Received: 9 July 2021 | Accepted: 4 December 2021 | Published online: 21 March 2022

DOI: 10.1002/csc2.20696

Crop Science

ORIGINAL RESEARCH ARTICLE

Crop Breeding & Genetics

Bayesian GGE model for heteroscedastic multienvironmental trials

Luciano Antonio de Oliveira[1] | **Carlos Pereira da Silva**[2] | **Alessandra Querino da Silva**[1] | **Cristian Tiago Erazo Mendes**[2] | **Joel Jorge Nuvunga**[3] | **José Airton Rodrigues Nunes**[4] | **Rafael Augusto da Costa Parrella**[5] | **Marcio Baleste**[2] | **Júlio Sílvio de Sousa Bueno Filho**[2]

[1]Faculty of Exact Sciences and Technology, Univ. Federal da Grande Dourados, Dourados, Mato Grosso do Sul CP 364, Brazil

[2]Dep. of Statistics, Institute of Exact Sciences, Univ. Federal de Lavras, Lavras, Minas Gerais CP 3037, Brazil

[3]School of Business and Entrepreneuership Chibuto, Univ. Eduardo Mondlane, Maputo, Gaza, Moçambique

[4]Biology Dep., Institute of Natural Sciences, Univ. Federal de Lavras, Lavras, Minas Gerais CP 3037, Brazil

[5]Embrapa Maize and Sorghum, Sete Lagoas, Minas Gerais, Brazil

Correspondence
Luciano Antonio de Oliveira, Faculty of Exact Sciences and Technology, Univ. Federal da Grande Dourados, Dourados, Mato Grosso do Sul, Brazil, CP 364.
Email: lucianoantonio@ufgd.edu.br

Assigned to Associate Editor Paulino Pérez-Rodríguez.

Abstract

The dissection of genotype × environment interaction (GEI) is a crucial aspect of the final stages of plant breeding pipelines and recommendation of cultivars. Linear-bilinear models used to analyze this interaction, such as the additive main effects and multiplicative interaction (AMMI) and genotype plus GEI (GGE), often assume homogeneity of the residual variances across environments which affects the estimates and therefore, interpretations and conclusions. Our main objective was to propose a GGE model that considers heteroscedasticity across environments using Bayesian inference and to evaluate its implications in the interpretation of real and simulated data. The GGE model assuming common variance was also fitted for comparison purposes. The great flexibility of the Bayesian inference is transferred to the biplots, allowing the construction of credible regions for genotypic and environmental scores. The inference on the stability and adaptability of genotypes might change when heteroscedasticity is ignored. When real data are used, different patterns of correlations between environments also affect the representativeness and discrimination of the target environment. The modeling of heteroscedasticity allowed the clustering of environments into subgroups, with similar effects for GEI. The proposed GGE model was more adequate and realistic to deal with scenarios of heterogeneous variance in multienvironment trials, which can be useful for exploiting the GEI.

Abbreviations: AE, average environment; AEA, axis of the average environment; AEC, average environment coordination; AMMI, additive main effects and multiplicative interaction model; BGGE, Bayesian GGE model under homogeneity of residual variances across sites; BGGEH, Bayesian GGE model under heterogeneity of residual variances across sites; FA, Factor-Analytic Model; G, Genotype; GEI, Genotype × environment interaction; GGE, Genotype main effect + genotype × environment interaction; HPD, Highest posterior density; IG, ideal genotype; MCMC, Markov chain Monte Carlo; MET, Multienvironmental trial; MLM, mixed linear model; SS_{GGE}, total sum of squares for G + GEI; SVD, Singular value decomposition.

1 | INTRODUCTION

Multienvironment trials (MET) are conducted over several years. They are used to estimate the yield of cultivars in a set of locations (or environments) of interest, with the aim of selecting and recommending superior genotypes. The differential response of cultivars across environments is referred to as the genotype × environment interaction (GEI). The crossover GEI makes the plant breeder's job difficult because it causes changes in the ranking of genotypes from one environment to another. As a result, assessments and recommendations cannot be performed in general. In this sense, different methodologies were developed to analyze MET data, seeking to identify highly adaptable genotypes or those that adapt in a specific way to take advantage of the positive effects of GEI for regionalized recommendations.

Multiplicative (or linear-bilinear) models are useful for analyzing data organized in two-way tables and have gained great popularity in the study of GEI. The models that include the main effect of genotypes (G) plus GEI, referred to in the literature as GGE (Xu et al., 2014; Yan et al., 2000) and the additive main effects and multiplicative interaction (AMMI; Gauch et al., 2008) deserve special mention due to the wide applicability by researchers and breeders in the analysis of MET data.

The GGE model and other linear-bilinear models combine the study of adaptability and stability in a single approach, providing more possibilities for analysis compared with regression methods, such as those suggested by Finlay and Wilkinson (1963) and Eberhart and Russell (1966). Moreover, bilinear term selection methods allow obtaining more parsimonious models because of the GEI (AMMI) or G + GEI (GGE) matrices are obtained by singular value decomposition (SVD) and approximated by lower dimensional matrices, although they still describe the pattern obtained by the original matrices.

By simplifying data in a few dimensions, the patterns of genotype responses across environments can be graphically visualized using biplots (Gabriel, 1971). The graphical representation of genotypic and environmental scores of the first main axes has become a standard procedure when using multiplicative models. Its interpretation has been the basis for important decisions on the analyzed data. The GGE biplot model, a variant of GGE, when using only the first two singular axes, has become a popular method for evaluating the responses of cultivars across environments (Yan et al., 2000; Yang et al., 2009).

Conventional implementations of AMMI and GGE have many limitations associated with the standard method of estimating fixed effects. They do not consider heteroscedasticity, although theoretically, there are procedures based on SVD for the same purpose (Rodrigues et al., 2014; Yan, 2014); however, such procedures are adjusted in two stages and may lead to loss of information (Gogel et al., 2018; Romão et al., 2019).

Core Ideas

- The GGE model is useful for studying genotype responses across environments.
- Heterogeneity of residual variances across environments occurs routinely in MET trials.
- To assume homogeneity of the residual variances across environments affects the interpretations.
- Bayesian modeling offers wide flexibility to model complex variance-covariance structures.
- The Bayesian GGE model brings promising perspectives for MET data analysis.

Mixed linear models offer greater flexibility when compared to fixed-effects modeling, and this class of models has been well developed (Oakey et al., 2006; Smith et al., 2005; Smith & Cullis, 2018). Mixed models are versatile in dealing with the heterogeneity of variances, incomplete data, and the spatial correlation within and between environments (Crossa, 2012; Smith et al., 2015).

Versions of mixed linear-bilinear models for AMMI and GGE have been proposed and result from assuming a factor-analytic structure for the genetic variance-covariance matrix and, for this reason, they are referred to as Factor-Analytic Models (FA; Piepho, 1997, 1998; Piepho & Mohring, 2006; Smith et al., 2001). In these models, it is unclear how parametric confidence regions can be constructed for biplot points, and approximate alternatives have been proposed, as highlighted by Crossa et al. (2011). Some of these problems were overcome by Nuvunga et al. (2019), who applied Bayesian FA.

Bayesian modeling offers practical advantages and wide flexibility to model complex variance-covariance structures and better exploit interactions. The first studies on Bayesian inference applied to the AMMI model were conducted by Viele and Srinivasan (2000) and Liu (2001), who showed how the sampling process could be conducted correctly, using the Markov Chain Monte Carlo (MCMC) method, mainly for the parameters that describe GEI, whose support for a posteriori distribution is not trivial.

Crossa et al. (2011) and Perez-Elizalde et al. (2012) empirically demonstrated the flexibility of the Bayesian model to incorporate uncertainty into the AMMI model fitted with two first principal components biplot through bivariate credibility regions built for genotypic and environmental scores, as well as using information from previous experiments incorporated through prior distributions. Other contributions to this method have been recently published (da Silva et al., 2015, 2019; de

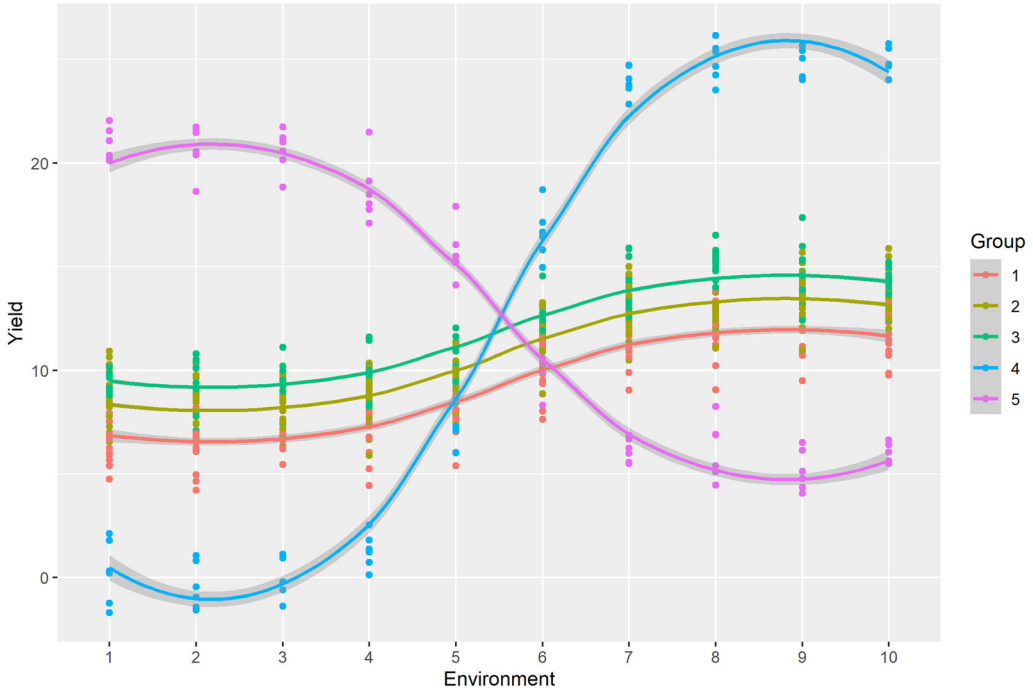

FIGURE 1 Average yields of genotype subgroups, relative to the simulated genotype × environment interaction pattern, in each environment

Oliveira et al., 2015, 2016; Jarquin et al., 2016; Romão et al., 2019).

Most recent studies involving the application of Bayesian inference to multiplicative models only use the AMMI model, and research and applications with the Bayesian version of GGE are still scarce in the literature. Jarquin et al. (2016), de Oliveira et al. (2016), Oliveira et al. (2021), and Omer and Singh (2017) are exceptions and addressed aspects related to the Bayesian GGE model; however, specific prior distributions have not been explicitly assumed for residual variance components at these locations. The implications and importance of such assumptions in modeling the main effects of genotypes and GEI effect, as well as details of their implementations in the Bayesian AMMI analysis, were presented by da Silva et al. (2019).

The GGE model has a strong graphic appeal and presents a plausible biological explanation for the use of the first two main axes in the biplot analysis (Yan et al., 2000); moreover, there is the potential for a more precise estimation of the singular values, and the genotypes can be evaluated with different experimental precisions in each environment.

Thus, the main objectives of this study were to (a) propose a Bayesian GGE model to deal with heteroscedasticity in MET and (b) verify the implications of this model on the interrelationship of genotypes and environments in biplot representations using simulated and real data, when compared with the homoscedastic model.

2 | MATERIALS AND METHODS

2.1 | Simulated data

The environmental effects were initially simulated proportional to the normal density for quantiles derived from environmental deviations equally spaced between −4.5 and 4.5. The effects are produced as $5p_j - 1$, where p_j is the normal density for the jth quantile. For the effect of GEI, three distinct genotype response patterns were simulated: (a) stable genotypes (lower, G1 to G4; median, G5 to G12; and higher, G13 to G16), (b) unstable genotypes with positive environmental reinforcement (G17 and G18), and (c) unstable genotypes with negative environmental reinforcement (G19 and G20). The matrix of the simulated data was corrected for the average of the rows and columns. The main effects of genotypes were obtained through a Gaussian N(0, 3) distribution and ordered, in ascending order, from G1 to G20.

The average response patterns of each of the subgroups (in relation to the simulated GEI pattern) highlighted above, in each environment, added to a common constant ($\mu = 10$), are shown in Figure 1.

To emulate the homogeneous variance, we considered $\sigma_e^2 = 1$. To emulate the heterogeneous variance, the diagonal of the covariance matrix was as follows: $diag\{V\} = [10.5, 0.5, 3.0, 2.0, 3.0, 5.0, 9.0, 9.0, 15.0, 17.0]$. The data were simulated using a multivariate normal distribution.

TABLE 1 Geographic characteristics of the locations where the sorghum trials were carried out in the 2014–2015 crop season

Environments	Lat	Lon	ALT[a]	PD	HD
			m		
E1 - Dourados-MS	22°13'S	54°48'W	430	11 Nov. 2014	13 May 2015
E2 - Dracena-SP	21°28'S	51°31'W	421	15 Nov. 2014	28 Apr. 2015
E3 - Goiânia-GO	16°40'S	49°15'W	823	18 Dec. 2014	23 May 2015
E4 - Guaíra-SP	24°04'S	54°15'W	220	26 Nov. 2014	14 Apr. 2015
E5 - Lavras-MG	21°14'S	45°00'W	919	22 Nov. 2014	12 May 2015
E6-Nova Porterinha-MG	15°48'S	43°18'W	85	22 Nov. 2014	20 May 2015
E7 - Pelotas-RS	31°46'S	52°20'W	7	6 Dec. 2014	25 Mar. 2015
E8 - Sete Lagoas-MG	19°27'S	44°14'W	761	6 Nov. 2014	12 May 2015
E9 - Sinop-MT	11°50'S	55°38'W	384	4 Dec. 2014	11 May 2015
E10 - Uberlândia-MG	18°55'S	48°16'W	863	6 Dec. 2014	24 May 2015

Lat, latitude; Lon, longitude;[a]ALT, altitude; PD, planting dates; HD, harvest dates.

2.2 | Real data

The real dataset comes from a network of cultivation and use value trials coordinated by Embrapa Maize and Sorghum with 36 sorghum genotypes; of these, 33 are experimental hybrids developed by the biomass sorghum breeding program of Embrapa Maize and Sorghum, and the remaining genotypes are control cultivars. The genotypes were evaluated during the 2014–2015 crop season in 10 different locations in Brazil. The geographic characteristics and cultivation aspects of these locations are listed in Table 1.

The experiments were arranged in a 6 × 6 triple lattice design. The plots consisted of four rows of 5.0 m in length, spaced 0.7 m apart, considering only the two central rows as useful. The trait assessed was the dry mass yield of the plot, which was converted into tonne ha^{-1}. The crop management is described in detail by Delgado et al. (2019).

2.3 | Statistical model

The GGE model, in matrix notation, is given by

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

Where $\mathbf{y}_{n\times 1}$ is the vector of phenotypic data with $n = r \times c \times b$, where r is the number of genotypes, c is the number of locations (or environments), and b is the number of blocks or replications within each location; $\boldsymbol{\beta}_{l\times 1}$ where $l = c \times b$ is the vector of fixed effects of blocks within environments; λ_k is the kth singular value, with $k = 1, \ldots, t$. The $t = min\ (r - 1, c)$ is the rank of the interaction matrix ($\mathbf{GGE}_{r\times c}$) with errors of nonadditivity to the main effects of environments; $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$, of dimensions $r \times 1$ and $c \times 1$, respectively, are the kth singular vectors of genotypes and environments, respectively, related to the kth principal component.

In addition, $\mathbf{X}_{1(n\times l)}$, $\mathbf{X}_{2(n\times c)}$, and $\mathbf{Z}_{n\times c}$ are the design matrices associated with $\boldsymbol{\beta}$, $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$, respectively, and $\boldsymbol{\varepsilon}_{n\times 1}$ is the vector of random errors in which $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V})$, where $\mathbf{0}_{n\times 1}$ is the null vector and $\mathbf{V}_{n\times n}$ is a block diagonal matrix of dimension $n \times n$ composed of $\sigma^2_{e_j}$, with $j = 1,\ldots, c$. The bilinear part of the model is subject to order restrictions $\lambda_k \geq \lambda_{k+1}$ and the orthonormality of singular vectors, that is, $\boldsymbol{\alpha}_k^\top \boldsymbol{\alpha}_k = \boldsymbol{\gamma}_k^\top \boldsymbol{\gamma}_k = 1$ and $\boldsymbol{\alpha}_k^\top \boldsymbol{\alpha}_{k'} = \boldsymbol{\gamma}_k^\top \boldsymbol{\gamma}_{k'} = 0; k \neq k'$ for $k = 1, \ldots, t$. The distribution of $\mathbf{y}$, conditioned to the parametric vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda_k, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \mathbf{V})$, is multivariate normal denoted by $\mathbf{y} \mid \boldsymbol{\theta} \sim N(\boldsymbol{\mu}_y, \mathbf{V})$, with $\boldsymbol{\mu}_y = \mathbf{X_1}\boldsymbol{\beta} + \sum_{k=1}^{t} \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X_2}\boldsymbol{\gamma}_k$.

2.4 | Prior distributions for model parameters

The prior distributions assumed for the parameters $\boldsymbol{\beta}$, λ_k, $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$ of the model are the same as those described by de Oliveira et al. (2016) under the homogeneity of variances:

$$\boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \sigma^2_\beta \sim N\left(\mu_\beta, \mathbf{I}\sigma^2_\beta\right),\ \boldsymbol{\mu}_\beta = 0 \text{ and } \sigma^2_\beta = 10^8$$

$$\lambda_k|\mu_{\lambda_k}, \sigma^2_{\lambda_k} \sim N^+\left(\mu_{\lambda_k}, \sigma^2_{\lambda_k}\right),\ \mu_{\lambda_k} = 0 \text{ and } \sigma^2_{\lambda_k} = 10^8;$$

$\boldsymbol{\alpha}_k \sim$ spherical uniform on the correct subspace and $\boldsymbol{\gamma}_k \sim$ spherical uniform on the correct subspace.

For the residual variance component associated with environment j, with $j = (1, \ldots, c)$, a Jeffreys priori is assigned, that is, $p(\sigma^2_{e_j}) \propto 1/\sigma^2_{e_j}$.

2.5 | Full Conditional Posterior distributions

The likelihood function for $\boldsymbol{\theta}$ according to model (1) is given by

$$L(\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{V}|^{\frac{n}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_y)^{\top}\boldsymbol{V}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}_y\right)\right] \tag{2}$$

Connecting likelihood with a priori information (via Bayes' theorem), a posteriori joint distribution is obtained, which can be written as:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto L(\boldsymbol{\theta}|\mathbf{y})p\left(\boldsymbol{\beta}|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}}\right) \times$$

$$\times \left[\prod_{k=1}^{t} p(\lambda_k|\mu_{\lambda_k}, \sigma^2_{\lambda_k})p\left(\boldsymbol{\alpha}_k\right)p\left(\boldsymbol{\gamma}_k\right)\right] \prod_{j=1}^{c} p\left(\sigma^2_{e_j}\right) \tag{3}$$

The full conditional posterior distributions for the parameters of the model in Equation 1 are obtained by algebraic manipulations from the joint posteriori distribution in Equation 3, under the assumptions considered a priori, and are presented in Supplemental Appendix S1.

Owing to the orthogonality constraint ($\boldsymbol{\alpha}_k^{\top}\boldsymbol{\alpha}_{k'} = \boldsymbol{\gamma}_k^{\top}\boldsymbol{\gamma}_{k'} = 0$ for $k \neq k'$), the vectors $\boldsymbol{\alpha}_k$ (genotype scores) should be distributed only in a restricted subspace of the unitary sphere in R^r and $\boldsymbol{\gamma}_k$ (environmental scores) should be distributed in a restricted subspace of the unitary sphere in R^c. They must be orthogonal to s ($0 \leq s \leq t-1$) the directions indicated by the vectors in dimension p. The sampling of singular vectors is performed in the corrected subspace ($p - s$), with the correct support, by the definition of auxiliary variables $\tilde{\boldsymbol{\alpha}}_k = \mathbf{H}_k^{\top}\boldsymbol{\alpha}_k$ and $\tilde{\boldsymbol{\gamma}}_k = \mathbf{D}_k^{\top}\boldsymbol{\gamma}_k$, where $\mathbf{H}_k$ and $\mathbf{D}_k$ are orthogonal linear transformation matrices. The complete a posteriori conditional distributions for the singular vectors in the corrected subspace and algebraic details are presented in Supplemental Appendix S1.

The a posteriori conditional hypotheses and distributions described above refer to the Bayesian modeling of GGE under heterogeneity of residual variances across locations and this model will be referred to throughout the text as BGGEH.

For the purpose of comparison, the GGE model that assumes a single variance in all locations (BGGE) was also adjusted. For this analysis, the sample distribution is $\mathbf{y} \mid \boldsymbol{\theta} \sim N(\boldsymbol{\mu}_{\mathbf{y}}, \mathbf{I}\sigma^2_e)$, where $\mathbf{I}$ is the identity matrix of order n. The a priori densities attributed to the parameters are basically the same as those used for the heterogeneous case. The exception, as already pointed out, is the hypothesis of the variance structure. In BGGE, a residual variance-covariance matrix is a variance component structure in which a priori is $p(\sigma^2_e) \propto 1/\sigma^2_e$. Under this assumption, the conditional posterior densities are different and are extensively described by de Oliveira et al. (2016).

2.6 | Sampling and inference for linear and bilinear parameters

The implemented sampling process was the Gibbs sampler, which is a procedure that belongs to the class of MCMC method. The algorithm, analogous to that described by da Silva et al. (2019), is exemplified in Supplemental Appendix S1.

For Markov chains, values were discarded to eliminate the initial effect (*burn-in*). In addition, the autocorrelation function of the chains was analyzed by adopting fixed *thinning* intervals to select a sample approximately unrelated to the parameters. These corrections were based on training samples according to the Raftery and Lewis (1992) criteria. The convergence of the MCMC chains was monitored by the methods of Heidel-Berger and Welch (1983), Raftery and Lewis (1992), and Geweke (1992).

The estimates and intervals for $\boldsymbol{\beta}$, λ_k, and σ^2_e were obtained by the mean, maximum a posteriori (MAP) and median of the simulated MCMC samples. Highest posterior density intervals (HPD), at the 95% credibility level, were constructed using the method of Chen and Shao (1999). Estimates of $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$, were obtained by the method used by Liu (2001).

Bivariate credibility regions (95%) for genotypic $(\lambda_1^{1/2}\alpha_{i1}, \lambda_1^{1/2}\alpha_{i2})$ and environmental $(\lambda_1^{1/2}\gamma_{j1}, \lambda_1^{1/2}\gamma_{j2})$ scores, with $i = 1, \ldots, r$ and $j = 1, \ldots, c$, were incorporated into the biplot using the method of Hu and Yang (2013).

3 | RESULTS

The posterior distributions for all models showed good properties, as assessed by diagnostic tests. Trace plots of the chains of the residual variance components for the real data are presented in Supplemental Figures S1–S4. Random oscillations around a value are a good indication of the posterior sampling. This pattern was observed for all parameters in all scenarios. In Supplemental Table S4 (Supplemental Appendix S2), values obtained by applying the convergence tests for these same parameters are presented.

3.1 | Analysis of simulated data

The parametric value of error variance was covered by HPD regions from both models in homoscedastic simulated scenario (Table 2). Although greater precision is declared for the BGGE homoscedastic model, it is noteworthy that the BGGEH model also addressed the homoscedastic structure of errors in the locations.

The two Bayesian versions of the GGE model resulted in similar estimates for singular values, with the principal

TABLE 2 Summaries of the posterior densities of the residual variance components from Bayesian genotype main effect + genotype × environment interaction model under heterogeneity of residual variances across sites (BGGEH) and Bayesian genotype main effect + genotype × environment interaction model under homogeneity of residual variances across sites (BGGE) models for simulated data considering the homoscedastic scenario

						HPD intervals at 95% credibility	
Model	**PAR**[a]	**MAP**	**MED**	**M**	**SD**	**LL**	**UL**
BGGE	σ^2_e	1.02	1.04	1.04	0.07	0.91	1.17
BGGEH	$\sigma^2_{e_1}$	1.03	1.05	1.08	0.23	0.68	1.55
	$\sigma^2_{e_2}$	0.92	0.98	1.01	0.22	0.62	1.46
	$\sigma^2_{e_3}$	0.75	0.79	0.82	0.18	0.50	1.20
	$\sigma^2_{e_4}$	1.03	1.06	1.09	0.24	0.67	1.58
	$\sigma^2_{e_5}$	1.01	1.06	1.08	0.23	0.68	1.53
	$\sigma^2_{e_6}$	1.03	1.10	1.13	0.25	0.71	1.64
	$\sigma^2_{e_7}$	1.12	1.15	1.19	0.25	0.73	1.68
	$\sigma^2_{e_8}$	0.88	0.91	0.94	0.21	0.56	1.37
	$\sigma^2_{e_9}$	0.98	1.09	1.12	0.24	0.70	1.60
	$\sigma^2_{e_{10}}$	0.65	0.70	0.73	0.18	0.42	1.08

[a]PAR, parameter; MAP, maximum a posteriori; MED, median; M, mean; SD, standard deviation; HPD, highest posterity density; LL, lower limit; UL, upper limit.

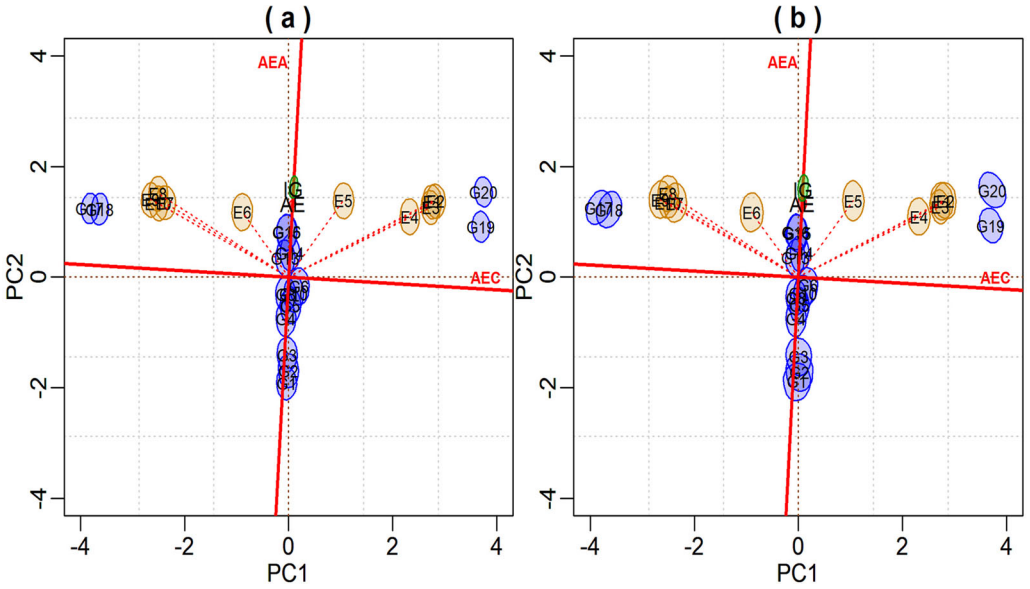

FIGURE 2 Genotype main effect + genotype × environment interaction (GGE) biplots with 95% bivariate credibility regions for genotypic (G), environmental (E), average environment (AE), and ideal genotype (IG) scores in the homoscedastic scenario Bayesian GGE model under homogeneity of residual variances across sites (a) and Bayesian GGE model under heterogeneity of residual variances across sites (b). AEA, axis of average environment; PC, principal component

components capturing, practically, the same information as shown in Supplemental Table S1.

In Figure 2, biplots are presented with their bivariate credibility regions for genotypic and environmental scores, this graphical configuration displays the "Mean vs. Instability" form of the GGE biplot (Yan, 2001). In each biplot, a secondary system of axes comprising the axis of the average environment (AEA) and its ordinate (AEC) was inserted. The average environment (AE) and the ideal genotype (IG), defined to have the maximum yield among environments, were also represented with their regions of credibility. The direction of the semi-straight line in the AEA (from the origin to the IG) indicates higher values of the average yield for genotypes. As for the AEC, the further away a point is from this axis (either in the positive or negative direction), the greater the contribution of the respective genotype (or environment) to the interaction. Environmental vectors were inserted to assist in visualizing the correlations between environments. The norms of these vectors are measures of the environment's ability to discriminate between genotypes.

TABLE 3 Point and interval summaries of a posteriori densities of the residual variance components of the Bayesian genotype main effect + genotype × environment interaction model under heterogeneity of residual variances across sites (BGGEH) and Bayesian genotype main effect + genotype × environment interaction model under homogeneity of residual variances across sites (BGGE) models for simulated data considering the heteroscedastic scenario

							HPD[a] intervals at 95% credibility	
PAR	**SPV**	**Model**	**MAP**	**MED**	**M**	**SD**	**LL**	**UL**
σ^2_e	–	**BGGE**	7.53	7.62	7.64	0.54	6.62	8.69
$\sigma^2_{e_1}$	10.50	**BGGEH**	10.41	10.75	11.02	2.22	7.19	15.35
$\sigma^2_{e_2}$	0.50		0.57	0.64	0.68	0.55	0.35	1.09
$\sigma^2_{e_3}$	3.00		2.20	2.36	2.44	1.01	1.52	3.49
$\sigma^2_{e_4}$	2.00		1.98	2.17	2.24	0.60	1.37	3.25
$\sigma^2_{e_5}$	3.00		3.21	3.37	3.46	0.81	2.19	4.89
$\sigma^2_{e_6}$	5.00		5.81	6.10	6.26	1.58	4.01	8.90
$\sigma^2_{e_7}$	9.00		12.86	13.78	14.15	3.01	8.98	20.30
$\sigma^2_{e_8}$	9.00		12.24	12.79	13.16	2.99	7.61	18.99
$\sigma^2_{e_9}$	15.00		19.29	20.56	21.10	4.48	13.01	29.77
$\sigma^2_{e_{10}}$	17.00		15.22	16.61	17.08	3.81	10.16	24.54

[a]HPD, highest posterior density; PAR, parameter; SPV, simulated parametric value; MAP, maximum a posteriori; MED, median; M, mean; SD, standard deviation; LL, lower limit; UL, upper limit.

The smaller the angle between a given environment, and the average environment, the more representative it will be in the target population of environments. Genotypes, in turn, are compared (in the average environment) according to their Euclidean distance to the ideal genotype. These concepts, in specific terms, are described by Yan and Kang (2003) and Yan (2014). Only genotypes and environments whose bivariate regions do not include the origin are represented to simplify visualization and interpretation.

Biplots exhibited almost the same pattern for both GGE models (Figure 2). The amplitudes are slightly higher for regions of credibility in the BGGEH biplot, especially for genotypes farther away from the origin (Figure 2). However, these differences in amplitudes did not affect the interpretation in terms of adaptability and stability. It is noteworthy that the BGGEH model was able to address the simulated homoscedastic scenario.

The estimates of the residual variance components at the locations obtained from the BGGEH model were similar to the parametric values, considering the heteroscedastic scenario (Table 3). Conversely, the estimate of the common residual variance by the BGGE model is included only in the credible intervals of the components associated with locations Environments 1, 6, and 8 (Table 3). In addition, none of the simulated values were included in the HPD interval of the common variance. These results indicate that the BGGEH model was efficient in capturing the present heterogeneity structure and the assumption of a common variance would not be consistent.

The point and interval estimate for singular values presented in Supplemental Table S2 indicate that the first principal component of the BGGEH explained approximately 97% of the sum of squares for genotype + GEI (SS_{GGE}), which was significantly higher than the percentage explained by the BGGE model (86.14%). In contrast, the second axis of the BGGEH model explained only 2%, with 11% explained by the second principal component of the BGGE model. The SS_{GGE} is obtained by summing the squares of the singular values for the full interaction model ($\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_t^2$) and the partial recovery of the SS_{GGE}, in each model, is obtained by the sum of squares of the singular values accumulated in each dimension k ($k = 1, \ldots t$). The model sum of squares recovered by the BGGE is greater (3,537.98) than that obtained by the BGGEH (2,983.55) and this difference is more accentuated when compared to the analysis with homogeneous data. This may be an indication that the interaction is being overestimated in the BGGE model.

The biplots shown in Figure 3a and 3b illustrate the heteroscedastic scenario. As it is possible to perceive the basic pattern is maintained in the two biplots, but different experimental precision is shown in the BGGEH biplot, capturing the residual variance heterogeneity simulated. Furthermore, differences can be observed, such as the greater distance from the average environment and ideal genotype in relation to the origin in the BGGE model (Figure 3a), and a more pronounced rotation of the secondary system of axes in the BGGEH biplot (Figure 3b). It is noteworthy that these and other changes indicate that the assumptions on the structure of the residual variance are important in the analysis and affect the interpretations in the biplot. For these simulated data, they affect more the analysis in average environment.

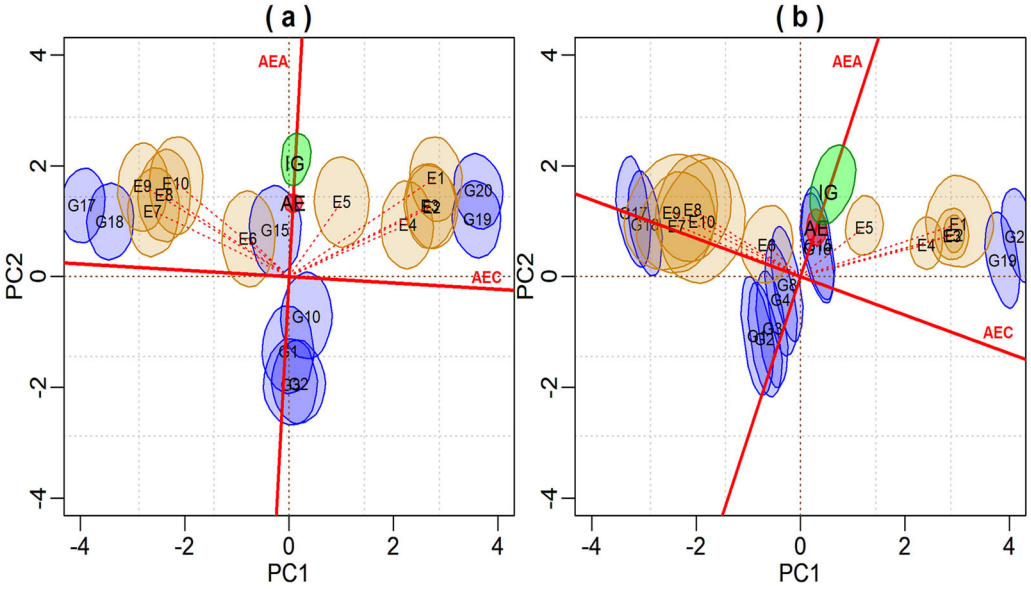

FIGURE 3 Biplot genotype main effect + genotype × environment interaction (GGE) with 95% credibility bivariate regions for genotypic (G), environmental (E), average environment (AE), and ideal genotype (IG) scores in a heteroscedastic scenario Bayesian GGE model under homogeneity of residual variances across sites (a) and Bayesian GGE model under heterogeneity of residual variances across sites (b). AEA, axis of average environment; PC, principal component

For the BGGE model, the genotype closest to the ideal was Genotype 15 (Figure 3). This genotype can be considered stable because its credibility region is intercepted by the AEA axis and presents a yield higher than the general mean, since the region of bivariate credibility is above the AEC (average yield is not likely). The Genotypes 17 and 18 could also be considered stable. For the BGGEH model, Genotypes 15 and 16 are the genotypes with the shortest distance to the ideal, but their regions of credibility are intercepted by the AEC axis, which indicates that the general mean of the yields is probable, which also occurs for Genotypes 17 and 18. These interpretations differ from those based on the BGGE biplot. This figure also shows that the simulated stable genotypes respond positively to environmental improvement (which corroborates the information in Figure 1). In general, this information was better identified using the BGGEH model.

Biplots showing the "who won where" pattern are shown in Supplemental Figures S5 and S6. Although the sectors have differentiated between biplots for heterogeneous data (Supplemental Figure S6), there were no strong implications in terms of mega-environments when comparing the two models, which is of great importance when generating polygons in the biplot.

3.2 | Analysis of real data

The common variance estimate of the BGGE model is included in credible intervals of only three components, which are those related to Environments 2, 5, and 10. For other environments, the HPD intervals for the components did not overlap with the common variance (Table 4).

Estimates of residual variances at each location obtained by adjusting the mixed linear model (MLM) using the lme4 package are included in the Bayesian intervals estimated for the respective environments, according to the BGGEH (Table 4). This indicates similarities between the BGGEH and MLM analyses. It was also observed that in nine of the 10 environments, the BGGEH model presented estimate of the residual variance smaller than the MLM analysis. Although we do not want to rigorously compare the joint analysis with the individual analysis (in each environment), this result suggests a better adjustment for the BGGEH model.

The first two principal components explained 93.09 and 91.41% of the SS_{GGE} for BGGE and BGGEH, respectively (Supplemental Table S3). However, the a posteriori estimate of the percentage explained by the first BGGEH component was slightly higher. Smaller magnitudes and more expressive shrinkage are also noted for a posteriori estimates of the first three singular values in the BGGEH model; from then on, the opposite behavior is observed, with an ever more marked shrinkage effect for BGGE. The SS_{GGE} for BGGEH was lower than that for BGGE as in the simulated scenario.

The results of the application of the information criteria for model selection (Akaike, 1974; Raftery et al. 2007; Schwarz, 1978) are shown in Figure 4. In general, the BGGEH models obtained significantly lower values for both criteria. In relation to the winning models (in each GGE version) and with two bilinear terms, BGGEH also achieved remarkably better

TABLE 4 Estimates obtained by mixed linear model and point and interval summaries of posterior densities for residual variance components of the BGGEH and BGGE models applied to the real MET of sorghum for the total dry biomass in ton/ha

							HPD[a] intervals at 95% credibility	
PAR	MLM	Model	MAP	MED	M	SD	LL	UL
σ^2_e	–	**BGGE**	141.17	141.90	142.19	7.42	128.69	157.60
$\sigma^2_{e_1}$	372.59	**BGGEH**	420.46	434.53	440.75	70.84	304.47	578.74
$\sigma^2_{e_2}$	123.84		122.15	126.80	129.20	22.56	87.34	172.53
$\sigma^2_{e_3}$	88.92		81.33	83.78	85.70	15.54	58.47	117.88
$\sigma^2_{e_4}$	303.02		262.76	266.35	270.37	40.80	197.38	351.28
$\sigma^2_{e_5}$	121.65		118.24	121.75	123.89	20.91	86.89	166.49
$\sigma^2_{e_6}$	60.28		53.55	55.20	56.24	9.61	39.17	74.90
$\sigma^2_{e_7}$	45.37		37.74	39.91	40.54	6.94	28.71	54.12
$\sigma^2_{e_8}$	89.51		85.23	85.51	87.00	15.11	59.21	116.42
$\sigma^2_{e_9}$	67.70		59.41	63.21	64.3	11.11	44.52	86.60
$\sigma^2_{e_{10}}$	150.96		126.42	131.57	133.93	21.12	97.14	176.78

[a]HPD, highest posterior density; PAR, parameter; MLM, individual analyses using the mixed linear model; MAP, maximum a posteriori; MED, median; M, mean; SD, standard deviation; LL, lower limit; UL, upper limit.

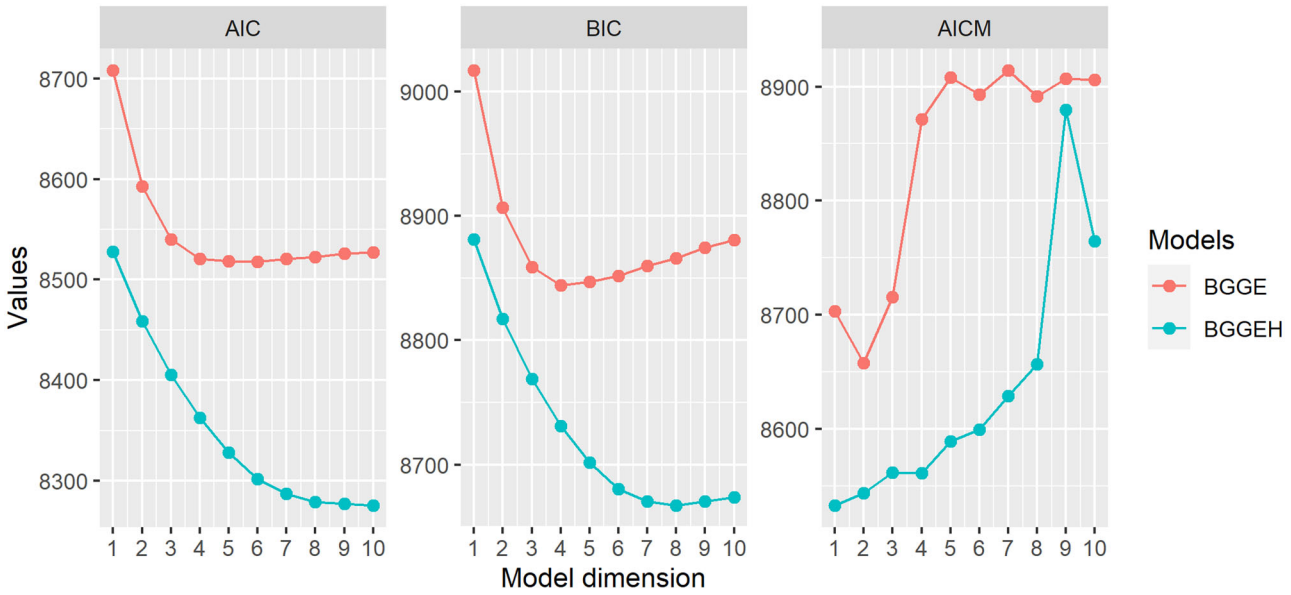

FIGURE 4 Results of the information criteria used to select models. AIC, Akaike information criterion; BIC, Bayesian information criterion; AICM, Akaike Monte Carlo information criterion; BGGE, Bayesian genotype main effect + genotype × environment interaction model under homogeneity of residual variances across sites; BGGEH, Bayesian genotype main effect + genotype × environment interaction model under heterogeneity of residual variances across sites

performance (with emphasis on Akaike Monte Carlo Information Criterion). These results suggest that the heteroscedastic model was better adjusted to the data and was more adequate than under assumption of homogeneity.

The biplots for the analysis considering real data are shown in Figure 5, and different experimental precisions are also addressed by the BGGEH, with emphasis on Environments 1 and 8. It is observed that Environment 1 is shrunk towards the average interaction in BGGEH but stands out as the environment with the greatest ability for discrimination of genotypes in BGGE, being considered as the most important environment for interaction. The opposite occurs with Environment 8, which has an apparently small contribution to the interaction in the BGGE and due to the relatively small variance, it appears more distant from the middle axis in the BGGEH.

Other changes are visible, such as that the average environment and the ideal genotype are further removed from the origin in BGGE (Figure 5), which also presents a more

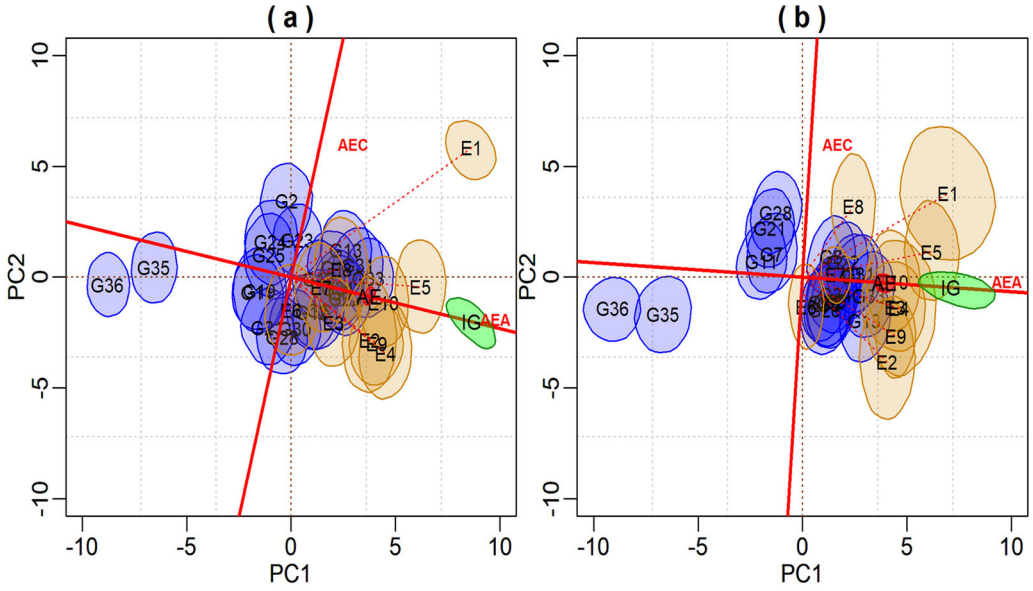

FIGURE 5 Genotype main effect + genotype × environment interaction (GGE) biplots with 95% credibility bivariate regions for genotypic scores (G), environmental scores (E), average environment (AE), and ideal genotype (IG); (a) Bayesian GGE model under homogeneity of residual variances across sites biplot and (b) Bayesian GGE model under heterogeneity of residual variances across sites biplot. AEC, average environment coordination; PC, principal component

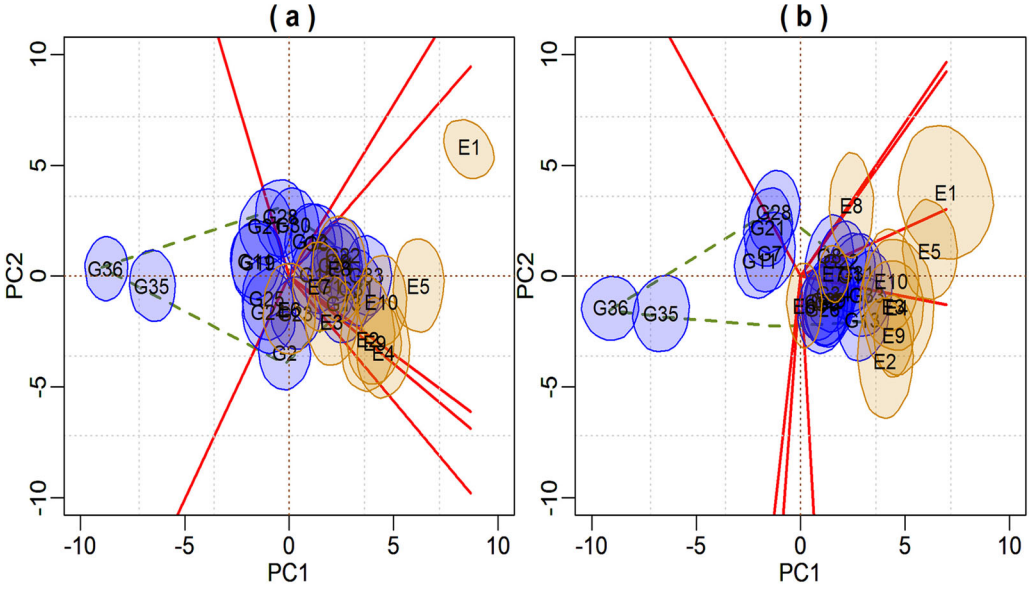

FIGURE 6 Genotype main effect + genotype × environment interaction (GGE) biplots with 95% credibility bivariate regions for genotypic and environmental scores showing "the standard who won where"; (a) Bayesian GGE model under homogeneity of residual variances across sites biplot and (b) Bayesian GGE model under heterogeneity of residual variances across sites biplot. PC, principal component

accentuated rotation of the secondary axis system (AEA × AEC). There is also a greater distance between the average environment and the IG in the BGGE analysis, in which more genotypes were interpreted as unstable. On the other hand, a greater subdivision of environments is observed in similar subgroups in the BGGEH biplot, which can be observed more clearly using the graphic presentation "who won where" (Figure 6). This fact can also be seen in Supplemental Figure S7, in which genotypic, polygon and semi-straight regions are removed to facilitate interpretation.

The slight differences in these biplots indicate that ignoring the presence of heteroscedasticity can compromise interpretations, mistakenly suggesting a simplification of the GEI structure. Better visualization of the pattern of environments

and genotypes can also be obtained in biplots built with a posteriori means (Supplemental Figure S8). When removing the regions of credibility, the differences in rotations of average scores are more clearly observed, which can affect interpretations and possible inferences.

The GGE biplots for fixed effects models are also shown in the supplemental material for comparison purposes. In Supplemental Figure S9a, the GGE biplot, obtained without any preliminary scaling procedure for the genotype + GEI matrix (scale or weight), exhibits a pattern similar to that observed for the BGGE model. Conversely, the GGE biplot (Supplemental Figure S9b) was obtained from the standardization of the cell means by the phenotypic standard deviation in each column, as well as by weighting the square root of heritability in each environment. The biplot from the "corrected" data exhibits a pattern more similar to that presented to the BGGEH. Yan (2014) emphasizes the need for this preliminary procedure (scale, weighting) to deal with heterogeneity in the biplot analysis, which is yet another argument in favor of the method proposed here.

4 | DISCUSSION

Heterogeneity of residual variances across test sites occurs routinely in MET trials and can lead to inefficient estimates if it is not addressed in the analysis (Crossa et al., 2006; Smith et al., 2001). This fact was confirmed here by the results of the information criteria applied to the real data and showed that methods for modeling specific variances in environments would be more appropriate.

The implications of the differences in the models can be seen in the biplots for the simulated and real data. It is observed in the BGGEH biplot that in an environment with greater variance, there will be greater shrinkage of interactions in relation to the general average, and in environments with less variance, the interaction will be more preserved. This was evident in Environments 1 and 8 of the real data (Figures 4, 5; Supplemental Figure S3). Such findings lead us to believe that the estimate in the BGGE model may overestimate the interaction and could lead to the belief that the model with homogeneous variances is superior (greater interaction variance), but the heterogeneous model is clearly more suitable.

Rotational changes in the system of additional axes and different precision, reflected in different ranges of the intervals for the environmental scores (Figures 2, 4, 5; Supplemental Figures S1, S2, S3), directly affect the interpretations of adaptability and stability. Graphical changes similar to those observed here (due to the structure of variances) were reported by Rodrigues et al. (2014) using the weighted AMMI, by da Silva et al. (2019) who modeled heterogeneity of variances in the AMMI-Bayesian model, and by Nuvunga et al. (2015), who compared the standard of the GGE model and the FA version of the sites regression model with two-factor fitted (FA-SREG2).

The method proposed in this study is an extension of the modeling by da Silva et al. (2019) to the GGE model to verify the impact of inferences on the bilinear parameters and, consequently, on the GGE biplot graphical representation. As already noted, studies addressing the GGE model from a Bayesian perspective (de Oliveira et al., 2016; Jarquín et al., 2016; Omer & Singh, 2017) did not consider heteroscedastic scenarios across locations.

Heteroscedasticity is disregarded in most applications involving the standard procedure of estimation in linear-bilinear models with fixed effects (such as AMMI and GGE). It is worth noting that it is a possible to model the phenomenon using algorithms based on SVD, such as the weighted AMMI proposed by Rodrigues et al. (2014). In the GGE biplot, Yan (2014) recommends the use of data dimensioning methods as a preliminary step to obtain the graphic representation. For the author, the essence of the process is to review the scale and weigh the data from different environments, with appropriate weights in the joint analysis. These weights are expressed as the lengths of the environmental vectors in the GGE biplot.

These procedures are, however, controversial because lower weights are attributed to environments that have a higher residual mean square. A disadvantage of this method is that the weights can be correlated to the responses in the environment. Thus, environments with high yields may present greater error variance and environments with low yields present reduced error variances which may mask the true performance of some genotypes in certain environments (da Silva et al., 2019; Crossa, 1990); moreover, the equations presented for fixed multiplicative models are valid only for ideal situations, that is, when complete and balanced data are available. Although there are procedures for imputing missing data based on the SVD (Yan, 2013), in nonorthogonal and unbalanced scenarios, components of variance and derived measures such as heritabilities are best estimated from direct modeling methods, such as mixed models (Kelly et al., 2007; Spilke et al., 2005).

Mixed models have become popular for MET analysis, especially the class of mixed multiplicative models (MMM), also referred to as the analytic factorial (FA) model. They are identified as the best among the mixed model classes for analyzing cultivar tests (Kelly et al., 2007). The FA models have been considered superior to fixed effects models in several aspects, including regular biplot analyses, especially the GGE biplot. They stand out for their ample flexibility to manipulate missing data, heterogeneity among environments, genotypes, and correlations within environments, in addition to being able to accommodate spatial dependence between environments (Beeck et al., 2010; Crossa, 2012; Cullis et al., 2010; Smith et al., 2001). Despite these advantages, MMM is

computationally demanding and eventually estimates parameters outside the parametric space, which is referred to as Heyhood cases. Although there are several alternatives to overcome these limitations, including the sparse matrix algorithm or modified Average Information by Thompson et al. (2003) and parameter expanded expectation–maximization algorithm (Diffey et al., 2017), its modeling remains a challenge.

Conversely, the Bayesian method has been shown to be promising for estimating parameters related to stability and genotypic adaptability in MET data. Nuvunga et al. (2019) showed that under the Bayesian approach, full-dimension models can be considered in the FA analysis without requiring axis rotation. Another advantage is that the estimates are restricted to the parametric space. The Bayesian method also allows the incorporation of more complex structures for the variance-covariance matrices (Cotes et al., 2006; Edwards & Jannink, 2006; Orellana et al., 2014). In our approach, the components of variance are considered directly in a single-stage analysis, only establishing hypotheses and a priori information. In addition, all dimensions of the GGE can be estimated, overcoming the limitations of the weighted approach as proposed by Rodrigues et al. (2014) in which, for weighted AMMI, the number of components must be decided before the algorithm is executed.

It is worth mentioning that the graphic analysis implemented here is an extension of the inference about the joint a posteriori distribution. Such an inference is not feasible in fixed effects models and is not usually presented in FA models. The Bayesian version of the FA-SERG (Bayesian Analytical Factor model) by Nuvunga et al. (2019), similar to our proposal, is also flexible and allows the incorporation of bivariate credible regions in the biplot. However, the authors themselves recommend caution regarding the interpretation, which must be performed separately for genotypes and environments. Their responses do not have the same scale and do not have the properties of the internal product. The incorporation of uncertainty about biplot scores, in the frequentist context, on the other hand, has been controversial in the literature (Crossa et al., 2011; Yan et al., 2010; Yang et al., 2009).

The BGGEH was better adjusted to the data for all the model selection criteria used. As pointed out by Piepho (1998), the usefulness of any measure of stability primarily depends on how well the model fits the data. This means that the choice of the candidate model class and the estimate of the variance component have direct implications for the stability measure (Hu et al., 2014). Thus, the BGGEH model would be a better alternative to represent the stability of genotypes in MET trials than its version that considers homogeneity of variances across environments.

The impacts of heteroscedasticity modeling on point estimates and inference on genotype effects have been recognized and presented in several studies (Casanoves et al., 2005; Hu & Spilke, 2011; Hu et al., 2013), with the classification of the genotype yield stability based on the estimates of the parameters of the model used. Thus, a reliable estimate of the model's parameters is also an important prerequisite for valid inferences regarding the analysis of genotype stability (Hu et al., 2014). Here, we show that assuming homogeneity of variances in a situation of clear deviation from this assumption, far from being a harmless practice, can lead to misinterpretations and influence the evaluation of genotype stability in MET trials. This was particularly striking in the analysis of real data.

The simulated genotype + GEI matrix with a low dimension and contrasting subgroups compared to the GEI effect may have influenced some similarity between the BGGE and BGGEH biplots, as the first two principal components explained practically the entire pattern in the data. In particular, the most interesting pattern is explained by the first axis; however, this simple example was sufficient to show the main differences between the models in terms of biplot analysis and information retrieved by singular values.

A slightly more complex interaction pattern can be observed for the real data, although the first two axes also have captured most of the variation. For this specific example, the effect of shrinkage is more clearly observed in the estimates of the first three singular values of BGGEH (Supplemental Table S3). Thereafter, however, the opposite behavior is observed, that is, a lower decay of the higher order estimates values of BGGEH model compared to BGGE. Thus, we expect that for other examples (and with more complex interaction pattern) higher order BGGEH models would be needed to better explain the interaction. This is consistent with da Silva et al. (2019) for the Bayesian AMMI model with heterogeneity model. From the aspects observed here, we believe that modeling the heterogeneous variances brings promising perspectives for multi-environmental data analysis.

ACKNOWLEDGMENTS

We would like to thank for financial support by the Universidade Federal da Grande Dourados (Edital PROPP No. 02/2021 – Programa de Apoio à Pesquisa–PAP/UFGD) and the Universidade Federal de Lavras. This Sweet Sorghum MET was funded by the Banco Nacional de Desenvolvimento Econômico e Social (BNDES) through the "Sorgo-Energia" Project coordinated by Embrapa Maize and Sorghum. The authors would also like to thank brazilian funding agency FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) for using equipment funded by process number APQ-01439-09, and the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financing part of this study.

AUTHOR CONTRIBUTIONS

Luciano Antonio de Oliveira: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Resources; Supervision; Validation; Visualization; Writing-original draft; Writing-review & editing. Alessandra Querino da Silva: Investigation; Methodology; Supervision; Visualization; Writing-original draft; Writing-review & editing. Joel Jorge Nuvunga: Conceptualization; Investigation; Methodology; Writing-review & editing. Rafael Augusto da Costa Parrella: Conceptualization; Writing-review & editing. Carlos Pereira da Silva: Data curation, Formal analysis, Software, Validation, Visualization, Writing-review & editing. Cristian Tiago Erazo Mendes: Conceptualization, Investigation, Visualization, Writing-review & editing. José Airton Rodrigues Nunes: Conceptualization, Investigation, Writing-review & editing. Marcio Baleste: Conceptualization, Investigation, Software. Júlio Sílvio de Sousa Bueno Filho: Conceptualization, Methodology, Supervision, Writing-original draft.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Luciano Antonio de Oliveira https://orcid.org/0000-0001-5054-2623
Carlos Pereira da Silva https://orcid.org/0000-0002-8053-1703
Alessandra Querino da Silva https://orcid.org/0000-0002-1702-0169
Cristian Tiago Erazo Mendes https://orcid.org/0000-0003-3486-0212
Joel Jorge Nuvunga https://orcid.org/0000-0001-6593-4136
José Airton Rodrigues Nunes https://orcid.org/0000-0002-6260-7890
Rafael Augusto da Costa Parrella https://orcid.org/0000-0001-6599-7487
Marcio Baleste https://orcid.org/0000-0001-8855-9674
Júlio Sílvio de Sousa Bueno Filho https://orcid.org/0000-0002-3243-0959

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Beeck, C. P., Cowling, W. A., Smith, A. B., & Cullis, B. R. (2010). Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. *Genome*, *53*(11), 992–1001. https://doi.org/10.1139/G10-051

Casanoves, F., Macchiavelli, R., & Balzarini, M. (2005). Error variation in multienvironment peanut trials: Within-trial spatial correlation and between-trial heterogeneity. *Crop Science*, *45*(5), 1927–1933. https://doi.org/10.2135/cropsci2004.0547

Chen, M. H., & Shao, Q. M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, *8*(1), 69–92. https://doi.org/10.2307/1390921

Cotes, J. M., Crossa, J., Sanches, A., & Cornelius, P. L. (2006). A Bayesian approach for assessing the stability of genotypes. *Crop Science*, *46*(6), 2654–2665. https://doi.org/10.2135/cropsci2006.04.0227

Crossa, J. (1990). Statistical analyses of multilocation trials. *Advances in Agronomy*, *44*, 55–85. https://doi.org/10.1016/S0065-2113(08)60818-4

Crossa, J. (2012). From genotype x environment interaction to gene x environment interaction. *Current Genomics*, *13*(3), 225–244. https://doi.org/10.2174/138920212800543066

Crossa, J., Burgueño, J., Cornelius, P. L., McLaren, G., Trethowan, R., & Krishnamachari, A. (2006). Modeling genotype× environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. *Crop science*, *46*(4), 1722–1733. https://doi.org/10.2135/cropsci2005.11-0427

Crossa, J., Perez-Elizalde, S., Jarquin, D., Cotes, J. M., Viele, K., Liu, G., & Cornelius, P. L. (2011). Bayesian estimation of the additive main effects and multiplicative interaction model. *Crop Science*, *51*(4), 1458–1469. https://doi.org/10.2135/cropsci2010.06.0343

Cullis, B. R., Smith, A. B., Beeck, C. P., & Cowling, W. A. (2010). Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. *Genome*, *53*(11), 1002–1016. https://doi.org/10.1139/G10-080

da Silva, C. P., de Oliveira, L. A., Nuvunga, J. J., Pamplona, A. K. A., & Balestre, M. (2015). A Bayesian shrinkage approach for AMMI models. *PLOS ONE*, *10*(7), e0131414. https://doi.org/10.1371/journal.pone.0131414

da Silva, C. P., de Oliveira, L. A., Nuvunga, J. J., Pamplona, A. K. A., & Balestre, M. (2019). Heterogeneity of variances in the Bayesian AMMI model for multienvironment trial studies. *Crop Science*, *59*(6), 2455–2472. https://doi.org/10.2135/cropsci2018.10.0641

de Oliveira, L. A., da Silva, C. P., Nuvunga, J. J., da Silva, A. Q., & Balestre, M. (2015). Credible intervals for scores in the AMMI with random effects for genotype. *Crop Science*, *55*(2), 465–476. https://doi.org/10.2135/cropsci2014.05.0369

de Oliveira, L. A., da Silva, C. P., Nuvunga, J. J., da Silva, A. Q., & Balestre, M. (2016). Bayesian GGE biplot models applied to maize multi-environments trials. *Genetics and Molecular Research [Electronic Resource]*, *15*(2). https://doi.org/10.4238/gmr.15028612

Delgado, I. D., Gonçalves, F. M. A., Parrella, R. A. D. C., Castro, F. M. R. D., & Nunes, J. A. R. (2019). Genotype by environment interaction and adaptability of photoperiod-sensitive biomass sorghum hybrids. *Bragantia*, *78*, 509–521. https://doi.org/10.1590/1678-4499.20190028

Diffey, S. M., Smith, A. B., Welsh, A. H., & Cullis, B. R. (2017). A new REML (parameter expanded) EM algorithm for linear mixed models. *Australian & New Zealand Journal of Statistics*, *59*(4), 433–448. https://doi.org/10.1111/anzs.12208

Eberhart, S. T., & Russell, W. A. (1966). Stability parameters for comparing varieties 1. *Crop science*, *6*(1), 36–40. https://doi.org/10.2135/cropsci1966.0011183X000600010011x

Edwards, J. W., & Jannink, J. L. (2006). Bayesian modeling of heterogeneous error and genotype× environment interaction variances. *Crop Science*, *46*(2), 820–833. https://doi.org/10.2135/cropsci2005.0164

Finlay, K. W., & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, *14*(6), 742–754. https://doi.org/10.1071/AR9630742

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*(3), 453–467. https://doi.org/10.1093/biomet/58.3.453

Gauch Jr, H. G., Piepho, H. P., & Annicchiarico, P. (2008). Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop science*, *48*(3), 866–889. https://doi.org/10.2135/cropsci2007.09.0513

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo,; J. O. Berger,; A. P. Dawid,;, & A. F. M. Smith (Eds). *Bayesian Statistics*, 4. New York: Oxford University Press, pp. 625–631.

Gogel, B., Smith, A., & Cullis, B. (2018). Comparison of a one-and two-stage mixed model analysis of Australia's National Variety Trial Southern Region wheat data. *Euphytica*, *214*(2), 1–21. https://doi.org/10.1007/s10681-018-2116-4

Heidelberger, P., & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, *31*(6), 1109–1144. https://doi.org/10.1287/opre.31.6.1109

Hu, X., & Spilke, J. (2011). Variance–covariance structure and its influence on variety assessment in regional crop trials. *Field Crops Research*, *120*(1), 1–8. https://doi.org/10.1016/j.fcr.2010.09.015

Hu, Z., & Yang, R. C. (2013). Improved statistical inference for graphical description and interpretation of genotype× environment interaction. *Crop Science*, *53*(6), 2400–2410. https://doi.org/10.2135/cropsci2013.04.0218

Hu, X., Yan, S., & Li, S. (2014). The influence of error variance variation on analysis of genotype stability in multi-environment trials. *Field Crops Research*, *156*, 84–90. https://doi.org/10.1016/j.fcr.2013.11.001

Hu, X., Yan, S., & Shen, K. (2013). Heterogeneity of error variance and its influence on genotype comparison in multi-location trials. *Field Crops Research*, *149*, 322–328. https://doi.org/10.1016/j.fcr.2013.05.011

Jarquín, D., Pérez-Elizalde, S., Burgueño, J., & Crossa, J. (2016). A hierarchical Bayesian estimation model for multienvironment plant breeding trials in successive years. *Crop science*, *56*, 2260–2276. https://doi.org/10.2135/cropci2015.08.0475

Kelly, A. M., Smith, A. B., Eccleston, J. A., & Cullis, B. R. (2007). The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Science*, *47*(3), 1063–1070. https://doi.org/10.2135/cropsci2006.08.0540

Liu, G. (2001). *Bayesian computations for general linear-bilinear models* (Doctoral dissertation). University of Kentucky.

Nuvunga, J. J., Oliveira, L. A., Pamplona, A. K. A., Silva, C. P., Lima, R. R., & Balestre, M. (2015). Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. *Genetics and Molecular Research*, *14*(4), 14262–14278. https://doi.org/10.4238/2015.November.13.10

Nuvunga, J. J., da Silva, C. P., de Oliveira, L. A., de Lima, R. R., & Balestre, M. (2019). Bayesian factor analytic model: An approach in multiple environment trials. *PLOS ONE*, *14*(8), e0220290. https://doi.org/10.1371/journal.pone.0220290

Oakey, H., Verbyla, A., Pitchford, W., Cullis, B., & Kuchel, H. (2006). Joint modeling of additive and non-additive genetic line effects in single field trials. *Theoretical and Applied Genetics*, *113*(5), 809–819. https://doi.org/10.1007/s00122-006-0333-z

Oliveira, L. A., Silva, C. P., Silva, A. Q., Mendes, C. T. E., Nuvunga, J. J., Muniz, J. A., Bueno Filho, J. S. S., & Balestre, M. (2021). Shrinkage in the Bayesian analysis of the GGE model: A case study with simulation. *PLOS ONE*, *16*(8), e0256882. https://doi.org/10.1371/journal.pone.0256882

Omer, S. O., & Singh, M. (2017). Comparing Bayesian and Frequentist Approaches for GGE Bi-plot Analysis in Multi-Environment Trials in Sorghum. *Eur Exp Biol*, *7*(6), 40. https://doi.org/10.21767/2248-9215.100040

Orellana, M. A., Edwards, J., & Carriquiry, A. L. (2014). Heterogeneous variances in multi-environment yield trials for corn hybrids. *Crop Science*, *54*(3), 1048. https://doi.org/10.2135/cropsci2013.09.0653

Perez-Elizalde, S., Jarquin, D., & Crossa, J. (2012). A general Bayesian estimation method of linear–bilinear models applied to plant breeding trials with genotype× environment interaction. *Journal of Agricultural, Biological, and Environmental Statistics*, *17*(1), 15–37. https://doi.org/10.1007/s13253-011-0063-9

Piepho, H. P. (1997). Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics*, 761–766. https://doi.org/10.5935/1806-6690.20170103

Piepho, H. P. (1998). Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theoretical and Applied Genetics*, *97*(1-2), 195–201. https://doi.org/10.1007/s001220050885

Piepho, H. P., & Möhring, J. (2006). Selection in cultivar trials—Is it ignorable? *Crop Science*, *46*(1), 192–201. https://doi.org/10.2135/cropsci2005.04-0038

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.Rproject.org/

Raftery, A. E., & Lewis, S. (1992). How many iterations in the Gibbs sampler?. In J. M. Bernardo et al. (Eds.), *Bayesian Statistics* (pp. 763–773). Oxford Univ. Press. https://doi.org/10.21236/ada640705

Raftery, A. E., Newton, M. A., Satagopan, J. M., & Krivitsky, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, & M. West (Eds.), *Bayesian Statistics*, (pp. 1–45). Oxford University Press.

Rodrigues, P. C., Malosetti, M., Gauch, H. G., & van Eeuwijk, F. A. (2014). A weighted AMMI algorithm to study genotype-by-environment interaction and QTL-by-environment interaction. *Crop Science*, *54*, 1555–1570. https://doi.org/10.2135/cropsci2013.07.0462

Romão, R. F., Nuvunga, J. J., Silva, C. P., Oliveira, L. A., Mendes, C. T. E., & Balestre, M. (2019). Predictive ability of AMMI and factorial analytical models in the study of unbalanced multi-environment data. *Genetics and Molecular Research*, *18*(3), 1–20. https://doi.org/10.4238/gmr18176

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464. https://doi.org/10.1214/aos/1176344136

Smith, A. B., & Cullis, B. R. (2018). Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica*, *214*(8), 1–19. https://doi.org/10.1007/s10681-018-2220-5

Smith, A., Cullis, B., & Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for

spatial field trend. *Biometrics*, *57*(4), 1138–1147. https://doi.org/10.1111/j.0006-341X.2001.01138.x

Smith, A. B., Cullis, B. R., & Thompson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *The Journal of Agricultural Science*, *143*(6), 449–462. https://doi.org/10.1017/S0021859605005587P

Smith, A. B., Ganesalingam, A., Kuchel, H., & Cullis, B. R. (2015). Factor analytic mixed models for the provision of grower information from national crop variety testing programs. *Theoretical and Applied Genetics*, *128*(1), 55–72. https://doi.org/10.1007/s00122-014-2412-x

Spilke, J., Piepho, H. P., & Hu, X. (2005). Analysis of unbalanced data by mixed linear models using the MIXED procedure of the SAS system. *Journal of Agronomy and Crop Science*, *191*(1), 47–54. https://doi.org/10.1111/j.1439-037X.2004.00120.x

Thompson, R., Cullis, B., Smith, A., & Gilmour, A. (2003). A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Australian & New Zealand Journal of Statistics*, *45*(4), 445–459. https://doi.org/10.1111/1467-842X.00297

Viele, K., & Srinivasan, C. (2000). Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback–Leibler Information. *Journal of Statistical Planning and Inference*, *84*(1-2), 201–219. https://doi.org/10.1016/S0378-3758(99)00151-2

Xu, N. Y., Fok, M., Zhang, G. W., Jian, L. I., & Zhou, Z. G. (2014). The application of GGE biplot analysis for evaluating test locations and mega-environment investigation of cotton regional trials. *Journal of Integrative Agriculture*, *13*(9), 1921–1933. https://doi.org/10.1016/S2095-3119(13)60656-5

Yan, W. (2001). GGEbiplot — a Windows application for graphical analysis of multi-environment trial data and other types of two-way data. *Agronomy Journal*, *93*(5), 1111–1118. https://doi.org/10.2134/agronj2001.9351111x

Yan, W. (2013). Biplot analysis of incomplete two-way data. *Crop Science*, *53*(1), 48–57. https://doi.org/10.2135/cropsci2012.05.0301

Yan, W. (2014). *Crop variety trials: Data management and analysis*. John Wiley & Sons.

Yan, W., Glover, K. D., Kang, M. S., Yang, R. C., Crossa, J., & Burgueño, J. (2010). Comment on" Biplot Analysis of Genotype× Environment Interaction: Proceed with Caution, by R.-C. Yang, J. Crossa, P.L. Cornelius, and J. Burgueno in 2009, 49:1564–1576. *Crop Science*, *50*, 1121–1123. https://doi.org/10.2135/cropsci2010.01.00011

Yan, W., Hunt, L. A., Sheng, Q., & Szlavnics, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop science*, *40*(3), 597–605. https://doi.org/10.2135/cropsci2000.403597x

Yan, W., & Kang, M. S. (2003). *GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists*. CRC Press.

Yang, R. C., Crossa, J., Cornelius, P. L., & Burgueño, J. (2009). Biplot analysis of genotype× environment interaction: Proceed with caution. *Crop Science*, *49*(5), 1564–1576. https://doi.org/10.2135/cropsci2008.11.0665

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: de Oliveira, L. A., da Silva, C. P., da Silva, A. Q., Mendes, C. T. E., Nuvunga, J. J., Nunes, J. A. R., Parrella, R. A. d. C., Baleste, M., & Filho, J. S. d. S. B. (2022). Bayesian GGE model for heteroscedastic multienvironmental trials. *Crop Science*, *62,* 982–996. https://doi.org/10.1002/csc2.20696