J Statist Softw* **33**: 1–22.
- Hallander J, Waldmann P, Wang C, Sillanpää MJ (2010). Bayesian inference of genetic parameters based on conditional decompositions of multivariate normal distributions. *Genetics* **185**: 645–654.
- Hanson WD (1963). Heritability. In: Hanson WD, Robinson HF (eds). *Statistical Genetics and Plant Breeding*. National Academy of Sciences-National Research Council: Washington, DC, pp 125–139.
- Hastings WK (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97–109.
- Henderson CR (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**: 69–83.
- Henderson CR (1985a). Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *J Anim Sci* **60**: 111–117.
- Henderson CR (1985b). MIVQUE and REML estimation of additive and nonadditive genetic variances. *J Anim Sci* **61**: 113–121.
- Hoeschele I, VanRaden PM (1991). Rapid inverse of dominance relationships matrices for noninbred populations by including sire by dam subclass effects. *J Dairy Sci* **74**: 557–569.
- Hoti FJ, Sillanpää MJ, Holmström L (2002). A note on estimating the posterior density of a quantitative trait locus from a Markov chain Monte Carlo sample. *Genet Epidemiol* **22**: 369–376.
- Kruuk LEB (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philos Trans R Soc Lond B* **359**: 873–890.
- Lund MS, Sahana G, de Koning DJ, SU G, Carlborg Ö (2009). Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection. *BMC Proc* **3**(Suppl 1) S1.
- Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates: Sunderland, MA.
- Matlab (2007). *High-Performance Numeric Computation and Visualization Software* vol 7. The Math Works Inc.: Natick, Mass, USA.
- Meyer K, Kirkpatrick M (2010). Better estimates of genetic covariance matrices by "bending" using penalized maximum likelihood. *Genetics* **185**: 1097–1110.

- Misztal I (1997). Estimation of variance components with large-scale dominance models. *J Dairy Sci* **80**: 965–974.
- Norris D, Varona L, Ngambi JW, Visser DP, Mbajjorgu CA, Voordewind SF (2010). Estimation of the additive and dominance variances in SA Duroc pigs. *Livest Sci* **131**: 144–147.
- Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H (2006). Joint modelling of additive and non-additive genetic line effects in single field trials. *Theor Appl Genet* **113**: 809–819.
- O'Hara RB, Cano JM, Ovaskainen O, Teplitsky C, Alho JS (2008). Bayesian approaches in evolutionary quantitative genetics. *J Evolution Biol* **21**: 949–957.
- O'Hara BR, Sillanpää MJ (2009). Review of Bayesian variable selection methods: what, how and which. *Bayesian Anal* **4**: 85–118.
- Piepho H-P (2009). Ridge regression and extensions for genomewide selection in maize. *Crop Sci* **49**: 1165–1176.
- Piepho H-P, Ogutu JO, Schulz-Streeck T, Estaghirou B, Gordillo A, Technow F (2012). Efficient computation of ridge-regression best linear unbiased prediction in genomic selection in plant breeding. *Crop Sci* **52**: 1093–1104.
- Plummer M, Best N, Cowles K, Vines K (2006). Coda: convergence diagnosis and output analysis for MCMC. *R News* **6**: 7–11 <http://CRAN.R-project.org/doc/Rnews/>.
- Quaas RL (1976). Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* **32**: 949–953.
- Robert C, Casella G (2004). *Monte Carlo Statistical Methods*, 2nd edn. Springer-Verlag: New York.
- Roberts GO, Gelman A, Gilks WR (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann Appl Prob* **7**: 110–120.
- Roberts GO, Rosenthal JS (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Stat Sci* **16**: 351–367.
- Roberts GO, Rosenthal JS (2007). Coupling and ergodicity of adaptive MCMC. *J Appl Prob* **44**: 458–475.
- Rosenthal JS (2011). Optimal proposal distributions and adaptive MCMC. In Brooks S, Gelman A, Jones GL, Meng X-L (eds). *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC: Boca Raton, pp 93–111.
- Schulz-Streeck T, Piepho H-P (2010). Genome-wide selection by mixed model ridge regression and extensions based on geostatistical models. *BMC Proc* **4**:(Suppl 1) S8.
- Sorensen D, Gianola D (2002). Likelihood, Bayesian and MCMC. *Methods in Quantitative Genetics*. Springer-Verlag: New York, pp 313–314.
- Waagepetersen R, Ibánñez-Escriche N, Sorensen D (2008). A comparison of strategies for Markov chain Monte Carlo computation in quantitative genetics. *Genet Sel Evol* **40**: 161–176.
- Waldmann P, Hallander J, Hoti F, Sillanpää MJ (2008). Efficient MCMC implementation of Bayesian analysis of additive and dominance genetic variances in non-inbred pedigrees. *Genetics* **179**: 1101–1112.
- Wall E, Brotherstone S, Kearney JF, Woolliams JA, Coffey MP (2005). Impact of nonadditive genetic effects in the estimation of breeding values for fertility and correlated traits. *J Dairy Sci* **88**: 376–385.
- Wang CS, Rutledge JJ, Gianola D (1993). Marginal inferences about variance components in a mixed linear models using Gibbs sampling. *Genet Sel Evol* **25**: 41–62.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>).

## APPENDIX

### Calculation of the likelihood ratio

Calculating the determinants of high-dimensional matrices is challenging, because numerical problems increase as the dimension increases. To calculate the likelihood ratio  $p(y|\psi^*)/p(y|\psi)$  in Equation (12), we needed to compute the determinants of the covariance matrices of  $y$  conditionally on the proposed and current point. We scaled these matrices to mitigate numerical problems. Scaling is based on the identity  $\det(s\mathbf{\Sigma}) = s^n \det(\mathbf{\Sigma})$  (valid whenever  $s$  is a scalar and  $\mathbf{\Sigma}$  is an  $n \times n$  matrix) and the identity  $(s\mathbf{\Sigma})^{-1} = \mathbf{\Sigma}^{-1}/s$  (valid whenever  $s$  is a scalar and  $\mathbf{\Sigma}$  is an invertible matrix). We set  $s = 1/\psi_e$ ,  $s^* = 1/\psi_e^*$  as the scaling factors for the current ( $\mathbf{\Sigma}$ ) and proposed ( $\mathbf{\Sigma}^*$ ) values, respectively. Let  $\psi^* = (\psi_a^*, \psi_b^*, \psi_e^*)$  be the proposed values of (inverses of) variance components and  $\psi = (\psi_a, \psi_b, \psi_e)$  be their current values. We calculate the logarithm of the likelihood ratio as

$$\begin{aligned} \log \frac{p(y|\psi^*)}{p(y|\psi)} &= -\frac{n}{2} (\log(s^*) - \log(s)) - \frac{1}{2} (\log(\det(\mathbf{\Sigma}^*/s^*)) \\ &\quad - (\log(\det \mathbf{\Sigma}/s))) - \frac{1}{2s^*} (y'(\mathbf{\Sigma}^*/s^*)^{-1}y \\ &\quad + \frac{1}{2s} (y'(\mathbf{\Sigma}/s)^{-1}y) \end{aligned} \quad (15)$$

Here, we calculate the determinants and quadratic forms using Cholesky decomposition. If  $\mathbf{M} = \mathbf{\Sigma}/s$  is an  $n \times n$  positively definite symmetric matrix, then its Cholesky decomposition is  $\mathbf{M} = \mathbf{L}\mathbf{L}'$ , where  $\mathbf{L}$  is the lower triangular Cholesky factor. The determinant is calculated as  $\log(\det(\mathbf{M})) = 2\sum_i \log(L_{i,i})$  where  $L_{i,i}$  is the  $i$ th diagonal element of  $\mathbf{L}$ . The quadratic form  $y'(\mathbf{\Sigma}/s)^{-1}y = y'\mathbf{M}^{-1}y$  is calculated using the identity  $y'\mathbf{M}^{-1}y = (\mathbf{L}^{-1}y)'(\mathbf{L}^{-1}y)$ , where  $\mathbf{L}^{-1}y$  is calculated by solving  $\mathbf{z}$  from the equation  $\mathbf{L}\mathbf{z} = y$ .

### Effective sample size (ESS)

ESS (Geyer, 1992; Waagepetersen *et al.*, 2008) is the approximate number of independent samples that would deliver the same estimation accuracy as our dependent MCMC samples. ESS is based

on the central limit theorem for MC. Let  $x_0, x_1, \dots$  be the MC and consider a scalar valued function  $h$  defined on the state space. If the MC satisfies a central limit theorem for this function, then as the sample size increases

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=0}^n h(x_i) - E_\pi h(x) \right) \xrightarrow{d} N(0, \tau_h \text{var}_\pi h(x)), \quad (16)$$

where  $\pi$  is the stationary density of the MC,  $E_\pi h(x)$  is the expected value of  $h(x)$  under  $\pi$ ,  $\text{var}_\pi h(x)$  is the variance of  $h(x)$  under  $\pi$  and  $\tau_h$  is the integrated autocorrelation time for estimating  $E_\pi h(x)$  for the given MC, defined as

$$\tau_h = 1 + 2 \sum_{i=0}^{\infty} \text{corr}_\pi(h(x_i), h(x_{i+k})), \quad (17)$$

Here  $\text{corr}_\pi$  is the correlation between the values when the chain is started from the stationary distribution ( $x_0 \sim \pi$ ). However, if  $y_0, y_1, \dots, y_n$  are i.i.d samples from the stationary distribution  $\pi$ , then by the central limit theorem for i.i.d. sequences

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=0}^n h(y_i) - E_\pi h(x) \right) \xrightarrow{d} N(0, \text{var}_\pi h(x)), \quad (18)$$

Comparing Equations (16) and (18) gives  $\text{ESS} = n/\tau_h$  as the ESS, when we estimate the expectation  $E_\pi h(x)$  using the arithmetic mean of a large number of values  $h(x_1), \dots, h(x_n)$  based on the history of the MC. There are different methods available for estimating  $\tau_h$  and ESS, but we simply used the R package coda.

### Adaptive MCMC algorithm

To learn an efficient proposal distribution, a reasonable number of MCMC samples is required. In the following, steps 2 and 3 describe the details of the learning phase, and steps 4 and 5 those of the adapted phase. Blocked Gibbs sampler (step 3) that uses block updates of parameters in  $\theta$  has good mixing properties compared with the single-site Gibbs sampler (step 2). However, then one needs to solve very large linear systems (step 3d) which makes this

computationally challenging. The hybrid Gibbs sampler, which uses block updates every 50th iteration, is much faster than plain blocked Gibbs sampling, and it has better mixing properties than the single-site Gibbs sampler. The use of block updates for variance components in the adaptive MCMC method helps the chain to converge reasonably fast to its equilibrium distribution.

The algorithm of the adaptive MCMC is as follows:

1. Initialize  $\psi_a, \psi_d$  and  $\psi_e$  with some reasonable positive values. Set  $k_a^* = k_a + q/2$ ,  $k_d^* = k_d + q/2$ , and  $k_e^* = k_e + n/2$ .

2. Single-site Gibbs sampling:

- (a) Sample  $\theta_i$  from  $N(\hat{\theta}, 1/(C_{i,i}\psi_e))$ , where  $\hat{\theta} = (\mathbf{W}'\mathbf{y} - \mathbf{C}_{i,-i}\theta_{-i})/C_{i,i}$ . Here,  $\theta_{-i}$  is  $\theta$  without its  $i$ th component,  $\mathbf{C}_{i,-i}$  is the  $i$ th row of  $\mathbf{C}$  without its  $i$ th component and finally,  $C_{i,i}$  is the  $i$ th diagonal element of  $\mathbf{C}$ .
- (b) Calculate  $\lambda_a^* = \lambda_a + (\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a})/2$ ,  $\lambda_d^* = \lambda_d + (\mathbf{d}^T\mathbf{D}^{-1}\mathbf{d})/2$  and  $\lambda_e^* = \lambda_e + 1/2\|\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}_1\mathbf{a} - \mathbf{Z}_2\mathbf{d}\|^2$ .
- (c) Sample the precision parameters  $\psi_i$  from Gamma( $k_i^*, \lambda_i^*$ ) for  $i = a, d, e$ .
- (d) Calculate  $\alpha_a = \psi_a/\psi_e$ ,  $\alpha_d = \psi_d/\psi_e$  and update the coefficient matrix  $\mathbf{C}$  using Equation (10).

3. Block Gibbs sampling (every 50th iteration):

- (a) Generate  $\mathbf{a}^*$  from  $\text{MVN}(\mathbf{0}, \mathbf{A}/\psi_a)$  and  $\mathbf{d}^*$  from  $\text{MVN}(\mathbf{0}, \mathbf{D}/\psi_d)$ .
- (b) Generate  $\mathbf{z}^*$  from  $\text{MVN}(\mathbf{Z}_1\mathbf{a}^* + \mathbf{Z}_2\mathbf{d}^*, \mathbf{I}/\psi_e)$ .
- (c) Calculate  $\mathbf{W}'(\mathbf{y} - \mathbf{z}^*)$ .
- (d) Calculate  $\theta$  as  $[\mathbf{0}', \mathbf{a}^{*'}, \mathbf{d}^{*'}] + \mathbf{C}^{-1}\mathbf{W}'(\mathbf{y} - \mathbf{z}^*)$ , where  $\mathbf{0}$  is the zero vector of the size of the fixed effects vector  $\beta$ .
- (e) Calculate  $\lambda_a^* = \lambda_a + (\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a})/2$ ,  $\lambda_d^* = \lambda_d + (\mathbf{d}^T\mathbf{D}^{-1}\mathbf{d})/2$  and  $\lambda_e^* = \lambda_e + 1/2\|\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}_1\mathbf{a} - \mathbf{Z}_2\mathbf{d}\|^2$ .
- (f) Sample the precision parameters  $\psi_i$  from Gamma( $k_i^*, \lambda_i^*$ ), for  $i = a, d, e$ .

- (g) Calculate  $\alpha_a = \psi_a/\psi_e$ ,  $\alpha_d = \psi_d/\psi_e$  and update the coefficient matrix  $\mathbf{C}$  using Equation (10).

4. Setting up the adapted MCMC (after the learning period):

- (a) Transform the samples from the learning period into logarithmic scale with the formula  $\tau_i = -\log(\psi_i)$ , for  $i = a, d, e$ .
- (b) Calculate the sample covariance matrix  $\hat{\mathbf{S}}$  from the transformed variables  $\tau_i$ . Calculate the proposal covariance matrix  $\mathbf{S}_p = (2.38)^2/d \hat{\mathbf{S}}$ , where  $d = 3$ . Initialize the current state  $\tau$  from the last state visited during the learning phase.

5. The iterations during the adapted phase:

- (a) Generate proposed values  $\tau^*$  from the Gaussian distribution  $\text{MVN}(\tau, \mathbf{S}_p)$ . Calculate the  $\psi$  values and  $\psi^*$  values corresponding to the current and the proposed vectors.
- (b) Calculate the logarithm of the M–H acceptance ratio  $r$  from Equation (11) by calculating the logarithm of the prior ratio  $p(\tau^*)/p(\tau)$ , where we take into account the Jacobian ratio from Equation (14), and the logarithm of the likelihood ratio from Equation (15).
- (c) Draw  $u$  from the uniform distribution over  $[0,1]$  and accept the proposed value  $\tau^*$ , if  $u < r$ . If the proposal is accepted, then we assign  $\tau = \tau^*$ , otherwise the current value is retained.

In a random-walk M–H algorithm, such as we used in the adapted phase, the acceptance rate (the ratio between the number of times the proposed value is accepted and the total number of iterations after the learning period) should be between 10 and 50%, but the optimal rate is around 23% (see Roberts and Rosenthal, 2001; Rosenthal, 2011).