

Aplicação do modelo AMMI-bayesiano no estudo de estabilidade e adaptabilidade genotípica em dados de mostarda

Application of the bayesian-AMMI model in the study of genotypic stability and adaptability in mustard data

Aplicación del modelo AMMI-bayesiano en el estudio de la estabilidad genotípica y la adaptabilidad en los datos de mostaza

Recebido: 22/07/2020 | Revisado: 04/08/2020 | Aceito: 08/08/2020 | Publicado: 14/08/2020

Alessandra Querino da Silva

ORCID: <https://orcid.org/0000-0002-1702-0169>

Universidade Federal da Grande Dourados, Brasil

E-mail: alessandrasilva@ufgd.edu.br

Luciano Antonio de Oliveira

ORCID: <https://orcid.org/0000-0001-5054-2623>

Universidade Federal da Grande Dourados, Brasil

E-mail: lucianoantonio@ufgd.edu.br

Carlos Pereira da Silva

ORCID: <https://orcid.org/0000-0002-8053-1703>

Universidade Federal de Lavras, Brasil

E-mail: ccpsilva81@hotmail.com

Cristian Tiago Erazo Mendes

ORCID: <https://orcid.org/0000-0003-3486-0212>

Universidade Federal de Lavras, Brasil

E-mail: tiagoerazo@gmail.com

Elias Silva de Medeiros

ORCID: <https://orcid.org/0000-0002-9694-4019>

Universidade Federal da Grande Dourados, Brasil

E-mail: eliasmedeiros@ufgd.edu.br

Thelma Sáfadi

ORCID: <https://orcid.org/0000-0002-4918-300X>

Universidade Federal de Lavras, Brasil

E-mail: safadi@ufla.br

Resumo

A análise de conjuntos de dados provenientes de ensaios multiambientais é de fundamental importância nas fases finais dos programas de melhoramento de plantas. Nesse contexto, o Modelo de Efeitos Principais Aditivos e Interação Multiplicativa (AMMI) se tornou um método popular para avaliar respostas de genótipos em diversos ambientes. No presente trabalho aplicou-se o modelo AMMI, sob o enfoque bayesiano, a um conjunto de dados provenientes de um experimento em blocos casualizados com 12 genótipos (variedades) de mostarda em 6 ambientes distintos. O objetivo foi analisar estabilidade e adaptabilidade genotípica por meio da representação biplot AMMI-2, ressaltando diferenças dessa abordagem em relação a análise AMMI-clássica. Os resultados evidenciaram a grande flexibilidade do método bayesiano para incorporar efeito aleatório para genótipos, bem como inferência ao biplot por meio de regiões de credibilidade para escores genotípicos e ambientais que descrevem o efeito da interação genótipos por ambientes (GEI). As regiões de credibilidade construídas para efeitos principais e parâmetros bilineares permitiram identificar genótipos mais produtivos e visualizar subgrupos homogêneos de genótipos e ambientes em relação ao efeito da GEI. Os genótipos mais produtivos foram G8 e G10 e apenas o G2 foi considerado estatisticamente estável.

Palavras-chave: Interação; Regiões de credibilidade; Biplot.

Abstract

The analysis of data sets from multienvironmental trials is of fundamental importance in the final stages of plant breeding programs. In this context, the Additive Main Effects and Multiplicative Interaction (AMMI) model has become a popular method for evaluating genotype responses in different environments. In the present work, the AMMI model was applied, under the bayesian approach, to a set of data from a randomized block experiment with 12 mustard genotypes (varieties) in 6 different environments. The objective was to analyze genotypic stability and adaptability through the AMMI-2 biplot representation, highlighting differences in this approach in relation to the classical AMMI analysis. The results showed the great flexibility of the bayesian method to incorporate a random effect for genotypes, as well as inference to the biplot through regions of credibility for genotypic and environmental scores that describe the effect of the interaction between genotypes by environments (GEI). The regions of credibility built for main effects and bilinear parameters allowed to identify more productive genotypes and to visualize homogeneous subgroups of

genotypes and environments in relation to the effect of GEI. The more productive genotypes were G8 and G10 and only G2 was considered statistically stable.

Keywords: Interaction; Credibility region; Biplot.

Resumen

El análisis de conjuntos de datos de pruebas multi-ambientales es de fundamental importancia en las etapas finales de los programas de fitomejoramiento. En este contexto, el Modelo de Efectos Aditivos Principales e Interacción Multiplicativa (AMMI) se ha convertido en un método popular para evaluar las respuestas de genotipo en diferentes entornos. En el presente trabajo, el modelo AMMI se aplicó, bajo el enfoque bayesiano, a un conjunto de datos de un experimento de bloques al azar con 12 genotipos (variedades) de mostaza en 6 entornos diferentes. El objetivo fue analizar la estabilidad genotípica y la adaptabilidad a través de la representación biplot AMMI-2, destacando las diferencias en este enfoque en relación con el análisis AMMI clásico. Los resultados mostraron la gran flexibilidad del método bayesiano para incorporar un efecto aleatorio para los genotipos, así como la inferencia al biplot a través de regiones de credibilidad para puntajes genotípicos y ambientales que describen el efecto de la interacción entre genotipos por medio ambiente (GEI). Las regiones de credibilidad creadas para los efectos principales y los parámetros bilineales permitieron identificar genotipos más productivos y visualizar subgrupos homogéneos de genotipos y entornos en relación con el efecto de GEI. Los genotipos más productivos fueron G8 y G10 y solo G2 se consideró estadísticamente estable.

Palabras clave: Interacción; Regiones de credibilidad; Biplot.

1. Introdução

Muitas pesquisas têm sido realizadas abordando o Modelo de Efeitos Principais Aditivos e Interação Multiplicativa (AMMI). Esse procedimento de análise se destacou pela ampla aplicabilidade em programas de melhoramento de plantas no estudo da interação entre genótipos e ambientes (GEI). O método permite isolar o padrão (mais diretamente associado a interação) e descartar o ruído, melhorando a capacidade preditiva do modelo. Além disso, os escores que descrevem a interação podem ser representados graficamente em biplots, que são considerados formas simples de interpretação e identificação de padrões em conjunto de dados (Gabriel, 1971; Kempton, 1984; Zobel et al., 1988).

Uma das principais dificuldades em se trabalhar com o modelo AMMI tradicional

(modelo com efeitos fixos) é a incorporação de inferência à representação biplot dos parâmetros que descrevem a GEI (Yang et al., 2009; Yan et al., 2010). Algumas propostas para contornar essa dificuldade têm sido relatadas na literatura. Denis e Gower (1994), por exemplo, propuseram a construção de regiões de confiança assintóticas para os escores genotípicos e ambientais, ao passo que Yang et al. (2009) atacaram esse problema por meio de procedimentos bootstrap. Contudo, esses métodos são suscetíveis a críticas, sejam por suposições restritivas dos métodos paramétricos ou ainda pelos processos de reamostragem problemáticos nas linhas e colunas da matriz da GEI em procedimentos não paramétricos (Yan et al., 2010; Oliveira et al., 2015).

Essas dificuldades têm sido contornadas por meio da inferência bayesiana, com a implementação de regiões de credibilidade no biplot AMMI-2, que permitem identificar genótipos e ambientes estáveis, bem como subgrupos homogêneos de genótipos (ou ambientes) com relação ao efeito da interação (Crossa et al., 2011; Perez-Elizalde et al., 2012; Oliveira et al., 2015; Silva et al., 2015; Jarquin et al., 2016). Aplicações e desenvolvimentos recentes são baseados nos trabalhos pioneiros de Viele e Srinivasan (2000) e Liu (2001) que mostraram como contornar as dificuldades na amostragem dos parâmetros bilineares.

Além de proporcionar um método paramétrico para incorporar inferência ao biplot, a abordagem bayesiana oferece vantagens sugestivas, como a flexibilidade para lidar com dados heterocedásticos e desbalanceados, além de permitir a incorporação de informações adicionais à análise (Perez-Elizalde et al., 2012; Silva et al., 2019).

Nesse trabalho, utilizou-se o modelo AMMI em sua versão bayesiana (AMMI-bayesiano) para analisar adaptabilidade e estabilidade de genótipos de mostarda avaliados em diferentes ambientes.

2. Material e métodos

O conjunto de dados multiambientais (MET) utilizado neste trabalho foi extraído do *Indian Agricultural Statistics Research Institute* (IASRI, 2014) e corresponde a 12 variedades (G1-G12) de mostarda que foram avaliadas em seis ambientes (A1-A6) distintos, em blocos casualizados com três repetições.

O modelo AMMI em notação vetorial é dado por:

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

sendo \mathbf{y} o vetor de respostas fenotípicas de dimensão $n \times 1$. As matrizes \mathbf{X}_1 , \mathbf{X}_2 e \mathbf{Z} são de delineamentos e $\boldsymbol{\beta}_{a \times 1}$ e $\mathbf{g}_{g \times 1}$ são vetores de efeitos principais de ambientes e de genótipos, respectivamente. Os termos λ_k , $\boldsymbol{\alpha}_k$ e $\boldsymbol{\gamma}_k$ representam, respectivamente, o k -ésimo valor singular e os k -ésimos vetores singulares genotípicos e ambientais. O vetor de erros experimentais $\boldsymbol{\epsilon}$ é admitido ter distribuição de probabilidade normal multivariada com vetor de média zero e matriz de variância e covariâncias $\mathbf{I}\sigma_e^2$ (sendo \mathbf{I}_n a matriz identidade).

Nesse contexto, o vetor de observações \mathbf{y} tem distribuição condicional normal multivariada, $\mathbf{y} | \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda, \mathbf{g}, \boldsymbol{\beta}, \sigma_e^2 \sim N(\boldsymbol{\theta}, \mathbf{I}_n \sigma_e^2)$, com matriz de variâncias $\mathbf{I}_n \sigma_e^2$, sendo σ_e^2 a variância residual e $\boldsymbol{\theta} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k$ o vetor de médias.

Densidades a priori

As densidades a priori para os parâmetros são as mesmas utilizadas por Oliveira et al. (2015):

$$\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \sigma_\beta^2 \sim N(0, \sigma_\beta^2); \sigma_\beta^2 \rightarrow \infty \text{ implica que } \boldsymbol{\beta} \sim \text{constante};$$

$$\mathbf{g} | \boldsymbol{\mu}_g, \sigma_g^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_g^2);$$

$$\sigma_g^2 \sim \frac{1}{\sigma_g^2};$$

$$\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2 \sim N^+(0, \sigma_{\lambda_k}^2); \sigma_{\lambda_k}^2 \rightarrow \infty, \text{ por conseguinte, } \lambda_k \sim \text{constante};$$

$$\boldsymbol{\alpha}_k \sim \text{uniforme esférica no subespaço corrigido};$$

$$\boldsymbol{\gamma}_k \sim \text{uniforme esférica no subespaço corrigido};$$

$$\sigma_e^2 \sim \frac{1}{\sigma_e^2}.$$

Distribuição conjunta a posteriori e distribuições condicionais completas

Considera-se inicialmente a distribuição de verossimilhança que é apresentada por:

$$L(\boldsymbol{\theta}, \sigma_e^2 | \mathbf{y}) = p(\mathbf{y} | \boldsymbol{\theta}, \sigma_e^2) = \frac{1}{(2\pi)^{\frac{n}{2}} |I\sigma_e^2|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\sigma_e^2} (\mathbf{y} - \boldsymbol{\theta})^\top (\mathbf{y} - \boldsymbol{\theta})\right\} \quad (2)$$

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

A distribuição conjunta a posteriori é obtida combinando a verossimilhança dos dados com as distribuições a priori via teorema de Bayes da seguinte forma:

$$p(\boldsymbol{\Phi} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}, \sigma_e^2) p(\mathbf{g} | \boldsymbol{\mu}_g, \sigma_g^2) p(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \sigma_\beta^2) p(\sigma_g^2 | \nu_g, S_g^2) \times \\ \times p(\sigma_e^2 | \nu_e, S_e^2) \prod_{k=1}^t p(\lambda_k | \boldsymbol{\mu}_{\lambda_k}, \sigma_{\lambda_k}^2) p(\boldsymbol{\alpha}_k) p(\boldsymbol{\gamma}_k) \quad (3)$$

em que $\boldsymbol{\Phi} = (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{g}, \boldsymbol{\beta}, \sigma_g^2, \sigma_e^2)$. Sob as hipóteses a priori estabelecidas, a equação (3) pode ser reescrita como:

$$p(\boldsymbol{\Phi} | \mathbf{y}) \propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_e^2} (\mathbf{y} - \boldsymbol{\theta})^\top (\mathbf{y} - \boldsymbol{\theta})\right\} (\sigma_g^2)^{-\frac{n_g}{2}} \exp\left\{-\frac{1}{2\sigma_g^2} \mathbf{g}^\top \mathbf{g}\right\} \frac{1}{\sigma_g^2} \frac{1}{\sigma_e^2} \quad (4).$$

As distribuições condicionais completas a posteriori para os parâmetros são obtidas a partir da distribuição a posteriori conjunta (equação 4) por manipulações algébricas. Aqui apresentamos apenas as densidades condicionais, detalhes algébricos podem ser consultados em Oliveira et al. (2015) e Silva et al. (2015), sendo:

$$\boldsymbol{\beta} | \dots \sim N\left[(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top (\mathbf{y} - \mathbf{Z}\mathbf{g} - \boldsymbol{\Theta}), (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \sigma_e^2\right] \quad \text{em que} \quad \boldsymbol{\Theta} = \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k$$

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

$$\sigma_g^2 | \dots \sim \text{inv-escalada} - \chi^2[n_g, \mathbf{g}^\top \mathbf{g}]$$

$$\sigma_e^2 | \dots \sim \text{inv-escalada} - \chi^2[n, (\mathbf{y} - \boldsymbol{\theta})^\top (\mathbf{y} - \boldsymbol{\theta})]$$

$$\lambda_k | \dots \sim N^+ \left[\left(\phi_k^\top \phi_k \right)^{-1} \phi_k^\top \Delta_k, \left(\phi_k^\top \phi_k \right)^{-1} \sigma_e^2 \right]$$

com $\Delta_k = \mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g} - \sum_{k' \neq k}^t \lambda_{k'} \text{diag}(\mathbf{Z} \boldsymbol{\alpha}_{k'}) \mathbf{X}_2 \boldsymbol{\gamma}_{k'}$, $\phi = \text{diag}(\mathbf{Z} \boldsymbol{\alpha}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k$ e $\lambda_1 \geq \dots \geq \lambda_t \geq 0$

$$p(\boldsymbol{\alpha}_k | \dots) \propto \exp \left\{ \frac{\lambda_k}{\sigma_e^2} \left[\boldsymbol{\alpha}_k^\top \boldsymbol{\Lambda}_k^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g}) \right] \right\} \text{ que é proporcional a uma von Mises-}$$

Fisher $\boldsymbol{\alpha}_k | \dots \sim \text{VMF}[k, \boldsymbol{\mu}_{\alpha_k}]$ com parâmetro de concentração λ_k / σ_e^2 e média direcional

$$\boldsymbol{\mu}_{\alpha_k} = \boldsymbol{\Lambda}_k^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g}) \text{ sendo } \boldsymbol{\Lambda}_k = \text{diag}(\mathbf{X}_2 \boldsymbol{\gamma}_k) \mathbf{Z}.$$

De forma análoga, $\boldsymbol{\gamma}_k | \dots \sim \text{VMF}[k, \boldsymbol{\mu}_{\gamma_k}]$ com média direcional $\boldsymbol{\mu}_{\gamma_k} = \boldsymbol{\Omega}_k^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g})$ e $\boldsymbol{\Omega}_k = \text{diag}(\mathbf{Z} \boldsymbol{\alpha}_k) \mathbf{X}_2$.

Condicionais completas para os vetores singulares no subespaço corrigido

Uma das principais limitações para a aplicação do método bayesiano é a amostragem dos vetores singulares devido a restrição de ortogonalidade imposta pela decomposição por valores singulares (DVS), fazendo com que o suporte da distribuição a posteriori não seja trivial. A solução para esse problema foi dada por Viele e Srinivasan (2000) que conduziram a amostragem via transformação linear ortogonal.

Um vetor ortogonal a t vetores na esfera unitária de raio um em \mathbb{R}^n pode ser transformado no mesmo vetor em \mathbb{R}^{n-t} , sendo amostrado sem a restrição de ortogonalidade. Essa transformação é bijetora (um a um), portanto a transformação inversa coloca o vetor no subespaço correto em \mathbb{R}^n .

Para o modelo considerado, $\boldsymbol{\alpha}_k$ deve ser ortogonal aos outros $\boldsymbol{\alpha}_{k's}$ que compõem as colunas da matriz $\mathbf{A}_{-\alpha_k}$. Logo, existe uma matriz \mathbf{H}_k tal que suas colunas formam um conjunto ortonormal de vetores, que são ortogonais a $\mathbf{A}_{-\alpha_k}$ (Liu, 2001). A distribuição condicional completa a posteriori para o vetor no subespaço corrigido $\boldsymbol{\alpha}_k^* = \mathbf{H}_k^\top \boldsymbol{\alpha}_k$ é dado por

$$\boldsymbol{\alpha}_k^* | \dots \sim \text{VMF} \left(r - t, \frac{c_k \lambda_k}{\sigma_e^2}, \boldsymbol{\mu}_{\alpha_k} \right) \quad (5)$$

em que VMF é a distribuição von Mises-Fisher sendo $\tilde{\boldsymbol{\mu}}_{\alpha_k} = c_k^{-1} \mathbf{H}_k^\top \boldsymbol{\mu}_{\alpha_k}$ e $c_k = \sqrt{(\mathbf{H}_k^\top \boldsymbol{\mu}_{\alpha_k})^\top \mathbf{H}_k \boldsymbol{\mu}_{\alpha_k}}$. Então, pode-se amostrar $\boldsymbol{\alpha}_k^*$ em \mathbb{R}^{r-t} e para obter o vetor no correto subespaço em \mathbb{R}^r basta realizar a transformação inversa dada por $\boldsymbol{\alpha}_k = \mathbf{H}_k \boldsymbol{\alpha}_k^*$.

Com cálculos análogos aos realizados para $\boldsymbol{\alpha}_k$ obtém-se $\boldsymbol{\gamma}_k^*$ tal que $\boldsymbol{\gamma}_k^* = \mathbf{R}_k^\top \boldsymbol{\gamma}_k$ sendo a posteriori condicional no subespaço corrigido dada por:

$$\boldsymbol{\gamma}_k^* | \dots \sim VMF \left(c-t, \frac{d_k \lambda_k}{\sigma_e^2}, \tilde{\boldsymbol{\mu}}_{\boldsymbol{\gamma}_k} \right) \quad (6)$$

com $\tilde{\boldsymbol{\mu}}_{\boldsymbol{\gamma}_k} = d_k^{-1} \mathbf{R}_k^\top \boldsymbol{\mu}_{\boldsymbol{\gamma}_k}$ e $d_k = \sqrt{(\mathbf{R}_k^\top \boldsymbol{\mu}_{\boldsymbol{\gamma}_k})^\top \mathbf{R}_k \boldsymbol{\mu}_{\boldsymbol{\gamma}_k}}$. Sendo a matriz \mathbf{R}_k de dimensão $c \times (c-t)$ com colunas ortonormais e ortogonais aos vetores $\boldsymbol{\gamma}_{k's}$ que compõe as colunas da matriz $\mathbf{D}_{-\boldsymbol{\gamma}_k}$. A transformação inversa é obtida por $\boldsymbol{\gamma}_k = \mathbf{R}_k \boldsymbol{\gamma}_k^*$.

Amostragem e inferência

Para obter estimativas pontuais e intervalares das quantidades desconhecidas de interesse foi utilizado um procedimento numérico estocástico baseado no método de Monte Carlo via Cadeias de Markov (MCMC). Como todas as condicionais são conhecidas e permitem amostragem direta utilizou-se um amostrador de Gibbs (Oliveira et al., 2015).

A convergência das cadeias foi verificada pelos critérios de Raftery e Lewis (1992) e Heidelberger e Welch (1983), implementados no pacote Boa (Smith, 2007).

Estimativas pontuais foram obtidas por médias a posteriori das cadeias construídas e estimativas intervalares foram implementadas por regiões a 95% de credibilidade com Máxima Densidade a Posteriori (HPD) (Box & Tiao, 1973) pelo método proposto por Chen e Shao (1999).

Além disso, regiões de credibilidade bivariadas para os escores genotípicos $(\boldsymbol{\alpha}_1 \sqrt{\lambda_1}, \boldsymbol{\alpha}_2 \sqrt{\lambda_2})$ e ambientais $(\boldsymbol{\gamma}_1 \sqrt{\lambda_1}, \boldsymbol{\gamma}_2 \sqrt{\lambda_2})$ foram incorporadas ao biplot utilizando o procedimento proposto por Ooms (2009).

Todas as análises foram realizadas utilizando o software estatístico R (R Core Team, 2018).

3. Resultados e Discussão

Na Tabela 1 é apresentada a análise conjunta de experimentos. Observa-se que todos os fatores são significativos ao nível de significância de 1%. Em particular, o fato de a interação ser significativa implica na não aditividade do modelo em relação aos efeitos principais e conclusões mais gerais não podem ser estabelecidas (Duarte & Vencovsky, 1999).

Tabela 1. Análise conjunta de experimentos de 6 genótipos de mostardas avaliados em 12 ambientes

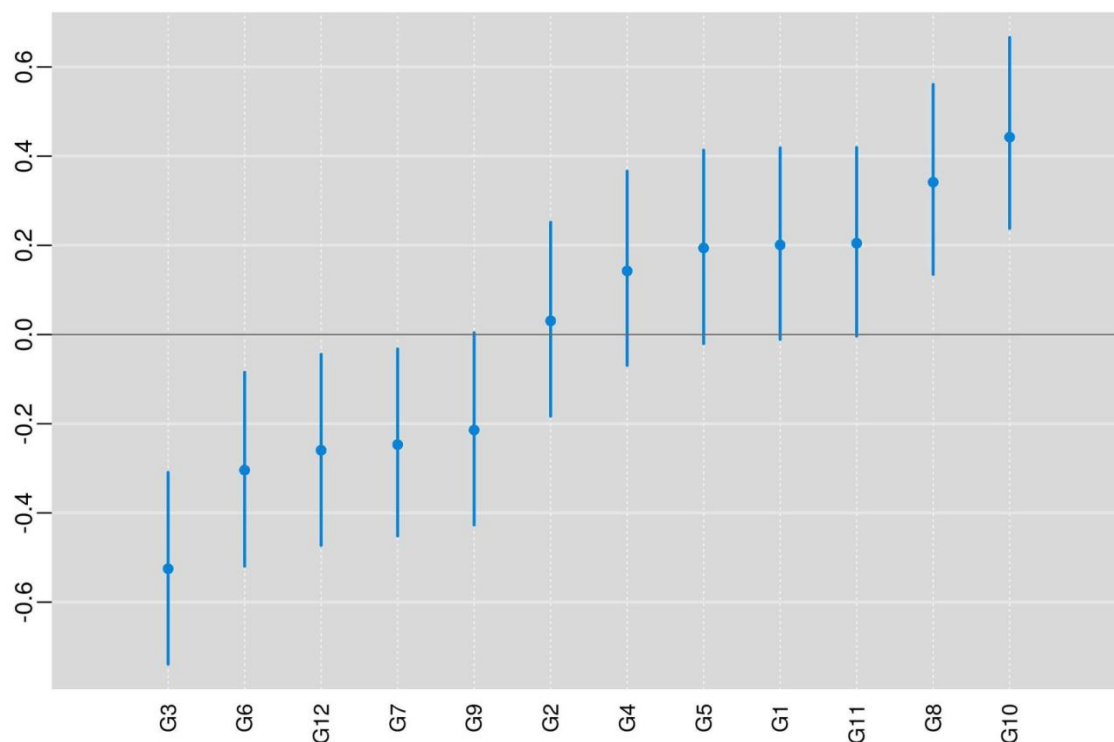
Fontes de Variação	G.L	S.Q.	Q.M.	Valor F	Valor-p
Ambientes	5	52,7940	10,5588	19,3274	$4,69 \times 10^{-11}$ *
Genótipos	11	18,9875	1,7261	3,1596	0,00226*
Blocos (Amb.)	12	2,5605	0,2134	5,0057	$8,22 \times 10^{-07}$ *
Interação	55	30,0471	0,5463	12,8163	$1,39 \times 10^{-32}$ *
Resíduo	132	5,6267	0,0426		
Total	2015	215	110,0158		

*significativo a 1% de probabilidade. Fonte: Autores.

Utilizando o critério de Raftery e Lewis (1992) observou-se que todos os parâmetros tiveram fator de dependência $I < 5$. Além disso, todos passaram pelo teste de estacionariedade (Heidelberger & Welch, 1983), indicando que a convergência foi alcançada. Para o processo de inferência, foi selecionada uma amostra de tamanho $n = 8000$ observações após a queima (*burn-in*) das primeiras 1000 observações e saltos (*thinning*) a cada 4 observações na coleta da amostra.

As médias a posteriori e os intervalos HPD a 95% de credibilidade para efeitos principais de genótipos, são visualizados na Figura 1. As médias estão ranqueadas em ordem decrescente da direita para esquerda e sobreposições das regiões HPD indicam efeitos semelhantes. Entretanto, percebe-se que G8 e G10 possuem efeitos principais maiores que zero.

Figura 1. Regiões de credibilidade HPD a 95% de probabilidade para o efeito principal de genótipos



Fonte: Autores.

Uma diferença fundamental entre as versões clássica e bayesiana para o efeito de genótipo é que as médias a posteriori têm valores encolhidos quando comparados as estimativas de quadrados mínimos ordinários (OLS). A priori hierárquica utilizada em dois níveis para genótipos faz com tenham efeitos aleatórios, semelhante ao que acontece em um modelo misto. Considerar genótipos como aleatórios oferece muitas vantagens como, por exemplo, determinar componentes de variância e incorporar matrizes com coeficientes de parentesco. Além disso, conduz a predições mais realistas, permitindo inferir sobre o valor genotípico das cultivares livres dos efeitos ambientais (Smith et al., 2005; Resende & Duarte, 2007).

Na Tabela 2 são apresentadas médias a posteriori e regiões HPD a 95% de credibilidade para valores singulares e também estimativas frequentistas em relação a abordagem AMMI clássica obtidas por DVS da matriz da GEI. Observam-se valores ligeiramente menores para médias a posteriori quando comparadas com soluções DVS.

Tabela 2. Estimativas obtidas por DVS, médias a posteriori e limites inferiores (LI) e superiores (LS) para regiões HPD a 95% de credibilidade para valores singulares.

AMMI-fixo	AMMI-bayesiano			
	Estimativas DVS	Médias a posteriori	LI	LS
	2,4640	2,4204	2,1667	2,6522
	1,5737	1,5119	1,2717	1,7504
	0,9318	0,8386	0,5773	1,0953
	0,6368	0,4924	0,2035	0,7715
	0,4409	0,2555	0,0008	0,4768

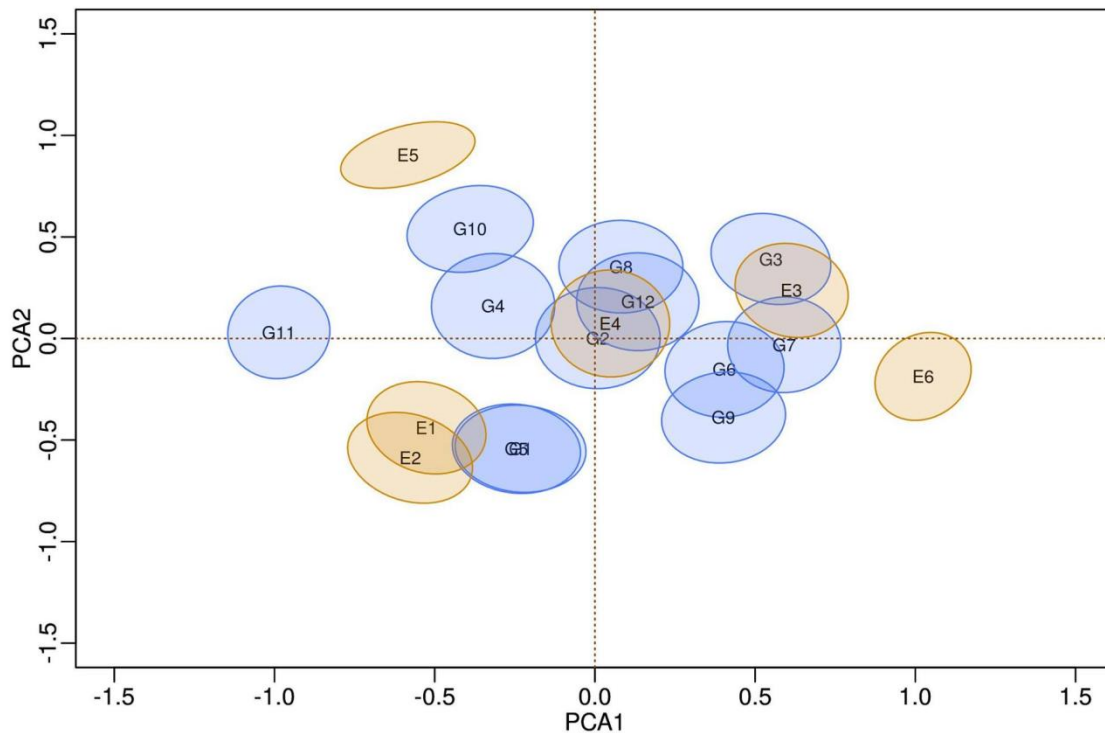
Fonte: Autores.

Efeitos de encolhimento em relação à abordagem convencional têm sido observados em diversos trabalhos, mesmo utilizando-se priori não informativas para a variância de λ (Liu, 2001; Crossa et al., 2011). Se prioris específicas para componentes de variância dos valores singulares fossem utilizadas era de se esperar encolhimentos ainda mais pronunciados para as estimativas, como, por exemplo, aqueles encontrados por Silva et al. (2015).

Os percentuais da soma de quadrado total da interação, recuperados pelos dois primeiros eixos principais, são de aproximadamente 85% e 90% para AMMI clássico e bayesiano, respectivamente.

Em termos de seleção e recomendação, além do ranqueamento das médias a posteriori, como apresentado na Figura 1, é necessário ainda o estudo da estabilidade e adaptabilidade, que é realizado pela interpretação de escores genotípicos e ambientais no biplot. A representação biplot AMMI-2, considerando os dois primeiros componentes principais, juntamente com regiões de credibilidade elípticas a 95% de probabilidade é apresentada na Figura 2.

Figura 2. Regiões elípticas a 95% de credibilidade para escores genotípicos e ambientais no biplot em relação aos dois primeiros eixos principais.



Fonte: Autores.

A interpretação do biplot é realizada de acordo com propriedades do produto interno entre vetores determinados pelos escores. Regiões de credibilidade que englobam a origem do sistema (0,0) indicam que os respectivos genótipos e/ou ambientes não possuem contribuições significativas para a GEI sendo referidos como estáveis (Crossa et al., 2011; Perez-Elizalde et al., 2012). Nesse sentido, para identificar genótipos estáveis ou sugerir subgrupos homogêneos (semelhantes) no que se refere ao efeito da GEI basta observar posições e sobreposições entre as regiões no biplot (Oliveira et al., 2017; Júnior et al., 2018).

O genótipo G2 é dito ser estável, já que a região de credibilidade para o respectivo escore genotípico engloba a origem do biplot e o mesmo pode ser dito sobre o ambiente E4. Por outro lado, o genótipo G11 e os ambientes E1, E2 e E6 são aqueles que mais contribuem para o efeito da GEI, pois apresentam maior distância em relação a origem.

Analisando simultaneamente o ranqueamento das médias a posteriori (Figura 1) e o padrão da GEI no biplot (Figura 2) é possível observar que os genótipos G8 e G10, cujas regiões HPD para efeitos principais possuem apenas valores positivos, não seriam amplamente adaptados a todos os ambientes, já que pela análise biplot são considerados instáveis (regiões de credibilidade não englobam a origem) e, portanto, possuem

contribuições expressivas para GEI. Por outro lado, o genótipo G2, considerado estável, não apresenta um desempenho interessante no que se refere ao efeito principal.

Subgrupos de genótipos e/ou ambientes (interseção entre as elipses) que possuem características semelhantes com relação ao efeito da interação (subgrupos homogêneos) são identificados observando as sobreposições entre as regiões. Nesse contexto, destacam-se os seguintes subgrupos homogêneos: {G1, G5}, {G3}, {G4, G10}, {G11}, {G8, G12}, {G6, G7, G9}, {E1, E2}, {E3}, {E4}, {E5} e {E6}. Além disso, adaptabilidade de genótipos a ambientes específicos pode ser observada no biplot.

É possível sugerir adaptabilidade dos genótipos G3, G6, G7 e G9 aos ambientes E3 e E6. De modo análogo, {G4, G10} ao ambiente E5 e {G1, G5} ao subgrupo {E1, E2}. Os genótipos G8 e G12 poderiam ser indicados para E4 e embora não se observe sobreposições destes com o ambiente E3 parece razoável sugerir uma combinação positiva devido ao fato de G8, G12 e E3 terem a maior parte da área de suas regiões dentro do primeiro quadrante.

Alguns questionamentos poderiam ser feitos sobre essas interpretações. Salienta-se que decisões, várias vezes, são determinadas por questões práticas (experiência do melhorista). Denis e Gower (1994), por exemplo, sugeriram subgrupos homogêneos de ambientes e genótipos mesmo diante de complexas sobreposições. Perez-Elizalde et al. (2012), por outro lado, parecem ter contornado essa dificuldade por meio de um algoritmo de análise de cluster hierárquico. Na análise clássica, as interpretações são mais simples, pois os escores são representados apenas por pontos no plano. Contudo, essa representação se configura apenas como uma ferramenta descritiva já que nenhuma medida da incerteza é considerada.

4. Conclusões

O modelo AMMI-bayesiano permitiu identificar os genótipos mais produtivos, bem como separar subgrupos homogêneos de genótipos e ambientes com relação ao efeito da GEI utilizando níveis de credibilidade. Como apontado, o método é flexível para incorporar efeitos aleatórios aos parâmetros (em nosso caso para genótipos) e regiões de credibilidade permitem identificar aqueles estatisticamente mais produtivos, ou com rendimento acima da média geral (significativamente diferente de zero) que foram os genótipos {G8, G10}.

Por outro lado, com base no biplot, apenas G2 foi considerado estatisticamente estável. Além disso, foi possível sugerir adaptabilidade de genótipos a ambientes específicos com incerteza dada pelas regiões de credibilidade bivariadas para escores bilineares. Em uma

análise AMMI convencional, essas conclusões têm sido consideradas apenas de forma descritiva.

Referências

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. New York: John Wiley.

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.

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>

Denis, J. B., & Gower, J. C. (1994). Asymptotic covariances for parameters of biadditive models. *Utilitas Mathematica*, v. 46, 193-205.

Duarte, J. B., & Vencovsky, R. (1999). *Interação genótipos × ambientes: uma introdução à análise "AMMI"*. Ribeirão Preto: Sociedade Brasileira de Genética.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal components analysis. *Biometrika*, 58(3), 453-467.

Heidelberger, P., & Welch, P. (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>

Indian Agricultural Statistics Research Institute. IASRI. (2014). Recuperado de http://www.iasri.res.in/design/Analysis%20of%20data/combined_anlysis_rcbd.html

Jarquin, D., Perez-Elizalde, S., Burgueño, J., & Crossa, J. (2016). A hierarchical Bayesian estimation model for multi-environment plant breeding trials in successive years. *Crop Science*, 56(5), 2260-2276. <https://doi.org/10.2135/cropsci2015.08.0475>

Júnior, L. A. Y. B., Silva, C. P., Oliveira, L. A., Nuvunga, J. J., Pires, L. P. M., Pinho, R. G. V., & Balestre, M. (2018). AMMI Bayesian Models to Study Stability and Adaptability in Maize. *Agronomy Journal*, 110(5), 1765-1776. <https://doi.org/10.2134/agronj2017.11.0668>

Kempton, R. A. (1984). The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science*, 103(1), 123-135. <https://doi.org/10.1017/S0021859600043392>

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

Oliveira, L. A., Silva, C. P., Nuvunga, J. J., 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>

Oliveira, L. A., Silva, C. P., Teodoro, P. E., Torres, F. E., Corrêa, A. M., & Bhering, L. L. (2017). Performance of Cowpea Genotypes in the Brazilian Midwest Using the Bayesian Additive Main Effects and Multiplicative Interaction Model. *Agronomy Journal*, 110(1), 147-154. <https://doi.org/10.2134/agronj2017.03.0183>

Ooms, J. C. L. (2009). *The highest posterior density posterior prior for Bayesian model selection*. (Master's thesis). <https://dspace.library.uu.nl/handle/1874/34234>

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>

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

Raftery, A. E., & Lewis, S. (1992). How many iterations in the Gibbs sampler? In: Bernardo, J. M. et al. (Ed.), *Bayesian statistics*, 763-773. Oxford: Oxford University.

Resende, M. D. V., & Duarte, J. B. (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesquisa Agropecuária Tropical*, 37(3), 182-194. Recuperado de <https://www.revistas.ufg.br/pat/article/view/1867>.

Silva, C. P., Oliveira, L. A., Nuvunga, J. J., Pamplona, A. K. A., & Balestre, M. (2015). A Bayesian Shrinkage approach for AMMI Models. *PLoS One*, 10(7), 1-27. <https://doi.org/10.1371/journal.pone.0131414>

Silva, C. P., 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>

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. *Journal of Agricultural Science*, 143(6), 449-462. <https://doi.org/10.1017/S0021859605005587>

Smith, B. J. (2007). Boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. *Journal of Statistical Software*, 21(11), 1-37.

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](https://doi.org/10.1016/S0378-3758(99)00151-2)

Yan, W., Glover, K. D., & Kang, M. S. (2010). 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 49, 1564-1576. *Crop Science*, 50(4), 1121-1123. <https://doi.org/10.2135/cropsci2010.01.00011e>

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>

Zobel, R. W., Wright, M. J., & Gauch, H. G. (1988). Statistical analysis of a yield trial. *Agronomy Journal*, 80(3), 388-393. <https://doi.org/10.2134/agronj1988.00021962008000030002x>

Porcentagem de contribuição de cada autor no manuscrito

Alessandra Querino da Silva – 30%

Luciano Antonio de Oliveira – 20%

Carlos Pereira da Silva – 20%

Cristian Tiago Erazo Mendes – 10%

Elias Silva de Medeiros – 10%

Thelma Sáfadi – 10%