Environmental stratification and genotype recommendation toward the soybean ideotype: a Bayesian approach

Jeniffer Santana Pinto Coelho Evangelista[1], Marco Antônio Peixoto[1], Igor Ferreira Coelho[1], Rodrigo Silva Alves[2], Fabyano Fonseca e Silva[1], Marcos Deon Vilela de Resende[3], Felipe Lopes da Silva[1] and Leonardo Lopes Bhering[1*]

Abstract: *The genotype × environment (G×E) interaction plays an essential role in phenotypic expression and can lead to difficulties in genotypes recommendation. Thus, the objectives of this study were: i) propose the Multi-Environment Index Based on Factor Analysis and Ideotype-Design/Markov Chain Monte Carlo (FAI/MCMC index), and ii) apply it for soybean genotypes recommendation. To this end, a data set with 30 soybean genotypes evaluated in 10 environments for grain yield trait was used. Variance components, genetic parameters and genetic values were estimated through MCMC algorithm. Environmental stratification was conducted by factor analyses and the selection of soybean genotypes was performed using the FAI/MCMC index. The results indicated the existence of genotypic variability and G×E interaction. The environments were grouped into three factors. The predicted genetic gains from indirect selection was 4.81%. Thus, our results suggest that the FAI/MCMC index can be successfully used in soybean breeding.*

Keywords: *MCMC, Glycine max, genotype × environment interaction, mega-environments, selection index.*

INTRODUCTION

Soybean [*Glycine max* (L.) Merrill] is one of the most widely cultivated crops worldwide (Silva et al. 2017). Brazil, the United States, and Argentina account for 81.65% of global production, according to the US Department of Agriculture (2020). As the environment influences the gene expression of each genotype, genetic evaluation should be conducted in different environments, that is, under different edaphoclimatic and crop management conditions. Thus, analyses of the genotype × environment (G×E) interaction, adaptability, and stability are key requirements for progress in soybean production programs.

The G×E interaction occurs when the performance of genotypes varies throughout the environment (Resende 2015, Liu et al. 2017), implying different genotype rankings in different environments. In soybean crops, the G×E interaction is of marked importance. This is especially true for the grain yield trait, in which phenotypic expression is the result of the action of a large number of genes highly influenced by the environment (quantitative trait). In this context, it is necessary to use accurate methodologies for genetic evaluation to select and recommend genotypes with high adaptability, stability, and productivity.

*Corresponding author:
E-mail: leonardo.bhering@ufv.br
ORCID: 0000-0003-0564-7068

[1] Universidade Federal de Viçosa, Viçosa, MG, Brazil
[2] INCT Café/Universidade Federal de Viçosa, Viçosa, MG, Brazil
[3] Embrapa Café/Universidade Federal de Viçosa, Viçosa, MG, Brazil

According to Resende and Duarte (2007), the main method to detect the adverse effects of the G×E interaction is to identify genotypes with wide stability, adaptability, and productivity. Another technique to understand the adverse effects of the G×E interaction is to perform environmental stratification. This approach relies on grouping together environments in which complex G×E interactions do not occur. Environmental stratification is of great practical importance in breeding programs because it allows the breeder to establish an efficient experimental network, maximizing resource allocation and permitting a larger number of cultivars to be evaluated (Cruz et al. 2012).

Currently, the standard methodology used for the estimation of variance components and prediction of genetic values in annual crops is restricted maximum likelihood/best linear unbiased prediction (REML/BLUP), developed by Patterson and Thompson (1971) and Henderson (1975), respectively. However, compared with Bayesian inference, Fisherian (or likelihood) inference has the following drawbacks: i) it does not consider the distribution of the components of variance and genetic parameters, but it considers that of the data; ii) the desirable properties of the point and interval estimation are based on hypothetical repetitions of the experiment (i.e., an infinite number of repetitions); and iii) there is no possibility of incorporating prior information into the model (Volpato et al. 2019, Silva et al. 2020). All these factors can interfere with genetic evaluation.

To identify genotypes close to the ideotypes desired by the breeder and group correlated variables, Rocha et al. (2018) proposed a multi-trait index based on factor analysis and ideotype-design (FAI/BLUP index). This index takes into account the genetic correlation structure (exploratory factor analysis), allowing genotypes closest to those idealized to be identified. Several studies have demonstrated the efficiency and applicability of this index in plant breeding (Rocha et al. 2018, Rocha et al. 2019, Woyann et al. 2020). Based on the above, the aims of the present study were as follows: i) to propose a multi-environment index based on factor analysis and ideotype-design/Markov chain Monte Carlo (FAI/MCMC index) and ii) apply it to soybean genotype recommendation.

MATERIAL AND METHODS

Experimental data

Thirty soybean lines belonging to the 6-7 relative maturity group were evaluated in 10 municipalities (environments E1 to E10) (Table 1) during the 2013/2014 crop year. The experimental design was a randomized complete block design with three replicates. Each plot consisted of four 5-meter lines spaced 0.5 m apart and between plots. At maturity, the two central lines were harvested, totaling a usable area of 5 m². Grain yield was evaluated (kg ha^{-1}), with humidity corrected to 13% moisture. Management techniques followed recommendations for soybean cultivation at each site.

Table 1. Geographic coordinates and climatic characteristics of the 10 environments and climatic characteristics average of mega-environments 1 (ME1), 2 (ME2) and 3 (ME3)

Environment	Altitude (m)	Latitude	Longitude	Annual Precipitation (mm day^{-1})	Relative Humidity (%)	Temperature annual average (°C)
1	590	-22.9948	-51.1858	5.23	78.05	21.35
2	563	-24.6181	-53.3218	4.99	78.51	21.07
3	430	-22.2211	-54.8056	3.40	71.23	23.76
4	762	-24.2749	-52.5114	4.60	77.62	21.34
5	364	-23.0652	-54.1906	3.41	73.43	23.32
6	333	-24.2817	-53.8304	4.63	76.78	22.27
7	655	-22.5365	-55.7267	4.37	76.74	22.52
8	736	-23.3105	-51.3695	3.89	74.00	21.82
9	384	-23.0597	-51.0332	3.15	68.24	23.23
10	560	-24.7199	-53.7433	5.23	78.05	21.35
ME1	-	-	-	4.21	75.94	22.31
ME2	-	-	-	4.70	76.85	21.41
ME3	-	-	-	4.04	73.57	22.72

Statistical analyses

The Bayesian statistical model associated with the evaluation of genotypes in a randomized complete block design, with one observation per plot in several (10) environments, is given by the following equation:

$$y = Xf + Zg + Ti + e,$$

where y is the vector of phenotypic data $[y|f,g,i, \hat{\sigma}^2_g, \hat{\sigma}^2_i, \hat{\sigma}^2_e, \sim N (Xf + Zg + Ti, R \otimes I]$, $\hat{\sigma}^2_g$ is the genetic variance, $\hat{\sigma}^2_i$ is the G×E interaction variance, $\hat{\sigma}^2_e$ is the residual variance, and I is an identity matrix, f is the vector of the replication-environment combination that comprises the systematic effects of environment and replication within the environment (added to the overall mean), g is the vector of genotype effects $[g |I\hat{\sigma}^2_g, \sim N (0, I\hat{\sigma}^2_g)]$, i is the vector of $G \times E$ interaction effects $[i |I\hat{\sigma}^2_i, \sim N (0, I\hat{\sigma}^2_i)]$, and e is the vector of residuals $[e |I\hat{\sigma}^2_e, \sim N (0, I\hat{\sigma}^2_e)]$. The uppercase letters X, Z, and T refer to the incidence matrices for f, g, and i effects, respectively.

The posterior densities for the parameters followed the joint posterior distribution, as calculated using the following equation:

$$P(f, g, i, \hat{\sigma}^2_g, \hat{\sigma}^2_i, \hat{\sigma}^2_e |y) = P(y|f, g, i, \hat{\sigma}^2_g, \hat{\sigma}^2_i, \hat{\sigma}^2_e) \times P(f) \times P(g|\hat{\sigma}^2_g) \times P(i| \hat{\sigma}^2_i) \times P(\hat{\sigma}^2_g) \times P(\hat{\sigma}^2_i) \times P(\hat{\sigma}^2_e)$$

where $P(f, g, i, \hat{\sigma}^2_g, \hat{\sigma}^2_i, \hat{\sigma}^2_e |y)$ is the joint posterior distribution obtained by the multiplication of the likelihood function, $P(y|f, g, i, \hat{\sigma}^2_g, \hat{\sigma}^2_i, \hat{\sigma}^2_e)$, with the following prior distributions: $P(f)$, $P(g|\hat{\sigma}^2_g)$, $P(i| \hat{\sigma}^2_i)$, $P(\hat{\sigma}^2_g)$, $P(\hat{\sigma}^2_i)$, and $P(\hat{\sigma}^2_e)$. In the Bayesian statistical model, a uniform distribution was assumed for the systematic effects, normal distributions for the random effects, and scaled inverted Chi-squared distributions for the variance components. The number of iterations was 4.000.000, with an assumed burn-in period of 400.000 and sampling interval (thin) of 40 iterations, which provided 90.000 chains. The package boa (Smith 2007) was used to check the convergence according to Geweke (1992), Raftery and Lewis (Raftery and Lewis 1992)'s criteria.

The variance components, genetic and non-genetic parameters, genetic and high posterior probability (HPD) values were estimated by Gibbs sampling algorithm using Markov chain Monte Carlo (MCMC) methods, accounting for a vague prior, using the MCMCglmm package (Hadfield 2010) in R (R Development Core Team 2019). The degree of reliability parameter (nu) was defined as 0.02 (Hadfield 2010).

The random effects significance was tested by removing one effect of the full model at a time (genetic and G×E interaction) and calculating the deviance information criterion (DIC). The DIC was calculated using the following equation (Spiegelhalter et al. 2002):

$$DIC = D(\theta) + 2pD,$$

where $D(\theta)$ is a point estimate of the deviance obtained by replacing the parameters with their posterior mean estimates in the likelihood function, and pD is the effective number of parameters in the model. The lower the value of DIC, the better is the model adjustment.

The phenotypic variance ($\hat{\sigma}^2_{phen}$), heritability of individual plots in the broad sense (h^2_g), coefficient of determination of the G×E interaction effect (c^2_i), coefficient of determination of residual effects (c^2_{res}), genotypic correlation through environments (r_{genv}), and genotypic correlations between pairs of environments (r_{penv}) were obtained using the mean of the posterior distribution of the variance components. Contrastingly, the selective accuracy ($\tilde{r}_{\hat{g}g}$) was calculated using the standard errors of the estimated genetic values (Resende et al. 2014).

A correlation network was used to graphically express the genetic correlations between environments. In such networks, the proximity between the nodes (traces) is proportional to the absolute value of the correlation between these nodes (Fruchterman and Reingold 1991). The thickness of the edge was controlled by applying a cut-off value of 0.60; Thus, only the edges of $|r_{ij}| \geq 0.60$ were highlighted. This analysis was performed using the Rbio software (Bhering 2017).

Environmental stratification and genetic recommendation

The genetic values estimated by Bayesian inference were used to compose the FAI/MCMC index for environmental stratification and genetic recommendation. This index uses factor analysis, and the number of factors is determined based on the minimum eigenvalue (0.7) (Rocha et al. 2018). The factor analysis model used is given by the following equation:

$$X_j = I_{j1}F_1 + I_{j1}F_1 + ... + I_mF_m + \varepsilon_j$$

where X_j is the jth environment, with $j = 1,2,...,k$; I_{jk} is the factorial load for the jth environment associated with the kth factor ($k = 1,2,...,m$); F_k is the kth common factor; and ε is the specific factor. In this case, the number of factors was established such that the average proportion of variance of the environments explained by the common factors (commonality) reached approximately 80% (Rocha et al. 2018). The main goal of factor analysis is to describe the original variability of random vector *Y* (genetic values) in terms of a smaller number of random variables, called factors, which are initially unknown (Mingoti 2007).

To recommend genotypes based on the grain yield trait, considering all environments together, the FAI-MCMC index was used. This index uses factor analysis with the proposition of ideotypes to explore the covariance between variables, which in the current study refers to the grain yield. The FAI-MCMC index aims to select genotypes with the shortest distance from the ideotypes of interest, considering maximum productivity in all environments as an ideotype, using the following equation (Rocha et al. 2018):

$$P_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{i=1;j=1}^{i=n;j=m} \frac{1}{d_{ij}}},$$

where P_{ij} is the likelihood of the genotype $i(i = 1,2,...,n)$ being similar to the ideotype $j(j = 1,2,...,m)$, and *dij* is the mean standardized Euclidian distance between genotype *i* and ideotype *j*. Environmental stratification was performed using the routine developed by Rocha et al. (2018) in R (R Development Core Team 2020). Selection gains, considering the genotypes selected by the FAI-MCMC index, were obtained using the following equation (Evangelista et al. 2021):

$$SG = \frac{\sum_{i=1}^n GV_i}{p},$$

where GV_i is the genetic value of genotype *i* and *p* is the number of selected genotypes. A selection intensity of 20% was considered, indicating the selection of six genotypes.

RESULTS

For all variance components as well as genetic and non-genetic parameters, convergence of the chains was achieved according to the absolute values of the Z statistic (between -1.96 and 1.96) by Geweke's convergence criterion (Geweke 1992) and the values of the dependency factor (below 5) by Raftery and Lewis convergence criterion (Raftery and Lewis 1992) (Table 2). DIC values indicated that the best fit model was the complete model, indicating significance for the random effects of the model (genotype and G×E interaction effects). The HPD intervals, from the posterior densities of the genetic and non-genetic parameters, indicated significance for all parameters no interval contained the zero value.

The genotypic, G×E interaction, and residual effects explained 19%, 21%, and 60% of the phenotypic variance, respectively (Table 2). The genetic correlation across environments (r_{genv}) was 0.47, and the selective accuracy was

Table 2. Convergence diagnostic for the random effects of the model, based on Geweke and Raftery & Lewis criteria, model selection by deviance information criterion (DIC) and variance components estimates, with their respective HPD intervals, for grain yield trait evaluated in 30 soybean genotypes in ten environments

VC[a]	Convergence criterion		DIC	Estimate	R[2b]
	Geweke	Raftery & Lewis			
$\hat{\sigma}^2_g$	0.594	0.999	13582.86	46953.31 (20400.22 - 79209.96)	— 0.19
$\hat{\sigma}^2_i$	-0.265	1.014	13489.91	53100.32 (34435.53 - 71892.32)	0.21
$\hat{\sigma}^2_e$	-0.605	0.998	-	189872.28 (131517.69 - 165606.04)	0.60
$\hat{\sigma}^2_{phen}$	-	-	13460.57[#]	201828.273 (213782.90 - 285961.80)	-

$\hat{\sigma}^2_g$: genetic variance; $\hat{\sigma}^2_i$: genotype × environment interaction variance; $\hat{\sigma}^2_e$: residual variance; $\hat{\sigma}^2_{phen}$: phenotypic variance; [#]: selected model. [a] Variance component; [b] Coefficient of determination.
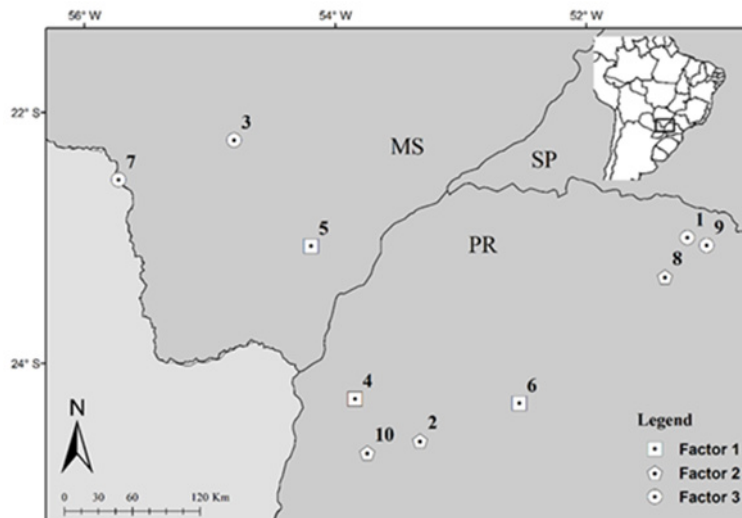
Figure 1. Environmental stratification, based on the factor analysis considering the ten environments (1-10). MS: Mato Grosso State; PR: Paraná State; SP: São Paulo State.

0.90. The environments were grouped into three factors, in which each factor represented a group of environments (Figure 1). Four of the ten environments were grouped in factor one (1, 3, 7, and 9), three environments were grouped in factor two (2, 8, and 10), and three environments were grouped in factor three (4, 5, and 6) (Table 3). The average commonality (i.e., the common variance explained by the factors) accounted for 84% of genetic variability present in the 30 soybean genotypes evaluated across the 10 environments.

Genetic correlations between pairs of environments were graphically expressed using correlation networks (Figure 2). The mean correlation values obtained in mega-environments 1 (ME1), 2 (ME2), and 3 (ME3) were 0.76, 0.64, and 0.73 respectively. The mean precipitation, relative humidity, and temperature for each mega-environment are presented in Table 1. It can be observed that ME2 presented higher precipitation and relative humidity, whereas ME3 presented the lowest values of these environmental characteristics.

The estimated genotypic values in each environment were used to design an ideotype for the grain yield trait. Considering a selection intensity of 20%, genotypes 5, 6, 9, 15, 21, and 28 were selected (Figure 3) as those with excellent adaptability, stability, and productivity. The predicted genetic gain from indirect selection (for all environments

Table 3. Environmental stratification by factor analysis, considering 30 soybean genotypes evaluated in 10 environments based on estimated genetic values; and predicted selection gain (SG), considering each environment individually and all environments, by FAI-MCMC index, with selection intensity of 20%

Environment	FA1	FA2	FA3	Communality	SG (kg ha^{-1})	SG (%)
E1	**-0.843**	0.317	-0.265	0.882 (0.07)	188.378	5.069
E2	-0.424	**0.58**	-0.574	0.847 (0.07)	215.337	5.816
E3	**-0.816**	0.185	-0.407	0.866 (0.07)	178.108	4.783
E4	-0.19	0.176	**-0.918**	0.909 (0.07)	70.976	1.912
E5	-0.435	0.3	**-0.749**	0.841 (0.07)	186.532	5.02
E6	-0.485	0.478	**-0.584**	0.806 (0.07)	330.125	8.878
E7	**-0.624**	0.189	-0.58	0.762 (0.07)	191.164	5.144
E8	-0.431	**0.639**	-0.288	0.700 (0.07)	133.311	3.596
E9	**-0.886**	0.231	-0.201	0.878 (0.07)	199.344	5.364
E10	-0.137	**0.932**	-0.163	0.914 (0.07)	92.771	2.498
Average	-	-	-	0.84	178.605	4.808

Bold numbers indicated the factor that the environments belong to. FA1: Factor 1; FA2: Factor 2; and FA3: Factor 3. The standard deviations are indicated between the parentheses.

simultaneously) was 4.81% (Table 3). Moreover, genetic gain was verified in all the environments with indirect selection.

DISCUSSION

The DIC criterion is widely used for model selection in Bayesian inference and is largely applied in plant breeding (Nascimento et al. 2020, Silva et al. 2020, Peixoto et al. 2021). The values of DIC and HPD indicate the significance of the genotype and G×E interaction effects on grain yield traits. Therefore, there is genetic variability between the genotypes evaluated, in addition to a G×E interaction. Resende (2015) found that a broad estimate of heritability presents a moderate magnitude ($0.15 < h_g^2 < 0.50$). This value represents the expected grain yield trait in soybean, once the trait assessed is considered a quantitative trait and is largely influenced by environmental effects (Soares et al. 2020).

In addition to the presence of G×E interaction, the observed genotypic correlations across environments showed moderate magnitude (0.47) (Resende and Alves 2020), and the coefficient of determination for G×E interaction was higher than 0.10. The experimental network analyzed was distributed in a tropical area with a marked latitudinal variation among the sites; hence, the existence of a complex G×E interaction was indicated. This result suggests the need for the development of efficient statistical methods for the selection and recommendation of soybean genotypes.

The results of the factor analyses indicated that the first three factors accounted for more than 84% of the variation in the data. This split the ten environments into three mega-environments. Each environment that contributed high values to each factor belonged to this factor. Furthermore, the groups of environments in each factor did not present complex G×E interactions. This led to the evaluation and recommendation of any of those genotypes to such areas and demonstrated that G×E interactions did not affect the ranking of the genotypes due to the environments inside the mega-environments.

Genetic correlations between pairs of environments were estimated to verify the consistency of the mega-environments. The environments of ME1 and ME3 showed strong correlations. Correlations of moderate magnitude were presented by environments of ME2. The magnitude of genetic correlations observed between pairs of environments in each mega-environment suggests that similar information about the genotypes can be retrieved with a smaller number of environments (Teodoro et al. 2018). This makes it possible to reduce the environments
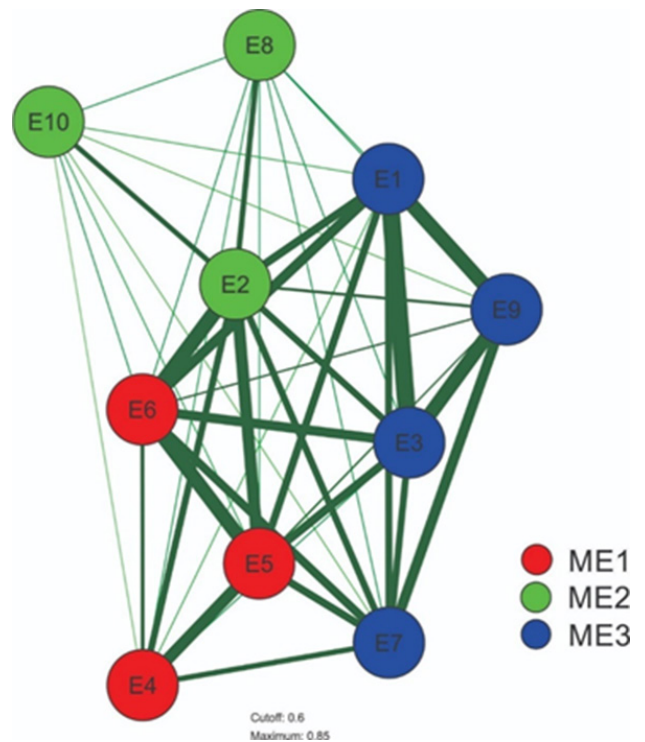

Figure 2. Correlations network among the evaluated environments. ME1, ME2 and ME3 represent the Factor 1, Factor 2, and Factor 3, respectively, grouped by the factor analysis; the thickness of the traces is proportional to the magnitude of the correlation.
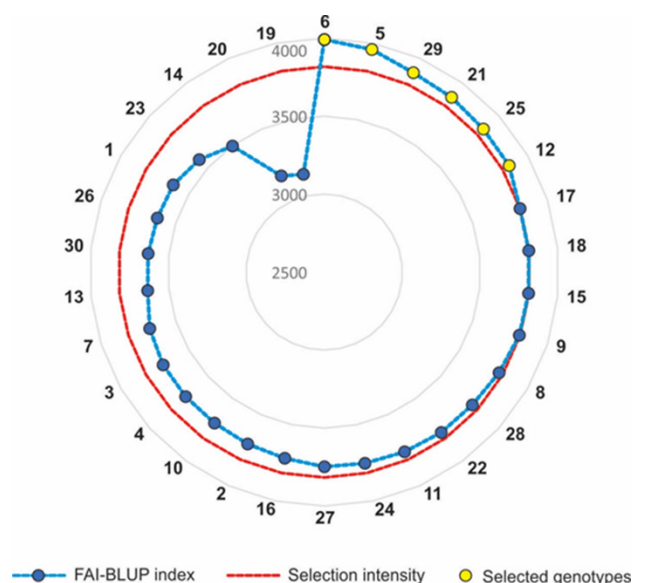

Figure 3. Genotype ranking, and the six genotypes selected by the FAI-MCMC index.

used as sites for the evaluation of soybean lines and reduce costs in breeding programs.

Bayesian inference requires that the estimated parameters must be based on mean posterior marginal distributions, such that any uncertainty about the parameters is considered (Wakefield 2013). This inference also allows estimation in samples of reduced finite size, and the Bayesian inference credibility or HPD intervals are constructed to qualify in terms of final probability, which is the valid probability for the observed sample rather than possible repetitions or hypothetical results, thus leading to more accurate interval estimates (Gianola and Fernando 1986, Ehlers and Brooks 2004). Accordingly, using Bayesian inference estimates in mega-environment delimitation allows the breeder to eliminate similar environments within each mega-environment, without losing efficiency or even precision in the selection process.

Selective accuracy is a parameter that estimates the reliability in the prediction of genetic values. This has the characteristic of informing the reliability of the inference from a genetic perspective, measuring the correlation between the estimated and real genotypic values. According to the criteria of Resende and Alves (2020), the selective accuracies estimated in this study were considered very high ($r_{\hat{g}g} \geq 0.90$), which is a favorable scenario for genotype recommendation. In recent years, increases in soybean yields have been reported (Gonçalves et al. 2020), and the main factor responsible for this result is the increased accuracy of evaluation methods and a better understanding of G×E interactions (Van Eeuwijk et al. 2016).

The FAI-MCMC index combines factor analysis with the proposition of ideotypes, exploring covariance between environments and enabling the use of genotypic values capitalized by G×E interaction effects (Rocha et al. 2018), thus leading to the selection of genotypes with greater adaptability, stability, and productivity. By selecting the six best genotypes using the FAI-MCMC index, gains were achieved in the desired direction for all environments evaluated, in addition to a considerable gain with indirect selection (for all environments simultaneously).

This index was also useful in selecting soybean genotypes, since it led to the selection of adaptable and stable genotypes that provided considerable predicted gains with selection for the grain yield trait. The use of the FAI-MCMC index in the recognition of mega environments and in the recommendation of genotypes with high stability, adaptability, and productivity in each mega-environment, as proposed in this study, demonstrates the potential to improve the recommendation of soybean genotypes.

CONCLUSIONS

The FAI-MCMC index was efficient and can be used for selecting adaptable, stable, and productive genotypes for mega-environments, achieving considerable gains for all environments. The selected genotypes that had excellent adaptability, stability, and productivity were 5, 6, 9, 15, and 21. Thus, the use of the FAI-MCMC index has the potential to improve genotype recommendations in soybean breeding.

ACKNOWLEDGEMENTS

REFERENCES

Bhering LL (2017) Rbio: A tool for biometric and statistical analysis using the R platform. **Crop Breeding and Applied Biotechnology 17**: 187-190.

Cruz CD, Regazzi AJ and Carneiro PCS (2012) **Modelos biométricos aplicados ao melhoramento genético**. UFV, Viçosa, 514p.

Ehlers RS and Brooks SP (2004) **Bayesian analysis of order uncertainty in ARIMA models**. Universidade Tecnológica Federal do Paraná, Paraná, 19p.

Evangelista JSPC, Alves RS, Peixoto MA, Resende MDV, Teodoro PE, Silva FL and Bhering LL (2021) Soybean productivity, stability, and adaptability through mixed model methodology. **Ciência Rural 51**: e20200406.

Fruchterman TMJ and Reingold EM (1991) Graph drawing by force-directed placement. **Software: Practice and Experience 21**: 1129-1164.

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. **Bayesian statistics 4**: 641-649.

Gianola D and Fernando RL (1986) Bayesian methods in animal breeding theory. **Journal of Animal Science 63**: 217-244.

Gonçalves GMC, Ferreira GRL, Lopes ACA and Vieira PFMJ (2020) Adaptability and yield stability of soybean genotypes by REML/BLUP and GGE Biplot. **Crop Breeding and Applied Biotechnology 20**: e282920217.

Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. **Journal of Statistical Software 33**: 1-22.

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. **Biometrics:** 423-447.

Liu X, Wu J, Ren H, Qi Y, Li C, Cao J, Zhang X, Zhang Z, Cai Z and Gai J (2017) Genetic variation of world soybean maturity date and geographic distribution of maturity groups. **Breeding Science 67**: 221-232.

Mingoti SA (2007) **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Editora UFMG, Belo Horizonte, 295p.

Nascimento M, Nascimento ACC, Silva FF, Teodoro PE, Azevedo CF, Oliveira TRA, Amaral Junior AT, Cruz CD, Farias FJC and Carvalho LP (2020) Bayesian segmented regression model for adaptability and stability evaluation of cotton genotypes. **Euphytica 216**: 30.

Patterson HD and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. **Biometrika 58**: 545-554.

Peixoto MA, Evangelista JSPC, Coelho IF, Alves RS, Laviola BG, Silva FF, Resende MD and Bhering LL (2021) Multiple-trait model through Bayesian inference applied to *Jatropha curcas* breeding for bioenergy. **PLOS ONE 16**: e0247775.

R Core Team (2019) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raftery AE and Lewis S (1992) How many iterations in the Gibbs sampler? **Bayesian Statistics 4**: 763-773.

Resende MDV (2015) **Genética quantitativa e de populações**. Suprema, Visconde do Rio Branco, 463p.

Resende MDV and Alves RS (2020) Linear, generalized, hierarchical, bayesian and random regression mixed models in genetics/genomics in plant breeding. **Functional Plant Breeding Journal 2**: 1-31.

Resende MDV and Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical 37**: 182-194.

Resende MDV, Silva FF and Azevedo CF (2014) **Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-QWAS, estatística espacial e temporal, competição, sobrevivência**. Suprema, Visconde do Rio Branco, 881p.

Rocha JRAS, Machado JC and Carneiro PCS (2018) Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. **GCB Bioenergy 10**: 52-60.

Rocha JRAS, Nunes KV, Carneiro ALN, Marçal TS, Salvador FV, Carneiro PCS and Carneiro JES (2019) Selection of superior inbred progenies toward the common bean ideotype. **Agronomy Journal 111**: 1181-1189.

Silva FA, Viana AP, Corrêa CCG, Carvalho BM, Sousa CMB, Amaral BD, Ambrósio M and Glória LS (2020) Impact of bayesian inference on the selection of *Psidium guajava*. **Scientific Reports 10**: 1-10.

Silva FL, Borém A, Sediyama T and Ludke WH (2017) **Soybean breeding**. Springer International Publishing, Brazil, 563p.

Smith BJ (2007) boa: an R package for MCMC output convergence assessment and posterior inference. **Journal of Statistical Software 21**: 1-37.

Spiegelhalter DJ, Best NG, Carlin BP and van der Linde A (2002) Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64**: 583-639.

Soares IO, Bianchi MC, Bruzi AT, Gesteira G de S, Silva KB, Guilherme SR and Cianzio SR (2020) Genetic and phenotypic parameters associated with soybean progenies in a recurrent selection program. **Crop Breeding and Applied Biotechnology 20**: e28092046.

Teodoro PE, Farias FJC, Carvalho LP, Nascimento M, Peixoto LA, Cruz CD and Bhering LL (2018) Identification of optimal environments for cotton cultivars in the Brazilian Cerrado. **Agronomy Journal 110**: 1226-1232.

US Departament of Agriculture (2020) Oilseeds: World markets and trade. Available at https://www.usda.gov/. Accessed on February 10, 2020.

Van Eeuwijk FA, Bustos-Korts DV and Malosetti M (2016) What should students in plant breeding know about the statistical aspects of genotype × Environment interactions? **Crop Science 56**: 2119-2140.

Volpato L, Alves RS, Teodoro PE, De Resende MDV, Nascimento M, Nascimento ACC, Ludke WH, Silva FL and Borém A (2019) Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. **PLOS ONE 14**: 1-22.

Wakefield J (2013) **Bayesian and frequentist regression methods**. Springer, New York, 700p.

Woyann LG, Meira D, Matei G, Zdziarski AD, Dallacorte LV, Madella LA and Benin G (2020) Selection indexes based on linear-bilinear models applied to soybean breeding. **Agronomy Journal 112**: 175-182.