

Predictive ability of AMMI and factorial analytical models in the study of unbalanced multi-environment data

R.F. Romão^{1,2}, J.J. Nuvunga^{1,3}, C.P. Silva¹, L.A. Oliveira⁴,
C.T.E. Mendes¹ and M. Balestre¹

¹ Departamento de Agricultura, Universidade Federal de Lavras, Lavras, MG, Brasil

² Instituto Superior Politécnico de Gaza, Gaza, Mozambique

³ Eduardo Mondlane University, Chibuto College of Business and Entrepreneurship, Maputo, Mozambique

⁴ Faculdade de Ciências Exatas e Tecnologia, Universidade Federal da Grande Dourados, Dourados, MS, Brasil

Corresponding author: M. Balestre
E-mail: marciobalestre@des.ufla.br

Genet. Mol. Res. 18 (3): gmr18176

Received October 11, 2018

Accepted May 13, 2019

Published July 04, 2019

DOI <http://dx.doi.org/10.4238/gmr18176>

ABSTRACT. Efficient analysis of datasets from multi-environment trials (MET) is of paramount importance in plant breeding programs. Several methods have been proposed for this purpose, each of them having advantages and disadvantages, depending on the objectives of the study. We examined the robustness in the predictive power of models that have been widely used in the study of genotype-by-environment interaction such as AMMI (additive main-effects and multiplicative interaction) models via EM algorithm, Bayesian AMMI models with homogeneity (BAMMI), heterogeneity of variances (BAMMI-H) and the Analytical Factorial model (FA). To check the efficiency of these methods, genotype and genotype-by-environment interaction effects were simulated and further unbalances were included at levels of 10, 33 and 50% loss of genotypes in the environments. To evaluate the predictive power of the proposed models, the PRESS (prediction error sum square) statistics and the Cor (correlation between predicted and observed

value) were used. The genotype-environment interaction models had low sensitivity to missing data since all models showed correlations above 0.5 in all scenarios - even with high unbalance levels (50%). In general, there were differences in predictive accuracy among the models in different scenarios, with a slight advantage for the Bayesian models in the correlation among observed and predicted data ranging from 0.79 to 0.855 compared to 0.591 to 0.853 obtained from the competing models. Similar results were observed for the PRESS (4.988 to 8.027) in Bayesian models compared to competing models (5.411 to 23,361). Overall, there was slight advantage of the Bayesian models in unbalanced scenarios.

Key words: Genotype-by-environment interaction; Bayesian models; Analytical factorial model; Bayesian AMMI models

INTRODUCTION

Analyzing multi-environment datasets (METs) has been a central issue in plant breeding programs (Yan and Kang, 2003; Studnicki et al., 2017). The importance of quantifying the genotype-by-environment interaction (GEI) and identifying stable genotypes with wide adaptability is a routine task for the plant breeder. Many methods have been developed for this purpose, providing breeders with information that will support decision making and guide the selection of superior genotypes and/or cultivars.

In general, GEI quantification methods aim to reduce the complex relationships between genotypes and environments to a few parameters. Generally, it can be said that the presence of GEI makes the work of breeders difficult, impeding the recommendations of cultivars for a wide range of environments. However, if GEI patterns are established, this source of response variation can be applied to regionalized recommendations, leading to more efficient modeling (Heslot et al., 2013; van Eeuwijk et al., 2016; Studnicki et al., 2017).

Currently there is a great diversity of statistical methods for analyzing MET data, including parametric and non-parametric analysis, based on uni- and multivariate procedures (Crossa, 1990). Among the procedures that consider fixed effects, the linear-bilinear models stand out, such as additive main effects and multiplicative interaction (AMMI) and the main effects of genotype plus interaction effects (SREG or GGE) (Cornelius et al., 1996; Crossa et al., 2002, Crossa, 2012). These are the most popular among multiplicative models and have wide applicability for data analysis in plant breeding programs and multi-environment trials in general.

Despite the advantages of these procedures, they have limitations, such as the difficulty in modeling different variances between environments and dealing with unbalanced sets or missing data (Smith et al., 2001; So and Edwards, 2009; Hu and Spilke, 2011; Studnicki et al., 2017;). Although there are proposals to treat heteroscedastic data in fixed AMMI, as presented by Rodrigues et al. (2014), little work has been done in the context of AMMI-Bayesian models. To circumvent the difficulty of analyzing datasets with missing observations, versions of the AMMI model via the EM algorithm (AMMI-EM) were proposed to impute or predict non-observed interactions (Gauch and Zobel, 1990).

These limitations motivated the accomplishment of MET analyses in the context of mixed models, which theoretically have greater flexibility to deal with missing data in which is commonly observed in datasets obtained from MET experiments (Curti et al., 2014; Orellana et al., 2014). In this sense, we can highlight the versions of the linear-bilinear mixed AMMI and GGE models that were initially proposed by Piepho (1997, 1998), Smith et al. (2001), and Piepho and Mohring (2006). These mixed models versions of AMMI and GGE rely on the establishment of a factor analytic structure of the variance-covariance genetic matrix for environments and are referred to as factor analytic (FA) models.

Naturally, these versions have the flexibility to model the heterogeneity of variances between environments. Other parameters of interest such as spatial correlation within the environment and imbalance can be easily treated with these models. In addition, when considering random effects for genotypes, the kinship coefficients can be incorporated into the factor analytic structure for GEI or G+GEI modeling of the AMMI and SREG, respectively, obtaining more accurate estimates of the genotypic values (Crossa et al., 2006; Oakey et al., 2006; Burgueño et al., 2007, Crossa et al., 2011).

Another attractive, but still incipient, proposal is to treat multiplicative models within the Bayesian perspective. Viele and Srivivasan (2000) were the first to propose and implement the AMMI analysis using the Bayesian method, showing how to overcome the difficulties that arise in sampling the singular vectors, whose support of the complete conditional posterior is not trivial. Bayesian versions of the AMMI and GGE can be found (Crossa et al., 2011; Perez-Elizalde et al., 2011; Josse et al., 2014, Oliveira et al., 2015; Silva et al., 2015, Oliveira et al., 2016), which demonstrated the flexibility of the method to incorporate inference to the parameters and information to the analyzes. AMMI-Bayesian is able to handle complex data offering broad prospects for research and applicability in plant breeding programs.

Consequently, the analysis of MET data can be driven by several procedures with their advantages and disadvantages. However, when studying models with differences in variance-covariance structure, it is necessary to search for criteria for comparison between them, which vary according to the objective, including the investigation of predictive accuracy (Stefanova and Buirchell, 2010; Burgueno et al., 2011, Sa'diyah and Hadi, 2016). Evaluation of models with different variance-covariance structures (including FA) has been performed in several studies (Piepho, 1998; Smith et al., 2001; Kelly et al., 2007; Hu and Spilke, 2011; Nuvunga et al., 2015).

Given that unbalanced trials are commonly observed in plant breeding experiments and the associated importance of quantifying the GEI to identify stable genotypes in these scenarios, we evaluated the predictive ability of GEI using AMMI (Frequentist and Bayesian versions) and analytical factorial models in simulated unbalanced framework.

MATERIAL AND METHODS

Simulated data

A dataset with 20 genotypes (G1-G20) that were evaluated in seven environments (E1-E7) was simulated using a randomized complete block design with three replicates. Five genotypes had interactions simulated from the Gaussian distribution with large

variances (unstable genotypes) and positive marginal effects. Five more genotypes were marginally negative and simulated from Gaussian with large variances (unstable). The variances of response in each environment related to stable genotypes were 4 times larger than stable genotypes. The others 10 genotypes were sampling from a standard Gaussian distribution. Therefore, the stability and instability in this study were considered as a function of the size of the variability across the environments and the genotypes marginal effects as Gaussian realizations (Figure 1) (The simulation of the data was done in Software R).

The vector of the residuals ε was sampled by a multivariate normal $\varepsilon \sim N(\mathbf{0}, \mathbf{R})$, in which $\mathbf{0}$ is a null vector and \mathbf{R} is a diagonal matrix $\mathbf{R} = \text{diag}(\sigma_{e1}^2, \dots, \sigma_{ek}^2)$, where i ($i = 1, \dots, k$) denotes the environment. The genotype effects were sampled from a normal multivariate Gaussian distribution with zero mean and variance-covariance matrix $\mathbf{I}_g \sigma_g^2$. The environmental and repetition effects were fixed using a uniform distribution. The GEI response pattern is illustrated in Figure 1.

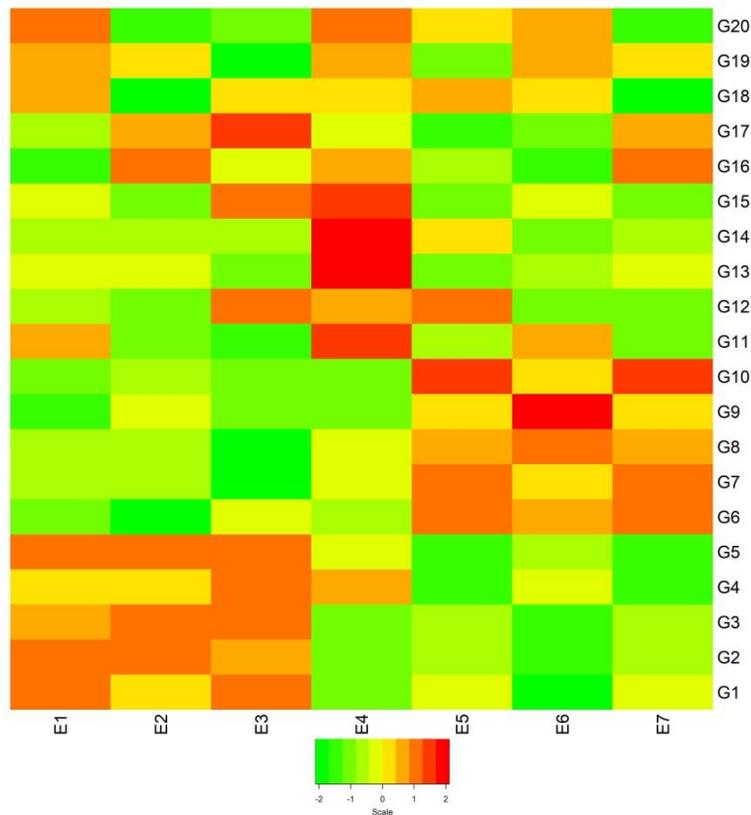


Figure 1. Heatmap of the genotype x environment interaction pattern of simulated data.

The clusters observed in the heatmap were obtained by the Euclidean distance between environments and genotypes. The green, yellow and red colors, in this order, indicate the magnitude of the simulated GEI, the more red the more positive the interaction.

MET analysis models

The competing models used for MET analysis with unbalanced data used were the following: classical AMMI fixed model for genotypes, environments and interactions (EM-AMMI-F), AMMI mixed model with fixed environments and random genotypes and GEI (EM-AMMI-M). In both AMMI models the missing cells in the two-way table were calculated via an expectation-maximization-EM algorithm as described by Gauch and Zobel (1990). In addition, we evaluated the AMMI Bayesian model with homogeneity of variance (Oliveira et al., 2015), and with variance heterogeneity - BAMMIB-H. The AMMI models were compared with the factor analytic structured mixed models described by Smith et al. (2001). Below we briefly describe each of the methods.

EM-AMMI Model

The AMMI-EM model proposed by Gauch and Zobel (1988) uses the average response of the genotype i in the environment j given by:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^t \lambda_k \gamma_{ik} \alpha_{jk} + \varepsilon_{ij}$$

where:

Y_{ij} is the mean response of the i -th genotype in the j -th environment;

μ is the constant, overall mean of observations;

g_i and e_j are the effects of the i -th genotype and j -th environment, respectively;

λ_k is k -th singular value;

γ_{ik} and α_{jk} are the singular vectors related to the elements of the i -th genotype and j -th environment, respectively;

ε_{ij} is the experimental error and $t = \min(r - 1, c - 1)$, where r and c indicate the number of rows and columns, respectively, of the interaction matrix.

The EM-AMMI adjustment was performed according to the algorithm proposed by Paderewski and Rodrigues (2014), which involves 5 steps:

Provide the initial values for the missing cells; in this case the mean of the row referring to genotype i .

The parameters of the AMMI models are estimated by singular value decomposition;

The interactions with the AMMI model are adjusted;

The missing cells are filled with the means adjusted with the AMMI model using the result of the reconstruction of the GEI matrix by the inner product \mathbf{ULV}^* where \mathbf{U} is

the matrix of singular vectors for genotypes, L is the matrix of singular values and V is the matrix of singular vectors referring to environments;

Based on the predicted values, steps ii to iv are repeated until the convergence.

Factor Analytic Model

The mixed model with factor analytic structure (FA) was constructed according to Smith et al. (2001). In MET analysis, the observed phenotypic values (\mathbf{y}), in different environments, from the perspective of unstructured mixed model, are modeled as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon}$$

In which \mathbf{X} and \mathbf{Z} are design matrices, $\boldsymbol{\beta}$ is the vector of fixed effects, \mathbf{g} is the vector with effects of genotypes considered random, with $\mathbf{g} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\varepsilon}$ is the vector of random errors, with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$ where $\boldsymbol{\Sigma}$ and \mathbf{R} are the matrices related to the genetic and residual covariances, respectively.

The mixed model, with factor analytic structure for $\boldsymbol{\Sigma}$, proposed by Smith et al. (2001) is denoted by:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\boldsymbol{\Gamma} \otimes \mathbf{I}_g)\mathbf{f} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon}$$

\mathbf{f} is the vector of factorial scores for genotypes in the factors;

$\boldsymbol{\delta}$ is the vector of specific errors representing the lack of adjustment of the factorial model;

$\boldsymbol{\Gamma}$ is the matrix of factor loadings in the variables represented by the environments;

$\boldsymbol{\Gamma} \otimes \mathbf{I}_g$ is the matrix of factor loads.

Assuming that the genetic covariance matrix $\boldsymbol{\Sigma}$ is modeled by the FA structure,

$$(\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})$$

where $\boldsymbol{\Psi} = \text{COV}(\boldsymbol{\delta}_i)$, solutions for fixed and random effects are obtained in a similar way to the known solutions of the conventional mixed models (Mayer, 2009). Further structural details on the FA approach can be found in Nuvunga et al. (2015). The models were adjusted using the ASReml-R package in R software (R Development Core Team, 2016). The variance and model parameters were estimated using restricted maximum likelihood (REML).

Bayesian AMMI Model with Homogeneous Variances

The Bayesian AMMI model with homogeneity of variance (BAMMI), as in Oliveira et al. (2015), is expressed by:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^l \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}$$

The vector \mathbf{y} has $n = lr$ phenotypic responses, being r the number of genotypes and $l = bc$ the number of genotypic repetitions, where b and c are the numbers of blocks

and environments, respectively. The vectors $\boldsymbol{\beta}$ and \mathbf{g} contain the main effects of blocks and environments, and main effects of genotypes, respectively.

The terms λ_k , \mathbf{a}_k and $\boldsymbol{\gamma}_k$ are the k -th singular values and singular vectors related to genotypes and environments, respectively, subjected to the constraints of the singular value decomposition, being $t = \min(r, c)$ the rank of interaction matrix between genotypes and environments $GE_{r \times c}$. The matrices \mathbf{X}_1 , \mathbf{X}_2 and \mathbf{Z} are the model design matrices. The vector $\boldsymbol{\varepsilon}_{n \times 1}$ contains the effects of random errors, with $\boldsymbol{\varepsilon} \sim N_n(0, \sigma_e^2 \mathbf{I})$.

The prior distributions for the model parameters are the same as those used by Oliveira et al. (2015). The likelihood function is given by:

$$L(\boldsymbol{\theta} | \mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{I} \sigma_e^2|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \boldsymbol{\theta})' (\mathbf{y} - \boldsymbol{\theta}) \right\}$$

being $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k) \mathbf{X}\boldsymbol{\gamma}_k$

Connecting the likelihood with the prior describes above via Bayes theorem gives the joint posterior distribution can be given by:

$$p(\boldsymbol{\Phi} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}, \sigma_e^2) p(\mathbf{g} | \boldsymbol{\mu}_g, \sigma_g^2) p(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \sigma_\beta^2) p(\sigma_g^2 | v_g, S_g^2) \times \\ \times p(\sigma_e^2 | v_e, S_e^2) \prod_{k=1}^t p(\lambda_k | \boldsymbol{\mu}_{\lambda_k}, \sigma_{\lambda_k}^2) p(\mathbf{a}_k) p(\boldsymbol{\gamma}_k)$$

where $\boldsymbol{\Phi} = (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{g}, \boldsymbol{\beta}, \sigma_g^2, \sigma_e^2)$, or:

$$p(\boldsymbol{\Phi} | \mathbf{y}) \propto (\sigma_e^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \boldsymbol{\theta})' (\mathbf{y} - \boldsymbol{\theta}) \right\} (\sigma_g^2)^{-\frac{n_g}{2}} \exp \left\{ -\frac{1}{2\sigma_g^2} \mathbf{g}' \mathbf{g} \right\} \times \\ \times \frac{1}{\sigma_g^2} \frac{1}{\sigma_e^2}$$

The complete conditional distributions for all model parameters are closed as shown below:

$$\boldsymbol{\beta} | \dots \sim N \left[(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{y} - \mathbf{Z}\mathbf{g} - \boldsymbol{\Theta}), (\mathbf{X}' \mathbf{X})^{-1} \sigma_e^2 \right],$$

where $\boldsymbol{\Theta} = \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k) \mathbf{X}\boldsymbol{\gamma}_k$.

$$\mathbf{g} | \dots \sim N \left[\left(\mathbf{Z}' \mathbf{Z} + \mathbf{I} \frac{\sigma_e^2}{\sigma_g^2} \right)^{-1} \mathbf{Z}' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Theta}), \left(\mathbf{Z}' \mathbf{Z} + \mathbf{I} \frac{\sigma_e^2}{\sigma_g^2} \right)^{-1} \sigma_e^2 \right]$$

$$\sigma_e^2 | \dots \sim \text{inv-escalada} - \chi^2 [n, (\mathbf{y} - \boldsymbol{\theta})' (\mathbf{y} - \boldsymbol{\theta})].$$

$$\sigma_g^2 | \dots \sim \text{inv-escalada} - \chi^2 [n_g, \mathbf{g}' \mathbf{g}].$$

$$\lambda_k | \dots \sim N^+ \left[(\boldsymbol{\phi}' \boldsymbol{\phi})^{-1} \boldsymbol{\phi}' \boldsymbol{\Delta}, (\boldsymbol{\phi}' \boldsymbol{\phi})^{-1} \sigma_e^2 \right]$$

Considering:

$$\Delta = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \sum_{k \neq k}^p \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}\boldsymbol{\gamma}_k \text{ and } \boldsymbol{\phi} = \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}\boldsymbol{\gamma}_k.$$

where: $\lambda_1 \geq \dots \geq \lambda_p \geq 0$.

For the singular vectors, the complete conditional posterior are proportional to a von Mises-Fisher distribution (VMF):

$\boldsymbol{\alpha}_k | \dots \sim \text{VMF}[k, \boldsymbol{\mu}_{\alpha_k}]$, with concentration parameter $k = \lambda_k / \sigma_e^2$ and directional mean $\boldsymbol{\mu}_{\alpha_k} = \boldsymbol{\Lambda}' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g})$ being $\boldsymbol{\Lambda} = \text{diag}(\mathbf{X}\boldsymbol{\gamma}_k) \mathbf{Z}$.

$\boldsymbol{\gamma}_k | \dots \sim \text{VMF}[k, \boldsymbol{\mu}_{\gamma_k}]$, with concentration parameter $k = \lambda_k / \sigma_e^2$ and directional mean $\boldsymbol{\mu}_{\gamma_k} = \boldsymbol{\Omega}' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g})$ being $\boldsymbol{\Omega} = \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}$.

Bayesian AMMI Model with Heterogeneous Variances (BAMMI-H)

The Bayesian AMMI model with variance heterogeneity (BAMMI-H) is obtained from (1) but considering different variances $\sigma_{e_r}^2$ between environments, where $r = \{1, 2, \dots, E\}$ indicates the environmental index. Thus, the matrix $\mathbf{I}_n \sigma_e^2$ in model (1) is replaced by \mathbf{V} , which is a block diagonal matrix with dimension $n \times n$ composed of the environmental variances and n_r corresponds to the number of observations in the

environment r (here we introduce the notation $n = \sum_{r=1}^E n_r$). In this way, the conditional

distribution of \mathbf{y} is a multivariate normal: $\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}, \lambda_k, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \sigma_{e_r}^2 \sim N(\boldsymbol{\theta}, \mathbf{V})$, com

$$\boldsymbol{\theta} = \mathbf{X}_1 \boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^p \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k.$$

The priors for the parameters are the same of the BAMMI model, only emphasizing that in BAMMI-H, it is considered different variances for the environments requesting more than one prior for this parameters. The prior densities for the environmental variances are individually an inverse-scaled chi-square distribution. It was considered $\nu_{e_r} = S_{e_r}^2 = 0$, which are respectively the degree of freedom and scale parameter; therefore the prior distributions converge to $\sigma_{e_r}^2 | \nu_{e_r}, S_{e_r}^2 \sim \frac{1}{\sigma_{e_r}^2}$ (Jeffrey's prior).

The complete conditional posterior distributions for the BAMMI-H model are obtained from the joint posterior distribution considering the matrix \mathbf{V} , as defined above. The likelihood distribution is given by:

$$L(\boldsymbol{\phi} | \mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{V}|^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\theta})' \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\theta}) \right\}$$

where

$$\boldsymbol{\theta} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^p \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k)\mathbf{X}_2\boldsymbol{\gamma}_k \text{ and } \boldsymbol{\varphi} = \{\boldsymbol{\beta}, \mathbf{g}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}\}.$$

In turn, the joint posterior distribution is expressed by:

$$P(\boldsymbol{\Phi} | \mathbf{y}) \propto L(\boldsymbol{\varphi} | \mathbf{y})P(\boldsymbol{\beta} | \mu_{\boldsymbol{\beta}}, \sigma_{\boldsymbol{\beta}}^2)P(\mathbf{g} | \mu_{\mathbf{g}}, \sigma_{\mathbf{g}}^2)P(\sigma_{\mathbf{g}}^2 | v_{\mathbf{g}}, S_{\mathbf{g}}^2) \\ \left\{ \prod_{k=1}^p P(\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2)P(\mathbf{a}_k)P(\boldsymbol{\gamma}_k) \right\} \prod_{r=1}^E P(\sigma_{e_r}^2 | v_{e_r}, S_{e_r}^2)$$

in which

$$\boldsymbol{\Phi} = (\boldsymbol{\beta}, \mathbf{g}, \lambda_k, \mathbf{a}_k, \boldsymbol{\gamma}_k, \sigma_{\mathbf{g}}^2, \sigma_{e_r}^2).$$

From algebraic manipulations, it is possible to demonstrate that the conditional posterior distributions for the model parameters as follows.

$$\boldsymbol{\beta} | \dots \sim N \left[(\mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{V}^{-1} \mathbf{A}_1; (\mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_1)^{-1} \right]; \text{ being}$$

$$\mathbf{A}_1 = \mathbf{y} - \mathbf{Z}\mathbf{g} - \sum \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k)\mathbf{X}_2\boldsymbol{\gamma}_k.$$

$$\mathbf{g} | \dots \sim N \left[\left(\mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z} + \frac{1}{\sigma_{\mathbf{g}}^2} I_{\mathbf{g}} \right)^{-1} \mathbf{Z}' \mathbf{V}^{-1} \mathbf{A}_2; \left(\mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z} + \frac{1}{\sigma_{\mathbf{g}}^2} I_{\mathbf{g}} \right)^{-1} \right]$$

where

$$\mathbf{A}_2 = \mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \sum \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k)\mathbf{X}_2\boldsymbol{\gamma}_k.$$

$$\sigma_{\mathbf{g}}^2 | \dots \sim \text{Inv} - \text{Esc} - \chi^2 \left[n_{\mathbf{g}}, \frac{\mathbf{g}' I_{\mathbf{g}} \mathbf{g}}{n_{\mathbf{g}}} \right]$$

$$\lambda_k | \dots \sim N^+ \left[\left(\lambda_k - (\boldsymbol{\phi}_k' \mathbf{V}^{-1} \boldsymbol{\phi}_k)^{-1} \boldsymbol{\phi}_k' \mathbf{V}^{-1} \mathbf{A}_{4k} \right); (\boldsymbol{\phi}_k' \mathbf{V}^{-1} \boldsymbol{\phi}_k)^{-1} \right]$$

where

$$\mathbf{A}_{4k} = \mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \sum_{k' \neq k}^p \lambda_{k'} \text{diag}(\mathbf{Z}\mathbf{a}_{k'})\mathbf{X}_2\boldsymbol{\gamma}_{k'}, \boldsymbol{\phi}_k = \text{diag}(\mathbf{Z}\mathbf{a}_k)\mathbf{X}_2\boldsymbol{\gamma}_k.$$

$$\mathbf{a}_k | \dots \sim N \left[\left((\boldsymbol{\Delta}_{1k}' \mathbf{V}^{-1} \boldsymbol{\Delta}_{1k})^{-1} \boldsymbol{\Delta}_{1k}' \mathbf{V}^{-1} \mathbf{A}_{4k} \right); (\boldsymbol{\Delta}_{1k}' \mathbf{V}^{-1} \boldsymbol{\Delta}_{1k})^{-1} \right]$$

in which

$$\boldsymbol{\Delta}_{1k} = \lambda_k \text{diag}(\mathbf{X}_2\boldsymbol{\gamma}_k)\mathbf{Z}.$$

When the assumption of homogeneity of variances is assumed, we observe that the posterior conditional distributions for the singular vectors are proportional to a von Mises Fisher. However, assuming different variances in each environment, the von Mises Fisher distributions are changed for Gaussian distributions. Thus, the singular vectors are not sampled in hyperspheres but in hyperellipses in the corrected subspace and placed in the correct subspace by means of orthogonal linear transformation.

The complete conditional posterior for the genotype singular vector in the correct subspace is given by:

$$\tilde{\boldsymbol{\alpha}}_k | \dots \sim N(\mathbf{m}_{\alpha_k}, \mathbf{I}\sigma_{\alpha_k}^2)$$

where \mathbf{H}_k is a matrix of linear transformation, being

$$\mathbf{m}_{\alpha_k} = \Delta_{\alpha_k} \left[\Delta_{\alpha_k} \Delta_{\alpha_k} \right]^{\left(\frac{-1}{2}\right)}, \Delta_{\alpha_k} = \left(\mathbf{H}_k \left(\Delta_{1k} \mathbf{V}^{-1} \Delta_{1k} \right)^{-1} \Delta_{1k} \mathbf{V}^{-1} \mathbf{A}_{4k} \right), \text{ and}$$

$$\sigma_{\alpha_k}^2 = \left[\left(\left[\Delta_{\alpha_k} \Delta_{\alpha_k} \right]^{\left(\frac{1}{2}\right)} \right)' \mathbf{H}_k \left(\Delta_{1k} \mathbf{V}^{-1} \Delta_{1k} \right) \mathbf{H}_k \bar{m}_{\alpha} \left[\Delta_{\alpha_k} \Delta_{\alpha_k} \right]^{\left(\frac{-1}{2}\right)} \right]^{-1}$$

$$\boldsymbol{\gamma}_k | \dots \sim N \left[\left(\left(\Delta_{2k} \mathbf{V}^{-1} \Delta_{2k} \right)^{-1} \Delta_{2k} \mathbf{V}^{-1} \mathbf{A}_{4k} \right); \left(\Delta_{2k} \mathbf{V}^{-1} \Delta_{2k} \right)^{-1} \right]$$

where

$$\Delta_{2k} = \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}_2.$$

The conditional posterior for the environment singular vector $\tilde{\boldsymbol{\gamma}}_k$ in the correct subspace is given by:

$$\tilde{\boldsymbol{\gamma}}_k | \dots \sim N(\mathbf{m}_{\gamma_k}, \mathbf{I}\sigma_{\gamma_k}^2)$$

where \mathbf{W}_k is the orthogonal linear transformation matrix,

$$\mathbf{m}_{\gamma_k} = \Delta_{\gamma_k} \left[\Delta_{\gamma_k} \Delta_{\gamma_k} \right]^{\left(\frac{-1}{2}\right)},$$

$$\Delta_{\gamma_k} = \left(\mathbf{W}_k \left(\Delta_{2k} \mathbf{V}^{-1} \Delta_{2k} \right)^{-1} \Delta_{2k} \mathbf{V}^{-1} \mathbf{A}_{4k} \right), \text{ and}$$

$$\sigma_{\gamma_k}^2 = \left[\left(\left[\Delta_{\gamma_k} \Delta_{\gamma_k} \right]^{\left(\frac{1}{2}\right)} \right)' \mathbf{W}_k \left(\Delta_{2k} \mathbf{V}^{-1} \Delta_{2k} \right) \mathbf{W}_k \bar{m}_{\gamma} \left[\Delta_{\gamma_k} \Delta_{\gamma_k} \right]^{\left(\frac{-1}{2}\right)} \right]^{-1}$$

Further details on the a posteriori complete conditionals for the singular vectors as well as the sampling process can be found in Silva et al. (2018) and supplementary material.

For the residual variances, the conditional posterior distributions are given as following:

$$\sigma_{e_r}^2 | \dots \sim \text{Inv} - \text{Esc} - \chi^2 \left[n_r, \frac{(\mathbf{y}_{e_r} - \boldsymbol{\theta}_{e_r})' (\mathbf{y}_{e_r} - \boldsymbol{\theta}_{e_r})}{n_r} \right]$$

In which the degree of freedom and scale parameter of the scaled chi-square are, respectively,

$$v_{e_r} = n_r \text{ and}$$

$$S_{e_r}^2 = \frac{(\mathbf{y}_{e_r} - \boldsymbol{\theta}_{e_r})' (\mathbf{y}_{e_r} - \boldsymbol{\theta}_{e_r})}{n_r}.$$

The common mean variance between the environments may be estimated by the geometric mean of the residual variances between environments $\hat{\sigma}_e^2 = \sqrt[E]{\prod_{r=1}^E \sigma_{e_r}^2}$, in which E is the number of environments considered in the simulation.

Sampling and estimation process

All conditionals were obtained in a closed form allowing direct sampling using the Gibbs sampler during the MCMC (Monte Carlo via Markov Chains). The convergence of the chains produced was monitored by methods of Raftery and Lewis (1992) and by Heidelberger and Welch (1983) criterion, implemented in the Bayesian Output Analysis (BOA) package.

MCMC chains with 88,000 iterations were simulated for the AMMI-Bayesian model parameters. In addition, the first 8000 sampling observations were burned (burn-in) and samples were taken at intervals of four observations (thinning).

All inference process was performed using R statistical software (R Development Core Team, 2016).

Comparison between models

In order to check the efficiency of these methods, random imbalances were performed in the genotypes-by-environment interaction matrix, which corresponds to the total loss of the genotype in the specific environment. Further, the analyses on unbalanced data were performed and the GEI predictions were carried out. In the two-way table, simulated imbalances on the cells were performed using the levels of 10%, 33% and 50% of loss.

The predictive ability of the models was measured using prediction error sum square (PRESS) and the correlation between the predicted and observed (Cor) GEI using cross-validation procedures. The higher the correlation and lower the PRESS statistics the more robust are genotype to unbalance. Here robustness is inversely related to the sensibility to missing data. All analyses were done using the R software.

RESULTS

Table 1 presents the simulated values for the residual variances of each environment, as well as the values estimated by each candidate model. It is possible to verify that the EM-AMMI models (EM-AMMI-F or EM-AMMI-M) presented greater residual variance when considering the simulated mean value. The models FA2 and FA7, on the other hand, presented variations in the precision of the estimates of the residual variance by environments; however, the mean residual variance estimated by the FA2 model was very close to the simulated variance. Among the Bayesian models, BAMMI-H presented better results than the model that considers homogeneity of variance. In general, the Bayesian models had a higher mean residual variance when compared with the factor analytic structure models.

Table 1. Simulated values for the residual variances and estimates obtained considering the different MET analysis models and geometric means of the residual variances.

Environment	Real values	Models				
		BAMMI	BAMMI-H	EM-AMMI [*]	FA2	FA7
1	0.2188	3.9643	0.4111	5.56	0.201	0.202
2	0.985	3.9643	1.8807	5.56	1.721	1.722
3	2.036	3.9643	1.7070	5.56	1.552	1.555
4	5.012	3.9643	4.3815	5.56	4.416	4.212
5	7.509	3.9643	6.7934	5.56	7.016	6.678
6	9.641	3.9643	9.1451	5.56	9.405	9.289
7	16.55	3.9643	13.0065	5.56	13.962	14.038
Mean [#]	3.081	3.9643	3.3437	5.56	2.999	2.959
logLik					-573.9	-607.9

^{*}The estimates of EM-AMMI-F and EM-AMMI-M were the same; [#]geometric mean as presented in methods.

Although the FA models were able to recover the residual variance through the environments, the GEI predictive ability related to models was not superior to simpler models such as AMMI-EM. Figure 2 illustrates the performance of each model at the 10% level of losses of genotypes at the environments (10-fold). In general, the models were equivalent in the prediction of data not observed. However, there was a small predictive disadvantage in FA7 model followed by EM-AMMI-F. The BAMMI and FA2 models presented practically the same predictive power whereas the Bayesian AMMI model with heterogeneity (BAMMI-H) together with the EM-AMMI-M imputation model presented a relatively superior performance.

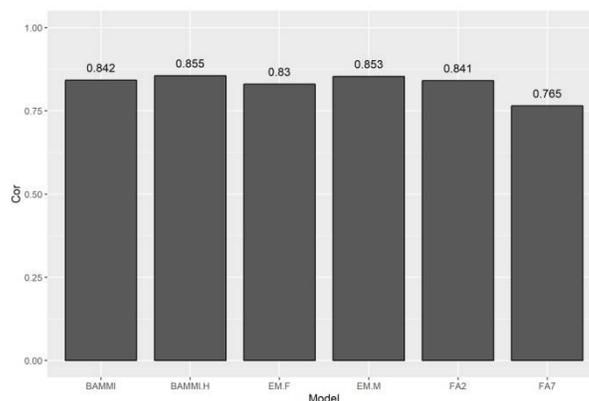


Figure 2. Mean correlation bar chart from cross validation considering 10% imbalance of genotype loss in the environments using the model FA, EM-AMMI-F, EM-AMMI-M, BAMMI and BAMMI-H.

Figure 3 shows the mean performances of the models with 33% loss in the genotypes (3-fold). It is possible to verify that in this scenario, the Bayesian models (BAMMI-H) and (BAMMI) presented superior performance to the FA models. Again, the complete FA model FA7 presented lower predictive power. It is important to emphasize that the EM methods were superior to the FA adjustments in the prediction of missing data, even violating the assumption of homogeneity. In addition, a slight superiority of BAMMI over BAMMI-H was found. In general, the predictions of the models can be considered good, since the correlation was superior to 0.7.

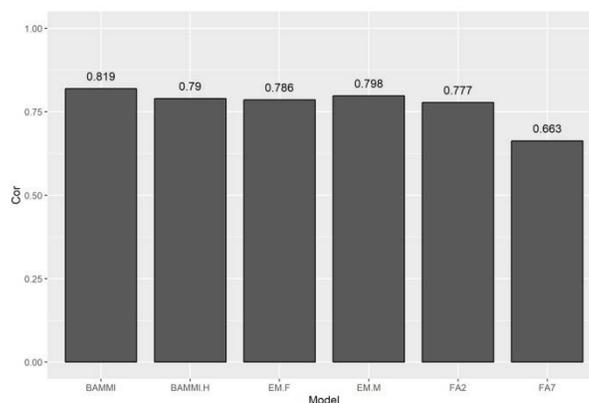


Figure 3. Mean correlation bar chart from the cross validation considering 33% imbalance of genotype loss in the environments using the models FA, EM-AMMI-F, EM-AMMI-M, BAMMI and BAMMI-H.

Comparing the mean performance of the models (Figure 4), at 50% losses (2-fold), it is verified that the EM-AMMI models, unlike that observed in 3-fold, presented the worst performance among all the proposed approaches. On the other hand, as already observed in lower levels of missing data, the Bayesian models presented better predictive ability in relation to the competing models. The FA7 model showed no convergence for the scenario with 50% of hybrids lost in the environments. The FA2 model presented convergence overcoming the AMMI-EM models. In this scenario, the superiority of Bayesian AMMI models in predicting missing data became more evident. The BAMMI models were 6 points higher in accuracy than the FA2 models and 14 points in relation to the EM-AMMI-M model and 20 points in relation to the EM-AMMI-F model. It is also worth to highlight the similar performances of the BAMMI and BAMMI-H models, evidencing the robustness of the AMMI analysis even assuming homogeneity of variances. Moreover, under high levels of imbalance, the EM-AMMI-F model presents a low predictive power and the EM-AMMI-M presents higher robustness.

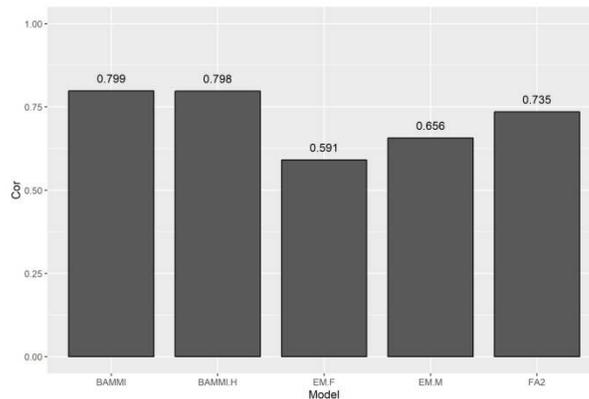


Figure 4. Mean correlation bar chart from the cross validation considering 50% imbalance of genotype loss in the environments using the model FA, EM-AMMI-F, EM-AMMI-M, BAMMI and BAMMI-H.

Figures 5 and 6 show the PRESS results, at 10% and 33% losses, respectively. Again, the results indicate that the Bayesian models present predictions closer to the real values observed in relation to the competing models.

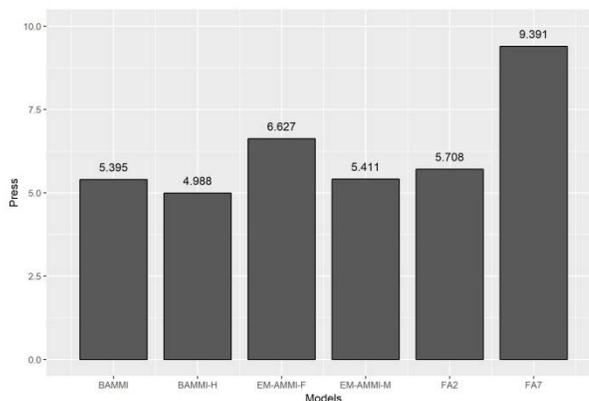


Figure 5. Mean PRESS bar chart from the cross validation considering 10% imbalance of genotype loss in the environments using the model FA, EM-AMMI-F, EM-AMMI-M, BAMMI and BAMMI-H.

It should be noted that although the AMMI models under Bayesian perspective have similar predictive power, the PRESS values between the two models were different in two imbalance scenarios (10 and 33%). Regardless of the levels of losses, the FA7 model again presented worse performance according to the PRESS criterion, in comparison with the other models. In addition, the fixed model EM-AMMI-F presented lower predictive power than its version using mixed models.

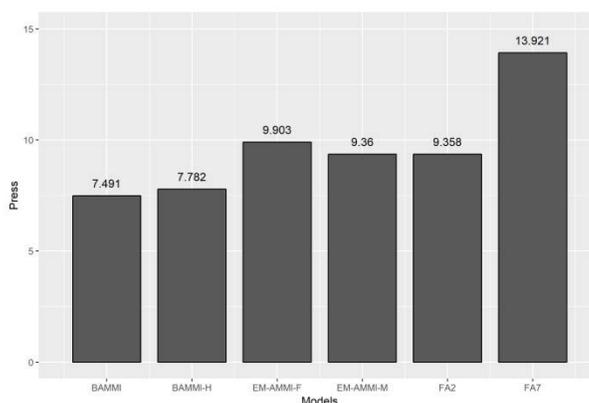


Figure 6. Mean PRESS bar chart from the cross validation considering 33% imbalance of genotype loss in the environments using the model FA, EM-AMMI-F, EM-AMMI-M, BAMMI and BAMMI-H.

At 50% loss level (Figure 7), the Bayesian models presented again better performance, in which the BAMMI-H was slightly superior in this level of imbalance, followed by FA2. The EM-AMMI-F model presented the worst performance among competing models. As already highlighted, in this scenario it was not possible to adjust the FA7 model, due to convergence problems. FA models presenting higher dimension than FA2 also did not converge and were not also explored for this parameter.

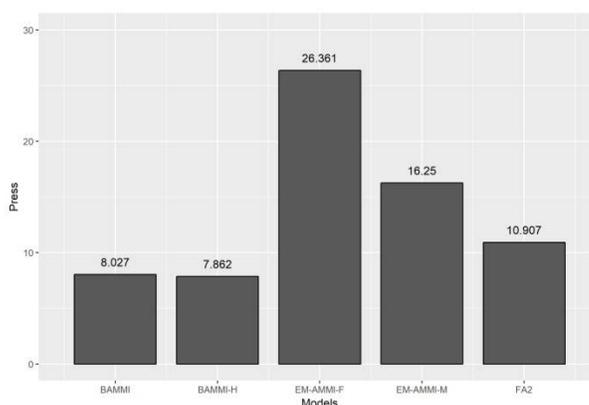


Figure 7. Mean PRESS bar chart from the cross validation considering 50% imbalance of genotype loss in the environments using the model FA, EM-AMMI-F, EM-AMMI-M, BAMMI and BAMMI-H.

A better comparison between the models can be obtained by looking at the efficiency gain of the candidates models in relation to FA7-based model, which presented the worst performance in all scenarios. For 50% losses, in which FA7 did not converged it was used the

EM-AMMI-F. In relation to the correlation (Figure 8), at 10% and 33%, we observed a greater efficiency gain of the BAMMI-H and BAMMI models, respectively. For the scenario with the greatest imbalance (55%), the Bayesian models presented a mean efficiency gain above all others (35%), followed by FA2 (24%).

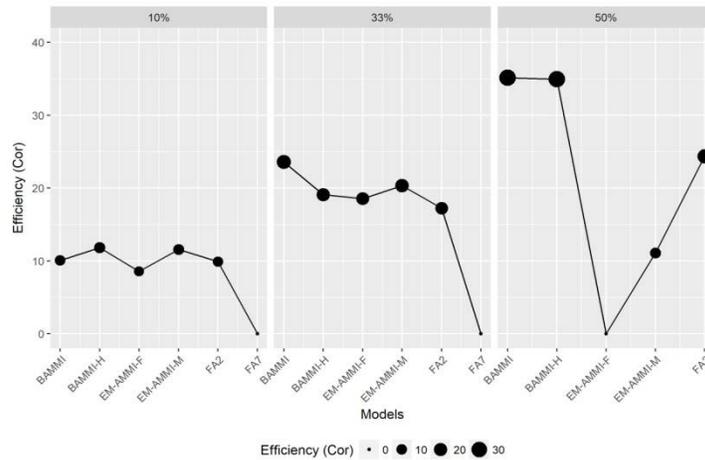


Figure 8. Gain of efficiency of the models in relation to the correlation between predicted and observed values.

In the PRESS criterion (Figure 9), there is a greater efficiency gain for the AMM-Bayesian models in all scenarios. This superiority was evident at levels with greater losses. At 33%, gains of 23% was observed for BAMMI and 19% for BAMMI-H, whereas EM-AMMI-M and FA2 obtained similar performances of about 20% and 17%, respectively. At 50%, the AMMI-Bayesian models obtained on average 35% efficiency gain, whereas FA2 presented a gain of 24%. Again, the efficiency was calculated based on the gain of the F7 model (at 10% and 33%) and the EM-AMMI-F model, at the 50% level of imbalance.

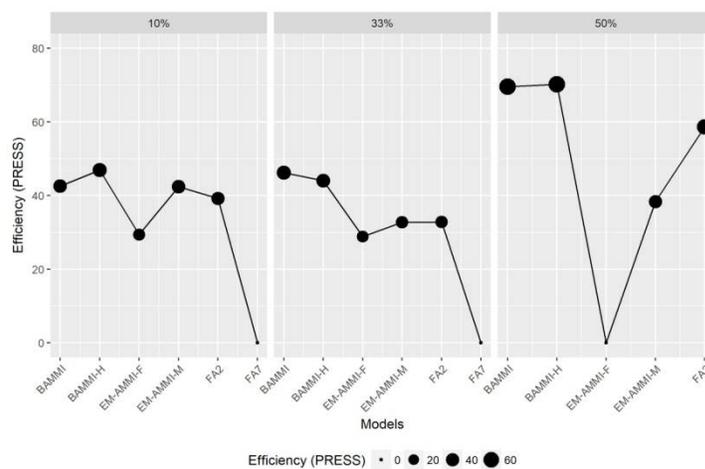


Figure 9. Gain of efficiency of the models in relation to PRESS.

DISCUSSION

In recent years, several models have been employed in the analysis of experiments in multiples environments. Among these techniques, model with FA structure have been widely used given to user friendly programs such as ASReml (Smith et al., 2015, Studnicki et al., 2017). One of the main appeals to use simplified GEI structures is parsimony and parameter estimability since the complexity of unstructured matrices requires highly parameterized models. Another important issue in favor of models with simplified FA structure like is that the available data are not always sufficient to make a good estimates of the dispersion parameters; a very common scenario in MET data with a high level of imbalance.

The simplification of the FA structure is not only important (the parsimonious principle), but it is occasionally necessary when high dimensional models present low rates of convergence. In this study, the model with two factorial loadings- presented convergence in all the proposed unbalance scenarios (Figures 2 to 7). While the complete FA model - which recovers the UN matrix with a low specific variance - did not present problems in the balanced analysis or low level of losses of genotypes, it has presented not convergence for the parameters in any 2-fold scenarios (Figures 4 and 7).

In the modeling of FA models, a single-stage approach is used where all the effects of the models (fixed and/or random) are combined in a single analysis (Smith et al., 2001; Mohring and Piepho, 2009). As observed during the cross validation process, high levels of imbalance can lead to computational problems when the model is adjusted for a large number of cultivars and environments or when the data set is highly unbalanced as observed at the level of 50% imbalance.

As reported by Smith et al. (2015) and Beeck et al. (2010), FA models are sensitive to the initial guesses and may present great problems of convergence. The FA3 model, under 50% imbalance in the environments, also did not present convergence, therefore, it was not possible to proceed in the adjustment of higher order models. Computational difficulties may have been caused by problems with the model algorithm (Thompson et al., 2003). Experience indicates that convergence is usually achieved through a sensitive sequential adjustment of models, i.e. starting from the smaller to higher model order.

In relation to FA models presenting high dimension, it was not observed parameter convergence in FA models presenting structures above of three loadings when the unbalance level was 50%. If so, in this scenario only FA(2) was used and presented good predictive ability in all scenarios (Figures 2 to 7). In this study, we sought to compare simple alternatives for MET analysis under imbalance, such as classical EM-AMMI methods, with more complex models based in Bayesian and mixed model theory. Our results indicate that this method can handle scenarios of small GEI imbalance and does not require high computational demand and complex parameter estimation. This is because the EM iterative process occurs in the GEI matrix converging the GEI imputation in a few steps (Paderewski, Rodrigues, 2014). Another important issue observed in the AMMI-EM is that considering genotypes and GEI matrix as random effects may increase the predictions index. Similar results were obtained by Balestre et al. (2010) and Smith et al (2001).

Thus, our results indicate that under low level of imbalance, it is possible to use the AMMI method without having to use sophisticated statistical methods. The AMMI also could be used in genomic selection (GS) context given its ability to predict GEI interaction.

For instance, Heslot et al. (2013) observed that when environmental covariates are included in the GS models was possible to predict genotypes in untested environments. Also, Sa'diyah and Hadi (2016) using rice data observed the AMMI accuracy to predict genotypic values in GEI trials..

It is worth mentioning that, at the 10% level of imbalance, the EM-AMMI models were superior to the FA7 model (Figure 2) and with prediction similar to the FA2 model. At 33%, again the EM-AMMI models exceeded the FA model, with respect to predictive accuracy, and had similar performances to BAMMI-H. However, under a more severe level of imbalance (50% losses), the AMMI-EM models were overcome by the others. The good performance of AMMI-EM models, at 10% and 33% of losses, may have been influenced by the efficiency of the algorithm. According to Paderewski (2014), this algorithm (EM) is very good at imputation of missing data, which may have contributed to the quality of the fit.

According to the PRESS criterion, which indicates the error of prediction (Dias and Krazanoski, 2003; Kelly et al., 2007,), the BAMMI-H model generally obtained the lowest values, except for 33% of missing where it was overcome by BAMMI. In general, the Bayesian AMMI presented lower predictive error when compared with the others candidate models. It was also observed that the models assuming random effects were always better than those with fixed effect as observed by Burgueno et al., (2008).

The question of whether or not to consider variance heterogeneity in the Bayesian AMMI model may be very relevant to model parameter estimation and to find GEI pattern. However, in this study, it was not observed prediction superiority in ranking genotypes using BAMMI-H and BAMMI. It is important to note that the difference between the variance of environment 1 and environment 7 was approximately 79 times suggesting that even under high level of heterogeneity of variances between environments, the BAMMI method is very robust in prediction missing GEI values. Similar results were found by So and Edwards (2011) that using different variance structures and cross-validation and simulation analyses in maize MET showed no significant improvement by including heterogeneous (co)variance structures.

In general, the FA model was overcome by Bayesian models even with low levels of imbalance. A possible justification would be the existence of a weak genetic covariance between the environments that may have been caused by the nature of the data used in the simulation. In addition, the relatively small number of environments and genotypes used in this work may not reflect the complexity of the MET data described in Kelly et al. (2007) and Smith et al. (2015).

Regarding parameter estimation, the models that assumed heterogeneity of variances, such as FA and BAMMI-H, presented better performance and captured the fluctuations of the residual variances throughout the environments (Table 1). These results support the idea that assuming homogeneity of variance may not be reasonable when the objective is parameter estimation (Crossa et al., 1990; Edwards and Jannnik, 2006; Orellana, 2012 and Orellana et al., 2014).

The model ability to capture environmental nuances converge to a better selection precision altering the ranking of genotypes as highlighted by Smith et al. (2001). In this study, we did not estimate the changes in the ranking of genotypes under the different techniques as in Burgueno et al. (2012); instead the model ability to ranking the GEI effects. As reported by Kelly et al. (2007), the most important, from the point of view of the

breeding program, is the model predictive precision, since it directly affects the gains obtained in the selection process. Furthermore, the cross-validation process is the most drastic way of verifying the model ability to learn from data or to be robust under certain violations of assumptions.

Generally, although the observed differences between the methods compared, it was found that even fixed models are robust to data losses in the MET analyses. Nevertheless, although still little used in breeding, Bayesian AMMI methods were generally superior as can be seen in the efficiency gain in relation to prediction analyses (Figures 8 and 9) demonstrating the potential of the techniques in the study of genotype-by-environment interaction and prediction of missing genotypes.

Some authors such as Kelly et al. (2007), Burgueno et al. (2008, 2012) and Nuvunga et al. (2015) argue that the FA2 model is the best predictor of missing genotypes or analysis of MET assays with unbalanced data, however, to date the Bayesian methods suggested by Crossa et al. (2011), Oliveira et al. (2015) and Silva et al. (2015) were not compared with AF models. Silva et al. (2015) shows, for example, that BAMMI models exhibit shrinkage characteristics at single values. This feature suggests that low-dimensional BAMMI models with random singular values may present higher predictive power for missing data. This configuration was not adopted in this study, although it may be incorporated in later studies.

The magnitude of simulated heritability in this study was 0.87. The magnitudes of heritabilities estimated on the basis of the competing models were, in general, moderate (between 0.69 and 0.89). The relationship between the estimated heritability and the observed prediction accuracy reflects the predictive power of the methods achieving the maximum accuracy allowed. This result reinforces what has been discussed in (Crossa et al., 2010; Oakey et al., 2016; Kelly et al., 2009; Heslot et al., 2013) on the importance of incorporating GEI in genomic selection studies. Since current methods are able to predict the behavior of genotypes in untested environments, the inclusion of molecular marker information in FA or BAMMI models can greatly contribute to the prediction of non-evaluated genotypes.

The results of the analyses indicate that all models are generally robust under imbalance. Even the classical versions of the AMMI model, using the EM method, presented good performances in the prediction analyses. In relation to the AMMI-Bayesian approach, it was verified the robustness of the BAMMI model, which presented similar performance to its version that models different variances between the environments, although the comparison of models should also be based on inference criteria and in this sense is of consensus that the assumption of homogeneity of variance, as already emphasized in the course of this text, may not be reasonable.

ACKNOWLEDGMENTS

We thank Eduardo Mondlane University for the financial support.

REFERENCES

- Balestre M, Von Pinho RG and Souza JC (2010). Prediction of maize single-cross performance by mixed linear models with microsatellite marker information. *Genet. Mol. Res.* 9: 1054-1068.

- Beeck CP, Cowling WA, Smith AB and Cullis BR (2010). Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. *Genome*. 53: 992-1001.
- Burgueño J, Crossa J, Cornelius PL, Trethowan R, et al. (2007). Modeling additive x environment and additive x additive x environment using genetic covariances of relatives of wheat genotypes. *Crop Sci*. 47: 311-320.
- Burgueño J, Crossa J, Cornelius PL and Yang RC (2008). Using factor analytic models for joining environments and genotypes without crossover genotype x environment interaction. *Crop Sci*. 48: 1291-1305.
- Burgueño J, Crossa J, Cotes JM, Vicente FS, et al. (2011). Prediction assessment of linear mixed models for multi-environment trials. *Crop Sci*. 51: 944-954.
- Burgueño J, de los Campos G, Weigel K, and Crossa J (2012). Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. *Crop Sci*. 52: 707-719.
- Cornelius PL, Crossa J and Seyedsadr MS (1996). Statistical test and estimator of multiplicative models for genotype-by-environment interaction. In: Gauch HG, Kang M (eds). *Genotype-by-Environment Interaction*. CRC Press. Boca Raton. 199-234.
- Crossa J (1990). Statistical analyses of multilocation trials. *Adv. Agron*. 44: 55-85.
- Crossa J, Gauch HG and Zobel RW (1990). Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. *Crop Sci*. 30: 493-500.
- Crossa J, Cornelius PL and Yan W (2002). Biplot of linear-bilinear models for studying crossover genotype x environment interaction. *Crop Sci*. 42: 619-633.
- Crossa J, Burgueño J, Cornelius PL, McLaren G, et al. (2006). Modeling genotype x environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. *Crop Sci*. 46: 1722-1733.
- Crossa J, de Los Campos G, Pérez P, Gianola D, et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*. 186: 713-724.
- Crossa J, Pérez P, De Los Campos G, Mahuku G, et al. (2011). Genomic selection and prediction in plant breeding. *J. Crop. Improv.* 25: 239-261.
- Crossa J (2012). From genotype x environment interaction to gene x environment interaction. *Curr. Genomics*. 13: 225-244.
- Curti RN, De la Vega AJ, Andrade AJ, Bramardi SJ, et al. (2014). Multi-environmental evaluation for grain yield and its physiological determinants of quinoa genotypes across Northwest Argentina. *Field Crops Res*. 166: 46-57.
- Dias CTS and Krzanowski WJ (2003). Model selection and cross-validation in additive main effect and multiplicative interaction models. *Crop Sci*. 43: 865-873.
- Edwards JW and Jannink JL (2006). Bayesian modeling of heterogeneous error and genotype x environment interaction variances. *Crop Sci*. 46: 820-833.
- Gauch HG and Zobel RW (1988). Predictive and postdictive success of statistical analyses of yield trials. *Theor. Appl. Genet.* 76: 1-10.
- Gauch HG and Zobel RW (1990). Imputing missing yield trial data. *Theor. Appl. Genet.* 79: 753-761.
- Sa'diyah H and Hadi AF (2016). AMMI Model for Yield Estimation in Multi-Environment Trials: A Comparison to BLUP. *Agric. and Agric. Sci Proc*. 9: 163-169.
- Heidelberger P and Welch PD (1983). Simulation run length control in the presence of an initial transient. *Oper. Res* 31: 1109-1144.
- Hu X and Spilke J (2011). Variance-covariance structure and its influence on variety assessment in regional crop trials. *Field Crops Res*. 120: 1-8.
- Josse J, van Eeuwijk FA, Piepho HP and Denis JB (2014). Another look at Bayesian analysis of AMMI models for genotype-environment data. *J. Agric. Biol. Environ. Stat.* 19: 240-257.
- Kelly AM, Smith AB, Eccleston JA and Cullis BR (2007). The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Sci*. 47: 1063-1070.
- Kelly AM, Cullis, BR, Gilmour AR, Eccleston JA, et al. (2009). Estimation in a multiplicative mixed model involving a genetic relationship matrix. *Genet. Sel. Evol.* 41: 1-9.
- Mayer K (2009). Factor-analytic models for genotype x environment type problems and structured covariance matrices. *Genet. Sel. Evol.* 41. Available at [<http://www.gsejournal.biomedcentral.com/articles/10.1186/1297-9686-41-21>].
- Möhring J, Piepho HP (2009). Comparison of weighting in two-stage analyses of series of experiments. *Crop Sci*. 49: 1977-1988.
- Heslot N, Akdemir D, Sorrells ME and Jannink JL (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127: 463-480.
- Nuvunga JJ, Oliveira LA, Pamplona AKA, Silva CP, et al. (2015). Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. *Genet. Mol. Res.* 14: 14262-14278.
- Oakey H, Verbyla A, Pitchford W, Cullis B, et al. (2006). Joint modeling of additive and non-additive genetic line effects in single field trials. *Theor. Appl. Genet.* 113: 809-819.
- Oakey H, Cullis B, Thompson R, Comadran J, et al. (2016). Genomic selection in multi-environment crop trials. *G3 (Bethesda)*. 6: 1313-1326.

- Oliveira LA, Silva CP, Nuvunga JJ, Silva AQ, et al. (2015). Credible intervals for scores in the AMMI with random effects for genotype. *Crop Sci.* 55: 465-476.
- Oliveira LA, Silva CP, Nuvunga JJ, Da Silva AQ, et al. (2016). Bayesian GGE biplot models applied to maize multi-environments trials. *Genet. Mol. Res.* 15: gmr.15028612
- Orellana MA (2012). Bayesian prediction of crop performance modeling genotype by environment interaction with heterogeneous variances. Doctoral thesis. Iowa State University. Available at [<http://lib.dr.iastate.edu/etd/12740>].
- Orellana M, Edwards J and Carriquiry A (2014). Heterogeneous variances in multi-environment yield trials for corn hybrids. *Crop Sci.* 54: 1048-1056.
- Paderewski J and Rodrigues PC (2014). The usefulness of EM-AMMI to study the influence of missing data pattern and application to Polish post-registration winter wheat data. *Aust. J. Crop Sci.* 8: 640-645
- 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× environment interaction. *J. Agric. Biol. Environ. Stat.* 17: 15-37.
- Piepho HP (1997). Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics.* 53: 761-766.
- Piepho HP (1998). Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theor. Appl. Genet.* 97: 195-201.
- Piepho HP and Möhring J (2006). Selection in cultivar trials- Is it ignorable? *Crop Sci.* 46: 192-201.
- Raftery AE and Lewis S (1992). How Many Iterations in the Gibbs Sampler? In *Bayesian Statistics*. Oxford Univ. Press. Oxford. 763-773.
- R Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna.
- Rodrigues PC, Malosetti M, Gauch HG and van Eeuwijk FA (2014). A weighted AMMI algorithm to study genotype-by-environment interaction and QTL-by-environment interaction. *Crop Sci.* 54: 1555-1570.
- Silva CP, Oliveira LA, Nuvunga JJ, Pamplona AKA, et al. (2015). A Bayesian shrinkage approach for AMMI models. *PLoS One.* 10: 1-27.
- Silva CP, Oliveira LA, Nuvunga JJ, Pamplona AKA, et al. (2018). Heterogeneity of variance in AMMI-Bayesian model in the study of multi-environment data. *Crop Sci.* (Accepted)
- Smith AB, Cullis BR and Thompson R (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics.* 57: 1138-1147.
- Smith AB, Ganesalingam A, Kuchel H, and Cullis BR (2015). Factor analytic mixed models for the provision of grower information from national crop variety testing programs. *Theor. Appl. Genet.* 128: 55-72.
- So YS and Edwards J (2009). A comparison of mixed-model analyses of the Iowa Crop Performance Test for Corn. *Crop Sci.* 49: 1593-1601.
- Stefanova KT and Buirchell B (2010). Multiplicative mixed models for genetic gain assessment in lupin breeding. *Crop Sci.* 50: 880-891.
- Studnicki M, Paderewski J, Piepho HP and Wójcik-Gront E (2017). Prediction accuracy and consistency in cultivar ranking for factor-analytic linear mixed models for winter wheat multi-environmental trials. *Crop Sci.* 57: 2506-2516.
- Thompson R, Cullis B, Smith A and Gilmour A (2003). A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Aust. N. Z. J. Stat.* 45: 445-459.
- Van Eeuwijk FA, Bustos-Korts DV, Malosetti M, Boer MP, et al. (2016). Will high throughput phenotyping and genotyping techniques help us to better predict G×E interactions? Some considerations from statistics and crop growth modelling. In 4th International Plant Phenotyping Symposium. CIMMYT. Mexico. 8-9
- Viele K and Srinivasan C (2000). Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback-Leibler information. *J. Statist. Plann. Inference.* 84: 201-219.
- Yan W and Kang MS (2003). GGE Biplot Analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press. Boca Raton.