A Bayesian Poisson-lognormal Model for Count Data for Multiple-Trait Multiple-Environment Genomic-Enabled Prediction

Osval A. Montesinos-López,* Abelardo Montesinos-López,† José Crossa,‡,1 Fernando H. Toledo,‡ José C. Montesinos-López,§ Pawan Singh,‡ Philomin Juliana,‡ and Josafhat Salinas-Ruiz**

*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, ‡International Maize and Wheat Improvement Center (CIMMYT), 06600 México, D.F., México, §Departamento de Estadística, Centro de Investigación en Matemáticas (CIMAT), Guanajuato 36240, México, and **Colegio de Postgraduados, Campus Córdoba, Km. 348 Carretera Federal Córdoba-Veracruz, Amatlán de los Reyes, 94946, México

ABSTRACT When a plant scientist wishes to make genomic-enabled predictions of multiple traits measured in multiple individuals in multiple environments, the most common strategy for performing the analysis is to use a single trait at a time taking into account genotype × environment interaction (G × E), because there is a lack of comprehensive models that simultaneously take into account the correlated counting traits and G × E. For this reason, in this study we propose a multiple-trait and multiple-environment model for count data. The proposed model was developed under the Bayesian paradigm for which we developed a Markov Chain Monte Carlo (MCMC) with noninformative priors. This allows obtaining all required full conditional distributions of the parameters leading to an exact Gibbs sampler for the posterior distribution. Our model was tested with simulated data and a real data set. Results show that the proposed multi-trait, multi-environment model is an attractive alternative for modeling multiple count traits measured in multiple environments.

Plant breeders need more efficient models for performing genomic selection for multiple-traits and multiple-environments for count data. Count data are those dependent variables that take values 0, 1, 2,... without an explicit upper limit. These types of dependent variables are common in genomic selection, for example: panicle number per plant, seed number per plant, number of infected spikelets per plant, *etc*. Due to its simplicity and its ability to generate samples from high-dimensional probability distributions, the Gibbs sampler is one of the most popular computationally intensive methods for fitting complex multilevel models (Park and van Dyk 2009). This method is also very popular for modeling

normal and binary responses when efficient closed-form Gibbs samplers have been developed. However, obtaining a closed-form Gibbs sampler for count data is not straightforward. For this reason, Montesinos-López *et al.* (2015, 2016a) in the context of genomic-enabled prediction and genomic selection proposed closed-form Gibbs samplers for multi-level models for univariate count responses with and without the genotype × environment interaction (G × E) term that helps fill the lack of closed-form Gibbs samplers for count data. Although these models are helpful for modeling univariate count responses, many times breeders record phenotypic data for multiple counts. Scientists must take advantage of correlated traits to improve the prediction of unobserved genotypes and to increase the prediction accuracy of other count traits that are difficult to measure but that are associated with traits that are easy to measure. The available univariate count models are not appropriate for dealing with these situations.

Since prediction problems are ubiquitous and of great interest and importance in statistical science, more attention has been given to parametric inference than to predictive inference (Harville 2014). However, thanks to the efforts of scientists like C.R. Henderson, currently there is a lot of evidence that the model-based approach

[1]Corresponding author: Apdo. Postal 6-641, Mexico DF, 06600, Mexico. E-mail: j.crossa@cgiar.org

to prediction is a useful tool for predicting future observations, and many linear mixed effects models have been developed for predicting future observations (Harville 2014). Most of these models are valid for normal responses, but few have been developed for discrete outcomes and very few for predicting more than one trait at a time. Thus there is a great need for developing more univariate and multivariate models for non-normal traits. Developing these models is very challenging since in predictive inference, inferences are sought about tangible quantities that are regarded as unobservable at the time of the inferences; the focus is on quantities that will, or could, become observable in the future, or that can be observed at the time of the inferences but only with an unacceptable delay or with undue effort or expense (Harville 2014). These quantities are regarded as the realizations of the elements of an unobservable random vector, say an M-dimensional unobservable random column vector **w**, and/or as realizations of linear combinations (or other functions) of the elements of **w** (Harville 2014).

Poisson and negative binomial (NB) distributions are the most common random variables used for modeling count data; NB distribution is preferred when there is evidence of considerable overdispersion. In the Bayesian context, estimating the overdispersion parameter in the NB distribution is challenging because Metropolis-Hastings algorithms are used most of the time; this is computationally expensive and not practical for use in genomic-enabled prediction where data sets are large. Montesinos-López et al. (2015, 2016a) proposed a Gibbs sampler for NB distribution but it is not computationally efficient. Other authors have proposed using the Poisson-lognormal distribution to model count data to account for overdispersion (Williams and Ebel 2012). The Poisson component of the Poisson-lognormal distribution accommodates integer inputs (or outputs) to describe the actual number of counts observed within a single unit or sample, while the lognormal component of the distribution describes the overdispersion in the Poisson rate parameter due to clustering of some factors and describes how the average of these factors varies across the population (Williams and Ebel 2012). Adding this lognormal component to the predictor of a Poisson model is very helpful for accommodating a general correlation structure between traits when more than one trait is under study (Ma et al. 2008).

Based on the previous considerations, the main goal of this research is to extend the genomic-enabled Bayesian prediction model for count data with genotype × environment (G × E) interaction to the context of multiple traits under a Poisson-lognormal model. Since nowadays scientists measure multiple count traits in multiple environments, the joint modeling of multiple count traits can help to increase prediction accuracy and parameter estimation accuracy, and reduce trait selection bias. This argument is very well documented for continuous phenotypes; see, for example, Henderson and Quaas (1976), Pollak et al. (1984), Schaeffer (1984), Jiang et al. (2015), and Montesinos-López et al. (2016b). It is reasonable that it could also substantially help in the context of multivariate count phenotypes.

MATERIALS AND METHODS

Experimental data

First we describe real phenotypic and genotypic experimental data used to illustrate the results of the new Poisson-lognormal model. Then we explain the theory (joint posterior density and prior specification) of the proposed model, the Gibb sampler, and its implementation. We also describe how to assess prediction accuracy and simulated data sets, genomic-enabled prediction models, and give the link to the data and software availability.

Phenotypic data

The phenotypic data set used included 182 spring wheat lines developed by the International Maize and Wheat Improvement Center (CIMMYT) that were assembled and evaluated for resistance to *Fusarium graminearum* at El Batan experiment station in Mexico in 2011 in three experiments. For the application, we call these three experiments Env1, Env2, and Env3. In all the experiments (environments), the genotypes were arranged in a randomized complete block design, in which each plot comprised two 1-m double rows separated by a 0.25-m space. *Fusarium* head blight (FHB) severity data were collected 20 and 30 d before maturity by counting symptomatic spikelets on five randomly selected spikes in each plot. We used the counts collected at 20 d as trait 1 and the counts collected at 30 d as trait 2. These data sets were taken from the data used by Montesinos-López et al. (2016a) in their paper for count data with genotype × environment interaction.

Genotypic data

DNA samples were extracted from young leaves 2–3 wk old taken from each line, using Wizard Genomic DNA purification (Promega) following the manufacturer's protocol. DNA samples were genotyped using an Illumina 9K SNP chip with 8632 single nucleotide polymorphisms (SNPs) (Cavanagh et al. 2013). For a given marker, the genotype for the ith line was coded as the number of copies of a designated marker-specific allele carried by the ith line (absence = zero and presence = one). SNP markers with unexpected AB (heterozygous) genotype were recoded as either AA or BB based on the graphical interface visualization tool of GenomeStudio (Illumina) software. SNP markers that did not show clear clustering patterns were excluded. In addition, 66 simple sequence repeat markers were screened. After filtering the markers for 0.05 minor allele frequency and deleting markers with >10% of no calls, the final set of SNPs included 1635 SNPs.

Statistical model

Let $Y_{ijk}^{(l)}$ be the phenotypic response of trait l, in replication k, in environment i for genotype j. Conditionally on $b_j^{(l)}$, $b_{ij}^{(l)}$, and $c_{ijk}^{(l)}$, we assume that the distribution of $Y_{ijk}^{(l)}$ is a Poisson distribution with mean equal to $\mu_{ijk}^{(l)} = \exp\left(\eta_{ijk}^{(l)}\right)$, with the following linear predictor:

$$\eta_{ijk}^{(l)} = \beta_i^{(l)} + b_j^{(l)} + b_{ij}^{(l)} + c_{ijk}^{(l)} \tag{1}$$

where $\beta_i^{(l)}$ is the intercept of environment i in trait l, $b_j^{(l)}$ is the random effect of line j in trait l, $b_{ij}^{(l)}$ is the interaction random effect of environment i, line j and trait l, and $c_{ijk}^{(l)}$ is an individual random effect for the response in the replication k, environment i, line j, and trait l; the $\boldsymbol{c}_{ijk}$ are jointly distributed as $\boldsymbol{c}_{ijk} = \begin{bmatrix} c_{ijk}^{(1)} \\ \vdots \\ c_{ijk}^{(n_T)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_c\right)$, in order to take into account part of the correlation between traits not explained by the other random effects. It is interesting to point out that marginally the proposed model is equivalent to a Poisson-lognormal model with random effects. Conditioning on $b_j^{(l)}$, $b_{ij}^{(l)}$, and $c_{ijk}^{(l)}$, the distribution of $Y_{ijk}^{(l)}$ can be approximated using the NB distribution as:

$$P\left(Y_{ijk}^{(l)} = y_{ijk}^{(l)}\right) = \frac{\Gamma\left(y_{ijk}^{(l)} + r\right)}{y_{ijk}^{(l)}!\Gamma(r)}\left(1 - \frac{\mu_{ijk}^{(l)}}{r + \mu_{ijk}^{(l)}}\right)^r\left(\frac{\mu_{ijk}^{(l)}}{r + \mu_{ijk}^{(l)}}\right)^{y_{ijk}^{(l)}} \quad (2)$$

with a large enough value of r (for example, $r = 1000$). This probability can be expressed as:

$$\begin{aligned} P\left(Y_{ijk}^{(l)} = y_{ijk}^{(l)}\right) &= \frac{\Gamma\left(y_{ijk}^{(l)} + r\right)}{y_{ijk}^{(l)}!\Gamma(r)}\left(1 - \frac{\mu_{ijk}^{(l)}}{r + \mu_{ijk}^{(l)}}\right)^r\left(\frac{\mu_{ijk}^{(l)}}{r + \mu_{ijk}^{(l)}}\right)^{y_{ijk}^{(l)}} \\ &= \frac{\Gamma\left(y_{ijk}^{(l)} + r\right)}{y_{ijk}^{(l)}!\Gamma(r)}\frac{\left[\exp\left(\eta_{ijk}^{*(l)}\right)\right]^{y_{ijk}^{(l)}}}{\left[1 + \exp\left(\eta_{ijk}^{*(l)}\right)\right]^{y_{ijk}^{(l)}+r}} \end{aligned}$$

where $\eta_{ijk}^{*(l)} = x_i^{(l)T}\beta^{*(l)} + b_j^{(l)} + b_{ij}^{(l)} + c_{ijk}^{(l)}$, and $\beta^{*(l)} = \beta^{(l)} - \log(r)$. Then, using the identity of Polson *et al.* (2013), this expression (Equation 2) can be expressed as:

$$\begin{aligned} P\left(Y_{ijk}^{(l)} = y_{ijk}^{(l)}\right) &= \frac{\Gamma\left(y_{ijk}^{(l)} + r\right)}{y_{ijk}^{(l)}!\Gamma(r)}\frac{\left[\exp\left(\eta_{ijk}^{*(l)}\right)\right]^{y_{ijk}^{(l)}}}{\left[1 + \exp\left(\eta_{ijk}^{*(l)}\right)\right]^{y_{ijk}^{(l)}+r}} \\ &= \frac{\Gamma\left(y_{ijk}^{(l)} + r\right)}{y_{ijk}^{(l)}!\Gamma(r)}2^{-y_{ijk}^{(l)}-r}\exp\left(\frac{y_{ijk}^{(l)} - r}{2}\eta_{ijk}^{*(l)}\right) \\ &\quad \times \int_0^\infty \exp\left(\frac{-\omega_{ijk}^{(l)}\left(\eta_{ijk}^{*(l)}\right)^2}{2}\right)f\left(\omega_{ijk}^{(l)}; y_{ijk}^{(l)} + r, 0\right)d\omega_{ijk}^{(l)} \end{aligned}$$

where $f(\cdot\ ; b, d)$ denotes the probability density function of a Pólya-Gamma random variable ($\omega_{ijk}^{(l)}$) with parameters b and d. Then conditioning on $b_j^{(l)}$, $b_{ij}^{(l)}$, $c_{ijk}^{(l)}$, and $\omega_{ijk}^{(l)}$, the distribution of $Y_{ijk}^{(l)}$ is expressed as:

$$\begin{aligned} &P\left(Y_{ijk}^{(l)} = y_{ijk}^{(l)}\Big|ELSE\right) \\ &= \frac{\Gamma\left(y_{ijk}^{(l)} + r\right)}{y_{ijk}^{(l)}!\Gamma(r)}2^{-y_{ijk}^{(l)}-r}\exp\left(\frac{y_{ijk}^{(l)} - r}{2}\eta_{ijk}^{*(l)}\right)\ \exp\left(\frac{-\omega_{ijk}^{(l)}\left(\eta_{ijk}^{*(l)}\right)^2}{2}\right) \end{aligned}$$

ELSE means that the conditioning was done on $b_j^{(l)}$, $b_{ij}^{(l)}$, $c_{ijk}^{(l)}$, and $\omega_{ijk}^{(l)}$. Therefore, the joint conditional distribution of all phenotypic count responses in the same individual is given by:

$$\begin{aligned} &P\left(Y_{ijk}^{(1)} = y_{ijk}^{(1)}, \cdots, Y_{ijk}^{(n_T)} = y_{ijk}^{(n_T)}\Big|ELSE\right) = \prod_{l=1}^{n_T} P\left(Y_{ijk}^{(l)} = y_{ijk}^{(l)}\Big|ELSE\right) \\ &= \prod_{l=1}^{n_T}\frac{\Gamma\left(y_{ijk}^{(l)} + r\right)}{y_{ijk}^{(l)}!\Gamma(r)}2^{-y_{ijk}^{(l)}-r}\exp\left(\frac{y_{ijk}^{(l)} - r}{2}\eta_{ijk}^{*(l)} - \frac{\omega_{ijk}^{(l)}\left(\eta_{ijk}^{*(l)}\right)^2}{2}\right) \\ &\propto \exp\left(\sum_{l=1}^{n_T}\frac{y_{ijk}^{(l)} - r}{2}\eta_{ijk}^{*(l)} - \sum_{t=1}^{n_T}\frac{\omega_{ijk}^{(l)}\left(\eta_{ijk}^{*(l)}\right)^2}{2}\right)\ \propto \exp\left(y_{ijk}^{*T}\eta_{ijk}^* - \frac{1}{2}\eta_{ijk}^{*T}D_{ijk}\eta_{ijk}^*\right) \end{aligned}$$

$$(3)$$

where $y_{ijk}^* = \frac{1}{2}\left[y_{ijk}^{(1)} - r, \cdots, y_{ijk}^{(n_T)} - r\right]^T$, $\eta_{ijk}^* = \left[\eta_{ijk}^{(1)}, \cdots, \eta_{ijk}^{(T)}\right]^T$ and $D_{ijk} = \text{Diag}\left(\omega_{ijk}^{(1)}, \cdots, \omega_{ijk}^{(n_T)}\right)$. Here *ELSE* has a similar meaning as before, but adds the corresponding random effects of each response. In this case, we are conditioning on $b_{1j} = \left[b_j^{(1)}, \ldots, b_j^{(n_T)}\right]^T$, $b_{2ij} = \left[b_{ij}^{(1)}, \ldots, b_{ij}^{(n_T)}\right]^T$, $c_{ijk} = \left[c_{ijk}^{(1)}, \ldots, c_{ijk}^{(n_T)}\right]^T$, and $\omega_{ijk} =$

$\left[\omega_{ijk}^{(1)}, \ldots, \omega_{ijk}^{(n_T)}\right]^T$. Therefore, according to Appendix A, the joint conditional distribution for all phenotypic count responses is expressed as:

$$\begin{aligned} P(Y|\beta^*, b_1, b_2, c, \omega) &\propto \exp\Big(y^{*T}(X\beta^* + Z_1b_1 + Z_2b_2 + c) \\ &\quad - \frac{1}{2}(X\beta^* + Z_1b_1 + Z_2b_2 + c)^T \\ &\quad \times D(X\beta^* + Z_1b_1 + Z_2b_2 + c)\Big). \end{aligned}$$

where $\omega = [\omega_1, \ldots, \omega_I]^T$, $\omega_i = [\omega_{i1}, \ldots, \omega_{iJ}]^T$, $\omega_{ij} = [\omega_{ij1}, \ldots, \omega_{ijK}]^T$, and the others terms are defined in the Appendix A. Hereafter we assume that the joint distribution of the random line-trait effects is $b_1|\Sigma_t \sim N_{JL}(0, G_1)$, with $G_1 = G \otimes \Sigma_t$, the joint distribution of the interaction random effect of environment i and line j in trait l is $b_2|\Sigma_t, \Sigma_E \sim N_{IJL}(0, G_2)$, with $G_2 = \Sigma_E \otimes G \otimes \Sigma_t$, and $c|\Sigma_c \sim N_{JL}(0, \Sigma_{c(a)})$ with $\Sigma_{c(a)} = I_{KJI} \otimes \Sigma_c$, assuming that all the extra random effects (c_{ijk}) are independent and identically distributed. It is important to point out that the correlation between traits is taken into account in both parts of the model: (a) in random effect b_1 with general covariance matrix Σ_t, and (b) in random effect c with general covariance matrix Σ_c, while the correlation between environments is taken into account in random effect b_2 with general covariance matrix Σ_E. Normal random effect c resulted because we paired a lognormal distribution, $\exp(c)$, with the Poisson distribution to create an overdispersed distribution, which is referred to as the Poisson-lognormal distribution. The above statement is clear if we remember that if W has a normal distribution, then $Z = \exp(W)$ has a lognormal distribution.

Prior specification and joint posterior

In this section, we provide the joint posterior density and prior specification for the Bayesian Poisson Multiple-Trait and Multiple-Environment (BPMTME) model. The joint posterior density of the parameter vector becomes:

$$\begin{aligned} &f(\beta^*, b_1, b_2, c, \omega, \Sigma_t, \Sigma_E, \Sigma_c, a, a_E, a_c|Y) \\ &\propto P(Y|\beta^*, b_1, b_2, c, \omega)f(\beta^*)f(b_1|\Sigma_t)f(\Sigma_t|a_1, \ldots, a_L)f(a_1, \ldots, a_L) \\ &\times f(b_2|\Sigma_t, \Sigma_E)f(\Sigma_E|a_{E_1}, \ldots, a_{E_I})f(a_{E_1}, \ldots, a_{E_I})f(c|\Sigma_c)f(\Sigma_c|a_{c_1}, \ldots, a_{c_L}) \\ &\times f(a_{c_1}, \ldots, a_{c_L})f(\omega; Y+r, 0) \end{aligned}$$

$$(4)$$

where we are implicitly assuming the following structure of the prior density of the hyper-parameters:

$$\begin{aligned} &f(\beta^*, \Sigma_t, \Sigma_E, \Sigma_c, a, a_E, a_c) \\ &= f(\beta^*)f(\Sigma_t|a_1, \ldots, a_L)\prod_{l=1}^L f(a_l)f(\Sigma_E|a_{E_1}, \ldots, a_{E_I}) \\ &\times \prod_{i=1}^I f(a_{E_i}) \times f(\Sigma_c|a_{c_1}, \ldots, a_{c_L})\prod_{l=1}^L f(a_{c_l}). \end{aligned}$$

More specifically, we assume that $\beta^* \sim N_p\left(\beta_\nu^*, \Sigma_\nu^{-1}\right)$, $\Sigma_t|a_1, \ldots, a_L \sim IW\left(\nu_t + L - 1, 2\nu_t\,\text{diag}\left(\frac{1}{a_1}, \ldots \frac{1}{a_L}\right)\right)$, $a_l \sim IG\left(\frac{1}{2}, 1/A_l^2\right)$, $l = 1, \ldots, L$, $\Sigma_E|a_{E1}, \ldots, a_L \sim IW\left(\nu_{Ei} + I - 1, 2\nu_E\,\text{diag}\left(\frac{1}{a_{E1}}, \ldots \frac{1}{a_{EI}}\right)\right)$, $a_{Ei} \sim IG\left(\frac{1}{2}, 1/A_{Ei}^2\right)$, $i = 1, \ldots, I$, $\Sigma_c|a_{c1}, \ldots, a_{cL} \sim IW$ $\left(\nu_{cl} + L - 1, 2\nu_c\,\text{diag}\left(\frac{1}{a_{c1}}, \ldots \frac{1}{a_{cL}}\right)\right)$, and $a_{cl} \sim IG\left(\frac{1}{2}, 1/A_{cl}^2\right)$. Here

$\mathbf{\Omega} \sim IW(\kappa, \mathbf{B})$ indicates an inverse Wishart random matrix distribution with density function $|\mathbf{\Omega}|^{-\frac{\kappa+p+1}{2}} \exp\left[-\frac{1}{2}tr(\mathbf{B}\mathbf{\Omega}^{-1})\right], \kappa > 0,\ \mathbf{B},\ \mathbf{\Omega}$; both are positive definite matrices, and $IG(a, b)$ denotes an inverse gamma distribution with shape parameter a and rate parameter b.

Using the specified priors we ended up with all full conditional distributions given in Appendix B.

Gibbs sampler

To produce posterior means for all relevant model parameters, below we outline the exact Gibbs sampler procedure that we propose for estimating the parameters of interest. As is the case with Markov Chain Monte Carlo (MCMC) techniques, the ordering of draws is somewhat arbitrary; however, we suggest the following order:

Step 1. Simulate $\boldsymbol{\beta}^*$ according to the normal distribution given in Appendix B (B1).

Step 2. Simulate $\boldsymbol{b}_1$ according to the normal distribution given in Appendix B (B2).

Step 3. Simulate $\boldsymbol{b}_2$ according to the normal distribution given in Appendix B (B3).

Step 4. Simulate $\boldsymbol{c}$ according to the normal distribution given in Appendix B (B4).

Step 5. Simulate $\boldsymbol{\Sigma}_t$ according to the IW distribution given in Appendix B (B5).

Step 6. Simulate a_{lt} according to the IG distribution given in Appendix B (B6).

Step 7. Simulate $\boldsymbol{\Sigma}_E$ according to the IW distribution given in Appendix B (B7).

Step 8. Simulate a_{lE} according to the IG distribution given in Appendix B (B8).

Step 9. Simulate $\omega_{ijk}^{(l)}$ according to the Pólya-gamma distribution given in Appendix B (B9).

Step 10. Simulate $\boldsymbol{\Sigma}_c$ according to the IW distribution given in Appendix B (B10).

Step 11. Simulate a_{lc} according to the IG distribution given in Appendix B (B11).

Step 12. Return to step 1 or terminate when chain length is adequate for meeting convergence diagnostics.

Model implementation

The Gibbs sampler described above for the BPMTME model was implemented as an R-software package (R Core Team 2016). We performed a total of 40,000 iterations; 20,000 samples were used for inference because the first 20,000 were used as burn-in to decrease MCMC errors in prediction accuracy. We used a thinning of five, so that 4000 samples were used for inference. The chain convergence diagnostic was done by visual checks using the trace plots and autocorrelation functions of each component of β coefficients and variance components, and in general we observed that they stabilize very quickly for the real and simulated data. We implemented the prior specification given in the previous section where the BPMTME model was defined. The hyper-parameters used were: $\boldsymbol{\beta}_v^* = \mathbf{0}_{IL}$, $\boldsymbol{\Sigma}_v = 1 \times 10^4 \boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_0 = \boldsymbol{I}_{IL}$, $\nu_t = \nu_{Ei} = \nu_{cl} = 2$, and $A_l = A_{Ei} = A_{cl} = 1 \times 10^4$, where $\boldsymbol{I}_{IL}$ is the identity matrix of dimension $IL \times IL$. All these hyper-parameters were chosen to lead weakly informative priors. Also, it is important to point out that for the implementation with real and simulated data sets, we worked with the sum for each line resulting from its corresponding

number of replicates. This was done to save time in the implementation of the proposed model.

Assessing the prediction accuracy of the models

We used cross-validation 1 (CV1), which mimics a situation where lines were evaluated in some environments for all traits but some lines are missing in other environments, as proposed by Montesinos-López et al. (2016b). For real and simulated data, prediction ability was assessed using 10 trn-tst (trn = training and tst = testing) random partitions; we used this approach because it provides higher precision in the predictive estimates than the framework that uses different numbers of folds. However, for the simulated data we used five different partitions (percentages) for the training and testing sets, and the corresponding testing sets were assumed as missed values. The percentages for the training set were 90, 80, 70, 60, and 50% and their corresponding complements were used as testing (tst) sets (tst = 10, tst = 20, tst = 30, tst = 40, and tst = 50%). Of the variety of methods for comparing the predictive posterior distribution to the observed data (generally termed "posterior predictive checks"), we used Spearman's correlation because the phenotype is not normally distributed. Models with values closer to one indicate better predictions. The predicted observations were calculated with S collected Gibbs samplers as: $\hat{\boldsymbol{Y}}^{(s)} = \exp(\boldsymbol{X}\hat{\boldsymbol{\beta}}^{*(s)} + \boldsymbol{Z}_1\hat{\boldsymbol{b}}_1^{(s)} + \boldsymbol{Z}_2\hat{\boldsymbol{b}}_2^{(s)} + \hat{\boldsymbol{c}})$, where $\hat{\boldsymbol{\beta}}^{*(s)}, \hat{\boldsymbol{b}}_1^{(s)}, \hat{\boldsymbol{b}}_2^{(s)}$, and $\hat{\boldsymbol{c}}$ are estimates of $\boldsymbol{\beta}^*$, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{c}$ in the sth collected sample.

Simulated data sets

The proposed model was also tested with two simulated data sets (S1 and S2). Both data sets were simulated under the proposed model (described in the statistical model section) with three environments, two traits, 200 genotypes, and five replications. The β coefficients used for both data sets were $\boldsymbol{\beta}^T = [0.20, 0.25, 0.15, 0.20, 0.30,$ and $0.32]$. The first two β coefficients belong to traits 1 and 2 in environment 1, the third and fourth values belong to traits 1 and 2 in environment 2, and the last two traits for environment 3. The variance-covariance matrices used for the first data set (scenario S1) gave rise to a matrix of correlation between traits and environments of 0.8; these matrices were:

$$\boldsymbol{\Sigma}_t = \begin{bmatrix} 0.005 & 0.0031 \\ 0.0031 & 0.003 \end{bmatrix}, \quad \boldsymbol{\Sigma}_E = \begin{bmatrix} 0.003 & 0.0022 & 0.0024 \\ 0.0022 & 0.0022 & 0.0012 \\ 0.0024 & 0.0012 & 0.0030 \end{bmatrix}, \text{ and}$$

$$\boldsymbol{\Sigma}_c = \begin{bmatrix} 0.0003 & 0.0003 \\ 0.0003 & 0.0004 \end{bmatrix}.$$ The variance-covariance matrices used for the second data set (scenario 2) gave rise to correlation matrices with correlations equal to 0.3; the corresponding covariances were:

$$\boldsymbol{\Sigma}_t = \begin{bmatrix} 0.005 & 0.0012 \\ 0.0012 & 0.003 \end{bmatrix}, \quad \boldsymbol{\Sigma}_E = \begin{bmatrix} 0.003 & 0.0007 & 0.0009 \\ 0.0007 & 0.0022 & 0.0007 \\ 0.0009 & 0.0007 & 0.0030 \end{bmatrix}, \text{ and}$$

$$\boldsymbol{\Sigma}_c = \begin{bmatrix} 0.0003 & 0.0001 \\ 0.0001 & 0.0004 \end{bmatrix}.$$ It is also important to point out that for both simulated data sets, we assumed independence between genotypes, i.e., $\mathbf{G}_g = \boldsymbol{I}_{200}$. The implementation of the BPMTME model for these two simulated data sets was the same as that used with the real data sets described in the section on model implementation.

Data availability

The Gibbs algorithm described in this paper was implemented in C++ using the Armadillo linear algebra Library (Eddelbuettel and Sanderson

■ Table 1 Prediction accuracy measured with the average of the 10 random partitions using the Spearman correlation (ASC) for each testing (tst) percentage for the BPMTME and BPME models

		BPMTME					BPME				
		tst = 10%	tst = 20%	tst = 30%	tst = 40%	tst = 50%	tst = 10%	tst = 20%	tst = 30%	tst = 40%	tst = 50%
	Trait–Env	ASC	ASC	ASC	ASC	ASC	ASC	ASC	ASC	ASC	ASC
	11	0.625	0.575	0.710	0.635	0.597	0.480	0.574	0.574	0.551	0.553
	21	0.693	0.568	0.705	0.566	0.542	0.533	0.527	0.560	0.572	0.546
	12	0.815	0.571	0.694	0.615	0.551	0.492	0.395	0.452	0.423	0.397
	22	0.548	0.553	0.685	0.612	0.541	0.597	0.627	0.608	0.609	0.590
	13	0.754	0.575	0.715	0.623	0.518	0.613	0.635	0.616	0.581	0.569
	23	0.610	0.590	0.706	0.628	0.559	0.539	0.601	0.556	0.600	0.597
	Average	0.674	0.572	0.702	0.613	0.551	0.542	0.560	0.561	0.556	0.542
S1		SD	SD	SD	SD	SD	SD	SD	SD	SD	SD
	11	0.118	0.104	0.060	0.045	0.050	0.053	0.018	0.018	0.013	0.016
	21	0.104	0.118	0.046	0.059	0.045	0.033	0.026	0.025	0.013	0.007
	12	0.048	0.113	0.032	0.061	0.051	0.026	0.025	0.016	0.014	0.009
	22	0.136	0.116	0.072	0.065	0.045	0.032	0.032	0.015	0.010	0.011
	13	0.077	0.135	0.045	0.047	0.048	0.033	0.030	0.018	0.012	0.011
	23	0.102	0.118	0.062	0.050	0.043	0.033	0.019	0.014	0.013	0.008
	Trait–Env	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
	11	0.728	0.514	0.740	0.643	0.581	0.499	0.581	0.584	0.548	0.561
	21	0.561	0.630	0.711	0.587	0.544	0.511	0.550	0.561	0.584	0.561
	12	0.537	0.623	0.668	0.647	0.567	0.529	0.402	0.437	0.439	0.416
	22	0.531	0.575	0.721	0.625	0.550	0.613	0.641	0.591	0.616	0.600
	13	0.571	0.633	0.674	0.613	0.522	0.607	0.618	0.593	0.564	0.570
	23	0.640	0.538	0.711	0.636	0.568	0.570	0.574	0.544	0.588	0.585
	Average	0.595	0.585	0.704	0.625	0.555	0.555	0.561	0.552	0.557	0.549
S2		SD	SD	SD	SD	SD	SD	SD	SD	SD	SD
	11	0.074	0.122	0.044	0.050	0.054	0.046	0.023	0.016	0.015	0.012
	21	0.154	0.098	0.057	0.055	0.061	0.039	0.028	0.026	0.014	0.007
	12	0.144	0.088	0.041	0.056	0.039	0.023	0.025	0.014	0.013	0.010
	22	0.142	0.108	0.055	0.054	0.035	0.025	0.029	0.020	0.012	0.012
	13	0.137	0.112	0.065	0.058	0.040	0.030	0.027	0.017	0.013	0.010
	23	0.105	0.135	0.064	0.051	0.043	0.032	0.020	0.014	0.014	0.006

SD denotes standard deviations.

2014). It was included in the R package BMTME as an extension of its current features. With the updated version (0.0.4), the user can also fit count data by means of the same pipeline presented by Montesino-López *et al.* (2016b). However, due to the need for sampling from a Pólya-gamma distribution, it was necessary to include a dependency regarding the R package BayesLogit (Polson *et al.* 2013) and for the moment, we are only distributing the package through CIMMYT's repository http://hdl.handle.net/11529/10866. Users must have BayesLogit properly installed in their computers before installing BMTME v0.0.4, as well as the R version 3.2.4. or higher. The phenotypic (FHB) and genotypic (marker) data used in this study can also be downloaded from that link.

RESULTS

The results are given in two sections: the first section gives the results of the simulated data and the second section gives the results of the real data set.

Prediction accuracy in the simulated data

Table 1 shows the results for both scenarios (S1 and S2) described in the simulation study for the BPMTME model and for the Bayesian Poisson multi-environment (BPME) model (univariate version of the BPMTME model). Since for each scenario we studied the prediction accuracy of the models using 10 random partitions for each of the five training-testing percentages, we provided the average of the Spearman correlation of the 10 random partitions for each testing percentage (10,

20, 30, 40, and 50%) for each scenario and each model (BPMTME and BPME models). First, we compare, for each scenario (S1), the prediction accuracies for each testing percentage between the BPMTME model and the BPME model. Then we compare the prediction accuracy for each testing percentage between the two scenarios.

The first comparison is important since we hypothesized that the BPMTME model produces better predictions than the BPME model. Under S1, on average, the BPMTME model produced better predictions than the BPME model. For the 10% testing percentage, the BPMTME model was on average 17.89% better than the BPME model. When the testing percentages were 20 and 30%, the BPMTME model was on average 2.1% better than the BPME model. When the testing percentage was 40%, the BPMTME model was on average 9.16% better than the BPME model, and when the testing percentage was 50%, the BPMTME model was on average 1.4% better than the BPME model. It is also important to point out that for trait–environment combination 12, the best model was the BPMTME model for the five testing percentages; the superiority of this model over the BPME model in the five cases was >27%.

For the second scenario (S2) and the 10% testing percentage, the BPMTME model was on average 5.16% better than the BPME model, whereas for the 20, 30, 40, and 50% testing percentages, the BPMTME model was on average 3.17, 21.75, 10.75, and 0.9% better than the BPME model, respectively.

When we compare the prediction accuracies of scenarios S1 and S2, we see that there are no significant differences between the two scenarios, although, on average, for all the testing percentages the BPME model was

■ **Table 2 The posterior means (Mean) and SD of parameter estimates of $\boldsymbol{\beta}^*$, Σ_t, Σ_c, and Σ_E of the BPMTME model**

Coefficient Estimates of $\boldsymbol{\beta}^*$			Estimates of Σ_t and Σ_c			Estimates of Σ_E		
Parameter	Mean	SD	Parameter	Mean	SD	Parameter	Mean	SD
$\hat{\beta}_{11}^*$	0.018	0.119	$\hat{\Sigma}_{t11}$	0.393	0.086	$\hat{\Sigma}_{E11}$	0.0198	0.0416
$\hat{\beta}_{12}^*$	−0.151	0.133	$\hat{\Sigma}_{t12}$	0.367	0.063	$\hat{\Sigma}_{E12}$	0.0014	0.0079
$\hat{\beta}_{21}^*$	1.806	0.099	$\hat{\Sigma}_{t22}$	0.381	0.069	$\hat{\Sigma}_{E13}$	3.92E−05	0.0003
$\hat{\beta}_{22}^*$	1.680	0.102	$\hat{\Sigma}_{c11}$	0.702	0.058	$\hat{\Sigma}_{E22}$	0.003	0.0041
$\hat{\beta}_{31}^*$	2.806	0.097	$\hat{\Sigma}_{c12}$	0.102	0.039	$\hat{\Sigma}_{E23}$	1.45E−05	0.0002
$\hat{\beta}_{32}^*$	2.751	0.091	$\hat{\Sigma}_{c22}$	0.726	0.060	$\hat{\Sigma}_{E33}$	4.94E−05	0.0001

SD denotes standard deviations.

a little better than BPMTME. To make sense of this result, we need to remember that under scenario S1, the correlation of the three variance-covariance matrices (Σ_t, Σ_E, and Σ_c) was 0.8, while under scenario S2 the correlation of these three covariance matrices was 0.30.

Estimation and prediction accuracy in the real data set

Table 2 gives the parameter estimates of the real data set. The β coefficients between traits and between environments are different, which implies that the counts between environments and between traits are different. The larger counts are observed in environment 3 and the lower counts are observed in environment 1; also, the counts in trait 1 are a little larger in comparison to trait 2. The variances for each trait (that belong to Σ_t) are very similar (0.393 for trait 1 and 0.381 for trait 2) and the correlation between these two traits is 0.948 (with a covariance of 0.367 in Σ_t); that is, the correlation between traits is high since the resulting correlation between traits is close to one. The variances of traits estimated under the variance-covariance matrix (Σ_c) were 0.702 for trait 1 and 0.726 for trait 2, with a correlation of 0.1418 (and a covariance of 0.102); in this case, the correlation between traits that is not explained by Σ_t is very low (0.1418). However, the variances of environments were very low: 0.0198, 0.003, and 4.94E−05 for environments 1, 2, and 3, respectively (obtained from Σ_E). The correlations between environments were also very low since the correlation between environments 1 and 2 was 0.182 (with covariance 0.0014 from Σ_E), the correlation between environments 1 and 3 was 0.039 (with covariance 3.92E−05 from Σ_E), and the correlation between environments 2 and 3 was 0.037 (with covariance 1.45E−05, from Σ_E). Σ_c is the general variance-covariance matrix of traits that were not explained by variance-covariance matrix Σ_t.

Table 3 shows that, in general, there are no significant differences in terms of prediction accuracy between the two models (BPMTME and BPME); however, it is very interesting that predictions are pretty high under both models, since in all cases the predictions are >0.5. For example, under both models the lowest predictions were for trait–environment combination 11, while the best predictions were for trait–environment combinations 13 and 23.

These results can be explained in part because the correlation observed between traits in correlation matrix Σ_c was very low (0.1418), and because the correlation observed in the correlation matrix between environments was also low (see Table 2). However, although the correlation between traits observed in correlation matrix Σ_t was very high (0.948), this alone did not contribute to superior predictions under the BPMTME model in comparison to the BPME model. It is important to point out that these results with the real data do not indicate that the BPMTME model is not useful, but rather that there is no advantage in using this model over the BPME model when the correlation structures of traits (Σ_t, Σ_c) and environments (Σ_E) are not high. Another possible explanation of why the BPMTME model did not

outperform the BPME model is that the number of wheat lines in this data set is limited (only 182) and only 18 (10%) of them are predicted in some environments. When our proposed BPMTME model is used with limited data sets, it is likely to capture idiosyncrasies of the data rather than the true underlying processes; this problem may be entirely due to limitations the data impose on our ability to detect the underlying processes, rather than to any inherent value of simple models. Although the results in Table 3 do not favor the BPMTME model, we can argue that the proposed multivariate model has a low risk of doing worse in terms of prediction accuracy. For this reason, we are convinced that when dealing with prediction of count phenotypes for multiple traits and multiple environments, the savviest solution is to run both models and choose the one with better prediction performance. Another possible reason for the difference in performance of both models may be the presence of more zeros than our model can support in the phenotypes under study, since Figure 1 shows that for trait–environment combinations, Trait1_Environment1, Trait1_Environment2, Trait2_Environment1, and Trait2_Environment2, there is overdispersion for the excess zeros. In the statistical literature, zero-inflated and hurdle models have been proposed for coping with zero-inflated outcome data with or without overdispersion. For this reason, as one reviewer suggested, the proposed models for multiple-traits and multiple-environments can be expanded to deal with the high occurrence of zeros in the observed dependent variables to end up with zero-inflated and hurdle versions for multiple-traits and multiple-environments.

DISCUSSION

In this paper, we propose a Poisson-lognormal multivariate model that takes into account the $T \times E$, $T \times G$, and $T \times G \times E$ interaction terms, and helps to improve prediction capacity in comparison to the univariate Poisson-lognormal model, which only takes into account the $G \times E$ interaction term. The proposed model is original in the sense that an

■ **Table 3 Prediction accuracy of the real data set measured based on the average of the 10 random partitions using the Spearman correlation (ASC) for each testing (tst) percentage for the BPMTME and BPME models in bold are the best predictions.**

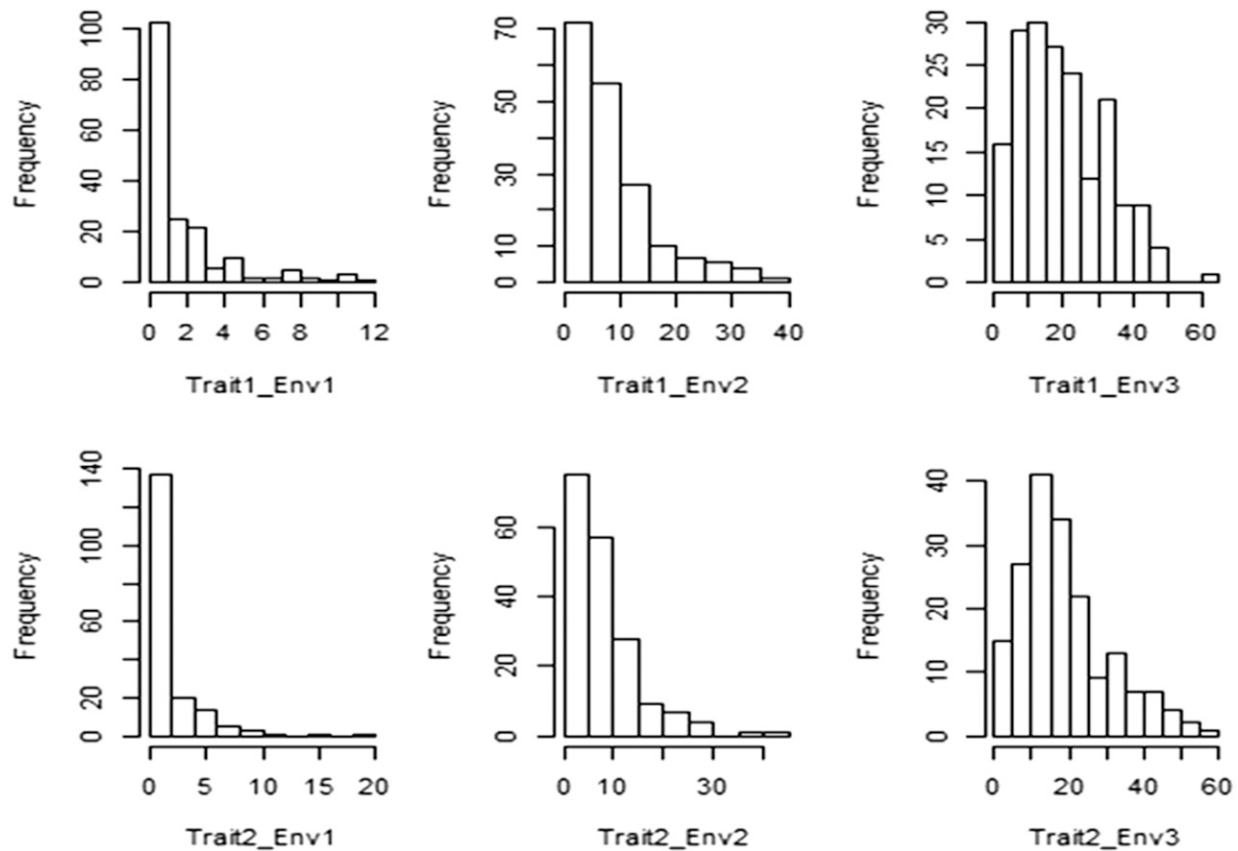| Trait–Env	BPMTME		BPME	
	Mean	SE	Mean	SE
11	0.5008	0.0666	**0.5145**	0.0595
21	0.5309	0.0375	**0.5814**	0.0290
12	0.6805	0.0287	**0.7064**	0.0383
22	**0.7025**	0.0462	0.6715	0.0394
13	0.7447	0.0158	**0.7672**	0.0195
23	**0.7576**	0.0262	0.7187	0.0214
Average	0.6528	0.0368	0.6600	0.0345

SE denotes the standard error.

Figure 1 Histogram of count frequencies of the real data set for the two traits under study for each environment.

exact Gibbs sampler was obtained, which makes this model more efficient than existing Poisson-lognormal multivariate models that do not take into account interaction terms and use the Metropolis-Hastings algorithm to estimate parameters. Also, the lognormal random effect $\exp(\boldsymbol{c})$ that was introduced in the linear predictor allows having an approximate Poisson-lognormal model. The advantage of introducing these lognormal random effects was twofold: (1) accounting for overdispersion (more than the Poisson variance) and (2) exploiting the correlation between traits that was not captured by the variance-covariance matrix of traits $(\boldsymbol{\Sigma}_t)$. We introduced a general covariance structure between the traits for the extra random effect $(\boldsymbol{\Sigma}_c)$, which plays a really important role because it also allows borrowing information between traits. Additionally, the proposed model also has a general covariance structure between environments which also allows borrowing information between environments. It is important to point out that if a gamma distribution was assumed for the random effect $\boldsymbol{c}$, this would produce a multivariate negative binomial model which, in addition to being very inefficient for estimating the scale parameter, only allows positive correlation structures because the NB model (univariate and multivariate) was motivated solely for mathematical convenience.

Another advantage of the proposed BPMTME model is that it can be used for performing univariate analysis employing the predictor given in Equation (1), using the appropriate design matrices of environments, lines, and genotype by environment. This alternative is really helpful since in addition to implementing the multiple-trait and multi-environment model for count data, it also allows implementing a univariate multiple-environment model with the same proposed code under a Poisson-lognormal model. Here it is important to point out that the resulting univariate

Poisson-lognormal model is different from the Poisson and NB models with the $G \times E$ interaction term proposed by Montesinos-López *et al.* (2016a), given that the models with $G \times E$ proposed by these authors do not have the normal random effect $c_{ijk}^{(l)}$. For this reason, a comparison with this model was not included. However, according to the existing statistical literature, there is evidence that when there is overdispersion, the Poisson-lognormal model does a better job than the Poisson only model.

According to our results, the proposed model may be useful for breeders interested in modeling more than one count trait in many environments simultaneously. Although in the simulation study the BPMTME model was better in terms of prediction accuracy than the BPME model, in the real data set both models had similar prediction performance. Although the proposed BPMTME model showed no superiority in the real data set, we believe that when the correlations between traits and between environments are high, the proposed model should be superior. However, more empirical evidence is necessary to be sure that the proposed model helps to increase prediction accuracy under these circumstances.

As previously mentioned, our results with the real data sets indicated that the univariate model was slightly better than our multivariate model in terms of prediction accuracy. This result, which was not expected, has been documented in areas like econometrics where multivariate models have been better at modeling interdependencies and achieve better fit. Within a given sample, it has been found that univariate methods outperform multivariate methods in sample predictions. Some of the reasons that may explain these results are given here. (a) Multivariate models have more parameters than univariate models (parsimonious). Every additional parameter is an unknown quantity and has to be

estimated. This estimation brings an additional source of error due to sampling variation. (b) The number of potential candidates for multivariate models exceeds its univariate counterpart. Model selection is therefore more complex, lengthier, and more prone to errors, which affect prediction. (c) It is difficult to generalize nonlinear procedures in the multivariate case. Generally, multivariate models must have a simpler structure than univariate ones to overcome the additional complexity of being multivariate. For example, while a researcher may use a nonlinear model for univariate data, she/he may refrain from using the multivariate counterpart or such a generalization may not have been developed. Then, multivariate models will miss the nonlinearities that are handled properly by the univariate models. (d) Outliers can have a more serious effect on multivariate predictions than on univariate predictions. Moreover, it is easier to spot and control outliers in the univariate context, *etc.* For these reasons, we believe it is necessary to explore other approaches to prediction when dealing with multivariate responses (count and normal phenotypes) and multiple environments, for when the number of traits, lines, and environments increased significantly even with normal phenotypes, the implementation of this model became very difficult.

Finally, although the proposed model has some advantages, it also has some disadvantages. The main disadvantage of the proposed BPMTME model is that it is computationally demanding. Although we implemented this model in C++ code, when the number of lines, environments, and traits grows substantially, the time required for its implementation increases greatly and can be untenable for large numbers. The second disadvantage is that our proposed BPMTME model is in reality an approximation to the true BPMTME model: instead of working directly with the Poisson model, we approximated it with the negative NB distribution using $r = 1000$ as a scale parameter, which works very well for small counts, as was shown by Montesinos-López *et al.* (2016a). Thanks to this approximation, it was possible to obtain an exact Gibbs sampler that is more powerful for complex and large data sets.

Conclusions

In this paper, we propose a multiple-trait, multiple-environment model for count data under an approximate Bayesian Poisson-lognormal model. The Gibbs sampler obtained is an approximation since the Poisson-lognormal model is approximated using the NB distribution in conjunction with a lognormal random effect. The proposed BPMTME model was compared with its univariate counterpart (BPME), and in the simulation examples, the proposed multivariate model (BPMTME) was considerably better in terms of prediction accuracy. This can be expected when there is a moderate or large correlation between traits, which allows borrowing information between traits. However, the prediction accuracy of the proposed multivariate model was not superior to that of the univariate model with the real data set. The proposed BPMTME model allows borrowing information between environments because it takes into account a general covariance matrix between environments. For these reasons, we believe that more empirical work is needed to determine whether the proposed BPMTME model is really a better tool than the BPME model tool for breeders interested in simultaneously improving more than one count trait in many environments. However, although an exact Gibbs sampler was

proposed and the program was implemented in C++, it is computationally demanding and more work is required for increasing its speed. Nonetheless, we are convinced that this limitation will disappear or diminish considerably in the coming years as computer science is improving very quickly.

LITERATURE CITED

Cavanagh, C. R., S. Chao, S. Wang, B. E. Huang, S. Stephen *et al.*, 2013 Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. USA 110(20): 8057–8062.

Eddelbuettel, D., and C. Sanderson, 2014 Rcpparmadillo: accelerating R with high-performance C++ linear algebra. Comput. Stat. Data Anal. 71: 1054–1063 10.1016/j.csda.2013.02.005.

Harville, D. A., 2014 The need for more emphasis on prediction: a "non-denominational" model-based approach. Am. Stat. 68(2): 71–83.

Henderson, C. R., and R. L. Quaas, 1976 Multiple trait evaluation using relatives' records. J. Anim. Sci. 43(6): 1188–1197.

Jiang, J., Q. Zhang, L. Ma, J. Li, Z. Wang *et al.*, 2015 Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. Heredity 115(1): 29–36.

Ma, J., K. M. Kockelman, and P. Damien, 2008 A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accid. Anal. Prev. 40(3): 964–975.

Montesinos-López, O. A., A. Montesinos-López, P. Pérez-Rodríguez, K. Eskridge, X. He *et al.*, 2015 Genomic prediction models for count data. J. Agric. Biol. Environ. Stat. 20(4): 533–554.

Montesinos-López, A., O. A. Montesinos-López, J. Crossa, J. Burgueño, K. Eskridge *et al.*, 2016a Genomic bayesian prediction model for count data with genotype × environment interaction. G3: 6(5): 1165–1177.

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

Park, T., and D. A. van Dyk, 2009 Partially collapsed Gibbs samplers: illustrations and applications. J. Comput. Graph. Stat. 18: 283–305.

Pollak, E. J., J. Van der Werf, and R. L. Quaas, 1984 Selection bias and multiple trait evaluation. J. Dairy Sci. 67(7): 1590–1595.

Polson, N. G., J. G. Scott, and J. Windle, 2013 Bayesian inference for logistic models using Pólya–Gamma latent variables. J. Am. Stat. Assoc. 108: 1339–1349.

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/. Accessed: January 10, 2017.

Schaeffer, L. R., 1984 Sire and cow evaluation under multiple trait models. J. Dairy Sci. 67(7): 1567–1580.

Williams, M. S., and E. D. Ebel, 2012 Methods for fitting the Poisson-lognormal distribution to microbial testing data. Food Control 27(1): 73–80.

Communicating editor: D. J. de Koning

APPENDIX A: JOINT DISTRIBUTION OF ALL PHENOTYPIC COUNT RESPONSES

Conditioning on b_{1j}, b_{2ij}, $\boldsymbol{c}_{ij} = [c_{ij1}, \ldots, c_{ijK}]^T$ and $\boldsymbol{\omega}_{ij} = [\omega_{ij1}, \ldots, \omega_{ijK}]^T$, the joint conditional distribution for all the replications in environment i and genotype j is

$$L_{ij} = \prod_{k=1}^{K} L_{ijk} \propto \prod_{l=1}^{n_T} \exp\left(\sum_{k=1}^{K} \boldsymbol{y}_{ijk}^{*T} \boldsymbol{\eta}_{ijk}^{*} - \frac{1}{2}\sum_{k=1}^{K} \boldsymbol{\eta}_{ijk}^{*T} \boldsymbol{D}_{ijk} \boldsymbol{\eta}_{ijk}^{*} \right) = \exp\left(\boldsymbol{y}_{ij}^{*T} \boldsymbol{\eta}_{ij}^{*} - \frac{1}{2} \boldsymbol{\eta}_{ij}^{*T} \boldsymbol{D}_{ij} \boldsymbol{\eta}_{ij}^{*} \right)$$

where $\boldsymbol{y}_{ij}^{*} = \left[\boldsymbol{y}_{ij1}^{*T}, \cdots, \boldsymbol{y}_{ijK}^{*T}\right]^T$, $\boldsymbol{\eta}_{ij}^{*} = \left[\boldsymbol{\eta}_{ij1}^{*T}, \cdots, \boldsymbol{\eta}_{ijK}^{*T}\right]^T$ and $\boldsymbol{D}_{ij} = \mathrm{diag}(\boldsymbol{D}_{ij1}, \cdots, \boldsymbol{D}_{ijK})$. Then, the joint conditional distribution for all the observations in environment i is given by

$$L_{i} = \prod_{j=1}^{J} L_{ij} \propto \exp\left(\boldsymbol{y}_{i}^{*T} \boldsymbol{\eta}_{i}^{*} - \frac{1}{2} \boldsymbol{\eta}_{i}^{*T} \boldsymbol{D}_{i} \boldsymbol{\eta}_{i}^{*} \right)$$

where $\boldsymbol{y}_{i}^{*} = \left[\boldsymbol{y}_{i1}^{*T}, \cdots, \boldsymbol{y}_{iJ}^{*T}\right]^T$, $\boldsymbol{\eta}_{i}^{*} = \left[\boldsymbol{\eta}_{i1}^{*T}, \cdots, \boldsymbol{\eta}_{iJ}^{*T}\right]^T$, and $\boldsymbol{D}_{i} = \mathrm{diag}(\boldsymbol{D}_{i1}, \cdots, \boldsymbol{D}_{iJ})$. Finally, the joint conditional distribution for all the observations is

$$L = \prod_{i=1}^{I} L_{i} \propto \exp\left(\boldsymbol{y}^{*T} \boldsymbol{\eta}^{*} + \frac{1}{2} \boldsymbol{\eta}_{i}^{*T} \boldsymbol{D} \ \boldsymbol{\eta}^{*} \right)$$

where $\boldsymbol{y}* = \left[\boldsymbol{y}_{1}^{*T}, \cdots, \boldsymbol{y}_{I}^{*T}\right]^T$, $\boldsymbol{\eta}_{i}^{*} = \left[\boldsymbol{\eta}_{1}^{*T}, \cdots, \boldsymbol{\eta}_{I}^{*T}\right]^T$, and $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{D}_{1}, \cdots, \boldsymbol{D}_{I})$.

Note that the predictor (1) can be expressed as

$$\eta_{ijk}^{*(l)} = \beta_{i}^{(l)} + b_{1j}^{(l)} + b_{2ij}^{(l)} + c_{ijk}^{(l)} - \log(r) = \beta_{i}^{*(l)} + b_{1j}^{(l)} + b_{2ij}^{(l)} + c_{ijk}^{(l)} = \boldsymbol{x}_{i}^{(l)T}\boldsymbol{\beta}^{*(l)} + b_{1j}^{(l)} + b_{2ij}^{(l)} + c_{ijk}^{(l)}$$

with $\boldsymbol{\beta}^{*(l)T} = \left[\beta_{1}^{*(l)}, \cdots, \beta_{I}^{*(l)}\right]$, $\boldsymbol{x}_{i}^{(l)T} = \left[x_{i1}^{(l)}, \cdots, x_{iI}^{(l)}\right]$ and where $x_{ii'}^{(l)} = 1$ if $i = i'$ and $x_{ii'}^{(l)} = 0$ otherwise. This implies that

$$\boldsymbol{\eta}_{ijk}^{*} = \begin{bmatrix} \eta_{ijk}^{*(1)} \\ \vdots \\ \eta_{ijk}^{*(n_T)} \end{bmatrix} = \boldsymbol{X}_{ijk}\boldsymbol{\beta}^{*} + \boldsymbol{b}_{1j} + \boldsymbol{b}_{2ij} + \boldsymbol{c}_{ijk},$$

where $\boldsymbol{X}_{ijk} = \begin{bmatrix} \boldsymbol{e}_{1}^{T} \otimes \boldsymbol{x}_{i}^{(1)^{T}} \\ \vdots \\ \boldsymbol{e}_{n_T}^{T} \otimes \boldsymbol{x}_{i}^{(n_T)^{T}} \end{bmatrix}$, where $\boldsymbol{e}_{l}^{T}$ is the l-th element of the canonical basis of $\mathbb{R}^{n_T}$. $\boldsymbol{\beta}^{*} = \begin{bmatrix} \boldsymbol{\beta}^{*(1)} \\ \vdots \\ \boldsymbol{\beta}^{*(n_T)} \end{bmatrix}$, $\boldsymbol{b}_{1j} = \begin{bmatrix} b_{1j}^{(1)} \\ \vdots \\ b_{1j}^{(n_T)} \end{bmatrix}$, $\boldsymbol{b}_{2ij} = \begin{bmatrix} b_{2ij}^{(1)} \\ \vdots \\ b_{2ij}^{(n_T)} \end{bmatrix}$, and

$\boldsymbol{c}_{ijk} = \begin{bmatrix} c_{ijk}^{(1)} \\ \vdots \\ c_{ijk}^{(n_T)} \end{bmatrix}$. From here,

$$\boldsymbol{\eta}_{ij}^{*} = \begin{bmatrix} \boldsymbol{\eta}_{ij1}^{*} \\ \vdots \\ \boldsymbol{\eta}_{ijK}^{*} \end{bmatrix} = \boldsymbol{X}_{ij}\boldsymbol{\beta}^{*} + \boldsymbol{Z}_{1ij}\boldsymbol{b}_{1j} + \boldsymbol{Z}_{2ij}\boldsymbol{b}_{2ij} + \boldsymbol{c}_{ij},$$

where $\boldsymbol{X}_{ij} = \begin{bmatrix} \boldsymbol{X}_{ij1} \\ \vdots \\ \boldsymbol{X}_{ijK} \end{bmatrix}$, $\boldsymbol{Z}_{1ij} = \boldsymbol{1}_K \otimes \boldsymbol{I}_{n_T}$, $\boldsymbol{Z}_{2ij} = \boldsymbol{1}_K \otimes \boldsymbol{I}_{n_T}$, $\boldsymbol{c}_{ij} = \begin{bmatrix} \boldsymbol{c}_{ij1} \\ \vdots \\ \boldsymbol{c}_{ijK} \end{bmatrix}$, and $\boldsymbol{1}_K$ is a column vector of ones of dimension K.

This implies that

$$\boldsymbol{\eta}_{i}^{*} = \begin{bmatrix} \boldsymbol{\eta}_{i1}^{*} \\ \vdots \\ \boldsymbol{\eta}_{iJ}^{*} \end{bmatrix} = \boldsymbol{X}_{i}\boldsymbol{\beta}^{*} + \boldsymbol{Z}_{1i}\boldsymbol{b}_{1} + \boldsymbol{Z}_{2i}\boldsymbol{b}_{2i} + \boldsymbol{c}_{i}$$

where $\boldsymbol{X}_{i} = \begin{bmatrix} \boldsymbol{X}_{i1} \\ \vdots \\ \boldsymbol{X}_{iJ} \end{bmatrix}$, $\boldsymbol{Z}_{1i} = \begin{bmatrix} \boldsymbol{Z}_{1i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{Z}_{1iJ} \end{bmatrix}$, $\boldsymbol{Z}_{2i} = \begin{bmatrix} \boldsymbol{Z}_{2i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{Z}_{2iJ} \end{bmatrix}$, $\boldsymbol{b}_{1} = \begin{bmatrix} \boldsymbol{b}_{11} \\ \vdots \\ \boldsymbol{b}_{1J} \end{bmatrix}$, $\boldsymbol{b}_{2i} = \begin{bmatrix} \boldsymbol{b}_{2i1} \\ \vdots \\ \boldsymbol{b}_{2iJ} \end{bmatrix}$, and $\boldsymbol{c}_{i} = \begin{bmatrix} \boldsymbol{c}_{ij} \\ \vdots \\ \boldsymbol{c}_{iJ} \end{bmatrix}$. Then

$$\boldsymbol{\eta}^{*} = \begin{bmatrix} \boldsymbol{\eta}_{1}^{*} \\ \vdots \\ \boldsymbol{\eta}_{I}^{*} \end{bmatrix} = \boldsymbol{X}\boldsymbol{\beta}^{*} + \boldsymbol{Z}_{1}\boldsymbol{b}_{1} + \boldsymbol{Z}_{2}\boldsymbol{b}_{2} + \boldsymbol{c}$$

where $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{1} \\ \vdots \\ \boldsymbol{X}_{I} \end{bmatrix}$, $\boldsymbol{Z}_{1} = \begin{bmatrix} \boldsymbol{Z}_{11} \\ \vdots \\ \boldsymbol{Z}_{1I} \end{bmatrix}$, $\boldsymbol{Z}_{2} = \begin{bmatrix} \boldsymbol{Z}_{21} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{Z}_{2I} \end{bmatrix}$, $\boldsymbol{b}_{2} = \begin{bmatrix} \boldsymbol{b}_{21} \\ \vdots \\ \boldsymbol{b}_{2I} \end{bmatrix}$, and $\boldsymbol{c} = \begin{bmatrix} \boldsymbol{c}_{1} \\ \vdots \\ \boldsymbol{c}_{I} \end{bmatrix}$. Finally,

$$L = \exp(y^{*T}\boldsymbol{\eta}^* - \boldsymbol{\eta}^{*T}D\boldsymbol{\eta}^*) \propto \exp\left(y^{*T}(X\boldsymbol{\beta}^* + Z_1b_1 + Z_2b_2 + c) - \frac{1}{2}(X\boldsymbol{\beta}^* + Z_1b_1 + Z_2b_2 + c)^T D(X\boldsymbol{\beta}^* + Z_1b_1 + Z_2b_2 + c)\right).$$

APPENDIX B: DERIVING FULL CONDITIONAL DISTRIBUTION FOR ALL PARAMETERS

Full conditional for $\boldsymbol{\beta}^*$

$$f(\boldsymbol{\beta}^*|-) \propto \exp\left(y^{*T}(X\boldsymbol{\beta}^*) - \frac{1}{2}(X\boldsymbol{\beta}^* + Z_1b_1 + Z_2b_2 + c)^T D(X\boldsymbol{\beta}^* + Z_1b_1 + Z_2b_2 + c) - \frac{1}{2}(\boldsymbol{\beta}^* - \boldsymbol{\beta}_v^*)^T \Sigma_v^{-1}(\boldsymbol{\beta}^* - \boldsymbol{\beta}_v^*)\right)$$

$$\propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}^{*T}(\Sigma_v^{-1} + X^T DX)\boldsymbol{\beta}^* - \frac{1}{2}(-2)\left[\boldsymbol{\beta}_v^{*T}\Sigma_v^{-1} + y^{*T}X - (Z_1b_1 + Z_2b_2 + c)^T DX\right]\boldsymbol{\beta}^*\right)$$

$$\propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}^{*T}(\Sigma_v^*)^{-1}\boldsymbol{\beta}^* - \frac{1}{2}(-2)\boldsymbol{\beta}_v^{**T}(\Sigma_v^*)^{-1}\boldsymbol{\beta}^*\right)$$

$$\propto \exp\left[\frac{1}{2}(\boldsymbol{\beta}^* - \boldsymbol{\beta}_v^{**})^T(\Sigma_v^*)^{-1}(\boldsymbol{\beta}^* - \boldsymbol{\beta}_v^{**})\right] \sim N(\boldsymbol{\beta}_v^{**}, \Sigma_v^*) \qquad (B1)$$

where $\Sigma_v^* = (\Sigma_v^{-1} + X^T DX)^{-1}$ and $\boldsymbol{\beta}_v^{**} = \Sigma_v^*[\Sigma_v^{-1}\boldsymbol{\beta}_v^* + X^T y^* - X^T D(Z_1b_1 + Z_2b_2 + c)]$

Full conditional for b_1

$$f(b_1|-) \propto \exp\left(y^{*T}Z_1b_1 - \frac{1}{2}\boldsymbol{\eta}^{*T}D\boldsymbol{\eta}^* - \frac{1}{2}b_1^T G_1^{-1}b_1\right)$$

$$\propto \exp\left(-\frac{1}{2}b_1^T(G_1^{-1} + Z_1^T DZ_1)b_1 - \frac{1}{2}(-2)\left[y^{*T}Z_1 - (X\boldsymbol{\beta}* + Z_2b_2 + c)^T DZ_1\right]b_1\right)$$

$$\propto \exp\left(-\frac{1}{2}b_1^T(G_1^*)b_1 - \frac{1}{2}(-2)b_1(G_1^*)^{-1}b_1\right)$$

$$\propto \exp\left(-\frac{1}{2}(b_1 - b_1^*)^T(G_1^*)^{-1}(b_1 - b_1^*)\right) \sim N(b_1^*, G_1^*) \qquad (B2)$$

where $G_1^* = (G_1^{-1} + Z_1^T DZ)^{-1}$ and $b_1^* = G_1^*(Z_1^T y^* - Z_1^T D(X\boldsymbol{\beta}^* + Z_2b_2 + c))$.

Full conditional for b_2

In a similar way as the full conditional of b_1, then

$$b_2| \sim N(b_2^*, G_2^*) \qquad (B3)$$

where $G_2^* = (G_2^{-1} + Z_2^T DZ_2)^{-1}$ and $b_2^* = G_2^*(Z_2^T y^* - Z_2^T D(X\boldsymbol{\beta}^* + Z_1b_1 + c))$.

Full conditional for c

First note that c is equal to $c = [c_1 \quad \ldots \quad c_I]^T = [c_{11}\ldots c_{1J}\ldots c_{I1} \quad \ldots \quad c_{IJ}]^T$, and we need to remember that $c_{ijk} = \begin{bmatrix} c_{ijk}^{(1)} & \ldots & c_{ijk}^{(n_T)} \end{bmatrix}^T \sim N(0, \Sigma_c)$, for all i, j, and k. This implies that $c_{ij} = \begin{bmatrix} c_{ij1} & \ldots & c_{ijK} \end{bmatrix}^T \sim N(0, I_K \otimes \Sigma_c)$, $c_i = \begin{bmatrix} c_{i1} & \ldots & c_{iJ} \end{bmatrix}^T \sim N(0, I_{KJ} \otimes \Sigma_c)$, and $c = \begin{bmatrix} c_1 & \ldots & c_I \end{bmatrix}^T \sim N(0, \Sigma_{c(a)})$, with $\Sigma_{c(a)} = I_{KJI} \otimes \Sigma_c$. Therefore,

$$f(c|-) \propto \exp\left(y^{*T}c - \frac{1}{2}\boldsymbol{\eta}^{*T}D\boldsymbol{\eta}^* - \frac{1}{2}c^T\Sigma_{c(a)}^{-1}c\right)$$

$$\propto \exp\left(-\frac{1}{2}c^T(\Sigma_{c(a)}^{-1} + D)c - \frac{1}{2}(-2)\left[y^{*T} - \left(X\boldsymbol{\beta}^* + Z_1b_1 + Z_2b_2\right)^T D\right]c\right) \qquad (B4)$$

$$\propto \exp\left(-\frac{1}{2}\left(c - c^*\right)^T(\Sigma_{c(a)}^*)^{-1}\left(c - c^*\right)\right) \sim N\left(c^*, \Sigma_{c(a)}^*\right)$$

where $\Sigma_{c(a)}^* = (\Sigma_{c(a)}^{-1} + D)^{-1}$ and $c^* = \Sigma_{c(a)}^*[y^* - D(X\boldsymbol{\beta}^* + Z_1b_1 + Z_2b_2)]$

Full conditional for Σ_t

$$f(\Sigma_t|-) \propto f(\mathbf{b}_1|\Sigma_t)\, f(\mathbf{b}_2|\Sigma_E,\Sigma_t)\, f(\Sigma_t|a_t) \propto \frac{1}{|\Sigma_t|^{J/2}}\frac{1}{|\Sigma_t|^{IJ/2}}\exp\left(-\frac{1}{2}\mathbf{b}_1^T \mathbf{G}_1^{-1}\mathbf{b}_1 - \frac{1}{2}\mathbf{b}_2^T \mathbf{G}_2^{-1}\mathbf{b}_2\right) f(\Sigma_t|a_t)$$

$$\propto \frac{1}{|\Sigma_t|^{\frac{I+IJ}{2}}}\exp\left(-\frac{1}{2}tr\left(\Sigma_t^{-1}\left(\mathbf{B}_1^*\mathbf{G}^{-1}\mathbf{B}_1^{*T} + \mathbf{B}_2^*\mathbf{G}_3^{-1}\mathbf{B}_2^{*T}\right)\right)\right)|\mathbf{C}_t|^{\frac{\nu_t+n_t-1}{2}}|\Sigma_t|^{\frac{\nu_t+n_t}{2}}\exp\left(-\frac{1}{2}tr(\mathbf{C}_t\Sigma_t^{-1})\right)$$

$$\propto |\Sigma_t|^{-\frac{\nu_t+J(1+I)+2n_t}{2}}\exp\left(-\frac{1}{2}tr(\mathbf{C}_t^*\Sigma_t^{-1})\right)$$

$$\sim IW(\nu = \nu_t + (1+I)J + n_T - 1, \mathbf{C}_t^*) \tag{B5}$$

where $\mathbf{C}_t^* = \mathbf{C}_t + \mathbf{B}_1^*\mathbf{G}_g^{-1}\mathbf{B}_1^{*T} + \mathbf{B}_2^*\mathbf{G}_3^{-1}\mathbf{B}_2^{*T}$ and $\mathbf{C}_t = 2\nu_t \text{diag}\left(\frac{1}{a_{1t}} \cdots \frac{1}{a_{n_T t}}\right)$, and $\mathbf{B}_1^*$, $\mathbf{B}_2^*$, and $\mathbf{G}_3^{-1}$ are matrices defined below. To see this, note that $\mathbf{b}_1^T\mathbf{G}_1^{-1}\mathbf{b}_1 = \|\mathbf{G}_1^{-1/2}\mathbf{b}_1\|^2$ and $\mathbf{G}_1^{-1/2} = \mathbf{G}^{-1/2}\otimes\Sigma_t^{-(1/2)}$. Using the property $(\mathbf{B}^T\otimes\mathbf{A})\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{AXB})$, we have that

$$\mathbf{G}_1^{-1/2}\mathbf{b}_1 = \left(\mathbf{G}^{-1/2}\otimes\Sigma_t^{-(1/2)}\right)\mathbf{b}_1 = \left(\mathbf{G}^{-1/2}\otimes\Sigma_t^{-(1/2)}\right)\text{vec}(\mathbf{B}_1^*) = \text{vec}\left(\Sigma_t^{-1/2}\mathbf{B}_1^*\mathbf{G}^{-1/2}\right) = \begin{bmatrix}\Sigma_t^{-(1/2)}\mathbf{d}_1 \\ \vdots \\ \Sigma_t^{-1/2}\mathbf{d}_J\end{bmatrix}, \quad \text{where} \quad \mathbf{B}_1^* = [\mathbf{b}_{11}\cdots\mathbf{b}_{1J}] \quad \text{and}$$

$\mathbf{d}_j$, $j = 1, \ldots, J$, are the columns of $\mathbf{B}_1^*\mathbf{G}^{-1/2}$. Then $\mathbf{b}_1^T\mathbf{G}_1^{-1}\mathbf{b}_1 = \sum_{j=1}^{J}\mathbf{d}_j^T\Sigma_t^{-1/2}\Sigma_t^{-1/2}\mathbf{d}_j = tr\left(\Sigma_t^{-1}\sum_{j=1}^{J}\mathbf{d}_j\mathbf{d}_j^T\right) = tr\left(\Sigma_t^{-1}\right.$
$\left.\mathbf{B}_1^*\mathbf{G}^{-1/2}\mathbf{G}^{-1/2}\mathbf{B}_1^{*T}\right) = tr(\mathbf{B}_1^*\mathbf{G}^{-1}\mathbf{B}_1^{*T}\Sigma_t^{-1})$. Defining $\mathbf{G}_3 = \Sigma_E\otimes\mathbf{G}$, in a similar way, we see that

$$\mathbf{b}_2^T\mathbf{G}_2^{-1}\mathbf{b}_2 = tr(\Sigma_t^{-1}\mathbf{B}_2^*\mathbf{G}_3^{-1}\mathbf{B}_2^{*T}) = tr(\mathbf{B}_2^*\mathbf{G}_3^{-1}\mathbf{B}_2^{*T}\Sigma_t^{-1}),$$

where $\mathbf{B}_2^* = [\mathbf{b}_{211},\cdots,\mathbf{b}_{21J},\cdots,\mathbf{b}_{2I1},\cdots,\mathbf{b}_{2IJ}]$.

Full conditional for a_{lt}

$$f(a_{lt}|ELSE) \propto f(\Sigma_t|a_{lt})\, f(a_{lt}) \propto \frac{1}{(a_l)^{\frac{\nu_t+n_T}{2}+1}}\exp\left(-\frac{\frac{1}{A_{lt}^2}+\nu_t(\Sigma_t^{-1})_{ll}}{a_{lt}}\right)$$

$$\propto IG\left(\frac{\nu_t+n_T}{2}, 1/A_{lt}^2 + \nu_t(\Sigma_t^{-1})_{ll}\right) \tag{B6}$$

Full conditional for Σ_E

$$f(\Sigma_E|-) \propto \frac{1}{|\Sigma_E|^{\frac{n_TJ}{2}}}\exp\left(-\frac{1}{2}tr(\mathbf{B}_2^{*T}\mathbf{G}_1^{-1}\mathbf{B}_3^*\Sigma_E^{-1})\right)\frac{1}{|\Sigma_E|^{\frac{\nu_E+2I}{2}}}\exp\left[-\frac{1}{2}tr(\mathbf{C}_E\Sigma_E^{-1})\right] \propto \frac{1}{|\Sigma_E|^{\frac{\nu_E+n_TJ+2I}{2}}}\exp\left[-\frac{1}{2}tr(\mathbf{C}_E^*\Sigma_E^{-1})\right] \propto |\Sigma_E|^{-\frac{\nu_E+n_TJ+I+I-1+1}{2}}\exp\left[-\frac{1}{2}tr(\mathbf{C}_E^*\Sigma_E^{-1})\right]$$

$$\sim IW(\nu_E + n_TJ + I - 1, \mathbf{C}_E^*) \tag{B7}$$

where $\mathbf{C}_E^* = \mathbf{C}_E + \mathbf{B}_3^{*T}\mathbf{G}_1^{-1}\mathbf{B}_3^*$ and $\mathbf{B}_3^*$ is defined below. First, this is because $\mathbf{b}_2^T(\Sigma_E^{-1}\otimes\mathbf{G}_1^{-1})\mathbf{b}_2 = \|(\Sigma_E^{-1/2}\otimes\mathbf{G}_1^{-1/2})\mathbf{b}_2\|^2$ and $(\Sigma_E^{-1/2}\otimes\mathbf{G}_1^{-1/2})$ $\mathbf{b}_2 = (\Sigma_E^{-1/2}\otimes\mathbf{G}_1^{-1/2})\text{vec}(\mathbf{B}_3^*) = \text{vec}(\mathbf{G}_1^{-1/2}\mathbf{B}_2^*\Sigma_E^{-1/2})$, where $\mathbf{B}_3^*$ is a matrix with I columns such that $\text{vec}(\mathbf{B}_3^*) = \mathbf{b}_2$. Then, expressing $\mathbf{G}_1^{-1/2}\mathbf{B}_3^*\Sigma_E^{-1/2} = [\mathbf{f}_1,\cdots,\mathbf{f}_I]$, $\mathbf{b}_2^T\mathbf{G}_1^{-1}\mathbf{b}_2 = \sum_{i=1}^{I}\mathbf{f}_i^T\mathbf{f}_i = \sum_{i=1}^{I}\mathbf{e}_i^T\Sigma_E^{-1/2}\mathbf{B}_3^{*T}\mathbf{G}_1^{-1/2}\mathbf{G}_1^{-1/2}\mathbf{B}_3^*\Sigma_E^{-1/2}\mathbf{e}_i = \sum_{i=1}^{I}\mathbf{e}_i^T\Sigma_E^{-1/2}\mathbf{B}_3^{*T}\mathbf{G}_1^{-1}\mathbf{B}_3^*\Sigma_E^{-1/2}\mathbf{e}_i = tr(\Sigma_E^{-1/2}$
$\mathbf{B}_3^{*T}\mathbf{G}_1^{-1}\mathbf{B}_3^*\Sigma_E^{-1/2}) = tr(\mathbf{B}_2^{*T}\mathbf{G}_1^{-1}\mathbf{B}_3^*\Sigma_E^{-1})$.

Full conditional for a_{lE}

$$f(a_{lE}|ELSE) \propto f(\Sigma_E|a_{lE})\, f(a_{lE}) \propto \frac{1}{(a_{lE})^{\frac{\nu_E+I}{2}+1}}\exp\left(-\frac{\frac{1}{A_{lE}^2}+\nu_E(\Sigma_E^{-1})_{ll}}{a_{lE}}\right)$$

$$\propto IG\left(\frac{\nu_E+I}{2}, 1/A_{lE}^2 + \nu_E(\Sigma_E^{-1})_{ll}\right) \tag{B8}$$

Full conditional for $\omega_{ijk}^{(l)}$

$$f\left(\omega_{ijk}^{(l)}\middle|ELSE\right) \sim PG\left(b = y_{ijk}^{(l)} + r, d = \eta_{ijk}^{*(l)}\right) \tag{B9}$$

Full conditional for Σ_c

$$f(\Sigma_c|-) \propto \frac{1}{|\Sigma_c|^{\frac{KJI}{2}}} \exp\left[-\frac{1}{2}\boldsymbol{c}^T(\boldsymbol{I}_{KJI} \otimes \Sigma_c^{-1})\boldsymbol{c}\right]f(\Sigma_c)$$

$$\propto \frac{1}{|\Sigma_c|^{\frac{KJI}{2}}} \exp\left[-\frac{1}{2}\boldsymbol{c}^T(\boldsymbol{I}_{KJI} \otimes \Sigma_c^{-1})\boldsymbol{c}\right]\frac{1}{|\Sigma_c|^{\frac{\nu_c+2n_T}{2}}} \exp\left(-\frac{1}{2}tr(\boldsymbol{C}_c\Sigma_c^{-1})\right) \propto \tag{B10}$$

$$|\Sigma_c|^{-\frac{\nu_c+KJI+2n_T}{2}}\exp\left(-\frac{1}{2}tr(\boldsymbol{C}_c^*\Sigma_c^{-1})\right) \sim IW(\nu_c + KJI + n_T - 1, \boldsymbol{C}_c^*)$$

where $\boldsymbol{C}_c^* = \boldsymbol{C}_c + \boldsymbol{B}_4^*\boldsymbol{B}_4^{*T}$, where $\boldsymbol{B}_4^*$ is a matrix with n_T rows such that $\text{vec}(\boldsymbol{B}_4^*) = \boldsymbol{c}$.

Full conditional for a_{lc}

$$f(a_{lc}|ELSE) \propto f(\Sigma_c|a_{lc}) f(a_{lc}) \propto \frac{1}{(a_{lc})^{\frac{\nu_c+I}{2}+1}} \exp\left(-\frac{\frac{1}{A_{lc}^2}+\nu_c(\Sigma_c^{-1})_{ll}}{a_{lc}}\right) \tag{B11}$$

$$\propto IG\left(\frac{\nu_c+I}{2}, 1/A_{lc}^2 + \nu_c(\Sigma_c^{-1})_{ll}\right)$$