

Efeito de encolhimento na análise bayesiana do modelo GGE utilizando priori de máxima entropia

Luciano A. Oliveira¹, Carlos P. Silva², Alessandra Q. Silva^{1†}, Cristian T. E. Mendes³, Joel J. Nuvunga⁴, Júlio S. S. Bueno Filho⁵

¹Universidade Federal da Grande Dourados (UFGD).

²Universidade Federal de Alfenas (UNIFAL-MG).

³Doutorando em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras (UFLA).

⁴Universidade Eduardo Mondlane (UEM), Moçambique.

⁵ Universidade Federal de Lavras (UFLA).

Resumo: Neste trabalho, utilizou-se o princípio de máxima entropia para atribuir informação a priori às componentes de variância dos valores singulares no modelo de efeitos principais de genótipos mais interação genótipo×ambientes (GGE). O método foi exemplificado a partir de um conjunto de dados simulados. Os resultados mostraram que o modelo GGE com priori de máxima entropia (BGGE) produz efeito de encolhimento nas estimativas dos valores singulares, quando comparado com a análise GGE frequentista ou com a versão bayesiana GGE que utiliza priori não informativa (referida por BGGE). O BGGE mostrou maior parcimônia, estimando os valores singulares com maior contribuição para o efeito da interação e encolhendo as estimativas de valores singulares associados a maiores dimensões a zero. Assim, o BGGE capturou mais padrão e descartou mais ruído do que a versão Bayesiana típica. Ao utilizar priori de máxima entropia, constatou-se que o modelo completo e aquele com apenas dois termos bilineares são quase indistinguíveis. Isso sinaliza que a seleção do modelo no ajuste BGGE não seria uma etapa necessária. O método também evita problemas de amostragem observados quando a priori de Jeffreys é utilizada, resultando em distribuições marginais a posteriori próprias e uni modais.

Palavras-chave: Modelo GGE; bayesiana; máxima entropia.

Abstract: In this work, the maximum entropy principle was used to assign a priori information to the variance components of the singular values in the genotype main effects model plus genotype×environment interaction (GGE). The method was exemplified from simulated data. The results showed that the GGE model with maximum entropy priori (BGGE) produces a shrinkage effect on the estimates of singular values, when compared with the frequentist GGE analysis or with the Bayesian GGE version that uses non-informative priors (referred to by BGGE). The BGGE showed greater parsimony, estimating the singular values with greater contribution to the interaction effect and shrinking the estimates of singular values associated with larger dimensions to zero. Thus, BGGE captured more pattern and discarded more noise than the typical Bayesian version. When using maximum entropy priori, it was found that the complete model and the one with only two bilinear terms are almost indistinguishable. This signals that model selection in the BGGE fit would not be a necessary step. The method also avoids sampling problems observed when Jeffreys prior are used, resulting in proper and unimodal posterior marginal distributions.

Keywords: GGE model; Bayesian inference; Maximum entropy.

† Autora correspondente: alessandrasilva@ufgd.edu.br.

Introdução

Ensaio de avaliação de cultivares em vários ambientes (MET-multienvironmental trials) são de fundamental importância em programas de melhoramento de plantas e devem ser realizados no decorrer de vários anos. Os resultados observados podem ser utilizados para classificar, selecionar e recomendar genótipos superiores segundo características de interesse. Esses ensaios permitem modelar a interação genótipo×ambiente (GEI- genotype × environment interaction) que representa um dos maiores desafios ao trabalho do melhorista de plantas.

A GEI é percebida pela resposta diferenciada de genótipos frente à diversidade ambiental e pode ocasionar mudanças na classificação das respostas fenotípicas de um ambiente a outro (CROSSA; YANG; CORNELIUS, 2004; YANG et al., 2009). A presença de GEI significativa dificulta a avaliação e recomendação ampla de cultivares, mas oferece a possibilidade de explorá-la em seus efeitos positivos, visando identificar adaptações de genótipos a ambientes específicos (BURGUENÑO et al., 2008; CRUZ; REGAZZI; CARNEIRO, 2012).

Modelos lineares-bilineares (ou multiplicativos) de efeitos fixos são úteis para analisar dados bidirecionais dispostos em uma tabela de dupla entrada. Nessa classe de modelos, estimam-se, inicialmente, os efeitos principais por mínimos quadrados ordinários (ordinary least squares-OLS) e os parâmetros bilineares são obtidos, em um segundo estágio, pela aplicação da decomposição por valores singulares (DVS) à matriz de erros de não aditividade aos efeitos principais (CORNELIUS; SEYEDSADR, 1997).

O padrão de respostas de genótipos entre os ambientes pode ser visualizado diretamente em biplots, que são representações gráficas baseadas nas propriedades do produto interno de vetores compostos pelos escores genotípicos e ambientais. Dentre os modelos lineares-bilineares, o modelo de efeitos principais de genótipos (G) mais GEI (GGE), merece especial destaque pela grande aplicabilidade em programas de melhoramento de plantas (CORNELIUS; CROSSA; SEYEDSADR, 1996; CROSSA; CORNELIUS, 1997). Outros modelos multiplicativos podem ser obtidos a partir de um modelo linear-bilinear geral (CORNELIUS; SEYEDSADR, 1997).

No GGE, o efeito G é confundido com o efeito da GEI. O efeito combinado G+GEI é absorvido pelos parâmetros bilineares do modelo e obtidos por decomposição de valores singulares (DVS) da matriz de resíduos ao ajuste dos efeitos de ambientes. O GGE biplot utiliza os dois primeiros componentes principais oferecendo uma análise gráfica do comportamento das cultivares nos diferentes ambientes. Pela simplicidade das interpretações, o GGE biplot se tornou amplamente utilizado em ensaios MET, oferecendo diversas possibilidades para o estudo da GEI (YAN, 2014).

Embora o GGE, bem como os demais modelos multiplicativos, traga grandes vantagens à análise, sua versão frequentista (clássica) tem reconhecidas limitações dos modelos que consideram os efeitos como sendo fixos (CROSSA et al., 2011; OLIVEIRA et al., 2015). Dentre essas limitações destaca-se a dificuldade para incorporar incerteza aos escores genotípicos e ambientais nas representações biplots (YANG et al., 2009; YAN et al., 2010; OLIVEIRA et al., 2015).

Crossa et al. (2011) contornaram essa dificuldade incorporando regiões de credibilidade aos escores no biplot AMMI-2. Para tanto resgataram os trabalhos pioneiros de Viele e Srivivasan (2000) e Liu (2001), que utilizaram a inferência bayesiana para estimar parâmetros com o uso do método MCMC (Markov chain Monte Carlo) no processo de amostragem. Várias contribuições foram dadas ao método desde então (PEREZ-ELIZALDE et al., 2012; OLIVEIRA et al., 2015; SILVA et al., 2015; SILVA et al., 2019).

A análise bayesiana do modelo GGE também foi proposta por Jarquin et al. (2016) e Oliveira et al. (2016). Contudo, esses trabalhos utilizaram densidades a priori não informativas para todos os parâmetros. Oliveira et al. (2021) estenderam o método dos estimadores de encolhimentos, proposto para o modelo AMMI por Silva et al. (2019), aos parâmetros que descrevem G+GEI no modelo GGE bayesiano obtendo, dessa forma, modelos mais parcimoniosos. A análise bayesiana considerando heterogeneidade de variâncias também foi implementada para o AMMI e GGE (SILVA et al., 2019; OLIVEIRA et al., 2022).

O objetivo desse trabalho foi utilizar o método proposto por Oliveira et al. (2021) e analisar um conjunto de dados simulados, ilustrando diferenças que surgem ao comparar seus resultados com o GGE frequentista e com a análise GGE-bayesiana que utiliza distribuição a priori não informativa para os parâmetros bilineares.

Materiais e Métodos

Dados simulados

Foi simulado um conjunto de dados com 35 genótipos (G1-G35), avaliados em 13 ambientes (E1-E13), usando o delineamento de blocos ao acaso com duas repetições. Os efeitos principais foram simulados com distribuições Gaussianas, genótipos (G)~N(0,5), ambientes (E)~N(0,4) e blocos (B)~N(0,1).

Três padrões de resposta distintos foram simulados para o GEI: i) 12 genótipos com interação positiva em todos ambientes a partir de uma gaussiana +N(0,3), ii) 12 genótipos com interação negativa, em todos os ambientes com -N(0,3) e iii) 11 genótipos com interações simuladas de uma normal padrão N(0,1).

Os dois primeiros subgrupos de genótipos são aqueles que possuem um padrão de resposta diferente entre os ambientes e são chamados de genótipos instáveis. No primeiro subgrupo + N(0, 3), significa que foram considerados os módulos dos valores. Para o segundo grupo multiplicou-se os módulos dos valores simulados de N(0, 3) pelo valor (-1). No último subgrupo iii) os genótipos não possuem padrão de resposta diferente entre os ambientes e, portanto, são considerados genótipos estáveis.

Os valores simulados foram organizados em uma tabela de dupla entrada e corrigidos pelos efeitos de linha (genótipos) e pelos efeitos de coluna (ambientes), constituindo assim a matriz de interação GEI. Cada observação no conjunto simulado foi obtida adicionando os respectivos efeitos de genótipos, ambientes, blocos e interações a uma média geral e um erro que foi simulado de uma gaussiana com média 0 e variância 6 (N(0, 6)).

Modelo GGE

No modelo GGE, o vetor \mathbf{y} contendo $n = g \times l$ respostas fenotípicas, em que g é o número de genótipos repetidos $l = b \times c$ nas combinações de blocos (b é número de blocos) dentro de locais (c é o número de locais) pode ser representado por:

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

em que $\boldsymbol{\beta}_{l \times 1}$ é o vetor de efeitos de blocos dentro de ambientes; λ_k , $\boldsymbol{\alpha}_k$ e $\boldsymbol{\gamma}_k$ são, respectivamente, o valor singular e vetores singulares (genotípico e ambiental) associados ao k -ésimo componente principal ($k = 1, \dots, t$), com $t = \min(r-1, c)$ sendo o posto da matriz $\mathbf{GGE}_{r \times l}$. Na análise GGE, $\mathbf{GGE}_{r \times l}$ é a matriz composta pelos erros de não-aditividade aos efeitos principais de ambientes. Por sua vez, \mathbf{X}_1 , \mathbf{X}_2 e \mathbf{Z} são as matrizes de delineamento e o vetor $\boldsymbol{\epsilon}_{n \times 1}$ contém os efeitos dos erros experimentais, com $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \mathbf{I}\sigma_\epsilon^2)$, sendo σ_ϵ^2 a variância residual, $\mathbf{0}_{n \times 1}$ o vetor nulo e \mathbf{I}_n a matriz identidade de ordem n . A distribuição condicional dos dados amostrais é normal multivariada $\mathbf{y} | \boldsymbol{\beta}, \lambda, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{I}, \sigma_\epsilon^2 \sim N_n(\boldsymbol{\theta}, \mathbf{I}\sigma_\epsilon^2)$, em que $\boldsymbol{\theta} = \mathbf{X}_1\boldsymbol{\beta} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}_2\boldsymbol{\gamma}_k$.

As distribuições *a priori* assumidas são as mesmas propostas por Oliveira et al. (2021):

$$\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \sigma_\beta^2 \sim N(\boldsymbol{\mu}_\beta, \mathbf{I}\sigma_\beta^2), \quad \boldsymbol{\mu}_\beta = \mathbf{0} \text{ e } \sigma_\beta^2 = 10^8;$$

$\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2 \sim N^+(\mu_{\lambda_k}, \sigma_{\lambda_k}^2)$, $\mu_{\lambda_k} = 0$, sendo consideradas duas suposições para $\sigma_{\lambda_k}^2$:

- i) $\sigma_{\lambda_k}^2 = 10^8$ utilizada por Oliveira et al. (2016) cujo modelo é referido por BGGE (*Bayezian GGE*) e
- ii) $\sigma_{\lambda_k}^2 \sim GI(a, b)$ que corresponde a suposição utilizada por Oliveira et al. (2021). O modelo que utiliza essa distribuição a priori será designado daqui por diante por BGGEE (*Bayesian GGE model with Entropy*).

$\alpha_k \sim$ Uniforme esférica no subespaço corrigido;

$\gamma_k \sim$ Uniforme esférica no subespaço corrigido.

Assumindo as hipóteses sobre os hiperparâmetros das densidades a priori, as densidades condicionais completas a posteriori são as mesmas do modelo BGGE proposto por Oliveira et al. (2016), com exceção daquela sob a hipótese ii) para o modelo com efeito de encolhimento. São elas:

$$\beta | \dots \sim N[(X_1^t X_1)^{-1} X_1^t (\mathbf{y} - \Theta), (X_1^t X_1)^{-1} \sigma_e^2], \text{ sendo } \Theta = \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z} \alpha_k) X_2 \gamma_k,$$

$$\lambda | \dots \sim N^+[(\Phi_k^t \Phi_k)^{-1} \Phi_1^t \Delta_{k'} (\Phi_k^t \Phi_k)^{-1} \sigma_e^2], \quad \text{em que } \Phi_k = \text{diag}(\mathbf{Z} \alpha_k) X_2 \gamma_k,$$

$$\Delta_{k'} = \sum_{k'=k'}^t \lambda_{k'} \text{diag}(\mathbf{Z} \alpha_{k'}) X_2 \gamma_{k'} \text{ e } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_t \geq 0.$$

Para os vetores singulares são obtidas distribuições a posteriori proporcionais a *Von Mises Fisher-VMF* (CROSSA et al., 2011).

$$p(\alpha_k | \dots) \propto \exp\{k \alpha_k^t \Lambda_k^t (\mathbf{y} - X_1 \beta)\}, \text{ com } \Lambda_k = \text{diag}(X_2 \gamma_k) \mathbf{Z} \text{ e}$$

$$p(\gamma_k | \dots) \propto \exp\{k \gamma_k^t \Omega_k^t (\mathbf{y} - X_1 \beta)\}, \text{ com } \Omega_k = \text{diag}(\mathbf{Z} \alpha_k) X_2.$$

Por fim, a distribuição condicional da variância residual é uma inversa qui-quadrado escalada $\sigma_e^2 \sim \text{inv-escl} - \chi^2[n, (\mathbf{y} - \Theta)^t (\mathbf{y} - \Theta)]$.

Para o modelo BGGEE, é assumida uma priori hierárquica para λ_k em dois níveis, sendo a densidade a priori de $\sigma_{\lambda_k}^2$ uma gama-invertida obtida pelo método de máxima entropia (OLIVEIRA et al., 2021). Nesse sentido, a distribuição condicional completa a posteriori para λ_k , sob a hipótese ii) (BGGEE), é dada por:

$$\lambda_k | \dots \sim N^+ \left[\left(\Lambda_k^t \Lambda_k + \frac{\sigma_e^2}{\sigma_{\lambda_k}^2} \right)^{-1} \Lambda_k^t \mathbf{M}_{k'}^t, (\Lambda_k^t \Lambda_k)^{-1} \sigma_e^2 \right] \quad (3)$$

em que $\mathbf{M}_{k'} = \mathbf{y} - X_1 \beta - \sum_{k'=k'}^t \lambda_{k'} \text{diag}(\mathbf{Z} \alpha_{k'}) X_2 \gamma_{k'}$.

Por sua vez, $\sigma_{\lambda_k}^2 | \dots \sim GI(a + 1, \lambda_k^2 + 2b)$ com parâmetros de escala $\lambda_k^2 + 2b$ e com $a + 2b$ graus de liberdade. Considerando as hipóteses *a priori* ($a = 1$ e $b = 0$) obtém-se $\sigma_{\lambda_k}^2 | \dots \sim GI(2, \lambda_k^2)$. Detalhes algébricos podem ser consultado em Oliveira et al. (2021).

Estimativas para todos os parâmetros foram obtidas pela media a posteriori das cadeias de Markov com a utilização de um amostrador de Gibbs. Os passos para o processo iterativo são os mesmos utilizados por Oliveira et al. (2015) e Oliveira et al. (2021).

Devido à restrição de ortonormalidade, a amostragem para os vetores singulares é realizada em um espaço corrigido utilizando variáveis auxiliares por transformações lineares ortogonais (OLIVEIRA et al., 2015; VIELE; SRNIVASAN, 2000).

A convergência das cadeias de Markov obtidas pelo amostrador de Gibbs foram monitoradas utilizando os critérios de Raftery e Lewis (1992) e Heidelberger e Welch (1983).

Os vetores singulares estimados não satisfazem as restrições de ortonormalidade. Nesse sentido, é necessário realizar uma transformação nas matrizes formadas pelas médias das coordenadas e para isso, utilizou-se o método proposto por Liu (2001).

Regiões de máxima densidade a posteriori (HPD) a 95% de credibilidade para os parâmetros do modelo foram obtidas pelo método proposto por Chen e Shao (1999), sendo implementadas com auxílio do pacote estatístico Bayesian output analysis (BOA) disponível no software R (R CORE TEAM, 2019).

Regiões de credibilidade para os escores $(\lambda_1^{1/2}\alpha_{i1}, \lambda_2^{1/2}\alpha_{i2})$ e $(\lambda_1^{1/2}\gamma_{j1}, \lambda_2^{1/2}\gamma_{j2})$, sendo $i = 1, \dots, 35$ e $j = 1, \dots, 35$, foram incorporadas ao biplot utilizando-se o método de Hu e Yang (2013). As interpretações são realizadas observando posições e sobreposições entre as regiões de credibilidade para os escores genotípicos e ambientais no biplot.

A seleção de modelos no contexto bayesiano foi realizada utilizando o critério de informação de Akaike Monte Carlo-AICM (RAFTERY et al., 2007).

O modelo GGE frequentista também foi ajustado para fins de comparação. Estimativas para os parâmetros lineares foram obtidas por mínimos quadrados e para os parâmetros bilineares por meio da DVS da matriz de interação. A seleção de modelos foi conduzida utilizando o teste Fr de Cornélius (CORNELIUS et al., 1993).

Todo processo de inferência e simulação foi realizado utilizando o software estatístico R (R CORE TEAM, 2019).

Resultados e discussão

Análise frequentista do modelo GGE

Na análise conjunta de experimentos (Tabela 1), o efeito da interação foi significativo e, portanto, o uso de modelos lineares-bilineares seria recomendado. Utilizando-se o teste Fr de Cornelius (CORNELIUS, 1993) percebe-se que o modelo GGE3, ou seja, o modelo em que os três primeiros componentes principais são retidos é a melhor escolha para representar o conjunto de dados. O resultado do teste Fr é apresentado na Tabela 2.

Tabela 1. Análise conjunta de experimentos referentes aos dados simulados em que 35 genótipos foram considerados em 13 ambientes.

F.V.	G.L.	SQ	QM	Valor F	Pr (>F)
Ambientes	12	992,0787	82,6732	7,8922	0 ***
Genótipos	34	5267,5585	154,9282	14,7899	0 ***
Blocos(amb)	1	0,0311	0,0311	0,0048	0,9448
Interação	408	4273,9031	10,4753	1,6147	0 ***
Resíduos	454	2945,3653	6,4876		
Total	909	13478,9366			

Fonte: Autores.

Embora o GGE3 seja o modelo selecionado, ao utilizar o teste Fr, a análise gráfica biplot mais popularmente utilizada é a versão GGE que considera apenas os dois primeiros componentes principais (PC1 e PC2). Essa análise foi proposta por Yan et al. (2000), sendo referida na literatura por análise GGE biplot. A representação GGE biplot do conjunto de dados simulados é apresentada na Figura 1.

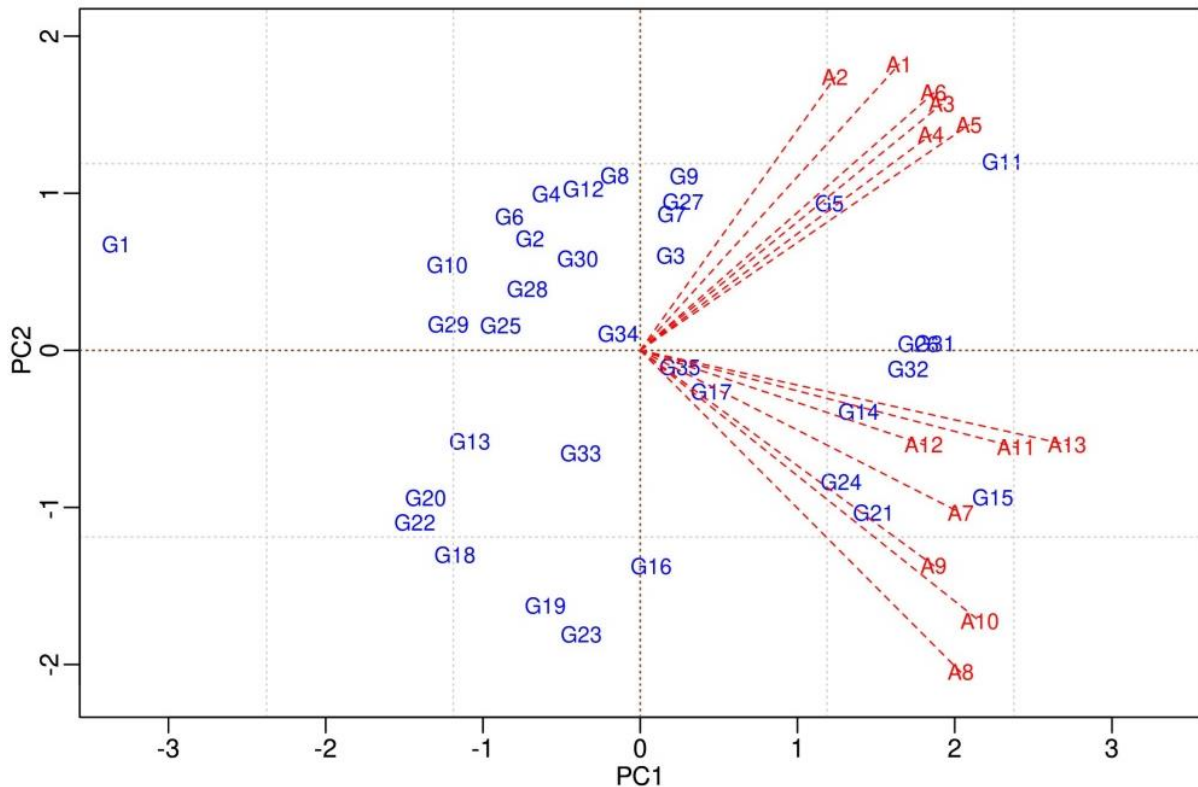
Tabela 2. Seleção de modelos utilizando o teste Fr de cornélius

Modelos	Per. %	Per. Ac. %	G.L.	SQ	QM	Valor F	P(>F)
GGE1	56,65	56,65	363	5404,8272	11,3957	1,7565	0 ***
GGE2	14,87	71,51	320	1418,6074	8,4938	1,3092	0,0043**
GGE3	6,43	77,94	279	613,637	7,5426	1,1626	0,0784
GGE4	5,34	83,29	240	509,7103	6,6445	1,0242	0,4115
GGE5	4,05	87,33	203	385,9898	5,9541	0,9178	0,7576
GGE6	3,01	90,35	168	287,4728	5,4834	0,8452	0,8995
GGE7	2,91	93,26	135	277,9256	4,7651	0,7345	0,9836
GGE8	1,68	94,94	104	160,7566	4,6398	0,7152	0,9807
GGE9	1,50	96,45	75	143,3937	4,5219	0,697	0,9723
GGE10	1,32	97,77	48	125,9966	4,4405	0,6845	0,9471
GGE11	0,10	98,77	23	95,5794	5,1115	0,7879	0,7478
GGE12	0,79	99,56	0	75,5849	0	0	1

Legenda: 0 '***' 0.001 '**' 0.01 '*'

Fonte: Autores.

Figura 1. Representação GGE biplot de 35 genótipos simulados em 13 ambientes.



Fonte: Autores.

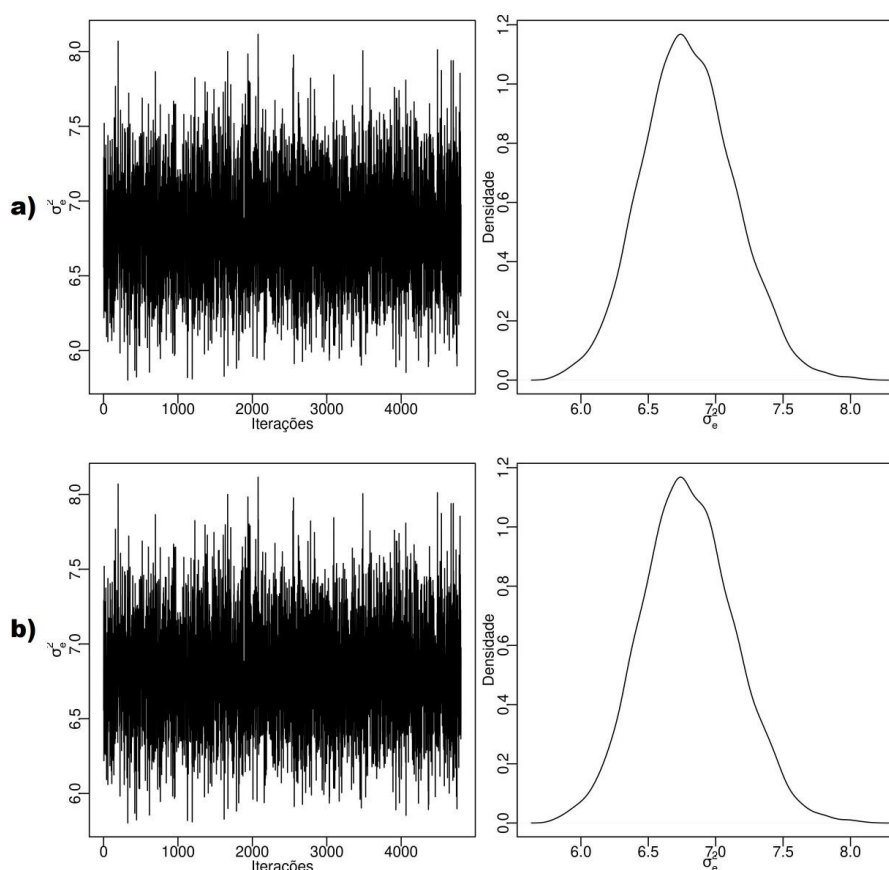
Considerando uma análise em um ambiente médio os genótipos G26, G31 e G32 seriam os de melhor desempenho, pois possuem alto valor para PC1 e valores pequenos para o PC2. Contudo, observa-se nitidamente a formação de dois subgrupos de ambientes (Mega-ambientes) em relação as posições no biplot, são eles: Mega-ambiente1 (A1, A2, A3, A4, A5, A6) e Mega-ambiente2 (A7, A8, A9, A10, A11, A12, A13). Para o Mega-ambiente1, G5 e G11 seriam os de melhor desempenho, enquanto para o Mega-ambiente2 os escolhidos seriam G15 e G21

Análise Bayesiana do modelo BGGE

Os números de iterações das cadeias MCMC, valores de queima e de desbastes, para garantir a convergência, foram definidos a partir de amostras piloto utilizando o critério de Raftery e Lewis (1992). Para todos os parâmetros observaram-se fator de dependência sempre abaixo de cinco ($I < 5$). Além disso, todos os parâmetros passaram pelo teste de estacionariedade de acordo com o critério de Heidelberger e Welch (1983).

Na Figura 2 são apresentados gráficos dos traços e de densidade das cadeias para as variâncias residuais dos modelos BGGE e BGEE obtidas a partir do ajuste do modelo completo (todos os componentes foram retidos). Nesses gráficos observam-se oscilações aleatórias em torno de um valor estável, corroborando os resultados positivos dos critérios utilizados para monitorar as cadeias e sugerindo que a convergência foi alcançada. Esse comportamento foi observado para todas as cadeias geradas pelo método MCMC em todos os ajustes conduzidos.

Figura 2. Representação gráfica dos traços e densidades das variâncias residuais dos modelos BGGE e BGEE, considerando modelos completos.



Fonte: Autores.

Na Tabela 3 são apresentados resumos das distribuições a posteriori dos valores singulares para os modelos BGGE e BGEE, em função do número de eixos mantidos no ajuste. É possível observar encolhimento acentuado nas estimativas dos valores singulares no modelo BGEE, a partir do modelo com três eixos ajustados (BGEE3), quando comparado às soluções de mínimos quadrados do modelo fixo e estimativas a posteriori do modelo BGGE. O BGEE pode ser considerado uma versão bayesiana dos estimadores de encolhimento propostos por Cornelius e Crossa (1999). Mas o método proposto por esses autores não faz qualquer suposição de efeito aleatório e pode ser utilizado apenas em um cenário balanceado e homogêneo. Além disso, o procedimento para

obtenção desses preditores, eventualmente, viola as restrições do modelo e requer passos adicionais de correção no algoritmo (SILVA et al., 2015, OLIVEIRA et al., 2021).

Observa-se também encolhimento, não tão acentuado, das estimativas bayesianas do BGGE em relação à solução DVS. A utilização de priori não informativa produz encolhimento semelhante ao que ocorre nos Blups de modelos mistos (SILVA et al., 2015; OLIVEIRA et al., 2015). Contudo, o Blup encolhe as médias de células relacionadas a GEI segundo uma razão constante ao passo que o BGGEE aplica diferentes proporções de encolhimentos diretamente no valor singular. Liu (2001) e Crossa et al. (2011) relataram que a utilização de priori não informativa produz resultado semelhante ao método do estimador de encolhimento aplicado ao modelo de efeitos fixos.

Tabela 3. Estimativas dos valores singulares para a família de modelos BGGE e BGGEE em função do número de termos bilineares ($k = 1, 2, \dots, 13$).

BGGE														
Modelo	Dim	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}
BGGE	1	50,14												
BGGE	2	50,46	23,54											
BGGE	3	50,59	23,72	11,36										
BGGE	4	50,56	23,79	11,70	6,77									
BGGE	5	50,62	23,80	11,86	6,92	3,07								
BGGE	6	50,62	23,83	11,89	7,02	3,09	1,36							
BGGE	7	50,60	23,86	11,93	7,05	3,12	1,37	0,63						
BGGE	8	50,61	23,85	11,97	7,10	3,11	1,37	0,63	0,30					
BGGE	9	50,60	23,81	11,88	6,88	3,10	1,37	0,64	0,31	0,15				
BGGE	10	50,63	23,82	11,95	7,09	3,13	1,38	0,63	0,31	0,15	0,07			
BGGE	11	50,60	23,89	11,82	7,02	3,08	1,36	0,63	0,31	0,15	0,07	0,04		
BGGE	12	50,58	23,79	11,94	7,02	3,13	1,36	0,63	0,31	0,15	0,08	0,04	0,02	
BGGE	13	50,63	23,86	11,87	6,83	3,00	1,29	0,61	0,30	0,14	0,07	0,04	0,02	0,01

BGGEE														
Modelo	Dim	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}
BGGEE	1	49,96												
BGGEE	2	50,26	23,00											
BGGEE	3	50,23	23,03	0,00*										
BGGEE	4	50,25	23,03	0,00*	0,00**									
BGGEE	5	50,26	22,97	0,00*	0,00**	0,00**								
BGGEE	6	50,28	22,97	0,00*	0,00**	0,00**	0							
BGGEE	7	50,24	23,00	0,00*	0,00**	0,00**	0	0						
BGGEE	8	50,27	23,02	0,00*	0,00**	0,00**	0	0	0					
BGGEE	9	50,30	22,97	0,00*	0,00**	0,00**	0	0	0	0				
BGGEE	10	50,27	22,99	0,00*	0,00**	0,00**	0	0	0	0	0			
BGGEE	11	50,29	23,02	0,00*	0,00**	0,00**	0	0	0	0	0	0		
BGGEE	12	50,29	22,99	0,00*	0,00**	0,00**	0	0	0	0	0	0	0	
BGGEE	13	50,30	23,01	0,00*	0,00**	0,00**	0	0	0	0	0	0	0	0

* < 0,0006; ** < 0,0002; 0 < 0,0001.

Fonte: Autores.

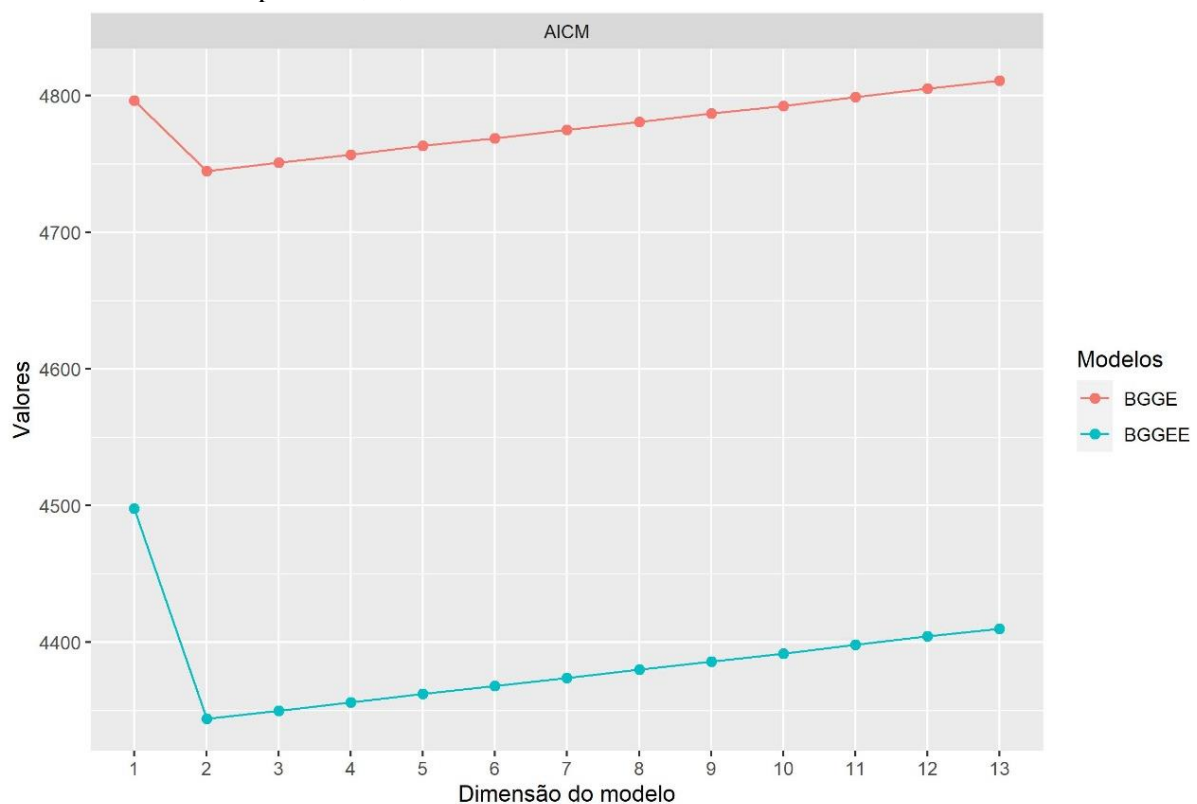
Silva et al. (2015) obtiveram encolhimento nas estimativas de valores singulares do modelo AMMI semelhante ao obtido aqui. Contudo, utilizaram prioris de Jeffreys como distribuições a priori para as variâncias dos valores singulares. A utilização de tais prioris resulta em densidades condicionais qui-quadrado-inversas escaladas com apenas um grau de liberdade e, conseqüentemente, em distribuições marginais a posteriori impróprias (SILVA et al., 2015). Para contornar esse

problema, os autores supracitados utilizaram a priori estendida de Ter Braak et al. (2005) que corrige os graus de liberdade. Uma consequência dessa correção é que a densidade a posteriori se torna bimodal. Esses problemas não ocorrem com a utilização da priori de máxima entropia como argumentam Oliveira et al. (2021) e, portanto, essa escolha flexibiliza a análise bayesiana do modelo GGE.

Seleção de modelos

Os valores da estatística AICM para os modelos BGGE e BGGE são representados na Figura 3, sendo que menores valores indicam melhores ajustes. Para ambas as versões do modelo GGE, o modelo com dois componentes multiplicativos retidos (BGGE2 e BGGE2) obteve melhor desempenho. Além disso, é possível observar que os modelos da família BGGE obtiveram valores expressivamente menores para AICM.

Figura 3. Resultados da seleção de modelos utilizando o cálculo do AICM para a família de modelos BGGE- k e BGGE- k para $k=1, \dots, 13$.



Fonte: Autores.

A seleção de modelos na análise bayesiana de modelos multiplicativos tem sido realizada pela utilização de critérios de informação ou fator de bayes (LIU et al., 2001; SILVA et al., 2015; OLIVEIRA et al., 2021). Aqui optamos por utilizar o AICM, pois o mesmo tem mostrado mais consistência que o BIC ou o AIC, conforme argumentado por Silva et al. (2023).

Crerios de informação e fator de bayes são métodos que exigem a exploração de toda a dimensão do modelo e sua aplicação pode ser tornar difícil. Uma alternativa foi proposta por Silva et al. (2023) que utilizaram saltos reversíveis em cadeias de Markov para selecionar componentes multiplicativos no modelo AMMI. Além de se mostrar mais viável, esse procedimento pode ser visto como uma generalização do algoritmo Metropolis-Hastings o que lhe dá certo apelo conceitual.

Seleção de modelos Inferência a partir dos modelos selecionados

Estimativas pontuais e intervalares (regiões HPD à 95% de credibilidade) dos valores singulares e componentes de variância para os modelos BGGE2 e BGGEE2 (modelos selecionados com base no AICM) são apresentadas na Tabela 4. As estimativas para os dois primeiros valores singulares e variância residual são bem próximas para os dois modelos bayesianos em análise.

Tabela 4. Estimativas pontuais e intervalares para valores singulares e componentes de variância para os modelos BGGE-3 e BGGEE-2.

Modelo BGGE				
Parâmetros	Média	SD	LI	LS
σ_e^2	67,997	3,436	6,121	74,577
λ_1	504,585	18,505	467,289	539,704
λ_2	235,422	1,985	195,904	237,714
Modelo BGGEE				
Parâmetros	Média	SD	LI	LS
σ_e^2	68,14	3,447	61,895	75,273
λ_1	502,647	18,867	467,321	540,029
λ_2	229,962	2,034	190,366	269,735
$\sigma_{\lambda_1}^2$	309,952	265,678,772	1,114,115	7,027,101
$\sigma_{\lambda_2}^2$	5,813,495	4,303,775	218,894	1,553,961

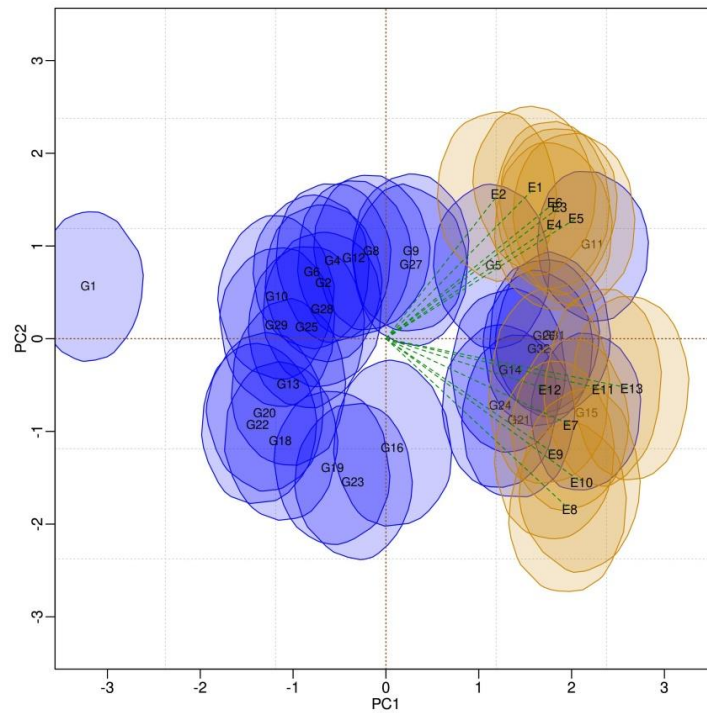
Fonte: Autores.

A representação biplot BGGE e BGGEEH com regiões bivariadas a 95% de credibilidade para os escores que descrevem G+GEI são apresentadas nas Figuras 4 e 5, respectivamente. As configurações dos escores nos biplots, com posições e sobreposições das regiões de credibilidade, são praticamente idênticas. Contudo, no biplot BGGEE, as regiões de credibilidade para G9 e G27 incluem a origem. No BGGE espera-se maior encolhimento das distâncias dos pontos em relação à origem, embora para esse conjunto de dados esse encolhimento não tenha sido acentuado. Os biplots, portanto, permitem basicamente as mesmas inferências. Observando as sobreposições das regiões de credibilidade para ambientes constatamos a presença de dois grupos homogêneos, ou seja, os mesmos Mega-ambientes vistos no biplot GGE para efeitos fixos.

Algumas vantagens do método bayesiano em comparação ao método frequentistas são listadas na literatura, como, por exemplo, a flexibilidade para lidar com dados desbalanceados e heterocedásticos (SILVA et al., 2018; ROMÃO et al., 2019; OLIVEIRA et al., 2021) e incorporar informações de experimentos anteriores à análise (PEREZ-ELIZALDE et al., 2012). Aqui, ressaltamos a versatilidade da metodologia Bayesiana em oferecer um procedimento paramétrico para incorporar inferência no Biplot. Existem métodos frequentistas para esse fim, mas os mesmos são suscetíveis a críticas seja pela incorporação de suposições restritivas sobre a distribuições dos escores genotípicos e ambientais (DENIS; GOWER, 1994), ou ainda pelos procedimentos de reamostragem problemáticos nas linhas e colunas da matriz de interação (YANG et al., 2009; YAN et al., 2010).

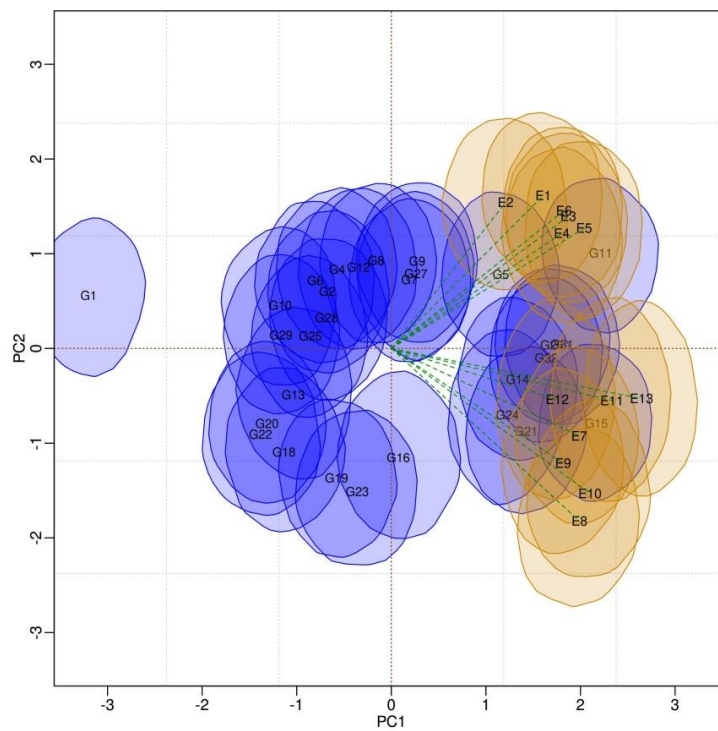
Os dois modelos abordados se diferem apenas pelas distribuições a priori atribuídas aos valores singulares. Como enfatizam Oliveira et al. (2021), uma crítica comum a métodos bayesianos é sua sensibilidade às especificações de distribuições a priori. Isso que pode levar a diferenças inferenciais, mesmo quando se utiliza uma mesma verossimilhança. Assim, argumenta-se que a obtenção de densidades a priori consistentes com o princípio da máxima entropia pode ser um caminho para evitar este problema (JAYNES, 1957, 2003).

Figura 4. Biplot para o modelo BGGE2 com regiões de credibilidade bivariadas a 95% para escores genotípicos e ambientais.



Fonte: Autores.

Figura 5. Biplot para o modelo BGEE2 com regiões de credibilidade bivariadas a 95% para escores genotípicos e ambientais.



Fonte: Autores.

Mais trabalhos são necessários para que a aplicação do método bayesiano a modelos lineares se tornem populares na análise de dados multiambientais. Uma desvantagem seria a demanda por tempo e recursos computacionais, o que pode ser amenizado pelo grande avanço tecnológico que estamos vivendo. Nesse sentido, pesquisas visando à otimização de softwares em interfaces mais amigáveis devem ser tema de futuras pesquisas.

Referências

- BURGUEÑO, J. et al. Using factor analytic models for joining environments and genotypes without crossover genotype _ environment interaction. *Crop Science*, v. 48, n. 4, p. 1291-1305, 2008.
- CHEN, M. H.; SHAO, Q. M. Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, v. 8, n. 1, p. 69-92, 1999.
- CORNELIUS, P.; SEYEDSADR, M. Estimation of general linear-bilinear models for two-way tables. *Journal of Statistical Computation and Simulation*, v. 58, n. 4, p. 287-322, 1997.
- CORNELIUS, P.; CROSSA, J.; SEYEDSADR, M. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In:----- . Genotype-byenvironment interaction. Boca Raton: CRC, 1996. p. 199-234.
- CORNELIUS, P. L.; CROSSA, J. Prediction assessment of shrinkage estimators of multiplicative models for multi-environment cultivar trials. *Crop Science*, v. 39, n. 4, p. 998-1009, 1999.
- CORNELIUS, P. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. *Crop Science*, v. 33, n. 6, p. 1186-1193, 1993.
- CROSSA, J.; YANG, R.-C.; CORNELIUS, P. L. Studying crossover genotype environment interaction using linear-bilinear models and mixed models. *Journal of Agricultural, Biological, and Environmental Statistics*, v. 9, n. 3, p. 362-380, 2004.
- CROSSA, J. et al. Bayesian estimation of the additive main effects and multiplicative interaction model. *Crop Science*, v. 51, n. 4, p. 1458-1469, 2011.
- CRUZ, C.; REGAZZI, A.; CARNEIRO, P. C. *Modelos biométricos aplicados ao melhoramento genético*. 4. ed. Viçosa: UFV, 2012. 514 p.
- CROSSA, J.; CORNELIUS, P. L. Sites regression and shifted multiplicative model clustering of cultivar trial sites under heterogeneity of error variances. *Crop Science*, v. 37, n. 2, p. 406-415, 1997.
- DENIS, J.-B.; GOWER, J. C. Asymptotic covariances for the parameters of biadditive models. *Utilitas Mathematica*, v. 46, p. 193-205, 1994.
- HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. *Operations Research*, v. 31, n. 6, p. 1109-1144, 1983.
- HU, Z.; YANG, R. Improved statistical inference for graphical description and interpretation of genotype times environment interaction. *Crop Science*, v. 53, n. 6, p.2400-2410, 2013.

- JARQUÍN, D. et al. A hierarchical Bayesian estimation model for multi-environment plant breeding trials in successive years. *Crop Science*, v. 56, n.5, p. 2260-2276, 2016.
- JAYNES, E. T. Information theory and statistical mechanics. *Physical Review*, APS, v. 106, n. 4, p. 620-630, 1957.
- JAYNES, E. T. *Probability theory: The logic of science*. New York: Cambridge University Press, 2003. 757 p.
- JEFFREYS, H. *Theory of Probability*. 3. ed. Oxford: Oxford Classic Texts in the Physical Sciences, 1961. 459 p.
- LIU, G. *Bayesian computations for general linear-bilinear models*. 2001. 169 p. Thesis (Doctor of Philosophy) - University of Kentucky, Lexington, 2001.
- OLIVEIRA, L. A. et al. Credible intervals for scores in the AMMI with random effects for genotype. *Crop Science*, v. 55, n. 2, p. 465-476, 2015.
- OLIVEIRA, L. A. de et al. Bayesian GGE biplot models applied to maize multienvironments trials. *Genetics and Molecular Research*, v. 15, n. 2, p. 1-21, 2016.
- PEREZ-ELIZALDE, S.; JARQUIN, D.; CROSSA, J. 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*, v. 17, n. 1, p. 15-37, 2012.
- R CORE TEAM. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2019. URL: <http://www.R-project.org/>
- RAFTERY, A. E. et al. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In: _____. *Bayesian Statistics*. [S.l.], 2007. v. 8, p. 1-45.
- RAFTERY, A. E.; LEWIS, S. M. [Practical Markov Chain Monte Carlo]: comment: one long run with diagnostics: implementation strategies for Markov Chain Monte Carlo. *Statistical Science*, v. 7, n. 4, p. 493-497, 1992.
- ROMÃO, R. F. et al. Predictive ability of AMMI and factorial analytical models in the study of unbalanced multi-environment data. *Genetics and Molecular Research*, v. 18, n. 3, p. 1-20, 2019.
- SILVA, C. P. et al. A Bayesian Shrinkage approach for AMMI Models. *PLoS One*, v.10, n. 7, p. e0131414, 2015.
- SILVA, C. P. et al. Heterogeneity of variances in the Bayesian AMMI model for multienvironment trial studies. *Crop Science*, v. 59, n. 6, p. 2455-2472, 2019.
- TER BRAAK, C.J.F; BOER, M.P.; BINK, M. C.A.M. Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. *Genetics*, v. 170, n. 3, p. 1435-1438, 2005.
- VIELE, K.; SRINIVASAN, C. Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback-Leibler information. *Journal of Statistical Planning and Inference*, v. 84, n. 1-2, p. 201-219, 2000.

YAN, W.; GLOVER, K. D.; KANG, M. S. Comment on “biplot analysis of genotype _environment interaction: proceed with caution” by R.-C. Yang, J. Crossa, P.L. Cornelius, and J. Burgueño in 2009, v.49, n.5: 1564-1576. *Crop Science*, v. 50, n. 4, p. 1121-1123, 2010.

YANG, R.-C. et al. Biplot analysis of genotype _ environment interaction: proceed with caution. *Crop Science*, v. 49, n. 5, p. 1564-1576, 2009.

YAN, W. *Crop Variety Trials: Data Management and Analysis*. New York: John Wiley & Sons, 2014. 361 p.

YAN, W et al. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science*, v. 40, n. 3, p. 597-605, 2000.