Another Look at Bayesian Analysis of AMMI Models for Genotype-Environment Data

Julie JOSSE, Fred van EEUWIJK, Hans-Peter PIEPHO, and Jean-Baptiste DENIS

Linear–bilinear models are frequently used to analyze two-way data such as genotype-by-environment data. A well-known example of this class of models is the additive main effects and multiplicative interaction effects model (AMMI). We propose a new Bayesian treatment of such models offering a proper way to deal with the major problem of overparameterization. The rationale is to ignore the issue at the prior level and apply an appropriate processing at the posterior level to be able to arrive at easily interpretable inferences. Compared to previous attempts, this new strategy has the great advantage of being directly implementable in standard software packages devoted to Bayesian statistics such as WinBUGS/OpenBUGS/JAGS. The method is assessed using simulated datasets and a real dataset from plant breeding. We discuss the benefits of a Bayesian perspective to the analysis of genotype-by-environment interactions, focusing on practical questions related to general and local adaptation and stability of genotypes. We also suggest a new solution to the estimation of the risk of a genotype not exceeding a given threshold.

Key Words: Adaptation; AMMI models; Bayesian inference; Genotype-by-environment interaction; Overparameterization; Singular-value decomposition; Stability.

1. INTRODUCTION

Linear–bilinear models (Mandel 1969), also named biadditive models by Denis and Gower (1994, 1996), are often used to model two-way data with interaction. For instance in plant breeding, it is common to study with these models the interaction between genotypes and environments for various purposes such as the evaluation and selection of new varieties (Gauch 1990; Gauch and Zobel 1996). A well-known example of this class of models is

Julie Josse (✉) is a Ph.D. (E-mail: *josse@agrocampus-ouest.fr*), Applied Mathematics Department, Agrocampus Ouest, Rennes, France. Fred van Eeuwijk is Professor, Plant Sciences Department, Wageningen University, Wageningen, The Netherlands. Hans-Peter Piepho is Professor, Crop Science Institute, Hohenheim University, Hohenheim, Germany. Jean-Baptiste Denis is Researcher, Applied Mathematics and Informatics Unit, INRA, Jouy-en-Josas, France.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 19, Number 2, Pages 240–257
DOI: 10.1007/s13253-014-0168-z

the AMMI (additive main effects and multiplicative interaction effects) model which can be defined as follows:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq} + E_{ij}, \quad E_{ij} \sim \mathcal{N}\left(0, \sigma_E^2\right), \tag{1.1}$$

where $\mathbf{Y} = (Y_{ij})$ represents the responses (such as the yield) for genotypes $i = 1, \ldots, I$ and environments $j = 1, \ldots, J$; the additive part (the linear part) is represented by the grand mean μ, the parameters for the main effects $\boldsymbol{\alpha} = (\alpha_i)_{i=1,\ldots,I}$ and $\boldsymbol{\beta} = (\beta_j)_{j=1,\ldots,J}$, and the interaction (the bilinear part) is represented by the Q terms $\sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}$.

Such models are overparameterized, and it is necessary to add constraints to obtain unique estimates for the parameters. In practice, the following constraints are often used:

$$\begin{aligned} \mathbf{1}_I' \boldsymbol{\alpha} = \mathbf{1}_J' \boldsymbol{\beta} &= 0, \\ \mathbf{1}_I' \boldsymbol{\gamma}_q = \mathbf{1}_J' \boldsymbol{\delta}_q &= 0 \quad \text{for } q = 1, \ldots, Q, \\ \boldsymbol{\gamma}' \boldsymbol{\gamma} = \boldsymbol{\delta}' \boldsymbol{\delta} &= \mathbb{I}_Q, \\ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_Q &\geq 0, \\ \gamma_{1q} &\geq 0 \quad \text{for } q = 1, \ldots, Q, \end{aligned} \tag{1.2}$$

where $\mathbf{1}_I'$ and $\mathbf{1}_J'$ are vectors of sizes I and J containing ones, $\boldsymbol{\gamma}_q$ and $\boldsymbol{\delta}_q$ are the qth columns of the matrices $\boldsymbol{\gamma}_{I \times Q}$ and $\boldsymbol{\delta}_{J \times Q}$, and $\mathbb{I}_Q$ is the identity matrix of size Q. The constraints $\lambda_Q \geq 0$ and $\gamma_{1q} \geq 0$ (for $q = 1, \ldots, Q$) allow one to avoid sign indeterminacies in the interaction terms.

With the chosen constraints (1.2) and for a complete dataset, the least squares estimates for the additive part are the same as in the usual analysis of variance framework, i.e., μ is estimated by the grand mean of the data, $\boldsymbol{\alpha}$ is estimated as the vector of column means minus the grand mean, and $\boldsymbol{\beta}$ as the vector of row means minus the grand mean. The least squares estimates for the interaction terms are given by the singular-value decomposition (SVD) of the matrix of interaction residuals, i.e., the row and column centered data matrix. It gives the Q singular values $(\lambda_q)_{(q=1,\ldots,Q)}$ and the left and right singular vectors in the matrices $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$. The number of underlying dimensions Q is here considered to be known.

Recently, some authors (Viele and Srinivasan 2000; Perez-Elizalde, Jarquin, and Crossa 2011) suggested a Bayesian treatment of AMMI models in the framework of genotype-environment (GE) data. Two main points motivate this approach. Firstly, a Bayesian strategy offers the possibility to incorporate in the analysis prior information on the phenomenon under study (experts knowledge, historical data, etc.). Secondly, distributions of any quantity of interest are available through the posterior distributions. These distributions may be straightforwardly used, for instance, to derive credible areas in the biplot representations, which can be helpful for the user to interpret the results. Note that Denis and Gower (1996) suggested a method to obtain confidence regions in the biplot representations in the frequentist framework. However, their proposal is based on asymptotic considerations, which may be far from reality in practice.

Due to a number of clear advantages of Bayesian approaches, their use is significantly increasing for many statistical applications. References in the field of variety trials in-

clude among others Theobald, Talbot, and Nabugoomu (2002) and Edwards and Jannink (2006). However, we can observe that in comparison to the abundant Bayesian literature available for many statistical problems, only a few proposals (Viele and Srinivasan 2000; Perez-Elizalde, Jarquin, and Crossa 2011) have been made for AMMI models. This can be explained by the difficulty to work in an overparameterization framework, i.e., to ensure that priors and posteriors on the parameters take into account constraints as the ones defined in (1.2). Difficulties specifically arise for the interaction terms.

Viele and Srinivasan (2000) put a uniform prior on the first column of the matrix $\boldsymbol{\gamma}$ and on the first column of the matrix $\boldsymbol{\delta}$. Since each vector is a centered vector of unit length, this corresponds to putting a uniform prior on an I-dimensional (respectively on a J-dimensional) unit sphere. For second and later columns $(\boldsymbol{\gamma}_q)$ $((\boldsymbol{\delta}_q))$, uniform distributions are defined conditional on the earlier column(s). Due to the constraints (the vectors must have unit length and be orthogonal to the previous ones), it is not easy to define the supports and to sample from uniform distributions with correct supports. After defining these priors that meet the constraints, the authors implemented a specific Gibbs sampler to obtain draws from the posterior distribution since no closed-form expression is available for the joint posterior distribution. Crossa et al. (2011) further elaborated the approach by Viele and Srinivasan (2000) and suggested another Gibbs sampler, which stabilizes the algorithm, and applied it on a real GE dataset. Crossa et al. (2011) showed that their approach seems promising for, among other things, incorporating inferential statistics in biplot representations.

Smidl and Quinn (2007) and Hoff (2009) suggested a Bayesian treatment of principal components analysis models. From a computational point of view, these models are close to the AMMI one, the main difference being that the linear part only includes grand mean and column main effect (PCA is often defined as the SVD of a column-centered matrix). Smidl and Quinn (2007) and Hoff (2009) impose a uniform prior on the matrix $\boldsymbol{\gamma}$ and on the matrix $\boldsymbol{\delta}$. Such uniform distributions are special cases of von Mises–Fisher (VMF) distributions (Khatri and Mardia 1977), which are distributions over the Stiefel manifold representing the set of orthonormal matrices (Chikuse 2003). Compared to the approach suggested by Viele and Srinivasan (2000), it directly works at the matrix level and not column by column. Consequently, the priors meet the constraints. Using such priors ensures orthonormality constraints at the posterior level because the conditional posterior distributions for the matrices $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are also VMF distributions (but not anymore uniform). Since, again, no closed-form expression is available for the joint posterior distribution, the authors implemented a specific Gibbs sampler. Hoff (2009) suggested a method to draw matrices from VMF distributions, which is implemented in the package rstiefel (Hoff 2012) of the free R software (R Core Team 2013). Using a VMF distribution is thus a way to solve the overparameterization problem for bilinear models.

More recently, in the framework of the analysis of GE data, Perez-Elizalde, Jarquin, and Crossa (2011) also suggested to use as priors for matrices $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ specific VMF distributions that are not uniform ones. More precisely, they suggested to work with an initial guess for the response matrix $\mathbf{Y}$, denoted by $\mathbf{Y}_0$. Such a matrix can be obtained from historical data. Then, they centered this matrix by row and by column and performed

the SVD on the resulting matrix to obtain matrices $\boldsymbol{\gamma}_0$ and $\boldsymbol{\delta}_0$. They defined their prior distributions for $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ as VMF distributions depending on $\boldsymbol{\gamma}_0$ and $\boldsymbol{\delta}_0$. A specific Gibbs sampler was built to obtain draws from the posterior distributions, which are again VMF distributions.

In this paper, we introduce a Bayesian treatment of AMMI models based on a new solution to deal with overparameterization in bilinear models. Compared to the approaches suggested by Viele and Srinivasan (2000) and Perez-Elizalde, Jarquin, and Crossa (2011), our proposal has the great advantage to be easily implementable in the standard BUGS (Bayesian inference Using Gibbs Sampling) softwares. Consequently, it avoids the burden of implementing a specific Gibbs algorithm and possible coding errors. The rationale of the approach and the definition of the priors are presented in Section 2. Then, the method is illustrated through a small simulation study and applied to a real plant breeding dataset in Section 3. Finally, we highlight the benefits of using a Bayesian point of view to answer practical questions raised when analyzing GE data.

2. A BAYESIAN TREATMENT OF AMMI MODELS

2.1. RATIONALE OF THE METHOD

As discussed in the introduction, the difficulty when suggesting a Bayesian approach for the AMMI models is to work in a overparameterization framework and to comply with the model constraints (1.2) at both the prior and posterior levels. Overparameterization can be defined as follows: there is overparameterization when modifications of any nonconstant function of the parameters do not modify the likelihood. The likelihood of model (1.1) is defined by the distribution of the data:

$$[Y_{ij} \mid \mu_{ij}] \sim \mathcal{N}\left(\mu_{ij}, \sigma_E^2\right)$$
$$\text{with } \mu_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}. \tag{2.1}$$

Model (1.1) is overparameterized: the functions of the parameters associated with the constraints used to avoid indeterminacy (such as $\sum_i \alpha_i$) do not influence the likelihood. Indeed, in an analysis of variance framework, a well-known result is that the estimated response $\hat{\mu}_{ij}$ is unique whatever the constraints are (Nelder 1994).

The rationale of the suggested approach is to get rid of the overparameterization issue simply by disregarding the constraints. More precisely, priors are defined for the complete set of parameters (on all the α_i, β_j, γ_{iq}, and δ_{jq}) without considering the constraints. This means that, contrary to the previous attempts (Viele and Srinivasan 2000; Perez-Elizalde, Jarquin, and Crossa 2011), the orthonormality constraints for the interaction terms are not ensured at the prior level. This strategy also implies that more random variables than required are defined, but it is without any consequences on the subsequent analyses.

Since priors are defined on an overparameterized setup, the posteriors will relate to the same overparameterized setup: they do not comply with the constraints (the posterior distributions for the interaction terms do not meet the orthonormality constraints), which

may be problematic for the interpretation of the results. To address this issue, the suggested solution consists in considering and working with functions of the parameters that are identifiable. More precisely, we will consider the expectation of the data μ_{ij} (defined in (2.1)), which contains all the information coming from the data. Since a posterior distribution is available for all of the parameters, a posterior distribution is also available for the μ_{ij}.

As is often the case, no closed-form expressions exist for the posterior distributions, but an algorithm (detailed in Section 2.4) is available to obtain S draws from the posteriors, leading to a Markov Chain Monte Carlo (MCMC) sample of size S to estimate the posterior distributions. This means that, concretely, S matrices of size $I \times J$ are available as draws from the posterior distributions of the μ_{ij} (respecting formula (2.1)). Thus, it is possible to apply a postprocessing on each matrix ($s = 1, \ldots, S$) performing the classical procedure (in accordance with the chosen constraints): each matrix is centered by row and by column, and an SVD is applied on the resulting matrix. Consequently, for each s, new parameters $(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, (\lambda_q)_{(q=1,\ldots,Q)})$ meeting the constraints (1.2) are available. Consequently, draws in the posterior distribution of the parameters (taking the S new values) are available. Such a postprocessing makes it easier to interpret the results.

2.2. Prior Distributions on the Parameters

The following independent prior distributions are used:

$$\begin{aligned}
\mu &\sim \mathcal{N}\big(m, s_\mu^2\big),\\
\alpha_i &\sim \mathcal{N}\big(0, s_\alpha^2\big),\\
\beta_j &\sim \mathcal{N}\big(0, s_\beta^2\big),\\
(\lambda_q)_{q=1,\ldots,Q} &\sim \text{ordered sample of } Q \text{ independent } \mathcal{N}^+\big(0, s_\lambda^2\big),\\
\gamma_{1q} &\sim \mathcal{N}^+(0,1) \quad \text{for } q = 1, \ldots, Q,\\
\gamma_{iq} &\sim \mathcal{N}(0,1) \quad \text{for } i > 1 \text{ and } q = 1, \ldots, Q,\\
\delta_{jq} &\sim \mathcal{N}(0,1) \quad \text{for } j \geq 1 \text{ and } q = 1, \ldots, Q,\\
\sigma_E &\sim U(0, S_{\mathrm{ME}}),
\end{aligned} \tag{2.2}$$

where $\mathcal{N}^+$ stands for the truncated normal distribution on the positive values, and U is the uniform distribution. The constant m is the prior mean of Y_{ij}, s_μ is the uncertainty around m, s_α is the uncertainty of the genotype main effect, s_β is the uncertainty of the environment main effect, s_λ is the uncertainty of the genotype-environment interaction on each of its Q components, and S_{ME} is the remaining uncertainty of the measurement.

A few comments can be made regarding these specific choices of priors. Any supplementary information available from experts or from historical data can be taken into account by assigning specific values to the constants. Each eigenvalue can also have its own specific prior to express the fact that different amounts of knowledge are available for these eigenvalues. It is common in the Bayesian literature to use gamma and inverse gamma distributions for variance parameters. However, some papers (Gelman 2006) showed that it is more efficient to use uniform priors, which explains our choice for σ_E. Finally, we remark

that it would be possible to use $\gamma_{iq} \sim N(0, I^{-1})$ and $\delta_{jq} \sim N(0, J^{-1})$ to meet closer the chosen constraints (1.2). However, as mentioned previously, a strong point of our approach is that it is not necessary to define priors that comply with the constraints.

2.3. PRIOR DISTRIBUTION ON THE DATA

The impact of the chosen priors on the parameters (2.2) can be assessed by inspecting the induced prior distribution on the data. Due to the presence of the bilinear terms in model (1.1), it is not possible to obtain a closed-form expression for the distribution of Y_{ij}. However, it is possible to simulate draws from the prior distribution of Y_{ij} and to compute explicitly its first two moments. The calculations are detailed in Josse and Denis (2012) and lead to

$$\begin{aligned}
E(Y_{ij}) &= m, \\
V(Y_{ij}) &= s_\mu^2 + s_\alpha^2 + s_\beta^2 + Q s_\lambda^2 + \frac{1}{3} S_{\mathrm{ME}}^2, \\
\operatorname{cov}(Y_{ij}, Y_{ij'}) &= s_\mu^2 + s_\alpha^2, \quad j \neq j', \\
\operatorname{cov}(Y_{ij}, Y_{i'j}) &= s_\mu^2 + s_\beta^2, \quad i \neq i', \\
\operatorname{cov}(Y_{ij}, Y_{i'j'}) &= s_\mu^2, \quad j \neq j' \text{ and } i \neq i'.
\end{aligned}$$

Note that the term $\frac{1}{3} S_{\mathrm{ME}}^2$ in $V(Y_{ij})$ results from the uniform prior distribution on the standard deviation of the errors (2.2).

Inspecting the priors Y_{ij} is an important step in the analysis. Indeed, it is a way for the user to better understand what a priori information is included in the analysis. He can, for instance, see what values of the yield are a priori considered for the genotypes and the environments. Thus, if the user is not satisfied with these values, he can adjust the constants in the definition of the priors (2.2).

From a more technical point of view, we can note that the suggested priors on $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, and $(\lambda_q)_{q=1,\ldots,Q}$ lead to a specific prior for the interaction part of each cell represented by the term $\sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}$: all terms have the same expectation and variance, whereas the covariance between terms is null.

2.4. POSTERIOR DISTRIBUTIONS

Combining the likelihood defined by the distribution of the data (2.1) and the priors (2.2) gives the joint posterior distribution. As mentioned previously, no closed-form solution is available for the posterior distribution, and simulated values from the posterior can be obtained using a Gibbs sampler. However, one of the strongest points of the suggested approach is that it is not necessary to build and implement a specific Gibbs sampler as in Viele and Srinivasan (2000) and in Perez-Elizalde, Jarquin, and Crossa (2011); all that is required is the use of a standard software for Bayesian methods. Consequently, JAGS (Martyn 2003) for instance, one of the softwares using the BUGS syntax to specify Bayesian models, can be used to get draws from the posterior. Thus, in practice, the user can easily implement our approach and reproduce our results. He only needs to define the priors (2.2) by choosing appropriate constants to reflect his degree of knowledge.

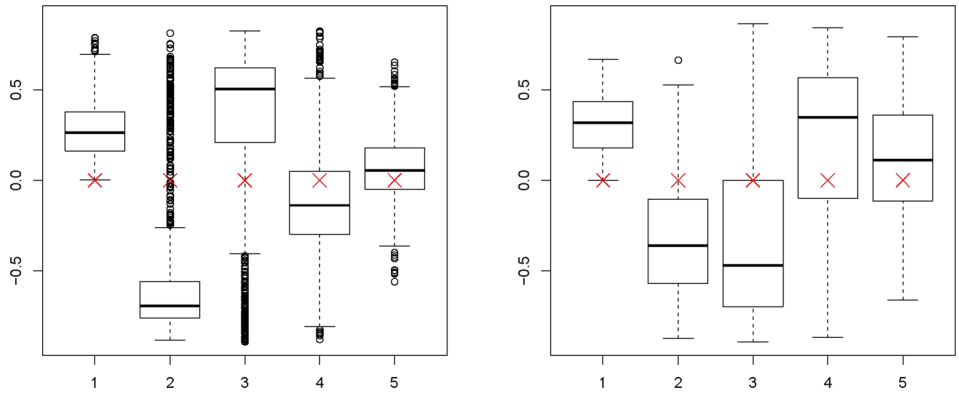

Figure 1. Posterior distribution of $\boldsymbol{\alpha}$ for two simulated datasets based on 1000 draws. A profile is defined as the lines linking $(i, \alpha_i^{(s)})_{i=1,\ldots,5}$ for each draw $s_{(s=1,\ldots,1000)}$; the red line represents the true profile: $(\boldsymbol{\alpha}) = (-1, -1, 0, 1, 1)$. (Color figure online.)

3. RESULTS

3.1. SIMULATION STUDY

To assess the performance of the suggested method, a small simulation study was conducted. Fifty datasets with $I = 5$ rows and $J = 9$ columns were generated according to model (1.1) with one multiplicative term for the interaction and true values for the parameters equal to $\mu = 100$, $(\boldsymbol{\alpha}) = (-1, -1, 0, 1, 1)$, $(\boldsymbol{\beta}) = (-4, -3, -2, -1, 0, 1, 2, 3, 4)$, $\lambda_1 = 12$, $(\boldsymbol{\gamma}) = (\frac{2}{\sqrt{10}}, \frac{1}{\sqrt{10}}, 0, -\frac{1}{\sqrt{10}}, -\frac{2}{\sqrt{10}})$, $(\boldsymbol{\delta}) = (\frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, -\frac{1}{2}, -\frac{1}{2})$, and $\sigma_E = \frac{3}{2}$. The Bayesian method is applied to each of the 50 datasets with the following constants for the priors defined in (2.2): $m = 90$, $s_\mu = 20$, $s_\alpha = 10$, $s_\beta = 10$, $s_\lambda = 10$, and $S_{\text{ME}} = 10$. We run the method with $Q = 2$ components for the interaction part even though only one component was simulated. Consequently, we can assess the capability of the method to detect that the second component is random noise. One chain of size 100,000 is simulated with a burn-in period of size 10,000 and a thinning period of 100. Consequently, a sample of 1000 draws is available for the posterior distributions.

Figure 1 represents the posterior distributions of the $(\alpha_i)_{i=1,\ldots,5}$ for two of the 50 simulations. For each draw $s_{(s=1,\ldots,1000)}$, the points of coordinates $(i, \alpha_i^{(s)})_{i=1,\ldots,5}$ are connected with a line to define a profile. The profile defined by the true values of the parameters is represented in red. Figures 2 and 3 provide the same representation for the interaction terms γ_{i1} and γ_{i2}. The Bayesian strategy can provide a point estimate for the parameters by the mean (or the median) of the posterior distribution. For instance, the point estimate for the parameter α_1 is the mean of the 1000 draws from the posterior distribution of α_1. In addition, the Bayesian strategy has the great advantage of directly providing an estimate for the uncertainty in the parameters. In Figure 1, the uncertainty in α_1 can be defined by a function (credible interval) of the points in the vertical axis having 1 as a coordinate on the horizontal axis. This simulation study shows that with our method, the parameters for

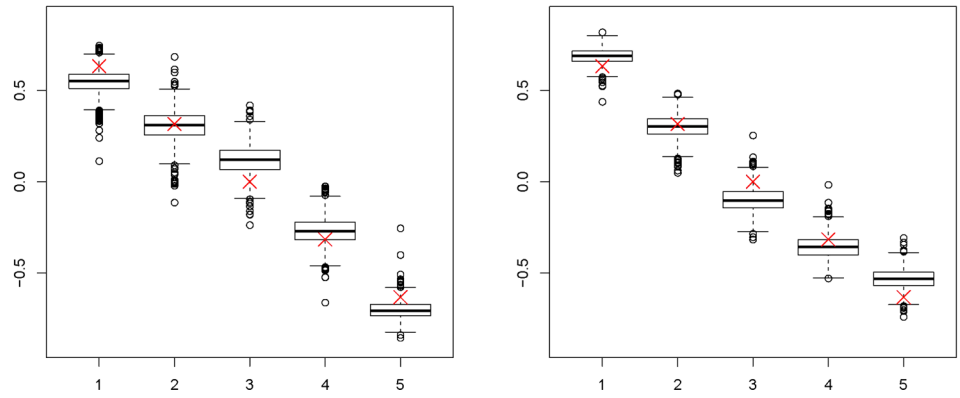

Figure 2. Posterior distribution of $\boldsymbol{\gamma}_1$ for two simulated datasets based on 1000 draws. A profile is defined as the lines linking $(i, \gamma_{i1})_{i=1,\ldots,5}$ for each draw $s_{(s=1,\ldots,1000)}$; the red line represents the true profile $(\boldsymbol{\gamma}) = (\frac{2}{\sqrt{10}}, \frac{1}{\sqrt{10}}, 0, -\frac{1}{\sqrt{10}}, -\frac{2}{\sqrt{10}})$. (Color figure online.)

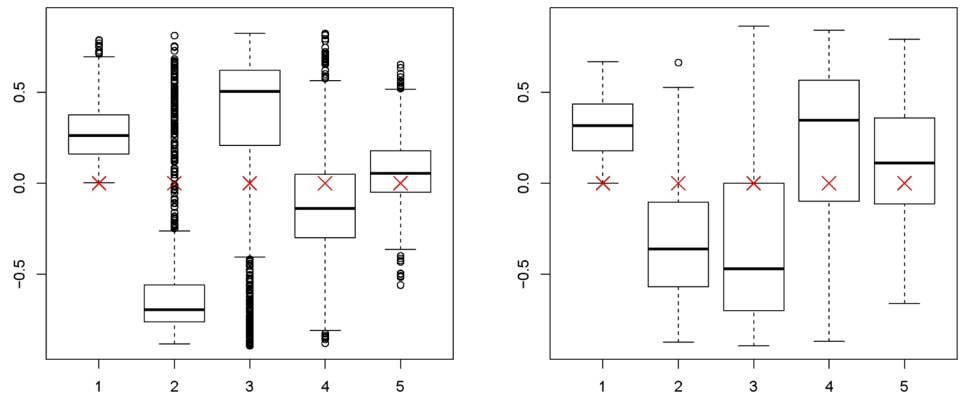

Figure 3. Posterior distribution of $\boldsymbol{\gamma}_2$ for two simulated datasets based on 1000 draws. A profile is defined as the line linking $(i, \gamma_{i2})_{i=1,\ldots,5}$ for each draw $s_{(s=1,\ldots,1000)}$; the red line is the true profile which is the zero line since there is no second multiplicative term in the underlying model. (Color figure online.)

the linear part and the parameters corresponding to the first interaction term are well assessed since the true profiles are within the posterior distributions (Figures 1 and 2). The erratic behavior of the $\boldsymbol{\gamma}_2$ profiles (Figure 3) is in accordance with the fact that the data are simulated with $Q = 1$.

3.2. REAL DATA SET

The method is illustrated on real genotype-environment data for grain yield (kilograms per hectare) in $I = 16$ triticale lines or genotypes and $J = 10$ experiments (environments) across Spain in 1989. Each experiment consisted of a randomized complete block experiment with four replicates. In this paper, we look at the means across the replicates. The data were analyzed with a conventional bilinear model in Royo, Rodriguez, and Romagosa (1993). The six first genotypes correspond to so-called "complete" triticales, and then there were eight so-called "substituted" lines. Finally, two check genotypes were included. The

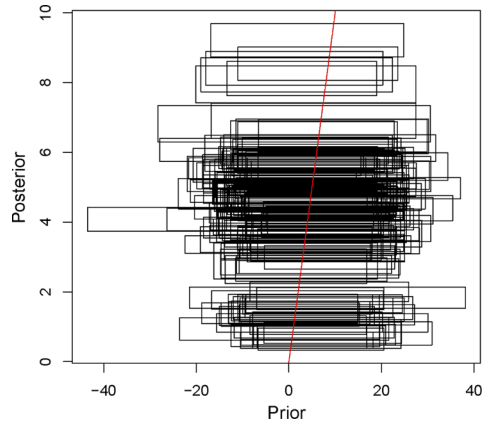

Figure 4. Prior-posterior credible boxes for the μ_{ij} ($i = 1, \ldots, 16$; $j = 1, \ldots, 10$); 2.5 % and 97.5 % quantiles are used to define the boxes around the median positions. Each box corresponds to one μ_{ij}. The red line is the first bisector. (Color figure online.)

difference between the complete and substituted lines resided in the fact that the substituted genotypes have in their full set of chromosomes one chromosome stemming from rye replaced by a chromosome stemming from wheat. This substitution is thought to produce day-length insensitivity and reduce drought stress tolerance and acidity tolerance. The substitution will thus play an important role in mechanisms underlying the GE interaction. Under nonstress conditions, substituted genotypes will out-yield complete genotypes.

The aim of this example is to show how the Bayesian approach may help in answering important questions arising in the context of analyses of GE interactions. We will consider the five following questions: (Q1) What is the genotype with the best performance across environments? (Q2) What is the genotype with the best performance for a specific environment? (Q3) Are the genotypes stable across environments? (Q4) Is it possible to rank the genotypes? (Q5) Can we estimate the probability that a genotype produces less than a certain threshold? To answer these questions, we propose some new graphs. Note that Crossa et al. (2011) and Perez-Elizalde, Jarquin, and Crossa (2011) focused on other practical questions (such as the visualization issues of the uncertainties on the biplot representation), and especially they did not consider the last question.

Assuming that no information is available from experts, the following vague priors are used to illustrate the method: $m = 5$, $s_\mu = 0.8$, $s_\alpha = 0.5$, $s_\beta = 0.5$, $s_\lambda = 0.5\sqrt{IJ}$, and $S_{\text{ME}} = 2$. Our approach also requires as an input the number Q of terms of the interaction part. Many methods are available in the literature (Cornelius, Crossa, and Seyedsadr 1996; Jolliffe 2002; Dias and Krzanowski 2003; Hoff 2007) to select this number. A cross-validation procedure (Josse and Husson 2012) applied to the data matrix centered by row and column (to represent only the interaction part) suggested two terms for the interaction. Note that only one term was studied in Royo, Rodriguez, and Romagosa (1993). The Bayesian approach is performed with one chain of size 100,000 and a thinning period of 100. A sample of 1000 draws is thus available to assess the posterior distributions.

Figure 4 shows the prior (horizontal axis) and posterior (vertical axis) distributions of the μ_{ij} ($i = 1, \ldots, 16$; $j = 1, \ldots, 10$). Instead of directly representing the 1000 values

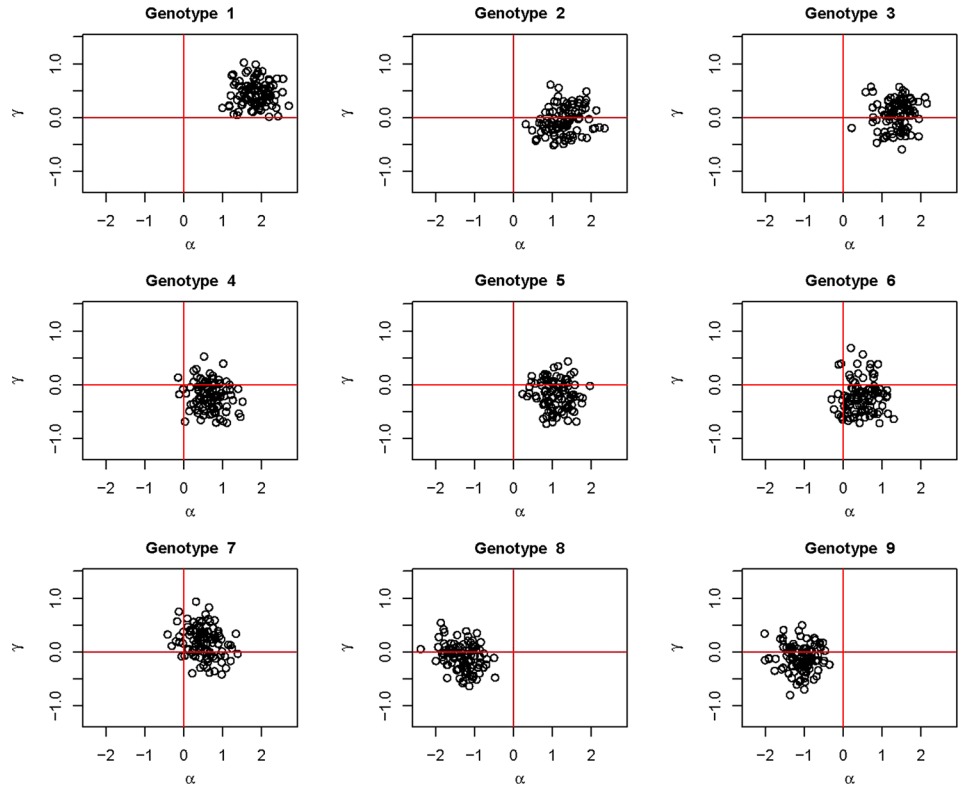

Figure 5. Genotype effect: for each genotype $i_{(i=1,\ldots,9)}$ representation of $(\alpha_i^s\sqrt{J},\gamma_{i,1}^s\lambda_1^s)$, $s=1,\ldots,1000$.

drawn using the algorithm, we built credible boxes using the 2.5 % and 97.5 % quantiles of the distributions. For instance, the box at the top of the figure corresponds to the credible box of μ_{21}. The label of the cell is not included in the graph for the sake of clarity. This representation is helpful to see both the impact of the parameters' priors (defined in (2.2)) on the μ_{ij} priors (looking at the coordinates on the horizontal axis) and the level of involvement of the dataset in the posteriors (looking at the coordinates on the vertical axis). Here, the priors on the parameters lead to priors on the μ_{ij} outside the permissible range (negative yield), which suggests to reduce the variability of the priors. Consequently, we use for the subsequent outputs the following constants: $s_\mu = 0.5$, $s_\alpha = 0.3$ and $s_\beta = 0.3$, $s_\lambda = 0.25\sqrt{IJ}$. More relevant prior information (with historical information for instance) will minimize the probability of generating negative yields. Figure 4 also shows that the data bring substantial information. Indeed, despite the relatively uninformative values for the priors, the posteriors μ_{ij} (which are obtained combining the priors and the data) have values between 0 and 10. In addition, no structure is visible in the prior dimension; this is not the case for the posterior dimension where three distinct clusters appear.

Figures 5 and 6 focus on the genotype effect. For each genotype i, each point has as coordinates the pair (main effect term, first interaction effect term) $(\alpha_i^s\sqrt{J},\lambda_1^s\gamma_{i,1}^s)$ with $s=1,\ldots,1000$. There are 1000 points representing the posterior distribution (instead of having one point for each genotype as in the frequentist framework) that provides a di-

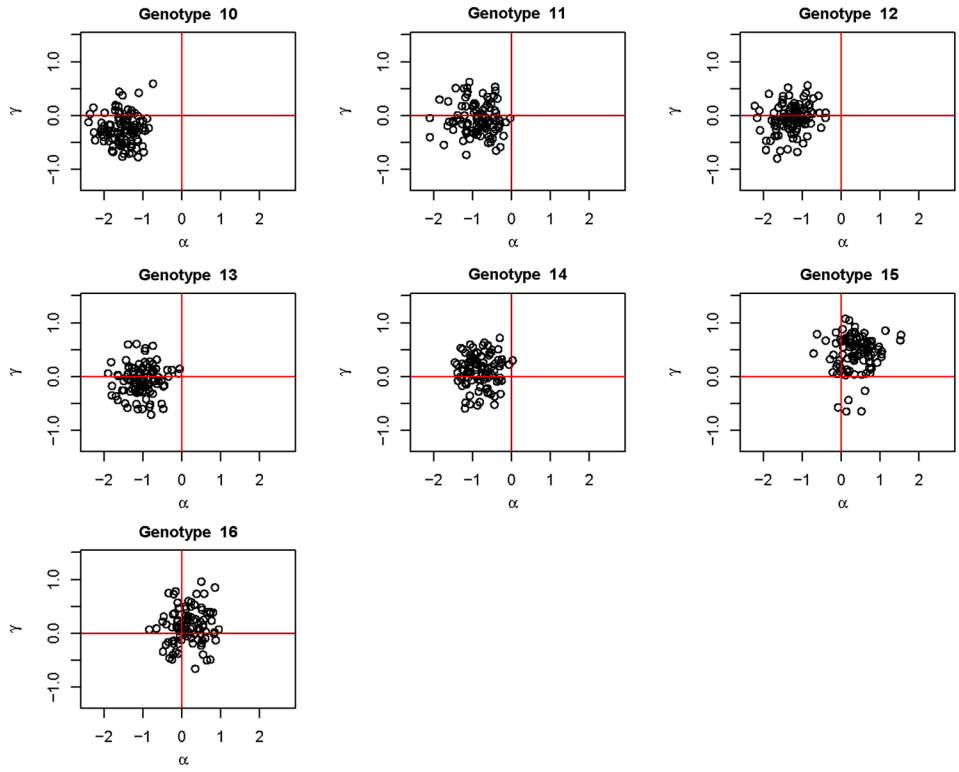

Figure 6. Genotype effect: for each genotype $i_{(i=10,\ldots,16)}$ representation of $(\alpha_i^s\sqrt{J}, \gamma_{i,1}^s\lambda_1^s)$, $s = 1, \ldots, 1000$.

rect view of the variability. The multiplication of the parameters by $\sqrt{J}$ and λ_1 ensures that their squared norms are equal to the sum of squares (SS) explained by the associated terms in the model ($\text{SS}_{\text{Genotype}} = J\sum_i \widehat{\alpha}_i^2$ and $\text{SS}_{\text{GE(1)}} = \lambda_1^2 = \lambda_1^2\sum_i \widehat{\gamma_{i,1}}^2$, respectively), which seems a fair scaling to relate them. The results for the second term of the interaction are given in the Appendix, Figures A.1 and A.2. Genotypes 1 to 6, the complete genotypes, have a positive main effect (positive values for α_i), whereas genotypes 8 to 14, the substituted genotypes, have a negative main effect. This is in agreement with the results obtained by Royo, Rodriguez, and Romagosa (1993), who concluded that complete genotypes tend to out-yield the substitute ones. Genotype 1 appears to be the most stable since its values for the interaction terms are the smallest in absolute value (around 0) for the first (see Figure 5) and second dimensions (see Figure A.1). The high values for the interaction terms on the second dimension (Figures A.1 and A.2) for many genotypes reinforce the fact that studying two dimensions may be appropriate to describe the genotype-environment interaction. However, the magnitude of the second term is really inferior compared to the first one, which can explain that the original paper (Royo, Rodriguez, and Romagosa 1993) only focused on one term. These graphical representations allow us to answer three of the questions raised at the beginning of this section. Concerning the global performance of the genotypes (Q1) figures show that genotype 1 has a better performance across the set of all environments. Genotype 1 is also one of the most stable genotypes, together with

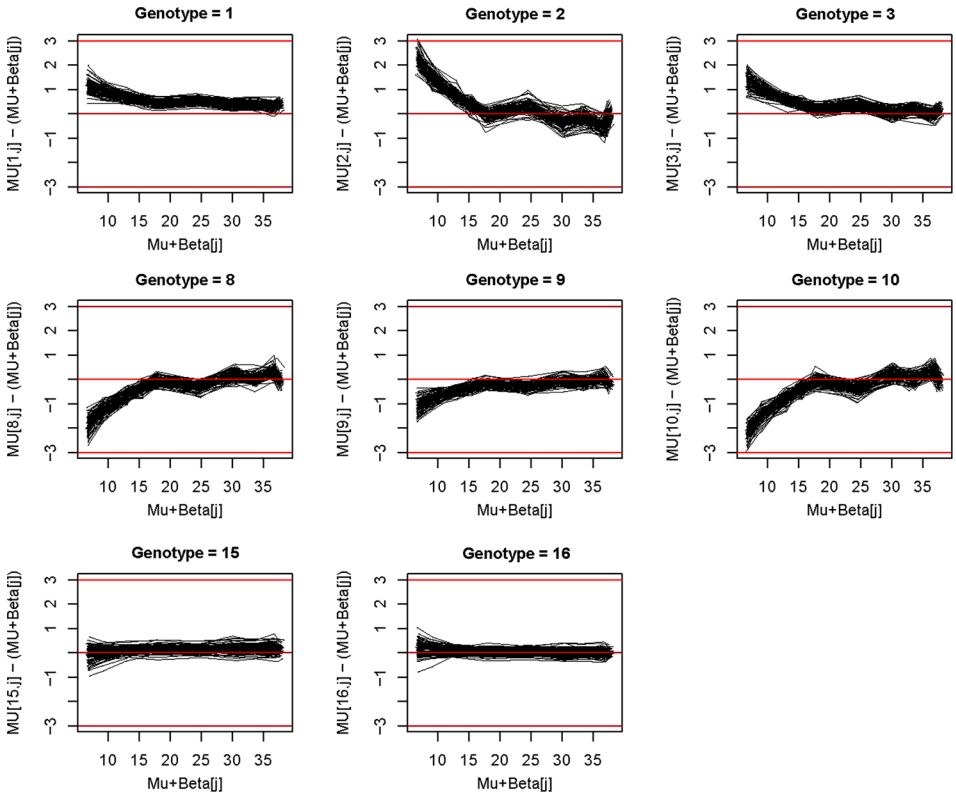

Figure 7. Posterior genotype profiles: for each genotype i ($i = 1, 2, 3, 8, 9, 10, 15, 16$), representation of $(\frac{1}{I}\sum_i \mu_{ij}^s; \mu_{ij} - \frac{1}{I}\sum_i \mu_{ij}^s)$, $s = 1, \ldots, 1000$.

genotypes 7, 15, and 16 (Q3), where it needs to be emphasized that genotypes 15 and 16 were check genotypes that typically are chosen for their stability. Finally, concerning the rank of the genotypes (Q4), it is not possible to decide between genotypes 1 and 3 for the first rank since the posterior distributions overlap.

Figure 7 focuses on the behavior of the genotypes across specific environments. For each genotype i, each point has as coordinates $(\frac{1}{I}\sum_i \mu_{ij}^s; \mu_{ij} - \frac{1}{I}\sum_i \mu_{ij}^s)$ which is equal to $(\mu^s + \beta_j^s; \alpha_i^s + \sum_{q=1}^{Q} \lambda_q^s \gamma_{iq}^s \delta_{js}^s)$, $s = 1, \ldots, 1000$. Only a subset of genotypes are represented in this figure but they are representative of the behavior of the others (genotypes 1 to 3 represent the first six, complete, genotypes, while genotypes 8 and 9 represent the eight substituted genotypes and genotypes 15 and 16 are checks). Each pane for a genotype i can be read as follow: the poorest environments (with small β_j) are on the left of the horizontal axis and high values on the vertical axis correspond to high values for the yield for the genotype i in a particular environment. There are strong differences between genotypes: according to expectation, the complete genotypes 1 to 6 tend to perform better in the poorest environments, whereas the substituted genotypes 8 to 14 show good behavior in the best environments, again according to expectation, see Royo, Rodriguez, and Romagosa (1993). Indeed, environments with small β_j are environments with acid soil

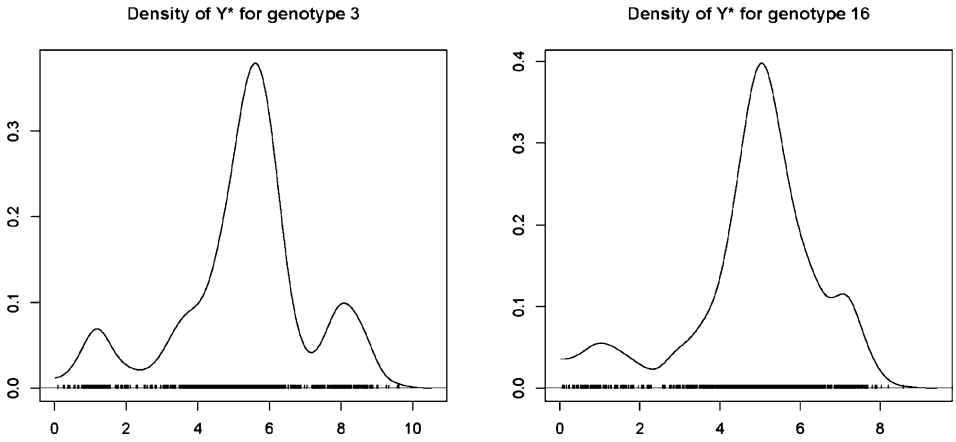

Figure 8. Performances of the genotypes 3 and 16 in a randomly selected environment. Representation of the posterior distributions of Y_i^* $(i = 3, 16)$.

and the authors noted that "the superiority of complete over substituted decreased when the soil pH increased." Genotypes 15 and 16 are very stable across all the environments. As mentioned, these genotypes were official checks with "very low interaction and yield similar or slightly superior to the grand mean." These graphical representations allow us to answer the question Q2 on the local performances, i.e., to determine which are the best genotypes for a given environment.

Finally, we consider the last question Q5, which is known as the risk assessment of a genotype (Eskridge and Mumm 1992; Piepho 1996). It corresponds to the risk that the yield of a genotype falls below a certain level in a randomly selected environment. The Bayesian approach produces a straightforward answer with respect to that issue. Thus, it can be viewed as a new solution to this important question. We simply use conditioning and define a new variable Y_i^* corresponding to the yield of a genotype i grown in a randomly selected environment. This new variable can easily be introduced into the model, and a posterior distribution for Y_i^* can be obtained (since a posterior distribution is available for the Y_{ij}, $j = 1, \ldots, 10$). Consequently, the risk is defined as the proportion of the posterior simulations of Y_i^* below the given level. The procedure is illustrated using a threshold equal to 4. Results are displayed Figure 8 for genotypes 3 and 16. It can be seen that genotypes 3 and 16 have a probability respectively estimated to 0.16 and 0.19 (by the posterior mean of the proportions) to get a yield lower than 4. The distributions are multimodal, which is expected since it takes into account poor and rich environments.

4. CONCLUSION

The Bayesian treatment of the AMMI model as proposed in this paper offers a new solution to tackle the problem of overparameterization by defining directly priors on all the parameters without taking into account the constraints and by applying an appropriate

postprocessing at the posterior level. This proposal can straightforwardly be extended to any overparameterized model and is directly tractable with standard MCMC algorithms. The method showed good results based on a small simulation and on a real dataset. In addition, we showed how the Bayesian approach can be used to answer many practical questions arising when analyzing genotype-environment data, such as general and specific performance and the risk assessment of a genotype. The code to perform the method, the simulations, and the real data analysis is available on the web page of the first authors.

In the plant breeding field, breeders have often a good idea about what the average of the yield should be, what the genetic and environmental variance is, and even about the magnitude of the interaction when some historical information is available. It would be interesting to carry out the Bayesian analysis with such informative priors to assess the impact of the insertion of knowledge on the results and to compare it with classical least squares or restricted maximum likelihood solutions. In our real data analysis, using vague priors implies that the results will be close to those of the least squares solution detailed in Royo, Rodriguez, and Romagosa (1993). However, the point estimates of the parameters are different since the means of the posterior distributions correspond to shrinkage versions of the least squares estimates (Cornelius and Crossa 1999; Moreno-Gonzalez, Crossa, and Cornelius 2003). Shrinkage estimation can be considered to be an advantage of the Bayesian approach since shrinkage estimators are known (Hastie, Tibshirani, and Friedman 2009; Hoff 2007) to be often better (i.e., closer to the true parameters) in terms of the mean squared error compared to classical estimators.

Another strong point of the Bayesian approach is that it can be straightforwardly used on incomplete datasets. Consequently, an interesting research perspective of our work is to assess our proposal in relation to unbalanced data. Many methods are available to deal with missing values in linear–bilinear models. A common approach to obtain point estimates for parameters consists in using alternating weighted least squares algorithms or iterative imputation algorithms (Gabriel and Zamir 1979; Kiers 1997; Gauch and Zobel 1990; Josse and Husson 2013). Estimating the variability of the parameters in a missing data framework is more difficult, and some authors (Adams et al. 2002; Josse and Husson 2011) suggested approaches based on bootstrap simulations. Using a Bayesian point of view will directly provide credible regions for the parameters through the posterior distributions, which is very appealing. In addition, credible regions will also be available for missing entries. We mention that Josse and Denis (2012) draw a parallel between the way missing values are handled and the overparameterization issue.

Finally, we want to make a comment regarding the nature, fixed or random, of the effects in AMMI models. This issue has induced many debates and discussions, and no general consensus seems to exist. Using a random version of the AMMI models (Piepho 1997; Smith, Cullis, and Thompson 2001) with the environment effect as random seems appropriate to answer the first question (Q1) about the global performances of the genotypes. However, a fixed effect model seems more suitable to answer the second question (Q2) about the local performances of the genotypes, even if a solution is available with a random effects model (Piepho 1998) using BLUP (Robinson 1991) estimates (known as shrinkage

estimates, Cornelius and Crossa 1999). Consequently, different models arise from the different questions and imply different results, which may be confusing. An advantage of the Bayesian strategy, also mentioned in Crossa et al. (2011), is that these choices are no longer necessary since there are no unknown parameters anymore, but only random variables. The issue can thus be tackled using conditioning. However, it is not so simple, and it is difficult to claim that the Bayesian approach solves the issue that exists with respect to taking terms fixed or random. Indeed, firstly, a Bayesian model can be seen as a random effects model with all the effects considering as random. However, the difference between fixed and random effects can be made using different priors such as uniform priors for the former case and Gaussian priors for the latter one. Secondly, one can argue that a Bayesian approach consists in putting priors on the parameters on which one want to do inference. Consequently, in a "random effect view" of the model, a prior distribution could be added on the variance parameters (such as s_β^2), which would also lead to different results.

APPENDIX: GENOTYPES EFFECT FOR THE SECOND INTERACTION TERM

See Figures A.1 and A.2.

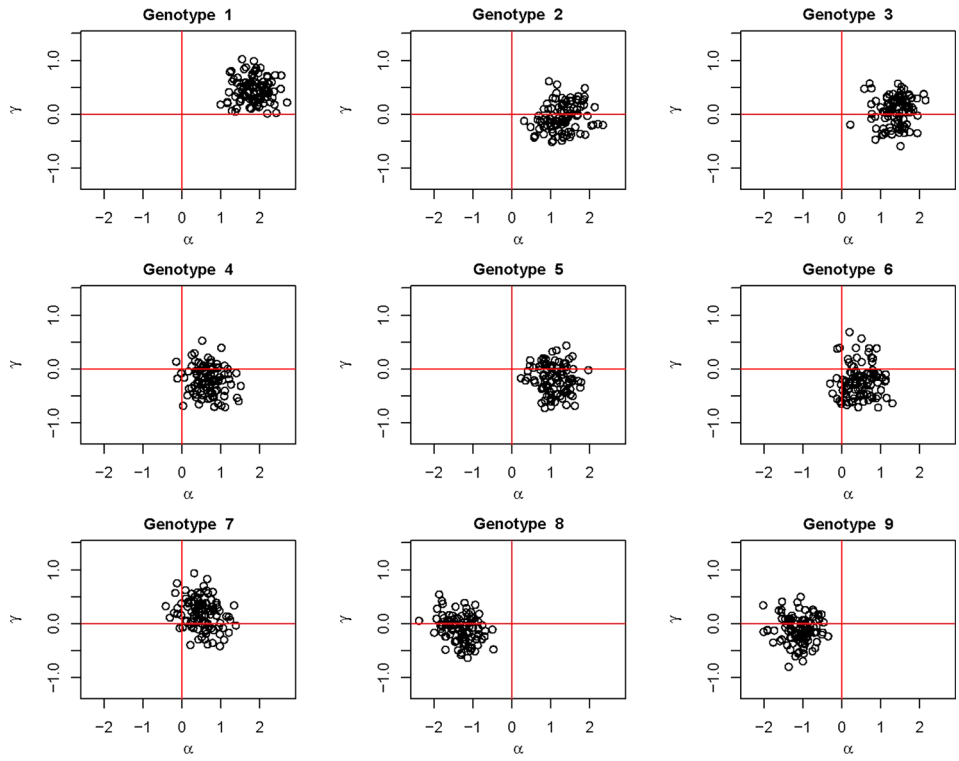

Figure A.1. Genotypes effect: for each genotype $i_{(i=1,\ldots,9)}$, representation of $(\alpha_i^s\sqrt{J}, \gamma_{i,2}^s\lambda_2^s)$, $s = 1, \ldots, 1000$.

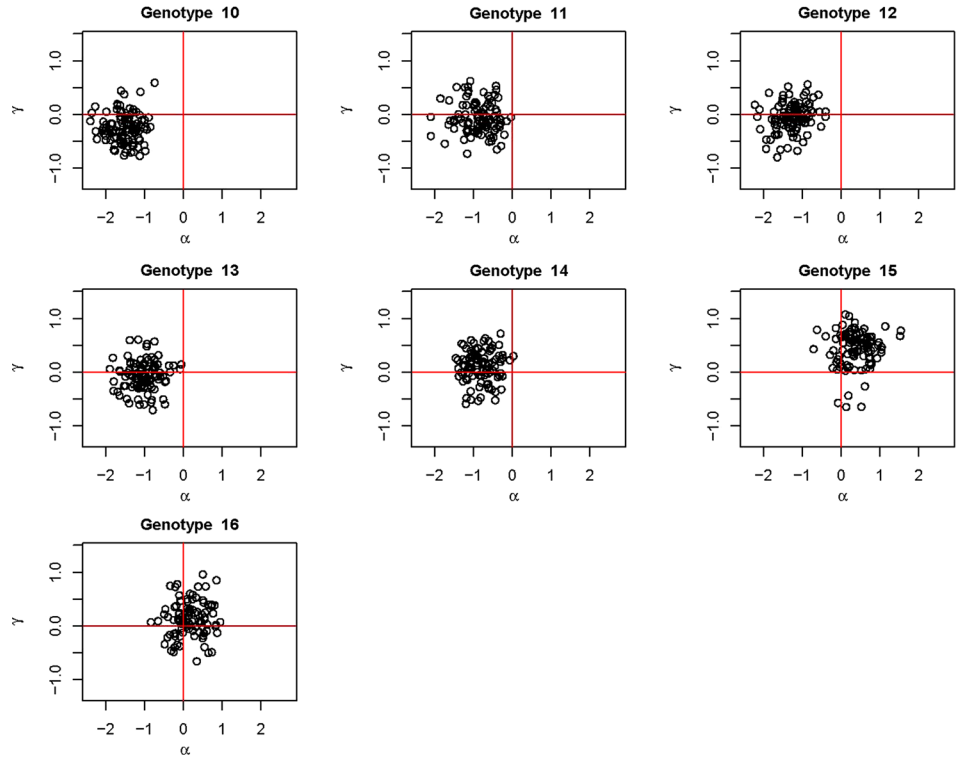

Figure A.2. Genotypes effect: for each genotype $i_{(i=10,\ldots,16)}$, representation of $(\alpha_i^s\sqrt{J}, \gamma_{i,2}^s\lambda_2^s)$, $s = 1,\ldots,1000$.

[Received October 2013. Accepted February 2014. Published Online March 2014.]

REFERENCES

Adams, E., Walczak, B., Vervaet, C., Risha, P. G., and Massart, D. L. (2002), "Principal Component Analysis of Dissolution Data with Missing Elements," *International Journal of Pharmaceutics*, 234, 169–178.

Chikuse, Y. (2003), *Statistics on Special Manifolds*, New York: Springer.

Cornelius, P., and Crossa, J. (1999), "Prediction Assessment of Shrinkage Estimators of Multiplicative Models for Multi-environment Cultivar Trials," *Crop Science*, 39, 998–1009.

Cornelius, P., Crossa, J., and Seyedsadr, M. (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, FL: CRC Press, pp. 199–234.

Crossa, J., Perez-Elizalde, S., Jarquin, D., Miguel Cotes, J., Viele, K., Liu, G., and Cornelius, P. (2011), "Bayesian Estimation of the Additive Main Effects and Multiplicative Interaction Model," *Crop Science*, 51, 1468–1469.

Denis, J. B., and Gower, J. C. (1994), "Asymptotic Covariances for the Parameters of Biadditive Models," *Utilitas Mathematica*, 46, 193–205.

——— (1996), "Asymptotic Confidence Regions for Biadditive Models: Interpreting Genotype-Environment Interactions," *Applied Statistics*, 45 (4), 479–493.

Dias, S., and Krzanowski, W. (2003), "Model Selection and Cross Validation in Additive Main Effect and Multiplicative Interaction Models," *Crop Science*, 43 (3), 865–873.

Edwards, J. W., and Jannink, J. L. (2006), "Bayesian Modeling of Heterogeneous Error and Genotype X Environment Interaction Variances," *Crop Science*, 46, 820–833.

Eskridge, K., and Mumm, R. (1992), "Choosing Plant Cultivars Based on the Probability of Outperforming a Check," *Theoretical and Applied Genetics*, 84, 494—500.

Gabriel, K. R., and Zamir, S. (1979), "Lower Rank Approximation of Matrices by Least Squares with Any Choice of Weights," *Technometrics*, 21 (4), 236–246.

Gauch, H. (1990), "Using Interaction to Improve Yield Estimates," in *Genotype by Environment Interaction*, ed. M. S. Kang, Boca Raton, FL: CRC Press, pp. 141–150.

Gauch, H., and Zobel, R. (1990), "Imputing Missing Yield Trial Data," *Theoretical and Applied Genetics*, 79, 753–761.

——— (1996), "AMMI Analysis of Yield Trials," in *Genotype by environment interaction*, eds. M. S. Kang, and H. G. Gauch, Boca Raton, FL: CRC Press, pp. 141–150.

Gelman, A. (2006), "Prior Distributions for Variance Parameters in Hierarchical Models," *Bayesian Analysis*, 3, 515–533.

Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning. Data Mining, Inference and Prediction. Springer Series in Statistics* (2nd ed.).

Hoff, P. D. (2007), "Model Averaging and Dimension Selection for the Singular Value Decomposition," *Journal of the American Statistical Association*, 102 (478), 674–685.

——— (2009), "Simulation of the Matrix Bingham–von Mises–Fisher Distribution, with Applications to Multivariate and Relational Data," *Journal of Computational and Graphical Statistics*, 18 (2), 438–456.

——— (2012), "rstiefel: Random Orthonormal Matrix Generation on the Stiefel Manifold," available at *http://CRAN.R-project.org/package=rstiefel*, R Package Version 0.9.

Jolliffe, I. T. (2002), *Principal Component Analysis*, New York: Springer.

Josse, J., and Denis, J. (2012), "Inferring Biadditive Models Within the Bayesian Paradigm," Tech. Rep., INRA, MIA.

Josse, J., and Husson, F. (2011), "Multiple Imputation in PCA," *Advances in Data Analysis and Classification*, 5 (3), 231–246.

——— (2012), "Selecting the Number of Components in PCA Using Cross-Validation Approximations," *Computational Statistics & Data Analysis*, 56 (6), 1869–1879.

——— (2013), "Handling Missing Values in Exploratory Multivariate Data Analysis Methods," *Journal de la Société Française de Statistique*, 153 (2), 79–99.

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*, 39 (1), 95–106.

Kiers, H. A. L. (1997), "Weighted Least Squares Fitting Using Ordinary Least Squares Algorithms," *Psychometrika*, 62 (2), 251–266.

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

Martyn, P. (2003), "Jags: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, March 20–22, Vienna, Austria.

Moreno-Gonzalez, J., Crossa, J., and Cornelius, P. L. (2003), "Additive Main Effects and Multiplicative Interaction Model: II. Theory on Shrinkage Factors for Predicting Cell Means," *Crop Science*, 43, 1976–1982.

Nelder, J. A. (1994), "The Statistics of Linear Models: Back to Basics," *Statistics and Computing*, 4 (4), 221–234.

Perez-Elizalde, S., Jarquin, D., and Crossa, J. (2011), "A General Bayesian Estimation Method of Linear-Bilinear Models Applied to Plant Breeding Trials with Genotype X Environment Interaction," *Journal of Agricultural, Biological, and Environmental Statistics*, 17 (1), 15–37.

Piepho, H. (1996), "A Simplified Procedure for Comparing the Stability of Cropping Systems," *Biometrics*, 52, 315–320.

——— (1997), "Analyzing Genotype-Environment Data by Mixed Models with Multiplicative Effects," *Biometrics*, 53, 761–766.

——— (1998), "Empirical Best Linear Unbiased Prediction in Cultivar Trials Using Factor Analytic Variance-Covariance Structures," *Theoretical and Applied Genetics*, 97, 195–201.

R Core Team (2013), "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, available at *http://www.R-project.org/*.

Robinson, G. K. (1991), "That BLUP Is a Good Thing: The Estimation of Random Effects," *Statistical Science*, 6 (1), 15–51.

Royo, C., Rodriguez, A., and Romagosa, I. (1993), "Differential Adaptation of Complete and Substitute Triticale," *Plant Breeding*, 111, 113–119.

Smidl, V., and Quinn, A. (2007), "On Bayesian Principal Component Analysis," *Computational Statistics & Data Analysis*, 51, 4101–4123.

Smith, A., Cullis, B., and Thompson, R. (2001), "Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend," *Biometrics*, 57 (4), 1138–1147.

Theobald, C. M., Talbot, M., and Nabugoomu, F. (2002), "A Bayesian Approach to Regional and Local-Area Prediction from Crop Variety Trials," *Journal of Agricultural, Biological, and Environmental Statistics*, 7 (3), 403–419.

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.