A General Bayesian Estimation Method of Linear–Bilinear Models Applied to Plant Breeding Trials With Genotype × Environment Interaction

3 authors:

Sergio Perez-Elizalde
Colegio de Postgraduados
50 PUBLICATIONS · **1,052** CITATIONS

Diego Jarquin
University of Nebraska at Lincoln
79 PUBLICATIONS · **2,060** CITATIONS

Jose Crossa
Consultative Group on International Agricultural Research
637 PUBLICATIONS · **24,595** CITATIONS

Some of the authors of this publication are also working on these related projects:

Nubap proje View project

Seeds of Discovery - MasAgro Biodiversidad View project

A General Bayesian Estimation Method of Linear–Bilinear Models Applied to Plant Breeding Trials With Genotype × Environment Interaction

Sergio Perez-Elizalde, Diego Jarquin, and Jose Crossa

Statistical analyses of two-way tables with interaction arise in many different fields of research. This study proposes the von Mises–Fisher distribution as a prior on the set of orthogonal matrices in a linear–bilinear model for studying and interpreting interaction in a two-way table. Simulated and empirical plant breeding data were used for illustration; the empirical data consist of a multi-environment trial established in two consecutive years. For the simulated data, vague but proper prior distributions were used, and for the real plant breeding data, observations from the first year were used to elicit a prior for parameters of the model for data of the second year trial. Bivariate Highest Posterior Density (HPD) regions for the posterior scores are shown in the biplots, and the significance of the bilinear terms was tested using the Bayes factor. Results of the plant breeding trials show the usefulness of this general Bayesian approach for breeding trials and for detecting groups of genotypes and environments that cause significant genotype × environment interaction. The present Bayes inference methodology is general and may be extended to other linear–bilinear models by fixing certain parameters equal to zero and relaxing some model constraints.

Key Words: Bayesian inference; Bilinear interaction terms; Two-way table with interaction; von Mises–Fisher.

1. INTRODUCTION

Statistical analyses of two-way tables with interactions are performed in different areas of research—for example, in agriculture, plant breeding and genetics, medicine, and the

Sergio Perez-Elizalde is a Professor of Statistics (E-mail: *sergiop@colpos.mx*) and Diego Jarquin is a Ph.D. Student of Statistics (E-mail: *diego.jarquin@gmail.com*), Programa de Posgrado en Socieconomía, Estadística e Informática, Colegio de Postgraduados, Montecillos, Edo. de Mexico, Mexico. Jose Crossa (⊠) is a Biometrician at the Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600, Mexico DF, Mexico (E-mail: *j.crossa@cgiar.org*).

social sciences. Models combining linear and bilinear terms have proved to be useful for analyzing two-factor studies with interaction, especially when the row and column factors do not have specific structures that might suggest contrasts between rows and columns or response functions (Cornelius and Seyedsadr 1997). This is particularly important in plant breeding, where genotypes (rows) are evaluated in several environments (columns), and genotype × environment interactions (GE) usually complicate selection decisions for the next cycle of improvement.

The usual two-way analysis of variance model is

$$\bar{y}_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \bar{\varepsilon}_{ij} \tag{1.1}$$

where μ, α_i, β_j, and $(\alpha\beta)_{ij}$ (for $i = 1, 2, \ldots, r$; and $j = 1, 2, \ldots, c,$) are the grand mean, the effect of the ith row, the effect of the jth column, and the effect of the interaction of the ith row on the jth column, respectively. The residuals $\bar{\varepsilon}_{ij}$ are identically and independently distributed with $N(0, \sigma_e^2/n_{ij})$ (for simplicity in what follows, we assumed an equal number of observations n in each cell). Parsimonious modeling of the interaction can be considered by the singular value decomposition (SVD) of $(\alpha\beta)_{ij}$ and by retaining only the first few components. This gives rise to the usual linear (additive) bilinear (non-additive) two-way model originally introduced by Gollob (1968) and Mandel (1969, 1971) and extensively used in plant breeding trials for assessing adaptation and stability (Kempton 1984; Gauch 1988; Crossa, Yang, and Cornelius 2004; Cornelius, Crossa, and Seyedsadr 1996). This is known as the Additive Main effect and Multiplicative Interaction (AMMI) model

$$\bar{y}_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=1}^{t} \lambda_k u_{ik} v_{jk} + \bar{\varepsilon}_{ij} \tag{1.2}$$

where λ_k is the singular value for the k principal component axis subject to $\lambda_1 \geq \cdots \lambda_t \geq 0$; u_{ik} and v_{jk} are elements of the left and right singular vectors, respectively, with the side condition that $\sum_i u_{ik}^2 = \sum_j v_{jk}^2 = 1$ and, for $k \neq k'$, $\sum_i u_{ik} u_{ik'} = \sum_j v_{jk} v_{jk'} = 0$ and $t = \min(r, c) - 1$. In matrix notation, (1.2) can be expressed as

$$\mathbf{Y} = \mu \mathbf{1}_r \mathbf{1}_c' + \boldsymbol{\alpha} \otimes \mathbf{1}_c' + \boldsymbol{\beta}' \otimes \mathbf{1}_r + \mathbf{UDV}' + \mathbf{E} \tag{1.3}$$

where $\mathbf{Y} = [\bar{y}_{ij}]$, $\boldsymbol{\alpha} = [\alpha_i]$, $\boldsymbol{\beta} = [\beta_j]$, $\mathbf{D} = \text{diag}(\lambda_k, k = 1, 2, \ldots, t)$, $\mathbf{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_t)$, $\boldsymbol{u}_k = [u_{ik}]$, $\mathbf{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_t)$, $\boldsymbol{v}_k = [v_{jk}]$, and $\mathbf{E} = [\varepsilon_{ij}]$. Note that from model (1.3) [or (1.2)] and allowing $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\beta} = \mathbf{0}$, other linear–bilinear models can be obtained. For example, dropping α_i in (1.2) and writing $\mu_j = \mu + \beta_j$ gives the column regression model that has proved to be very useful in plant breeding for assessing the stability of genotypes (rows) when tested under different environmental conditions (columns). Similarly, dropping β_j in (1.2) and writing $\mu_i = \mu + \alpha_i$ gives the row regression model, and dropping μ, α_i and β_j gives the complete multiplicative model.

Commonly, the parameters in (1.3) are estimated by an iterative least squares (LS) method that first fits the linear terms while ignoring the bilinear terms, which are subsequently fitted as the first t components of the SVD of the residual matrix $\mathbf{Z} = \mathbf{Y} - \hat{\mu} \mathbf{1}_r \mathbf{1}_c' - \hat{\boldsymbol{\alpha}} \otimes \mathbf{1}_c' - \hat{\boldsymbol{\beta}}' \otimes \mathbf{1}_r$, where $\hat{\mu}, \hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are the LS estimates obtained in the first step (Gabriel 1978). Interestingly, Seyedsadr and Cornelius (1992) showed the LS estimates of

the model $\bar{y}_{ij} = \gamma + \sum_{k=1}^{t} \lambda_k u_{ik} v_{jk} + \bar{\varepsilon}_{ij}$, which was originally a problem unsolved by Gabriel (1978) (and named it the Shifted Multiplicative Model, SHMM). The main difficulties of the standard frequentist fixed or mixed linear–bilinear model (1.3) and other related models are: insufficient flexibility to handle heterogeneity of variances and unequal cell size; incorporation of previous information is not possible; only approximate tests for determining the number of components to be retained in the model are available; and inferential statistics to the interaction parameters λ_k, u_{ik}, and v_{jk} not easily developed.

Viele and Srinivasan (2000) were the first to propose Bayesian estimation of parameters for model (1.3) using MCMC techniques through Gibbs sampling with embedded Metropolis-Hastings random walks. The authors proposed spherical uniform prior distributions for the bilinear effects and used the posterior means as shrinkage estimates. Some practical and theoretical issues unresolved by Viele and Srinivasan (2000), such as whether the MCMC will always converge on the target posterior distribution, or whether the bilinear terms of model (1.3) can be estimated from the MCMC sample without violating model constraints, were investigated in the unpublished Ph.D. thesis of Liu (2001). Liu used the same prior distributions as Viele and Srinivasan (2000) to derive the posterior conditional distributions for model (1.3) parameters, such that a Gibbs sampler (without using embedded Metropolis-Hastings steps) for sampling from the joint posterior distribution could be used. Recently, Crossa et al. (2011) applied this approach to practical data resulting from plant breeding multi-environment trials and showed that inferential statistics can be incorporated naturally by adopting the Bayesian approach for estimating the GE interaction parameters including confidence regions in the biplot of the first two components of the bilinear terms.

The spherical uniform distribution is a special case of the von Mises–Fisher distribution (Mardia, Kent, and Bibbi 1979) and was used as a prior by Viele and Srinivasan (2000) for estimating the interaction parameters of (1.3). The authors referred to the constraints on the vectors $\boldsymbol{u}_k = [u_{ik}]$ and $\boldsymbol{v}_k = [v_{jk}]$, which must have unit length and zero sum; in other words, $\boldsymbol{u}_k = [u_{ik}]$ and $\boldsymbol{v}_k = [v_{jk}]$ must be orthonormal and orthogonal to the $\mathbf{1}$ vector. However, the support of the joint posterior distribution of $\boldsymbol{u}_k = [u_{ik}]$ and $\boldsymbol{v}_k = [v_{jk}]$ is not trivial, and Viele and Srinivasan (2000) described a solution for sampling from it using the correct supports. Conceptually, this approach to sampling the conditional posterior distributions of $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$, which are spherical distributions, while maintaining the constraints on the parameters, was performed within the vector framework, that is, sequentially for column vectors of $\mathbf{U}$ and $\mathbf{V}$.

However, there is a need to consider probability models for data from higher dimensionality that allow generalization of the vector approach to a matrix approach; specifically, a useful method would be to use a von Mises–Fisher distribution as a prior on the set of orthonormal matrices whose terms are the bilinear coefficients. Hoff (2009) showed how to sample from the von Mises–Fisher distribution on the multi-dimensional sphere considering the posterior distributions of orthonormal matrices that arise in the analysis of multivariate data.

In this paper we show how to adopt and use the multivariate von Mises–Fisher distribution as a prior on the set of orthonormal matrices that produce the singular value decomposition of interaction matrices as those suitable for model (1.3). In Section 2 we define

the joint prior for unknowns in model (1.3) using the multivariate von Mises–Fisher distribution as a conditional prior for the bilinear effects, the conditional posterior distributions, the Gibbs sampling scheme to simulate MCMC samples from the joint posterior distribution, and the Bayes factor for testing the number of bilinear terms to be retained in the model. Section 3 shows the results and discussion of the application in the context of one simulated data set (five rows and three columns with interaction) and one plant breeding multi-environment multi-year trial comprising 12 genotypes and 25 environments evaluated in two consecutive years, where data from the first year are formally incorporated into the Bayesian inference process through the prior distribution when analyzing the new data from the second year. Bivariate confidence regions (HPD) were estimated for the first two components of the GE bilinear interaction parameters, and only those with HPD not covering the origin $(0, 0)$ are shown in graphical form (biplot). Extension to other types of linear–bilinear model is discussed. Section 4 gives the conclusions.

2. BAYESIAN INFERENCE FOR THE AMMI MODEL

2.1. LIKELIHOOD FUNCTION

The likelihood function for parameters of model (1.3) is

$$L(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y}) \propto \tau^{\frac{nrc}{2}} \exp\left\{-\frac{\tau}{2}\left[n \operatorname{tr}(\mathbf{EE}') + (n-1)\operatorname{tr}(\mathbf{SS}')\right]\right\} \quad (2.1)$$

where $\tau = 1/\sigma^2$, $\mathbf{S} = \left\{\sqrt{s_{ij}^2}\right\}$, $s_{ij}^2 = \sum_{l=1}^{n} \frac{(\overline{Y}_{ij} - Y_{ijl})^2}{n-1}$ and $\mathbf{E} = \mathbf{Y} - \mu \mathbf{1}_r \mathbf{1}_c' - \boldsymbol{\alpha} \otimes \mathbf{1}_c' - \boldsymbol{\beta}' \otimes \mathbf{1}_r - \mathbf{UDV}'$.

2.2. PRIOR DISTRIBUTION

For assessing the prior distributions of the unknowns, we used conditional conjugate prior distributions such that the posterior distribution is proper and can be used to incorporate valuable prior information from experimenters' expertise or from information generated by previous trials. Note that, since the matrices $\mathbf{U}$ and $\mathbf{V}$ are orthonormal and $\mathbf{D}$ is diagonal,

$$\operatorname{tr}(\mathbf{1}_r \mathbf{1}_c' \mathbf{VDU}') = 0$$

$$\operatorname{tr}\left((\boldsymbol{\alpha} \otimes \mathbf{1}_c')\mathbf{VDU}'\right) = 0$$

$$\operatorname{tr}\left((\boldsymbol{\beta}' \otimes \mathbf{1}_r)\mathbf{VDU}'\right) = 0 \quad\quad (2.2)$$

$$\operatorname{tr}\left((\mathbf{UDV}')(\mathbf{UDV}')'\right) = \operatorname{tr}(\mathbf{D}'\mathbf{D}) = \sum_{k=1}^{t} \lambda_k^2$$

$$\operatorname{tr}\left\{(-2\mathbf{Y} + \mathbf{UDV}')(\mathbf{UDV}')'\right\} = \operatorname{tr}\left\{(\mathbf{D} - \mathbf{U}'\mathbf{YV})'(\mathbf{D} - \mathbf{U}'\mathbf{YV}) - (\mathbf{U}'\mathbf{YV})'(\mathbf{U}'\mathbf{YV})\right\}$$

thus it can be shown from (1.3) that, given $\boldsymbol{\theta} = (\mu, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and τ, the conditional likelihood function for the matrices $(\mathbf{U}, \mathbf{D}, \mathbf{V})$ is

$$L(\mathbf{U}, \mathbf{V}, \mathbf{D} \mid \boldsymbol{\theta}, \tau, \mathbf{Y}) = L(\mathbf{U}, \mathbf{V}, \mathbf{D} \mid \tau, \mathbf{Y}) \propto \exp\left\{-\frac{n\tau}{2}\operatorname{tr}\bigl((-2\mathbf{Y} + \mathbf{U}\mathbf{D}\mathbf{V}')(\mathbf{U}\mathbf{D}\mathbf{V}')'\bigr)\right\}$$

$$= \operatorname{etr}\left\{-\frac{n\tau}{2}(-2\mathbf{Y} + \mathbf{U}\mathbf{D}\mathbf{V}')(\mathbf{U}\mathbf{D}\mathbf{V}')'\right\} \tag{2.3}$$

where 'etr' is the exponential of the trace. From (2.2) and (2.3), the conditional likelihoods for $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$ are

$$L(\mathbf{U} \mid \mathbf{V}, \mathbf{D}, \tau, \mathbf{Y}) \propto \operatorname{etr}\{n\tau\mathbf{Y}\mathbf{V}\mathbf{D}\mathbf{U}'\} \tag{2.4}$$

$$L(\mathbf{V} \mid \mathbf{U}, \mathbf{D}, \tau, \mathbf{Y}) \propto \operatorname{etr}\{n\tau\mathbf{Y}'\mathbf{U}\mathbf{D}\mathbf{V}'\} \tag{2.5}$$

$$L(\mathbf{D} \mid \mathbf{U}, \mathbf{V}, \tau, \mathbf{Y}) \propto \operatorname{etr}\left\{-\frac{n\tau}{2}(\mathbf{D} - \mathbf{U}'\mathbf{Y}\mathbf{V})'(\mathbf{D} - \mathbf{U}'\mathbf{Y}\mathbf{V})\right\}$$

$$\propto \exp\left\{-\frac{n\tau}{2}\sum_{k=1}^{t}(\lambda_k - l_k)^2\right\}, \qquad \lambda_1 > \lambda_2 > \cdots > \lambda_t \tag{2.6}$$

respectively, where $(l_1, \ldots, l_k) = \operatorname{diag}(\mathbf{U}'\mathbf{Y}\mathbf{V})$.

From expression (2.4) it follows that a conditional conjugate prior for $\mathbf{U}$ is

$$\pi(\mathbf{U} \mid \tau) \propto \operatorname{etr}\{\tau n_0 \mathbf{Y}_0 \mathbf{V}_0 \mathbf{D}_0 \mathbf{U}'\} \tag{2.7}$$

where $\mathbf{Y}_0$ could be interpreted as the matrix of prior cell averages such that

$$\mathbf{Y}_0 = \mu_0 \mathbf{1}_r \mathbf{1}_c' + \boldsymbol{\alpha}_0 \otimes \mathbf{1}_c' + \boldsymbol{\beta}_0' \otimes \mathbf{1}_r + \mathbf{U}_0 \mathbf{D}_0 \mathbf{V}_0';$$

that is, $\mathbf{D}_0$ is a diagonal matrix of prior singular values, and $\mathbf{U}_0$ and $\mathbf{V}_0$ are orthonormal matrices such that $\mathbf{U}_0 \mathbf{D}_0 \mathbf{V}_0'$ is the SVD of $\mathbf{Z}_0 = \mathbf{Y}_0 - \mu_0 \mathbf{1}_r \mathbf{1}_c' - \boldsymbol{\alpha}_0 \otimes \mathbf{1}_c' - \boldsymbol{\beta}_0' \otimes \mathbf{1}_r$, where μ_0, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are prior values for the linear effects.

Similarly, from (2.5), a prior for $\mathbf{V}$ is

$$\pi(\mathbf{V} \mid \tau) \propto \operatorname{etr}\{\tau n_0 \mathbf{Y}_0' \mathbf{U}_0 \mathbf{D}_0 \mathbf{V}'\} \tag{2.8}$$

Both (2.7) and (2.8) are von Mises–Fisher distributions (see Appendix A). From (2.6) for each one of the elements $\lambda_1 > \lambda_2 > \cdots > \lambda_k$ on the diagonal of $\mathbf{D}$, the conditional conjugate prior distributions are left truncated normal with marginal densities of the form

$$\pi(\lambda_k \mid \tau) = \bigl\{1 - \Phi\bigl(\sqrt{n_0\tau}(\lambda_{k+1} - l_k^0)\bigr)\bigr\}^{-1}\mathrm{N}\bigl(\lambda_k \mid l_k^0, (n_0\tau)^{-1}\bigr), \qquad k = 1, \ldots, t, \ \lambda_{t+1} = 0 \tag{2.9}$$

For the linear terms $\boldsymbol{\theta} = (\mu, \boldsymbol{\alpha}, \boldsymbol{\beta})$ of model (1.1), a conditional conjugate prior is a $(1 + r + c)$-multivariate normal distribution with mean $\boldsymbol{\theta}_0 = (\mu_0, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and singular block diagonal covariance matrix

$$(n_0\tau)^{-1}\begin{bmatrix} (r_0 c_0)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & c_0^{-1}\mathbf{K}_r\mathbf{K}_r' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & r_0^{-1}\mathbf{K}_c\mathbf{K}_c' \end{bmatrix}$$

where $\mathbf{K}_w$ is a matrix such that $\mathbf{K}_w'\mathbf{K}_w = \mathbf{I}_{w-1}$ and $\mathbf{K}_w\mathbf{K}_w' = \mathbf{I}_w - \frac{1}{w}\mathbf{J}_w$, where $\mathbf{J}_w$ is an $w \times w$ matrix with all its elements equal to one. Because of the restrictions $\boldsymbol{\alpha}'1_r = 0$ and $\boldsymbol{\beta}'1_c = 0$, the distribution characterized by the covariance matrix above is a singular multivariate normal distribution that does not have a density. For a prior density we need to choose a one to one transformation like $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = (\mathbf{K}_r'\boldsymbol{\alpha}, \mathbf{K}_c'\boldsymbol{\beta})$.

Let $\boldsymbol{\theta}^* = (\mu, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$; then the prior density of $\boldsymbol{\theta}^*$ is

$$\pi(\boldsymbol{\theta}^* \mid \tau) \propto \left| \boldsymbol{\Sigma}_0 \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0^*)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0^*) \right\}$$

$$\boldsymbol{\Sigma}_0 = (n_0\tau)^{-1} \begin{bmatrix} (r_0c_0)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & c_0^{-1}\mathbf{I}_{r-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & r_0^{-1}\mathbf{I}_{c-1} \end{bmatrix} \tag{2.10}$$

which is the density of a $(1 + r - 1 + c - 1)$-multivariate normal distribution with mean $\boldsymbol{\theta}_0^* = (\mu_0, \boldsymbol{\alpha}_0^*, \boldsymbol{\beta}_0^*) = (\mu, \mathbf{K}_r'\boldsymbol{\alpha}_0, \mathbf{K}_c'\boldsymbol{\beta}_0)$ and covariance matrix $\boldsymbol{\Sigma}_0$.

The joint likelihood (2.1) suggests that a conjugate prior for τ is a gamma distribution with parameters $a/2$, and $s_0^2/2$; that is,

$$\pi(\tau) \propto \tau^{\frac{a}{2}-1} \exp\left\{ -\frac{as_0^2}{2}\tau \right\} \tag{2.11}$$

or equivalently, $as_0^2\tau \sim \chi_a^2$; thus, a and s_0^2 may be considered as prior values for sample size and variance, respectively.

Finally, the joint prior for $(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau)$ is

$$\pi(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau) = \pi(\boldsymbol{\theta}^* \mid \tau)\pi(\mathbf{U} \mid \tau)\pi(\mathbf{D} \mid \tau)\pi(\mathbf{V} \mid \tau)\pi(\tau) \tag{2.12}$$

The proposed prior has practical advantages and is flexible enough to incorporate prior uncertainty about unknown parameters. On the other hand, the main disadvantage of the prior used by Liu (2001) and Crossa et al. (2011) for implementing their Bayesian approach was the elicitation of the distribution of each element on the matrices given by the SVD decomposition of the interaction. In our proposal, the incorporation of prior information is straightforward and intuitive, as it only needs to express our beliefs in the prior cell averages $\mathbf{Y}_0$ and prior linear effects $\boldsymbol{\theta}_0$. Then $\mathbf{U}_0$, $\mathbf{V}_0$ and $\mathbf{D}_0$ are obtained from the SVD decomposition of $\mathbf{Z}_0$, under the restrictions $\mathbf{U}_0'1_r = \mathbf{0}$ and $\mathbf{V}_0'1_c = \mathbf{0}$. The prior distribution of the linear effects is completely specified by giving a belief $\boldsymbol{\theta}_0^*$ about $\boldsymbol{\theta}^*$. This prior belief may be expressed as a function of $\mathbf{Y}_0$; for example, $\boldsymbol{\theta}_0^*(\mathbf{Y}_0) = (\frac{1_r'\mathbf{Y}_01_c}{r_0c_0}, \frac{\mathbf{K}_r'\mathbf{Y}_01_c}{c_0}, \frac{\mathbf{K}_c'\mathbf{Y}_0'1_r}{r_0})$. Also, it is important to note that a vague prior for τ implies diffuse priors for all the other parameters. Then, an objective or default Bayesian analysis could be performed by setting small values to the hyper-parameter a in the prior for τ given by (2.11). Therefore, we may summarize our prior information by giving, a priori, a prediction of the two-way array of means $\mathbf{Y}_0$ and a measure of our prior uncertainty s_0^2 given a prior sample size a.

For the analysis of the simulated data set below, the priors used were vague but proper, whereas for the plant breeding data set, the prior hyper-parameters used were obtained from the first-year evaluation of the 12 genotypes in 25 environments, and the data set

analyzed was from the second year of evaluation. This example illustrates how to use this approach within a practical breeding program and how to draw useful biological inferences on the interaction parameters.

2.3. POSTERIOR DISTRIBUTION AND GIBBS SAMPLER

The joint posterior distribution is obtained by combining the likelihood function (2.1) and the prior distribution (2.12),

$$\pi(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y}) \propto L(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y})\pi(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau)$$

where $L(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y})$ is a re-parameterization of $L(\boldsymbol{\theta}, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y})$. The marginal posterior distribution, which involves high dimensional integration on complex spaces, is needed for marginal inference about the unknowns. In order to use a Markov Chain Monte Carlo (MCMC) method through the Gibbs sampler, the full conditional posterior distributions, which are known except for the proportionality constants, are needed. These distributions are computed by considering the joint posterior as a function only of a variable when fixing the others. Thus, the conditional posterior for $\boldsymbol{\theta}^*$ is

$$\pi(\boldsymbol{\theta}^* \mid \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau, \mathbf{Y}) \propto L(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y})\pi(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau)$$

Knowing that the conditional likelihood function of $(\mathbf{U}, \mathbf{V}, \mathbf{D})$ does not depend on $\boldsymbol{\theta}^*$, and that given τ, the prior for $\boldsymbol{\theta}^*$ is independent of $(\mathbf{U}, \mathbf{V}, \mathbf{D})$, then

$$\pi(\boldsymbol{\theta}^* \mid \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau, \mathbf{Y}) = \pi(\boldsymbol{\theta}^* \mid \tau, \mathbf{Y}) \propto L(\boldsymbol{\theta}^* \mid \tau, \mathbf{Y})\pi(\boldsymbol{\theta}^* \mid \tau)$$

$$\propto L(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y})\pi(\boldsymbol{\theta}^* \mid \tau)/L(\mathbf{U}, \mathbf{V}, \mathbf{D} \mid \tau, \mathbf{Y})$$

$$\propto \exp\left\{-\frac{n\tau}{2}\operatorname{tr}(\mathbf{Z}\mathbf{Z}')\right\}\pi(\boldsymbol{\theta}^* \mid \tau)$$

where $\mathbf{Z} = \mathbf{Y} - \mu\mathbf{1}_r\mathbf{1}_c' - \boldsymbol{\alpha} \otimes \mathbf{1}_c' - \boldsymbol{\beta}' \otimes \mathbf{1}_r|_{\boldsymbol{\alpha}=\mathbf{K}_r\boldsymbol{\alpha}^*, \boldsymbol{\beta}=\mathbf{K}_c\boldsymbol{\beta}^*}$. It can be shown that the conditional posterior of $\boldsymbol{\theta}^*$ is multivariate normal with density $\pi(\boldsymbol{\theta}^* \mid \tau, \mathbf{Y}) = N_{r+c-1}(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}_n^*, \boldsymbol{\Sigma}_n^*)$, covariance matrix $\boldsymbol{\Sigma}_n^* = (\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_n^{-1})^{-1}$ and mean $(\boldsymbol{\theta}_n^* = (\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_n^{-1})^{-1} \times (\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_0^* + \boldsymbol{\Sigma}_n^{-1}\hat{\boldsymbol{\theta}}_n^*)$ with $\hat{\boldsymbol{\theta}}_n^* = (\frac{\mathbf{1}_r'\mathbf{Y}\mathbf{1}_c}{rc}, \frac{\mathbf{K}_r'\mathbf{Y}\mathbf{1}_c}{c}, \frac{\mathbf{K}_c'\mathbf{Y}'\mathbf{1}_r}{r})$ and

$$\boldsymbol{\Sigma}_n = (n\tau)^{-1}\begin{bmatrix} (rc)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & c^{-1}\mathbf{I}_{r-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & r^{-1}\mathbf{I}_{c-1} \end{bmatrix}$$

We may use the conditional likelihoods given by (2.4)–(2.6) to calculate the conditional posteriors for $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{D}$; i.e.,

$$\pi(\mathbf{U} \mid \mathbf{V}, \mathbf{D}, \tau, \mathbf{Y}) \propto L(\mathbf{U} \mid \mathbf{V}, \mathbf{D}, \tau, \mathbf{Y})\pi(\mathbf{U} \mid \tau)$$

$$\propto \text{etr}\{\tau(n_0\mathbf{Y}_0\mathbf{V}_0\mathbf{D}_0 + n\mathbf{YVD})\mathbf{U}'\} \tag{2.13}$$

$$\pi(\mathbf{V} \mid \mathbf{U}, \mathbf{D}, \tau, \mathbf{Y}) \propto L(\mathbf{V} \mid \mathbf{U}, \mathbf{D}, \tau, \mathbf{Y})\pi(\mathbf{V} \mid \tau)$$

$$\propto \text{etr}\{\tau(n_0\mathbf{Y}_0'\mathbf{U}_0\mathbf{D} + n\mathbf{Y}'\mathbf{UD})\mathbf{V}'\} \tag{2.14}$$

$$\pi(\mathbf{D} \mid \mathbf{U}, \mathbf{V}, \tau, \mathbf{Y}) \propto L(\mathbf{D} \mid \mathbf{U}, \mathbf{V}, \tau, \mathbf{Y})\pi(\mathbf{D} \mid \mathbf{U}, \mathbf{V}, \tau)$$

$$\propto \exp\left\{-\frac{n\tau}{2}\sum_{k=1}^{t}(\lambda_k - l_k)^2\right\}\prod_{k=1}^{t}\pi(\lambda_k \mid \tau)$$

$$\propto \prod_{k=1}^{t}\left\{1 - \Phi\left(\sqrt{\frac{\tau^{-1}}{n_0+n}}\left(\lambda_{k+1} - \frac{n_0l_k^0 + nl_k}{n_0+n}\right)\right)\right\}^{-1}$$

$$\times \text{N}\left(\lambda_k \,\middle|\, \frac{n_0l_k^0 + nl_k}{n_0+n}, \frac{\tau^{-1}}{n_0+n}\right)$$

$$\lambda_1 > \lambda_2 > \cdots > \lambda_t > \lambda_{t+1} = 0 \tag{2.15}$$

Finally, the conditional posterior for the precision τ is a gamma with density

$$\pi(\tau \mid \boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \mathbf{Y}) \propto L(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y})\pi(\tau)$$

$$\propto \text{Ga}\left(\tau \,\middle|\, \frac{a_n}{2}, \frac{b_n}{2}\right)$$

where

$$a_n = a + nrc$$

$$b_n = as_0^2 + n\big(\text{tr}(\mathbf{EE}')\big) + (n-1)\text{tr}(\mathbf{SS}')$$

2.3.1. Gibbs Sampler

The Gibbs Sampler is implemented by sequentially drawing simulated samples from the full conditional posterior distributions; thus we may proceed with the following algorithm.

Let s be the desired length of the Markov chain to be simulated. Let $\mathbf{U}^{(0)}$, $\mathbf{V}^{(0)}$ and $\mathbf{D}^{(0)}$ be the initial values of the simulated Markov chain.

For $i = 0, \ldots, s$ simulate

$$\tau^{(i+1)} \sim \pi\big(\tau \mid \boldsymbol{\theta}^{*(i)}, \mathbf{U}^{(i)}, \mathbf{D}^{(i)}, \mathbf{Y}\big)$$

$$\boldsymbol{\theta}^{*(i+1)} \sim \pi\big(\boldsymbol{\theta}^* \mid \tau^{(i+1)}, \mathbf{Y}\big)$$

$$\mathbf{U}^{(i+1)} \sim \pi\big(\mathbf{U} \mid \tau^{(i+1)}\big)$$

$$\mathbf{D}^{(i+1)} \sim \pi\big(\mathbf{D} \mid \tau^{(i+1)}\big)$$

$$\mathbf{V}^{(i+1)} \sim \pi\big(\mathbf{V} \mid \tau^{(i+1)}\big)$$

After a burn-in period, we assume that the generated samples arise from the stationary distribution, i.e., joint posterior distribution $\pi(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y})$. For reversibility of the Markov chain, we may permute the order of simulation, but in what follows we use the order indicated above. Some standard convergence diagnostic tools may be used to determine an effective sample size; in the examples below, criteria for convergence of simulated Markov chains from Raftery and Lewis (1995) and Gelman and Rubin (1992) were used.

2.4. TESTING THE SIGNIFICANCE OF THE BILINEAR TERMS USING THE BAYES FACTOR

Concerning the inference on the parameters of liner-bilinear models, a central question is how many linear components the model should retain. Several frequentist model selection approaches exist for linear–bilinear models of a two-way table without replicated data in each cell. Johnson and Graybill (1972) developed a hypothesis test for a model with one bilinear term and Marasinghe (1985) and Schott (1986) proposed a sequential test for the bilinear terms beyond the first one. For two-way tables with replications in each cell, Gollob (1968) proposed using an approximate F statistics for the hypothesis H_1: $\lambda_t = 0$ versus H_2: $\lambda_t \neq 0$. However, this test is too liberal for practical use and Cornelius, Seyedsadr, and Crossa (1992), and Cornelius, Crossa, and Seyedsadr (1994) introduced a series of sequential F approximations that effectively control Type I error rates.

Several approaches for Bayesian model selection can be found in the literature, the most widely used being the one based on the Bayes factor. Suppose a data sample $(\boldsymbol{Y})$ coming from two competing models M_0 and M_1 according to probabilities $p(\boldsymbol{Y} \mid M_0)$ and $p(\boldsymbol{Y} \mid M_1)$, whose prior probabilities are such that $p(M_0)$ and $p(M_1) = 1 - p(M_0)$, and their respective posterior probabilities are $p(M_0 \mid \boldsymbol{Y})$ and $p(M_1 \mid \boldsymbol{Y})$. The Bayes factor B_{01}, defined as the ratio of the posterior to prior odds, is the weight of evidence in favor of model M_1 provided by the data (Kass and Raftery 1995); that is,

$$B_{01} = \frac{p(M_0 \mid \mathbf{Y})}{p(M_1 \mid \mathbf{Y})} \bigg/ \frac{p(M_0)}{p(M_1)} = \frac{p(\mathbf{Y} \mid M_0)}{p(\mathbf{Y} \mid M_1)}.$$

The conventional decision rule suggested by Jeffreys (1961) is that M_0 is selected if $B_{01} > 10$; otherwise, the model selected is M_1. Note that this reduces to the posterior odds in favor of M_0 when M_0 and M_1 have equal prior probabilities, $p(M_0) = p(M_1) = 0.5$ (representing the usual non-informative prior on two competitive models). Considering that $\lambda_1 > \lambda_2 \cdots > \lambda_t$, it is sufficient to sequentially compare model M_0 with $\lambda_k \neq 0$ and $\lambda_{k+1} = 0$ against the alternative model M_1 with $\lambda_{k+1} \neq 0$ [for $k = 1, 2, \ldots, t$] in order to choose the model with the highest posterior probability. By the above definition of the Bayes factor, the interest lies in obtaining the marginal density of the data $p(\mathbf{Y} \mid M_i)$; Chib (1995) provided a method for computing this when the full conditionals of blocks of parameters are available in closed form. For a model where $\mathbf{Y}$ is the observed data and $\boldsymbol{\theta}$ denotes the unknown parameter, Chib (1995) noted that by Bayes' rule, $\forall \boldsymbol{\theta} \in \Theta$ $p(\mathbf{Y} \mid M_i) = f(\mathbf{Y} \mid \boldsymbol{\theta}, M_i) p(\boldsymbol{\theta} \mid M_i) / p(\boldsymbol{\theta} \mid \mathbf{Y}, M_i)$. Then, an estimate of $\log p(\mathbf{Y} \mid M_i)$ is

$$\log \widehat{p}(\mathbf{Y} \mid M_i) = \log f(\mathbf{Y} \mid \widetilde{\boldsymbol{\theta}}, M_i) + \log p(\widetilde{\boldsymbol{\theta}} \mid M_i) - \log \widehat{p}(\widetilde{\boldsymbol{\theta}} \mid \mathbf{Y}, M_i)$$

where $\widehat{p}(\boldsymbol{\theta} \mid \mathbf{Y}, M_i)$ is an estimate of the posterior distribution under model M_i and $\widetilde{\boldsymbol{\theta}}$ is a value of $\boldsymbol{\theta}$ with high posterior density to ensure accuracy. It should be clear that normalization constants need to be known or at least estimated for each model. Therefore, an approximation of B_{01} is given by

$$\widehat{B}_{01} = \frac{\widehat{p}(\mathbf{Y} \mid M_0)}{\widehat{p}(\mathbf{Y} \mid M_1)}$$

For a detailed explanation of the algorithm and alternative approaches for computing Bayes factors, see Han and Carlin (2001). Since this algorithm works with the models one at a time, model indicators are dropped to simplify the notation. Note that in our problem

$$\pi(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau \mid \mathbf{Y}) = \pi(\mathbf{D} \mid \mathbf{U}, \mathbf{V}, \boldsymbol{\theta}^*, \tau, \mathbf{Y})\pi(\mathbf{U} \mid \mathbf{V}, \boldsymbol{\theta}^*, \tau, \mathbf{Y})$$
$$\times \pi(\mathbf{V} \mid \boldsymbol{\theta}, \tau, \mathbf{Y})\pi(\boldsymbol{\theta}^* \mid \tau, \mathbf{Y})\pi(\tau \mid \mathbf{Y})$$

Thus, an estimate of the joint posterior at the high density point $(\widetilde{\boldsymbol{\theta}}, \widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\mathbf{D}}, \widetilde{\tau})$ is

$$\widehat{\pi}(\widetilde{\boldsymbol{\theta}}, \widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\mathbf{D}}, \widetilde{\tau} \mid \mathbf{Y}) = \pi(\widetilde{\mathbf{D}} \mid \widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\boldsymbol{\theta}}, \widetilde{\tau}, \mathbf{Y})\widehat{\pi}(\widetilde{\mathbf{U}} \mid \widetilde{\mathbf{V}}, \widetilde{\boldsymbol{\theta}}, \widetilde{\tau}, \mathbf{Y})\widehat{\pi}(\widetilde{\mathbf{V}} \mid \widetilde{\boldsymbol{\theta}}, \widetilde{\tau}, \mathbf{Y})$$
$$\times \pi(\widetilde{\boldsymbol{\theta}} \mid \widetilde{\tau}, \mathbf{Y})\widehat{\pi}(\widetilde{\tau} \mid \mathbf{Y})$$
$$= \pi(\widetilde{\mathbf{D}} \mid \widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\tau}, \mathbf{Y})\widehat{\pi}(\widetilde{\mathbf{U}} \mid \widetilde{\mathbf{V}}, \widetilde{\tau}, \mathbf{Y})\widehat{\pi}(\widetilde{\mathbf{V}} \mid \widetilde{\tau}, \mathbf{Y})\pi(\widetilde{\boldsymbol{\theta}} \mid \widetilde{\tau}, \mathbf{Y})\widehat{\pi}(\widetilde{\tau} \mid \mathbf{Y})$$

where

$$\widehat{\pi}(\widetilde{\mathbf{U}} \mid \widetilde{\mathbf{V}}, \widetilde{\tau}, \mathbf{Y}) = \sum_{g=1}^{G} \pi(\widetilde{\mathbf{U}} \mid \mathbf{D}^{(g)}, \widetilde{\mathbf{V}}, \widetilde{\tau}, \mathbf{Y})/G$$

$$\widehat{\pi}(\widetilde{\mathbf{V}} \mid \widetilde{\tau}, \mathbf{Y}) = \sum_{g=1}^{G} \pi(\widetilde{\mathbf{V}} \mid \mathbf{D}^{(g)}, \mathbf{U}^{(g)}, \widetilde{\tau}, \mathbf{Y})/G$$

$$\widehat{\pi}(\widetilde{\tau} \mid \mathbf{Y}) = \sum_{g=1}^{G} \pi(\widetilde{\tau} \mid \mathbf{U}^{(g)}, \mathbf{D}^{(g)}, \mathbf{V}^{(g)}, \boldsymbol{\theta}^{*(g)}, \mathbf{Y})/G$$

are Monte Carlo estimates of the conditional densities based on a Gibbs sample of size G. We do not need an estimate of $\pi(\boldsymbol{\theta}^* \mid \tau, \mathbf{Y})$, since the full conditional does not depend on $(\mathbf{U}, \mathbf{V}, \mathbf{D})$. In this study, the most difficult task is to estimate the normalization constant of the full conditionals of $\mathbf{U}$ and $\mathbf{V}$, which are von Mises–Fisher distributions, because no closed form of the normalization constant is given by the hypergeometric function of a matrix argument (Herz 1955; James 1964); here we used the approximation given by Khatri and Mardia (1977) for either small or large eigenvalues of the matrix argument. A more accurate approximation was suggested by Koev and Edelman (2006).

With the above approximation of the joint posterior, using the priors given in Section 2.2 and adjusting them by model size, an estimate of the marginal log-density of $\mathbf{Y}$ conditional on M_i is

$$\log \widehat{p}(\mathbf{Y} \mid M_i) = \log \mathrm{L}(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\mathbf{D}}, \widetilde{\tau} \mid \mathbf{Y}, M_i) + \log \pi(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\mathbf{D}}, \widetilde{\tau} \mid M_i)$$
$$- \log \widehat{\pi}(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\mathbf{D}}, \widetilde{\tau} \mid \mathbf{Y}, M_i)$$

Note that by (2.1)–(2.3) and (2.12), the approximation may be written as

$$
\begin{aligned}
\log \widehat{p}(\mathbf{Y} \mid M_i) = {} & \log \mathrm{L}(\widetilde{\boldsymbol{\theta}}^*, \widetilde{\tau} \mid \mathbf{Y}) + \log \mathrm{L}(\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\mathbf{D}} \mid \widetilde{\tau}, \mathbf{Y}, M_i) \\
& + \log \pi(\widetilde{\boldsymbol{\theta}}^* \mid \widetilde{\tau}) + \log \pi(\widetilde{\mathbf{U}} \mid \widetilde{\tau}, M_i) \\
& + \log \pi(\widetilde{\mathbf{V}} \mid \widetilde{\tau}, M_i) + \log \pi(\widetilde{\mathbf{D}} \mid \widetilde{\tau}, M_i) \\
& + \log \pi(\widetilde{\tau} \mid M_i) - \log \pi(\widetilde{\mathbf{D}} \mid \widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\tau}, \mathbf{Y}, M_i) \\
& - \log \widehat{\pi}(\widetilde{\mathbf{U}} \mid \widetilde{\mathbf{V}}, \widetilde{\tau}, \mathbf{Y}, M_i) - \log \widehat{\pi}(\widetilde{\mathbf{V}} \mid \widetilde{\tau}, \mathbf{Y}, M_i) - \log \pi(\widetilde{\boldsymbol{\theta}}^* \mid \widetilde{\tau}, \mathbf{Y}) \\
& - \log \widehat{\pi}(\widetilde{\tau} \mid \mathbf{Y}, M_i).
\end{aligned}
$$

In the equation above, the terms involving $\widetilde{\boldsymbol{\theta}}^*$ do not depend on M_i; therefore, it is not necessary to evaluate those terms when estimating the Bayes factor, since it will be canceled out.

3. RESULTS AND DISCUSSION

3.1. SIMULATED DATA

To illustrate the implementation of our proposed model, we analyzed a simulated data set comprising five rows and three columns, each with an $n = 4$ sample size from a normal distribution with mean $\mu + a_i + \beta_j + \sum_{k=1}^{2} \lambda_k u_{ik} v_{jk}$ ($i = 1, \ldots, 5; j = 1, \ldots, 3$) and variance $\sigma^2 = 1$. Thus the overall sample size is 60. The true values of the fixed linear effects are $\boldsymbol{\alpha} = (-1, -0.5, 0.5, 1, 0)$ for the factor row and $\boldsymbol{\beta} = (-1, 1, 0)$ for the factor column, while the bilinear terms are given by the SVD decomposition with components:

$$
\mathbf{U} = \begin{bmatrix}
-0.8250 & -0.3372 \\
-0.0943 & 0.7041 \\
0.1785 & 0.3049 \\
0.3246 & -0.1463 \\
0.4162 & -0.5255
\end{bmatrix}, \qquad
\mathbf{D} = \begin{bmatrix}
1.4805 & 0 \\
0 & 0.1579
\end{bmatrix}
$$

$$
\mathbf{V} = \begin{bmatrix}
-0.7865 & -0.2191 \\
0.2035 & 0.7907 \\
0.5830 & -0.5716
\end{bmatrix}
$$

The interaction plot of the simulated data shows that the row effects are not equal for all levels of the factor column, with clear evidence of interaction between row and column effects (data not shown). The true biplot of the first (Component 1) and second (Component 2) principal components, i.e., the columns of $\mathbf{PC} = \left(\begin{smallmatrix} \mathbf{U}'\mathbf{D}^{1/2} \\ \mathbf{V}'\mathbf{D}^{1/2} \end{smallmatrix} \right)$, shows interaction effects between rows and columns (data not shown).

We used $\boldsymbol{\alpha}_0 = (0, 0, 0, 0, 0)$, $\boldsymbol{\beta}_0 = (0, 0, 0)$, $a = 1$, $s_0^2 = 1000$ as prior information, and samples of the posterior distribution were drawn with the Gibbs sampler described above. Two parallel chains of size $s = 20{,}000$ were simulated using a burn-in period of size 10,000. Finally, a thinning period of 1 was used. Therefore, a final MCMC sample of size 5000 was used to estimate the posterior distribution.

Table 1. Posterior summary (Mean), standard deviation (SD), quartiles ($q_{0.25}$, $q_{0.50}$, and $q_{0.75}$), 0.90 HPD, and 0.95 HPD intervals computed with 5000 approximately independent samples from the joint posterior distribution of the simulated data (five rows and three columns).

Parameter	Mean	SD	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	0.90 HPD interval		0.95 HPD interval	
						Lower	Upper	Lower	Upper
μ	5.0184	0.1342	4.7523	5.0175	5.2878	4.8055	5.2414	4.7584	5.2898
σ	1.1572	0.1312	0.9401	1.1461	1.4498	0.9504	1.3617	0.9224	1.4211
α_1	-1.2282	0.2713	-1.7687	-1.2268	-0.6991	-1.6437	-0.7530	-1.7421	-0.6755
α_2	-0.6733	0.2703	-1.2027	-0.6744	-0.1334	-1.1158	-0.2299	-1.2169	-0.1493
α_3	0.6239	0.2704	0.0897	0.6234	1.1536	0.1879	1.0654	0.0825	1.1429
α_4	0.8053	0.2707	0.2635	0.8078	1.3387	0.3514	1.2457	0.2674	1.3406
α_5	0.4723	0.2693	-0.0597	0.4756	0.9977	0.0283	0.9101	-0.0524	0.9990
β_1	-0.6824	0.1891	-1.0512	-0.6835	-0.3060	-0.9833	-0.3607	-1.0442	-0.3006
β_2	0.9110	0.1919	0.5323	0.9113	1.2878	0.5876	1.2178	0.5247	1.2801
β_3	-0.2286	0.1885	-0.5925	-0.2262	0.1382	-0.5313	0.0796	-0.5984	0.1299
$u_{1,1}$	-0.7251	0.1456	-0.8822	-0.7654	-0.3227	-0.8934	-0.5411	-0.8936	-0.4429
$u_{2,1}$	0.0486	0.2471	-0.4350	0.0491	0.5363	-0.3761	0.4418	-0.4145	0.5553
$u_{3,1}$	0.3568	0.2278	-0.1516	0.3785	0.7387	0.0148	0.7348	-0.0980	0.7848
$u_{4,1}$	0.0291	0.2664	-0.5018	0.0339	0.5419	-0.4052	0.4702	-0.4965	0.5463
$u_{5,1}$	0.2906	0.2329	-0.2194	0.3065	0.7075	-0.0625	0.6817	-0.1559	0.7540
$u_{1,2}$	-0.0794	0.3027	-0.6419	-0.0856	0.5312	-0.4084	0.5854	-0.6491	0.5206
$u_{2,2}$	0.4202	0.2532	0.0020	0.4207	0.8463	-0.8109	-0.0279	0.0060	0.8490
$u_{3,2}$	-0.1020	0.4324	-0.8099	-0.1219	0.6999	-0.5690	0.8001	-0.8616	0.6398
$u_{4,2}$	-0.1439	0.4835	-0.8506	-0.2008	0.7448	-0.5732	0.8802	-0.8802	0.6927
$u_{5,2}$	-0.0949	0.4480	-0.8220	-0.1154	0.7156	-0.5929	0.8072	-0.8734	0.6487
$v_{1,1}$	-0.7217	0.1321	-0.8164	-0.7698	-0.3394	-0.8165	-0.5704	-0.8165	-0.4665
$v_{2,1}$	0.4129	0.3096	-0.3190	0.4766	0.8095	-0.0152	0.8165	-0.1733	0.8165
$v_{3,1}$	0.3088	0.3162	-0.3635	0.3356	0.7980	-0.1241	0.8165	-0.2461	0.8165
$v_{1,2}$	0.0602	0.3532	-0.6508	0.0814	0.6803	-0.6089	0.5396	-0.6057	0.7196
$v_{2,2}$	0.5941	0.2177	0.0368	0.6540	0.8160	-0.8165	-0.2917	0.1700	0.8165
$v_{3,2}$	-0.6543	0.2080	-0.8163	-0.7326	-0.0962	0.3934	0.8165	-0.8165	-0.2449
λ_1	2.2079	0.4741	1.0096	2.2793	2.9221	1.5010	2.9257	1.2120	3.0237
λ_2	0.5073	0.3744	0.0132	0.4461	1.3127	0.0003	1.0357	0.0003	1.1863

Table 1 gives a summary of the marginal posterior distribution for all the parameters in the model. The true values are always within their interval estimations. Moreover, relative to the estimated standard deviations, point estimations are close enough to the true values. Thus the proposed Bayesian approach to model interaction in a two-way table gives reliable inferential answers about unknowns in the model. The second singular value is about 4 times smaller than the first singular value, and also had smaller SD than the first one. The eigenvectors for $u_{1,1}$, $v_{1,1}$, $v_{2,2}$ and $v_{3,2}$ are the only bilinear terms that do not include the null point $(0,0)$ for the bivariate 0.90 HPD region and thus cause most of the significant interaction. Other posterior 0.90 HPD intervals for eigenvector elements such as $u_{3,1}$ and $u_{2,2}$ do not contain zero, but the 0.90 HPD region for their corresponding scores $(u_{3,1}\sqrt{\lambda_1}, u_{3,2}\sqrt{\lambda_2})$ and $(u_{2,1}\sqrt{\lambda_1}, u_{2,2}\sqrt{\lambda_2})$ covers the null point $(0,0)$.

The HPD regions of the row (column) scores that are statistically different from the null point $(0,0)$ can be seen in the biplot depicted in Figures 1(a) and (b). For example, Figure 1(a) shows the plot of the row scores $\mathbf{U}'\mathbf{D}^{1/2}$, and the outer and inner shaded areas of the graph are the bivariate 0.95 and 0.90 HPD posterior regions, respectively, for the
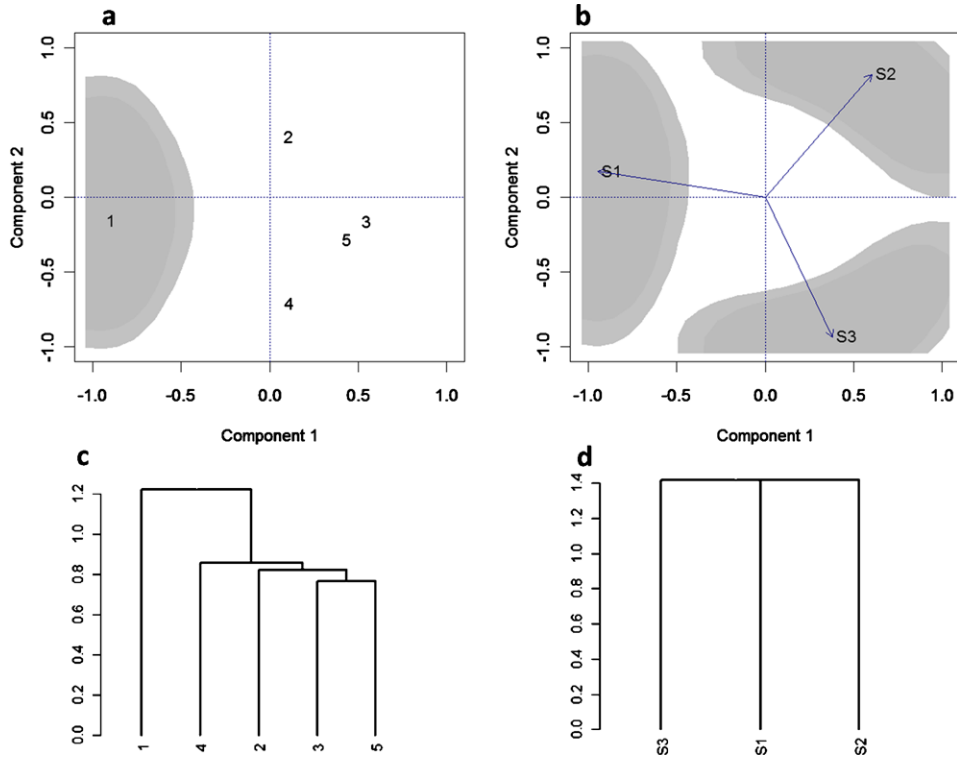
Figure 1.   Simulated data with five rows (1–5) and three columns (S1–S3): (a) plot of the row scores $\mathbf{U}'\mathbf{D}^{1/2}$ and the bivariate 95% (gray external contour) and 90% (gray internal contour) HPD regions; (b) plot of the column scores $\mathbf{V}'\mathbf{D}^{1/2}$ and the bivariate 0.95 (gray external contour) and 0.90 (gray internal contour) HPD regions. Only row 1 and columns S1, S2, and S3 which do not include the null point $(0, 0)$ at the 0.90 HPD probability level are depicted; (c) dendrogram of the five rows using the first two left singular vectors; (d) dendrogram of the three columns using the first two right singular vectors.

scores in row 1. For clarity, the 0.90 HPD and 0.95 HPD regions for the other row scores were not drawn because they contain the null point $(0, 0)$, which is evidence that their contribution to the interaction was not statistically significant. Analogously, from Figure 1(b) it can be seen that the posterior densities for the scores of S1, S2 and S3 do not include the null point $(0, 0)$ at either of the two probability levels; thus we can conclude that, given the data, there is enough evidence that the multiplicative column effect is significant at all its levels. Furthermore, since there is no overlapping of the interaction scores for columns at any of the 0.90 HPD and 0.95 HPD regions, we can conclude that among these regions there are different interaction effects that are statistically significant.

As an additional descriptive tool, we performed a hierarchical cluster algorithm with a complete linkage strategy based on the posterior means of the Euclidean distances between the rows (and columns) of the matrix $\mathbf{U}$ ($\mathbf{V}$); their dendrograms are presented in Figures 1(c) and (d), respectively. Row 1 is the farthest score from $(0, 0)$ and significant for the interaction; it does not cluster with any of the other non-significant genotypes [Figure 1(c)]. The three columns do not form any clear clusters [Figure 1(d)].
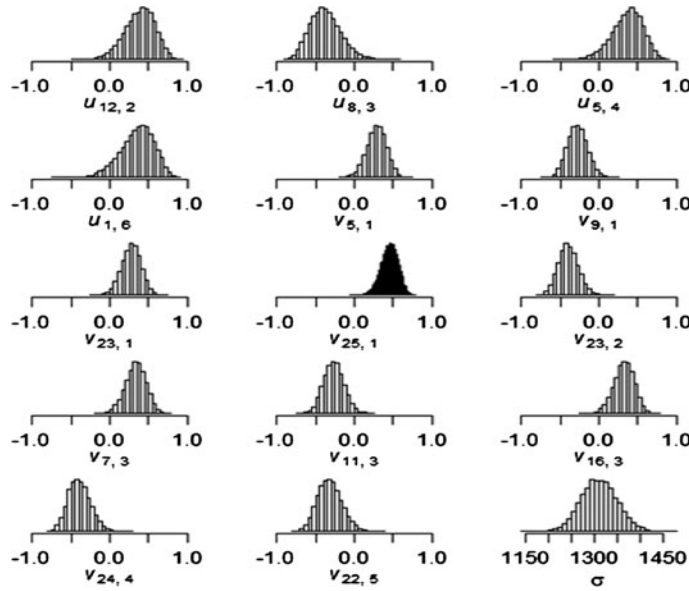
Figure 2.   Plant breeding data of 12 genotypes in 25 environments. Histograms of the MCMC samples from the marginal posterior distributions of $u_{12,2}$, $u_{8,3}$, $u_{5,4}$, $u_{1,6}$, $v_{5,1}$, $v_{9,1}$, $v_{23,1}$, $v_{25,1}$, $v_{23,2}$, $v_{7,3}$, $v_{11,3}$, $v_{16,3}$, $v_{24,4}$, $v_{22,5}$ and σ.

In the next section, we illustrate the results of our prior selection and MCMC sampling strategy to analyze the genotype $\times$ environment interaction in the context of a multi-environment and multi-year plant breeding trial with genotypes evaluated under different environmental conditions during two consecutive years.

3.2. PLANT BREEDING DATA

The multi-environment multi-year plant breeding trial analyzed in this section comprises 12 maize hybrids evaluated for grain yield in 25 environments for two consecutive years. The 12 hybrids were arranged in a randomized complete block design with two replicates in each environment and year. The effect of the design (i.e., complete or incomplete block) is easily incorporated into the model.

The first-year data were used to elicit the prior to analyze data from the second year. The posterior means of the 38 linear parameters (overall mean, 12 genotypic effects and 25 environmental effects) are given in Table B.1 (Appendix B). The histograms of the posterior means of several bilinear interaction parameters depicted in Figure 2 show bell-shaped marginal posterior distributions on $(-1, 1)$. The histograms of the MCMC samples of the marginal posteriors of $\lambda_1 - \lambda_{11}$ are also shown and, as expected, the posterior densities of the late singular values move toward zero [Figure 3(a)]. The posterior densities of the cumulative proportion of variance explained by the eigenvalues show that five components explained about 90% of the interaction variance [Figure 3(b)].

The Bayes factor computed for determining the number of λ's indicated that the model with three GE bilinear components is appropriate. Assuming that the prior odds is 1, the Bayes factor in favor of model with two components ($\lambda_2 \neq 0, \lambda_3 = 0$) when comparing
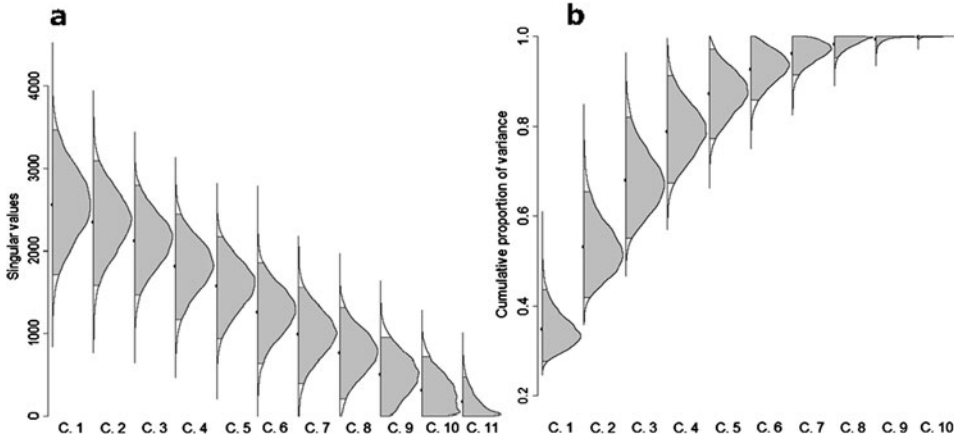
Figure 3. Plant breeding data of 12 genotypes in 25 environments: (a) Posterior densities and 0.95 HPD regions of the singular values, $\lambda_1, \ldots, \lambda_{11}$ (C.1–C.11); (b) posterior densities and 0.95 HPD regions of the cumulative proportion of variance $\phi_t = \dfrac{\sum_{k=1}^{t} \lambda_k^2}{\sum_{k=1}^{\min(r,c)-1} \lambda_k^2}, t = 1, \ldots, \min(r, c) - 2$.

against the alternative model with at least three components ($\lambda_3 \neq 0$) is 0.0022, whereas the value of the Bayes factor indicated that the model with three components ($\lambda_3 \neq 0, \lambda_4 = 0$) is 104.15 times more probable than the model with at least four components ($\lambda_4 \neq 0$). These results of three significant bilinear components ($\lambda_3 \neq 0$) is in agreement with those usually found in the analyses of plant breeding trials where the complexity of the GE requires more than one bilinear component to be retained in the model.

In general, the values of u_{i1} and v_{j1} were, in absolute terms, larger than the values of u_{i2} and v_{j2}, whereas the standard deviations (SD) of u_{i1} and v_{j1} were smaller than those of u_{i2} and v_{j2}. Consequently, the lengths of the HPD regions were narrower for the first bilinear components of genotypes and environments (u_{i1} and v_{j1}) than for the second bilinear components, u_{i2} and v_{j2} (data not shown). In summary, there is more posterior uncertainty in the second bilinear components for genotypes and sites than in the first bilinear components.

3.2.1. Credible Regions of the First Two Bilinear Terms of the Linear–Bilinear Model

Given in Table 2 are the posterior means of the 11 singular values, together with the values of the eigenvectors of genotypes and environments whose 0.95 HPD for their corresponding scores do not contain the null value. As the singular values decrease in size, their SDs also decrease. The scores of genotype 12 ($u_{12,2}$) and the scores of environments S5, S9, S23, and S25 ($v_{5,1}, v_{9,1}, v_{23,1}, v_{25,1}, v_{23,2}$) significantly contributed to genotype $\times$ environment variability, as shown by the bivariate 0.90 HPD regions, whereas the other genotypes and environments did not significantly contribute to that variability.

Posterior modes of the first and second components for the scores of environments and genotypes, with their associated HPD regions, indicate the scores that significantly contributed to the interaction between genotypes and environments (Figure 4). In this biplot,

Table 2. Posterior summary (Mean), standard deviation (SD), quartiles ($q_{0.25}$, $q_{0.50}$, and $q_{0.75}$), 0.90 HPD, and 0.95 HPD intervals computed with 20,000 approximately independent samples simulated from the posterior distribution of the residual variance (σ), all singular values ($\lambda_1, \ldots, \lambda_{11}$) and the right and the left singular vector elements of genotypes and environments, respectively, whose 0.90 HPD and 0.95 HPD intervals do not contain the null value $(0, 0)$. Data for grain yield measured in kilograms per hectare.

Parameter	Mean	SD	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	0.90 HPD interval		0.95 HPD interval	
						Lower	Upper	Lower	Upper
σ	1311.00	39.13	1237.00	1310.00	1389.00	1247.40	1375.18	1234.90	1386.49
λ_1	2572.00	451.76	1683.00	2574.00	3446.00	1835.96	3323.21	1709.64	3468.89
λ_2	2354.00	388.40	1574.00	2364.00	3088.00	1730.65	3005.55	1585.22	3098.23
λ_3	2123.00	340.34	1429.00	2132.00	2763.00	1549.53	2662.10	1467.47	2794.07
λ_4	1815.00	328.85	1146.00	1823.00	2435.00	1277.49	2358.62	1167.92	2452.60
λ_5	1576.00	317.49	933.40	1586.00	2173.00	1058.67	2096.67	935.42	2173.17
λ_6	1262.00	311.66	630.20	1270.00	1847.00	758.68	1780.72	643.20	1856.85
λ_7	992.00	297.34	385.50	998.90	1555.00	506.17	1479.93	394.04	1562.37
λ_8	762.60	283.71	182.70	767.10	1308.00	299.76	1242.21	199.20	1321.56
λ_9	506.10	260.09	42.12	501.30	1027.00	15.90	865.98	0.04	949.55
λ_{10}	315.40	217.51	8.89	289.60	794.40	0.01	614.78	0.01	714.54
λ_{11}	167.40	153.72	1.82	124.30	553.80	0.00	388.81	0.00	472.82
$u_{12,2}$	0.3850	0.1966	−0.0371	0.4009	0.7242	0.0815	0.7176	0.0005	0.7524
$u_{8,3}$	−0.3714	0.1988	−0.7131	−0.3873	0.0618	−0.7054	−0.0679	−0.7464	0.0129
$u_{5,4}$	0.3671	0.2001	−0.0775	0.3838	0.7096	0.0490	0.6941	−0.0361	0.7358
$u_{1,6}$	0.3533	0.2163	−0.1170	0.3747	0.7167	0.0156	0.7073	−0.0702	0.7510
$v_{5,1}$	0.2878	0.1235	0.0389	0.2908	0.5203	0.0868	0.4911	0.0482	0.5271
$v_{9,1}$	−0.2769	0.1281	−0.5182	−0.2803	−0.0182	−0.4805	−0.0587	−0.5220	−0.0234
$v_{23,1}$	0.2779	0.1234	0.0276	0.2818	0.5106	0.0797	0.4846	0.0367	0.5167
$v_{25,1}$	0.4589	0.1104	0.2288	0.4637	0.6607	0.2804	0.6400	0.2431	0.6713
$v_{23,2}$	−0.3958	0.1268	−0.6263	−0.4017	−0.1312	−0.6090	−0.1960	−0.6338	−0.1419
$v_{7,3}$	0.3350	0.1323	0.0636	0.3402	0.5773	0.1221	0.5558	0.0758	0.5849
$v_{11,3}$	−0.2652	0.1319	−0.5161	−0.2668	0.0005	−0.4895	−0.0576	−0.5224	−0.0104
$v_{16,3}$	0.3249	0.1304	0.0547	0.3301	0.5674	0.1104	0.5372	0.0679	0.5755
$v_{24,4}$	−0.3918	0.1418	−0.6427	−0.4011	−0.0907	−0.6151	−0.1559	−0.6624	−0.1165
$v_{22,5}$	−0.3166	0.1581	−0.6007	−0.3248	0.0139	−0.5735	−0.0564	−0.6060	0.0055

the bivariate 0.90 HPD and 0.95 HPD regions are shown only for genotype 12 [Figure 4(a)] and environments S5, S9, S23, and S25 [Figure 4(b)]. These are the genotypes and environments that contributed significantly to the interaction, as their HPD regions did not include the null point $(0, 0)$.

Some description of the joint response of genotypes and environments can be given; for example, environments S5 and S25 and genotype 12 are located far from the center on the upper right-hand side of Figures 4(a) and (b), whereas genotype 2 is the farthest point from the center of the figure on the lower left-hand side, and is negatively related to genotype 12 and to environments S5 and S25. This result indicated that genotype 12 has significant positive interaction with S5 and S25, whereas genotype 2 has negative interaction with those environments. Concerning the relationships among environments, S9 and S23 are located in opposite directions of the biplot [Figure 4(b)] and can be considered as two very different environments in terms of genotype × environment interaction. Furthermore, environments S5, S23, and S25 do not overlap with environment S9; therefore, they
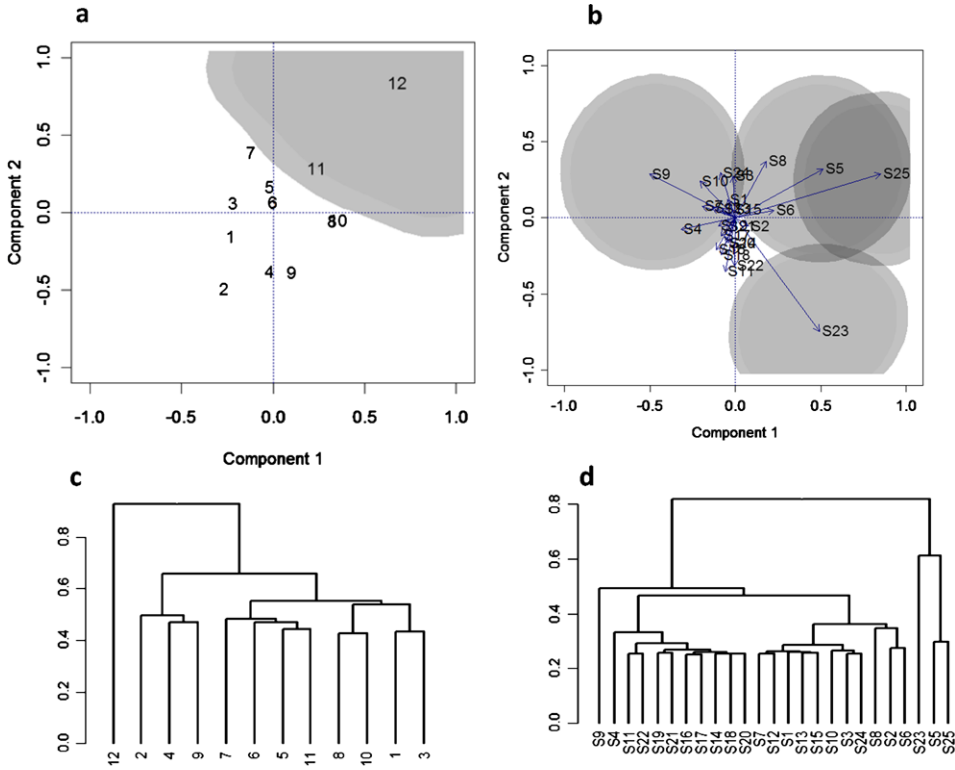
Figure 4. Plant breeding data of 12 genotypes in 25 environments: (a) plot of the bivariate row scores $\mathbf{U}'\mathbf{D}^{1/2}$ and the bivariate 0.95 (gray external contour) and 0.90 (gray internal contour) HPD regions; (b) plot of the bivariate column scores $\mathbf{V}'\mathbf{D}^{1/2}$ and the bivariate 0.95 (gray external contour) and 0.90 (gray internal contour) HPD regions. Only genotype 12, and environments S5, S8, S9, S23, and S25 which do not include the null point $(0,0)$ at the 90 HPD probability level are depicted; (c) dendrogram of the 12 genotypes using the first two left singular vectors; (d) dendrogram of the 25 environments using the first two right singular vectors.

are also significantly different from environment S9 in terms of genotype × environment interaction variability. Since environments S5 and S25 do overlap with each other, they form a homogeneous group of environments but different from S9. Environments S5 and S25 did show some degree of similarity with S9.

Since the degree of overlap between significant environments based on the 0.90 HPD and 0.95 HP regions differed and may not be clearly represented in the biplot, we performed a hierarchical cluster algorithm with a complete linkage strategy based on the posterior means of the Euclidean distance between the rows of the interaction matrices for genotypes and environments; the dendrograms are presented in Figures 4(c) and (d), respectively. This complements what is depicted in Figures 4(a) and (b) very well. For example, genotype 12 forms a singleton opposite to the groups formed by the other genotypes [Figure 4(c)].

Concerning the environments, S5, S23, S25, and S9 are in opposite groups [Figure 4(d)]. Furthermore, as already pointed out in the biplots of environments, environments S5, S23, and S25 formed a group in which each group member significantly contributes to genotype × environment variability; this group shows that S5 and S25 clustered earlier and S23

joined them later. Environments S5, S25 and S23 are different from environment S9. The overlapping of S5 and S25 is much more pronounced than the overlapping of S23; thus they joined early in the cluster. This approach for identifying subsets of homogeneous genotypes and environments that cause significant genotype × environment interaction is analogous (but not the same) to that presented by Burgueño et al. (2008) using a frequentist inference on a linear–bilinear model that belongs to the family of linear–bilinear models employed in this study.

Genotypes had longer HPD regions along the first and second bilinear terms than environments, indicating their large SD and great length of the HPD, which reflected the uncertainty of these estimates at both probability levels.

3.2.2. Implications for Breeding Trials of Bayesian Inference of Linear–Bilinear Models

The conditional posterior estimates of Bayesian linear–bilinear models for plant breeding data have the following advantages: (i) they provide a natural method for deriving confidence regions around the genotypic and environmental interaction parameters given by their scores, as represented in the biplot (and/or dendrogram); (ii) they facilitate the identification of genotypes and environments that cause significant interaction and allow detecting groups of genotypes and environments with similar responses; (iii) they deal with unbalanced data (always present in plant breeding trials) in a natural manner; (iv) they can be used to efficiently incorporate information from historical plant breeding trials (prior) on environmental and genotypic means, or on dispersion parameters such as environmental, genotypic, or error variances; (v) they can be used naturally with unequal cell size; and (vi) they provide an efficient test for the significance of the number of GE bilinear components to be retained in the model.

Linear-bilinear models such as AMMI offer a family of models, rather than a single model; the general Bayesian computational methodology developed in this study can be applied to other linear–bilinear models by fixing certain parameters equal to 0 and relaxing some model constraints. For example, for $\boldsymbol{\alpha} = 0$, $\mathbf{Y} = \mu \mathbf{1}_r \mathbf{1}_c' + \boldsymbol{\beta}' \otimes \mathbf{1}_r + \mathbf{UDV}' + \mathbf{E}$ is the column (site) regression model (SREG); for $\boldsymbol{\beta} = 0$, $\mathbf{Y} = \mu \mathbf{1}_r \mathbf{1}_c' + \boldsymbol{\alpha} \otimes \mathbf{1}_c' + \mathbf{UDV}' + \mathbf{E}$ is the row (genotype) regression model (GREG); and for $\boldsymbol{\alpha} = \boldsymbol{\beta} = 0$, $\mathbf{Y} = \mu \mathbf{1}_r \mathbf{1}_c' + \mathbf{UDV}' + \mathbf{E}$ is the complete multiplicative model (COMM). As the frequentist mixed-effect linear–bilinear model leads to a factor analytic structure of rows, columns and/or their interaction, the Bayesian paradigm of linear–bilinear models can also be represented in a factor analytic form.

This new approach offers new opportunities for efficiently incorporating historical data on environments and genotypes that should be useful for breeder's objectives, as well as for forming density regions around the estimated interaction parameters. Furthermore, with the methodology presented in this article, meta-analysis (hierarchical analysis) involving year or other factors can be analyzed naturally. Although the computer time needed to process large plant breeding trials can be substantially greater than that needed to fit frequentist fixed or mixed linear–bilinear models, the continuous increase in computer power will

minimize this disadvantage of the Bayesian estimation of linear–bilinear models over time. The Bayesian inference methodology described here is available in R in the following page of the CIMMYT website: *http://www.cimmyt.org/english/wps/biometrics/index.htm*.

4. CONCLUSIONS

In this research, we applied Bayesian inference for linear–bilinear models by using the multivariate von Mises–Fisher distribution as a prior for interaction parameters. In contrast to previous approaches, this estimation is not performed on the orthonormal eigenvectors but rather on the orthonormal matrices $\mathbf{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_t)$ and $\mathbf{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_t)$, and is done based on an MCMC sample from their posterior distribution; this satisfies model constraints and offers statistical inferential tools such as confidence regions for interaction parameters. Two data sets were used, one containing simulated data and one real plant breeding data. Results of the plant breeding trials show the usefulness of this general Bayesian approach for breeding trials and for detecting groups of genotypes and environments that cause interaction. For similar-structures data, this method is promising in that confidence regions for the GE interaction terms may be derived, unbalanced data are handled well, and the number of components to be retained in the model can be assessed by the Bayes factor. The Bayesian inference methodology described here can be extended to other linear–bilinear models by fixing certain parameters equal to zero and relaxing some model constraints.

APPENDIX A: THE VON MISES–FISHER DISTRIBUTION

The set of $r \times k$ orthonormal matrices is called the Stiefel manifold, which is denoted as $\mathcal{V}_{k,r}$. A probability distribution on $\mathcal{V}_{k,r}$, whose density has exponential form with linear and quadratic terms, is the matrix Bingham–von Mises–Fisher Distribution. The density function is given by

$$p(X \mid \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) \propto \text{etr}(\boldsymbol{C}'X + \boldsymbol{B}X'\boldsymbol{A}X)$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ may be assumed symmetric and diagonal matrices, respectively. A random variable X with von Mises–Fisher distribution (Khatri and Mardia 1977) is denoted as $X \sim \text{BMF}(\boldsymbol{A} = \boldsymbol{0}, \boldsymbol{B} = \boldsymbol{0}, \boldsymbol{C})$. The normalization constant of the von Mises–Fisher density is given by the hypergeometric function of a matrix argument ${}_0\text{F}_1(\frac{1}{2}r, \frac{1}{4}D_\phi^2)$, where D_ϕ is the diagonal matrix of singular values of $\boldsymbol{C}$ (Herz 1955; James 1964).

APPENDIX B: RESULTS FROM PLANT BREEDING DATA

Table B.1. Posterior summary (Mean), standard deviation (SD), quartiles ($q_{0.25}$, $q_{0.50}$ and $q_{0.75}$) and 0.95 HPD intervals computed with 5000 approximately independent samples simulated from the joint posterior distribution of the 38 linear effects for plant breeding data of grain yield measured in kilograms per hectare.

Parameter	Mean	SD	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	0.90 HPD interval		0.95 HPD interval	
						Lower	Upper	Lower	Upper
μ	3848.00	30.68	3788.00	3848.00	3909.00	3797.57	3898.59	3786.54	3907.27
α_1	135.90	102.92	−64.98	135.40	338.50	−25.06	312.32	−65.81	337.29
α_2	317.50	102.29	115.70	318.10	516.70	151.89	487.33	116.23	517.01
α_3	−64.68	103.26	−269.50	−64.43	134.80	−235.00	103.67	−265.60	138.61
α_4	2.75	102.62	−199.90	2.90	204.10	−157.74	178.77	−202.18	200.93
α_5	93.39	101.90	−106.50	93.61	292.90	−73.23	261.02	−106.95	292.18
α_6	181.00	101.81	−19.13	181.50	379.60	12.22	347.25	−16.59	381.85
α_7	499.30	103.21	297.00	499.20	704.60	329.03	668.15	291.42	697.90
α_8	64.30	102.72	−137.90	64.94	265.00	−99.60	237.96	−141.13	261.47
α_9	129.30	102.34	−72.10	129.60	328.40	−34.99	300.58	−75.84	324.12
α_{10}	−479.00	103.10	−682.50	−478.10	−278.20	−652.14	−311.59	−683.30	−279.28
α_{11}	−270.10	103.02	−471.50	−270.00	−66.31	−436.12	−99.78	−471.13	−66.11
α_{12}	−609.70	102.61	−809.80	−609.80	−405.00	−783.02	−447.03	−811.81	−407.89
β_1	134.50	151.21	−163.50	134.60	431.70	−109.56	386.25	−154.36	438.26
β_2	−533.60	151.39	−828.60	−532.60	−238.70	−784.48	−287.41	−831.18	−242.05
β_3	864.20	151.39	565.70	864.00	1163.00	611.24	1111.18	556.00	1150.01
β_4	−377.40	151.37	−677.60	−376.30	−76.97	−626.95	−130.58	−683.14	−84.26
β_5	1537.00	150.94	1240.00	1537.00	1837.00	1283.94	1782.23	1253.61	1848.75
β_6	−96.78	150.28	−390.40	−96.82	196.90	−344.40	150.31	−385.25	199.82
β_7	−1761.00	151.22	−2059.00	−1761.00	−1465.00	−2004.18	−1506.47	−2054.37	−1460.12
β_8	348.20	151.61	46.56	347.90	645.20	98.69	596.65	60.80	658.62
β_9	−1468.00	150.96	−1763.00	−1468.00	−1176.00	−1714.35	−1219.74	−1762.29	−1175.01

Table B.1. (Continued.)

Parameter	Mean	SD	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	0.90 HPD interval		0.95 HPD interval	
						Lower	Upper	Lower	Upper
β_{10}	−745.10	150.44	−1039.00	−745.00	−447.60	−997.52	−504.73	−1042.34	−452.14
β_{11}	−1202.00	151.64	−1497.00	−1203.00	−904.80	−1445.44	−949.06	−1496.57	−904.36
β_{12}	−1900.00	152.11	−2200.00	−1899.00	−1604.00	−2156.86	−1656.97	−2198.63	−1603.46
β_{13}	−971.60	150.75	−1268.00	−972.30	−676.10	−1217.06	−722.63	−1257.52	−668.61
β_{14}	−548.60	151.84	−851.10	−548.20	−252.70	−801.42	−301.01	−850.60	−252.51
β_{15}	−2215.00	151.62	−2513.00	−2214.00	−1920.00	−2461.53	−1962.21	−2517.25	−1925.26
β_{16}	87.58	150.95	−208.80	88.78	382.50	−158.01	338.84	−200.56	389.50
β_{17}	160.40	152.11	−138.50	161.60	460.70	−78.12	420.05	−138.58	460.65
β_{18}	1848.00	152.18	1546.00	1849.00	2148.00	1601.28	2102.67	1540.92	2140.93
β_{19}	−865.20	150.62	−1159.00	−865.60	−566.70	−1100.99	−607.88	−1161.24	−569.44
β_{20}	−1112.00	151.82	−1407.00	−1112.00	−810.00	−1362.40	−859.85	−1411.64	−815.98
β_{21}	−1627.00	151.76	−1930.00	−1627.00	−1332.00	−1878.18	−1378.41	−1929.72	−1331.88
β_{22}	4917.00	151.56	4621.00	4917.00	5212.00	4663.11	5158.02	4624.65	5214.83
β_{23}	2865.00	151.21	2570.00	2866.00	3160.00	2612.72	3110.06	2565.08	3155.47
β_{24}	911.80	149.76	614.40	911.10	1204.00	668.50	1158.86	625.06	1212.78
β_{25}	1749.00	150.27	1457.00	1747.00	2045.00	1492.73	1985.95	1456.93	2044.38

ACKNOWLEDGEMENTS

REFERENCES

Burgueño, J., Crossa, J., Cornelius, P. L., and Yang, R.-C. (2008), "Using Factor Analytic Models for Joining Environments and Genotypes Without Crossover Genotype × Environment Interaction," *Crop Science*, doi:*10.2135/cropsci2007.11.0632*.

Chib, S. (1995), "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

Cornelius, P. L., and Seyedsadr, M. S. (1997), "Estimation of General Linear–Bilinear Models for Two-Way Tables," *Journal of Statistical Computation and Simulation*, 58, 287–322.

Cornelius, P. L., Seyedsadr, M. S., and Crossa, J. (1992), "Using the Shifted Multiplicative Model to Search for 'Separability' in Crop Cultivar Trials," *Theoretical and Applied Genetics*, 84, 161–172.

Cornelius, P. L., Crossa, J., and Seyedsadr, M. S. (1994), "Tests and Estimators of Multiplicative Models for Variety Trials," in *Proceedings of the 1993 Kansas State University Conference on Applied Statistics in Agriculture*, Manhattan, Kansas.

Cornelius, P. L., Crossa, J., and Seyedsadr, M. S. (1996), "Statistical Tests and Estimators of Multiplicative Models for Genotype-by-Environment Interaction," in *Genotype-by-Environment Interaction*, eds. M. S. Kang and H. G. Gauch, Boca Raton: CRC Press, pp. 199–234.

Crossa, J., Yang, R.-C., and Cornelius, P. L. (2004), "Studying Crossover Genotype × Environment Interaction Using Linear–Bilinear Models and Mixed Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 9, 362–380.

Crossa, J., Perez-Elizalde, S., Jarquin, D., Cotes, J. M., Viele, K., Liu, G., and Cornelius, P. L. (2011), "Bayesian Estimation of the Additive Main Effects and Multiplicative Interaction (AMMI) Model," *Crop Science*, 51, 1458–1469.

Gabriel, K. R. (1978), "Least Squares Approximation of Matrices by Additive and Multiplicative Models," *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 40, 186–196.

Gauch, H. G. (1988), "Model Selection and Validation for Yield Trials With Interaction," *Biometrics*, 44, 705–715.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.

Gollob, H. F. (1968), "A Statistical Model Which Combines Features of Factor Analytic and Analysis of Variance," *Psychometrika*, 33, 73–115.

Han, C., and Carlin, B. P. (2001), "Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative Review," *Journal of the American Statistical Association*, 96, 1122–1132.

Herz, C. S. (1955), "Bessel Functions of Matrix Argument," *Annals of Mathematics*, 61, 474–523.

Hoff, P. D. (2009), "Simulation of the Matrix Bingham–von Mises–Fisher Distribution, With Applications to Multivariate and Relational Data," *Journal of Computational and Graphical Statistics*, 18, 438–456.

James, A. T. (1964), "Distributions of Matrix Variates and Latent Roots Derived From Normal Samples," *Annals of Mathematical Statistics*, 35, 475–501.

Jeffreys, H. (1961), *Theory of Probability*, Oxford: Clarendon Press.

Johnson, D. E., and Graybill, G. A. (1972), "An Analysis of a Two-Way Model With Interaction and No Replication," *Journal of the American Statistical Association*, 67, 862–868.

Kass, R. E., and Raftery, A. (1995), "Bayesian Factors," *Journal of the American Statistical Association*, 90, 773–795.

Kempton, R. A. (1984), "The Use of Biplots in Interpreting Variety by Environment Interactions," *Journal of Agriculture Science*, 103, 123–135.

Khatri, C. G., and Mardia, K. V. (1977), "The von Mises–Fisher Matrix Distribution in Orientation Statistics," *Journal of the Royal Statistical Society. Series B. Methodological*, 39 (1), 95–106.

Koev, P., and Edelman, A. (2006), "The Efficient Evaluation of the Hypergeometric Function of a Matrix Argument," *Mathematics of Computation*, 75, 833–845.

Liu, G. (2001). "Bayesian Computation for General Linear–Bilinear Models," Ph.D. dissertation, Department of Statistics, University of Kentucky, Lexington, KY.

Mandel, J. (1969), "The Partitioning of Interaction in Analysis of Variance," *Journal of Research of the National Bureau of Standards, Series B*, 73, 309–328.

Mandel, J. (1971), "A New Analysis of Variance Models for Non-Additive Data," *Technometrics*, 13, 1–18.

Marasinghe, M. G. (1985), "Asymptotic Tests and Monte Carlo Studies Associated With the Multiplicative Interaction Model," *Communications in Statistics. Theory and Methods*, 14, 2219–2231.

Mardia, K. V., Kent, J. T., and Bibbi, J. B. (1979), *Multivariate Analysis*, London: Academic Press.

Raftery, A. E., and Lewis, S. M. (1995), "The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms," in *Practical Markov Chain Monte Carlo*, eds. W. R. Gilks, D. J. Spiegelhalter, and S. Richardson, London: Chapman and Hall.

Schott, J. R. (1986), "A Note on the Critical Value in Stepwise Tests of Multiplicative Components of Interactions," *Communications in Statistics. Theory and Methods*, 15, 1561–1570.

Seyedsadr, M. S., and Cornelius, P. L. (1992), "Shifted Multiplicative Models for Non-additive Two-Way Tables," *Communications in Statistics. Simulation and Computation*, 21, 807–822.

Viele, K., and Srinivasan, C. (2000), "Parsimonious Estimation of Multiplicative Interaction in Analysis of Variance Using Kullback–Leibler Information," *Journal of Statistical Planning and Inference*, 84, 201–219.