A Variational Bayes Genomic-Enabled Prediction Model with Genotype × Environment Interaction

Osval A. Montesinos-López,* Abelardo Montesinos-López,† José Crossa,‡ José Cricelio Montesinos-López,§ Francisco Javier Luna-Vázquez,* J osafhat Salinas-Ruiz,** José R. Herrera-Morales,* and Raymundo Buenrostro-Mariscal*

*Facultad de Telemática, Universidad de Colima, 28040, México, †Departamento de Matemáticas, Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), Universidad de Guadalajara, 44430, Jalisco, México, ‡Biometrics and Statistics Unit and Global Wheat Program of the International Maize and Wheat Improvement Center (CIMMYT), 06600 México D.F., México, §Departamento de Estadística, Centro de Investigación en Matemáticas (CIMAT), 36240 Guanajuato, México, **Colegio de Postgraduados, Campus Córdoba, Km. 348 Carretera Federal Córdoba-Veracruz, Amatlan de los Reyes, 94946, México

ABSTRACT There are Bayesian and non-Bayesian genomic models that take into account G×E interactions. However, the computational cost of implementing Bayesian models is high, and becomes almost impossible when the number of genotypes, environments, and traits is very large, while, in non-Bayesian models, there are often important and unsolved convergence problems. The variational Bayes method is popular in machine learning, and, by approximating the probability distributions through optimization, it tends to be faster than Markov Chain Monte Carlo methods. For this reason, in this paper, we propose a new genomic variational Bayes version of the Bayesian genomic model with G×E using half-t priors on each standard deviation (SD) term to guarantee highly noninformative and posterior inferences that are not sensitive to the choice of hyper-parameters. We show the complete theoretical derivation of the full conditional and the variational posterior distributions, and their implementations. We used eight experimental genomic maize and wheat data sets to illustrate the new proposed variational Bayes approximation, and compared its predictions and implementation time with a standard Bayesian genomic model with G×E. Results indicated that prediction accuracies are slightly higher in the standard Bayesian model with G×E than in its variational counterpart, but, in terms of computation time, the variational Bayes genomic model with G×E is, in general, 10 times faster than the conventional Bayesian genomic model with G×E. For this reason, the proposed model may be a useful tool for researchers who need to predict and select genotypes in several environments.

Nowadays, genomic-assisted selection (GS) is a very attractive breeding tool for selecting candidate genotypes early, using only dense molecular markers without the need to phenotype the validation population for several complex traits. However, to take advantage of all the available information (markers, multiple-traits, multiple-environments, etc.), more efficient statistical methods are desirable for genome-enabled prediction models. Markov Chain Monte Carlo (MCMC) methods provide an effective means of dealing with complex models, and are widely used in genomic prediction models because they allow simulation of high-dimensional distributions to arbitrary levels of accuracy. However, MCMC methods are computationally expensive, especially in the presence of intractable multivariate integrals in posterior densities and likelihood (Ormerod and Wand 2012).

For the above reasons, the variational Bayes method (VBM), which is very popular in computer science and the machine learning community, has been adopted by the statistical science community, because, although variational approximations kill some accuracy of the MCMC methods, they offer vast improvements in terms of computational speed and time. The VBM facilitates approximate inference for the parameters in complex statistical models, and provides fast, deterministic alternatives to Monte Carlo methods (Ormerod and Wand 2010). Most VBM
applications are found in the machine learning community, and there is enough empirical evidence that the VBM tends to be faster than classic methods such as MCMC sampling. Also, this method is easier to scale to large data—it has been applied to problems such as large-scale document analysis, high-dimensional neuroscience, and computer vision, as well as in computational biology, robotics, speech recognition, marketing, optimal control, reinforcement learning, statistical network analysis, astrophysics, and the social sciences (Blei et al. 2016). However, the VBM has been studied less rigorously than MCMC, and its statistical properties are less understood (Blei et al. 2016).

In Gs and prediction, we have found some applications for the VBM, for: (1) clustering analysis of gene-expression data (Teschendorf et al. 2005); (2) conducting genome-wide association studies (Logsdon et al. 2010; Carbonetto and Stephens 2012); (3) predicting genomic breeding values in univariate and multiple traits (Hayashi and Iwata 2013; Yamamoto et al. 2017); (4) estimating quantitative trait locus effects (Li and Sillanpää 2012, 2013), and (5) estimating variance components in the context of animal science (Arakawa et al. 2016).

Most VBMs applied in genomic-enabled prediction use a simple linear model of the form \( y_i = \mu + \sum_{j=1}^{m} x_{ij} \beta_j + e_i \), where \( y_i \) is the phenotype of the \( i \)th individual, \( \mu \) is the sample mean, \( x_{ij} \) defines the genotype of the \( j \)th marker, and \( e_i \) is an error term normally distributed with mean zero and variance \( \sigma^2_e \) (Logsdon et al. 2010; Carbonetto and Stephens 2012; Yamamoto et al. 2017; Li and Sillanpää 2012; Teschendorf et al. 2005). Exceptions to the use of simple linear models in VBM were studied by Hayashi and Iwata (2013), who modeled multiple traits without effects of environments and without interaction terms; by Arakawa et al. (2016), who used a model for univariate responses with mixed effects without interaction terms; and in the research of Li and Sillanpää (2013), who used a model for functional regression without random effects and without interaction terms.

Based on the above, the purpose of this study was to propose a univariate genomic VBM that takes into account genotype × environment interaction (G×E), and could be an alternative to the existing Bayesian and non-Bayesian models with G×E described by Jarquin et al. (2014) in the reaction norm model. Since animal and plant breeding programs today perform genetic evaluations of large numbers of lines, or animals in many environments, for several traits and genotyped with thousands (or millions) of genetic markers, complex Bayesian models may take days or weeks to do this, and, as a result, the existing models become untenable.

For this reason, in order to retain all the useful features of the Bayesian genomic models with G×E at a reduced computational cost, we propose, within the general framework of variational Bayes, a method for estimating the univariate genomic model with G×E, which we call the variational Bayes multi-environment (VBM) model. This means that the VBME will recast the problem of computing posterior densities by the Gibbs sampling of the conventional Bayesian methods with G×E—referred to here as the Bayesian multi-environment (BME) model—as an optimization problem using the VBM, which consists of introducing a class of approximating distributions of the latent variables, then optimizing some criteria to find the distribution within this class that best matches the posterior (Carbonetto and Stephens 2012). Also, it is important to point out that the newly proposed method uses, as a prior for all standard deviations (SDs), a half-t distribution that is a high noninformative prior distribution (Huang and Wand 2013). We derived the full conditional and variational posterior distribution of the VBME model, explained its implementation, and applied it to eight experimental genomic (maize and wheat) data sets with two main objectives: (1) to compare the accuracy of the VBME with that of the standard BME model; and (2) to compare the computational processing time of VBME with that of BME.

**MATERIALS AND METHODS**

**The variational Bayes method using the density transform approach**

The most common variants of the VBM are the density transform approach and the tangent transform approach. We will explain the basis of the density transform approach; readers interested in understanding the tangent transform approach can consult Ormerod and Wand (2010). The density transform approach consists of approximating the posterior densities through other densities for which inference is more tractable. In Bayesian statistics, the inference problem is to compute the conditional distribution of unknown parameters given the observations \( p(\theta | y) = p(y|\theta)p(\theta) \); the denominator is known as the marginal distribution of the observations or model evidence, \( \theta \in \Theta \) is the parameter vector, and \( p(y, \theta) \) is the joint likelihood of the data and model parameters. The essence of the density transform approach is to approximate the posterior density \( p(\theta | y) \) by a function \( q(\theta) \), for which the \( q \)-dependent lower bound \( \int p(y; q) \) on the marginal likelihood is given by

\[
\mathcal{L}(\theta; y) = \int q(\theta) \log \left( \frac{p(y, \theta)}{q(\theta)} \right) d\theta
\]

Expression (1) is more tractable than the marginal likelihood, \( p(y) \). Indeed, the maximization of \( \mathcal{L}(\theta; y) \) with respect to \( q \) is equivalent to minimizing the Kullback-Leibler divergence between \( q(\theta) \) and \( p(y, \theta) \), K(L(q(·), p(·|y))). This is because the marginal likelihood satisfies

\[
p(y) = \exp \left( \text{KL} \left( q(\cdot), p(\cdot|y) \right) \right) \mathcal{L}(\theta; y) \geq p(y; q)
\]

which can be derived through Jensen’s inequality, and the equality holds if, and only if, \( q(\theta) = p(\theta|y) \) almost everywhere (Ormerod and Wand 2010). To achieve tractability, \( p(y; q) \) is maximized over a restricted, more manageable, \( q \) class of densities.

Here, for the implementation, we chose, as restrictions for the \( q \) class of densities, the mean field variational family restriction, where the latent variables are mutually independent, and each is governed by a distinct factor in the variational distribution, that is, \( q(\theta) \) factorizes into \( \prod_{i=1}^{n} q_i(\theta_i) \) for some partitions \( \{ \theta_1, \ldots, \theta_m \} \) of \( \Theta \). Each \( \theta_i \) is governed by its own distribution \( q_i(\theta_i) \), and, in principle, each variational factor can take on any parametric form that is appropriate for the corresponding random variable. In the optimization process, these variational factors are chosen to maximize the evidence lower bound given in Equation (2). One advantage of using the mean field variational family restriction over \( q(\theta) \) is that it is possible to obtain explicit solutions for each product component in terms of the others when conjugate priors are used. These lead to an interactive scheme for obtaining simultaneous solutions. The solutions rely on result 2.1.

Result 2.1. Let \( v \) and \( w \) be continuous random vectors with joint density \( p(v, w) \). Then

\[
\sup_q \left\{ \int q(v) \log \frac{p(v, w)}{q(v)} dv \right\}
\]

is attained by \( q^*(v) = p(v|w) \).

One disadvantage of the mean field variational family restriction is its tendency to underestimate the variability of parameter estimates; it is also not a good option when the parameters under study have a high degree of dependence, which causes degradation in the resulting inference. Conversely, if the posterior dependence between parameters is weak, then the product density restriction could lead to very accurate approximate inferences (Ormerod and Wand 2010; Blei et al. 2016). When conditional conjugate priors are used, given the other variational
factor components, the optimal densities of each variational factor, \( q_i(\theta_i) \), are proportional to the exponentiated expected log of the full conditionals obtained with Gibbs sampling, that is,

\[
q_i(\theta_i) \propto \exp \left\{ E_{-,\theta_i} \log p(\theta_i) | \text{ELSE} \right\}
\]

where \( p(\theta_i | \text{ELSE}) \) are the known full conditionals in the MCMC literature, and the expectations on the right-hand side of Equation (3) are computed with respect to the density \( q(\theta)/q(\theta_i) \). From this equation, an interactive scheme can be used to obtain the optimal densities. Equation (3) allows obtaining all the variational posteriors for implementing the VBM using mean variational approximation.

### Statistical model

Since we wanted to develop a variational Bayes version of the BME model, we used \( y_{ijk} \) to represent the normal phenotype from the \( k \)th replication of the \( j \)th line in the \( i \)th environment \((i = 1, 2, \ldots, I; j = 1, 2, \ldots, J; k = 1, 2, \ldots, K)\), where \( K \) represents the number of replicates of each line in each environment under study. Therefore, the total number of observations is \( n = I \times J \times K \). Therefore, the BME model is

\[
y_{ijk} = E_i + g_j + g_E g_{ij} + e_{ijk}
\]

where \( E_i \) represents the effect of \( i \)th environment, \( g_j \) represents the genomic effect of \( j \)th line, and is assumed as a random effect, \( g_E g_{ij} \) is the interaction between the genomic effect of the \( j \)th line and the \( i \)th environment, and is assumed a random effect, and \( e_{ijk} \) is a random error term associated with the \( k \)th replication of the \( j \)th line in the \( i \)th environment.

In matrix notation, the model given in Equation (4) including all the information is expressed as:

\[
Y = XB + Z_1b_1 + Z_2b_2 + e
\]

where \( Y \) is the response vector of order \( n \times J \), \( X \) is the design matrix of environment effects of order \( n \times I \), \( B \) is the regression coefficient vector of order \( I \times 1 \) associated with the environment effects, \( Z_1 \) is the design matrix of order \( n \times J \), associated with the random effects of lines, \( b_1 \) is the vector of random effects of lines of order \( J \times 1 \), \( Z_2 \) is the design matrix of order \( n \times J \), associated with the interaction random effects between lines and environments, \( b_2 \) is the vector of interaction random effects of lines and environments of order \( J \times 1 \), and \( e \) is the vector of random error terms of order \( nx1 \). Then, we assumed that \( b_1 \sim N(0, G_1) \), \( b_2 \sim N(0, G_2) \), and \( e \sim N(0, I, \sigma^2_e) \), also that this vector is statistically independent, where \( G \) denotes the genomic relationship matrix, and was calculated as suggested by VanRaden (2008). \( G_2 = I \times G \), \( \otimes \) denotes a Kronecker product, where \( I_1 \) is an identity matrix of order \( I \times I \), which indicates that we are assuming independence between environments.

### BME model

The joint posterior density of the parameter vector of the BME model is given by

\[
p(B_1, B_2, \sigma^2_1, \sigma^2_2, \sigma^2_3, \sigma^2_4, \text{ab}, a_{b1}, a_{a2}, a_{c}) \\
\propto p(Y | B_1, B_2, \sigma^2_1) p(B_1 | \sigma^2_2) p(\text{ab} | \sigma^2_3) p(a_{b1}) p(a_{a2}) p(a_{c})
\]

which is not a known distribution that is easy to simulate, so we will use MCMC techniques to obtain samples of this by means of the Gibbs sampler. The notation \( x \sim IG(A, B) \) represents a random variable, \( x \), that has an inverse Gamma distribution with shape \( A \) and scale \( B \) parameters, \( A, B > 0 \), with density function \( P(x) = B^A x^{-A-1} \exp[-(B/x)], x > 0 \). The following prior specification is given:

\[
\beta | \sigma^2_\beta \sim N_{p}(\beta, \Sigma_{p} \sigma^2_\beta), \sigma^2_\beta | a_{b} \sim IG(v_{b}/2, v_{b}/a_{b}), a_{b} \sim IG(1/2, 1/A^2_b), \sigma^2_1 | a_{c} \sim IG(v_{c}/2, v_{c}/a_{c}), a_{c} \sim IG(1/2, 1/A^2_c), \sigma^2_2 | a_{b2} \sim IG(v_{b2}/2, v_{b2}/a_{b2}) \text{ and } a_{b2} \sim IG(1/2, 1/A^2_{b2}).
\]

It is important to point out that the hierarchical priors given to the SDs induce Half-f priors to achieve arbitrary high noninformativity in this parameter (Huang and Wand 2013). Note that the use of this type of priors of our proposed model for genomic-enabled prediction is different than the priors used in the context of the existing Bayesian univariate models with GxE. Details of all full conditional distributions derived from the Gibbs sampler for implementing these BME and VBME models are given in Appendices A and B, whereas the R code for implementing the VBME model is provided in Appendix D.

### Variational version of the BME model

As an alternative to the Gibbs sampler for estimating the parameters, in this section, we will derive the optimal densities under the mean field variational family restriction, assuming the following factorization of \( q(\theta) \):

\[
q(\beta, b_1, b_2, \sigma^2_1, \sigma^2_2, \sigma^2_3, \sigma^2_4, \text{ab}, a_{b1}, a_{b2}, a_{c}) = q(\beta)q(b_1)q(b_2)q(\sigma^2_1)q(\sigma^2_2)q(\sigma^2_3)q(\sigma^2_4)q(\text{ab})q(a_{b1})q(a_{b2})q(a_{c})
\]

The difficult problem of finding the exact posterior (6) became less difficult, i.e., it now consists of finding an approximate parametric posterior \( q(\beta, b_1, b_2, \sigma^2_1, \sigma^2_2, \sigma^2_3, \sigma^2_4, \text{ab}, a_{b1}, a_{b2}, a_{c}) \) with moments (i.e., parameters). Inference then reduces to finding a density \( q \) that minimizes a measure of dissimilarity between \( q \) and \( p \). This can be achieved by maximizing the log of the evidence lower bound that is an approximation of the BME model using Equation (6) with respect to (the moments of) \( q \). For details, see MacKay (1995), Attias (2000), Ghahramani and Beal (2001) and Bishop (1999).

Because in the Bayesian formulation of the model given in Equation (5), we use conditional conjugate priors over parameters of the corresponding partition of \( \theta \), in the factorization assumed for \( q(\theta) \) in Equation (7), all the optimal distributions (Equation 3) of the BME model of each variational factor, \( q(\theta_i) \), belong to the same family as the full conditional distributions for the BME model. Because the same Bayesian model (BME) was converted to its variational counterpart (VBME), both models are comparable. The derivation details of all these densities are given in Appendix A.

In summary, the iterative scheme used to obtain the optimal density \( q(\theta) \) over the class of densities given in Equation (7), is given by the following algorithm. Before presenting the algorithm, we provide some basic notations. We will use \( \mu_\beta(\theta_i) \) and \( \Sigma_\beta(\theta_i) \) to denote the mean and variance of \( q_i \) distribution. For vector parameter \( \theta \), we use the analogously defined \( \mu_{\beta}(\theta) \) and \( \Sigma_{\beta}(\theta) \). Therefore, the expected values for the inverse of the parameters \( \sigma^2_1, \sigma^2_2, \sigma^2_3, \sigma^2_4, \text{ab}, a_{b1}, a_{b2}, a_{c} \) and \( a_{b} \) used were:

\[
\mu_{q}(\sigma^2_1) = 0.5(\nu_{b1} + 1)/(B_{q}(\sigma^2_1)), \quad \mu_{q}(\sigma^2_2) = 0.5(\nu_{b2} + 1)/(B_{q}(\sigma^2_2)), \quad \mu_{q}(\sigma^2_3) = 0.5(\nu_{c} + 1)/(B_{q}(\sigma^2_3)), \quad \mu_{q}(\sigma^2_4) = 0.5(\nu_{c} + 1)/(B_{q}(\sigma^2_4)), \quad \mu_{q}(\text{ab}) = 0.5(\nu_{b} + 1)/(B_{q}(\text{ab})), \quad \mu_{q}(\text{ab}) = 0.5(\nu_{b} + 1)/(B_{q}(\text{ab}))
\]

respectively, while, for the scale parameters, we will use the notation \( B_{q}(\theta_i) \). Therefore, \( B_{q}(\sigma^2_1), B_{q}(\sigma^2_2), B_{q}(\sigma^2_3), B_{q}(\sigma^2_4), B_{q}((\text{ab})) \)
and $B_q(a_k)$ represent the scale parameters of the variational posterior distribution of $\alpha_1^2, \alpha_2^2, \alpha_3^2, \alpha_4^2, \alpha_5^2, \alpha_6^2, \alpha_7^2, \alpha_8^2$, and $\alpha_9^2$ respectively. Also, $e_0 = Y - X\mu_0|\beta| = \Sigma_1 \mu_0|\beta| - \Sigma_2 \mu_0|\beta|$, and $\beta_0|\beta|$ is the prior of the beta regression coefficients.

**Algorithm for implementing the VBME**

Initialize: $\mu_0|\beta| = 0, \mu_0|\beta| = 0, \mu_0|\beta| = 0, \Sigma_0|\beta| = I_1, \Sigma_0|\beta| = I_1, B_q(a_1), B_q(a_2), B_q(a_3), B_q(a_4), B_q(a_5), B_q(a_6), B_q(a_7) > 0$. Cycle:

$$
\Sigma_0|\beta| \leftarrow \left( \Sigma_0|\beta|^{-1} \mu_0|\beta|^{-1} X^T \left[ I_n \mu_0|\beta|^{-1} \right] X \right)^{-1}
$$

$$
\mu_0|\beta| \leftarrow \Sigma_0|\beta| \left( \Sigma_0|\beta|^{-1} \beta_0|\beta| + X^T \left[ I_n \mu_0|\beta|^{-1} \right] \right)
$$

$$
\times \left[ Y - \sum_{h=1}^{2} Z_h \mu_0|\beta| \right]
$$

$$
B_q(a_0) \leftarrow 1/A_{a_0}^2 + \nu_r \mu_0|\beta| \left( 1/a_0 \right)
$$

$$
\Sigma_0|\beta| \leftarrow \left( G_{0.1}^{-1} \mu_0|\beta|^{-1} + Z_1^T \left[ I_n \mu_0|\beta|^{-1} \right] Z_1 \right)^{-1}
$$

$$
\mu_0|\beta| \leftarrow \Sigma_0|\beta| \left( Z_1^T \left[ I_n \mu_0|\beta|^{-1} \right] \right) \left[ Y - X \mu_0|\beta| - Z_2 \mu_0|\beta| \right]
$$

$$
B_q(a_1) \leftarrow \mu_0|\beta| \left( G_{0.1}^{-1} \mu_0|\beta|^{-1} \right) + \nu_r \mu_0|\beta| \left( a_0 \right)
$$

$$
B_q(a_2) \leftarrow 1/A_{a_2}^2 + \nu_r \mu_0|\beta| \left( 1/a_2 \right)
$$

To obtain (8), several substitutions were made for simplification; it is valid only for the VBME model proposed. In this algorithm, the updates of the parameters should be as given in the order proposed for the algorithm.

**Model implementation**

Recall that the hyperparameters of the prior distributions are values given by the user in the Bayesian paradigm. In this case, the hyperparameters used for the BME model given in the previous section were $\beta_0|\beta| = 0_1, \Sigma_0 = I_1, \nu_r = \nu_l = \nu_s = 2, A_0 = 2, A_0 = 2, A_0 = 10, 000$. All these hyperparameters were chosen to lead weakly informative priors. The implementation of both models (BME and VBME) was done in the R statistical software (R Core Team 2016). The full conditional and posterior distributions given in Appendix A were used for the BME and VBME models, respectively. Also, for the VBME model, an R code is given in Appendix D. The BME model was implemented with the MCMC method with 40,000 iterations with a burn-in of 20,000 and a thinning of five; therefore, 4000 samples were used for inference.

**Assessing prediction accuracy**

Prediction accuracy was assessed using 20 training (trn)-testing (tst) random partitions; we used this approach because one can obtain as many partitions as one needs with a replicated trn-tst design. The implemented cross-validation (CV1) mimics a situation where lines were evaluated in some environments for the trait of interest, but some lines are missing from all the other environments. We assigned 80% of the observations to the trn set and the remaining 20% to the tst set. Of the variety of methods for comparing the predictive posterior distribution to the observed data (generally termed “posterior predictive checks”), we used the Pearson correlation. Models with higher correlation values indicate better predictions. Under the BME, the predicted observations were calculated as $\hat{y}_i = X\hat{\beta}_i + Z_1\hat{b}_1 + Z_2\hat{b}_2$, where $\hat{\beta}_i$, $\hat{b}_1$, and $\hat{b}_2$ are estimates of $\beta_i$, $b_1$, and $b_2$, these estimates correspond to the posterior mean of $s$ collected Gibbs samplers to estimate the predicted values after discarding those used for the burning period, while under the
VBME, \( \hat{Y} = X\mu + Z_1\mu_{(b_1)} + Z_2\mu_{(b_2)} \), as described in the algorithm for implementing the VBME.

Experimental data sets
Here, we present the information on the data sets used for implementing the proposed models. A total of eight data sets were used (three maize and five wheat).

Maize data sets 1–3: A total of 309 doubled haploid maize lines was phenotyped and genotyped; this is part of the data set used by Crossa et al. (2013) and Montesinos-López et al. (2016), which comprised a total of 504 doubled haploid lines derived by crossing and backcrossing eight inbred lines to form several full-sib families. Traits available in this data set include grain yield (MaizeGY), anthesis-silking interval (MaizeASI), and plant height (MaizePH); each of these traits was evaluated in three optimum rainfed environments (Env1, Env2, and Env3). The experimental field design used in each of the three environments was an alpha lattice incomplete block design with two replicates. Data were preadjusted using estimates of block and environmental effects derived from a linear model that accounted for the incomplete block design within environment, and for environmental effects. The genomic data were obtained through genotyping-by-sequencing (GBS) for each maize chromosome; the number of markers after initial filtering and the number of markers after imputation were summarized in Crossa et al. (2013). Filtering was first done by removing markers that had >80% of the maize lines with missing values, and then markers with minor allele frequency ≤0.05 were deleted. The total numbers of GBS data were 681,257 single nucleotide polymorphisms (SNPs), and, after filtering for missing values and minor allele frequency, 158,281 SNPs were used for the analyses. About 20% of cells were missing in the filtered GBS information used for prediction; these missing values were replaced by their expected values before doing the prediction.

Wheat data sets 4–5: The first two wheat data sets came from a total of 250 wheat lines that were extracted from a large set of 39 yield trials grown during the 2013–2014 crop season in Ciudad Obregon, Sonora, Mexico (Rutkoski et al. 2016). The trials were sown in mid-November and grown on beds with five and two irrigations plus drip irrigation. Days to heading (WheatDTHD) were recorded as the number of days from germination until 50% of spikes had emerged in each plot, in the first replicate of each trial, while plant height (WheatPTH) was recorded in centimeters.

For the first two wheat data sets, GBS was used for genome-wide genotyping. Single nucleotide polymorphisms were called across all lines using the TASSEL GBS pipeline anchored to the genome assembly of Chinese Spring. Single nucleotide polymorphisms were extracted and markers were filtered so that percent missing data did not exceed 80 and 20%, respectively. Individuals with >80% missing marker data were removed, and markers were recorded as −1, 0 and 1, which indicate homozygous for the minor allele, heterozygous, and homozygous for the major allele, respectively. Next, markers with <0.01 minor allele frequency were removed, and missing data were imputed with the marker mean. A total of 12,083 markers remained after marker editing.

Wheat data set 6: This data set, from CIMMYT’s Global Wheat Program, was used by Crossa et al. (2010) and Cuevas et al. (2016a,b) and includes 599 wheat lines derived from 25 yr (1979–2005) of Elite Spring Wheat Yield Trials (ESWYT). The environments represented in these trials are four agroclimatic regions (mega-environments). The phenotypic trait considered here was grain yield (WheatYield) of the 599 wheat lines evaluated in each of the four mega-environments. The 599 wheat lines were genotyped using 1447 Diversity Array Technology (DArT) markers generated by Triticarte Pty. Ltd. (Canberra, Australia; http://www.triticarte.com.au). Markers with a minor allele frequency (MAF) (0.05) were removed, and missing genotypes were imputed using samples from the marginal distribution of marker genotypes. The number of DArT markers after edition was 1279.

Wheat data sets 7–8: These two data sets were described and used by Lopez-Cruz et al. (2015) and Cuevas et al. (2016b) for proposing a marker × environment interaction model. The phenotypic data consisted of adjusted grain yield (tons/hectare) records collected during three evaluation cycles of different inbred lines evaluated in different environments, and denoted as WheatYield2 and WheatYield3. The environments were three irrigation regimes (moderate drought stress, optimal irrigation, and drought stress), two planting systems (bed and flat planting), and two different planting dates (normal and late). All trials were planted using a lattice design with three replicates in each environment at CIMMYT’s main wheat breeding station in Cd. Obregón, México. The phenotype used in the analysis was the Best Linear Unbiased Estimate (BLUE) of grain yield obtained from a linear model applied to the alpha-lattice design of each cycle-environment combination. The trait WheatYield2 had 693 wheat lines evaluated in four environments, and the trait WheatYield3 included 670 wheat lines evaluated in four environments. Genotypes were derived using GBS technology, and markers with a MAF of 0.05 were removed. All markers had a high incidence of uncalled genotypes, so we applied thresholds for incidence of missing values, and focused on maintaining relatively large and similar numbers of markers per data set. After editing the missing markers, we had a total of 15,744 GBS markers for analyzing traits WheatYield2 and WheatYield3.

Data availability
Phenotypic and genotypic data on the eight data sets included in this study can be downloaded from http://hdl.handle.net/11529/10907.

RESULTS
The results of this research are given in two sections. The first section compares the parameter estimates and implementation time under both models (BME and VBME). The second section compares the prediction accuracy of both models in each of the eight data sets under study.

Comparing parameter estimates and the implementation time of models BME and VBME
Table 1 gives the parameter estimates under both models (BME and VBME) of the \( \beta \) coefficients, variance components (\( \sigma_1^2 \), \( \sigma_2^2 \) and \( \sigma_3^2 \)), Pearson correlation between the observed and predicted phenotypes of the full data sets, and the time in minutes required to implement each model on the data sets under study. In general, the beta coefficient estimates obtained under the BME and VBME models are very similar, but there are substantial differences in the estimates of variance components, given that a little more than half of them were lower under the VBME model as compared to those estimates obtained under the BME. Also, in seven out of eight data sets, the implementation time was ~10 times lower under the VBME than under BME.

Table 2 shows details of the comparison between the proposed VBME model vs. the BME model. The smallest difference between the beta coefficients was observed in \( \beta_1 \) for data set WheatYield1 (1%), while the largest difference was observed in \( \beta_3 \) for data set...
Wheat_Yield2 (25.3%); the estimates of the BME model were lower. With regard to the variance components, Table 2 shows that the smallest difference was observed in $\sigma^2_i$ for Wheat_Yield3 (3.8%), while the largest difference was also in $\sigma^2_i$ for data set Maize_PH, where the estimated variance component of the VBME was 5.68 times higher than the estimated variance component of the BME model.

In terms of the Pearson correlation, in seven out of eight data sets, the best correlations were observed in the BME model; the minimum difference (1.6%) was observed in data set Wheat_Yield3, while the maximum difference (15.4%) was observed in Wheat_PHT. Finally, in terms of implementation time (in minutes), in seven out of eight data sets the VBME model was faster than the BME model, with a minimum difference of 16.9%, and a maximum difference of 98.2%. The computer used to implement both models for each of the eight data sets (reported in Table 1 and Table 2) was a Windows operating system with processor Intel(R) Core(TM) i7-4610M CPU @ 3.0 GHz under a 64-bit operating system.

Comparing the prediction ability of models BME and VBME

Table 3 presents prediction accuracies for both models BME and VBME.

Table 2 Relative comparison using the full data sets between the BME and VBME calculated as parameter estimate under BME minus parameter estimate under VBME divided by parameter estimate under BME

| Data Set   | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\sigma^2_1$ | $\sigma^2_2$ | $\sigma^2_3$ | Corr | Time |
|------------|-----------|-----------|-----------|-----------|--------------|--------------|--------------|------|------|
| Maize_GY   | 0.030     | -0.022    | 0.009     | —         | 0.302        | 0.847        | -0.346       | 0.122| 0.885|
| Maize_ASI  | -0.012    | 0.022     | -0.064    | —         | 0.372        | 0.781        | -0.152       | 0.087| 0.890|
| Maize_PH   | -0.003    | -0.015    | 0.004     | —         | -0.568       | -5.289       | 0.330        | -0.153| 0.936|
| Wheat_DTHD | -0.028    | -0.015    | -0.075    | —         | 0.050        | 0.944        | -14.272      | 0.056| 0.982|
| Wheat_PHT  | 0.012     | -0.008    | 0.213     | —         | 0.233        | 0.886        | -41.243      | 0.154| 0.961|
| Wheat_Yield1| -0.001   | -0.010    | -0.068    | -0.029   | -0.235       | 0.222        | -0.087       | 0.022| 0.426|
| Wheat_Yield2| -0.010   | 0.047     | -0.253    | -0.110   | 0.153        | 0.488        | -0.115       | 0.038| -0.169|
| Wheat_Yield3| 0.160    | 0.148     | 0.144     | 0.055    | -0.038       | 0.221        | -0.079       | 0.016| 0.190|

The smallest and largest differences in $\beta$ coefficients, variance components, Pearson correlation and implementation time are in boldface.
We proposed and implemented the variational Bayes version of the BME model. We found that the $\beta$ coefficients were estimated in a similar way in both models (BME and VBME), but the variance components were considerably underestimated in the VBME model. However, in terms of prediction accuracy, the BME was a little better than the VBME model. The VBME model is a quick and a good approximation to the BME model. These results were expected, since, under MCMC methods, we estimate the parameters using the mean posterior from samples of the target distribution (posterior distribution), while, in the second approach for estimating the parameters, we also use the mean, but from the approximated target distribution, here obtained under the VB and called variational posterior distribution.

The above is documented in some papers that reported that the VBM divided by the APC under BME calculated as APC under BME minus APC

The smallest and largest differences in terms of prediction accuracy for both models are in boldface. In the bottom half of this table, 1 means that the best model was BME and 0 means that the best model was VBME in terms of prediction accuracy.

### Table 3 Prediction accuracy of the BME and VBME models resulting from the 20 trn-tst random partitions implemented

| Model   | Data set | Env1 | SE  | Env2 | SE  | Env3 | SE  | Env4 | SE  |
|---------|----------|------|-----|------|-----|------|-----|------|-----|
| BME     | Maize_GY | 0.319| 0.041| 0.378| 0.032| 0.362| 0.040| —    | —   |
|         | Maize_AS1| 0.466| 0.027| 0.380| 0.045| 0.285| 0.048| —    | —   |
|         | Maize_PH | 0.391| 0.024| 0.352| 0.029| 0.585| 0.024| —    | —   |
|         | Wheat_DTHD| 0.905| 0.011| 0.835| 0.012| 0.597| 0.008| —    | —   |
|         | Wheat_PTHT| 0.523| 0.020| 0.347| 0.027| 0.499| 0.026| —    | —   |
|         | Wheat_Yield1| 0.191| 0.013| 0.644| 0.013| 0.593| 0.007| 0.556| 0.000|
|         | Wheat_Yield2| 0.690| 0.015| 0.785| 0.010| 0.617| 0.021| 0.637| 0.020|
|         | Wheat_Yield3| 0.627| 0.030| 0.711| 0.019| 0.669| 0.024| 0.719| 0.022|
| VBME    | Maize_GY | 0.341| 0.025| 0.383| 0.027| 0.283| 0.021| —    | —   |
|         | Maize_AS1| 0.473| 0.022| 0.427| 0.027| 0.369| 0.025| —    | —   |
|         | Maize_PH | 0.313| 0.032| 0.380| 0.029| 0.389| 0.027| —    | —   |
|         | Wheat_DTHD| 0.920| 0.010| 0.860| 0.012| 0.908| 0.008| —    | —   |
|         | Wheat_PTHT| 0.425| 0.034| 0.439| 0.040| 0.398| 0.035| —    | —   |
|         | Wheat_Yield1| 0.455| 0.015| 0.527| 0.012| 0.434| 0.010| 0.262| 0.006|
|         | Wheat_Yield2| 0.661| 0.007| 0.706| 0.007| 0.557| 0.008| 0.598| 0.003|
|         | Wheat_Yield3| 0.432| 0.012| 0.490| 0.011| 0.480| 0.011| 0.514| 0.010|

### Table 4 Relative comparison of the prediction accuracies of the BME and VBME models calculated as APC under BME minus APC

| Data set | Env1  | Env2  | Env3  | Env4  |
|----------|-------|-------|-------|-------|
| Maize_GY | 0.069 | 0.015 | 0.217 | —     |
| Maize_AS1| 0.015 | 0.121 | 0.298 | —     |
| Maize_PH | 0.198 | 0.079 | 0.334 | —     |
| Wheat_DTHD| 0.017| 0.030 | 0.012 | —     |
| Wheat_PTHT| 0.187| 0.266 | 0.203 | —     |
| Wheat_Yield1| 1.381| 0.182 | 0.268 | 0.529|
| Wheat_Yield2| 0.042| 0.101| 0.097| 0.061|
| Wheat_Yield3| 0.311| 0.311| 0.282| 0.285|
| Maize_GY | 0     | 0     | 1     | —     |
| Maize_AS1| 0     | 0     | 0     | —     |
| Maize_PH | 1     | 0     | 1     | —     |
| Wheat_DTHD| 0    | 0     | 0     | —     |
| Wheat_PTHT| 1    | 0     | 1     | —     |
| Wheat_Yield1| 1    | 1     | 1     | 1     |
| Wheat_Yield2| 1    | 1     | 1     | 1     |
| Wheat_Yield3| 1    | 1     | 1     | 1     |

### Discussion

We proposed and implemented the variational Bayes version of the BME model. We found that the $\beta$ coefficients were estimated in a similar way in both models (BME and VBME), but the variance components were considerably underestimated in the VBME model. However, in terms of prediction accuracy, the BME was a little better than the VBME model. The VBME model is a quick and a good approximation to the BME model. These results were expected, since, under MCMC methods, we estimate the parameters using the mean posterior from samples of the target distribution (posterior distribution), while, in the second approach for estimating the parameters, we also use the mean, but from the approximated target distribution, here obtained under the VB and called variational posterior distribution. The above is documented in some papers that reported that the BVM showed less accuracy than its MCMC counterpart. This is due to the fact that variational Bayes methods approximate the target posterior distribution by a factorized distribution that has minimum KL distance to the target posterior. The factorized distribution we used was the mean field variational family restriction, which makes strong independent assumptions to achieve great flexibility and scalable optimization at the price of low accuracy and underestimation of posterior variances (Blei et al. 2016). In contrast, MCMC methods such as the BME model have a high computational cost with the advantage that they generate samples directly from the target posterior distribution that are guaranteed to converge asymptotically to the true posterior distribution. The new proposed VBME model for univariate phenotypes with G×E interaction is a step ahead of the models proposed until now, in the context of genomic-enabled prediction models proposed by other authors and described in the Introduction. In this paper, we propose a genomic VBM model using a half-t priors distribution, and give clear theoretical details of the derivation of the full conditional and posterior distributions, and account for G×E interaction.

Advances in genetic engineering allow collecting and storing millions of markers (SNPs, etc.) from plant and animal genomes, and many breeding programs are interested in improving many correlated traits simultaneously. For these reasons, in genomic selection there is a great need to use genome-enabled prediction models that take into account all available markers, traits, and environments under study to improve the prediction of genotypes early in time for multiple correlated traits. However, due to the large amount of input information for Bayesian methods (such as those used in the BME model), these statistical models and methods might become intractable, or impose

---

**Table 3** Prediction accuracy of the BME and VBME models resulting from the 20 trn-tst random partitions implemented

**Table 4** Relative comparison of the prediction accuracies of the BME and VBME models calculated as APC under BME minus APC

---

The smallest and largest differences in terms of prediction accuracy for both models are in boldface. In the bottom half of this table, 1 means that the best model was BME and 0 means that the best model was VBME in terms of prediction accuracy.
a very high computational cost in terms of time. For these reasons, and because it offers a significant improvement in computational speed at the cost of a small loss in prediction accuracy, we believe that the proposed VBME model is an attractive alternative to the BME model.

Advantages and disadvantages of the variational model (VBME)

As already mentioned, the variational model (VBME) was proposed due to the need to improve the speed of the BME. For this reason, we propose the iterative algorithm given in the section Algorithm for implementing the VBME, which approximates the BME model in a reasonable way. Results show that the VBME method: (1) is sometimes more flexible than its Bayesian counterpart; (2) offers a reasonable approximation to the posterior distributions of all model parameters; (3) most of the time is much less demanding in computational time than the BME built using MCMC techniques; and (4) maintains most of the advantages of Bayesian inference. However, the VBME has some limitations: (1) the approximation cannot become more accurate by increasing computational time; (2) part of the VBME’s advantage in computing time is because the model is simple; however, this algorithm needs to be adapted to cases that are different from the one used here; (3) the VBME, like all variational Bayes methods, does not guarantee that it will converge to a global maximum; (4) the convergence time depends strongly on the level of tolerance desired and the nature of the data; and (5) it has been documented, and was observed here, that variational models underestimate some covariance parameters. However, the time reduction depends strongly on the tolerance level: the lower the value, the longer the convergence time. For this reason, although our results show a significant reduction in time, this gain can be reduced substantially if the chosen tolerance level is smaller. Our empirical results when implementing the VBME in eight real data sets show that a tolerance level of 0.00001 is enough, because when we increased it to 1E−8, no significant gain in prediction accuracy or accuracy in parameter estimates was observed. Also, since in developing our VBME, we used the mean field variational family restriction that approximates \( q(\theta) \) by the product of \( q \) latent independent variables to gain tractability, this may underestimate the variability of parameter estimates when there is a considerable degree of dependence. Another important disadvantage of the VBME is that deriving the set of equations (variational posterior and lower bound) used to iteratively update the parameters often requires a large amount of algebra compared with deriving the comparable Gibbs sampling equations. This is evident in Appendix A for our proposed VBME model, which, despite being a simple model, required a lot of algebra and notation to obtain the variational posterior distribution and the corresponding lower bound. This amount of algebra needs to be done each time for every specific model.

To take advantage of the proposed model, we provide some guidance on when to use the VBME or the BME. The BME (an MCMC method) tends to be more computationally intense than the VBME (a variational method) but guarantees that exact samples from the target density are (asymptotically) produced (Robert 2004). While the VBME does not guarantee exact samples, only approximate ones (that is, it can only find a density close to the target), most of the time it is faster than MCMC methods. Since the VBME is based on optimization, it can more easily take advantages of stochastic optimization and distributed optimization. For these reasons, the VBME is expected to be better suited to larger data sets than the BME and for scenarios where we want to quickly explore many models. Also, Blei et al. (2016) point out that the VBME is preferred when the posterior of a mixture model admits multiple modes, since MCMC techniques are not an option for this scenario, even with small data sets.

Finally, we believe that the VBME method, as proposed here, is a useful addition to the existing arsenal of genome-enabled prediction models; it is also an alternative to the BME model, since it considerably reduces implementation time. Additionally, this research opens a new branch of methodologies for fitting genomic models to genomic data that need novel methods to take advantage of the huge amount of data available nowadays. For these reasons, we agree with Ormerod and Wand (2010), who pointed out that the usefulness of variational approximations increases as the size of the problem increases, and Monte Carlo methods such as MCMC start to become untenable. We also believe that our research contributes significantly to researchers’ understanding of the BME, because, in developing the proposed VBME, we provided all the details of the derivation of the variational posterior distributions needed to derive the q-specific evidence lower bound for building the algorithm, and the R scripts for implementing the proposed VBME model. This information gives the scientist the basic elements needed to develop new models under the variational Bayes approach. It is also important to point out that we provide all the details for deriving the full conditionals needed for implementing the BME model since the proposed model uses Half-t priors that produce non-informative variance components (Huang and Wand 2013).

Conclusions

In this paper, we propose a new alternative to the BME via variational Bayes, which we called the VBME model. The proposed method underestimated the variance components and similar estimates of the beta coefficients were found; in terms of prediction accuracy, the VBME had lower prediction accuracies than the BME. However, according to our results, the loss in prediction accuracy is compensated for by the significant gain in implementation time, since the VBME is much faster than the BME. However, it is important to point out that if a very low tolerance value is selected, the gain in terms of implementation speed will vanish. For this reason, to take advantage of the proposed VBME, a reasonable tolerance value should be chosen that guarantees the quality of the parameter estimates and predictions, as well as a significant reduction in implementation time.

We also believe that the proposed model should be tested on many real data sets to obtain more empirical evidence of its performance; this evidence would allow us to extend the proposed variational Bayes model to take into account nonnormal responses (binary, Poisson, negative binomial, etc.), for multiple traits and multiple-environment models (Montesinos-López et al. 2016), which would enable scientists to make a more precise selection of candidate genotypes at early crop stages. Also, models that avoid the product density restriction can be proposed. Finally, the proposed VBME model is a useful addition to the existing arsenal of genome-enabled prediction models, and an excellent alternative to the BME model, since it produces competitive predictions at a lower computational cost.

LITERATURE CITED

Arakawa, A., M. Taniguchi, T. Hayashi, and S. Mikawa, 2016 Variational Bayesian method of estimating variance components. Anim. Sci. J. 87: 863–872.
Attias, H., 2000 A variational Bayesian framework for graphical models. Adv. Neural Inf. Process. Syst. 12(1–2): 209–215.
Bishop, C. M., 1999 Variational principal components, pp. 509–514 in Proceedings Ninth International Conference on Artificial Neural Networks, ICANN’99. Institute of Electrical and Electronics Engineers, New York.
Blei, D. M., A. Kucukelbir, and J. D. McAuliffe, 2016 Variational inference: a review for statisticians. arXiv. Available at: https://arxiv.org/abs/1601.00670.
Carbonetto, P., and M. Stephens, 2012 Scalable variational inference for Bayesian variable selection in regression, and its accuracy in genetic association studies. Bayesian Anal. 7(1): 73–108.

Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño et al., 2010 Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186(2): 713–724.

Crossa, J., Y. Beyene, S. Kassa, P. Pérez, J. M. Hickey et al., 2013 Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 (Bethesda) 3(11): 1903–1926.

Cuevas, J., J. Crossa, O. A. Montesinos-Lopez, J. Burgueno, P. Perez-Rodriguez et al., 2016a Bayesian genomic prediction with genotype × environment interaction Kernel models. G3 (Bethesda) 7: 41–53.

Cuevas, J., J. Crossa, V. Soberanis, S. Pérez-Elizalde, P. Pérez-Rodríguez et al., 2016b Genomic prediction of genotype × environment interaction Kernel regression models. Plant Genome 9(3): 1–20.

Ghahramani, Z., and M. J. Beal, 2001 Graphical models and variational methods. In Advanced Mean Field Method—Theory and Practice, edited by Saad, D., and M. Opper. MIT Press, Cambridge, MA.

Hayashi, T., and H. Iwata, 2013 A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinformatics 14(1): 1–14.

Huang, A., and M. P. Wand, 2013 Simple marginally noninformative prior distributions for covariance matrices. Bayesian Anal. 8(2): 439–452.

Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt et al., 2014 A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. 127(3): 595–607.

Li, Z., and M. J. Sillanpää, 2012 Estimation of quantitative trait locus effects with epistasis by variational Bayes algorithms. Genetics 190: 231–249.

Li, Z., and M. J. Sillanpää, 2013 A Bayesian nonparametric approach for mapping dynamic quantitative traits. Genetics 194: 997–1016.

Logsdon, B. A., G. E. Hoffman, and J. G. Mezey, 2010 A variational Bayes algorithm for fast and accurate multiple locus genome-wide association analysis. BMC Bioinformatics 11(1): 58.

Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland et al., 2015 Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 (Bethesda) 5(4): 569–582.

MacKay, D., 1995 Developments in Probabilistic Modelling with Neural Networks – Ensemble Learning. Neural Networks: Artificial Intelligence and Industrial Applications. Springer, London. pp. 14–15.

Montesinos-López, O.A., A. Montesinos-López, J. Crossa, F. H. Toledo, O. Pérez-Hernández et al., 2016 A genomic Bayesian multi-trait and multi-environment model. G3 (Bethesda) 6(9): 2725–2744.

Ormerod, J. T., and M. P. Wand, 2010 Explaining variational approximations. Am. Stat. 64(2): 140–153.

Ormerod, J. T., and M. P. Wand, 2012 Gaussian variational approximate inference for generalized linear mixed models. J. Comput. Graph. Stat. 21(1): 2–17.

R Core Team, 2016 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.

Robert, C. P., 2004 Monte Carlo Methods. John Wiley & Sons, Ltd., NY.

Rutkoski, J., J. Poland, S. Mondal, E. Autrique, J. Crossa et al., 2016 Predictor traits from high-throughput phenotyping improve accuracy of pedigree and genomic selection for yield in wheat. G3 (Bethesda) 6(9): 2799–2808.

Teschendorff, A. E., Y. Wang, N. L. Barbosa-Morais, J. D. Brenton, and C. Caldas, 2005 A variational Bayesian mixture modelling framework for cluster analysis of gene-expression data. Bioinformatics 21(13): 3025–3033.

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11): 4414–4423.

Yamamoto, E., H. Matsuunga, A. Onogi, A. Ohyama, K. Miyatake et al., 2017 Efficiency of genomic selection for breeding population design and phenotype prediction in tomato. Heredity 118(2): 202–209.

Communicating editor: D. J. de Koning
APPENDIX A

DERIVATION OF FULL CONDITIONAL AND VARIATIONAL POSTERIOR DISTRIBUTIONS FOR THE BME AND VBME MODELS

Full conditional for $\beta$

$$P(\beta \mid \text{ELSE}) \propto P(Y \mid \beta, b_1, b_2, \sigma_b^2)P(\beta \mid \sigma_b^2)$$

$$\propto \exp \left\{ -\frac{1}{2} \left[ -\frac{1}{2} \left( \beta - \beta_{\beta} \right)^T \Sigma_\beta^{-1} \left( \beta - \beta_{\beta} \right) \right] \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left( \beta - \beta_{\beta} \right)^T \Sigma_\beta^{-1} \left( \beta - \beta_{\beta} \right) \right\} \propto N(\beta, \Sigma_\beta)$$

where:

$$\Sigma_\beta = \Sigma_b^{-1} \sigma_b^2 + X^T (I_n \sigma_b^2)^{-1} X$$

$$\tilde{\beta} = \Sigma_\beta \left( \Sigma_b^{-1} \beta \sigma_b^2 + X^T (I_n \sigma_b^2)^{-1} Y - X^T (I_n \sigma_b^2)^{-1} Z_b - X^T (I_n \sigma_b^2)^{-1} b_2 \right)$$

Variational posterior for $\beta$

$$q(\beta \mid \text{ELSE}) = \exp \left\{ E_{-\beta} \log P(\beta \mid \text{ELSE}) \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left( E_{-\beta} \left[ \beta - \mu_{q(\beta)} \right]^T \Sigma_{q(\beta)}^{-1} \left( \beta - \mu_{q(\beta)} \right) \right) \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left( \beta - \mu_{q(\beta)} \right)^T \Sigma_{q(\beta)}^{-1} \left( \beta - \mu_{q(\beta)} \right) \right\} \propto N(\mu_{q(\beta)}, \Sigma_{q(\beta)}) (A.11)$$

where

$$\Sigma_{q(\beta)} = \Sigma_b^{-1} \mu_q(1/\sigma_b^2) + X^T (I_n \mu_q(\sigma_b^2)) X^{-1} \cdot \mu_{q(\beta)} = \Sigma_q(\beta) \left( \Sigma_b^{-1} \beta \mu_q(1/\sigma_b^2) + X^T [I_n \mu_q(\sigma_b^2)] [Y - \sum_{i=1}^2 Z_b \mu_q(\sigma_b^2)] \right)$$

Full conditional for $\sigma_b^2$

$$P(\sigma_b^2 \mid \text{ELSE}) \propto P(\beta \mid \sigma_b^2)P(\sigma_b^2 \mid a_\beta)$$

$$\propto |\Sigma_\beta|^{-\frac{1}{2}} \left( \sigma_b^2 \right)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left( \beta^T \Sigma_\beta^{-1} \beta \right) \right\} \left( \sigma_b^2 \right)^{-\frac{1}{2}} \exp \left\{ -\frac{\nu_\beta}{\sigma_b^2} \right\} \left( \sigma_b^2 \right)^{\nu_\beta - 1} \exp \left\{ -\frac{\nu_\beta}{\sigma_b^2} \right\}$$

$$\propto \sigma_b^{2-1} \exp \left\{ -\frac{\nu_\beta}{\sigma_b^2} \right\} \left( \beta - \beta_{\beta} \right)^T \Sigma_\beta^{-1} \left( \beta - \beta_{\beta} \right) + \nu_\beta a_\beta$$

$$\propto IG \left( \frac{\nu_\beta + 1}{2}, \frac{\beta - \beta_{\beta}}{\Sigma_\beta^{-1} \beta - \beta_{\beta}} + \nu_\beta a_\beta \right)$$
Variational posterior for $\sigma_\beta^2$

$$q(\sigma_\beta^2 | ELSE) \propto \exp \left\{ E_{-\sigma_\beta^2} \left[ \log P(\sigma_\beta^2 | ELSE) \right] \right\}$$

$$\propto \frac{1}{\sigma_\beta^2 \gamma_\beta a} \exp \left\{ -\frac{1}{\sigma_\beta^2} \left[ E_{-\sigma_\beta^2} \left( \frac{1}{A_\beta^2} \right) \right] \right\}$$

Full conditional for $a_\beta$

$$P(a_\beta | ELSE) \propto P(\sigma_\beta^2 | a_\beta) P(a_\beta)$$

$$\propto \left( \frac{a_\beta}{\gamma_\beta a} \right)^{a_\beta - 1} \exp \left\{ -\frac{a_\beta}{\gamma_\beta a} \left( a_\beta + \frac{1}{A_\beta^2} \right) \right\} \exp \left\{ -\frac{1}{A_\beta^2} \right\}$$

Variational posterior for $a_\beta$

$$q(a_\beta | ELSE) \propto \exp \left\{ E_{-a_\beta} \left[ \log P(a_\beta | ELSE) \right] \right\}$$

$$\propto \frac{1}{a_\beta^{a_\beta + 1}} \exp \left\{ -\frac{a_\beta}{\gamma_\beta a} \left( a_\beta + \frac{1}{A_\beta^2} \right) \right\}$$

Full conditional for $b_1$

$$P(b_1 | ELSE) \propto P(Y | \beta, b_1, b_2, \sigma_\varepsilon^2) P \left( b_1 | \sigma_\beta^2 \right)$$

$$\propto \exp \left\{ -\frac{1}{2} (Y - X\beta - Z_1 b_1 - Z_2 b_2)^T (I_n \sigma_\varepsilon^2)^{-1} (Y - X\beta - Z_1 b_1 - Z_2 b_2) \right\} \times \exp \left\{ -\frac{1}{2} \left( b_1^T G^{-1} \sigma_\varepsilon^{-2} b_1 \right) \right\}$$
**Variational posterior for \( b_1 \)**

Defining \( \eta^1 = X\beta + Z_2 b_2 \), the variational posterior distribution of \( b_1 \) is given as

\[
q(b_1 \mid \text{ELSE}) \propto \exp \left\{ -E_{-b_1} \left[ \log P(b_1 \mid \text{ELSE}) \right] \right\} \\
\propto \exp \left\{ -\frac{1}{2} \left( b_1 - \mu_{q(b_1)} \right)^T \Sigma_{q(b_1)}^{-1} \left( b_1 - \mu_{q(b_1)} \right) - \frac{1}{2} \Sigma_{q(b_1)}^{-1} \left( Z_i^T (I_n \sigma_{\eta^1}^2)^{-1} Y - Z_i^T (I_n \sigma_{\eta^1}^2)^{-1} \eta^1 \right)^T \right\} \left( b_1 \right)
\]

where

\[
\Sigma_{q(b_1)} = (G^{-1} \mu_{q(\sigma_{\eta^1}^2)} + Z_i^T [I_n \mu_{q(\sigma_{\eta^1}^2)}] Z_i)^{-1} \text{ and } \mu_{q(b_1)} = \Sigma_{q(b_1)} \left( Z_i^T [I_n \mu_{q(\sigma_{\eta^1}^2)}] \right) \left( Y - X \mu_{q(\beta)} - Z_2 \mu_{q(b_2)} \right)
\]

**Full conditional for \( \sigma_{r_1}^2 \)**

\[
P(\sigma_{r_1}^2 \mid \text{ELSE}) \propto P(b_1 \mid \sigma_{r_1}^2) P(\sigma_{r_1}^2 \mid a_{b_1}) \\
\propto \left| G \right|^{-\frac{1}{2}} \left( \sigma_{r_1}^2 \right)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left( b_1^T G^{-1} \sigma_{r_1}^{-2} b_1 \right) \right\} \left( \sigma_{r_1}^2 \right)^{-1} \left( \frac{\nu_b}{\nu_a} \right)^{\frac{\nu_a}{2}} \exp \left\{ -\frac{\nu_b}{\nu_a} \frac{1}{\sigma_{r_1}^2} \right\} \\
\propto \left( \sigma_{r_1}^2 \right)^{-\frac{\nu_a}{2} - 1} \exp \left\{ -\frac{b_1^T G^{-1} b_1}{2} + \frac{\nu_b}{\nu_a} \frac{1}{\sigma_{r_1}^2} \right\} \propto IG \left( \nu_b + \frac{1}{2}, \frac{b_1^T G^{-1} b_1}{2} + \frac{\nu_b}{\nu_a} \right)
\]

**Variational posterior for \( \sigma_{r_1}^2 \)**

\[
q(\sigma_{r_1}^2 \mid \text{ELSE}) \propto \exp \left\{ E_{-\sigma_{r_1}^2} \left[ \log P(\sigma_{r_1}^2 \mid \text{ELSE}) \right] \right\} \\
\propto \frac{1}{(\sigma_{r_1}^2)^{\frac{\nu_a}{2} + 1}} \exp \left\{ \frac{1}{2} \left( E_{-\sigma_{r_1}^2} \left( \mu_{q(\sigma_{r_1}^2)}^T G^{-1} \left( \sigma_{r_1}^2 \right)^{-1} \mu_{q(\sigma_{r_1}^2)} + \text{tr} \left( G^{-1} \Sigma_{q(\sigma_{r_1}^2)} \right) \right) + \nu_b \cdot \mu_{q(1/\sigma_{r_1}^2)} \right) \right\} \\
\propto \left( \sigma_{r_1}^2 \right)^{-\frac{\nu_a}{2} - 1} \exp \left\{ -\frac{1}{2} \mu_{q(\sigma_{r_1}^2)}^T G^{-1} \mu_{q(\sigma_{r_1}^2)} + \nu_b \cdot \mu_{q(1/\sigma_{r_1}^2)} \right\} \propto IG \left( \kappa_{q(\sigma_{r_1}^2)} = \frac{\nu_b + 1}{2}, B_{q(\sigma_{r_1}^2)} = \frac{\nu_b}{2} \cdot \mu_{q(\sigma_{r_1}^2)} + \nu_b \cdot \mu_{q(1/\sigma_{r_1}^2)} \right)
\]

**Full conditional for \( a_{b_1} \)**

\[
P(a_{b_1} \mid \text{ELSE}) \propto P(\sigma_{r_1}^2 \mid a_{b_1}) P(a_{b_1}) \\
\propto (\sigma_{r_1}^2)^{-\frac{\nu_a}{2} - 1} \left( \frac{\nu_b}{\nu_a} \right)^{\frac{\nu_a}{2}} \exp \left\{ -\frac{\nu_b}{\nu_a} \frac{1}{\sigma_{r_1}^2} \right\} \left( \sigma_{r_1}^2 \right)^{-\frac{1}{2} - 1} \left( \frac{1}{A_{b_1}^2} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{A_{b_1}^2} \frac{1}{a_{b_1}} \right\} \\
\propto (a_{b_1})^{-\frac{\nu_a}{2} - 1} \exp \left\{ -\left( \frac{\nu_b}{\nu_a} \frac{1}{\sigma_{r_1}^2} + \frac{1}{A_{b_1}^2} \right) \frac{1}{a_{b_1}} \right\} \\
\propto IG \left( \frac{\nu_b}{\nu_a} + \frac{1}{2}, \frac{1}{A_{b_1}^2} \right)
\]
Variational posterior for $a_{b_{1}}$

$q(a_{b_{1}} | \text{ELSE}) \propto \exp\left\{ E_{-a_{b_{1}}} \left[ \log P(a_{b_{1}} | \text{ELSE}) \right] \right\} \\
\propto \frac{1}{a_{b_{1}}} \exp\left( -\frac{1}{a_{b_{1}}} \left( E_{-a_{b_{1}}} \left( \frac{1}{A_{b_{1}}^{2}} + \frac{\nu_{b_{1}}}{\sigma_{b_{1}}^{2}} \right) \right) \right) \\
\propto IG(k_{q(a_{b_{1}})} = \frac{\nu_{b_{1}}+1}{2}, B_{q(a_{b_{1}})} = 1/\left( A_{b_{1}}^{2} + \nu_{b_{1}}/\sigma_{b_{1}}^{2} \right) )$

Full conditional for $b_{2}$

$P(b_{2} | \text{ELSE}) \propto P(Y | \beta, b_{1}, b_{2}, \sigma_{b_{2}}^{2}) P(b_{2} | \sigma_{b_{2}}^{2})$ \\
$\propto \exp\left\{ -\frac{1}{2} \left( Y - X \beta - Z_{1} b_{1} - Z_{2} b_{2} \right)^{T} \left( I_{n} \sigma_{b_{2}}^{2} \right)^{-1} \left( Y - X \beta - Z_{1} b_{1} - Z_{2} b_{2} \right) \right\} \\
\propto \exp\left\{ -\frac{1}{2} \left( b_{2}^{T} G_{b_{2}}^{-1} \sigma_{b_{2}}^{2} b_{2} \right) \right\} \\
\propto \exp\left\{ -\frac{1}{2} \left( b_{2} - \tilde{b}_{2} \right)^{T} \Sigma_{b_{2}}^{-1} \left( b_{2} - \tilde{b}_{2} \right) \right\} \propto N\left( \tilde{b}_{2}, \Sigma_{b_{2}} \right)$

where:

$\Sigma_{b_{2}} = \left( G_{2}^{-1} \sigma_{2}^{-2} + Z_{2}^{T} I_{n} \sigma_{2}^{-2} Z_{2} \right)^{-1} ; \tilde{b}_{2} = \Sigma_{b_{2}} \left( Z_{2}^{T} I_{n} \sigma_{2}^{-2} - X \beta - Z_{2}^{T} I_{n} \sigma_{2}^{-2} - X \beta - Z_{2}^{T} I_{n} \sigma_{2}^{-2} - Z_{1} b_{1} \right)$

Variational posterior for $b_{2}$

Defining $\eta^{2} = X \beta + Z_{1} b_{1}$, the variational posterior distribution of $b_{2}$ is given as

$q(b_{2} | \text{ELSE}) \propto \exp\left\{ E_{-b_{2}} \left[ \log P(b_{2} | \text{ELSE}) \right] \right\} \\
\propto \exp\left\{ -\frac{1}{2} \left( b_{2} - \mu_{q(b_{2})} \right)^{T} \Sigma_{q(b_{2})}^{-1} \left( b_{2} - \mu_{q(b_{2})} \right) \right\} \propto N\left( \mu_{q(b_{2})}, \Sigma_{q(b_{2})} \right)$

where

$\Sigma_{q(b_{2})} = \left( G_{2}^{-1} \mu_{q(\sigma_{b_{2}}^{2})} + Z_{2}^{T} I_{n} \mu_{q(\sigma_{b_{2}}^{2})} Z_{2} \right)^{-1} \text{ and } \mu_{q(b_{2})} = \Sigma_{q(b_{2})} \left( Z_{2}^{T} I_{n} \mu_{q(\sigma_{b_{2}}^{2})} \right) \left[ Y - X \mu_{q(\beta)} - Z_{2} \mu_{q(b_{1})} \right]$}

Full conditional for $\sigma_{b_{2}}^{2}$

$P(\sigma_{b_{2}}^{2} | \text{ELSE}) \propto P(b_{2} | \sigma_{b_{2}}^{2}) P(\sigma_{b_{2}}^{2}) P(a_{b_{1}})$ \\
$\propto | G_{2} | \sigma_{b_{2}}^{2 N_{2}} \exp\left( -\frac{1}{2} \left( b_{2}^{T} G_{b_{2}}^{-1} \sigma_{b_{2}}^{2} b_{2} \right) \right), (\sigma_{b_{2}}^{2})^{-\frac{N_{2}}{2}} \exp\left( -\frac{\nu_{b_{1}}}{2} \frac{1}{\sigma_{b_{2}}^{2}} \right)$ \\
$\propto (\sigma_{b_{2}}^{2})^{-\frac{N_{2} + \nu_{b_{1}}}{2}} \exp\left( -\frac{\nu_{b_{1}}}{2} \frac{1}{\sigma_{b_{2}}^{2}} \right)$ \\
$\propto IG\left( \frac{\nu_{b_{1}} + \nu_{1}}{2}, \frac{\nu_{b_{1}}}{2} \right)$
Variational posterior for $\sigma^2_2$

$$q(\sigma^2_2 \mid \text{ELSE}) \propto \exp \left\{ E_{-\sigma^2_2} \left[ \log P(\sigma^2_2 \mid \text{ELSE}) \right] \right\}$$

$$\propto \frac{1}{(\sigma^2_2)^{\nu_2/2+1}} \exp \left( -\frac{1}{\sigma^2_2} \left\{ E_{-\sigma^2_2} \left( \frac{b_2^T G_2^{-1} b_2}{2} + \nu_{b_2} \right) \right\} \right)$$

$$\propto \frac{1}{(\sigma^2_2)^{\nu_2/2+1}} \exp \left( -\frac{1}{\sigma^2_2} \left\{ \mu_{q(b_2)}^T G_2^{-1} \mu_{q(b_2)} + \text{tr} \left( G_2^{-1} \Sigma_{q(b_2)} \right) + \nu_{b_2} \mu_{q(1/\sigma_2)} \right\} \right)$$

$$\propto IG \left( k_q(\sigma^2_2) = \frac{\nu_{b_2} + n}{2}, B_q(\sigma^2_2) = \frac{\mu_{q(b_2)}^T G_2^{-1} \mu_{q(b_2)} + \text{tr} \left( G_2^{-1} \Sigma_{q(b_2)} \right) + \nu_{b_2} \mu_{q(1/\sigma_2)}}{2} + \nu_{b_2} \mu_{q(1/\sigma_2)} \right)$$

Full conditional for $a_{b_2}$

$$P(a_{b_2} \mid \text{ELSE}) \propto P(\sigma^2_2 \mid a_{b_2}) P(a_{b_2})$$

$$\propto (\sigma^2_2)^{-\nu_2/2-1} \exp \left( -\frac{\nu_{b_2} + 1}{\sigma^2_2} \left[ \frac{1}{A_{b_2}^2} \right] \right)$$

$$\propto IG \left( k_q(a_{b_2}) = \frac{\nu_{b_2} + 1}{2}, B_q(a_{b_2}) = \frac{1}{A_{b_2}^2} + \nu_{b_2} \mu_{q(1/\sigma_2)} \right)$$

Variational posterior for $a_{b_2}$

$$q(\sigma^2_2 \mid \text{ELSE}) \propto \exp \left\{ E_{-a_{b_2}} \left[ \log P(a_{b_2} \mid \text{ELSE}) \right] \right\}$$

$$\propto \frac{1}{(a_{b_2})^{\nu_2/2+1}} \exp \left( -\frac{1}{a_{b_2}} \left\{ E_{-a_{b_2}} \left( \frac{1}{A_{b_2}^2} + \frac{\nu_{b_2}}{\sigma^2_2} \right) \right\} \right)$$

$$\propto IG \left( k_q(a_{b_2}) = \frac{\nu_{b_2} + 1}{2}, B_q(a_{b_2}) = \frac{1}{A_{b_2}^2} + \nu_{b_2} \mu_{q(1/\sigma_2)} \right)$$

Full conditional for $\sigma^2_2$

$$P(\sigma^2_2 \mid \text{ELSE}) \propto P(Y \mid \beta, b_1, b_2, \sigma^2_2) P(\sigma^2_2 \mid a_{c})$$

$$\propto I_n^{-1/2} (\sigma^2_2)^{-\nu_2/2} \exp \left\{ -\frac{1}{2} (Y - X\beta - Z_1 b_1 - Z_2 b_2)^T (I_n \sigma^2_2)^{-1} (Y - X\beta - Z_1 b_1 - Z_2 b_2) \right\} \left( \sigma^2_2 \right)^{-\nu_2/2-1} \left( \frac{\nu_c}{a_c} \right)^{\nu_c/2} \exp \left\{ -\frac{\nu_c}{a_c} \frac{1}{\sigma^2_2} \right\}$$

$$\propto (\sigma^2_2)^{-\nu_2/2-1} \exp \left\{ -\frac{1}{2} (Y - X\beta - Z_1 b_1 - Z_2 b_2)^T (Y - X\beta - Z_1 b_1 - Z_2 b_2) + \frac{\nu_c}{a_c} \left( \frac{1}{\sigma^2_2} \right) \right\}$$

$$\propto IG \left( \frac{\nu_c + n}{2}, \frac{1}{2} (Y - X\beta - Z_1 b_1 - Z_2 b_2)^T (Y - X\beta - Z_1 b_1 - Z_2 b_2) + \frac{\nu_c}{a_c} \right)$$
Variational posterior for $\alpha_c^2$

$$q(\alpha_c^2 \mid \text{ELSE}) \propto \exp\left\{ E_{-\alpha_c}[\log P(\alpha_c^2 \mid \text{ELSE})]\right\}$$

$$\propto \frac{1}{(\alpha_c^2)^{-\nu_c/2}} \exp\left\{-\frac{\nu_c}{\alpha_c^2} \left(\frac{\nu_c}{\alpha_c^2} + \frac{1}{\nu_c}\right) - \frac{1}{\alpha_c^2} \right\}$$

Full conditional for $\alpha_c$

$$P(\alpha_c \mid \text{ELSE}) \propto P(\alpha_c^2 \mid \text{ELSE}) P(\alpha_c)$$

$$\propto (\alpha_c^2)^{-\nu_c/2-1} \left(\frac{\nu_c}{\alpha_c^2} + \frac{1}{\nu_c}\right) - \frac{1}{\alpha_c^2} \right\}$$

Variational posterior for $\alpha_c$

$$P(\alpha_c \mid \text{ELSE}) \propto \exp\{E_{-\alpha_c}[\log P(\alpha_c \mid \text{ELSE})]\}$$

$$\propto \frac{1}{(\alpha_c)^{\nu_c/2+1}} \exp\left\{-\frac{\nu_c}{\alpha_c} \left(\frac{\nu_c}{\alpha_c} + \frac{1}{\nu_c}\right) - \frac{1}{\alpha_c^2} \right\}$$

APPENDIX B

EXPRESSION OF THE LOWER BOUND OF THE LOG LIKELIHOOD FOR THE VARIATIONAL BAYES MODEL

$$\log P(y \mid q^*) = \int q(\theta) \log \left(\frac{P(y, \theta)}{q^*(\theta)}\right) d\theta = E_{q^*}[\log P(Y, \theta) - \log q^*(\theta)]$$

$$= E_{q^*}[\log P(Y \mid \beta, b_1, b_2, \sigma^2) + E_{q^*}[\log P(\beta \mid \sigma^2) - \log q^*(\beta)]$$

$$+ E_{q^*}[\log P(\sigma^2 \mid a_\sigma) - \log q^*(\sigma^2)] + E_{q^*}[\log P(a_\sigma) - \log q^*(a_\sigma)]$$

$$+ E_{q^*}[\log P(b_1 \mid \sigma^2) - \log q^*(b_1)] + E_{q^*}[\log P(b_2 \mid \sigma^2) - \log q^*(b_2)]$$

$$+ E_{q^*}[\log P(\sigma^2_1 \mid a_{\sigma_1}) - \log q^*(\sigma^2_1)] + E_{q^*}[\log P(a_{\sigma_1}) - \log q^*(a_{\sigma_1})]$$

$$+ E_{q^*}[\log P(\sigma^2_2 \mid a_{\sigma_2}) - \log q^*(\sigma^2_2)] + E_{q^*}[\log P(a_{\sigma_2}) - \log q^*(a_{\sigma_2})]$$

$$+ E_{q^*}[\log P(a_c^2 \mid \alpha_c^2) - \log q^*(a_c^2)] + E_{q^*}[\log P(\alpha_c^2) - \log q^*(\alpha_c^2)]$$

(B1)
The first term in (B1) is
\[
E_q[ \log P(Y \mid \beta, b_1, b_2, \sigma^2)] = -\frac{n}{2} \log(2\pi) - \frac{n}{2} E_q[\log(\sigma^2)]
\]
\[
- \frac{1}{2} \left[ (Y - X_\mu \beta)^T (I_n \mu_\eta (\sigma^2)) \right]^T (I_n \mu_\eta (\sigma^2)) \left[ (Y - X_\mu \beta)^T (I_n \mu_\eta (\sigma^2)) \right]
\]
\[
- \frac{1}{2} \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \frac{1}{2} \sigma^2 \log p \left( q \right)
\]

The second term in B1 is
\[
E_q[ \log P(\sigma^2 \mid b) - \log q^*(\sigma^2)] = \frac{1}{2} E_q[\log \left( I_{n \mu_\eta (\sigma^2)} \right)] - \frac{1}{2} E_q[\log \left( I_{n \mu_\eta (\sigma^2)} \right)] - \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \sigma^2 \log p \left( q \right)
\]

The third term in B1 is
\[
E_q[ \log P(\sigma^2 \mid a_\beta) - \log q^*(\sigma^2)] = \frac{1}{2} E_q[\log \left( I_{n \mu_\eta (\sigma^2)} \right)] - \frac{1}{2} E_q[\log \left( I_{n \mu_\eta (\sigma^2)} \right)] - \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \sigma^2 \log p \left( q \right)
\]

The fourth term in B1 is
\[
E_q[ \log P(a_\beta) - \log q^*(a_\beta)] = \log \left( \frac{\nu_\beta + 1}{2} \right) - \log \left( \frac{\nu_\beta}{2} + \frac{1}{2} E_q[\log(a_\beta)] \right) - \log \left( \frac{\nu_\beta}{2} + \frac{1}{2} E_q[\log(a_\beta)] \right) - \log \left( \frac{\nu_\beta}{2} + \frac{1}{2} E_q[\log(a_\beta)] \right)
\]

The fifth term in B1 is
\[
E_q[ \log P(b_1 \mid \sigma^2) - \log q^*(b_1)] = \log \left( \frac{1}{2} \right) - \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \sigma^2 \log p \left( q \right)
\]

The sixth term in B1 is
\[
E_q[ \log P(b_2 \mid \sigma^2) - \log q^*(b_2)] = \log \left( \frac{1}{2} \right) - \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \sigma^2 \log p \left( q \right) + \frac{1}{2} \sigma^2 \log p \left( q \right)
\]

The seventh term in B1 is
\[
E_q[ \log P(\sigma^2 \mid a_{b_1}) - \log q^*(\sigma^2)] = \frac{1}{2} E_q[\log(2\nu_{b_1})]
\]
\[
- \frac{1}{2} \frac{1}{2} \nu_{b_1} \log \left( q \right) + \frac{1}{2} \frac{1}{2} \nu_{b_1} \log \left( q \right) + \frac{1}{2} \frac{1}{2} \nu_{b_1} \log \left( q \right) + \frac{1}{2} \frac{1}{2} \nu_{b_1} \log \left( q \right)
\]
The eighth term in B1 is

$$E_q [\log P(a_{b1}) - \log q^*(a_{b1})] = \log \Gamma \left( \frac{\nu_{b1} + 1}{2} \right) - \log \Gamma \left( \frac{1}{2} \right) - \frac{1}{2} \log \left( \frac{1}{A_{b1}^2} \right) + \left( \frac{\nu_{b1}}{2} \right) E_q \log (a_{b1}) - \left( \frac{\nu_{b1} + 1}{2} \right)$$

$$\times E_q \log \left( \frac{1}{A_{b1}^2} + \nu_{b1} \mu_q(x_{i+}) \right) + \mu_{q(1/a_{b1})} \nu_{b1} \mu_q(x_{i+})$$

The ninth term in B1 is

$$E_q [\log P(\sigma_{b1}^2 | a_{b2}) - \log q^*(\sigma_{b1}^2)] = \frac{\nu_{b2}}{2} \log (2\nu_{b2}) - \frac{\nu_{b2} + \nu_{b2} E_q \log (a_{b2})}{2}$$

$$- \frac{\nu_{b2} + \nu_{b2} E_q \log \left( \mu_{q(\sigma_{b1}^2)} \right)}{2}$$

The tenth term in B1 is

$$E_q [\log P(\sigma_{b2}^2 | a_{b2}) - \log q^*(\sigma_{b2}^2)] = \log \Gamma \left( \frac{\nu_{b2} + 1}{2} \right) - \log \Gamma \left( \frac{1}{2} \right)$$

$$- \frac{1}{2} \log \left( \frac{1}{A_{b2}^2} \right) + \left( \frac{\nu_{b2}}{2} \right) E_q \log (a_{b2})$$

$$- \left( \frac{\nu_{b2} + 1}{2} \right) E_q \log \left( \frac{1}{A_{b2}^2} + \nu_{b2} \mu_q(x_{i+}) \right) + \mu_{q(1/a_{b2})} \nu_{b2} \mu_q(x_{i+})$$

The eleventh term in B1 is

$$E_q [\log P(\sigma_{b2}^2 | a_e) - \log q^*(\sigma_{b2}^2)] = \frac{\nu_e}{2} \log (2\nu_e) - \frac{\nu_e E_q \log (a_e)}{2}$$

$$+ \frac{\nu_e + \nu_e E_q \log \left( \mu_{q(\sigma_{b2}^2)} \right)}{2}$$

$$+ \frac{\nu_e + \nu_e E_q \log \left( \mu_{q(\sigma_{b2}^2)} \right)}{2}$$

The twelfth term in B1 is

$$E_q [\log P(a_e) - \log q^*(a_e)] = \log \Gamma \left( \frac{\nu_e + 1}{2} \right) - \log \Gamma \left( \frac{1}{2} \right)$$

$$- \frac{1}{2} \log \left( \frac{1}{A_{e2}^2} \right) + \left( \frac{\nu_e}{2} \right) E_q \log (a_e) - \left( \frac{\nu_e + 1}{2} \right)$$

$$\times E_q \log \left( \frac{1}{A_{e2}^2} + \nu_e \mu_q(x_{i+}) \right) + \mu_{q(1/a_e)} \nu_e \mu_q(x_{i+})$$

**APPENDIX C**

**ALTERNATIVE VERSION OF THE ALGORITHM FOR IMPLEMENTING THE VBM E**

For the VBME algorithm given here is an alternative version that ignores the priors of the components $B_{q(\sigma_{b1}^2)}$, $B_{q(\sigma_{b2}^2)}$ since the variance component $\sigma_{b1}^2$ was fixed. Also, the components $B_{q(a_{b1})}$, $B_{q(a_{b2})}$, $B_{q(a_e)}$ were ignored since the components $\nu_{b1} \mu_{q(a_{b1})}$, $\nu_{b2} \mu_{q(1/a_{b2})}$ and $\nu_e \mu_{q(a_e)}$ were replaced by the hyperparameters $Sb1$, $Sb2$ and $Sbe$. To provide informative priors for these parameters, you can use $Sb1=0.25 \times \text{var} (y)^*(\nu_{b1}+1)$, $Sb2=0.25 \times \text{var} (y)^*(\nu_{b2}+1)$ and $Sbe= 0.5 \times \text{var} (y)^*(\nu_e+1)$. 

G3 Genomes | Genomes | Genomics  | Volume 7 | June 2017 | A Genomic-Enabled Prediction Model | 1849
Initialize: \( \mu_{q(b)} = 0, \Sigma_{q(b)} = 0, \Sigma_{q(b)} = I_f, \sigma_{\beta}^2 = 10,000 \) \( \Sigma_{q(b)} = I_f, B_{q(\sigma^2)}, B_{q(\sigma^2)} > 0. \)

Cycle:

\[
\Sigma_{q(b)} \leftarrow \left( \Sigma_{\beta_0}^{-1} \sigma_{\beta}^{-2} + X^T \left( I_n \mu_{q(\sigma^2)} \right) X \right)^{-1}
\]

\[
\mu_{q(b)} \leftarrow \Sigma_{q(b)} \left( \Sigma_{\beta_0}^{-1} \beta_0 \sigma_{\beta}^{-2} + X^T \left[ I_n \mu_{q(\sigma^2)} \right] \left[ Y - \sum_{h=1}^{2} Z_h \mu_{q(b)} \right] \right)
\]

\[
\Sigma_{q(b)} \leftarrow \left( G^{-1} \mu_{q(\sigma^2)} + Z_1^T \left[ I_n \mu_{q(\sigma^2)} \right] Z_1 \right)^{-1}
\]

\[
\mu_{q(b)} \leftarrow \Sigma_{q(b)} \left( Z_1^T \left[ I_n \mu_{q(\sigma^2)} \right] \left[ Y - X \mu_{q(b)} - Z_2 \mu_{q(b)} \right] \right)
\]

\[
\Sigma_{q(b)} \leftarrow \left( G_2^{-1} \mu_{q(\sigma^2)} + Z_2^T \left[ I_n \mu_{q(\sigma^2)} \right] Z_2 \right)^{-1}
\]

\[
\mu_{q(b)} \leftarrow \Sigma_{q(b)} \left( Z_2^T \left[ I_n \mu_{q(\sigma^2)} \right] \left[ Y - X \mu_{q(b)} - Z_1 \mu_{q(b)} \right] \right)
\]

\[
B_{q(\sigma^2)} \leftarrow \frac{\mu_{q(b)}^T G^{-1} \mu_{q(b)}}{2} + Sb1
\]

\[
B_{q(\sigma^2)} \leftarrow \frac{\mu_{q(b)}^T G_2^{-1} \mu_{q(b)}}{2} + Sb2
\]

\[
B_{q(\sigma^2)} \leftarrow e^T e + tr \left( X' X \Sigma_{q(b)} \right) + tr \left( Z_1^T Z_1 \Sigma_{q(b)} \right) \left( Z_2^T Z_2 \Sigma_{q(b)} \right) + Sbc
\]

until the change in the lower bound, \( \log p(y; q) \), between iterations is less than a tolerance value specified a priori. In this case, the corresponding \( \log(p(y; q)) \) has the following analytical expression:

\[
\log p(y; q) = 0.5 \log \left( \left| \Sigma_{q(b)} \right| \right) + 0.5 \log \left( \left| \Sigma_{q(b)} \right| \right) + 0.5 \log \left( \left| \Sigma_{q(b)} \right| \right) - 0.5(n_2 + f) \log \left( \left| B_{q(\sigma^2)} \right| \right)
\]

\[
- 0.5(n_2 + f) \log \left( \left| B_{q(\sigma^2)} \right| \right) - 0.5(n_2 + n) \log \left( \left| B_{q(\sigma^2)} \right| \right)
\]

**APPENDIX D**

**R CODE FOR IMPLEMENTING THE VBME MODEL**

```r
### Set the working directory
rm(list=ls())
setwd("C:\Variationa Bayes Version\VarBayes.Univariate")
### Data
load("DATASET1.Wheat_GY.RData")
Env1=rep(1,599)
Env2=rep(2,599)
Env3=rep(3,599)
Env4=rep(4,599)
Entry=c(1:599)
Env.All=Env1,Env2,Env3,Env4
Environment.All=Entry,Entry,Entry,Entry
YY.N=matrix(Y,nrow=1,byrow=F)
Hybrids=cbind(Env.All,Entry.All,Entry.All,c(YY.N+0))
colnames(Hybrids)=c("Env","Entry","Y")
```

1850 | O. A. Montesinos-López et al.
Hyb <- data.frame(Hybrids)

# Matrix of incidences
X <- model.matrix(~ 0 + as.factor(Hyb$Env))
Z1 <- model.matrix(~ 0 + as.factor(Hyb$Entry))

svda <- svd(G)
D <- diag(svda$d)
d_Inv <- 1 / svda$d
D_Inv <- diag(d_Inv)
U <- svda$u
tV <- t(svda$v)
LL <- U
Z1 <- Z1 %*% LL
Gg <- D
Z2 <- model.matrix(~ 0 + Z1:as.factor(Hyb$Env))
y <- Hyb$y
betav <- c(colMeans(Y))
tX <- t(X)
tZ1 <- t(Z1)
tZ2 <- t(Z2)
I <- ncol(X)
n <- nrow(X)
J <- ncol(Z1)
S_invB <- diag(I)
d2_Inv <- rep(d_Inv, I)
SigmaB <- Sigmab1 = Sigmab2 = 0.2
Sig.e <- 0.1
vb0 = vb1 = vb2 = ve = 2
data.set <- data.frame(Hyb)

K = 1

### Lower bound of the log of the marginal likelihood (log(y|q))
logpyq <- function(y, Mu.q.b0, Mu.q.b1, Mu.q.b2, Sig.q.b0, Sig.q.b1, Sig.q.b2, Bq.Sig.b0, Bq.Sig.b1, Bq.Sig.b2, Bq.Sig.be, X, Z1, Z2, Gg, vb0, vb1, vb2, ve) {
  n.T <- length(y)
  J <- dim(Gg)[1]  # Number of lines
  I <- dim(X)[2]  # Number of environments
  K = 1  # Number of replications
  Inv.Gg <- Gg
  Mu.q.inv.sig.be <- (0.5 * (ve + n.T)) / Bq.Sig.be
  Mu.q.inv.sig.b0 <- (0.5 * vb0) / Bq.Sig.b0
  Mu.q.inv.sig.b1 <- (0.5 * vb1) / Bq.Sig.b1
  Mu.q.inv.sig.b2 <- (0.5 * vb2) / Bq.Sig.b2
  Inv.GD2 <- diag(d2_Inv)
  tX %*% (X)

  log.lik <- 0.5 * log(det(Sig.q.b0)) + 0.5 * log(det(Sig.q.b1)) + 0.5 * sum(log(diag(Sig.q.b2))) - 0.5 * (vb1 + 1) * log(Bq.Sig.b1) - 0.5 * vb2 * log(Bq.Sig.b2) - 0.5 * ve * log(Bg.Sig.be)

  YMatrix <- matrix(y, ncol=4, byrow=F)
  beta_zero <- c(apply(YMatrix, 2, mean, na.rm=TRUE))
  Mu.q.b0 <- rep(1, I)
  Mu.q.b1 <- rep(0.1, J)
  Mu.q.b2 <- rep(0.1, I)
  Sig.q.b0 <- diag(I)
  Sig.q.b1 <- diag(J)
  Sig.q.b2 <- diag(I)
  Bq.Sig.b0 <- Bq.Sig.b1 <- Bq.Sig.b2 <- Bq.Sig.be <- 1
  vb1 = vb2 = ve = 5
  Sb1 = 0.25 * var(y) * (vb1 + 1)
\[ Sb_2 = 0.25 \text{var}(y)^{(vb+1)} \]
\[ Se = \text{var}(y-X\%_{\beta}\text{\_zero})^{(ve+1)} \]
\[ \text{Inv.Gg} = D_{\text{Inv}} \]
\[ n.T = \text{length}(y) \]
\[ tX = t(X) \]
\[ tZ_1 = t(Z_1) \]
\[ tZ_2 = t(Z_2) \]
\[ LL_0 = \log p(y; \mu.q.b_0 = \mu.q.b_0.sv, \mu.q.b_1 = \mu.q.b_1.sv, \mu.q.b_2 = \mu.q.b_2.sv, \text{sig.q.b}_0 = \text{sig.q.b}_0.sv, \text{sig.q.b}_1 = \text{sig.q.b}_1.sv, \text{sig.q.b}_2 = \text{sig.q.b}_2.sv, \text{bq.sig.b}_0 = \text{bq.sig.b}_0.sv, \text{bq.sig.b}_1 = \text{bq.sig.b}_1.sv, \text{bq.sig.b}_2 = \text{bq.sig.b}_2.sv, \text{bq.sig.be} = \text{bq.sig.be.sv}, X = X, Z_1 = Z_1, Z_2 = Z_2, Gg = D, vb_0 = vb_0, vb_1 = vb_1, vb_2 = vb_2, ve = ve) \]
\[ LL_0 \]

Tolerance \[ < -c(0.00001) \]  # Tolerance to declare convergence
\[ \text{converged} = \text{FALSE} \]
\[ \text{maxIter} = 1000 \]
\[ \text{itnum} = 0 \]

while(!converged) {
    \[ \text{itnum} = \text{itnum} + 1; \]

    Updating the parameters of interest

    if (itnum == 1) previous.LL_0 = LL_0

    if (itnum > 1) previous.LL_0 = previous.LL_0 - current.LL_0

    \[ \mu.q.inv.sig.be = (0.5^{(ve+n.T)}/(\text{diag}(I) + \mu.q.inv.sig.be^{(tX%X)})) \]

    \[ \mu.q.inv.sig.b_0 = 1/10000 \]

    \[ \text{mu.q.inv.sig.b}_0 = \text{solve}((\text{diag}(I))^{(\mu.q.inv.sig.b_0) + \mu.q.inv.sig.be^{(tX%X)})} \]

    \[ \mu.q.b_0 = \text{inv}(\mu.q.inv.sig.b_0 + \mu.q.inv.sig.be^{(tX%X)}) \]

    \[ \text{inv.IG.D}_2 = 1/(d_22_\text{Inv} \cdot \mu.q.inv.sig.b_2 + \mu.q.inv.sig.be^{(tZ_2%Y)}) \]

    \[ \text{sig.q.b}_2 = \text{inv}(\text{diag}(\text{inv.IG.D}_2))^{(\mu.q.inv.sig.b_2) + \text{sum}(\text{diag}(\text{inv.IG.D}_2^{(\mu.q.inv.sig.b_2)})) + Sb_2} \]

    \[ \text{ym}_0 = (y-X\%_{\mu.q.b_0-Z1\%_{\mu.q.b_1}}) \]

    \[ \text{log.lik} = 0.5 \cdot \log(\text{det}(\text{sig.q.b}_0)) + 0.5 \cdot \text{sum}(\text{log}(\text{ddd}_1)) + 0.5 \cdot \text{sum}(\text{log}(\text{ddd}_2)) \]

    \[ \text{relErr} = -c(0.00001) \]

    if (itnum >= maxIter) {
        converged = TRUE
        print("Warning: maximum number of iterations exceeded."
    }
    \[ \text{current.LL} = -\text{log.lik} \]
    \[ \text{relErr} = -c(0.00001) \]
    \[ \text{converged} = \text{TRUE} \]
    \[ \text{print}("\text{Warning: maximum number of iterations exceeded.}\)"

    if (relErr < tolerance) converged = TRUE
    \[ \text{bq.sig.b}_1 = \text{c}(\text{bq.sig.b}_1) \]
    \[ \text{mu.q.b}_1 = \text{mu.q.b}_1 \]
    \[ \text{mu.q.b}_2 = \text{mu.q.b}_2 \]
    \[ \text{bq.sig.b}_1 = \text{c}(\text{bq.sig.b}_1) \]
    \[ \text{bq.sig.b}_2 = \text{c}(\text{bq.sig.b}_2) \]

    Estimates of variance components

    \[ \text{sig.b}_1 = (\text{bq.sig.b}_1) / (0.5^{(vb+1+1)}) \]
    \[ \text{sig.b}_2 = (\text{bq.sig.b}_2) / (0.5^{(vb+2+1)}) \]
    \[ \text{sig.e} = (\text{bq.sig.be}) / (0.5^{(ve+n.T)}) \]
cat('Iteration:', itnum, 'n\n')
cat('Current LL0: ', round(current.LL0,3), 'n', 'Previous LL0: ',
    round(previous.LL0,3), 'n', 'Tolerance: ', round(relErr,3), 'n',
    'Converged: ', converged, 'n\n')
}

#################################Printing parameter estimates#################################
Par.estimates=rbind(Mu.q,b0,Sig.b1,Sig.b2,Sig.e)
Par.estimates