

Stability and adaptability of elite upland rice lines using Bayesian-AMMI model

Joel Jorge Nuvunga¹, Alessandra Querino da Silva², Cristian Tiago Erazo Mendes^{3*}, Gabriel Cossa¹, Luciano Antonio de Oliveira², Carlos Pereira da Silva³, Nelio Cândido¹, Hermínio Bento Inácio⁴, Júlio Sílvio de Sousa Bueno Filho³

¹Universidade Eduardo Mondlane-Escola Superior de Empreendedorismo de Chibuto, Moçambique, Av. Samora Machel Cx Postal 25/26, Gaza-Moçambique

²Faculdade de Ciências Exatas e Tecnologia (FACET), Universidade Federal da Grande Dourados, CP 364, Mato Grosso do Sul, Brazil

³Departamento de Estatística (DES), Universidade Federal de Lavras, CP 3037, Minas Gerais, Brasil

⁴Instituto de Investigação Agrária de Moçambique, Av FPLM PAN, cx postal 2590-Nampula-Moçambique

*Corresponding author: tiagoerazo@gmail.com

Abstract

Rice is one of the world's most important crops. The search for genotypes that are more productive and have wide adaptation to different environments is paramount. One of the major breeder's obstacles faced is identification of superior strains is the presence of Genotype \times Environment Interaction (GEI), which motivated the development of countless statistical procedures aiming to offer more efficient studies. In this work we analysed adaptability and stability of 13 upland rice lineages as part of a genetic improvement program in nine different environments, resulting from local combination and years of agriculture. The experiment was conducted in a completely randomized block design, with three replicates. The main variable is the grain storage in kg/ha. The model applied is the Bayesian Main Additive Effects and Multiplicative Interaction (Bayesian-AMMI). Our implementation implies an extra assumption of random effects from genotypes coming from a single population as opposed to previous works in the literature. Credibility regions with maximum posteriori density allowed identification of cultivars with higher average yield. Stable genotypes showed an initial evidence of adaptation to an environment in this rice breeding program. Bayesian-AMMI is flexible, and starts to be more widely used, but our suggestion is promising in making it a more powerful tool.

Keywords: Genetic improvement; Genotype \times Environment Interaction; inference; multi-environment; recommendations.

Abbreviations: AMMI_Additive Main Effects and Multiplicative Interaction, BOA_Bayesian Output Analysis, E_Environment, EMBRAPA_Empresa Brasileira de Pesquisa Agropecuária, EPAMIG_Empresa de Pesquisa Agropecuária de Minas Gerais, G_Genotype, GEI_Genotypes \times Environments Interaction, HPD_Highest Posterior Density, MCMC_Markov chain Monte Carlo, MET_Multi-environment trial, MG_Minas Gerais, PCA_Principal Component Analysis, SVD_Singular Value Decomposition, VMF_von Mises-Fisher, UFLA_Universidade Federal de Lavras.

Introduction

Rice is one of the world's major crops, being the primary food source to more than half of the world's population. Rice supplies 27% of energetic supplies and 20% of protein in human diets (Van Nguyen and Ferrero, 2006; Ramos et al., 2011).

Rice can be grown in both swamped soils such as uplands. In both cases, the great challenge to plant breeders is to identify genotypes with ideal performance and broad adaptability to different environments. Such identification is difficult due to the presence of a phenomenon known as genotypes \times environments interaction (GEI). This effect is very important for quantitative characteristics, among them many economic traits (Yang and Kang, 2003).

Thus, a successful breeding program needs to test crop adaptability to cultivation regions, especially elite lines (Cossa, 1990; Colombari Filho et al., 2013). Several studies in Asian countries showed that the GEI is relatively high in

the context of rice production (Cooper et al., 1999a, b; Inthapanya et al., 2000).

Adaptability can be viewed as the capacity of a genotype to favorably respond to environmental stimuli, while the stability is the capacity of genotype to have an expected performance in different environmental conditions (Resende and Duarte, 2007; Borém et al., 2017).

Several methodologies have been used to quantify adaptability and stability for decades, involving univariate, multivariate, parametric, and non-parametric analysis (Finlay and Wilkinson, 1963; Eberhart and Russell, 1966; Perkins and Jinks, 1968; Freeman and Perkins, 1971; Shukla, 1972; Lin and Binns, 1988; Annicchiarico, 1992; Colombari Filho et al., 2013). Among diverse statistical procedures developed for this purpose, the class of linear-bilinear models (Cornelius and Seyedsadr, 1997) has distinguished itself for the wide applicability in analysis of Multi-Environment Trials

(MET) data, especially the model of Additive Main Effects and Multiplicative interaction (AMMI).

The AMMI model with fixed effects is fitted in two stages, analyzing the main effects, and describing the GEI through principal component analysis (PCA). Resulting patterns from PCA can be graphically represented in a biplot (Kempton, 1984; Crossa, 1990; Gauch Jr. and Zobel, 1997). The efficiency of AMMI procedure was clearly demonstrated by several authors using the MET data in soybeans, maize, and wheat (Crossa et al., 1991; Yan and Hunt, 2001; Tarakanovas and Ruzgas, 2006; Tariku et al., 2013; Gauch Jr., 2013). AMMI has also been successfully used to evaluate rice genotypes in different environments (Tariku et al., 2013; Bose et al., 2014; Akter et al., 2015).

Despite those descriptive advantages, the AMMI implementation uses fixed effects, and therefore is bound to limitations such as the difficulty in dealing with heterogeneous and/or unbalanced data sets. Resulting biplot in classical AMMI analysis brings no uncertainty measures on plotted scores. On the other hand, pure frequentist's methods to incorporate inference to biplot have been criticized upon using restrictive assumptions in the randomization of statistics and for problematic intensive computation procedures in the rows and columns of the GEI matrix (Yang et al., 2009; Crossa et al., 2011).

These limitations can be avoided by using Bayesian methods. Crossa et al. (2011) have drawn credibility regions to AMMI-2 biplots. Other recent studies have demonstrated the Bayesian-AMMI flexibility for the GEI study (Perez-Elizalde et al. 2012; Oliveira et al., 2015; Silva et al., 2015; and others). Another advantage of this method is to avoid choosing parameters to be treated as fixed or random, which is a dilemma in mixed models (Smith et al., 2001; Nuvunga et al., 2015).

In this work we carried out a study on adaptability and stability of upland rice genotypes using the Bayesian-AMMI model. The main objectives were to identify genotypes with wide recommendations and/or specific adaptations using the more recent advances in Bayesian inference for AMMI models. This evaluation is extremely important, but usually messy due to multiple overlapping of confidence regions. Our implementation follows (Oliveira et al. 2015) and implies an extra assumption of random effects from genotypes coming from a single population as opposed to Crossa et al. (2011). We expect to obtain more realistic estimates for the genotypic values using a hierarchical prior distribution specification for the main effects of genotypes.

Results

A joint analysis of variance was presented for experiments from all considered environments (Table 1). *F* tests for environments, blocks and interaction sources of variation were all significant. It is commonplace that a significant interaction does not allow consideration of additive model for all main effects (Duarte and Vencovsky, 1999). However, evaluations should be conditional on levels of main effects.

In Table 2, estimates for the fixed effects AMMI model parameters are depicted. It can be observed that the first two main axes explain 81.51% of the GEI variability. The selected model was AMMI5 by the Cornelius Fr (Cornelius et al., 1992).

To approximate joint posterior distribution for Bayesian Inference, MCMC chains with 88 thousand iterations were simulated for each model parameters. The first 8000

observations (burn-in) were discarded and every fourth observation was sampled (thinning), resulting in a sample size of 5,000. Good convergence properties were observed for all parameters of diagnostic criteria with a dependence factor $k < 5$ (Raftery and Lewis, 1992). All parameters also pass stationary testing (Heidelberg and Welsh, 1983), signaling that samples could be used to carry on the inference.

On Fig. 1, summaries of the marginal distributions *a posteriori* are presented (averages and Higher Posterior Density intervals - 95% HPD). Overlaps in this graph suggest similar effects. Measures are ranked in descending magnitude (from left to right). Note that G4, G7 and G9 could be selected based on average productivity once their ranges do not include zero. From a breeder's viewpoint, it is likely that selection will increase average genetic values. In Table 3 point and interval estimates (at 95% credibility) for singular values and components of variance are presented. Frequentist estimates were obtained by singular value decomposition (SVD) of the interaction matrix and included for comparison purposes. It is observed that the two first main components explain about 97% of the interaction variability, higher than the percentage explained by the conventional analysis (81.51%). We used no procedure to adjust the number of components in the model but AMMI-2 biplot was presented based on two first components.

For the selection and recommendation of highest production genotypes, we choose combining ranking averages and HPD regions for GEI analysis. In the AMMI model, this is done by interpreting the biplot representation. In Bayesian-AMMI, adaptability and stability analysis are carried out to observe the positions and overlaps of credibility regions in the biplot. The AMMI-2 biplot along with the 95% credibility bivariate regions are shown in Fig. 2. Overlaps between regions indicate similar effects in relation to GEI. Genotypes and environments whose credibility regions include the origin were not represented in the biplot to simplify the interpretations.

From the visual analysis of the biplot, it is possible to suggest the formation of homogeneous subgroups {G7, G8, G10, G12, G13}, as well as the subgroups {G3, G6} and {G1, G5}, as well as {G2}, despite existence of some overlaps between regions. On the other hand, G4 and G9 form a fifth group whose credible regions for genotypic scores include the origin (point [0,0]); therefore, being considered stable (with no relevant contribution to GEI).

In the same way, it is possible to suggest homogeneous subgroups of environments, namely: {E2, E7, E9}, {E1}, {E5}, {E4} and {E3, E6, E8}, the latter being formed by stable environments (regions covering the origin).

In this interpretation, environments were considered independent. Knowledge that they are a composition of locations evaluated in different agricultural years (Table 2) must be considered in further interpretation as follows: First location (UFLA - Lavras) is represented in environments {E1, E3, E9}, two of those (E1 and E9) were contrasting in the biplot. Patos de Minas {E2, E4, E7} has greater similarity with a slightly different agricultural year 2015/16 (E7), whose credibility region is closer to the origin. Lambari {E5, E8} was the place that most contributed to the interaction in the agricultural year 2014/15 (E5), showing stability in the next harvest (2015/16) (E8). Muquem - Lavras (E6) tests were included in the agricultural year 2015/16 and was a stable environment in the biplot.

Suggesting genotypes adaptability or genotypes to target environments is another interesting decorrence of biplots. Interpretations based on inner products properties are done on resulting positions and overlaps of the respective credibility regions for the scores which describe the GEI. However, location consistency along years is paramount for cultivar recommendation and mega-environments identification.

Discussion

GEI's remarkable effect on genotypic performance verified by analysis of variance (Table 1) is reflected in biplot analysis. Some genotypes and environments do not contribute to the interaction, which is very common in MET tests that generally incorporate a mixture of crossing and non-crossing GEI patterns. In the present study, GEI apparently causes more dissimilarity in the genetic systems which control the physiological process, conferring yield stability in different environments. Similar results are found in Saied (2010) and Tariku et al. (2013).

G4 (CMG 2097) and G9 (CMG 1896) genotypes were superior, presenting significantly higher productivity than the general average. Based on biplot analysis, those genotypes are more stable and suitable for broad recommendation, having no significant contribution to GEI. G7 genotype (CMG 2089) has shown good adaptability to Lambari location. However, it is unstable and more testing would be important to confirm a consistent response. In fact, G7 would be considered adapted to Lavras in the agricultural year 2013/14, but had unfavorable responses in 2015/16. This important genotype \times location \times year interaction should be investigated.

Muquém - Lavras (E6) was evaluated in only one season and has shown relatively low discriminative capacity. Lambari and UFLA were also stable in 2015/16 (E8) - Lavras and 2014/15 (E3), respectively. This pattern has not been repeated in other years. Patos de Minas was the more consistent location resulting in more reliable recommendations. However, the subgroup adapted to this location is formed by the genotypes with the worst performance in terms of average yield ($\{G1, G3, G5, G6\}$ Figure 1). For these locations, productive genotypes such as G4 and G9 are recommend. Note that complex overlaps in which credibility regions spam more than a single quadrant can cause interpretation problems. Identification of homogeneous subgroups regarding GEI can be sometimes deceiving. Here, we use the criterion proposed by Júnior et al. (2018) interpreting positions and overlaps based on the first two PC lines and respective quadrants. Denis and Gower (1994) suggested separable subgroups even under intense overlaps. Perez-Elizalde et al. (2012), circumvented this problem by using a hierarchical algorithm based on Euclidean distances from posterior averages. However, this is a point estimate solution.

Data used in this study has been previously described and analyzed by Inácio et al. (2018) who used joint analysis of variance combined with other methods of GEI (Wricke, 1965; Lin and Binns, 1988; Annichiarico, 1992). In here, we advocate the use of the AMMI model, which offers a multivariate analysis and allows the disposal of additional residues present in the sum of squares of GEI (Duarte and Vencovsky, 1999). In addition, Bayesian-AMMI approximation by MCMC samples of the joint posterior

distribution allows for working out credibility regions to the biplot. This implies higher levels of reliability in discriminating and grouping genotypes and environments, as well as contributes to more accurate predictions (Smith et al., 2005; Oliveira et al., 2015).

In Bayesian predictions and SVD estimates for singular values (Table 3) a clearcut shrinking effect for the posterior averages is noted from λ_3 , despite the non-informative prior distributions choosed to variance components. This was also reported by several authors (Crossa et al., 2011; Oliveira et al., 2015). By using the Bayesian method, the GEI effect is pronounced for the first components, while others are shrunk to near zero. In this sense, more severe shrinking would be expected if informative priors were elicited as demonstrated by Silva et al. (2015).

Bayesian AMMI is a flexible inference tool. In this approach, random genotypes are considered (Oliveira et al., 2015). Dataset is balanced and the ranking of better genotypes will not change compared to the more traditional approaches. However, shrinking such as those obtained by Bayesian predictors allows for more accurate predictions of genotypic values (Piepho, 1995; Smith et al., 2001; Resende and Duarte, 2007).

Bayesian AMMI models are not yet widely used, but certainly offer broad perspectives for MET data analysis.

Materials and Methods

Plant materials

Thirteen upland rice lines were used, originating from a partnership between UFLA (Universidade Federal de Lavras), EPAMIG (Empresa de Pesquisa Agropecuária de Minas Gerais) and EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) Rice and Beans, are they : BRSMG Relâmpago (G1); CMG 1511 (G2), BRSMG Caçula (G3); CMG 2097 (G4); BRSMG Caravera (G5); Esmeralda (G6), CMG 2089 (G7); CMG 1977 (G8); CMG 1896 (G9); CMG 2085 (G10); CMG 1987 (G11), CMG 2170 (G12), CMG 1509 (G13).

Field trials

The experiments were conducted in nine environments that are combinations of locations and agricultural years (Table 4). The locations are in the cities of Lavras, Patos de Minas and Lambari, all belonging to the state of Minas Gerais (MG) -Brasil. In the 2013/14 harvest, experiments were conducted at Universidade Federal de Lavras (UFLA) located in Lavras-MG and Patos de Minas-MG. In the 2014/15 harvest, the experiments were conducted at UFLA / Lavras-MG, Muquém farm located in Lavras-MG, in Lambari-MG and in Patos de Minas-MG and in the 2015/16 harvest were conducted at UFLA / Lavras-MG, Lambari-MG and Patos de Minas-MG. The locations have a tropical semi-humid and tropical altitude climate, with quite different edaphic and climatic conditions, suitable for this type of study (Inacio et al., 2018).

More details on the locations of the experiments are available in Fig. 3 built using ArcGIS software version 10.8.1.

Trials

Experiments were conducted in a randomized complete block design with three blocks. Plots consisted of five rows

Table 1. Joint analysis of experiments for the rice production variable (kg. ha⁻¹). 13 rice genotypes evaluated in 9 environments combining four different locations and three agricultural seasons (2014/15/16/17) , in the state of Minas Gerais-Brazil

Source	DF	SS	MS	F-value	Pr(>F)
Environments	8	573.093	71.637	29.131	< 0.001***
Genotypes	12	47.529	3.961	1.611	0.101
Blocks (Env)	18	36.636	2.035	4.335	< 0.001***
GEI	96	236.078	2.459	5.237	< 0.001***
Residues	216	101.41			
Total	350	994.747			

DF: degree of freedom, SS: sum of squares, MS: means square, Level of significance: 0.001 = ***, 0.01 = **, 0.05 = *.

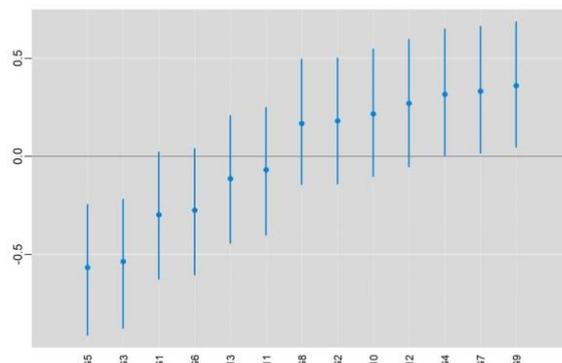


Fig 1. Highest Posterior Density (HPD) regions with 95% credibility for genotype main effects. The posteriori means are ranked in - order of magnitude from left to right.

Table 2. Summary of the AMMI model fitting to yield of 13 rice genotypes evaluated in 9 environments. Cornelius Fr test for the number of axes to be kept in the GEI model.

AMMI	P %	CP %	DF	SS	MS	Fr-value	Pr (>F)
AMMI1	64.98	64.98	19	153.415	8.074	17.198	<0.001 ***
AMMI2	16.53	81.51	17	39.023	2.295	4.889	<0.001 ***
AMMI3	6.07	87.58	15	14.323	0.955	2.034	0.014 *
AMMI4	4.73	92.31	13	11.156	0.858	1.828	0.040 *
AMMI5	3.84	96.14	11	9.058	0.823	1.754	0.064
AMMI6	2.01	98.15	9	4.744	0.527	1.123	0.348
AMMI7	1.09	99.25	7	2.581	0.369	0.785	0.6
AMMI8	0.75	100	5	1.778	0.356	0.758	0.581

P: percentage, CP: cumulative percentage, DF: degree of freedom, SS: sum of squares, MS: means square, Level of significance: 0.001 = ***, 0.01 = **, 0.05 = *.

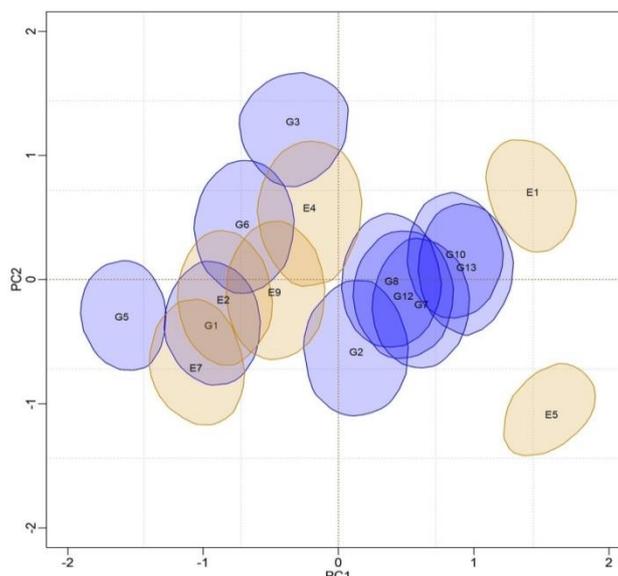


Fig 2. Biplot of Bayesian-AMMI model (AMMI-2). Genotypes (G) and Environment (E) scores, and respective bivariate credibility regions (95%) for the multi-environmental trial (MET) data of the rice. Only genotypes whose credibility regions do not include the origin (0.0) are depicted, to simplify the interpretations.

Table 3. SVD: likelihood estimates of singular values from the decomposition of GEI matrix. Bayesian -AMMI singular values and variance components estimates posterior means, standard deviations, lower and upper limits HPD (95%) regions.

Parameters	SVD	Mean	SD	LL	UL
λ_1	7.151	6.931	0.421	6.093	7739
λ_2	3.607	3.164	0.449	2.252	4.009
λ_3	2.185	1.189	0.593	0.024	2.156
λ_4	1.928	0.576	0.429	<0.001	1.362
λ_5	1.738	0.276	0.278	<0.001	0.849
λ_6	1.258	0.131	0.166	<0.001	0.485
λ_7	0.928	0.064	0.098	<0.001	0.262
λ_8	0.770	0.034	0.064	<0.001	0.156
σ_g^2	-	0.153	0.086	0.039	0.313
σ_e^2	-	0.509	0.048	0.418	0.602

SVD: Singular Value Decomposition, SD: Standard deviation, LL: Lower limit, UL: Upper limit, λ_k : k-th singular value, σ_g^2 : genotypic variance, σ_e^2 :residual variance

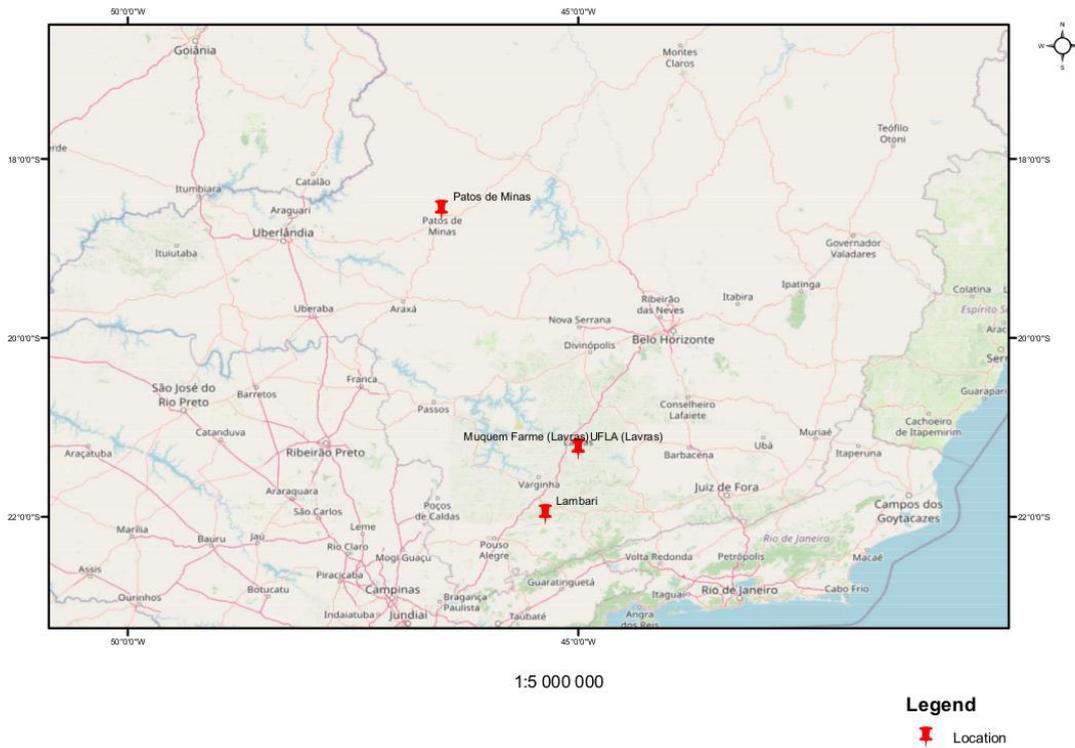


Fig 3. Places where the experiments with upland rice lines were installed

Table 4. Geographic coordinates of the locations (cities in the states of Minas Gerais-Brazil) where the tests were installed in different agricultural years. Environments (E) are constituted by a combination of places and agricultural years.

Places	Altitude	South Latitude	West Longitude
Lavras	919	21°14'45	44°59'59
Patos de Minas	832	18°34	46°31
Muquem	919	21°14'45	44°59'59
Lambari	845	21°58	45°22

Environments (E) composed of agricultural places and years	
E1	UFLA/Lavras 13/14
E2	Patos de Minas13/14
E3	UFLA/Lavras 14/15
E4	Patos de Minas 14/15
E5	Lambari 14/15
E6	Muquém/Lavras 14/15
E7	Patos de Minas 15/16
E8	Lambari 15/16
E9	UFLA/Lavras 15/16

of 4 m with sowing density of 80 seeds/meter. The crops treatment used in the tests were the same as those recommended for culture. Evaluated character was grain yield (kg.ha-1).

Statistical methods

Bayesian AMMI Model

The Bayesian AMMI Model is expressed by:

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

where the vector $\mathbf{y}_{n \times 1}$ is a vector of $n = lr$ phenotype realizations, being r the number of genotypes $l = bc$ the number of repetitions, where $b=3$ is the number of blocks and $c=9$ is the number of environments. Vector $\boldsymbol{\beta}$ has main effects of blocks and environments, and vector \mathbf{g} has genotypes main effects.

The terms λ_k , $\boldsymbol{\alpha}_k$, $\boldsymbol{\gamma}_k$ are, respectively, the k -th singular value and the k -th genotypic and environmental singular vectors related to the k -th main component, ($k=1, \dots, t$), subject to decomposition restrictions by singular value being, $t = \min(r, c)$, the post of the matrix interaction between genotypes and environments $\mathbf{GE}_{r \times c}$. The design matrices \mathbf{X}_1 , \mathbf{X}_2 and \mathbf{Z} describe, respectively, the linear model for Environments, Blocks and Genotypes main effects. The vector $\boldsymbol{\varepsilon}_{n \times 1}$ contain random error effects with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_n\sigma_\varepsilon^2)$.

The prior distribution specification for model parameters are the same used by Oliveira *et al.* (2015):

$$\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \sigma_\beta^2 \sim N(\boldsymbol{\mu}_\beta, \mathbf{I}_\beta \sigma_\beta^2), \sigma_\beta^2 \rightarrow \infty \text{ which is equivalent to } \beta \sim \text{constant}; \mathbf{g} | \boldsymbol{\mu}_g, \sigma_g^2 \sim N(\mathbf{0}, \mathbf{I}_g \sigma_g^2), \sigma_g^2 \sim \frac{1}{\sigma_g^2};$$

$$\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2 \sim N^+(\mu_{\lambda_k}, \sigma_{\lambda_k}^2), \sigma_{\lambda_k}^2 \rightarrow \infty \text{ in a way that } \lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2 \sim \text{constant};$$

$$\boldsymbol{\alpha}_k \sim \text{spherical uniform on the correct subspace};$$

$$\boldsymbol{\gamma}_k \sim \text{spherical uniform on the correct subspace};$$

$$\sigma_\varepsilon^2 \sim \frac{1}{\sigma_\varepsilon^2}.$$

The likelihood was assumed as:

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

$$\text{Where; } \bar{\boldsymbol{\theta}} = (\boldsymbol{\beta}, \mathbf{g}, \lambda_k, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \sigma_\varepsilon^2) \text{ and } \boldsymbol{\theta} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k.$$

The joint posterior distribution for all parameters is given by:

$$p(\boldsymbol{\Phi} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}, \sigma_\varepsilon^2) p(\mathbf{g} | \boldsymbol{\mu}_g, \sigma_g^2) p(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \sigma_\beta^2) p(\sigma_g^2 | v_g, S_g^2) \times$$

$$\times p(\sigma_\varepsilon^2 | v_\varepsilon, S_\varepsilon^2) \prod_{k=1}^t p(\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2) p(\boldsymbol{\alpha}_k) p(\boldsymbol{\gamma}_k)$$

$$\text{where; } \boldsymbol{\Phi} = (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda, \boldsymbol{\beta}, \sigma_\beta^2, \sigma_\varepsilon^2).$$

Considering the assumptions made regarding the hyperparameters of the prior densities one can rewrite the joint distribution *a posteriori* as follows:

$$p(\boldsymbol{\Phi} | \mathbf{y}) \propto (\sigma_\varepsilon^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} (\mathbf{y} - \boldsymbol{\theta})^\top (\mathbf{y} - \boldsymbol{\theta}) \right\} (\sigma_g^2)^{-\frac{ng}{2}} \exp \left\{ -\frac{1}{2\sigma_g^2} \mathbf{g}^\top \mathbf{g} \right\} \frac{1}{\sigma_\beta^2} \frac{1}{\sigma_\varepsilon^2}$$

Full conditional distribution are all obtained in closed forms as presented below:

$$\boldsymbol{\beta} | \dots \sim N[(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top (\mathbf{y} - \mathbf{Z}\mathbf{g} - \boldsymbol{\theta}), (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \sigma_\beta^2];$$

$$\text{where } \boldsymbol{\theta} = \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k;$$

$$\mathbf{g} | \dots \sim N \left[\left(\mathbf{Z}^\top \mathbf{Z} + \mathbf{I} \frac{\sigma_\varepsilon^2}{\sigma_g^2} \right)^{-1} \mathbf{Z}^\top (\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \boldsymbol{\theta}), \left(\mathbf{Z}^\top \mathbf{Z} + \mathbf{I} \frac{\sigma_\varepsilon^2}{\sigma_g^2} \right)^{-1} \sigma_g^2 \right];$$

$$\sigma_\varepsilon^2 | \dots \sim \text{Scale} - \chi^{-2}[n, (\mathbf{y} - \boldsymbol{\theta})^\top (\mathbf{y} - \boldsymbol{\theta})];$$

$$\sigma_g^2 | \dots \sim \text{Scale} - \chi^{-2}[ng, \mathbf{g}^\top \mathbf{g}];$$

$$\lambda_k | \dots \sim N^+[(\boldsymbol{\phi}_k^\top \boldsymbol{\phi}_k)^{-1} \boldsymbol{\phi}_k^\top \boldsymbol{\Delta}_{k'}, (\boldsymbol{\phi}_k^\top \boldsymbol{\phi}_k)^{-1} \sigma_\varepsilon^2];$$

$$\text{Considering that: } \boldsymbol{\Delta}_{k'} = \mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \sum_{k' \neq k} \lambda_{k'} \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_{k'})\mathbf{X}_2\boldsymbol{\gamma}_{k'},$$

$$\boldsymbol{\phi}_k = \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k \text{ and } \lambda_1 \geq \dots \geq \lambda_t \geq 0.$$

$p(\boldsymbol{\alpha}_k | \dots) \propto \exp\{k\boldsymbol{\alpha}_k^\top \boldsymbol{\Lambda}_k^\top (\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\mathbf{g})\}$; has a distribution that proportional to the Von Mises-Fisher distribution: $\boldsymbol{\alpha}_k | \dots \sim \text{VMF}[\rho_k, \boldsymbol{\mu}_{\boldsymbol{\alpha}_k}]$ with concentration parameter and directional mean proportional, respectively, to $\rho_k = \frac{\lambda_k}{\sigma_\varepsilon^2}$ and $\boldsymbol{\mu}_{\boldsymbol{\alpha}_k} = \boldsymbol{\Lambda}_k^\top (\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\mathbf{g})$, being $\boldsymbol{\Lambda}_k = \text{diag}(\mathbf{X}_2\boldsymbol{\gamma}_k)\mathbf{Z}$.

Similarly, $\boldsymbol{\gamma}_k$ has a full conditional posterior distribution that is proportional to the Von Mises-Fisher distribution, $\boldsymbol{\gamma}_k | \dots \sim \text{VMF}[\rho_{\boldsymbol{\gamma}_k}, \boldsymbol{\mu}_{\boldsymbol{\gamma}_k}]$; $\boldsymbol{\mu}_{\boldsymbol{\gamma}_k} = \boldsymbol{\Omega}_k^\top (\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\mathbf{g})$ and $\boldsymbol{\Omega}_k = \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2$.

Sampling and estimation process

To approximate the joint posterior distribution a Gibbs sampler was used via Markov chain Monte Carlo (MCMC) methods. Convergence of the resulting chains was verified using the Heidelberger and Welch (1983) and Raftery and Lewis (1992) methods. Resulting sample was used to produce marginal estimates for all parameters. All inference was performed using the R statistical software (R CORE TEAM, 2017).

Conclusion

Bayesian AMMI applied to this example allows for selecting stable and or adapted genotypes for yield such as: G4 (CMG 2097), G9 (CMG 1896, both stable) and G7 (CMG 2089, with initial evidence on adaptation to Lambari).

Producing credibility regions in biplot help to identify stable genotypes, as well as to group genotypes and environments according to existing interaction patterns with greater precision than previous techniques.

Acknowledgments

The authors would like to thank the Agência Sueca de Cooperação para o Desenvolvimento Internacional (ASDI), Fundo Nacional de Investigação de Moçambique (FNI), Universidade Eduardo Mondlane (UEM) and Escola Superior de Negócios e Empreendedorismo de Chibuto (ESNEC) for supporting this research.

References

- Akter A, Hasan MJ, Kulsum MU, Rahman MH, Paul AK, Lipi LF, Akter S (2015) Genotype × environment interaction and yield stability analysis in hybrid rice (*Oryza sativa* L.) by AMMI biplot. *Bangl Rice J.* 19(2):79-86.
- Annicchiarico P (1992) Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. *J Genet Breed.* 46:269-278.

- Borém A, Miranda GV, Fritsche-Neto R (2017) Melhoria de plantas. Editora UFV, Viçosa.
- Bose LK, Jambhulkar NN, Pande K, Singh ON (2014) Use of AMMI and other stability statistics in the simultaneous selection of rice genotypes for yield and stability under direct-seeded conditions. *Chil J Agric Res.* 74(1):3-9.
- Colombari Filho JM, de Resende MDV, de Moraes OP, de Castro AP, Guimarães EP, Pereira JA, Utumi MM, Breseghello F (2013) Upland rice breeding in Brazil: a simultaneous genotypic evaluation of stability, adaptability and grain yield. *Euphytica* 192:117-129.
- Cooper M, Rajatasereekul S, Immark S, Fukai S, Basnayake J (1999a) Rainfed lowland rice breeding strategies for Northeast Thailand. I. Genotypic variation and genotype \times environment interactions for grain yield. *Field Crops Res.* 64:131-151.
- Cooper M, Rajatasereekul S, Somrith B, Sriwisut S, Immark S, Boonwite C, Suwanwongse A, Ruangsook S, Hanviriyapant P, Romyen P, Porn-uraisanit P, Skulku E, Fukai S, Basnayake J, Podlich DW (1999b) Rainfed lowland rice breeding strategies for Northeast Thailand II. Comparison of intrastation and interstation selection. *Field Crops Res.* 64:153-176.
- Cornelius PL, Seyedsadr MS (1997) Estimation of general linear-bilinear models for two-way tables. *J Stat Comput Sim.* 58(4):287-322.
- Cornelius PL, Seyedsadr MS, Crossa J (1992) Using the shifted multiplicative model to search for "separability" in crop cultivar trials. *Theor Appl Genet.* 84:161-172.
- Crossa J (1990) Statistical analyses of multilocation trials. *Advances in Agron.* 44:55-85.
- Crossa J, Fox PN, Pfeiffer WH, Rajaram S, Gauch Jr. HG (1991) AMMI adjustment for statistical analysis of an international wheat yield trial. *Theor Appl Genet.* 81:27-37.
- Crossa J, Perez-Elizalde S, Jarquin D, Cotes JM, Viele K, Liu G, Cornelius PL (2011) Bayesian estimation of the additive main effects and multiplicative interaction model. *Crop Sci.* 51:1458-1469.
- Denis JB, Gower JC (1994) Asymptotic covariances for parameters of biadditive models. *Utilitas Math.* 46:193-205.
- Duarte JB, Vencovsky R (1999) Interação genótipos \times ambientes: uma introdução à análise AMMI. Sociedade Brasileira de Genética, Ribeirão Preto.
- Eberhart SA, Russell WA (1966) Stability parameters for comparing varieties. *Crop Sci.* 6:36-40.
- Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. *Aust J Agric Res.* 14(6):742-754.
- Freeman GH, Perkins JM (1971) Environmental and genotype-environmental components of variability VIII. Relations between genotypes grown in different environments and measures of these environments. *Heredity.* 27:15-23.
- Gauch Jr GH (2013) A simple protocol for AMMI analysis of yield trials. *Crop Sci.* 53: 1860-1869.
- Gauch Jr HG, Zobel RW (1997) Identifying mega-environments and targeting genotypes. *Crop Sci.* 37:311-326.
- Heidelberger P, Welch P (1983) Simulation run length control in the presence of an initial transient. *Operations Res.* 31(6):1109-1144.
- Inácio HB, Botelho FBS, de Moura AM, Mendes-Resende MP, dos Santos HO, de Castro AP, Reis MS, Bustamante FO (2018) Phenotypic plasticity of upland rice lines cultivated in Minas Gerais State. *Aust J Crop Sci.* 12(5):763-769.
- Inthapanya P, Sipaseuth, Sihavong P, Sihathep V, Chanphengsay M, Fukai S, Basnayake J (2000) Genotypic performance under fertilised and non-fertilised conditions in rainfed lowland rice. *Field Crops Res.* 65:1-14.
- Júnior LAYB, da Silva CP, de Oliveira LA, Nuvunga JJ, Pires LPM, Von Pinho RG, Balestre M (2018) AMMI bayesian models to study stability and adaptability in maize. *Agron J.* 110(5):1765-1776.
- Kempton RA (1984) The use of biplots in interpreting variety by environment interactions. *J Agric Sci.* 103:123-135.
- Lin CS, Binns MR (1988) A superiority measure of cultivar performance for cultivar \times location data. *Can J Plant Sci.* 68:193-198.
- Nuvunga JJ, Oliveira LA, Pamplona AKA, Silva CP, Lima RR, Balestre M (2015) Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. *Genet Mol Res.* 14(4):14262-14278.
- Oliveira LA, da Silva CP, Nuvunga JJ, da Silva AQ, Balestre M (2015) Credible intervals for scores in the AMMI with random effects for genotype. *Crop Sci.* 55:465-476.
- Perez-Elizalde S, Jarquin D, Crossa J (2012) A general bayesian estimation method of linear-bilinear models applied to plant breeding trials with genotype \times environment interaction. *J Agric Biol Environ Stat.* 17:15-37.
- Perkins JM, Jinks JL (1968) Environmental and genotype-environmental components of variability. *Heredity.* 23(3):339-356.
- Piepho HP (1995) Robustness of statistical tests for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trials. *Theor Appl Genet.* 90:438-443.
- R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Raftery AE, Lewis S (1992) How many iterations in the Gibbs sampler? In: Bernardo JM et al (ed.) *Bayesian statistics*, Oxford University, Oxford.
- Ramos LM, Sanches A, Cotes JM (2009) Testes multiambientais na seleção de genótipos de arroz utilizando o modelo de regressão nos sítios ou locais. *Ciênc Rural.* 39:52-57.
- Resende MDV, Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesq Agropec Trop.* 37(3):182-194.
- Saied ZF (2010) Survey of adaptation of thirty rice (*Oryza sativa* L.) genotypes to west guilan climatic conditions-Astara. *J Crop Ecophysiol.* 4(15):111-126.
- Silva CP, de Oliveira LA, Nuvunga JJ, Pamplona AKA, Balestre M (2015) A bayesian shrinkage approach for AMMI models. *PLoS One.* 10(7):1-27.
- Shukla GK (1972) Some statistical aspects of partitioning genotype environmental components of variability. *Heredity.* 29(2):237-245.
- Smith AB, Cullis BR, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics.* 57:1138-1147.
- Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *J Agric Sci.* 143(6): 449-462.
- Tarakanovas T, Ruzgas V (2006) Additive main effect and multiplicative interaction analysis of grain yield of wheat varieties in Lithuania. *Agron Res.* 4:91-98.
- Tariku S, Lakew T, Bitew M, Asfaw M (2013) Genotype by environment interaction and grain yield stability analysis of rice (*Oryza sativa* L.) genotypes evaluated in north western Ethiopia. *Net J Agric Sci.* 1:10-16.
- Van Nguyen N, Ferrero A (2006) Meeting the challenges of global rice production. *Paddy Water Environ.* 4:1-9.
- Wricke G (1965) Die Erfassung der Wechselwirkung zwischen Genotyp und Umwelt bei quantitativen Eigenschaften. *Z Pflanzenz.* 53:266-343.
- Yan W, Hunt LA (2001) Interpretation of genotype \times environment interaction for winter wheat yield in Ontario. *Crop Sci.* 41:19-25.
- Yan W, Kang MS (2003) GGE biplot analysis: a graphical tool for breeders, geneticists, and agronomists, 1st edn. Wiley CRC press, Boca Raton.
- Yang RC, Crossa J, Cornelius PL, Burgueño J (2009) Biplot analysis of genotype \times environment interaction: proceed with caution. *Crop Sci.* 49:1564-1576.