Revista Ciência Agronômica, v. 51, n. 4, e20207333, 2020
Centro de Ciências Agrárias - Universidade Federal do Ceará, Fortaleza, CE
www.ccarevista.ufc.br

Scientific Article
ISSN 1806-6690

Bayesian AMMI applied to food-type soybean multi-environment trials[1]

AMMI Bayesiano aplicado em ensaios multi-ambientes de soja tipo alimento

Gustavo Henrique Freiria[2]*, Leandro Simões Azeredo Gonçalves[3], Douglas Mariani Zeffa[4], Wilmar Ferreira Lima[5], Nelson da Silva Fonseca Júnior[5], Cássio Egídio Cavenaghi Prete[3] and Inês Cristina de Batista Fonseca[3]

ABSTRACT - A complicating factor for the selection of plant strains is the influence of a genotype-environment (GE) interaction. The Bayesian approach is a tool to increase the efficiency of adaptability and stability methodologies. In this context, the objective of this study was to evaluate the linear and bi-linear parameters of the additive main effects and multiplicative interaction (AMMI) analysis using the Bayesian approach for selection of food-type soybean genotypes in multi-environment trials. The grain yields of five lipoxygenase-free lines intended for human consumption of from the soybean breeding program of the Londrina State University and two commercial standards (BRS 257 and BMX Potência RR) were evaluated in four counties of the State of Paraná, Brazil, in the 2014/15, 2015/16 and 2016/17 growing seasons. Of the evaluated lines, only UEL 110 and UEL 122 had positive posterior genotypic effects, exceeding a probability of 95% against the commercial standard BRS 257. Only lines UEL 115 and UEL 123 did not contribute significantly to the GE interaction. Lines UEL 110 and UEL 122 proved adaptable to the largest number of environments with significant GE interaction and are therefore promising for the development of new food-type soybean cultivars. The use of AMMI1 (PC1 vs. effects genotypes) showed results for the stability of genotypes similar to AMMI2 (PC1 vs PC2), allowing a direct selection by the biplot for productivity and stability.

Key words: *Glycine max*. Bayesian inference. Genotype - environment. Functional food. Grain yield.

RESUMO - A interação genótipo–ambiente (GA) é um complicador para a seleção de novos genótipos. A abordagem Bayesiana é uma ferramenta que pode aumentar a eficiência das metodologias de adaptabilidade e estabilidade. Nesse contexto, o objetivo deste estudo foi avaliar os parâmetros lineares e bi-lineares da análise AMMI (*Additive Main Effects and Multiplicative Interaction*) pela abordagem Bayesiana na seleção de genótipos de soja tipo alimento em ensaios multi-ambientes. A produtividade de grãos de cinco linhagens livres das enzimas lipoxigenases e destinadas ao consumo humano do Programa de Melhoramento de Soja da Universidade Estadual de Londrina e duas cultivares comerciais (BRS 257 e BMX Potência RR) foram avaliadas em quatro municípios do Estado do Paraná, nas safras 2014/15, 2015/16 e 2016/17. Das linhagens avaliadas, apenas a UEL 110 e UEL 122 tiveram efeitos genotípicos *a posteriori* positivos, superiores a 95% de confiabilidade a cultivar comercial BRS 257. Somente as linhagens UEL 115 e UEL 123 não contribuíram significativamente para a interação GA. As linhagens UEL 110 e UEL 122 foram adaptáveis ao maior número de ambientes de interação GA significativas e, portanto, são promissoras para o desenvolvimento de novas cultivares de soja tipo alimento. A utilização da AMMI1 (CP1 *vs* efeitos genotípicos) mostrou resultados para a estabilidade semelhantes ao AMMI2 (CP1 *vs* CP2), o que possibilitou uma seleção direta pelo *biplot* para produtividade e estabilidade.

Palavras-chave: *Glycine max*. Inferência Bayesiana. Genótipo - ambiente. Alimento funcional. Produtividade de grãos.

DOI: 10.5935/1806-6690.20200077
Editor do artigo: Professor Salvador Barros Torres - sbtorres@ufersa.edu.br

*Author for correspondence

Received for publication in 28/04/2020; approved in 06/07/2020

[1] Parte da Tese do primeiro autor apresentada ao Programa de Pós-Graduação em Agronomia, Universidade Estadual de Londrina/UEL

[2] Departamento de Ciências Agrárias, Universidade do Estado de Minas Gerais/UEMG, Ituiutaba-MG, Brasil, gustavo.freiria@uemg.br (ORCID ID 0000-0003-2786-3810)

[3] Departamento de Agronomia, Universidade Estadual de Londrina/UEL, Londrina-PR, Brasil, leandrosag@uel.br (ORCID ID 0000-0001-9700-9375), cassio@uel.br (ORCID ID 0000-0003-0337-6874), inescbf@uel.br (ORCID ID 0000-0003-0129-8534)

[4] Programa de Pós-Graduação em Genética e Melhoramento, Universidade Estadual de Maringá/UEM, Maringá-PR, Brasil, douglas.mz@hotmail.com (ORCID ID 0000-0002-0808-7230)

[5] Instituto de Desenvolvimento Rural do Paraná/IAPAR-EMATER, Londrina-PR, Brasil, wilmar@idr.pr.gov.br (ORCID ID 0000-0003-4888-4282), nsfjr@idr.pr.gov.br (ORCID ID 0000-0003-2581-6017)

INTRODUCTION

Soybean [*Glycine max* (L.) Merrill] is a nutritious and functional food that is, rich in protein, vitamins, minerals, and dietary fiber. Moreover, it is also an important source of certain phytochemical compounds, specifically carotenoids and isoflavones, which are associated with antioxidant properties and the prevention or incidence reduction of several degenerative and tumoral chronic diseases (DAY, 2013; MA *et al*., 2010; RIGO *et al*., 2015).

The genetic elimination of lipoxygenase (LOX) enzymes improves the palatability of grains and derivatives by reducing the production of hexane compounds. Genotypes considered to be triple null, that is., those with a total absence of LOX in grains, can be classified as food-type and have special characteristics for human consumption (DESTRO *et al*., 2013; FREIRIA *et al*., 2016, 2018a). However, the number of food-type cultivars grown in Brazil is still small; less than 15 are registered by the Ministry of Agriculture, Livestock and Supply (MAPA); in other words, they account for less than 1% of the soybean cultivars registered in Brazil (BRASIL, 2019).

For the release of new cultivars, their performance must be tested in experiments under the environmental conditions of the recommended cultivation region (multi-environment trials). These experiments are essential to determine which cultivars have a good performance despite the genotype-environment (GE) interaction (VAN EEUWIJK; BUSTOS-KORTS; MALOSETTI, 2016). To minimize the GE interaction effects and improve the predictability of performance, the most stable genotypes adapted to the specific conditions of each environment must be identified and several statistical methodologies have been proposed to do so (FREIRIA *et al*., 2018b).

The additive main effects and multiplicative interaction analysis (AMMI) multivariate method is notable for its explanatory power of the GE interaction, owing to its greater ability to capture the sum of squares of the GE interaction compared to methods based on variance and regression analysis (GAUCH, 2013; ZOBEL; WRIGHT; GAUCH, 1988). However, this technique, along with other adaptability and stability methodologies, is used in a frequentist approach. This approach may reduce the efficiency of selection and indication of superior genotypes (VIELE; SRINIVASAN, 2000) and requires certain assumptions, such as homogeneity of variance, normal distribution and adjustment for data imbalance (COTES *et al*., 2006). To solve these problems, the use of AMMI with a Bayesian approach has been proposed (VIELE; SRINIVASAN, 2000).

Using the Bayesian approach, prior knowledge of the study parameters from previous experiments and the researcher's experience can be aggregated, resulting in the establishment of a prior distribution and more accurate estimates of the standard deviations and probability (so-called Highest Posterior Density -HPD) intervals, for the correct separation of genotypes and environments for the AMMI model (CROSSA, 2011; SILVA *et al*., 2015). The Bayesian approach to infer the linear and bi-linear parameters of the AMMI model has been used in studies with corn (*Zea mays*) (BERNARDO JÚNIOR *et al*., 2018; CROSSA *et al*., 2011; JARQUÍN *et al*., 2016; OLIVEIRA *et al*., 2015; PEREZ-ELIZALDE; JARQUÍN; CROSSA, 2012; SILVA *et al*., 2015) and cowpea (*Vigna unguiculata*) (OLIVEIRA *et al*., 2018). These authors described the effectiveness of the methodology to predict genotypic participation in GE interaction and can therefore contribute decisively to the choice and recommendation of new cultivars. In this context, this study evaluated linear and bi-linear parameters of the Bayesian AMMI analysis for the selection of soybean genotypes in multi-environment trials.

MATERIAL AND METHODS

Genotypes and experimental conditions

Five soybean lines (UEL 110, UEL 114, UEL 115, UEL 122, and UEL 123) from the Soybean Breeding Program of the State University of Londrina (PMSAH/UEL), all with grains free of lipoxygenase enzymes, and the commercial standards BRS 257 and BMX Potência RR, were evaluated in the counties of Londrina, Guarapuava, Ponta Grossa and Pato Branco in the 2014/2015, 2015/2016 and 2016/2017 growing seasons, for a total of 16 environments (Table 1). Sowing was carried out mechanically with a seeder in four rows, in a randomized complete block design with four replications. Base fertilization consisted of 250 kg ha^{-1} of the fertilizer mixture 00-20-20 (N-P-K). The seeds were treated with carboxanilide and dimethylditiocarbamate (Vitavax-Thiram®), at a concentration of 250 mL per 100 kg of seeds, and inoculated at sowing with *Bradyhizobium japonicum*, strains SEMIA 5079 and 5080, with 5.0×10^9 colony-forming units per mL of the commercial product.

A no-tillage management system with planting on crop residues, in plots with four 5 -m rows spaced 45 cm apart, with 15 plants m^{-1} was used. The harvest was carried out after the R_8 development stage, where the two border lines and 50 cm of the far end of the two central lines were eliminated from evaluations, corresponding to 3.6 m^2 of the assessed area. The grain yield was determined, corrected to a moisture content of 13%, and extrapolated to kg ha^{-1}.

Table 1 - Location and climate characterization of 16 environments in the state of Paraná, Brazil

Environments	County	Sowing date	Altitude (m)	Latitude (S)	Longitude (W)	Climate[1]
A1	Londrina	07/10/14	576	23°21’	51°09’	Cfa
A2	Londrina	04/12/14				
A3	Londrina	08/10/15				
A4	Londrina	11/10/16				
A5	Londrina	04/11/16				
A6	Guarapuava	15/10/14	1120	25°23’	52°27’	Cfb
A7	Guarapuava	05/11/14				
A8	Guarapuava	13/10/15				
A9	Guarapuava	08/11/16				
A10	Ponta Grossa	16/10/14	880	25°13’	50°01’	Cfb
A11	Ponta Grossa	03/11/14				
A12	Ponta Grossa	12/10/15				
A13	Pato Branco	14/10/14	760	26°11’	52°42’	Cfa
A14	Pato Branco	12/11/14				
A15	Pato Branco	13/10/15				
A16	Pato Branco	04/10/16				

[1] According to the classification of Köppen-Geifer

Statistical analysis

The inferences regarding the linear and bi-linear parameters of the Bayesian AMMI model were drawn as proposed by Crossa *et al.* (2011). In matrix notation, the frequentist AMMI model can be described by the formula:

$$y = 1_n\mu + X_1\tau + X_2\delta + \sum_{k=1}^{t} \lambda_k \, diag(X_1\alpha_k)X_2\gamma_k + \varepsilon \qquad (1)$$

The vector y contains $n = rge$ phenotypic responses of g genotypes in e environments and r replicates.

Where:

1_n is the vector of the order $n \times 1$;

μ is the overall mean;

X_1 is the matrix of genotypes of order $n \times g$;

τ is the fixed-effect vector $g \times 1$ for genotypes;

X_2 and δ are the matrices for environments of the order $n \times e$ and the fixed-effect vector $e \times 1$ for environment, respectively;

λ_k is the singular value for the kth principal component;

t is the number of multiplicative terms $[t \leq min\,(g, e)\, -1]$;

α_k and $\boldsymbol{\gamma}_k$ are the singular vectors of k for genotypes and environments, respectively;

And **ε** is the n-vector of random residual effects.

Vector **ε** has a multivariate normal distribution with zero mean and variance-covariance matrix $\sigma^2_\varepsilon I_n$. Thus, vector ***y*** also has a multivariate normal distribution.

For Bayesian AMMI, the estimation of the parameters of the above equation model assumes that the conditional distribution of y, given that μ, τ, δ, λ, α, γ, and σ^2_ε, is a multivariate normal distribution.

$$y \,|\, \mu,\tau,\delta,\lambda,\alpha,\gamma,\, \sigma^2_\varepsilon \sim N\,(1n\; 1_n\mu + X_1\tau + X_2\delta + \sum_{k=1}^{t} \lambda_k \, diag(X_1\alpha_k)X_2\gamma_k, I_n\sigma^2_\varepsilon \qquad (2)$$

Where:

I_n is the identity matrix of order n.

The prior distributions used for the parameters were presented by Crossa *et al.* (2011), and are the same as those used by Viele and Srinivasan (2000) (subscript symbols of μ and σ^2 represent the mean and variance, respectively, of the prior distribution):

$\mu|\, \mu_\mu,\, \sigma^2_\mu \sim N(\mu_\mu,\, \sigma^2_\mu)$

$\tau|\, \mu_\tau,\, \sigma^2_\tau \sim N(\mu_\tau,\, I_g\sigma^2_\tau)$

$\delta|\, \mu_\delta,\, \sigma^2_\delta \sim N(\mu_\delta,\, I_e\sigma^2_\delta)$

$\lambda_k|\, \mu_{\lambda k},\, \sigma^2_{\lambda k} \sim N^+(\mu_{\lambda k},\, I_e\sigma^2_{\lambda k})$ with the following restrictions $\lambda_k > 0$ and $\lambda_{k-1} \geq \lambda_k$

σ_k ~ spherical uniform distribution in the corrected subspace;

γ_k ~ spherical uniform distribution in the corrected subspace;

$\sigma^2_\varepsilon | v_\varepsilon, s^2_\varepsilon \sim Inv - Escala - x^2 (v_\varepsilon, s^2_\varepsilon)$

Where:

N() represents the normal distribution;

N^+ the positive normal distribution and $Inv – Escala – \chi^2$ the inverse scale chi-square distribution;

μ_τ and $I_g\ \sigma^2_\tau$ are the prior vectors of means and the prior covariance matrix of the genotypic main effects;

μ_δ and $I_\varepsilon \sigma^2_\delta$ are the prior vectors of means and the prior covariance matrix for the environmental main effects;

$\mu_{\lambda k}$ and $\sigma^2_{\lambda k}$ are the prior means and the variances for the singular value λ_k;

q and c are the number of effects and the variance of μ, respectively;

υ_ε and s^2_ε are the degree of belief and the scale factor σ^2_ε.

Details of the prior spherical uniform distribution of GE parameters are reported in Viele and Srinivasan (2000). The values of μ_μ, σ^2_μ, μ_τ, σ^2_τ, μ_δ, σ^2_δ, $\mu_{\lambda k}$, $\sigma^2_{\lambda k}$, q, υ_ε and s^2_ε were chosen with the aim of representing the prior belief about the model parameters. Therefore, they represent hyperparameters.

In this study, the prior distributions were uninformative. In a breeding program, the lines pass for extensive evaluation only in the final stages of evaluation for the selection and/or recommendation of a new cultivar, which comprises several locations and years. This organization chart of a breeding program results in an uncertainty regarding the genotypic variation of the lines in the face of the GE interaction, and with that, previous experimental data for composition of the prior may result in a bad prior. Other recent work with the use of Bayesian AMMI for the selection and/or recommendation of superior genotypes also opted for a non-informative prior (BERNARDO JÚNIOR *et al.*, 2018; OLIVEIRA *et al.*, 2018). Future work must be carried out to clarify these issues.

The value zero was used as the prior distribution for the mean in all genotypic and environmental effects and high values for the variances, resulting in: $\mu_\mu = 0$, $\mu_\tau = 1_g \times 0$, $\mu_\delta = 1_e \times 0$ and $\mu_{\lambda k} = 0$, and for the variances, σ^2_μ, σ^2_τ, σ^2_δ and $\sigma^2_{\lambda k} = 1 \times 10^{15}$. Multiplying the prior distributions by the likelihood function, we obtain the following joint posterior distribution:

$$p\ (\mu,\tau,\delta,\lambda,\alpha,\gamma,\ \sigma^2_\varepsilon | y)$$

$$\alpha\ exp\ [-(1/2\ \sigma^2_\mu)(\mu_\mu - \mu)'\ (\mu_\mu - \mu)]$$

$$x\ exp\ [-(1/2\ \sigma^2_\tau)(\mu_\tau - \tau)'\ (\mu_\tau - \tau)]$$

$$x\ exp\ [-(1/2\ \sigma^2_\delta)(\mu_\delta - \delta)'\ (\mu_\delta - \delta)]$$

$$\mathrm{x} \prod_{k=1}^{t} \exp\left[-(1/2\,\sigma^2_\lambda)(\mu_\lambda - \lambda_k)'(\mu_\lambda - \lambda_k)\right]$$

$$x\ exp\ \{-(1/2\ \sigma^2_\varepsilon)\ [y - 1_n\mu + X_1\tau + X_2\delta + \textstyle\sum^t_{k=1} \lambda_k\, diag\ (X_1\alpha_k)\ X_2\gamma_k]'\ [y - 1_n\mu + X_1\tau + X_2\delta + \textstyle\sum^t_{k=1} \lambda_k\, diag\ (X_1\alpha_k)\ X_2\gamma_k$$

$$\mathrm{x} \left\{ (\sigma^2_\varepsilon)^{-\left\{\left[\frac{(n+\upsilon_\varepsilon)}{2}\right]-1\right\}} \exp\left[-(1/2\sigma^2_\varepsilon)\upsilon_\varepsilon s^2_\varepsilon\right] \right\} \tag{3}$$

Restrictions were and (for ($k \neq k^*$). The description of the parameters has previously been provided.

The posterior marginal distribution for each parameter was obtained by Gibbs sampling, with 100,000 iterations. Bayesian AMMI analysis was carried out using the R code developed by Crossa *et al*. (2011), and requested by the last author of this manuscript, with the help of the packages: coda (version 0.19-1), ggplot2 (version 3.1.0), gmm (version 1.6-2) , mass (version 7.3-51.1), matrix (version 3.5.1), movMF (version 0.2-3), msm (version 1.1.6), mvtnorm (1.0-10), lme4 (version 1.1-21), rstiefel (version 0.20), sandwich (version 2.5-0) and stats4 (version 3.5.1).

Chain convergence was verified using the criteria of Raftery and Lewis (1992). This method finds the number of iterations needed to estimate P[U ≤ μ|data] to within an accuracy of ±r with probability s, where U is a quantity of interest and μ is the qth posterior quantile of U from the Markov chain data (a dataset set with 100,000 iterations per parameter). It returns the number M of initial iterations to be discarded (burn-in), the number N of additional iterations required, and k, where every kth iterate is used. This method also yields diagnostics. One can determine in advance the minimum number of iterations needed, M_{mim}, and so $I = (M + N)/N_{mim}$ measures the increase in the number of iterations due to dependence in the sequence. The thinning used was the 95% percentile of values.

The cumulative variance ratio (φ_t) was calculated as proposed by Jarquín *et al*. (2016):

$$\varphi_t = \frac{\sum^t_{k=1} \lambda^2_k}{\sum^{\min(g,e)-1}_{k-1} \lambda^2_k},\ \mathrm{t = mim}(g,\ e) - 2 \tag{4}$$

Where:

g is number of genotypes;

e is number of environments;

λ_k is the singular value for the kth principal component;

t is the number of multiplicative terms.

RESULT AND DISCUSSION

For each parameter, 100,000 Markov chains were generated. To avoid the selection of observations that had not yet reached convergence, the first 294 observations were discarded (burn-in), with sampling every 20 observations (thinning), resulting in a final 4,985 iterations, according to the criterion proposed by Raftery and Lewis (1992). The interpretation of the traces generated by the iterations, in relation to the model parameters, indicate that the values were randomly distributed around a midpoint, without trend and no deviations caused by a double solution were found, as reported by Oliveira *et al.* (2015) and Silva *et al*. (2015), (Figure 1).

According to Oliveira *et al.* (2015), “an initial attention to the Bayesian AMMI analysis is a possible convergence for two solutions, one positive and one negative”. Silva *et al.* (2015), reported that this change in the periodicity in the convergence does not prejudice the analysis, as long as one of the solutions (negative or positive) is chosen, this choice does not affect the result of the biplot, but represents an additional step in the analysis.

The means for the effects of genotypes (τ) and environments (δ), with their respective probability intervals (HPD), are shown in Table 2. The mean of the posterior tests (μ) was 2,928.48 kg ha^{-1}, and only cultivar BMX Potência RR (τ = 364.91) and lines UEL 110 (τ = 107.04) and UEL 122 (τ = 52.95) had positive genotypic effects, without zero in their HPDs. According to Crossa *et al*. (2011), at a probability of 95%, positive HPD values exceed the overall mean.

The lowest posterior mean for grain yield was found for cultivar BRS 257, with τ = -243.42 (HPD$_{2.5\%}$ = -253.28 and HPD$_{97.5\%}$ = -233.95) and with no overlapping of HPD in relation to the other genotypes. The fact that this cultivar is the main food-type soybean genotype registered by the Ministry of Agriculture, Livestock and Food Supply (MAPA) (BRASIL, 2019) reinforces the productive potential of the lines evaluated in this study

Figure 1 - Traces generated by Markov Chain Monte Carlo (MCMC) Gibbs sampling for the singular vectors α_{11}, α_{12}, γ_{11} and γ_{12} and singular values λ_1 and λ_2

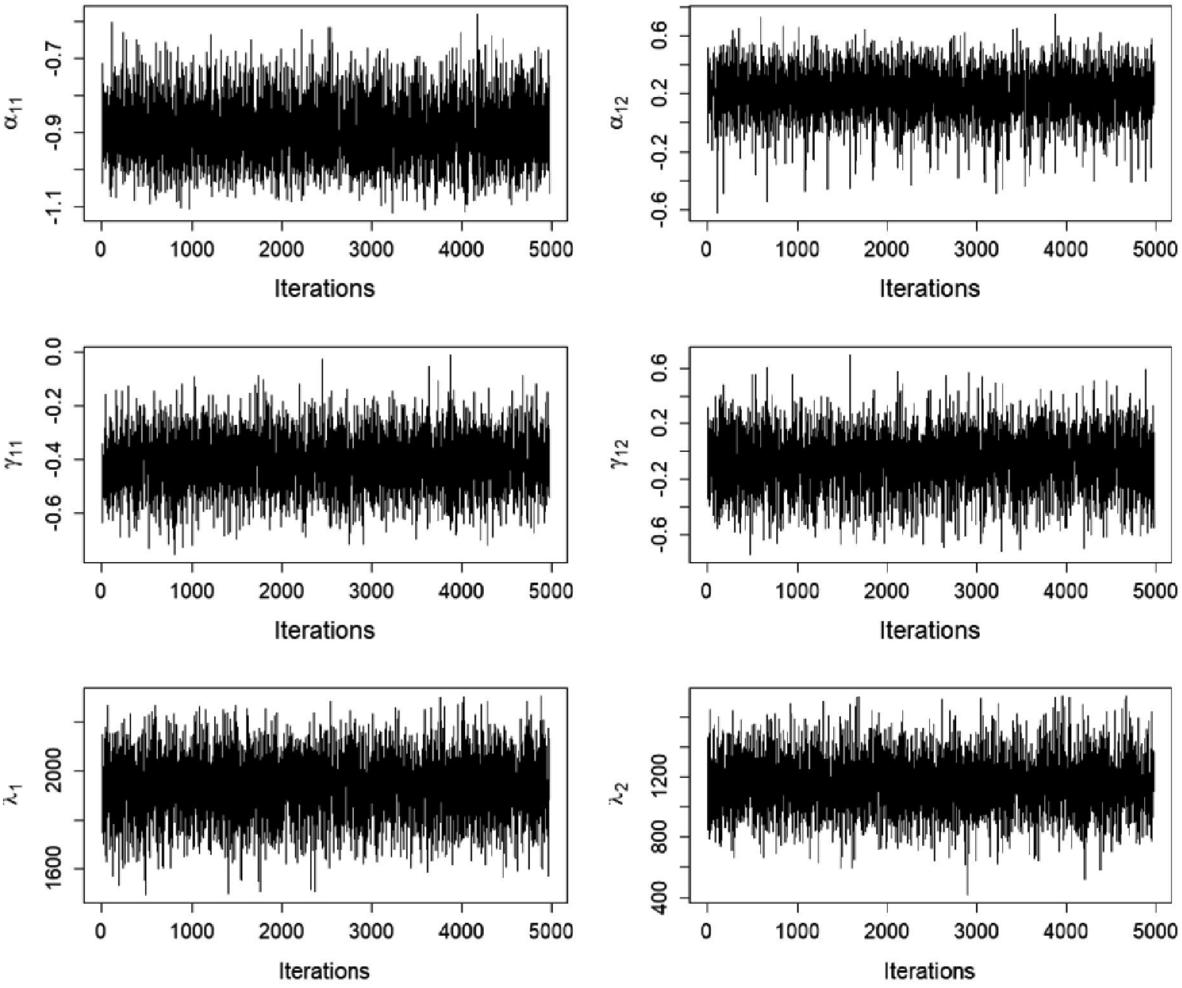


Table 2 - Posterior mean of the principal effects and multiplicative interaction (AMMI) for grain yield of seven food-type soybean genotypes (i = 1, 2, ..., 7) in 16 environments in the state of Paraná (j = 1, 2 , ..., 16) for the overall mean (μ), genotypic effects (τi), environmental effects (δj) and first and second singular values (λ1 and λ2) and their respective HPD at 95%

Parameters	Mean	HPD (2.5%)	HPD (97.5%)
μ	2928.48	2906.34	2949.49
τ1(BRS257)	-243.42	-253.28	-233.95
τ2(Potência)	364.91	355.33	374.44
τ3(UEL110)	107.04	97.37	116.96
τ4(UEL114)	-132.72	-142.35	-123.12
τ5(UEL115)	-96.17	-105.94	-86.68
τ6(UEL122)	52.95	43.34	62.74
τ7(UEL123)	-52.60	-62.54	-42.77
δ1(Londrina)	-978.90	-988.70	-968.85
δ2(Londrina)	-368.37	-378.73	-358.27
δ3(Londrina)	-829.72	-840.18	-819.59
δ4(Londrina)	1212.88	1202.84	1223.16
δ5(Londrina)	345.79	335.60	355.95
δ6(Guarapuava)	166.35	156.12	176.51
δ7(Guarapuava)	83.33	72.97	93.87
δ8(Guarapuava)	1403.81	1393.46	1414.06
δ9(Guarapuava)	77.06	66.94	87.11
δ10(P. Grossa)	-313.59	-323.93	-303.57
δ11(P. Grossa)	-218.91	-229.06	-208.89
δ12(P. Grossa)	944.43	934.33	954.65
δ13 (P. Branco)	-640.51	-650.72	-630.18
δ14 (P. Branco)	-1340.04	-1350.15	-1330.00
δ15 (P. Branco)	-1274.20	-1284.49	-1264.07
δ16 (P. Branco)	1730.59	1720.23	1740.71
λ1	1936.14	1687.66	2166.26
λ2	1145.58	815.64	1507.03

and its potential to compose a new food-type soybean cultivar.

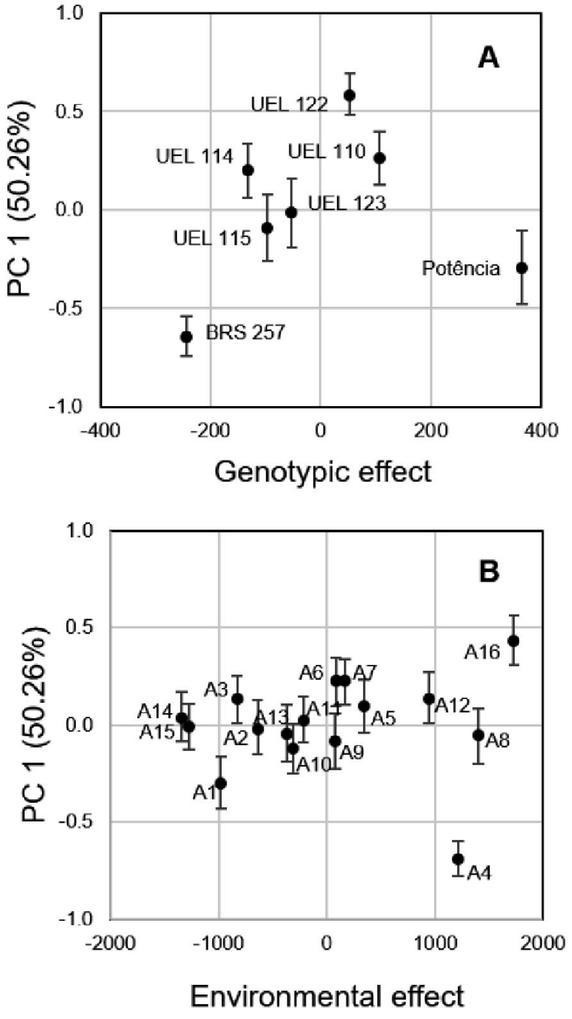
Among the 16 environments evaluated, eight of the posterior environmental effects were positive, in particular in Guarapuava (δ_6, δ_7, δ_8 and δ_9), where the means exceeded the overall mean in all trials. Differences in soybean grain yield between locations can be attributed to the climate, in particular to changes in temperature, rainfall, and photoperiod (MEOTTI *et al*., 2012).

According to Cruz, Regazzi and Carneiro (2012), superior genotypes would have a high mean productivity and combine broad adaptability and stability. To select such genotypes, apart from the genotypic effects, the GE interaction must be taken into account. The AMMI1 biplot was tailored for this purpose, with the first major component of the GE interaction (PC 1) for the genotypic scores ($\lambda^{1/2}_{1}\alpha_{i1}$), as opposed to their respective effects (Figure 2). According to Gauch (2013), in an AMMI1 biplot, stable genotypes would be those whose points are close to the origin of the PC 1 axis (coordinate 0 on the y-axis), that is, with practically zero scores for the first component of the GE interaction. PC 1 explained 50.26% of the interaction, and among the evaluated genotypes the lines UEL 110, UEL 114, UEL 115 and UEL 123 and cultivar BMX Potência RR contributed the least to this component (Figure 2A).

One of the main advantages of the Bayesian AMMI model is the possibility of also including HPDs for genotypic scores in the biplot, improving the precision of inferences on genotype stability, by reducing the subjectivity and proximity to the origin of the first interaction component. Thus, the only genotypes with coordinate 0 of PC 1 in their HPD at 95%, were lines UEL 115 and UEL 123. These results indicate that these genotypes are stable and do not significantly contribute to the GE interaction.

According to Cruz, Regazzi, and Carneiro (2012), for the recommendation of cultivars, stable genotypes should also have a desirable performance, which in this case, would be high grain yield. However, this condition was not met by these genotypes, as they had negative genotypic effects, i.e., grain yields below the overall mean (Figure 2A).

Figure 2 - Bayesian approach to the biplot AMMI1 for genotypic effects (A), environmental effects (B) and their contributions to the first principal component of the genotype x environment interaction (PC1), with their respective HPDs (at 95% probability)



Subtitle: the environments are described in table 1

These assessment criteria can be extended to the environment; AMMI1 with environmental scores ($\lambda^{1/2}_{1}\gamma_{i1}$) for PC 1, as opposed to its effects, are shown in Figure 2B. Among the 16 environments evaluated, only seven contributed significantly to the GE interaction (A1, A3, A4, A6, A7, A12, and A16), with HPD at 95% probability that coordinate PC 0 was not included. Environmental stability indicates the reliability of the genotype ordering, in relation to the classification of the mean of the tested environments in a given test environment. In this context, the nine stable environments tended to rank genotypes with little disagreement, compared to the mean classification at all sites.

All counties (Londrina, Guarapuava, Ponta Grossa, and Pato Branco) had at least one unstable environment (in terms of seasonal and/or crop variations), which shows that occasional variations in temperature and rainfall at the same test location are sufficient to contribute to the GE interaction. Variations at the same location between different years were also found by Bernardo Júnior *et al.* (2018), in a study of maize adaptability and stability by Bayesian AMMI. According to Figueiredo *et al.* (2015), in countries such as Brazil with wide climatic variation, the conditions at the same test site can vary greatly over the years, and may even exceed those observed between different geographical locations, suggesting that special attention should be paid to the environmental stratification in plant breeding programs in tropical and subtropical regions. Alternatively, multiway studies such as PARAFAC and Tucker3 can be employed in the construction of a triplot formed by the genotype × location × year interaction, as an ally in the interpretation of variations not only between test sites, but also variations in time at the same test site (ARAÚJO, 2009).

The inclusion of the second component (PC 2) ($\lambda^{1/2}_{2}\alpha_{i2}$ and $\lambda^{1/2}_{2}\gamma_{i2}$) in association with PC 1 allows the formation of the AMMI2 biplot, and is often used to explain more of the GE interaction. In the Bayesian AMMI2, the association of HPDs for the first and second components allows the creation of ellipses that also correspond to the 95% probability regions, where the stability is given by the overlapping ellipse and the intersection region of the coordinates 0.0 of PC 1 and PC 2, i.e., over the center of the biplot.

AMMI2 contributed to an explanation percentage of 67.92% of the GE interaction with the analysis (Figure 3). However, for all evaluated genotypes and environments, the ellipses covered coordinate 0 of the second component, classifying it as little discriminatory, resulting in identical responses to AMM1, with regard to the stability of genotypes and environments.

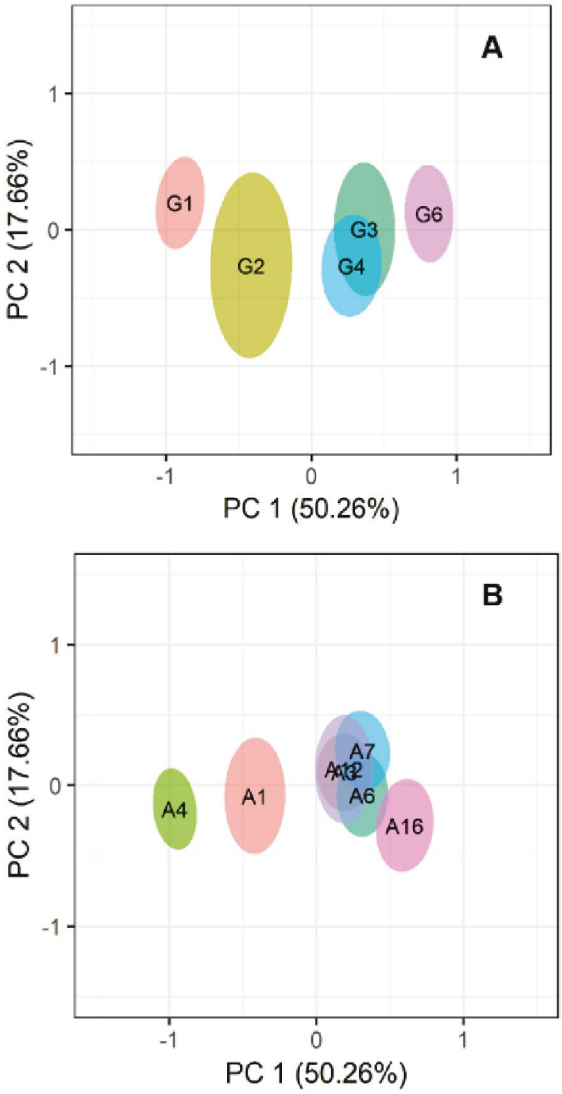
According to Crossa *et al*. (2011), PC1 discriminates genotypes better than PC2, in line with the theory of principal component analysis. The first component corresponds to greater variability, which results in less uncertainty, and consequently, in lower probability intervals associated with PC 1 compared to PC 2 and the other components. As mentioned by Gauch (2013) for the frequentist AMMI approach, as the number of selected axes rises, the "noise" percentage increases as well and reduces the predictive power of the analysis; this also applies to the Bayesian approach. This is because the sum of the squares of the GE interaction (SS_{GE}) that represents the usual data matrix in the AMMI analysis can be decomposed as $SS_{GE(standart)} + SS_{GE(noise)}$, the "noise" being related to the predictive error of the interaction. Thus, in this analysis the principal component (PC) should not necessarily explain 100% of the GE interaction, but only the percentage referring to the standard underlying the interaction. $SS_{GE} = \sum_{k=1}^{t}\lambda_k^2 = (\sum_{k=1}^{n}\lambda_k^2) + (\sum_{k=n+1}^{t}\lambda_k^2)$ PC_1 presents a greater explanation of the standard interaction in relation to PC_2 and $PC_2 > PC_3 > \ldots > PC_{min(g,e)-1}$.

This point is not taken into account in the most recent analysis of Bayesian AMMI, in which AMMI1 (PC1 vs. genotypic effects) is not being considered. With few distortions in relation to AMMI2 (PC1 vs PC2), the AMMI1 biplot becomes more indicated, due to the simultaneous selection of productive and stable genotypes.

Overlapping ellipses in AMMI2 indicate similar responses of genotypes or environments to the GE interaction (BERNARDO JUNIOR *et al*., 2018; CROSSA *et al*., 2011; JARQUÍN *et al*., 2016; OLIVEIRA *et al*., 2015, 2018; PEREZ-ELIZALDE; JARQUÍN; CROSSA, 2012; SILVA *et al*., 2015). In this sense, the performance of lines UEL 110 and UEL 114 was relatively unaffected by environmental variation (Figure 3A), and among the seven environments that significantly contributed to the interaction, four (A3, A6, A7, and A12) could be grouped in the same subgroup due to the overlap of their ellipses (Figure 3B).

Based on the PC1 and PC2 coordinates, conclusions could be drawn about the specific adaptation of genotypes to certain environments, taking into account the range of their ellipses and the quadrant sign to which they belong, resulting in the formation of the groups: (-,-), (-,0); (-,+); (0,-); (+,-); (+,0) and (+,+) (BERNARDO JÚNIOR *et al*., 2018). In this study, considering only the genotypes and environments with significant contributions to the GE interaction, two groups were formed: Group I (-, 0) formed by cultivars BRS 257 and BMX Potência RR, with specific adaptability to environments 1 and 4 (both in Londrina) and Group II (+, 0) consisting of lines UEL 110, UEL 114, and UEL 122, with specific adaptability to environments 3, 6, 7, 12, and 16, which covered the four evaluated counties (Figure 3 and Table 3).

Figure 3 - Bayesian approach to the biplot AMMI2 for the first (PC 1) and second (PC 2) principal component of the genotype x environment interaction for genotypic (A) and environmental scores (B), with their respective HPDs (at 95% probability), except for those with ellipses in the central region (0.0)



Subtitle: G1 - BRS 257, G2 - BMX Potência RR, G3 - UEL 110, G4 - UEL 114, G6 - UEL 122. The environments are described in table 1

Because of the positive genotypic effects, specific adaptability to the largest set of environments with significant contribution to the GE interaction (A3, A6, A7, A12, and A16) and maintenance of yield means in stable environments (A2 , A5, A8, A9, A10, A11, A13, A14, and A15), the lines UEL 110 and UEL 122

Table 3 - Genotypes clustering based on the first two principal components relates with 16 environments

Group	Genotypes	Environments
(-,-)	-	-
(-,0)	BRS 257; BMX Potência	A1, A4
(-,+)	-	-
(0,-)	-	-
(0,+)	-	-
(+,-)	-	-
(+,0)	UEL 110, UEL 114, UEL 122	A3, A6, A7, A12, A16
(+,+)	-	-
(0,0)	UEL 115, UEL 123	A2 , A5, A8, A9, A10, A11, A13, A14, A15

The clustering was build taking into account the 95% credible ellipse interval. The Cartesian coordinates indicate the ellipse signal and position in the biplot. The environments are described in Table 1

proved promising for the development of new soybean cultivars intended for human consumption.

CONCLUSIONS

1. The lines UEL 110 and UEL 122, with positive posterior genotypic effects and adaptability to a broad environmental set, are promising for the selection and release of new food-type soybean cultivars;
2. Bayesian AMMI analysis provides highly relevant information for the study of genotype adaptability and stability, contributing to the selection and/or recommendation of promising genotypes. The probability ranges (HPD) associated with genotypic and environmental effects as well as their biplot scores, allow for greater efficiency in selecting superior genotypes;
3. The use of AMMI1 (PC1 vs. effects genotypes) shows results for the stability of genotypes similar to AMMI2 (PC1 vs PC2) allowing a direct selection for productivity and stability by the biplot.

ACKNOWLEDGMENTS

The present study was financed the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Funding code 001.

REFERENCES

ARAÚJO, L. G. **Seleção e análise dos modelos PARAFAC e Tucker e gráfico *triplot* com aplicação em interação tripla**. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2009.

BERNARDO JÚNIOR, L. A. Y. *et al.* AMMI Bayesian models to study stability and adaptability in maize. **Agronomy Journal**, v. 110, n. 5, p. 1-12, 2018.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Cultivar Web**: gerenciamento de informação. 2019. Disponível em: http://extranet.agricultura.gov.br/php/snpc/cultivarweb/cultivares_registradas.php?txt_ordem=&postado=1&acao=pesquisar&. Acesso em: 5 jan. 2019.

COTES, J. M. *et al.* A Bayesian approach for assessing the stability of genotypes. **Crop Science**, v. 46, n. 6, p. 2654-2665, 2006.

CROSSA, J. *et al.* Bayesian estimation of the additive main effects and multiplicative interaction model. **Crop Science**, v. 51, n. 4, p. 1458-1469, 2011.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 4. ed. Viçosa, MG: Editora da UFV, 2012. 514 p. v. 1.

DAY, L. Proteins from land plants: potential resources for human nutrition's and food security. **Trends in Food Science & Technology**, v. 32, n. 1, p. 25-42, 2013.

DESTRO, D. *et al.* Food-type soybean cooking time: a review. **Crop Breeding and Applied Biotechnology**, v. 13, n. 3, p. 194-199, 2013.

FIGUEIREDO, A. G. *et al.* Application of mixed models for evaluating stability and adaptability of maize using unbalanced data. **Euphytica**, v. 202, p. 393-409, 2015.

FREIRIA, G. H. *et al.* Productivity and chemical composition of food-type soybeans sown on different dates. **Acta Scientiarum. Agronomy**, v. 38, n. 3, p. 371-377, 2016.

FREIRIA, G. H. *et al.* Protein, lipid and isoflavone contents of food-type soybean lines at two sowing dates. **Bioscience Journal**, v. 34, n. 6, p. 1540-1550, 2018a.

FREIRIA, G. H. *et al.* Statistical methods to study adaptability and stability in breeding lines of food-type soybeans. **Bragantia**, v. 77, n. 2, p. 253-264, 2018b.

GAUCH, H. G. A simple protocol for AMMI analysis of yield trials. **Crop Science**, v. 53, p. 1860-1869, 2013.

JARQUÍN, D. *et al.* A hierarchical Bayesian estimation model for multienvironment plant breending trials in successive years. **Crop Science**, v. 56, n. 5, p. 1-17, 2016.

MA, W. *et al.* Genistein as a neuroprotective antioxidant attenuates redox imbalance induced by β-amyloid peptides 25–35 in PC12 cells. **International Journal of Developmental Neuroscience**, v. 28, n. 4, p. 289-295, 2010.

MEOTTI, G. V. *et al.* Épocas de semeadura e desempenho agronômicos de cultivares de soja. **Pesquisa Agropecuária Brasileira**, v. 47, n. 1, p. 14-21, 2012.

OLIVEIRA, L. A. *et al.* Credible intervals for scores in the AMMI with radom effects for genotype. **Crop Science**, v. 55, n. 2, p. 1-12, 2015.

OLIVEIRA, L. A. *et al.* Performance of cowpea genotypes in the Brazilian Midwest using the Bayesian additive main effects and multiplicative interaction model. **Agronomy Journal**, v. 10, n. 1, p. 147-154, 2018.

PEREZ-ELIZALDE, S.; JARQUÍN, D.; CROSSA, J. A general Bayesian estimation method of linear-bilinear models applied to plant breeding trials genotype x environment interaction. **Journal of Agricultural, Biological and Environmental Statistics**, v. 17, n. 1, p. 15-37, 2012.

RAFTERY, A. E.; LEWIS, S. M. [Practical Markov Chain Monte Carlo]: comment: one long run with diagnostics: implementation strategies for Markov Chain Monte Carlo. **Statistical Science**, v. 7, n. 4, p. 494-497, 1992.

RIGO, A. A. *et al.* Characterization of soybean cultivars genetically improved for human consumption. **International Journal of Food Engineering**, v. 1, n. 1, p. 1-7, 2015.

SILVA, C. P. *et al.* A Bayesian shrinkage approach for AMMI models. **Plos one**, v. 10, n. 7, p. e0131414, 2015.

VAN EEUWIJK, F. A.; BUSTOS-KORTS, D. V.; MALOSETTI, M. M. What should students in plant breeding know about the statistical aspects of genotype x environment interactions? **Crop Science**, v. 56, n. 5, p. 2119-2140, 2016.

VIELE, K.; SRINIVASAN, C. Parsimoniuos estimation of multiplicative interaction in analysis of variance using Kullback-Leibler information. **Journal of Statistical Planning and Inference**, v. 84, n. 1/2, p. 201-219, 2000.

ZOBEL, R. W.; WRIGHT, M. J.; GAUCH, H. G. Statistical analysis of a yield trial. **Agronomy journal**, v. 80, n. 3, p. 388-393, 1988.