



LUCIANO ANTONIO DE OLIVEIRA

**MODELOS GGE: ESTIMADORES DE ENCOLHIMENTO E
HETEROCEDASTICIDADE**

LAVRAS – MG

2021

LUCIANO ANTONIO DE OLIVEIRA

**MODELOS GGE: ESTIMADORES DE ENCOLHIMENTO E
HETEROCEDASTICIDADE**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Prof. Dr. Márcio Balestre
(in memorian)

Orientador

Prof. Dr. Carlos Pereira da Silva

Coorientador

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho

Coorientador

LAVRAS – MG

2021

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Oliveira, Luciano Antonio de.

Modelos GGE: estimadores de encolhimento e heterocedasticidade / Luciano Antonio de Oliveira. - 2021.

173 p.

Orientador(a): Márcio Balestre.

Coorientador(a): Carlos Pereira da Silva, Júlio Sílvio de Sousa Bueno Filho.

Tese (doutorado) - Universidade Federal de Lavras, 2021.

Bibliografia.

1. Inferência bayesiana. 2. Modelos lineares-bilineares. 3.

Amostragem Gibbs. I. Balestre, Márcio. II. Silva, Carlos Pereira da. III. Bueno Filho, Júlio Sílvio de Sousa. IV. Título.

LUCIANO ANTONIO DE OLIVEIRA

**MODELOS GGE: ESTIMADORES DE ENCOLHIMENTO E
HETEROCEDASTICIDADE**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 18 de fevereiro de 2021.

Prof. Dr. Fernando Ribeiro Cassiano UNIMONTES
Prof. Dra. Kelly Cristina Mota Gonçalves UFRJ
Prof. Dr. Renzo Garcia Von Pinho UFLA
Profa. Dra. Thelma Sáfadi UFLA

Prof. Dr. Márcio Balestre
Orientador

Prof. Dr. Carlos Pereira da Silva
Coorientador

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho
Coorientador

**LAVRAS – MG
2021**

*Aos meus pais,
Onofre Maria de Oliveira e
Sebastiana Silva de Oliviera,
Por todo amor e carinho que recebi durante a realização desse trabalho.*

DEDICO

AGRADECIMENTOS

Primeiramente a Deus, presente em todos os momentos nesses quatro anos de doutorado.

À minha esposa Alessandra Querino, pelo carinho, compreensão e auxílio na realização deste sonho.

Aos meus irmãos Rita e José Mário, pelo companheirismo e acolhimento familiar.

À minha sobrinha Giovana, pelas orações e fiel torcida.

Aos colegas do curso de doutorado. Em especial ao Cristian Tiago, pela amizade, ajuda e parceria em estudos e pesquisas no decorrer dessa jornada.

Aos meus coorientadores, Carlos Pereira da Silva, grande companheiro de batalhas, cuja contribuição foi fundamental na conclusão desse projeto e professor Júlio Sílvio de Sousa Bueno Filho, que em um momento difícil se colocou, prontamente, à disposição para me auxiliar e com quem tive a oportunidade de aprender muito.

Ao saudoso professor Márcio Balestre, pela orientação e ensinamentos valiosos. Seu entusiasmo e determinação eram contagiantes. Sua originalidade e contribuições para a Ciência sempre estarão presentes.

À Universidade Federal da Grande Dourados (UFGD) pela concessão do afastamento para cursar o doutorado e ao apoio financeiro para a divulgação dos resultados desta tese no âmbito do Programa de Apoio à Pesquisa (PAP).

À Universidade Federal de Lavras (UFLA) e ao Departamento de Estatística (DES) pela oportunidade de cursar o doutorado.

*“Não a nós, Senhor,
nenhuma glória para nós,
mas sim ao teu nome,
por teu amor e por tua fidelidade”
(Salmos 115:1)*

RESUMO

Modelos multiplicativos (ou lineares-bilineares) são úteis em distintas áreas do conhecimento para analisar dados organizados em tabelas de dupla entrada com dois fatores e sua interação sendo estudados. Isso é particularmente importante nas fases finais de programas de melhoramento de plantas, em que, via de regra, vários genótipos são avaliados em diferentes ambientes e a interação genótipos×ambientes (GEI), geralmente, dificulta a seleção e recomendação ampla de cultivares superiores. Dentre esses modelos, o de efeitos principais de genótipos (G) mais GEI, referido na literatura como GGE ou SERG (*Sites “environments” Regression Model*) merece especial destaque devido à ampla aplicabilidade por pesquisadores e melhoristas na análise de dados resultantes de experimentos multiambientais (MET), especialmente a sua interpretação gráfica que considera apenas os dois primeiros componentes principais denominada GGE biplot. Muitos autores têm apontado as vantagens obtidas pela aplicação da inferência bayesiana nestes modelos em substituição à análise padrão, ou clássica, que considera os parâmetros como sendo de efeitos fixos (frequentista). Tanto o encolhimento de estimativas dos parâmetros que modelam a GEI quanto modelos que contemplam a heterogeneidade de variâncias entre locais foram tratados diretamente sob a metodologia bayesiana para o modelo AMMI (*Additive Main Effects and Multiplicative Interaction*), mas ainda não foram discutidos para modelos GGE. O principal objetivo desta tese foi estender a versão bayesiana do modelo GGE em dois sentidos: a) estudar distribuições *a priori* de máxima entropia para o encolhimento dos efeitos G+GEI e b) implementar a versão heterocedástica. Essas propostas foram verificadas em cenários simulados e exemplificadas com dados reais. Os resultados mostraram que a inferência bayesiana, com distribuições *a priori* de máxima entropia para componentes de variância dos valores singulares, flexibiliza o modelo GGE. Além disso, considerando variância específica em cada local obtém-se um melhor ajuste do modelo, permitindo, por outro lado, avaliar os genótipos com diferentes precisões, recuperando informações experimentais presentes em diferentes ensaios a partir da representação gráfica biplot. Ficou também evidenciada a versatilidade da modelagem bayesiana para incorporar inferência ao biplot (atribuindo regiões de credibilidade para ambiente médio, genótipo ideal, correlações entre ambientes e resumos do tipo “quem venceu onde”) difíceis de serem obtidas nas análises biplot usuais do modelo GGE. Os métodos aqui discutidos são bastante promissores e estão sendo implementados programas para disponibilizá-los aos pesquisadores.

Palavras-chave: Amostragem Gibbs. Inferência bayesiana. Modelos lineares-bilineares. Máxima Entropia.

ABSTRACT

Multiplicative (or linear-bilinear) models are useful in different areas of knowledge to analyze data in two-way tables in which both factors and their interaction are studied. This is particularly important in the final stages of plant breeding programs, where several genotypes are evaluated in different environments and the interaction between genotypes \times environments (GEI) generally makes it difficult to select and widely recommend cultivars. Among these models, the Main Effects of Genotypes (G) plus GEI, referred to in the literature as GGE or SERG (Sites “environments” Regression Model) has wide applicability by researchers and breeders in the analysis of data resulting from multi-environmental experiments (MET). Of main relevance is its graphical interpretation that considers only the first two main components called GGE biplot. Many authors have pointed out the advantages of applying Bayesian inference in these models, avoiding standard analysis which considers the parameters to be fixed effects. Shrinking estimates of the parameters that model the GEI and the heterogeneity of variances between locations were modeled directly under the Bayesian methodology for the AMMI model (Additive Main Effects and Multiplicative Interaction), another model belonging to the general class of multiplicative models and provided flexibility in the analysis of MET data. The main objective of this thesis was to extend the Bayesian version of the GGE model in two directions: a) to study maximum entropy principle to derive objective *a priori* distributions to G + GEI effects and b) to implement the heteroscedastic version. These proposals were exemplified with both simulated and actual data. We showed that maximum entropy prior distributions for variance components of singular values makes the GGE model more flexible. Furthermore, considering the specific variance in each location, a better fit of the model is obtained, allowing, on the other hand, to evaluate the genotypes with different precision, retrieving experimental information present in different tests from the biplot graphical representation. The versatility of Bayesian modeling to incorporate inference to the biplot was also shown, with many aspects such as credibility regions for medium environment and ideal genotype, correlations between environments and for the representation “who won where”, that are all difficult to obtain in the standard biplot analysis. In this sense, the method presented here looks quite promising and his implementation on statistical packages suitable for researchers is a work in progress.

Keywords: Gibbs Sampling. Linear–bilinear models. Maximum Entropy. Multi-environmental Trials

LISTA DE FIGURAS

Figura 2.1 – Relação entre coordenadas retangulares e polares	16
Figura 2.2 – Dispersão de dados amostrados de distribuições tridimensionais von Mises-Fisher para três valores dos parâmetros de concentração (azul: 4, verde: 16 e vermelho: 64).	17
Figura 2.3 – Representação geométrica do produto interno entre um marcador de linha \mathbf{l}_i e um marcado de coluna \mathbf{c}_j	38
Figura 2.4 – Biplot GGE para 20 genótipos avaliados em 10 ambientes sendo a variável o rendimento em <i>ton / ha</i>	41
Figura 2.5 – Biplot GGE exibindo a configuração "média versus instabilidade" para 20 genótipos em 10 ambientes.	42
Figura 2.6 – Biplot GGE para 20 genótipos em 10 ambientes exibindo o padrão "quem venceu onde".	43

SUMÁRIO

PRIMEIRA PARTE	9
1 INTRODUÇÃO	10
2 REFERENCIAL TEÓRICO	15
2.1 Dados direcionais	15
2.1.1 Distribuição von Mises-Fisher	16
2.2 Inferência bayesiana	18
2.2.1 Atribuição de informação <i>a priori</i>	20
2.2.2 Priori de máxima entropia	24
2.2.3 Resumos pontuais e intervalares <i>a posteriori</i>	25
2.2.4 Método de Monte Carlo via Cadeias de Markov (MCMC)	27
2.2.5 O algoritmo de Metropolis – Hastings	29
2.2.6 Amostrador de Gibbs	30
2.3 Interação genótipos por ambientes (GEI)	33
2.4 Modelos lineares-bilineares	34
2.4.1 O modelo GGE	36
2.4.2 Análise biplot em modelo GGE	37
2.5 Método bayesiano aplicado a modelos lineares-bilineares	44
2.5.1 Distribuições <i>a priori</i> e condicionais completas <i>a posteriori</i>	46
2.5.2 Distribuições condicionais para vetores singulares no subespaço corrigido	48
2.5.3 Amostragem e inferência para parâmetros lineares e bilineares	51
2.5.4 Regiões de credibilidade bivariadas para escores genotípicos e ambientais	54
2.6 Estimadores de encolhimento em modelos lineares-bilineares	55
2.7 Heterogeneidade de variância residual entre locais em análise de dados MET	61
REFERÊNCIAS	69
SEGUNDA PARTE	81
ARTIGO 1 Shrinkage in the Bayesian analysis of the GGE model: a case study with simulation	83
ARTIGO 2 Modelagem bayesiana GGE para ensaios multiambientais com variâncias heterogêneas	121
CONCLUSÃO	173

PRIMEIRA PARTE

1 INTRODUÇÃO

Ensaios de avaliação de cultivares são de fundamental importância em programas de melhoramento de plantas e devem ser realizados em diferentes locais, bem como no decorrer de vários anos. Os resultados observados podem ser utilizados para classificar, selecionar e recomendar genótipos superiores segundo características de interesse. Os dados resultantes de ensaios multiambientes (MET), permitem modelar a interação genótipos por ambientes (GEI) que representa um dos maiores desafios ao trabalho do melhorista. A GEI é percebida pela resposta diferencial dos genótipos frente a diversidade ambiental, sendo classificada como qualitativa ou cruzada (*crossover interactions*-COI) quando ocasiona mudanças na classificação das respostas genotípicas em função dos ambientes de teste. A presença de COI dificulta a avaliação e recomendação ampla de cultivares para todos os ambientes-alvo, mas oferece a possibilidade de explorá-la em seus efeitos positivos visando identificar adaptações de genótipos a ambientes específicos. Nesse sentido, a análise de uma tabela de dupla entrada (genótipos × ambientes) se tornou central para a avaliação de ensaios de cultivares.

Modelos lineares-bilineares (ou multiplicativos) de efeitos fixos são úteis para analisar dados bidirecionais e tem amplas aplicações para o estudo da GEI. 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. A grande virtude dos modelos multiplicativos está na possibilidade de analisar adaptabilidade e estabilidade em uma única abordagem e, principalmente, na capacidade do método em recuperar padrão e descartar ruído. Isso é possível devido as propriedades da DVS que permitem aproximar uma matriz de alta dimensionalidade por outra de menor dimensão. Além disso, 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 de efeitos principais aditivos e interação multiplicativa (*additive main effects and multiplicative interaction model*-AMMI) (GAUCH, 1988; GAUCH; ZOBEL, 1997) e o *site regression model* (SREG) (CORNELIUS; CROSSA; SEYED-SADR, 1996; CROSSA; CORNELIUS, 1997), também referido por modelo de efeitos principais de genótipos (G) mais GEI (GGE), merecem especial destaque pela grande aplicabilidade em programas de melhoramento de plantas. O AMMI corresponde ao modelo geral e estima

todos os efeitos separadamente, enquanto o GGE é obtido ao se confundir o efeito G com o efeito da GEI, que são absorvidos pelos parâmetros bilineares e obtidos por DVS. Além desses dois, outros modelos multiplicativos podem ser obtidos a partir de um modelo linear-bilinear geral (CORNELIUS; SEYEDSADR, 1997).

Um tema frequentemente pesquisado, no contexto da análise frequentista da classe de modelos lineares-bilineares, é a determinação do número de termos multiplicativos que devem ser retidos no modelo para explicar a interação. Muitos métodos estatísticos foram propostos para esse fim, sendo que os mais populares envolvem a aplicação de testes F aproximados, validação cruzada e procedimentos baseados em computação intensiva (GAUCH, 1988; CORNELIUS; SEYEDSADR; CROSSA, 1992; DIAS; KRZANOWSKI, 2003; HADASCH; FORKMAN; PIEPHO, 2017).

Outra proposta para o ajuste de modelos multiplicativos, sem envolver o uso de testes ou procedimentos de validação cruzada, é a utilização dos denominados estimadores de encolhimento (CORNELIUS; CROSSA; SEYEDSADR, 1996; CORNELIUS; CROSSA, 1999). A ideia é multiplicar as estimativas de OLS, referentes aos termos bilineares, por um fator que promove o encolhimento dessas estimativas, principalmente de componentes relacionados a dimensões mais elevadas. O método é visto como uma forma de recuperar informações interefeitos com o apelo de que os primeiros componentes sejam estimados de forma mais pronunciada e aqueles com pouca contribuição sejam encolhidos para o valor zero. Para tanto, nenhuma suposição de aleatoriedade é feita sobre quaisquer parâmetros do modelo e a única restrição utilizada é de que as estimativas minimizem o erro quadrático médio (EQM).

Não obstante as vantagens oferecidas, as versões usuais de modelos multiplicativos (efeitos fixos) como o GGE estão sujeitas a diversas limitações, como por exemplo, dificuldade em lidar com dados desbalanceados e heterogeneidade de variâncias, apesar de existirem procedimentos frequentistas para imputação de dados (DIAS; KRZANOWSKI, 2003; YAN, 2013) e possibilidade de dimensionamento dos dados e ponderações de médias de celulas baseadas em DVS (CROSSA; CORNELIUS, 1997; YAN, 2014; RODRIGUES et al., 2015). Além disso, a incorporação de inferência aos escores bilineares no biplot tem sido uma tarefa difícil e os métodos aplicados são suscetíveis a críticas nas abordagens para efeitos fixos do AMMI e GGE (DENIS; GOWER, 1994; DENIS; GOWER, 1996; MAIA et al., 2006; LAVORANTI; DIAS; KRAZNOWSKI, 2007; YANG et al., 2009; YAN; GLOVER; KANG, 2010).

Tais dificuldades motivaram a procura por métodos mais flexíveis. Versões em modelos mistos do GGE e AMMI foram propostas por Piepho (1997), Piepho (1998), Smith, Cullis e Thompson (2001) e Piepho e Möhring (2006) e resultam da análise fatorial analítica da matriz de variâncias-covariâncias genéticas entre ambientes. Esses modelos naturalmente oferecem mais vantagens em termos estatísticos e biológicos que suas versões fixas, já que lidam facilmente com a heterogeneidade de variâncias, dados incompletos e a correlação espacial dentro do ambiente. Além disso, considerar genótipos com efeitos aleatórios, abre a possibilidade de incorporar os coeficientes de parentesco na forma fatorial analítica, obtendo assim estimativas mais precisas de valores genéticos (CROSSA et al., 2006; CROSSA et al., 2011; OKEY et al., 2006; BURGUEÑO et al., 2007).

Outra opção vantajosa é a utilização da inferência bayesiana que tem como característica principal a possibilidade de incorporar informações além daquelas obtidas pelos dados experimentais utilizando informações *a priori*. Até pouco tempo atrás, a aplicação da inferência bayesiana aos modelos multiplicativos era limitada, devido às dificuldades envolvidas na amostragem dos vetores singulares (restritos a serem ortogonais entre si). Viele e Srinivasan (2000) e Liu (2001) foram os primeiros a ajustar parâmetros do AMMI sob a metodologia bayesiana e mostraram como conduzir o processo de amostragem, principalmente, para os termos bilineares, cujos suportes para distribuição a posteriori não são triviais. Crossa et al. (2011) e Perez-Elizalde, Jarquin e Crossa (2012), à luz do trabalho pioneiro de Viele e Srinivasan (2000), demonstraram empiricamente a flexibilidade do modelo bayesiano para incorporar inferência aos parâmetros bilineares, bem como informações de experimentos anteriores. Desde então, importantes contribuições foram publicadas (JOSSE et al., 2014; OLIVEIRA et al., 2015; SILVA et al., 2015; JARQUÍN et al., 2016; OLIVEIRA et al., 2016; ROMÃO et al., 20019; SILVA et al., 2019).

Silva et al. (2015) apresentaram uma justificativa bayesiana para o método de encolhimento (*shrinkage*) descrito por Cornelius, Crossa e Seyedsadr (1996) e Cornelius e Crossa (1999). Cornelius e Crossa (1999) constataram que modelos ajustados pelo método *shrinkage* eram superiores àqueles truncados para determinado número de termos multiplicativos a partir de testes *F*. Além disso, se mostraram tão bons e, as vezes, melhores que os melhores preditores lineares não viesados (BLUPs) de modelos considerando a interação aleatória. Contudo, o método *shrinkage* para modelos multiplicativos de efeitos fixos é recomendado para uma tabela de dupla entrada balanceada e o procedimento para obtenção dos valores singulares eventualmente

viola as restrições do modelo, sendo necessário a incorporação de procedimentos de correção *ad hoc* ao algoritmo. A modelagem bayesiana, por outro lado, é flexível o suficiente para lidar com todas essas limitações e oferece ajustes mais parcimoniosos, como ressaltado por Silva et al. (2015) no contexto do modelo AMMI.

Além disso, heterogeneidade de variâncias pode ser modelada de forma natural utilizando inferência bayesiana. Silva et al. (2019) consideraram variâncias residuais diferentes entre os ambientes em uma análise aplicada ao AMMI-bayesiano destacando as diferenças que podem surgir quando esse cenário é negligenciado. É de consenso que nem sempre é razoável adotar a suposição de homogeneidade de variâncias em dados MET, pois isso pode induzir a erros na avaliação e seleção dos melhores genótipos. Modelar heterogeneidade de variâncias, por outro lado, pode resultar em uma melhor estimativa de efeitos genotípicos, uma vez que as estimativas obtidas para os componentes de variância são mais precisas (EDWARDS; JANNINK, 2006; ORELLANA; EDWARDS; CARRIQUIRY, 2014).

A modelagem bayesiana, como aplicada ao AMMI, é geral podendo ser estendida a outros modelos multiplicativos, bastando assumir as pressuposições específicas em cada um deles (PEREZ-ELIZALDE; JARQUIN; CROSSA, 2012). Versões bayesianas do GGE foram apresentadas por Jarquín et al. (2016), Oliveira et al. (2016) e Omer e Singh (2017). Contudo, prioris mais específicas para valores singulares e componentes da variância nos locais não foram utilizadas. A extensão do método *shrinkage* e a modelagem da heterogeneidade da variância residual entre ambientes traria importantes contribuições ao modelo GGE, já que o mesmo possui um forte apelo à análise gráfica, principalmente a versão que considera apenas os dois primeiros componentes principais, denominada por GGE biplot (YAN et al., 2000), intensamente aplicada na análise MET. Com isso, os primeiros eixos podem ser estimados de forma mais pronunciada e certamente os modelos ajustados serão mais parcimoniosos. Além disso, sob a suposição de heterogeneidade, os genótipos podem ser avaliados com precisões diferentes (em cada ambiente), o que pode aumentar a capacidade de recuperar informações sobre diferenças experimentais entre os ensaios no biplot.

Os objetivos da presente tese foram: i) apresentar uma revisão do estado da arte na análise bayesiana GGE para ensaios MET; ii) relaxar a suposição de homogeneidade de variâncias entre ambientes na análise GGE; iii) estender o método de encolhimento ao GGE a partir de atribuição de priori específica para valores singulares e compará-lo com a análise bayesiana

usual e iv) incorporar regiões bivariadas ao GGE biplot e propor formas para interpretar o efeito G+GEI na análise gráfica.

A tese está organizada em duas partes. Além desse capítulo introdutório, a primeira parte é composta por mais um capítulo que é o referencial teórico, descrevendo aspectos fundamentais relacionados a conceitos utilizados na análise de dados resultantes de experimentos multiambientais, na aplicação de modelos multiplicativos a esses dados e, principalmente, apresenta o estado da arte da análise bayesiana desses modelos e, em especial, do modelo GGE. Na segunda parte, os objetivos propostos são contemplados em dois artigos. No primeiro deles estuda-se o efeito de encolhimento de parâmetros bilineares obtido pela atribuição de prioris de máxima entropia aos componentes de variância do valor singular. O segundo artigo teve por objetivo modelar a heterogeneidade de variância residual em locais e avaliar as implicações no inter-relacionamento de genótipos e ambientes, bem como o efeito das suposições feitas sobre a estrutura da variância residual em parâmetros que descrevem a G+GEI no biplot. Por fim, são apresentadas as conclusões gerais da tese.

2 REFERENCIAL TEÓRICO

Este capítulo tem por objetivo apresentar conceitos e definições, bem como os principais aspectos teóricos envolvidos na aplicação do método bayesiano aos modelos lineares-bilineares, em particular ao GGE que é o tema da presente pesquisa.

2.1 Dados direcionais

Em muitos contextos de investigações, os dados ou medidas sobre os fenômenos podem ser coletados, convenientemente, em termos de direções e dão origem aos denominados dados direcionais (MARDIA, 1972; BANERJEE et al., 2005; LEY; VERDEBOUT, 2017). Define-se uma direção aleatória, na dimensão p , por um vetor \mathbf{Y} de norma unitária ($\|\mathbf{Y}\| = 1$) tomando valores na superfície de uma hiperesfera (denotada por S_p) de raio unitário no espaço p -dimensional (\mathbb{R}^p), tendo o seu centro na origem (MARDIA; KENT; BIBBY, 1979).

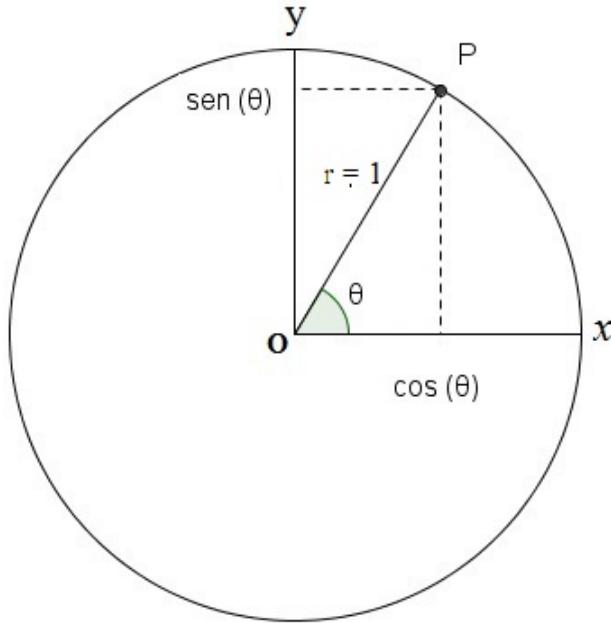
Exemplificando para o caso particular do espaço bidimensional, um ponto P no plano pode ser representado tanto no sistema de coordenadas retangulares por um par ordenado (x, y) , como também em termos de suas coordenadas polares por meio de um raio e um ângulo (r, θ) . A relação existente entre esses dois sistemas de representação é dado por:

$$\begin{cases} x = r\cos(\theta) \\ y = r\sin(\theta) \\ r = \sqrt{x^2 + y^2} \end{cases} \quad (2.1)$$

em que $0 \leq \theta < 2\pi$ e $r \geq 0$.

Ao considerar $r = 1$, tem-se $x = \cos(\theta)$ e $y = \sin(\theta)$, ou seja, cada ângulo (ou direção) corresponde a um ponto P na circunferência do círculo de raio unitário centrado na origem (Figura 2.1). Analogamente qualquer ponto na circunferência é determinado por uma direção (ângulo). Na situação ilustrada pela Figura 2.1, os dados são ditos circulares. Para $r = 0$, θ não está definido (nenhuma direção é indicada) (JAMMALAMADAKA; SENGUPTA, 2001). Essas propriedades são, naturalmente, estendidas ao espaço p -dimensional ($p = 1, 2, \dots$).

Figura 2.1 – Relação entre coordenadas retangulares e polares



Fonte: Jammalamadaka e Sengupta (2001)

Variáveis aleatórias direcionais são descritas pelas denominadas funções esféricas. Como já especificado, as direções, em geral, podem ser visualizadas como pontos na superfície de uma hiperesfera, mas elas são, obviamente, medidas angulares. Distribuições direcionais envolvem duas características básicas que são o vetor de referência, com um sentido de locação, ou direção média e um parâmetro que mede a dispersão dos pontos em relação a esse vetor. Dessa forma, correspondendo às distribuições paramétricas usuais para variáveis lineares, existem várias distribuições paramétricas para variáveis direcionais. Aqui será destacada uma das mais importantes funções esféricas que é a distribuição von Mises-Fisher (MARDIA; JUPP, 2000; NUÑEZ-ANTONIO; GUTIÉRREZ-PEÑA, 2005).

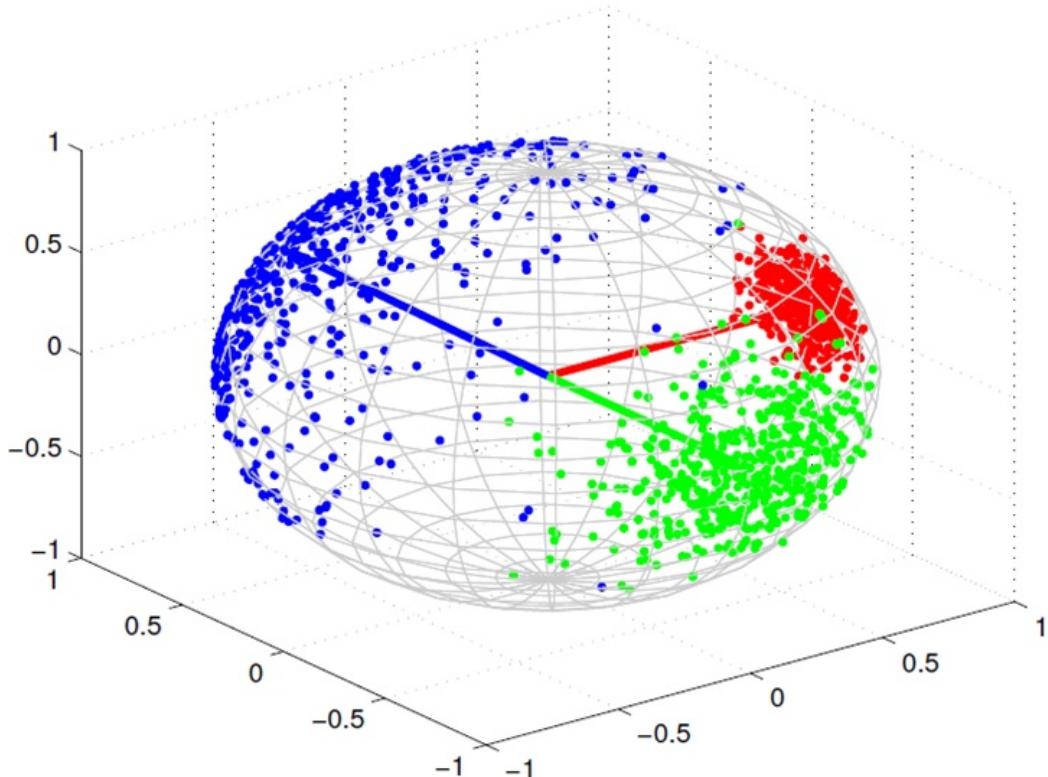
2.1.1 Distribuição von Mises-Fisher

A von Mises-Fisher (*VMF*) é uma extensão natural da distribuição gaussiana em uma hiperesfera unitária, sendo uma das principais distribuições utilizadas na estatística direcional. Diz-se que um vetor \mathbf{Y} tem uma distribuição p -variada von Mises-Fisher, indicado por $\mathbf{Y} \sim VMF(\boldsymbol{\mu}, k)$, se \mathbf{Y} é distribuído na superfície de S_p com função densidade de probabilidade dada por:

$$f(\mathbf{y}|\boldsymbol{\mu}, k) = c_p(k) e^{k\mathbf{u}^\top \mathbf{y}}; \quad \mathbf{y} \in S_p \quad (2.2)$$

em que o vetor μ ($\mu^\top \mu = 1$) denota a média direcional em relação a variável direcional Y . O termo $k \geq 0$ é denominado de parâmetro de concentração que, como o próprio nome sugere, especifica o quanto os vetores unitários estão concentrados em torno de μ . Como exemplificado na Figura 2.2, quanto maior o valor de k mais concentrados os pontos estão em torno da média direcional.

Figura 2.2 – Dispersão de dados amostrados de distribuições tridimensionais von Mises-Fisher para três valores dos parâmetros de concentração (azul: 4, verde: 16 e vermelho: 64).



Fonte: Mash'al e Hosseini (2015)

A constante de normalização é expressa por $c_p(k)$ com a seguinte equação:

$$c_p(k) = \frac{k^{(p-1)/2}}{\{(2\pi)^{p/2} I_{(p-1)/2}(k)\}}, \quad (2.3)$$

sendo $I_r(k)$ a função Bessel modificada do primeiro tipo e ordem r , cuja expressão é dada por

$$I_r(k) = \frac{\left(\frac{1}{2}k\right)^r}{\Gamma(r + \frac{1}{2}) \Gamma(\frac{1}{2})} \int_0^\pi e^{\pm k \cos \theta} \sin^{2r} \theta d\theta. \quad (2.4)$$

Em (2.4), $\Gamma(\cdot)$ denota a função gama definida por Mood, Graybill e Boes (1974) como

$$\Gamma(m) = \int_0^\infty y^{m-1} e^{-y} dy; \quad m > 0 \quad (2.5)$$

com as propriedades de que $\Gamma(m+1) = m\Gamma(m)$. No caso particular de m ser um inteiro positivo ($m \in \mathbb{Z}_+$) então $\Gamma(m) = m!$.

Como já ressaltado, um fato interessante da distribuição VMF é sua semelhança em relação a distribuição gaussiana. Assim, se um vetor \mathbf{Y} é composto por p variáveis aleatórias independentes e identicamente distribuídas com $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ e considerando $\|\mathbf{Y}\| = 1$, então é possível demonstrar que $\mathbf{Y} \sim VMF(\boldsymbol{\mu}, 1/\sigma^2)$ (JAMMALAMADAKA; SENGUPTA, 2001; OLIVEIRA et al., 2015; YASUTOMI; TANAKA, 2014). Além disso, se $\mathbf{Y} \sim VMF(\boldsymbol{\mu}, k)$, o vetor $\mathbf{X} = \mathbf{AY}$ (sendo \mathbf{A} uma matriz ortogonal conhecida) tem distribuição $VMF(\mathbf{A}\boldsymbol{\mu}, k)$, ou seja, a VMF é invariante sobre transformação linear ortogonal (MARDIA; JUPP, 2000).

Um caso especial da von Mises-Fisher surge quando o parâmetro de concentração é igual a zero ($k = 0$). A distribuição assim obtida é denominada de esférica uniforme e sua função densidade de probabilidade é dada por:

$$f(\mathbf{y}) = \frac{\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}} ; \quad \mathbf{y} \in S_p. \quad (2.6)$$

Como os pontos estão distribuídos uniformemente na superfície da hiperesfera, não é possível definir um vetor direcional médio.

2.2 Inferência bayesiana

Por razões históricas, para melhor descrever a modelagem bayesiana, é interessante inicialmente compará-la às versões clássicas ou frequentistas, que constituem um dos pilares da análise estatística de dados. A inferência frequentista, algumas vezes referida como inferência clássica, está intimamente ligada ao princípio da repetibilidade, ou seja, procedimentos tais como estimação e teste de hipóteses são desenvolvidos pressupondo seus desempenhos em todas as amostras aleatórias possíveis de serem obtidas em circunstâncias idênticas e não apenas com relação a amostra aleatória específica que foi obtida (BERNARDO; SMITH, 1994). A teoria da verossimilhança exerce papel de destaque na abordagem clássica e toda a informação sobre os parâmetros de interesse está contida na função de verossimilhança $L(\boldsymbol{\theta} | \mathbf{y})$, sendo $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ um vetor amostral observado e $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_m\}$ um vetor paramétrico, cujos valores são desconhecidos. Os parâmetros, embora desconhecidos, são considerados quantidades fixas e apenas funções da amostra (estimadores) são consideradas aleatórias.

A inferência bayesiana, por outro lado, adota uma postura subjetivista com uso explícito de probabilidades para quantificar o grau de incerteza ou crença acerca de quantidades de interesse não observadas que são representadas por modelos probabilísticos. Consequentemente, os parâmetros, bem como todas as quantidades desconhecidas em um modelo estatístico são considerados variáveis aleatórias. O objetivo da inferência bayesiana consiste em raciocinar da melhor maneira possível com base no conhecimento atual da informação, combinando todas as informações subjetivas que se tem sobre o fenômeno ou problema em estudo (antes de acessar os dados experimentais), com aquelas fornecidas pelos dados observados (ROAZZI; O'BRIEN; DIAS, 2003; BOLSTAD; CURRAN, 2017).

A inferência bayesiana é baseada em uma equação geral que representa um importante resultado relacionado a probabilidade condicional (BOX; TIAO, 1973), expressa por:

$$p(\boldsymbol{\theta} | \mathbf{y}) = \frac{p(\mathbf{y}, \boldsymbol{\theta})}{p(\mathbf{y})} = \frac{p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}. \quad (2.7)$$

Considera-se, sem perda de generalidade, que os valores, y_i e θ_j que compõem, respectivamente, os vetores \mathbf{y} e $\boldsymbol{\theta}$ são quantidades contínuas. Na equação (2.7), $p(\boldsymbol{\theta})$ é a densidade *a priori* de $\boldsymbol{\theta}$ e representa a crença anterior, aprendida antes que os dados sejam observados. A expressão $p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})$ é a densidade conjunta de $(\mathbf{y}, \boldsymbol{\theta})$, sendo $p(\mathbf{y} | \boldsymbol{\theta})$ a densidade conjunta dos dados condicionada no conhecimento de $\boldsymbol{\theta}$ (BAIO; BERARDI; HEATH, 2017). O fator $p(\boldsymbol{\theta} | \mathbf{y})$ é a denominada densidade *a posteriori* de $\boldsymbol{\theta}$ e sintetiza o conhecimento atual sobre os parâmetros, representando o modelo bayesiano de aprendizagem. O resultado expresso por (2.7) é comumente referido como teorema de Bayes (CASELLA; BERGER, 2011; STONE, 2013). A densidade marginal $p(\mathbf{y})$, por sua vez, é obtida integrando a função conjunta $p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})$, em relação ao vetor $\boldsymbol{\theta}$, ou seja:

$$p(\mathbf{y}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \quad (2.8)$$

em que $d\boldsymbol{\theta} = d\theta_1, \dots, d\theta_m$.

A integral definida em (2.8) frequentemente é substituída por uma constante (que será denotada pela letra c), porque depois da integração, c depende de \mathbf{y} mas não envolve mais o vetor aleatório $\boldsymbol{\theta}$. Essa integral, por vezes complicada, pode ser obtida notando que c é uma constante de normalização e a densidade *a posteriori* pode então ser representada por $p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})$, em que o simbolo \propto indica proporcionalidade. Além disso, $p(\mathbf{y} | \boldsymbol{\theta})$ pode ser

reinterpretada como a função de verossimilhança $L(\boldsymbol{\theta} | \mathbf{y})$ e a densidade *a posteriori* reescrita como $p(\boldsymbol{\theta} | \mathbf{y}) \propto L(\boldsymbol{\theta} | \mathbf{y})p(\boldsymbol{\theta})$ (YOUNG; SMITH, 2005). Assim, o teorema de Bayes determina que a densidade *a posteriori* seja igual ao produto da função de verossimilhança pela densidade *a priori* a menos de uma constante (indicado por \propto).

Toda inferência acerca de $\boldsymbol{\theta}$ é obtida diretamente da densidade *a posteriori* $p(\boldsymbol{\theta} | \mathbf{y})$, o que não inclui inferência sobre uma observação futura \mathbf{y}_0 . Na ótica bayesiana, \mathbf{y}_0 e \mathbf{y} não são independentes porque sua densidade depende de $\boldsymbol{\theta}$, um vetor aleatório cujo valor atual depende de \mathbf{y} . Dessa forma, cálculos probabilísticos envolvendo uma observação futura são realizados por meio da distribuição preditiva de \mathbf{y}_0 que é obtida integrando a densidade conjunta de \mathbf{y}_0 e \mathbf{y} em relação a $\boldsymbol{\theta}$:

$$f(\mathbf{y}_0 | \mathbf{y}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{y}_0, \boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{y}_0 | \boldsymbol{\theta}, \mathbf{y}) f(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (2.9)$$

expressando, de forma intuitiva, que a incerteza associada com o valor predito de futuras observações apresentam dois componentes: a variabilidade amostral e a incerteza nos parâmetros (RENCHER; SCHAALJE, 2008). Desde que \mathbf{y}_0 depende de \mathbf{y} somente por meio de $\boldsymbol{\theta}$, a equação (2.9) pode ser simplificada por:

$$p(\mathbf{y}_0 | \mathbf{y}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\mathbf{y}_0 | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}. \quad (2.10)$$

A distribuição preditiva é mais um dos ganhos da inferência bayesiana, já que previsões que consideram a incerteza inerente ao procedimento de estimação (uma característica marcante da análise bayesiana) são mais eficientes do que aquelas obtidas por métodos frequentistas que resultam de práticas do tipo “*plug-in*” dos parâmetros estimados.

2.2.1 Atribuição de informação *a priori*

É importante observar que o teorema de Bayes apresentado pela equação (2.7) não especifica quais devam ser nossas crenças, mas sim como elas devem mudar depois de acessar novas informações, ou seja, fornece a chave pelas quais as crenças originais devem se atualizar à luz da evidência (GHOSH; DELAMPADY; SAMANTA, 2007; BERNARDO; SMITH, 1994). Então, após definir um modelo para os dados, o próximo passo do pesquisador, em relação à modelagem bayesiana, é a atribuição de informação às quantidades desconhecidas, por meio de suas “convicções” ou hipóteses. Para tanto, a interpretação de probabilidade utilizada é a subje-

tiva em que o grau de credibilidade é medido, em função da evidencia atribuída à hipótese, por crenças individuais (PAULINO; TURKMAN; MURTEIRA, 2003; GILLIES, 2000; BERGER et al., 2006). Para a inferência bayesiana, a coerência interna dos modelos é um dos critérios de validação e métodos *ad hoc* como os da inferência frequentista ou clássica são evitados.

Em muitos casos, o estatístico pode obter informações *a priori* substanciais sobre os parâmetros do modelo, advindas de situações práticas podendo recorrer a outros indivíduos considerados especialistas em eventuais condições. A sintetização e quantificação dessas informações subjetivas em uma distribuição que traduza da melhor maneira os conhecimentos que se tem *a priori*, para posterior utilização na sequencia do método bayesiano, pode se dar por diversos procedimentos, envolvendo métodos estruturais e preditivos de eliciação (PAULINO; TURKMAN; MURTEIRA, 2003; ALBERT, 2007).

No entanto, atribuir uma distribuição de probabilidade, totalmente especificada a um parâmetro desconhecido, por procedimentos empíricos ou subjetivos, pode se tornar uma tarefa difícil. Em muitas situações, a seleção da distribuição *a priori* é bastante delicada na ausência de informações prévias confiáveis e soluções genéricas devem ser escolhidas (GHOSH; DELAMPADY; SAMANTA, 2007; MARIN; ROBERT, 2007). A escolha de $p(\theta)$, muitas vezes, é decidida por motivos práticos, em vez de fortes crenças subjetivas ou informações prévias contundentes. Nesse sentido, a utilização das denominadas prioris objetivas tem se tornado uma escolha comum, tendo como objetivo principal minimizar o impacto da seleção *a priori* na inferência (SIMPSON et al., 2017). As prioris objetivas, por vezes, são referidas como prioris não informativas (MARIN; ROBERT, 2007; ROBERT, 2007).

Como não houve consenso sobre uma definição precisa de “muito pouca” em relação a quantidade de informação a ser atribuída, diferentes propostas surgiram no intuito de responder a questão de como uma distribuição *a priori* objetiva deva ser definida. Uma primeira proposta para traduzir uma informação vaga (ignorância *a priori*) sobre um determinado parâmetro é utilizar o princípio da razão insuficiente de Bayes-Laplace, sugerindo que, na ausência de qualquer razão em contrário, todos os valores do parâmetro desconhecido devem ser considerados igualmente prováveis (BERNARDO; SMITH, 1994; JAYNES, 2003). Então, segundo esse princípio, considera-se uma distribuição *a priori* constante no espaço paramétrico (uniforme) $h(\theta) \propto k$, em que k assume valores em um subconjunto da reta. Essa priori será imprópria se o intervalo para $h(\theta)$ for ilimitado, ou seja, se $\int h(\theta)d(\theta) = \infty$.

Outra priori muito utilizada foi desenvolvida por Jeffreys (1961), que se baseia no princípio da invariância de transformações um a um, sendo esta uma vantagem em relação ao procedimento de Bayes-Laplace. A priori de Jeffreys se baseia na Informação de Fisher $I(\boldsymbol{\theta})$ que é obtida por:

$$I(\boldsymbol{\theta}) = E \left(\partial^2 \left[\frac{\ln h(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right] \right) \quad (2.11)$$

em que $\boldsymbol{\theta}$ é um vetor de parâmetros. A priori de Jeffreys é então dada pela raiz quadrada do determinante da matriz de informação de Fisher, ou seja, $h(\boldsymbol{\theta}) \propto [\det I(\boldsymbol{\theta})]^{\frac{1}{2}}$. No caso uniparamétrico tem-se $h(\theta) \propto I(\theta)^{\frac{1}{2}}$ (MARIN; ROBERT, 2007; PAULINO; TURKMAN; MURTEIRA, 2003).

Outras propostas surgiram para expressar a indiferença, ou a falta de informação, sobre o valor de um parâmetro em relação a outro. Essas propostas se basearam, por exemplo, no princípio de invariância de grupo e na teoria da informação (HARTIGAN et al., 1964; VILLEGRAS et al., 1969; BERNARDO, 1979; BERGER; BERNARDO, 1992). A razão para a diversidade de pontos de vista, como já referido, se deve ao fato de que a consideração sobre falta de informação *a priori* difere dependendo da situação, isto é, vários autores propõem diferentes prioris para diferentes grupos de transformações apropriadas em diferentes contextos (PRESS, 2003).

Prioris objetivas geralmente são impróprias e são utilizadas com o apelo de que os dados sejam informativos, o suficiente, para obter a densidade a posteriori própria. Essa dificuldade em utilizar prioris impróprias pode ser contornada incorporando prioris próprias com grandes valores para a variância. Isso pode ser feito trabalhando-se com prioris subjetivas, utilizando o conceito de famílias conjugadas. Ao se utilizar esse método impõe-se que a família de distribuições *a priori* H seja fechada sob amostragem de $F = \{f(\mathbf{y}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ sendo F a família de distribuições considerada e Θ o espaço paramétrico. Diz-se então, que a família H é conjugada de F se

$$p(\boldsymbol{\theta}) \in H \Rightarrow p(\boldsymbol{\theta}|\mathbf{y}) \propto h(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}) \in H. \quad (2.12)$$

A ideia é garantir que $p(\boldsymbol{\theta})$ e $p(\boldsymbol{\theta}|\mathbf{y})$ pertençam a mesma família geral de funções matemáticas, escolhendo $p(\boldsymbol{\theta})$ para ter a mesma “estrutura” que $p(\mathbf{y}|\boldsymbol{\theta})$, quando a última é visualizada como uma função de $\boldsymbol{\theta}$ (BERNARDO; SMITH, 1994). Prioris conjugadas oferecem maior flexibilidade algébrica, já que, as informações podem ser incorporadas pela atribuição de valores convenientes aos hiperparâmetros da densidade *a priori*. Para uma priori gaussiana, por exemplo, atribuir média nula e grandes valores para variância implica em uma densidade *a*

a priori aproximadamente não informativa. O mesmo ocorre, por exemplo, para uma densidade qui-quadrado escalada invertida ao se atribuir grau de liberdade igual a um e parâmetro de escala com altos valores (CROSSA et al., 2011). Prioris mais informativas podem ser atribuídas de forma análoga apenas declarando hipóteses sobre os valores dos hiperparâmetros (HOFF, 2009; ROBERT, 2007).

Uma estratégia flexível é decompor a distribuição *a priori* de forma hierárquica, dividindo o processo de atribuição de informações em diferentes estágios. Isso pode facilitar a especificação, além de ser uma abordagem natural em certas situações. Para exemplificar o método, considera-se que a distribuição *a priori* para um parâmetro θ , dependa de um hiperparâmetro ϕ , o que se representa por $p(\theta | \phi)$. A ideia é não fixar valores para ϕ e sim especificar uma distribuição *a priori* $p(\phi)$. Nesse contexto, a distribuição marginal $p(\theta)$ é dada por

$$p(\theta) = \int p(\theta, \phi)d\phi = \int p(\theta | \phi)p(\phi)d\phi. \quad (2.13)$$

Nesse exemplo, considerou-se apenas dois níveis na hierarquia, mas teoricamente não existe limitação quanto ao número de estágios. Embora, comumente, aplicações não excedam o segundo nível.

Cabe ressaltar que a questão de atribuição de informação *a priori* recai agora sobre as especificações dos hiperparâmetros. Um aspecto interessante e, de certa forma, atraente é que más especificações nas distribuições além do primeiro nível tem menor impacto inferencial do que uma má especificação na distribuição *a priori* padrão $p(\theta)$, como argumentam Paulino, Turkman e Murteira (2003). Essa hierarquização da distribuição justifica-se, de forma geral, pela impossibilidade em quantificar, exatamente, a informação *a priori*, bem como pelo interesse em considerar a incerteza sobre os hiperparâmetros. Além disso, dificuldades em especificar os hiperparâmetros, que se tornam maiores, conforme se aumentam os níveis da hierarquia, fazem com que, geralmente, se terminem essa decomposição com prioris não informativas em último nível.

Um novo conceito para construção de distribuições *a priori* é proposto por Simpson et al. (2017) pela definição de prioris próprias que penalizam a complexidade induzida pelo desvio em relação a um modelo base (mais simples), sendo formuladas pela entrada de um parâmetro de escala ξ definido como parâmetro de flexibilidade tal que $\pi(\xi)$ é a densidade de um dos componentes do modelo. Para o modelo base considera-se que $\xi = 0$. Além disso, esse parâmetro de flexibilidade pode ser um escalar, vários escalares ou ainda um vetor. Tais prioris

são invariantes a reparametrizações, tendo uma conexão natural com as prioris de Jeffreys e fornecem uma estrutura ampla e útil a uma grande classe de modelos hierárquicos (incluindo análises univariadas e multivariadas).

Essas prioris são referidas por prioris de Complexidade Penalizada (ou prioris PC) e são informativas, sendo obtidas sob certos princípios. Simpson et al. (2017) ressaltam as excelentes propriedades de robustez das prioris PC, que são projetadas para contemplar o princípio da parcimônia (navalha de Occam). A informação perdida ao aproximar um modelo flexível pelo modelo base é quantificada utilizando a divergência de Kullback-Leibler (KLD) (KULLBACK; LEIBLER, 1951). Nesse sentido, a medida de complexidade (d) do modelo f , comparado ao modelo g , é dada por $d(f||g) = \sqrt{2KLD(f||g)}$, que é a maneira natural de usar a KLD para definir a distância entre dois modelos com densidades f e g . Esses aspectos permitem utilizar a proposta PC para definir distribuições *a priori* padrão em modelos hierárquicos.

2.2.2 Priori de máxima entropia

A grande crítica a ser feita à diversidade de métodos para incorporar informações *a priori* é que diferentes escolhas podem levar a resultados expressivamente diferentes para uma mesma verossimilhança. A resposta a essa questão pode não estar na conveniente escolha de valores para hiperparâmetros em prioris conjugadas, por exemplo, mas sim na derivação de densidades *a priori* consistentes com o conhecimento que se tenha a disposição, o que pode ser tratado com a utilização do princípio da entropia máxima (JAYNES, 1957; JAYNES, 2003), sendo capaz de produzir estimativas menos tendenciosas o possível sob a informação fornecida. Em termos não-analíticos, a entropia pode ser definida como a “falta de informação” de uma determinada distribuição de probabilidade (FRANK; SMITH, 2010).

A medida de entropia de Shannon (1948) para uma variável aleatória contínua Y , com suporte em S , é quantificada por uma função $H(y)$ dada por:

$$H(y) = - \int_S f(y) \ln(f(y)) dy \quad (2.14)$$

sendo $f(y) \geq 0$, com a igualdade válida apenas fora do suporte S da variável. O problema em encontrar a distribuição que maximiza a entropia se reduz a maximizar a entropia de Shannon (2.14) sujeita aos vínculos do sistema, sendo que o primeiro deles é a condição de normalização de $f(y)$ ($\int f(y)dy = 1$). Outros vínculos podem ser incorporados dependendo das informações

que estão disponíveis e são representados por:

$$E[g_i(y)] = \int_S f(y)g_i(y)dy = a_i \quad (2.15)$$

para $1 \leq i \leq m$, em que a_i é o momento centrado de i -ésima ordem sendo g_i uma função que faz f , com suporte em S , atender as restrições de momentos (COVER; THOMAS, 2006). A maximização da entropia sujeita às condições ou vínculos (2.15) pode ser obtida pelo método dos multiplicadores de Lagrange (TEMPLEMAN; XINGSI, 1989). Então, considerando as $m+1$ condições de vínculo, maximizar $H(y)$ equivale a maximizar a Lagrangeana:

$$\begin{aligned} L(f, \zeta_0, \dots, \zeta_m) = J(f) = & - \int_S f(y) \ln(f(y)) dy + \\ & + \zeta_0 \left[\int_S f(y) dy - 1 \right] + \sum_{i=1}^m \zeta_i \left[\int_S f(y) g_i(y) dy - a_i \right] \end{aligned} \quad (2.16)$$

sendo $\zeta_0, \zeta_1, \dots, \zeta_m$ os multiplicadores de Lagrange. Derivando (2.16), em relação a $f(y)$, obtem-se:

$$\frac{\partial J}{\partial f(y)} = -\ln(f(y)) - 1 + \zeta_0 + \sum_{i=1}^m \zeta_i g_i(y). \quad (2.17)$$

Igualando a expressão (2.17) a zero, encontra-se a forma da densidade maximizada que é $f(y) = e^{\zeta_0 + \sum_{i=1}^m \zeta_i g_i(y)}$, em que $\zeta_0, \zeta_1, \dots, \zeta_m$ são escolhidos de forma que f satisfaça as restrições que foram incorporadas.

De acordo com a teoria da informação, a distribuição que melhor representa o estado de informação disponível é obtida com a maximização da entropia de Shannon.

2.2.3 Resumos pontuais e intervalares *a posteriori*

Estritamente falando, quando se utiliza o teorema de Bayes para determinar o que se pode aprender sobre os valores dos parâmetros a partir dos dados, o método bayesiano fornece uma distribuição de probabilidade *a posteriori* completa $p(\boldsymbol{\theta}|\mathbf{y})$ e não apenas um único ponto no espaço paramétrico. Obviamente é útil resumir essa distribuição (para exibição em texto, gráfico ou tabelas) em termos de um valor que represente o melhor ajuste, bem como quantificar a dispersão em relação ao mesmo (GREGORY, 2005). Nesse contexto, estimativas pontuais ou intervalares para os parâmetros são tomadas como resumos matemáticos da distribuição conjunta *a posteriori*, ou das densidades marginais *a posteriori* dos parâmetros individuais θ_j . A média, mediana ou moda da densidade marginal *a posteriori* dos parâmetros, por exemplo, são

frequentemente utilizadas como resumos pontuais, em que a escolha por uma ou outra dessas medidas obviamente dependente da natureza que envolve a modelagem, bem como, do valor que minimiza a função de perda utilizada pela regra de Bayes (HOFF, 2009; BERNARDO; SMITH, 1994; YOUNG; SMITH, 2005). Quando nenhuma função de perda específica está disponível, existem alternativas como, por exemplo, o estimador de máximo *a posteriori* (MAP), definido como $\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{y}) = \text{argmax}_{\boldsymbol{\theta}} L(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ (MARIN; ROBERT, 2007).

Por outro lado, um intervalo sobre o qual a densidade *a posteriori marginal* de um parâmetro θ_j integra $1 - w$ (com $0 \leq w \leq 1$) pode ser tomado como uma estimativa intervalar de $100(1-w)\%$ para o referido parâmetro (GELMAN et al., 2003). Formalmente um intervalo de densidade, ou região de credibilidade, como são comumente referidos em inferência bayesiana, a um nível de $100(1-w)\%$ de credibilidade, é obtido por meio de um subconjunto $C \subset \Theta$, em que Θ é o espaço de parâmetros, tal que $p\{C|Y=y\} = 1 - w$. Considerando θ contínuo e θ^1, θ^2 sendo $100w_1$ e $100(1-w_2)$ quantis da densidade marginal *a posteriori*, $[\theta^1, \theta^2]$ define um intervalo com um nível de credibilidade $(1 - w)$ sendo $w_1 + w_2 = w$. Um intervalo central é obtido com $w_1 = w_2 = w/2$ (GHOSH; DELAMPADY; SAMANTA, 2007).

Entretanto, em muitas situações busca-se determinar uma região (ou intervalo) que concentra a maior densidade no espaço paramétrico, sendo referida como região de credibilidade com máxima densidade *a posteriori* (HPD). Nesse contexto, uma região R do espaço paramétrico é denominada HPD a $(1 - w)$ de credibilidade se $p(\theta \in R|y) = (1 - w)$ de forma que se $\theta_1 \in R$ e $\theta_2 \notin R$, então $f(\theta_1|y) \geq f(\theta_2|y)$ (BOX; TIAO, 1973).

É importante observar que inferências sumárias, pontuais ou intervalares, como as referidas acima, são obtidas por valores esperados *a posteriori*, ou seja, pela integral $E[g(\theta)|y] = \int g(\theta)h(\theta|y)d\theta$ com escolhas apropriadas de $g(\theta)$. Contudo, expressões analíticas fechadas para as integrais envolvidas estão disponíveis apenas em casos específicos e modelos mais complexos, que envolvem funções n dimensionais, de forma geral, são intratáveis analiticamente. Além disso, a obtenção da distribuição *a posteriori* também exige o cálculo de uma integral analiticamente difícil. Essa dificuldade representou, durante longo tempo, uma grande limitação a aplicação generalizada do método bayesiano. Mas devido ao enorme avanço metodológico e computacional, métodos de integração numérica tem permitido enorme flexibilidade, tornando viável a aplicação e popularização da metodologia bayesiana.

2.2.4 Método de Monte Carlo via Cadeias de Markov (MCMC)

Como mencionado anteriormente, obter soluções analíticas em situações envolvendo modelagem multivariada, na maioria das vezes, não é possível e mesmo soluções numéricas diretas são impraticáveis para quatro ou cinco dimensões (YOUNG; SMITH, 2005). Essa dificuldade pode ser contornada pela utilização de procedimentos estocásticos e essa flexibilidade representa uma das principais vantagens práticas dos métodos bayesianos. Nesse contexto, os métodos de Monte Carlo adquiriram ampla aplicabilidade e buscam amostrar números aleatórios para simular amostras da distribuição *a posteriori* por meio de algoritmos computacionais (JIANG, 2010; ANDRIEU et al., 2003). Quando não é possível amostrar valores diretamente da densidade *a posteriori* a simulação de cadeias de Markov tem sido o método mais utilizado, para amostrar iterativamente valores dos parâmetros de interesse de uma distribuição que se aproxima cada vez mais de $p(\theta)$, sendo esta a estratégia dos denominados Método de Monte Carlo via Cadeia de Markov (MCMC) (ROBERT; CASELLA, 2004; GREGORY, 2005; KOCH, 2007).

Uma cadeia de Markov é um processo estocástico, ou seja, uma sequência de variáveis aleatórias (X_1, X_2, \dots) em que para qualquer tempo (ou estágio) n , a distribuição de X_n , dados todos os X anteriores, depende apenas do valor mais recente (X_{n-1}) (GELMAN et al., 2003). Define-se por espaço de estados S , o subconjunto de valores que $\{X_n\} = \{X_n, 0, 1, 2, \dots\}$ pode assumir. Assim, se $X_n = i$, diz-se que o processo está no estado i no tempo n (KULKARNI, 2011). Se a cadeia é independente do tempo ela é dita ser homogênea. Agora seja $S \subset \mathbb{R}^m$ e A uma σ -álgebra em S , a distribuição de uma cadeia de Markov homogênea com espaços de estados S é determinada pela sua distribuição inicial e pela função de transição que é uma aplicação $P : S \times A \rightarrow [0, 1]$ (PAULINO; TURKMAN; MURTEIRA, 2003).

Se a cadeia de Markov tem espaço de estado discreto sua probabilidade de transição é dada por:

$$\begin{aligned} P(X_n = j | X_{n-1} = i, X_{n-2} = k, \dots, X_0 = m) &= \\ &= P(X_n = j | X_{n-1} = i) = p_{i,j}; \quad i, j, k, \dots, m \in S \end{aligned} \tag{2.18}$$

em que o elemento $p_{i,j}$ representa a probabilidade condicional $p(j|i)$ de que a cadeia esteja no estado j no tempo n , dado que no tempo $n - 1$ estava no estado i (sendo j estocástico e i fixo). A representação $p_{i,j}$ é utilizada para indicar o elemento da linha i e da coluna j na matriz \mathbf{P} de probabilidades de transição da cadeia.

No caso de cadeias de Markov com espaço de estado contínuo não faz sentido falar em probabilidade de transição como em (2.18) e sim em kernels ou densidade de transição (PRESS, 2003). Então sendo \mathbf{X} um vetor aleatório de dimensão m com densidade f em \mathbb{R}^m e assumindo que \mathbf{B} é um subconjunto em \mathbb{R}^m , o kernel de transição especifica que a probabilidade condicional de \mathbf{X}_{n+1} pertencer ao subconjunto \mathbf{B} dado que $\mathbf{X}_n = \mathbf{x}$ é dada por (WAAGEPETERSEN; SORENSEN, 2001):

$$P(\mathbf{x}, \mathbf{B}) = P(\mathbf{X}_{n+1} \in \mathbf{B} | \mathbf{X}_n = \mathbf{x}) \quad (2.19)$$

Uma questão chave na simulação de uma cadeia de Markov é ela convergir para uma distribuição estacionária ou invariante π , independente de qualquer distribuição inicial adotada (GELMAN et al., 2003). Isso significa que se uma cadeia atingiu um estágio em que π é a distribuição estacionária, ela a retém nos movimentos subsequentes. A convergência e a unicidade exigem que a cadeia seja aperiódica e irredutível (KULKARNI, 2011). Uma cadeia é aperiódica se todos os estados tiverem o período igual a 1. Se todos os estados da cadeia de Markov se comunicarem, de modo que seja possível ir de qualquer estado i para qualquer estado j , em uma ou mais etapas ou transições, a cadeia de Markov é dita ser irredutível (JIANG, 2010; SORENSEN; GIANOLA, 2002).

Seja uma cadeia de Markov munida das propriedades ressaltadas anteriormente, com espaço de estados S , que converge para uma distribuição invariante π em que $\mathbf{x} \in S$ denota o estado atual do sistema e $\mathbf{y} \in S$ denota o estado na próxima etapa. Além disso, suponha que $p(\mathbf{x}, \mathbf{y})$ seja a probabilidade de uma transição de \mathbf{x} para \mathbf{y} e que $p(\mathbf{y}, \mathbf{x})$ denote a probabilidade de uma transição na direção oposta. Nesses termos, diz-se que uma cadeia de Markov é reversível se satisfizer a condição:

$$\pi(\mathbf{x})p(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y})p(\mathbf{y}, \mathbf{x}) \quad (2.20)$$

para todo $\mathbf{x}, \mathbf{y} \in S$, que é conhecida como a equação detalhada de equilíbrio (FARR; MANDEL; STEVENS, 2015). É possível provar que uma cadeia ergódica satisfazendo (2.20) tem π como sua distribuição estacionária única (SORENSEN; GIANOLA, 2002). Essa propriedade é de extrema importância para a construção de cadeias de Markov.

Os Métodos MCMC são utilizados para construir um núcleo de transição que gera uma cadeia de Markov com uma distribuição estacionária π de interesse. Então a cadeia gerada pode ser usada para a obtenção de estimativas de Monte Carlo para esperanças $E[h(X)]$ em relação a π (TIERNEY, 1994; HASTIE, 2005) obtidas por:

$$E(h(\mathbf{X})) = \int h(\mathbf{x})\pi(\mathbf{x})d\mathbf{x} = \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N h(X_i) \right) \quad (2.21)$$

Então, $E[h(\mathbf{X})]$ pode ser aproximado pela média amostral $E[h(\mathbf{X})] \approx \sum_{i=1}^N h(X_i)/N$ para grandes valores de N .

2.2.5 O algoritmo de Metropolis – Hastings

No algoritmo Metropolis – Hastings (M–H) a amostragem não é realizada de forma direta e a geração sucessiva de valores aleatórios candidatos é feita a partir de uma distribuição proposta conhecida $q(x|\cdot)$ envolvendo duas etapas. Suponha que no tempo n o estado da cadeia seja $X_n = x$ e para simplificar a notação admite-se, sem perda de generalidade, que X seja um escalar. Então a primeira etapa do algoritmo M–H consiste em gerar um valor candidato $Y_{n+1} = y$, para o próximo estado da cadeia $X_{n+1} = x'$, a partir da densidade proposta $q(x|\cdot)$. A segunda etapa envolve um teste com base em uma probabilidade de aceitação $a(x,y)$. Se o valor y for aceito, o próximo estado da cadeia será $X_{n+1} = y$, caso contrário a cadeia não se moverá e $X_{n+1} = x$.

A densidade conjunta do vetor aleatório (X_n, Y_{n+1}) , sendo X_n e Y_{n+1} como definidos anteriormente, é dada por:

$$g(x,y) = q(x,y)\pi(x) \quad (2.22)$$

em que π é a densidade estacionária dado que X_n tem o valor x . Considerando que a propriedade de reversibilidade é satisfeita por $q(x,\cdot)$, para todo (x,y) , tem se $q(x,y)\pi(x) = q(y,x)\pi(y)$, e a densidade da proposta é o núcleo de transição correto da cadeia produzida pelo M–H. Contudo, é provável que para alguns pontos (x,y) ocorrerá $q(x,y)\pi(x) > q(y,x)\pi(y)$.

Então para garantir a condição de reversibilidade introduz-se o fator $a(x,y) < 1$, que é a probabilidade de aceitação, multiplicando o lado esquerdo da desigualdade, definindo ainda $a(y,x) = 1$ que multiplica o lado direito, de modo a obter:

$$\begin{aligned} q(x,y)\pi(x)a(x,y) &= q(y,x)\pi(y)a(y,x) \\ q(x,y)\pi(x)a(x,y) &= q(y,x)\pi(y). \end{aligned}$$

O valor $a(x,y)$ é calculado por:

$$a(x,y) = \frac{q(y,x)\pi(y)}{q(x,y)\pi(x)}. \quad (2.23)$$

Para inverter o sentido da desigualdade (2.23), basta considerar $a(y,x) < 1$, definindo $a(x,y) = 1$. Como consequência dos argumentos apresentados para a derivação de $a(x,y) < 1$, tem-se que a probabilidade de aceitação deve ser:

$$a(x,y) = \min\left(1, \frac{q(y,x)\pi(y)}{q(x,y)\pi(x)}\right); \quad q(x,y)\pi(x) > 0. \quad (2.24)$$

Se a propriedade de reversibilidade é satisfeita $q(x,y)\pi(x) = q(y,x)\pi(y)$ então $a(x,y) = 1$ e o candidato gerado y é aceito, o que equivale a amostrar o ponto candidato da distribuição de equilíbrio π . É importante notar que para a aplicação desse método é necessário que se conheça a distribuição π apenas proporcionalmente, uma vez que a constante de normalização se cancela na razão $\pi(y)/\pi(x)$ (JIANG, 2010; SORENSEN; GIANOLA, 2002). O processo M–H garante a convergência para a distribuição de equilíbrio para um número N grande de iterações (na verdade tendendo ao infinito), embora esse resultado teórico não forneça nenhuma orientação prática sobre como decidir se a amostra simulada representa uma aproximação razoável. Na prática, as cadeias são monitoradas inspecionando-se o valor da taxa de aceitação, construindo gráficos e computando estatísticas de diagnóstico em relação aos valores simulados (ALBERT, 2007).

2.2.6 Amostrador de Gibbs

O amostrador de Gibbs é um algoritmo MCMC muito popular devido a sua simplicidade computacional, sendo um caso particular do algoritmo M–H quando o kernel de transição é conhecido (ROBERT; CASELLA, 2004). Para constatar esse fato, considera-se, inicialmente, que o estado atual da cadeia seja $X_n = \mathbf{x}$ e que $X_{n+1} = \mathbf{x}'$, em que \mathbf{x} e \mathbf{x}' sejam vetores p -dimensionais. Em muitas aplicações a atualização é realizada coordenada a coordenada, sendo esse o contexto aqui exemplificado. Então ao se passar de \mathbf{x} para \mathbf{x}' , a proposta é gerada a partir de $q_i(\mathbf{x}, z_i) = \pi(z_i | \mathbf{x}_{-i})$, em que z_i é a i -ésima coordenada de \mathbf{x}' e $\mathbf{x}_{-i} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$. Analogamente tem-se o movimento oposto dado por $q_i(\mathbf{x}', x_i) = \pi(x_i | \mathbf{x}'_{-i})$, sendo x_i a i -ésima coordenada de \mathbf{x} e $\mathbf{x}'_{-i} = (z_1, z_2, \dots, z_{i-1}, z_{i+1}, \dots, z_p)$.

No contexto da inferência bayesiana, as densidades $\pi(z_i | \mathbf{x}_{-i})$ e $\pi(x_i | \mathbf{x}'_{-i})$, em que os vetores são compostos por parâmetros de modelos propostos, são referidas como densidades condicionais completas *a posteriori* e são também representadas por $\pi(z_i | \dots)$ e $\pi(x_i | \dots)$. Reparando que $\pi(z_i | \mathbf{x}_{-i}) = \pi(\mathbf{x}') / \pi(\mathbf{x}_{-i})$ e $\pi(x_i | \mathbf{x}_{-i}) = \pi(\mathbf{x}) / \pi(\mathbf{x}_{-i})$ e substituindo essas quantidades na probabilidade de aceitação $a(x,y)$, referente a segunda etapa do algoritmo M–H, dada

por $(q_i(\mathbf{x}', x_i)\pi(\mathbf{x}')/q_i(\mathbf{x}, z_i)\pi(\mathbf{x}))$, tem se que (SORENSEN; GIANOLA, 2002):

$$\frac{q_i(\mathbf{x}', x_i)\pi(\mathbf{x}')}{q_i(\mathbf{x}, z_i)\pi(\mathbf{x})} = \frac{\pi(\mathbf{x})\pi(\mathbf{x}')\pi(\mathbf{x}_{-i})}{\pi(\mathbf{x})\pi(\mathbf{x}')\pi(\mathbf{x}_{-i})} = 1. \quad (2.25)$$

Assim, uma proposta de M–H gerada a partir da distribuição condicional completa apropriada é sempre aceita. Esse esquema é conhecido como amostrador de Gibbs (SORENSEN; GIANOLA, 2002; GELMAN et al., 2003; GAMERMAN; LOPES, 2006). O núcleo de transição do amostrador Gibbs para a atualização de todos os elementos de \mathbf{x} envolve o produto

$$\pi(z_1 | x_2, x_3, \dots, x_p) \pi(z_2 | z_1, x_3, \dots, x_p) \cdots \pi(z_p | z_1, z_3, \dots, z_{p-1}). \quad (2.26)$$

Para exemplificar o método, considera-se o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ de um determinado modelo de interesse. Nesse contexto, a distribuição *a posteriori* de $\boldsymbol{\theta}$ é dada por $p(\theta_1, \theta_2, \dots, \theta_p | \mathbf{y})$, sendo \mathbf{y} um vetor de dados experimentais n -dimensional. Seja $\boldsymbol{\theta}_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$, o vetor de dimensão $p - 1$ que se obtém ao excluir o i -ésimo elemento de $\boldsymbol{\theta}$. Então a distribuição condicional completa *a posteriori* de θ_i é dada por

$$p(\theta_i | \boldsymbol{\theta}_{-i}, \mathbf{y}) \propto p(\theta_1, \dots, \theta_{i-1}, \theta_i, \theta_{i+1}, \dots, \theta_p | \mathbf{y}). \quad (2.27)$$

Considerado o caso em que θ_i é um escalar, os movimentos são realizados ao longo de cada coordenada. De forma geral, poderia ser considerado um vetor $\boldsymbol{\theta}_i$ de dimensão $d > 1$ e a densidade condicional *a posteriori* seria d -dimensional $p(\boldsymbol{\theta}_i | \boldsymbol{\theta}_{-i}, \mathbf{y})$, o que pode resultar em convergência mais rápida, já que a estrutura de correlação entre os componentes é incorporada na densidade condicional *a posteriori* (LIU; WONG; KONG, 1994).

O amostrador de Gibbs fornece um processo iterativo para geração de sucessivas amostras das distribuições condicionais completas *a posteriori* (2.27). Gamerman e Lopes (2006) apresentam uma descrição desse procedimento que é exemplificado por meio do algoritmo apresentado abaixo.

O primeiro passo é a atribuição de valores arbitrários aos elementos do vetor $\boldsymbol{\theta}$, ou seja, $\boldsymbol{\theta}^0 = (\theta_1^0, \dots, \theta_p^0)$. Feito isso, a j -ésima iteração, referente a cadeia de Markov, é obtida pelo seguinte algoritmo:

1. Inicializar o contador de iteração de cada $j = 1$ e o conjunto de valores iniciais $\boldsymbol{\theta}^0 = (\theta_1^0, \dots, \theta_p^0)$;

2. Obter um novo valor $\boldsymbol{\theta}^j = (\theta_1^j, \dots, \theta_p^j)^\top$ de $\boldsymbol{\theta}^{j-1}$ através de sucessivas gerações dos valores:

$$\begin{aligned}\theta_1^j &\sim p\left(\theta_1 \mid \theta_2^{(j-1)}, \dots, \theta_p^{(j-1)}\right) \\ \theta_2^j &\sim p\left(\theta_2 \mid \theta_1^{(j)}, \dots, \theta_p^{(j-1)}\right) \\ &\vdots \\ \theta_p^j &\sim p\left(\theta_p \mid \theta_1^{(j)}, \dots, \theta_{p-1}^{(j-1)}\right)\end{aligned}\tag{2.28}$$

3. Mudar o contador de j para $j + 1$ e repetir o passo 2 até a convergência.

Após um período inicial durante o qual as amostras dependem do valor inicial e devem ser descartadas (queima), os valores $\theta_1^j, \dots, \theta_p^j$, para j suficientemente grande, são considerados amostras da distribuição *a posteriori* conjunta:

$$p(\theta_i \mid \boldsymbol{\theta}_{-i}, \mathbf{y}) = \frac{p(\theta_1, \theta_2, \dots, \theta_p \mid \mathbf{y})}{\int p(\theta_1, \theta_2, \dots, \theta_p \mid \mathbf{y}) d\theta_1 \cdots d\theta_p}\tag{2.29}$$

e a coordenada θ_i^j é considerada como se tivesse sido extraída da distribuição marginal *a posteriori* com densidade $p(\theta_i \mid \mathbf{y}) = p(\theta_i \mid \mathbf{y}) / \int p(\theta_i \mid \mathbf{y}) d\theta_i$. Isto é, o amostrador de Gibbs produz amostra a partir de uma distribuição conjunta, amostrando sucessivamente todas as distribuições condicionais completas *a posteriori* e isso implica que a forma das distribuições condicionais completas determina exclusivamente a forma da distribuição conjunta. À medida que o número de iterações aumenta, a distribuição de transição se aproxima do seu estado de equilíbrio e a convergência é assumida de forma aproximada. A amostra de Monte Carlo pode então ser usada diretamente para aproximar as probabilidades de interesse, bem como obter estimativas pontuais (YOUNG; SMITH, 2005). Regiões HPD podem ser obtidas por algoritmos iterativos a partir dos valores simulados como proposto por Chen e Shao (1999).

Dentre os métodos envolvendo técnicas de MCMC, o amostrador de Gibbs vem sendo o mais comumente empregado quando as distribuições condicionais completas permitem amostragem direta. O amostrador de Gibbs, explorando as distribuições condicionais completas através de algoritmo iterativo, foi proposto inicialmente por Geman e Geman (1984), no contexto de reconstrução de imagens, e posteriormente utilizado por Gelfand e Smith (1990) para simulação de distribuições *a posteriori*. Uma abordagem mais detalhada e rigorosa sobre procedimentos MCMC pode ser encontrada, por exemplo, em Sorensen e Gianola (2002), Kulkarni (2011) e Robert e Casella (2004).

2.3 Interação genótipos por ambientes (GEI)

O melhoramento de plantas é a ciência dedicada ao desenvolvimento de novas variedades de cultivares a partir da manipulação da sua constituição genética. Sendo assim, um dos principais objetivos dos melhoristas é a seleção e recomendação de genótipos superiores no que se refere a produtividade, estabilidade de produção e ampla adaptabilidade aos mais variados ambientes (SILVA; DUARTE, 2006; ASFAW et al., 2009). A adaptabilidade pode ser definida como a capacidade de genótipos em aproveitar vantajosamente as variações ambientais. A estabilidade, por sua vez, refere-se ao comportamento altamente previsível do genótipo, mesmo que avaliado em diferentes ambientes (BORÉM; MIRANDA; FRITSCHE-NETO, 2017; RESENDE; DUARTE, 2007).

O fenótipo (característica observada) é então influenciado pelo genótipo (que refere-se a constituição genética do indivíduo) e pelo ambiente, que pode ser definido como o conjunto das condições externas que interagem com o indivíduo e afetam seu processo de crescimento e desenvolvimento (YAN; KANG, 2003). Além disso, a avaliação fenotípica deve também levar em conta a influência de outro componente, de suma importância, denominado de interação genótipos por ambientes (GEI). A decisão de lançamento de novas cultivares normalmente é dificultada pela ocorrência da GEI que representa um dos maiores desafios aos melhoristas (CARVALHO et al., 2002). Assim, o processo de seleção, normalmente, se dá pela análise do desempenho dos genótipos em ambientes distintos, que podem ser ano, local, ou ambiente de semeadura (PIEPHO, 1996; CARGNIN et al., 2006; GUILLY et al., 2017).

De forma geral, a GEI é percebida pelas respostas diferenciadas de genótipos em função da variabilidade ambiental e faz com que genótipos não possuam performances semelhantes em todos os ambientes de teste (CROSSA; YANG; CORNELIUS, 2004). Ela é considerada simples ou quantitativa quando não se observam alterações de posições relativas das respostas dos genótipos frente a diversidade ambiental, ou seja, observam-se apenas diferenças na magnitude das respostas. Por outro lado, se existe ausência de correlação entre os desempenhos dos genótipos, fazendo com que esses apresentem respostas diferentes, ocasionando alteração no ranqueamento (em função dos ambientes em que foram avaliados), a GEI é dita ser cruzada ou qualitativa, sendo, por vezes, referida por COI (do inglês *Crossover Interaction*) (BURGUEÑO et al., 2008; YANG et al., 2009; CRUZ; REGAZZI; CARNEIRO, 2012).

Na agricultura, o tipo mais importante de GEI é a COI, já que a seleção de genótipos com o objetivo de maximizar a produtividade, no caso de mudanças de classificação nos am-

bientes, é complicada (COOPER; DELACY, 1994; SINGH; CECCARELLI; GRANDO, 1999; YANG et al., 2009). Isto motivou o desenvolvimento de métodos e testes estatísticos para identificar a presença de COI em conjunto de dados provenientes de ensaios multiambientes (MET - *multienvironment trials*) que são descritos sob o modelo fixo (BAKER, 1988; CORNELIUS; SEYEDSADR; CROSSA, 1992), modelo aleatório (YANG, 2002) e modelo misto (YANG, 2007).

Na presença de COI, a maneira de aumentar os ganhos genéticos é a identificação de genótipos especificamente adaptados. Assim, percebe-se que o tipo de GEI desempenha um papel importante na identificação dos genótipos adequados para adaptação específica ou ampla. Essa fonte de variação tem sido modelada à luz de diferentes métodos estatísticos incluindo procedimentos uni e multivariados (CROSSA, 1990; PIEPHO, 1997; SMITH; CULLIS; THOMPSON, 2001; PIEPHO; MÖHRING, 2006; CROSSA et al., 2011). Dentre esses métodos, os modelos lineares-bilineares como o AMMI e o GGE merecem especial destaque pela ampla aplicabilidade em programa de melhoramentos e ensaios multiambientes de forma geral (CROSSA; VARGAS; JOSHI, 2010).

2.4 Modelos lineares-bilineares

Para analisar dados envolvendo dois fatores que possam ser organizados em tabelas de dupla entrada, Gollob (1968), Mandel (1969) e Mandel (1971) introduziram o seguinte modelo para a média de caselas y_{ij} :

$$y_{ij} = \mu + \tau_i + \delta_j + \sum_{k=1}^t \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij} \quad (2.30)$$

sendo $i = 1, \dots, r$; $j = 1, \dots, c$ e $t \leq \min(r-1, c-1)$, em que r é o número de linhas, c o número de colunas e t o posto da matriz de dupla entrada. Na equação (2.30), μ representa a média geral; τ_i é o efeito do i -ésimo fator linha e δ_j é o efeito do j -ésimo fator coluna. Os termos $\lambda_k \alpha_{ik} \gamma_{jk}$, presentes no somatório, são chamados de componentes bilineares (ou multiplicativos) do modelo, em que λ_k , α_{ik} e γ_{jk} são, respectivamente, o valor singular e os elementos relacionados a linha i e coluna j dos vetores singulares associados ao k -ésimo componente principal ($k = 1, \dots, t$). Por fim, ε_{ij} denota o erro experimental que, usualmente, é admitido ser identicamente e normalmente distribuído com média zero e variância comum $\varepsilon_{ij} \sim N(0, \sigma^2)$. Esse modelo tem sido referido na literatura por modelo de efeitos principais aditivos e interação

multiplicativa ou simplesmente (AMMI) a partir do trabalho de Zobel, Wright e Gauch (1988) e adquiriu ampla aplicabilidade para análise de dados em experimentos multiambientais.

Outros modelos podem ser obtidos a partir do modelo geral (2.30) por eliminação de termos aditivos. Eliminando τ_i obtém-se o modelo de regressão de colunas (*column regression-CREG*) que também é referido como regressão de locais (*site regression-SREG*) ou ainda por efeito principais de genótipos mais GEI (GGE) (CROSSA; CORNELIUS, 1997). Por outro lado, o modelo resultante ao reter τ_i e eliminar δ_j é denominado regressão de genótipos (linhas) (*genotypes regression-GREG*). Ao eliminar conjuntamente os termos lineares e a média geral (μ, τ_i, δ_j) obtém-se o modelo completamente multiplicativo (*completely multiplicative model-COMM*). Outro membro dessa classe de modelos foi caracterizado por Seyedsadr e cornelius (1992), sendo referido por modelo multiplicativo deslocado (*shifted multiplicative model-SHMM*) em que os termos bilineares absorvem tanto os efeitos principais (ambientais e genotípicos), quanto o efeito de GEI, mas contém um parâmetro de deslocamento, cuja solução por mínimos quadrados não pode ser obtida analiticamente (CROSSA; CORNELIUS; YAN, 2002).

Quando os parâmetros lineares são eliminados, ou deixam de ser estimados, os respectivos efeitos de linha ou coluna são confundidos com o efeito da interação e absorvidos pelos parâmetros bilineares, diferentemente do que acontece com o modelo AMMI em que todos os efeitos são estimados separadamente. Cornelius e Seyedsadr (1997) formalizaram a classe de modelos para análise de tabela de dupla entrada em um modelo linear-bilinear geral (GLBM) (*General Linear-Bilinear Model*) representado por:

$$y_{ij} = \sum_{h=1}^w \beta_h x_{hij} + \sum_{k=1}^t \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij} \quad (2.31)$$

em que x_{hij} são constantes conhecidas, β_h é o h -ésimo parâmetro linear ($h=1, \dots, w$) e λ_k , α_{ik} e γ_{jk} são componentes bilineares relacionados ao k -ésimo componente principal ($k=1, \dots, t$), como apresentado no modelo (2.30), em que w e t são os números de termos aditivos e multiplicativos, respectivamente.

O ajuste da classe GLBM é realizado em dois estágios. No primeiro estágio são ajustados os efeitos principais pelo método de mínimos quadrados sujeitos às restrições marginais de identificabilidade ($\sum_i \tau_i = \sum_j \delta_j = 0$). O segundo estágio corresponde ao ajuste dos termos bilineares por decomposição por valores singulares (DVS) da matriz de erros de não aditividade ao modelo linear, sujeito as restrições de identificabilidade ($\sum_i \alpha_{ik} = \sum_j \gamma_{jk} = 0$); de ortonormali-

dade dos vetores singulares $(\sum_i \alpha_{ik}^2 = \sum_j \gamma_{jk}^2 = 1)$ e $(\sum_i \alpha_{ik} \alpha_{ik'} = \sum_j \gamma_{jk} \gamma_{jk'} = 0, k \neq k')$ e ainda à relação de ordem $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_t$. A grande virtude dos modelos lineares-bilineares está na possibilidade de simplificar os dados pela seleção de poucos eixos principais para aproximar o padrão da interação.

2.4.1 O modelo GGE

Como apontado na subseção anterior (2.4), o modelo SREG é obtido da equação geral (2.31) do GLBM retirando-se o parâmetro τ que representa o efeito do fator linha e que, daqui por diante, será utilizado para denotar o efeito de genótipo. Então, o efeito do genótipo (G) não é estimado separadamente do efeito da GEI, eles estão confundidos (G+GEI). O modelo SREG, assim como outros modelos multiplicativos, pode ser aplicado a qualquer matriz de dupla entrada em que a interação entre os fatores seja importante para o estudo do comportamento do fenômeno em questão. Como aqui está sendo abordado no contexto de análise multiambiental, esse modelo será referido por GGE (conforme definido anteriormente), cuja equação é dada por:

$$y_{ij} = \mu + \delta_j + \sum_{k=1}^t \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij} \quad (2.32)$$

em que:

μ é a média geral;

δ_j é o efeito do j -ésimo ambiente;

λ_k é o k -ésimo valor singular da matriz de $(\mathbf{GGE})^\top (\mathbf{GGE})$ ou $(\mathbf{GGE})(\mathbf{GGE})^\top$;

α_{ik} é o elemento referente ao i -ésimo genótipo relacionado ao k -ésimo vetor singular de $(\mathbf{GGE})(\mathbf{GGE})^\top$;

γ_{jk} é o elemento referente ao j -ésimo ambiente relacionado ao k -ésimo vetor singular de $(\mathbf{GGE})^\top (\mathbf{GGE})$;

ε_{ij} é o erro médio associado a observação tal que $\varepsilon_{ij} \sim N(0, \sigma_e^2)$.

O ajuste do modelo, como já relatado, é realizado em dois estágios. No primeiro estágio ajustam-se os efeitos principais pelo método de mínimos quadrados aplicados à matriz de médias $\mathbf{Y}_{r \times c} = [Y_{ij}]$. A média Y_{ij} é obtida a partir de l repetições de um ensaio em que genótipos são avaliados em c ambientes. As soluções OLS para a média geral e o efeito do

j -ésimo ambiente, sob as restrições de identificabilidade, são dadas por $\hat{\mu} = \bar{Y}_{..}$ e $\hat{\delta}_j = \bar{Y}_{.j} - \bar{Y}_{..}$, respectivamente, em que:

$$\bar{Y}_{..} = \frac{\sum_{ij} Y_{ij}}{r \times c} \quad \text{e} \quad \bar{Y}_{.j} = \frac{\sum_i Y_{ij}}{r}. \quad (2.33)$$

O segundo estágio corresponde ao ajuste do efeito $G + GEI$. Para tanto é necessário obter a matriz de resíduos de não aditividade aos efeitos principais ($\hat{Y}_{ij} = \bar{Y}_{.j}$). O desvio relacionado a cada combinação entre genótipo e ambiente é dado por $\hat{\varepsilon}_{ij} = \widehat{gge}_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{.j}$. Esses desvios determinam a matriz $\mathbf{GGE}_{r \times c}$, como representada abaixo.

$$\mathbf{GGE}_{(r \times c)} = \begin{bmatrix} \widehat{gge}_{11} & \widehat{gge}_{12} & \cdots & \widehat{gge}_{1c} \\ \widehat{gge}_{21} & \widehat{gge}_{22} & \cdots & \widehat{gge}_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{gge}_{r1} & \widehat{gge}_{r2} & \cdots & \widehat{gge}_{rc} \end{bmatrix}. \quad (2.34)$$

As estimativas para os parâmetros bilineares são então obtidas pela DVS da matriz **GGE** (JOHNSON; WICHERN, 1998; ANDERSON, 2003). A DVS constitui-se em um importante resultado da álgebra matricial e oferece um método (por meio de um algoritmo) para encontrar os componentes principais relativos às linhas e colunas de uma matriz especificada. Por DVS, a matriz $\mathbf{GGE}_{r \times c}$, de dimensões $(r \times c)$, pode ser decomposta da seguinte maneira:

$$\mathbf{GGE} = \mathbf{USV}^\top = \sum_{k=1}^t \lambda_k \boldsymbol{\alpha}_k \boldsymbol{\gamma}_k^\top \quad (2.35)$$

sendo $\mathbf{S}_{t \times t}$ a matriz diagonal contendo os valores singulares λ_k da matriz $(\mathbf{GGE})(\mathbf{GGE})^\top$ (os mesmos de $(\mathbf{GGE})^\top(\mathbf{GGE})$). As matrizes $\mathbf{U}_{r \times t}$ e $\mathbf{V}_{c \times t}$, por sua vez, contém em suas colunas os vetores singulares $\boldsymbol{\alpha}_k$ (autovetores de $(\mathbf{GGE})(\mathbf{GGE})^\top$) e $\boldsymbol{\gamma}_k$ (autovetores de $(\mathbf{GGE})^\top(\mathbf{GGE})$), respectivamente, com as propriedades já mencionadas no modelo (2.30). Tomando de forma cumulativa a soma dos sucessivos termos do desdobramento obtido pela DVS é possível obter aproximações cada vez melhores para soma de quadrado total de G+GEI (SQ_{GGE}).

2.4.2 Análise biplot em modelo GGE

A representação gráfica do efeito $G + GEI$, por meio de biplots centrado em locais, se tornou popular na análise GGE a partir do trabalho de Yan et al. (2000). O método biplot foi, originalmente, proposto por Gabriel (1971) e permite representar graficamente linhas (vetores

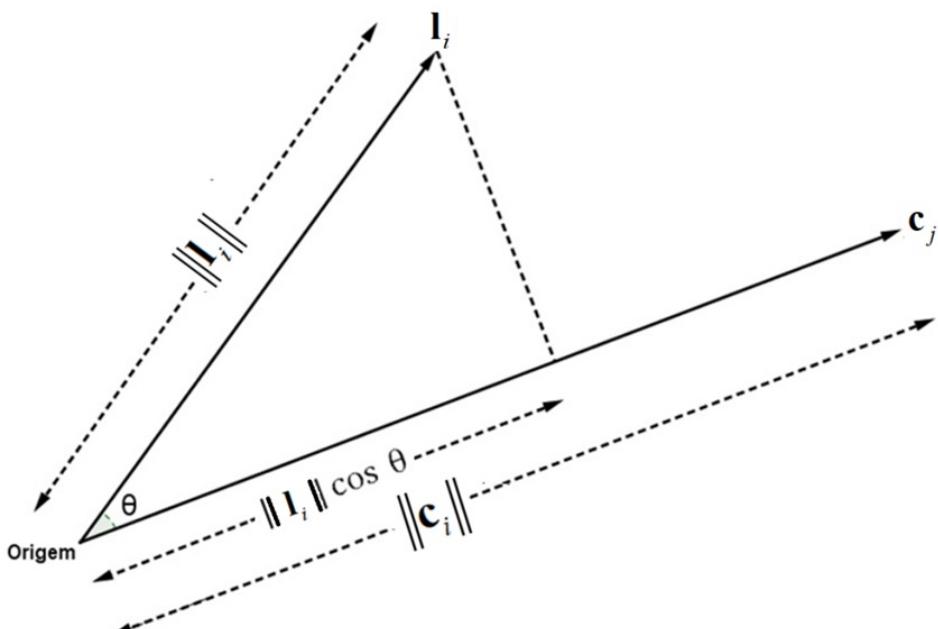
linhas) e colunas (vetores colunas) de uma matriz de interesse. Assim, esta técnica possibilita a representação simultânea de indivíduos e variáveis em baixas dimensões, de forma a tornar visíveis padrões e facilitar a observação e interpretação de relações existentes entre eles, são, portanto, considerados uma forma simples de observar resultados da análise estatística (GREENACRE, 2010). Biplots têm sido utilizados na análise de dados visuais por cientistas das mais distintas áreas de pesquisa, como: economia, sociologia, negócios, medicina, ecologia, genética e agronomia (YAN; TINKER, 2006).

Neste sentido, uma matriz $\mathbf{X}_{n \times p}$ pode ser fatorada como um produto entre duas matrizes $\mathbf{A}_{n \times r} \times \mathbf{B}_{r \times n}$ de forma que a matriz original \mathbf{X} possa ser recomposta como um produto interno entre os vetores marcadores de linhas ($\mathbf{l}_i, i = 1, \dots, n$) (linha da matriz \mathbf{A}) e os vetores marcadores de colunas ($\mathbf{c}_j, j = 1, \dots, p$) (colunas de \mathbf{B}) da seguinte forma:

$$\mathbf{X} = [x_{ij}] = [\langle \mathbf{l}_i, \mathbf{c}_j \rangle] = [\|\mathbf{l}_i\| \|\mathbf{c}_j\| \cos \theta_{ij}] \quad (2.36)$$

em que $\langle \cdot, \cdot \rangle$ denota o operador produto interno, $\|\cdot\|$ representa o módulo do vetor e θ_{ij} o ângulo formado entre os vetores (YAN, 2001). Uma interpretação geométrica, no plano, é oferecida pela Figura 2.3.

Figura 2.3 – Representação geométrica do produto interno entre um marcador de linha \mathbf{l}_i e um marcador de coluna \mathbf{c}_j .



Fonte: Oliveira (2014)

A interpretação é realizada pela observação do comprimento de cada vetor e do sinal do produto interno, que, em última instância, depende do ângulo formado entre eles (GABRIEL, 1971; YAN; KANG, 2003). Desta forma, o sinal de cada elemento x_{ij} , da matriz \mathbf{X} , depende do sinal do $\cos \theta_{ij}$, ou mais explicitamente, do ângulo θ_{ij} . Para $\theta < 90^\circ$, o $\cos \theta$ é positivo e, consequentemente, o produto escalar será positivo. Se $\theta > 90^\circ$, o $\cos \theta$ é negativo e o produto escalar também será negativo. No caso particular de $\theta = 90^\circ$, como $\cos \theta = 0$ o produto interno será nulo.

Assim, toda matriz de característica (ou posto) 2 ou 3 pode ser, integralmente, representada (graficamente) em um biplot bi ou tri-dimensional, respectivamente. Por outro lado, para matrizes com alta dimensionalidade ($\text{posto}(X) = \min(n, p) > 3$), os dados ainda podem ser representados, de forma aproximada, em duas ou três dimensões, utilizando biplots. Gabriel (1971) utilizou a decomposição de \mathbf{X} em valores e vetores próprios como técnica para obter os marcadores das linhas/indivíduos e das colunas/variáveis. A DVS de uma matriz \mathbf{X} é dada por $\mathbf{X} = \mathbf{USV}^\top$ em que \mathbf{S} é a matriz diagonal formada pelos valores singulares de $\mathbf{X}^\top \mathbf{X}$ (ou de \mathbf{XX}^\top) e as matrizes \mathbf{U} e \mathbf{V} contêm em suas colunas os autovetores de $\mathbf{X}^\top \mathbf{X}$ e \mathbf{XX}^\top , respectivamente. A partir da DVS é possível decompor \mathbf{X} em um produto de duas matrizes:

$$\mathbf{X} = (\mathbf{US}^f) (\mathbf{S}^{1-f} \mathbf{V}^\top) \quad (2.37)$$

com $\mathbf{A} = \mathbf{US}^f$ e $\mathbf{B} = \mathbf{S}^{1-f} \mathbf{V}^\top$, sendo f um número real, tal que $0 \leq f \leq 1$.

A análise biplot foi introduzida em pesquisas agronômicas a partir dos trabalhos de Bradu e Gabriel (1978), Kempton (1984), Gauch (1992) e Cooper e DeLacy (1994). Desde então, o método se tornou popular com ampla aplicabilidade em ensaios de competição de cultivares, dentre outras utilidades. Diferentes biplots podem ser considerados e dependem do modelo multiplicativo em uso, sendo que as duas configurações mais utilizadas, em ensaios agrícolas, são aquelas centradas nos efeitos principais e média geral, isolando o efeito da GEI (AMMI), e em ambientes (GGE).

O Modelo GGE biplot, como já mencionado, se configura em uma versão do GGE aproximada que retém apenas os dois primeiros componentes principais (PC1 e PC2), sendo expresso por:

$$y_{ij}^* = \alpha_{ik}^* + \gamma_{jk}^* + \varepsilon_{ijk} \quad (2.38)$$

em que $y_{ij}^* = y_{ij} - (\mu + \delta_j)$, $\alpha_{ik}^* = \lambda_k^f \alpha_{ik}$, $\gamma_{jk}^* = \lambda_k^{1-f} \gamma_{jk}$ e f como definido em (2.37), com $i = 1, \dots, r$, $j = 1, \dots, c$ e $k = 1, 2$. A aproximação em duas dimensões indicada na equação (2.38) reflete a restrição de que o terceiro e subsequentes valores singulares são iguais a zero $(\sum_{k=3}^t \alpha_{ik}^* \gamma_{jk}^* = 0)$ (HU; YANG, 2013). Marcadores de linhas $(\alpha_{i1}^*, \alpha_{i2}^*)$ e marcadores de colunas $(\gamma_{j1}^*, \gamma_{j2}^*)$ são plotados em relação ao plano determinado pelos dois primeiros PCs.

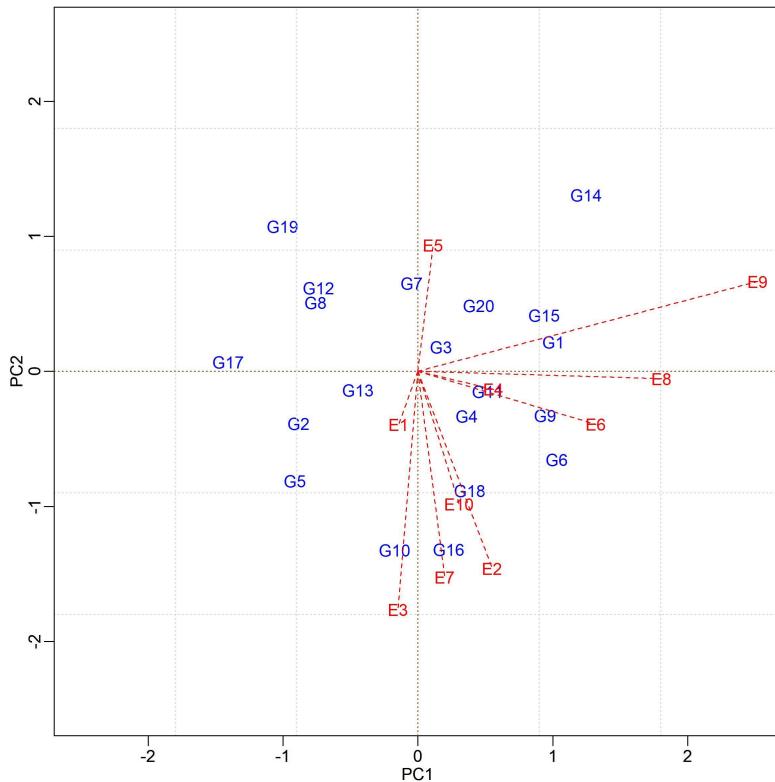
A justificativa para a utilização da configuração (2.38) se baseia na ideia de que apenas G e GEI sejam relevantes para a avaliação das cultivares. Assim, Yan e Kang (2003) advogam que na avaliação de cultivares, é essencial remover o efeito principal de ambientes (E) dos dados. Outro ponto importante, segundo os defensores do modelo GGE, é que G e GEI devem ser considerados simultaneamente para tomar decisões de seleção mais significativas (YAN et al., 2007; GAUCH, 2006; YAN; HOLLAND, 2010; XU et al., 2014).

Quando $f = 1$, diz-se que o GGE biplot tem escala focada em genótipo, pois $\alpha_{ik}^* = \lambda_k \alpha_{ik}$ e $\gamma_{jk}^* = \gamma_{jk}$ e os escores genotípicos possuem médias irrestritas, sendo a norma dos vetores iguais aos valores singulares associados (representam a unidade original dos dados) exibindo a inter-relação entre genótipos com mais precisão. Por outro lado, se $f = 0$, ocorre o oposto e o biplot é baseado em escala focada em ambiente, sendo esse o método mais frequentemente usado (referido como o dimensionamento da análise de componentes principais), naturalmente, é mais informativo das inter-relações entre os ambientes. Quando o fator de escala é $f = 1/2$, os escores resultantes da multiplicação entre vetores e valores singulares (marcadores de linhas e colunas) possuem mesma qualidade de representação para ambos, indivíduos (genótipos) e variáveis (ambientes), mas não a máxima que é possível separadamente (GABRIEL, 2002; YAN, 2002). Apesar de inúmeras possibilidades, a escolha de f no intervalo considerado, não altera relações entre genótipos e ambientes, embora as aparências dos biplots fiquem diferentes (YAN; KANG, 2003; YAN, 2014).

No GGE biplot, o PC1 informa sobre adaptabilidade. Assim quanto maior o valor relacionado à abcissa, maior será o rendimento da cultivar, enquanto pequenas projeções no eixo das ordenadas indicam que os mesmos são relativamente estáveis. Naturalmente, cultivares afastadas da abcissa, com grandes projeções para o eixo PC2 (abaixo ou acima da origem do biplot) são consideradas menos estáveis. Logo, o genótipo ideal seria aquele com alto valor para PC1 e pequenos valores para PC2. Essa proposta subtende a existência de uma alta correlação entre os escores do PC1 e efeitos principais de genótipos (YAN et al., 2000; YAN et al., 2001; YAN; HUNT, 2001; YAN, 2002; YAN; RAJCAN, 2002).

A Figura 2.4 exemplifica uma representação GGE biplot obtida para um conjunto simulados de dados MET em que 20 genótipos (linhas) são avaliados em 10 ambientes (colunas), cuja variável considerada é rendimento em tonelada por hectare (*ton/ha*). Esse biplot apresenta vetores ambientais que facilitam a visualização das correlações entre os ambientes permitindo separar subgrupos semelhantes (indicado por ângulos agudos entre os vetores).

Figura 2.4 – Biplot GGE para 20 genótipos avaliados em 10 ambientes sendo a variável o rendimento em *ton / ha*.

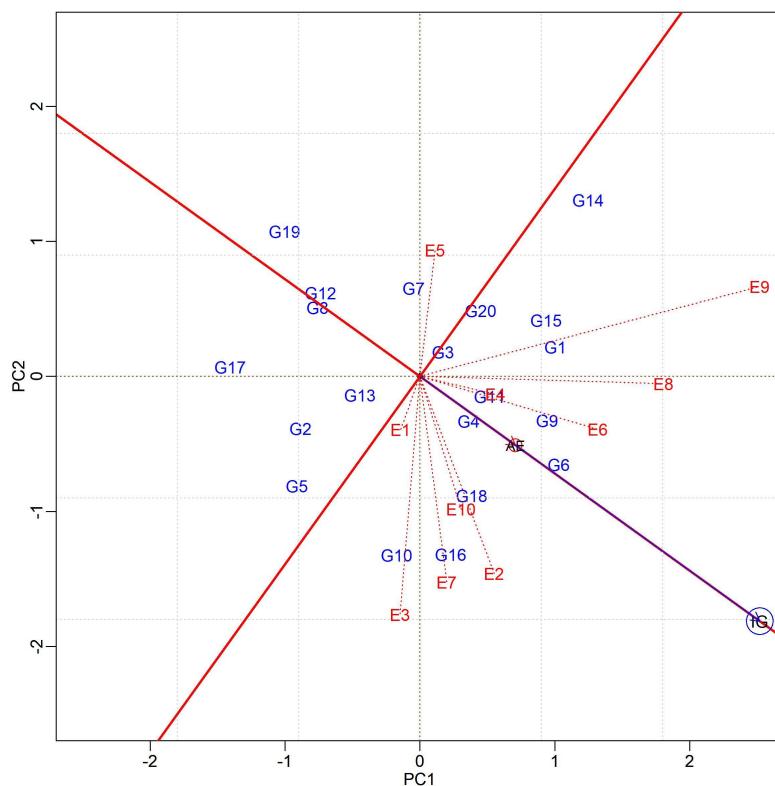


Outra proposta para separar graficamente G e GEI envolve a criação de eixos adicionais no biplot GGE2. Um eixo é traçado pela origem do biplot, passando pela coordenada média ambiental que será referido por AEA (do inglês *average environment axis*). Uma abscissa é obtida pela projeção do marcador de genótipos sobre AEA, de forma que o rendimento médio da cultivar é aproximado por essa projeção. Por outro lado, a estabilidade das cultivares é medida pela projeção em outro eixo que também passa pela origem, sendo perpendicular ao AEA, referido por AEC (*average environment coordination*). Quanto maior o comprimento absoluto da projeção de uma cultivar nessa ordenada, menos estável ela será (YAN, 2001; YAN; TINKER, 2005).

A Figura 2.5 apresenta o biplot exibindo essa configuração, que é referida por "média versus instabilidade", sendo útil para avaliar genótipos em um ambiente alvo. O eixo AEA

corta a origem passando pelo segundo e quarto quadrantes e sua ordenada AEC passa pelo primeiro e terceiro quadrantes. O pequeno círculo sobre a AEA, mais próximo à origem, representa o ambiente médio (*average environment-AE*) e quanto menor for o ângulo de um vetor ambiental em relação ao vetor do ambiente médio, mais representativo o respectivo ambiente será do ambiente de destino. O pequeno círculo, mais afastado da origem e ainda sobre o eixo AEA, representa o genótipo ideal (*ideal genotype-IG*) definido para ter o maior rendimento médio de todos os ambiente e sendo altamente estável. Nesse sentido, a distância de um genótipo a esse ideal é critério de seleção para o conjunto de ambientes considerados.

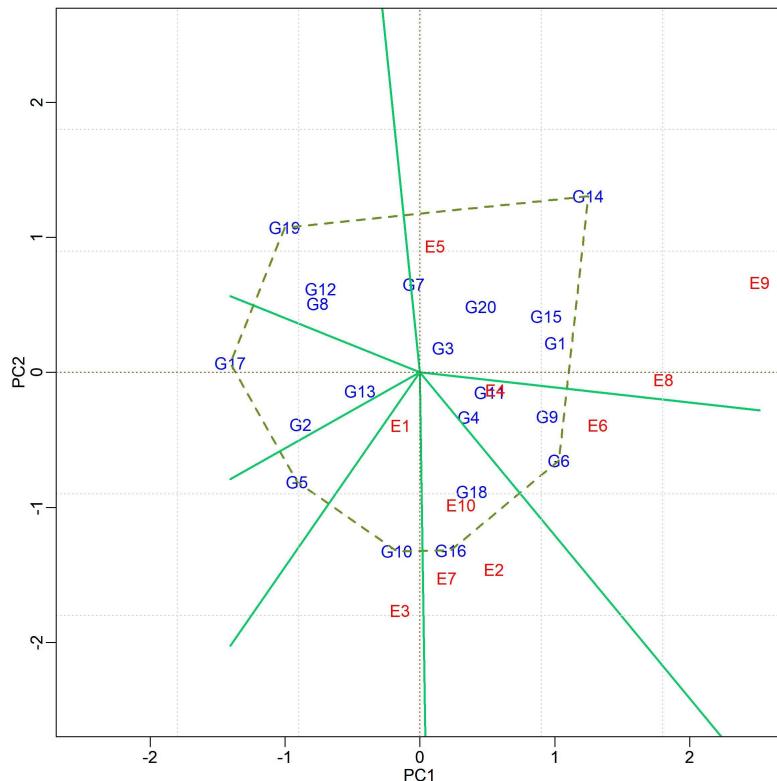
Figura 2.5 – Biplot GGE exibindo a configuração "média versus instabilidade" para 20 genótipos em 10 ambientes.



Além disso, o padrão “quem venceu onde”, que objetiva encontrar o genótipo com melhor desempenho em cada ambiente, é acessado diretamente no biplot do modelo GGE (YAN et al., 2000). Para tanto, é construído um polígono cujos vértices são os genótipos mais afastados em relação a origem do plano determinado pelos eixos PC1 e PC2 (os demais genótipos ficam localizados no interior desse polígono). Então, a partir do centro do biplot, traçam-se retas que são perpendiculares a cada lado do polígono dividindo o mesmo em setores. De forma geral, o fator de linha no vértice do polígono, em um determinado setor, possui os maiores valores para todos os fatores coluna presentes nesse setor (YAN; KANG, 2003). Essa representação é exem-

plificada na Figura 2.6. Assim, em um biplot, as linhas (genótipos) com os maiores valores para diferentes colunas (ambientes) podem ser facilmente visualizadas, independentemente de quantas linhas existem no conjunto de dados, ou mesmo do fator de partição (ou foco) utilizado (YAN, 2001; YAN; KANG, 2003).

Figura 2.6 – Biplot GGE para 20 genótipos em 10 ambientes exibindo o padrão "quem venceu onde".



O GGE biplot é um método gráfico versátil que permite interpretações diretas baseada nas propriedades do produto interno e pode ser utilizado para: i) classificar cultivares e ambientes com base em seus desempenhos; ii) identificar o melhor cultivar em cada ambiente; iii) agrupar os ambientes com base nas melhores cultivares; (vi) avaliar as cultivares, com base no rendimento médio e na estabilidade; (vii) avaliar os ambientes com base na capacidade de discriminação e na representatividade, dentre outras possibilidades (YAN, 2001; YAN et al., 2010; YAN, 2015; YAN, 2016). Além disso, é possível comparar todos os genótipos (ou ambientes) com um genótipo (ou ambiente) hipotético ideal, visualizando o desempenho e a estabilidade de cada um deles em relação ao ambiente de destino (YAN; KANG, 2003; YAN, 2014).

Esse procedimento de análise adquiriu grande aplicabilidade no contexto do melhoramento de plantas para o estudo da adaptabilidade e estabilidade genotípica possibilitando a identificação de genótipos altamente produtivos e amplamente adaptados, bem como combi-

nações positivas de genótipos a ambientes específicos (recomendações regionalizadas). Além disso, oferece um amplo arcabouço de procedimentos gráficos, possibilitando a visualização e interpretação de padrões de forma simples e ilustrativa. Vantagens e desvantagens do modelo GGE biplot são discutidas por diversos autores (GAUCH, 2006; YAN; TINKER, 2006; YAN et al., 2007; GAUCH; PIEPHO; ANNICHIAIRICO, 2008; BALESTRE et al., 2009; YAN et al., 2011; GAUCH, 2013).

2.5 Método bayesiano aplicado a modelos lineares-bilineares

As limitações presentes na análise clássica dos modelos lineares-bilineares motivaram a busca por procedimentos mais flexíveis, dentre os quais se destaca a utilização do método bayesiano. Viele e Srinivasan (2000) foram os primeiros a propor um procedimento metodológico para estimar parâmetros no modelo AMMI sob a ótica bayesiana e mostraram como conduzir o processo de amostragem por meio de procedimentos MCMC. Importantes contribuições foram dadas por Liu (2001), principalmente na obtenção de densidades completas *a posteriori* com formas conhecidas para os vetores singulares permitindo tornar o algoritmo de amostragem mais rápido e estável.

As dificuldades apresentadas por métodos paramétricos e não paramétricos frequentistas para incorporar inferência ao biplot motivaram Crossa et al. (2011) a resgatarem os trabalhos de Viele e Srinivasan (2000) e Liu (2001). Crossa et al. (2011) e Perez-Elizalde, Jarquin e Crossa (2012) mostraram como construir e interpretar regiões de credibilidade no biplot (AMMI-2), bem como incorporar ao modelo informações obtidas em experimentos anteriores.

O modelo AMMI-bayesiano, proposto por Viele e Srinivasan (2000), pode ser descrito a partir de uma notação vetorial da seguinte forma:

$$\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta} + \mathbf{Z} \mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z} \boldsymbol{\alpha}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k + \boldsymbol{\epsilon} \quad (2.39)$$

sendo $\mathbf{y}_{n \times 1}$ o vetor contendo $n = lcr$ respostas fenotípicas, em que os termos l , r e c são o número de repetições por ambiente, número de genótipos e o número de ambientes, respectivamente. Os vetores $\boldsymbol{\beta}_{cl \times 1}$ e $\mathbf{g}_{r \times 1}$ contêm, respectivamente, os parâmetros de efeitos de blocos hierarquizados dentro de ambientes e efeitos principais de genótipos. Os termos λ_k , $\boldsymbol{\alpha}_k$ e $\boldsymbol{\gamma}_k$ são o valor singular e vetores singulares relacionados ao k -ésimo componente principal, com $k = 1, \dots, t$; sendo $t = \min(r, c)$, o posto da matriz de interação **GE**. As matrizes \mathbf{X}_1 , \mathbf{X}_2 , e \mathbf{Z} de

dimensões $(n \times cl)$, $(n \times c)$ e $(n \times r)$, respectivamente, são de delineamento dos efeitos em $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_k$ e \mathbf{g} (e $\boldsymbol{\alpha}_k$), sendo $\boldsymbol{\epsilon}$ é o vetor de erros aleatórios, com $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$, em que $\mathbf{0}$ é vetor nulo, σ_e^2 a variância residual e \mathbf{I}_n a matriz identidade de ordem n .

Como especificado na subseção (2.4), o modelo AMMI está sujeito às restrições de ordem $\lambda_k \geq \lambda_{k+1} \geq 0$ e de ortonormalidade dos vetores singulares $\boldsymbol{\alpha}_k$ e $\boldsymbol{\gamma}_k$. A distribuição condicional dos dados é normal multivariada, ou seja, $\mathbf{y} | \boldsymbol{\lambda}_k, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \mathbf{g}, \boldsymbol{\beta}, \sigma_e^2 \sim N(\boldsymbol{\mu}_y, \sigma_e^2 \mathbf{I})$ com $\boldsymbol{\mu}_y = \mathbf{X}_1 \boldsymbol{\beta} + \mathbf{Z} \mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z} \boldsymbol{\alpha}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k$.

O modelo expresso pela equação (2.39) é a versão apresentada por Oliveira et al. (2015). Nas abordagens oferecidas por Viele e Srinivasan (2000), Liu (2001) e Crossa et al. (2011), uma média geral (μ) é estimada de forma a obter a equação (2.40),

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}\boldsymbol{\gamma}_k + \boldsymbol{\epsilon} \quad (2.40)$$

em que $\mathbf{1}$ é o vetor unitário de dimensão $n \times 1$ e o modelo apresenta as mesmas restrições de identificabilidade presentes na abordagem frequentista usual da análise AMMI.

Perez-Elizalde, Jarquin e Crossa (2012) introduziram uma notação matricial ao modelo AMMI utilizando a distribuição von Mises-Fisher como priori sobre o conjunto de matrizes ortonormais, cujos termos são os coeficientes bilineares dos vetores singulares. Essa nova abordagem permite a incorporação de informação *a priori* de forma mais simples e intuitiva, bastando expressar informações subjetivas *a priori* em relação às médias fenotípicas em uma matriz \mathbf{Y}_0 . Além disso, informações *a priori* relativas aos parâmetros bilineares são incorporadas pelas matrizes resultantes da decomposição por valores singulares da matriz de interação \mathbf{GE} e a amostragem não é realizada vetor por vetor e sim matricialmente.

A metodologia do AMMI-bayesiano pode ser estendida a outros modelos multiplicativos, fixando certos parâmetros iguais a zero e relaxando algumas restrições. Jarquín et al. (2016) mostraram como implementar a análise bayesiana do modelo GGE, utilizando um modelo multinível (hierárquico) a partir do método de amostragem matricial proposto por Perez-Elizalde, Jarquin e Crossa (2012). Oliveira et al. (2016), por sua vez, apresentam uma abordagem vetorial para o GGE apenas deletando do modelo (2.39) o parâmetro referente ao efeito principal de genótipos \mathbf{g} , obtendo a seguinte expressão:

$$\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z} \boldsymbol{\alpha}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k + \boldsymbol{\epsilon} \quad (2.41)$$

em que os termos bilineares λ_k , α_k e γ_k agora descrevem o efeito de G+GEI. As restrições são análogas àquelas ressaltadas para o modelo AMMI (2.39).

Outras contribuições foram dadas para este método recentemente. Josse et al. (2014), por exemplo, mostraram como conduzir a amostragem dos parâmetros bilineares livre das restrições de ortonormalidade dos vetores singulares. Além disso, diversos trabalhos têm abordado a flexibilidade da inferência bayesiana para superar limitações presentes em abordagens AMMI clássicas. A robustez do modelo AMMI-bayesiano, sob perdas aleatórias de genótipos em ambientes, foi verificada por Romão et al. (20019). Silva et al. (2019), por sua vez, mostraram como modelar variâncias heterogêneas entre os ambientes. Algumas aplicações do método também podem ser encontradas na literatura (OLIVEIRA et al., 2018; JÚNIOR et al., 2018; TEODORO et al., 2019; ZEFFA et al., 2020).

2.5.1 Distribuições *a priori* e condicionais completas *a posteriori*

Oliveira et al. (2015) atribuíram as seguintes densidades *a priori* para os parâmetros do modelo AMMI (2.39):

$$\beta | \mu_\beta, \sigma_\beta^2 \sim N(\mu_\beta, I_\beta \sigma_\beta^2), \mu_\beta = \mathbf{0} \text{ e } \sigma_\beta^2 = 10^8;$$

$$\mathbf{g} | \mu_\mathbf{g}, \sigma_\mathbf{g}^2 \sim N(\mathbf{0}, I_r \sigma_\mathbf{g}^2) \text{ com } p(\sigma_\mathbf{g}^2) \propto 1/\sigma_\mathbf{g}^2;$$

$$\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2 \sim N^+(\mu_{\lambda_k}, \sigma_{\lambda_k}^2), \mu_{\lambda_k} = 0 \text{ e } \sigma_{\lambda_k}^2 = 10^8;$$

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

$\gamma_k \sim$ uniforme esférica no subespaço corrigido;

$$p(\sigma_e^2) \propto 1/\sigma_e^2.$$

Como pode ser observado, pelas hipóteses sobre os hiperparâmetros, as prioris são não informativas para quase todos os parâmetros, com exceção ao parâmetro \mathbf{g} . Para o efeito de genótipo é atribuída uma densidade *a priori* hierárquica em dois níveis com priori de Jeffreys ($1/\sigma_\mathbf{g}^2$) para variância genotípica, o que implica em considerar uma população comum para efeitos genotípicos, tal como acontece nos modelos mistos, em que genótipos são considerados aleatórios. Essas mesmas densidades *a priori* foram utilizadas por Oliveira et al. (2016) no ajuste do GGE-bayesiano (exceto para \mathbf{g} que é retirado do modelo).

Crossa et al. (2011) utilizam densidades *a priori* semelhantes, assim como Viele e Srivivasan (2000) e Liu (2001), considerando valores para os hiperparâmetros, no modelo (2.40), iguais a: $\boldsymbol{\mu}_{\beta} = \mathbf{1}_{\beta} \times 0$, $\boldsymbol{\mu}_g = \mathbf{1}_g \times 0$, $\mu_{\lambda_k} = 0$ e 1×10^8 para cada componente de variância σ_{μ}^2 , σ_{β}^2 , σ_g^2 e $\sigma_{\lambda_k}^2$. Para distribuição de σ_e^2 foi assumida uma distribuição qui-quadrado-escalada-inversa considerando grau de liberdade $v_e = 1$ e parâmetro de $S_e^2 = 10^8$, respectivamente. Com isso, todas as distribuições *a priori* podem ser consideradas, aproximadamente, vagas ou não informativas, inclusive para efeitos de genótipos o que, consequentemente, resulta em estimativas semelhantes àquelas obtidas por OLS para efeitos fixos em todos os parâmetros do modelo AMMI.

Em relação aos componentes bilineares, $N^+(\cdot, \cdot)$ denota a distribuição normal positiva, obtida pelo truncamento da distribuição normal ao domínio $[0, \infty)$, pois os valores singulares devem assumir apenas valores positivos. A distribuição uniforme esférica, como já abordada na subseção (2.1), especifica que os vetores unitários estão distribuídos com mesma probabilidade na superfície de uma hiperesfera. Contudo, devido a restrição de ortogonalidade ($\boldsymbol{\alpha}_k^\top \boldsymbol{\alpha}_{k'} = \boldsymbol{\gamma}_k^\top \boldsymbol{\gamma}_{k'} = 0$ para $k \neq k'$), os vetores $\boldsymbol{\alpha}_k$ e $\boldsymbol{\gamma}_k$ devem ser distribuídos somente em um subespaço restrito da esfera unitária em \mathbb{R}^p para $p = r$ ou $p = c$, respectivamente.

Um vetor p -dimensional \mathbf{y} em \mathbb{R}^p é dito ter distribuição uniforme esférica na esfera $p-s$ ($0 \leq s \leq p-1$) dimensional contida em \mathbb{R}^p , denotada por $\mathbf{y} \sim U(S_p^{p-s})$, se ele é distribuído uniformemente na esfera unitária $S_p^{p-s} = \{ h; h \in \mathbb{R}^p \text{ e } h^\top h = 1 \text{ sendo } h \text{ ortogonal a } s \text{ vetores independentes em } \mathbb{R}^p \}$ (LIU, 2001). Existem dois casos triviais: quando $s = p-1$ em que \mathbf{y} assume apenas dois valores possíveis $\{-1, 1\}$ com igual probabilidade e para $s = 0$ que corresponde a definição de distribuição uniforme esférica na subcessão (2.1.1). A densidade esférica uniforme também é considerada *a priori* não informativa, utilizada para declarar conhecimento vago a cerca dos vetores singulares.

Outras hipóteses podem ser assumidas sobre os hiperparâmetros a partir das densidades *a priori* destacadas anteriormente. Silva et al. (2015), por exemplo, utilizaram prioris específicas para componentes de variância dos valores singulares com $p(\sigma_{\lambda_k}^2) \propto 1/\sigma_{\lambda_k}^2$ ($k = 1, \dots, t$), obtendo com isso efeito de encolhimento para os parâmetros bilineares, semelhantes àqueles obtidos pelo método *Shrinkage* proposto por Cornelius e Crossa (1999) e Cornelius, Crossa e Seyedsadr (1996) ao multiplicar as soluções de mínimo quadrados dos valores singulares (obtidas no modelo AMMI fixo) por um fator de encolhimento, tomando como critério a minimização do erro quadrático médio. A heterogeneidade de variância residual entre os ambientes

também pode ser modelada atribuindo-se prioris para as componentes de variância $\sigma_{e_j}^2$, em que $p(\sigma_{e_j}^2) \propto 1/\sigma_{e_j}^2$, para $j = 1, \dots, c$, sendo c o número de ambientes (SILVA et al., 2019).

Como foi visto na subseção (2.1), toda inferência em relação aos parâmetros do modelo é realizada pela distribuição conjunta *a posteriori* que combina as informações trazidas pelos dados experimentais, por meio da função de verossimilhança, com aquelas atribuídas *a priori* baseadas em conhecimentos subjetivos de pesquisadores. No caso do modelo AMMI (2.39), a função de verossimilhança é dada por:

$$L(\boldsymbol{\theta} | \mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{I}\boldsymbol{\sigma}_e^2|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \boldsymbol{\mu}_y)^\top (\mathbf{y} - \boldsymbol{\mu}_y) \right\} \quad (2.42)$$

sendo $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{g}, \lambda_k, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \sigma_e^2)$ e $\boldsymbol{\mu}_y = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k$.

A distribuição *a posteriori* é obtida então como resultado da aplicação do teorema de Bayes:

$$\begin{aligned} p(\Phi | \mathbf{y}) &\propto L(\boldsymbol{\theta} | \mathbf{y}) p(\mathbf{g} | \boldsymbol{\mu}_g, \sigma_g^2) p(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \sigma_\beta^2) p(\sigma_g^2 | v_g, S_g^2) \times \\ &\quad \times p(\sigma_e^2 | v_e, S_e^2) \prod_{k=1}^t p(\lambda_k | \boldsymbol{\mu}_{\lambda_k}, \sigma_{\lambda_k}^2) p(\boldsymbol{\alpha}_k) p(\boldsymbol{\gamma}_k) \end{aligned} \quad (2.43)$$

em que $\Phi = (\boldsymbol{\beta}, \mathbf{g}, \sigma_g^2, \lambda_k, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \sigma_e^2)$.

As distribuições *a posteriori* completamente condicionadas podem ser obtidas para cada parâmetro a partir da distribuição conjunta *a posteriori* (2.43). Sua forma é gaussiana para efeitos principais e valores singulares e inversa escalada qui-quadrado para componentes da variância. Maiores detalhes e o desenvolvimentos algébricos de sua obtenção podem ser encontrados em Oliveira et al. (2015), Silva et al. (2015) e Silva et al. (2019). Particularmente, para os vetores singulares as distribuições condicionais completas *a posteriori* são do tipo von Mises-Fisher.

Para o GGE, a verossimilhança $L(\boldsymbol{\theta}_{SREG} | \mathbf{y})$ é baseada no modelo (2.41) em que $\boldsymbol{\mu}_y = \mathbf{X}_1\boldsymbol{\beta} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k$. Além disso, na densidade *a posteriori* (2.43), considera-se $\Phi_{SREG} = (\boldsymbol{\beta}, \lambda_k, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \sigma_e^2)$. As densidades condicionais completas *a posteriori* são análogas àquelas obtidas para o AMMI deletando-se o vetor paramétrico \mathbf{g} e podem ser encontradas com maiores detalhes em Oliveira et al. (2016) ou em Jarquin et al. (2016).

2.5.2 Distribuições condicionais para vetores singulares no subespaço corrigido

Como pode ser observado na subseção (2.5.1), atribuíram-se aos vetores singulares distribuições esféricas uniformes no “subespaço corrigido”. A distribuição esférica uniforme,

como já referido, é obtida a partir da von Mises-Fisher quando o parâmetro de concentração é nulo ($k = 0$), ou seja, não existe uma “direção preferida” e os pontos estão dispersos uniformemente na superfície da hiperesfera. A principal dificuldade para aplicação do método bayesiano ao modelo AMMI são as restrições de ortogonalidade dos vetores singulares oriundas da DVS.

Os vetores singulares estão distribuídos em hiperesferas unitárias contida em \mathbb{R}^p , sendo $p = r$ e $p = c$ para genótipos e ambientes, respectivamente. Como já ressaltado, os vetores singulares devem ser ortogonais a $s(0 \leq s \leq t - 1)$ direções indicadas por vetores de uma base t dimensional na dimensão p , ou seja, são vetores $p - s$ dimensionais no espaço de dimensão p . O vetor α_k ($r \times 1$) dimensional, por exemplo, é ortogonal a s vetores na hiperesfera S_r^{r-s} contida em \mathbb{R}^r . Nesse contexto, o suporte para amostrar diretamente a partir das condicionais completas *a posteriori* não é trivial (VIELE; SRINIVASAN, 2000).

Viele e Srinivasan (2000) contornaram esse problema amostrando os vetores a partir de distribuições uniformes em \mathbb{R}^{r-s} , onde não existem restrições, para depois, por meio de transformação linear ortogonal, obtê-los normalizados com correto suporte em \mathbb{R}^r , utilizando o processo de ortonormalização de Gram-Schmidt. Aqui será exemplificado o processo para amostrar os vetores singulares genotípicos. Para vetores singulares ambientais o procedimento é análogo.

Admitindo, sem perda de generalidade, que $r > c$, o vetor α_k deve ser ortogonal aos demais s vetores $\alpha_{k'}s$ com $k' < k$, para $k = 1, \dots, t$. Esses $s(0 \leq s \leq k - 1)$ vetores, por sua vez, compõem as colunas de uma matriz que denotaremos por $\mathbf{A}_{k's}$, de dimensão $(r \times s)$. Pela propriedade do complemento ortogonal, existe \mathbf{H}_k ($r \times (r - s)$) tal que $\mathbf{H}_{r \times r} = [\mathbf{A}_{k's}, \mathbf{H}_k]$ é uma matriz ortonormal, ou seja, as colunas de \mathbf{H}_k formam um conjunto ortonormal de vetores que são ortogonais às colunas $\mathbf{A}_{k's}$.

Pode-se então determinar uma variável auxiliar, ou seja, o vetor α_k^* , ($r - s$) dimensional, por uma transformação linear ortogonal bijetora: $\alpha_k^* = \mathbf{H}_k^\top \alpha_k$. Assim, a distribuição condicional de α_k^* é obtida com suporte em \mathbb{R}^{r-s} . Para mostrar esse fato é necessário recordar que a distribuição condicional *a posteriori* de α_k em $S_r^{r-s} \subset \mathbb{R}^r$ é proporcional a uma von Mizes-Fisher (OLIVEIRA et al., 2016) e dada por:

$$p(\alpha_k | \dots) \propto \left\{ \frac{\lambda_k}{\sigma_e^2} (\alpha_k^\top \mu_{\alpha_k}) \right\} \quad (2.44)$$

em que $\mu_{\alpha_k} = \Lambda_{\alpha_k}^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g})$ para o AMMI, ou $\mu_{\alpha_k} = \Lambda_{\alpha_k}^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta})$ para GGE, com $\Lambda_{\alpha_k} = \text{diag}(\mathbf{X}_2 \boldsymbol{\gamma}_k) \mathbf{Z}$.

A transformação $\boldsymbol{\alpha}_k^* = \mathbf{H}_k^\top \boldsymbol{\alpha}_k$ é um a um, o que implica $\boldsymbol{\alpha}_k = \mathbf{H}_k \boldsymbol{\alpha}_k^*$, de forma que $\boldsymbol{\alpha}_k = \mathbf{H}_k \boldsymbol{\alpha}_k^* = \mathbf{H}_k (\mathbf{H}_k^\top \boldsymbol{\alpha}_k)$. Por outro lado, tem-se que $\boldsymbol{\alpha}_k^\top = \boldsymbol{\alpha}_k^\top \mathbf{H}_k \mathbf{H}_k^\top$, obtendo-se, por consequência, as seguintes expressões:

$$p(\boldsymbol{\alpha}_k | \dots) \propto \exp \left\{ \frac{\lambda_k}{\sigma_e^2} (\boldsymbol{\alpha}_k^\top \mathbf{H}_k \mathbf{H}_k^\top \boldsymbol{\mu}_{\boldsymbol{\alpha}_k}) \right\} = \exp \left\{ \frac{\lambda_k}{\sigma_e^2} ((\boldsymbol{\alpha}_k^*)^\top \mathbf{H}_k^\top \boldsymbol{\mu}_{\boldsymbol{\alpha}_k}) \right\} \quad (2.45)$$

O vetor de constantes $\mathbf{H}_k^\top \boldsymbol{\mu}_{\boldsymbol{\alpha}_k}$ deve ser normalizado. Dividindo e multiplicando a última expressão em (2.45), convenientemente, por $c_k = \sqrt{(\mathbf{H}_k^\top \boldsymbol{\mu}_{\boldsymbol{\alpha}_k})^\top \mathbf{H}_k^\top \boldsymbol{\mu}_{\boldsymbol{\alpha}_k}} = \sqrt{\boldsymbol{\mu}_{\boldsymbol{\alpha}_k}^\top \mathbf{H}_k \mathbf{H}_k^\top \boldsymbol{\mu}_{\boldsymbol{\alpha}_k}}$, resultando em

$$\frac{\lambda_k}{\sigma_e^2} (\boldsymbol{\alpha}_k^*)^\top \mathbf{H}_k \boldsymbol{\mu}_{\boldsymbol{\alpha}_k} = \frac{c_k}{c_k} \frac{\lambda_k}{\sigma_e^2} (\boldsymbol{\alpha}_k^*)^\top \mathbf{H}_k \boldsymbol{\mu}_{\boldsymbol{\alpha}_k} = \frac{c_k \lambda_k}{\sigma_e^2} \frac{(\boldsymbol{\alpha}_k^*)^\top \mathbf{H}_k \boldsymbol{\mu}_{\boldsymbol{\alpha}_k}}{c_k} = \frac{c_k \lambda_k}{\sigma_e^2} (\boldsymbol{\alpha}_k^*)^\top \tilde{\boldsymbol{\mu}}_{\boldsymbol{\alpha}_k}. \quad (2.46)$$

Substituindo (2.45) pela expressão obtida em (2.46) tem-se:

$$p(\boldsymbol{\alpha}_k^* | \dots) \propto \exp \left\{ \left(\frac{c_k \lambda_k}{\sigma_e^2} \right) (\boldsymbol{\alpha}_k^*)^\top \tilde{\boldsymbol{\mu}}_{\boldsymbol{\alpha}_k} \right\} \quad (2.47)$$

com parâmetro de concentração $c_k \lambda_k / \sigma_e^2$ e vetor média direcional $\tilde{\boldsymbol{\mu}}_{\boldsymbol{\alpha}_k} = c_k^{-1} \mathbf{H}_k^\top \boldsymbol{\mu}_{\boldsymbol{\alpha}_k}$. Assim, a distribuição de $\boldsymbol{\alpha}_k^*$ é uma von Mises-Fisher em $S_{r-s} \subset \mathbb{R}^{r-s}$, indicado por $\boldsymbol{\alpha}_k^* | \dots \sim VMF(r-s, c_k \lambda_k / \sigma_e^2, \tilde{\boldsymbol{\mu}}_{\boldsymbol{\alpha}_k})$. Por fim, é necessário obter o vetor no subespaço correto em \mathbb{R}^r e isso é realizado pela transformação inversa $\boldsymbol{\alpha}_k = \mathbf{H}_k \boldsymbol{\alpha}_k^* \in S_r^{r-s} \subset \mathbb{R}^r$, lembrando que a von Mises-Fisher é invariante sob transformação ortogonal e que a transformação ortogonal preserva a norma do vetor (LIU, 2001).

A partir de cálculos análogos obtém-se $\boldsymbol{\gamma}_k^* = \mathbf{D}_k^\top \boldsymbol{\gamma}_k$ tal que $\boldsymbol{\gamma}_k^* \in S_{c-s} \subset \mathbb{R}^{c-s}$ por:

$$p(\boldsymbol{\gamma}_k^* | \dots) \propto \exp \left\{ \left(\frac{d_k \lambda_k}{\sigma_e^2} \right) (\boldsymbol{\gamma}_k^*)^\top \tilde{\boldsymbol{\mu}}_{\boldsymbol{\gamma}_k} \right\} \quad (2.48)$$

em que $\tilde{\boldsymbol{\mu}}_{\boldsymbol{\gamma}_k} = d_k^{-1} \mathbf{D}_k^\top \boldsymbol{\mu}_{\boldsymbol{\gamma}_k}$, sendo $\boldsymbol{\mu}_{\boldsymbol{\gamma}_k} = \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top (\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z} \mathbf{g})$ com $\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} = diag(\mathbf{Z} \boldsymbol{\alpha}_k) \mathbf{X}_2$, $d_k = \sqrt{(\mathbf{D}_k^\top \boldsymbol{\mu}_{\boldsymbol{\gamma}_k})^\top \mathbf{D}_k^\top \boldsymbol{\mu}_{\boldsymbol{\gamma}_k}} = \sqrt{\boldsymbol{\mu}_{\boldsymbol{\gamma}_k}^\top \mathbf{D}_k \mathbf{D}_k^\top \boldsymbol{\mu}_{\boldsymbol{\gamma}_k}}$ e \mathbf{D}_k a matriz de transformação linear, $c \times (c-s)$ dimensional, com colunas ortonormais e ortogonais aos vetores singulares ambientais $\boldsymbol{\gamma}_k$'s que compõem as colunas da matriz $\mathbf{B}_{k'}$ s. A transformação inversa é obtida por $\boldsymbol{\gamma}_k = \mathbf{D}_k \boldsymbol{\gamma}_k^*$.

Para o modelo AMMI com heterogeneidade de variâncias entre os ambientes, as densidades condicionais dos vetores singulares não são esféricas e sim elípticas. As densidades no

subespaço corrigido foram obtidas em Silva et al. (2019) e mais detalhes serão apresentados no decorrer do texto.

2.5.3 Amostragem e inferência para parâmetros lineares e bilineares

Como já destacado, o método utilizado para obter estimativas para quantidades desconhecidas em modelos lineares-bilineares é o MCMC. Viele e Srinivasan (2000) utilizaram o amostrador de Gibbs com passos de Metropolis Hastings, já que a forma da distribuição condicional completa *a posteriori* para os vetores singulares não era conhecida. Liu (2001) utilizando as mesmas distribuições *a priori* derivou um conjunto de distribuições condicionais que permitem amostragem direta pelo amostrador de Gibbs, reduzindo assim o tempo computacional e tornando o algoritmo mais estável.

O algoritmo para amostragem envolve passos que devem ser repetidos iterativamente gerando as cadeias de Markov para os parâmetros e, considerando o modelo (2.39), pode ser conduzido da seguinte forma:

A1 - Atribuir valores iniciais aos parâmetros:

$$\Phi^0 = \left[\boldsymbol{\beta}^0, \mathbf{g}^0, (\sigma_g^2)^0, \lambda_k^0, \boldsymbol{\alpha}_k^0, \boldsymbol{\gamma}_k^0, (\sigma_e^2)^0 \right] \quad (2.49)$$

A2- A l -ésima iteração pode ser obtida a partir do seguinte esquema de atualização:

a) Gerar $\boldsymbol{\beta}^l | \mathbf{g}^{l-1}, (\sigma_g^2)^{l-1}, \lambda_k^{l-1}, \boldsymbol{\alpha}_k^{l-1}, \boldsymbol{\gamma}_k^{l-1}, (\sigma_e^2)^{l-1}$ a partir da distribuição condicional:

$$N \left[\left(\mathbf{X}_1^\top \mathbf{X}_1 \right)^{-1} \mathbf{X}_1^\top \left(\mathbf{y} - \mathbf{Z}\mathbf{g}^{l-1} - \boldsymbol{\psi} \right), \left(\mathbf{X}_1^\top \mathbf{X}_1 \right) (\sigma_e^2)^{l-1} \right] \quad (2.50)$$

$$\boldsymbol{\psi} = \sum_{k=1}^t \lambda_k^{l-1} \text{diag} \left(\mathbf{Z} \boldsymbol{\alpha}_k^{l-1} \right) \mathbf{X}_2 \boldsymbol{\gamma}_k^{l-1} \quad (2.51)$$

b) Gerar $\mathbf{g}^l | \boldsymbol{\beta}^l, (\sigma_g^2)^{l-1}, \lambda_k^{l-1}, \boldsymbol{\alpha}_k^{l-1}, \boldsymbol{\gamma}_k^{l-1}, (\sigma_e^2)^{l-1}$ a partir da distribuição:

$$N \left[\left(\mathbf{Z}^\top \mathbf{Z} + \mathbf{I} \frac{(\sigma_e^2)^{l-1}}{(\sigma_g^2)^{l-1}} \right)^{-1} \mathbf{Z}^\top \left(\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta}^l - \boldsymbol{\psi} \right), \left(\mathbf{Z}^\top \mathbf{Z} + \mathbf{I} \frac{(\sigma_e^2)^{l-1}}{(\sigma_g^2)^{l-1}} \right)^{-1} (\sigma_g^2)^{l-1} \right]. \quad (2.52)$$

c) Gerar $(\sigma_g^2)^l | \mathbf{g}^l, \boldsymbol{\beta}^l, \lambda_k^{l-1}, \boldsymbol{\alpha}_k^{l-1}, \boldsymbol{\gamma}_k^{l-1}, (\sigma_e^2)^{l-1}$ a partir da distribuição condicional:

$$inv-escalada - \chi^2 \left[n_{\mathbf{g}}, \frac{(\mathbf{g}^l)^\top \mathbf{g}}{n_{\mathbf{g}}} \right] \quad (2.53)$$

d) Para gerar a l -ésima observação dos parâmetros bilineares, a sequência d1), d2) e d3) deve ser seguida para $k = 1, 2, \dots, t$:

d1) Gerar $\lambda_k^l | \mathbf{g}^l, (\sigma_g^2)^l, \boldsymbol{\beta}^l, \boldsymbol{\alpha}_k^{l-1}, \boldsymbol{\gamma}_k^{l-1}, (\sigma_e^2)^{l-1}$, a partir da distribuição condicional:

$$N^+ \left[\left(\boldsymbol{\phi}_k^\top \boldsymbol{\phi}_k + \mathbf{I} \frac{(\sigma_e^2)^{l-1}}{\sigma_{\lambda_k}^2} \right)^{-1} \boldsymbol{\phi}_k^\top \mathbf{M}_{k'}, \left(\boldsymbol{\phi}_k^\top \boldsymbol{\phi}_k + \mathbf{I} \frac{(\sigma_e^2)^{l-1}}{\sigma_{\lambda_k}^2} \right)^{-1} \right] \quad (2.54)$$

$\mathbf{M}_{k'} = \mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta}^l - \mathbf{Z} \mathbf{g}^l - \sum_{k' \neq k}^t \lambda_{k'}^{l-1} diag(\mathbf{Z} \boldsymbol{\alpha}_k^{l-1}) \mathbf{X}_2 \boldsymbol{\gamma}_k^{l-1}$, $\boldsymbol{\phi}_k = diag(\mathbf{Z} \boldsymbol{\alpha}_k^{l-1}) \mathbf{X}_2 \boldsymbol{\gamma}_k^{l-1}$ e $\sigma_{\lambda_k}^2 = 10^8$.

d2) Gerar $\boldsymbol{\alpha}_k^l | \mathbf{g}^l, (\sigma_g^2)^l, \boldsymbol{\beta}^l, \lambda_k^l, \boldsymbol{\gamma}_k^{l-1}, (\sigma_e^2)^{l-1}$:

Gerar as coordenadas do vetor auxiliar $\boldsymbol{\alpha}_k^*$ da VFM $\left(r-s, \frac{c_k \lambda_k^l}{(\sigma_e^2)^{l-1}}, \tilde{\boldsymbol{\mu}}_{\boldsymbol{\alpha}_k} \right)$, tal como definida anteriormente, considerando a matriz ortogonal $\mathbf{H}_{k'}$, $r \times (r-s)$. Sendo que:

$$\tilde{\boldsymbol{\mu}}_{\boldsymbol{\alpha}_k} = c_k^{-1} \mathbf{H}_k^\top \boldsymbol{\mu}_{\boldsymbol{\alpha}_k} \text{ e } \sqrt{\boldsymbol{\mu}_{\boldsymbol{\alpha}_k}^\top \mathbf{H}_k \mathbf{H}_k^\top \boldsymbol{\mu}_{\boldsymbol{\alpha}_k}} \quad (2.55)$$

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}_k} = \left[diag(\mathbf{X}_2 \boldsymbol{\gamma}_k^{l-1}) \mathbf{Z} \right]^\top \left(\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta}^l - \mathbf{Z} \mathbf{g}^l \right) \quad (2.56)$$

obter $\boldsymbol{\alpha}_k = \mathbf{H}_k \boldsymbol{\alpha}_k^*$.

d3) Gerar $\boldsymbol{\gamma}_k^l | \mathbf{g}^l, (\sigma_g^2)^l, \boldsymbol{\beta}^l, \lambda_k^l, \boldsymbol{\alpha}_k^l, (\sigma_e^2)^{l-1}$:

Gerar $\boldsymbol{\gamma}_k^*$ da VFM $\left(c-s, \frac{d_k \lambda_k^l}{(\sigma_e^2)^{l-1}}, \tilde{\boldsymbol{\mu}}_{\boldsymbol{\gamma}_k} \right)$, tal como definida anteriormente sendo \mathbf{D}_k , $c \times (c-s)$, uma matriz ortogonal.

$$\tilde{\boldsymbol{\mu}}_{\gamma_k} = d_k^{-1} \mathbf{D}_k^\top \boldsymbol{\mu}_{\gamma_k} \text{ e } \sqrt{\boldsymbol{\mu}_{\gamma_k}^\top \mathbf{D}_k \mathbf{D}_k^\top \boldsymbol{\mu}_{\gamma_k}} \quad (2.57)$$

$$\boldsymbol{\mu}_{\gamma_k} = \left[\text{diag}(\mathbf{X}_2 \boldsymbol{\alpha}_k^l) \mathbf{Z} \right]^\top \left(\mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta}^l - \mathbf{Z} \mathbf{g}^l \right) \quad (2.58)$$

obter $\boldsymbol{\gamma}_k = \mathbf{D}_k \boldsymbol{\gamma}_k^*$.

e) Gerar $(\sigma_e^2)^l | \mathbf{g}^l, (\sigma_g^2)^l, \boldsymbol{\beta}^l, \lambda_k^l, \boldsymbol{\alpha}_k^l, \boldsymbol{\gamma}_k^l$ a partir da condicional *a posteriori*:

$$\text{inv-escalada} - \chi^2 \left[n, \frac{(\mathbf{y} - \boldsymbol{\mu}_y)^\top (\mathbf{y} - \boldsymbol{\mu}_y)}{n} \right] \quad (2.59)$$

$$\boldsymbol{\mu}_y = \mathbf{X}_1 \boldsymbol{\beta}^l + \mathbf{Z} \mathbf{g}^l + \sum_{k=1}^t \lambda_k^l \text{diag}(\mathbf{Z} \boldsymbol{\alpha}_k^l) \mathbf{X}_2 \boldsymbol{\gamma}_k^l \quad (2.60)$$

Os passos seguidos de “a” até “e” formam um ciclo de geração do amostrador de Gibbs e devem ser repetidos iterativamente para gerar as cadeias de Markov. Nesse algoritmo considerase o modelo com homogeneidade de variâncias e ainda com *priori* não informativa para todos os parâmetros bilineares. Versões considerando outras hipóteses *a priori* podem ser encontradas em Silva et al. (2015) e Silva et al. (2019).

A convergência das cadeias pode ser monitorada a partir de métodos formais, como os de Raftery e Lewis (1992) e Heidelberger e Welch (1983). Métodos informais, baseados em técnicas gráficas, também podem ser utilizados para avaliar a estacionariedade das cadeias MCMC. Nesse contexto, um dos instrumentos mais utilizados é a representação gráfica de cada uma das quantidades escalares. Esse diagrama, referido por gráfico dos traços, deve ser analisado visualmente, ao longo de várias janelas temporais, para inspecionar a ocorrência de eventuais alterações no padrão. Além disso, indicações de que as cadeias se estabilizaram e, portanto, convergiram podem ser obtidas, visualmente, pela geração de histogramas sequenciais (GELFAND; SMITH, 1990) ou pelo acompanhamento gráfico da evolução das médias ergóticas, à medida que o número de iterações aumenta (PAULINO; TURKMAN; MURTEIRA, 2003). Esses e outros métodos formais (e informais) podem ser consultados em Paulino, Turkman e Murteira (2003), Robert e Casella (2004) e Gamerman e Lopes (2006).

Após a convergência, as observações nas cadeias podem ser utilizadas para o processo de estimação e inferência. Com exceção aos vetores singulares, todos os parâmetros podem

ser estimados pela média ou moda *a posteriori* das observações geradas. Regiões com máxima densidade *a posteriori* para as quantidades univariadas, a $(1 - w)\%$ de credibilidade, podem ser construídas empiricamente por meio de algoritmos iterativos como o proposto por Chen e Shao (1999).

Para vetores singulares $\boldsymbol{\alpha}_k$ e $\boldsymbol{\gamma}_k$, a média *a posteriori* não satisfaz as restrições de ortonormalidade. Assumindo, por exemplo, $\bar{\boldsymbol{\Pi}}_{r \times c}$ como matriz relacionada a vetores singulares $\boldsymbol{\alpha}_k$ (análogo para $\boldsymbol{\gamma}_k$) cujas colunas são as médias a posteriores de $\boldsymbol{\alpha}_k$, pode ser demonstrado que esses vetores singulares, que denotaremos por $\bar{\boldsymbol{\alpha}}_k$, não são ortogonais, uma vez que esta restrição foi aplicada apenas ao processo MCMC, e tão pouco tem norma unitária (LIU, 2001). Uma matriz $\hat{\boldsymbol{\Pi}}$ deve satisfazer as seguintes restrições $\hat{\boldsymbol{\Pi}}^\top \mathbf{1} = \mathbf{0}$ e $\hat{\boldsymbol{\Pi}}^\top \hat{\boldsymbol{\Pi}} = \mathbf{I}$. Além disso, $\hat{\boldsymbol{\Pi}}$ deve ser tal que $\hat{\boldsymbol{\Pi}} = \text{argmin}_{\boldsymbol{\Pi}} ss(\boldsymbol{\Pi})$ sendo,

$$ss(\boldsymbol{\Pi}) = \frac{1}{m} \sum_{l=1}^m \text{trace} \left[(\boldsymbol{\Pi}^l - \boldsymbol{\Pi})^\top (\boldsymbol{\Pi}^l - \boldsymbol{\Pi}) \right] \quad (2.61)$$

em que os termos l e m denotam a l -ésima iteração do algoritmo de Gibbs e o número total de iterações (observações) das cadeias de Markov selecionadas para o processo de inferência, respectivamente.

Para encontrar as estimativas dos vetores singulares considera-se, inicialmente, a decomposição por valores singulares da matriz $\bar{\boldsymbol{\Pi}}$, que é dada por $\bar{\boldsymbol{\Pi}} = \mathbf{LSV}^\top$, em que $\mathbf{S}_{c \times c}$ é a matriz diagonal formada por valores singulares e $\mathbf{L}_{r \times c}$ e $\mathbf{V}_{c \times c}$ são as matrizes contendo os vetores singulares à esquerda e à direita, respectivamente (assumindo, sem perda de generalidade, que o número da linha é maior que o número de colunas, $r > c$). Liu (2001) demonstrou que a matriz $\hat{\boldsymbol{\Pi}}$ existe e é única, satisfazendo as referidas restrições, sendo obtida por $\hat{\boldsymbol{\Pi}} = \mathbf{LV}^\top$. A mesma argumentação se aplica a vetores singulares ambientais.

2.5.4 Regiões de credibilidade bivariadas para escores genotípicos e ambientais

Crossa et al. (2011) e Perez-Elizalde, Jarquin e Crossa (2012) mostraram como incorporar inferência aos escores genotípicos e ambientais no biplot AMMI-2 por meio de regiões de credibilidade bivariadas. As regiões de credibilidade no biplot são construídas para os dois primeiros componentes principais descritos pelos escores $(\lambda_1^{1/2} \boldsymbol{\alpha}_{i1}, \lambda_2^{1/2} \boldsymbol{\alpha}_{i2})$ e $(\lambda_1^{1/2} \boldsymbol{\gamma}_{j1}, \lambda_2^{1/2} \boldsymbol{\gamma}_{j2})$, com $i = 1, \dots, r$ e $j = 1, \dots, c$. Oliveira et al. (2015) e Silva et al. (2015) utilizaram as distâncias euclidianas dos pontos à origem das distribuições empíricas dos escores (OOMS, 2009) para

determinar elipses de credibilidade bivariadas, a 95 % de probabilidade, para os escores genotípicos e ambientais. Hu e Yang (2013) apresentam outros métodos para a construção de regiões de confiança bivariadas.

Interpretações, em relação a estabilidade e adaptabilidade, são baseadas na observação de posições e sobreposições entre as regiões de credibilidade nos quadrantes do biplot. Sobreposições entre essas regiões indicam que os valores dos escores não são estatisticamente diferentes entre si e, portanto, os genótipos ou ambientes podem ser interpretados como parte de um subgrupo homogêneo com características semelhantes relativas ao efeito da GEI. Por outro lado, valores de interação referentes aos escores cujas regiões de credibilidade incluem a origem (0,0) são considerados não estatisticamente diferentes de zero, no nível de credibilidade adotado, como destacam Crossa et al. (2011). Para o GGE, as interpretações no biplot são relativamente diferentes do AMMI, já que o efeito descrito pelos escores se referem a $G + GEI$, e devem ser realizadas como exemplificado na subseção 2.4.2.

Embora trabalhos como o de Crossa et al. (2011) ofereçam um procedimento metodológico para incorporar inferência ao biplot, trabalhos práticos de competição de cultivares ainda são escassos na literatura, mesmo já existindo rotinas disponíveis e até uma biblioteca específica no software R para isso, denominado “bayesammi” (YASEEN et al., 2018). Oliveira et al. (2018) e Júnior et al. (2018), por exemplo, se preocuparam em obter interpretações mais práticas para regiões de credibilidade. Contudo, é necessário mais pesquisas sobre o tema de forma a tornar as interpretações mais claras e imprimir maior aplicabilidade às versões bayesianas do AMMI e GGE.

2.6 Estimadores de encolhimento em modelos lineares-bilineares

Como já salientado, no decorrer do texto, parâmetros bilineares de modelos multiplicativos são obtidos por DVS e correspondem a uma mistura entre padrão e ruído. Por construção, os primeiros componentes principais recuperam mais padrão e descartam mais ruído em relação àqueles associados às dimensões subsequentes. Assim, a principal propriedade da classe GLBM é a possibilidade de melhorar a capacidade preditiva do modelo pela determinação do número ótimo de eixos a serem retidos na análise. Isso, geralmente, tem sido feito a partir de testes baseados em distribuições F , validação cruzada e procedimentos computacionais intensivos (CORNELIUS; SEYEDSADR; CROSSA, 1992; DIAS; KRZANOWSKI, 2003; GAUCH, 2006; FORKMAN; PIEPHO, 2014; HADASCH; FORKMAN; PIEPHO, 2017). Nesse sentido,

os termos bilineares dos modelos multiplicativos são truncados em algum ponto, acreditando-se que termos adicionais (além de tal ponto) têm pouca significância ou valor preditivo.

Outra proposta de ajuste, ainda no contexto de efeitos fixos, é baseada na utilização de estimadores de encolhimento para os termos bilineares (CORNELIUS; CROSSA; SEYED-SADR, 1996; CORNELIUS; CROSSA, 1999). Essa abordagem fundamenta-se na ideia de que médias fenotípicas OLS não são as melhores opções e não podem ser usadas em todas as situações (STEIN, 1956; JAMES; STEIN, 1961; EFRON, 1975). Todavia, relaxando a restrição de ausência de viés e utilizando o critério de mínimo erro quadrático médio (EQM) na avaliação de estimadores, é possível obter um estimador que produz um efeito de retração em estimativas médias convencionais, denominado estimador *shrinkage* (EFRON; MORRIS, 1977). Esses estimadores, geralmente, tem a forma de um produto entre um escalar, cujo valor está compreendido entre zero e um [0, 1], e um estimador de mínimos quadrados (ou de máxima verossimilhança).

Dentre os estimadores (ou preditores) que promovem encolhimento podem ser destacados os estimadores de James e Stein (1961), utilizado no contexto de efeitos fixos, os preditores bayesianos em que parâmetros são considerados variáveis aleatórias e o melhor preditor linear não viesado (BLUP) derivados em um modelo de efeito misto (RESENDE; DUARTE, 2007). Uma distinção conceitual entre modelos de efeitos fixos e aleatórios refere-se às estimativas (ou previsões) não encolhidas versus encolhidas dos efeitos. Geralmente, a propriedade de encolhimento está associada a efeitos aleatórios e não a efeitos fixos (CROSSA et al., 2011). Contudo, o método *shrinkage* descrito por Cornelius e Crossa (1999) não assume quaisquer suposições de efeitos aleatórios para os parâmetros e baseia-se apenas na restrição de mínimo EQM.

Para uma melhor exemplificação, considera-se aqui o exemplo dado por Silva (2014) abordando um modelo de interação entre dois fatores em que os efeitos são fixos, dado pela equação:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta} + \mathbf{W}\boldsymbol{\varphi} + \boldsymbol{\epsilon} \quad (2.62)$$

em que \mathbf{y} é o vetor de observações; $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ e $\boldsymbol{\varphi}$, são os vetores de efeitos de ambientes, genótipos e GEI, respectivamente. As matrizes \mathbf{X} , \mathbf{Z} e \mathbf{W} são de delineamento e o vetor $\boldsymbol{\epsilon}$ contém os erros experimentais, tal que $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, sendo $\mathbf{0}$ o vetor nulo, \mathbf{I} a matriz identidade e σ_e^2 a variância experimental.

Estimadores de mínimos quadrados da média ($\hat{\boldsymbol{\varphi}}$) e da variância (var($\hat{\boldsymbol{\varphi}}$)) para efeito da interação são dados por:

$$\begin{aligned}\hat{\boldsymbol{\varphi}} &= (\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top (\mathbf{y} - X\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta}) \\ \text{var}(\hat{\boldsymbol{\varphi}}) &= (\mathbf{W}^\top \mathbf{W})^{-1} \sigma_e^2\end{aligned}\quad (2.63)$$

Como ressaltado, o estimador shrinkage para $\boldsymbol{\varphi}$ é aquele que minimiza o EQM, sendo obtido da seguinte forma:

$$\begin{aligned}EQM(S_\varphi \hat{\boldsymbol{\varphi}}) &= E[(S_\varphi \hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi})(S_\varphi \hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi})^\top] = \\ &= (S_\varphi)^2 (\mathbf{W}^\top \mathbf{W})^{-1} \sigma_e^2 + (E(S_\varphi \hat{\boldsymbol{\varphi}}) - \boldsymbol{\varphi})(E(S_\varphi \hat{\boldsymbol{\varphi}}) - \boldsymbol{\varphi})^\top\end{aligned}\quad (2.64)$$

sendo S_φ o fator de encolhimento.

Tomando um termo genérico φ_{ij} , referente a combinação linha i e coluna j da matriz de interação, cujos elementos estão empilhados no vetor $\boldsymbol{\varphi}$ e sendo $\hat{\varphi}_{ij}$ o respectivo estimador, tem-se que:

$$\begin{aligned}EQM(S_\varphi \hat{\varphi}_{ij}) &= (S_\varphi)^2 \frac{\sigma_e^2}{l} + (E(S_\varphi \hat{\varphi}_{ij}) - \varphi_{ij})^2 = \\ &= (S_\varphi)^2 \frac{\sigma_e^2}{l} + (S_\varphi)^2 \varphi_{ij}^2 - 2S_\varphi \varphi_{ij}^2 + \varphi_{ij}^2.\end{aligned}\quad (2.65)$$

em que l é o número de repetições referente à combinação entre o fator β e o fator δ . O fator de encolhimento S_φ deve ser determinado de forma a minimizar o EQM. Para tanto, a expressão (2.65) deve ser derivada em relação a S_φ , ou seja

$$\frac{\partial (EQM(S_\varphi \hat{\varphi}_{ij}))}{\partial S_\varphi} = 2S_\varphi \frac{\sigma_e^2}{l} + 2k\varphi_{ij}^2 - 2\varphi_{ij}^2. \quad (2.66)$$

Igualando (2.65) a zero e resolvendo para S_φ tem-se a seguinte solução:

$$S_\varphi = \frac{l\varphi_{ij}^2}{\sigma_e^2 + l\varphi_{ij}^2} = \frac{\varphi_{ij}^2}{\varphi_{ij}^2 + \sigma_e^2/l}, \quad (2.67)$$

que pode ser reescrito como

$$\frac{\varphi_{ij}^2}{\varphi_{ij}^2 + \sigma_e^2/l} = \frac{\sigma_e^2/l - \sigma_e^2/l + \varphi_{ij}^2}{\varphi_{ij}^2 + \sigma_e^2/l} = \left(1 - \frac{\sigma_e^2/l}{\varphi_{ij}^2 + \sigma_e^2/l}\right). \quad (2.68)$$

O estimador shrinkage de φ_{ij} é então dado por

$$S_\varphi \hat{\varphi}_{ij} = \left(\frac{\varphi_{ij}^2}{\varphi_{ij}^2 + \sigma_e^2/l} \right) (y_{ij} - \bar{y}_{..} - \bar{y}_{.j} + \bar{y}_{..}) \quad (2.69)$$

ou, equivalentemente, por

$$\left(1 - \frac{\sigma_e^2/l}{\varphi_{ij}^2 + \sigma_e^2/l}\right) (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\dots}). \quad (2.70)$$

Essa solução, para efeitos fixos, é análoga ao BLUP de efeitos aleatórios. Sob um modelo linear de dois fatores com interação (como 2.62) para a resposta média da i -ésima cultivar (genótipo) com l repetições no j -ésimo local ($\bar{y}_{ij\cdot}$), supondo normalidade e independência, os BLUPs dos níveis de desempenhos realizados das cultivares em locais de testes são dados por (CORNELIUS; CROSSA; SEYEDSADR, 1996):

$$\bar{y}_{\dots} + S_\delta (\bar{y}_{i\cdot} - \bar{y}_{\dots}) + S_\beta (\bar{y}_{\cdot j} - \bar{y}_{\dots}) + S_\varphi (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\dots}) \quad (2.71)$$

em que \bar{y}_{\dots} , $\bar{y}_{i\cdot}$, $\bar{y}_{\cdot j}$ e $\bar{y}_{ij\cdot}$ são as médias: geral, do i -esimo genótipo, do j -ésimo ambiente e da célula referente a combinação ij , respectivamente. Os termos $S_\tau = (1 - 1/F_\tau)$, $S_\delta = (1 - 1/F_\delta)$ e $S_\varphi = (1 - 1/F_\varphi)$ são os fatores de encolhimento sendo que F_τ , F_δ e F_φ são estatísticas para teste F de efeitos principais de genótipos, de ambientes e interação, respectivamente, contra o quadrado médio do erro combinado, supondo que todas as razões F sejam maiores do que a unidade (CORNELIUS; CROSSA; SEYEDSADR, 1996; HENDERSON, 1984).

Os fatores de encolhimento são estimativas das funções dos componentes de variância, em especial, S_φ estima $l\sigma_\varphi^2 / (\sigma_e^2 + l\sigma_\varphi^2)$. No caso do modelo centrado em colunas (ou ambientes), ou seja, para o GGE, tem se que o BLUP de células é igual a $\bar{y}_{\dots} + S_\delta (\bar{y}_{i\cdot} - \bar{y}_{\dots}) + S_{GGE} (\bar{y}_{ij\cdot} - \bar{y}_{\cdot j})$, em que S_{GGE} estima $l\sigma_{GGE}^2 / (\sigma_e^2 + l\sigma_{GGE}^2)$.

Cornelius, Crossa e Seyedsadr (1996) argumentam contra a suposição de que as interações devam ser consideradas normais e identicamente distribuídas $N(0, \sigma_\varphi^2)$, como assumido no modelo linear aleatório. Segundo os autores supracitados, essa suposição implica que a interação em qualquer célula não é informativa sobre a interação que provavelmente ocorrerá em outra célula, ou seja, não contêm informações transferíveis (exceto que ela contribuirá para a estimativa da variância desconhecida σ_φ^2).

Para modelos multiplicativos, a ideia foi buscar um procedimento conciliador entre a metodologia de truncamento de modelos (para um determinado número de termos bilineares) com as propriedades do BLUP. Isto é, um método em que as estimativas obtidas explorem tanto os aspectos de reconhecimento de padrões, que estão implícitos em métodos utilizados para determinar o número de componentes a ser mantido no modelo, quanto os erros reduzidos

de estimativa (ou predição) resultantes da utilização de BLUPs, sem assumir uma distribuição comum para componentes principais.

Para obter os estimadores de encolhimento, considera-se que qualquer conjunto de valores para os parâmetros bilineares $\lambda_k \alpha_{ik} \gamma_{jk}$, que satisfaçam a restrições do modelo, tenha probabilidade diferente de zero de ser o conjunto realizado. Se tais conjuntos não forem todos igualmente prováveis pode-se, pelo menos, supor que α_{ik} e γ_{jk} , para qualquer k , sejam permutáveis. Além disso, cada componente bilinear de mínimos quadrados empíricos $\hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk}$, em que $\hat{\lambda}_k$, $\hat{\alpha}_{ik}$ e $\hat{\gamma}_{jk}$ são estimativas OLS, é visto como a soma de algum valor verdadeiro mais alguma perturbação aleatória $\eta_{ijk} = \hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk} - \lambda_k \alpha_{ik} \gamma_{jk}$ resultante da combinação dos valores realizados de efeitos bilineares com erros residuais. Nesse contexto, para os modelos AMMI, GREG e SREG tem-se que $\sum_i \sum_j \hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk} = \sum_i \sum_j \lambda_k \alpha_{ik} \gamma_{jk} = \sum_i \sum_j \eta_{ijk} = 0$.

Devido a pressuposição de permutabilidade, η_{ijk} e $\lambda_k \alpha_{ik} \gamma_{jk}$ são não correlacionados. Além disso, η_{ijk} também é não correlacionado com $\lambda_{k'} \alpha_{ik'} \gamma_{jk'}$ (para $k \neq k'$). Como consequência tem-se $E(\sum_i \sum_j \lambda_k \alpha_{ik} \gamma_{jk} \hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk}) = E(\lambda_k^2)$ e $E(\lambda_k^2) = E(\hat{\lambda}_k^2 - \lambda_k^2)$. Procedendo de forma análoga ao exemplo apresentado para o modelo (2.62), ou seja, considerando um estimador da forma $\sum_k S_k \hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk}$ para estimar a contribuição $\sum_k \lambda_k \alpha_{ik} \gamma_{jk}$ para a média de casela μ_{ij} , o erro quadrático médio de estimação por célula

$$EQM(\lambda_k) = E \left[\sum_i \sum_j \left(\sum_k S_k \hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk} - \sum_k \lambda_k \alpha_{ik} \gamma_{jk} \right)^2 \right] / rc \quad (2.72)$$

é minimizado se

$$S_k = \frac{E(\sum_i \sum_j \lambda_k \alpha_{ik} \gamma_{jk} \hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk})}{E(\sum_i \sum_j \hat{\lambda}_k^2 \hat{\alpha}_{ik}^2 \hat{\gamma}_{jk}^2)} = \frac{E(\lambda_k^2)}{E(\hat{\lambda}_k^2)}. \quad (2.73)$$

Demonstrações e detalhes algébricos relacionados às equações (2.72) e (2.73) podem ser encontrados em Cornelius, Crossa e Seyedsadr (1996), Cornelius e Crossa (1999) e Silva (2014).

Definindo u_k tal que $E(\eta_{ijk}^2) = u_k \sigma_e^2 / n$ e recordando que $E(\eta_{ijk}^2) = E(\hat{\lambda}_k^2) - E(\lambda_k^2)$, tem-se que $E(\hat{\lambda}_k^2) - E(\lambda_k^2) = u_k \sigma_e^2 / l = E(\hat{\lambda}_k^2) - S_k E(\hat{\lambda}_k^2)$. Na prática o termo u_k é desconhecido e alguma estimativa deve ser usada, sendo comum utilizar o número de parâmetros no k -ésimo termo multiplicativo menos o número de restrições sobre esses parâmetros. Essa escolha representa o grau de liberdade assumido para soma de quadrados $l\hat{\lambda}_k^2$ quando se utiliza o teste F de Gollob (1968) sob a hipótese de que $\lambda_k = 0$. Um conjunto de estimativas de encolhimento, baseadas no grau de liberdade de Gollob, pode ser utilizado para iniciar um

processo iterativo de simulação, visando melhorar as estimativas do grau de liberdade u_k e, subsequentemente, obter estimativas de encolhimento melhoradas com base nesse grau de liberdade. Isolando S_k na equação $E(\hat{\lambda}_k^2) - S_k E(\hat{\lambda}_k^2) = u_k \sigma_e^2 / l$ tem-se:

$$S_k = 1 - \frac{u_k \sigma_e^2 / l}{E(\hat{\lambda}_k^2)} = 1 - \frac{u_k \sigma_e^2}{l E(\hat{\lambda}_k^2)}. \quad (2.74)$$

Assim, $S_k = 1 - F_k^{-1}$, sendo que $F_k = l E(\hat{\lambda}_k^2) / u_k \sigma_e^2$. Estimando $E(\hat{\lambda}_k^2)$ por $\hat{\lambda}_k^2$ e σ_e^2 por S^2 (S^2 é a estimativa da variância média considerando todos os ensaios) tem-se $F_k = l \hat{\lambda}_k^2 / u_k S^2$. Como ressaltado, os graus de liberdade u_k são atribuídos segundo o método de Gollob (CORNELIUS; CROSSA, 1999) ou por simulação (CORNELIUS; CROSSA; SEYEDSADR, 1996) e produzem resultados similares. O uso de $\hat{\lambda}_k^2$ para estimar $E(\hat{\lambda}_k^2)$ é justificado ao se assumir que $\hat{\lambda}_k^2$ é o principal parâmetro para o conjunto de efeitos $\hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk}$ e $l \hat{\lambda}_k^2$ é a soma de quadrados devido a esse efeito. Isso condiz com a prática comum de usar somas de quadrados da análise de variância, ou quadrados médios, como estimativas próprias dos valores esperados (CORNELIUS; CROSSA, 1999).

Nessas circunstâncias, o fator de encolhimento é dado por $S_k = \max(1 - F_k^{-1}, 0)$, ou seja, será $1 - F_k^{-1}$ desde que se tenha $F_k > 1$ (consequentemente $S_k > 0$). Caso contrário tem-se $S_k = 0$. Como já referido, F_k resulta da soma de quadrados devido ao termo bilinear k dividido por uma estimativa da contribuição da variância dos erros àquela soma de quadrados. Assim, o fator de encolhimento é intuitivamente atraente, já que o mesmo estima a razão efeito/(efeito+ruído). Uma das particularidades do método é que eventualmente se verifica $S_k \hat{\lambda}_k \geq S_{k-1} \hat{\lambda}_{k-1}$, violando as restrições da DVS de que $\hat{\lambda}_{k-1} \geq \hat{\lambda}_k$, o que requer passos adicionais (*ad hoc*) para correções no algoritmo (CORNELIUS; CROSSA, 1999).

Uma versão bayesiana para o método *shrinkage* de efeitos fixos foi considerada por Silva et al. (2015) no ajuste do modelo AMMI. Para tanto, assume-se uma priori hierárquica amostrando o hiperparâmetro de escala dos valores singulares. Considerando o modelo (2.39) expresso na subseção 2.5 e atribuindo priori de Jeffreys para $\sigma_{\lambda_k}^2$ obtém-se a seguinte distribuição condicional completa *a posteriori* para λ_k :

$$\lambda_k | \dots \sim N^+ \left[\left(\mathbf{\Lambda}_k^\top \mathbf{\Lambda}_k + \frac{\sigma_e^2}{\sigma_{\lambda_k}^2} \right)^{-1} \mathbf{\Lambda}_k^\top \mathbf{M}_{k'}, \left(\mathbf{\Lambda}_k^\top \mathbf{\Lambda}_k + \frac{\sigma_e^2}{\sigma_{\lambda_k}^2} \right)^{-1} \sigma_e^2 \right] \quad (2.75)$$

sendo $\boldsymbol{\Lambda}_k = \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\gamma_k$, $\mathbf{M}_{k'} = \mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \sum_{k' \neq k} \lambda_{k'} \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_{k'})\mathbf{X}_2\gamma_{k'}$ e $\lambda_1 \geq \dots \geq \lambda_t \geq 0$.

A média da densidade (2.75) pode ser reescrita como

$$\begin{aligned} \left(\boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k + \frac{\sigma_e^2}{\sigma_{\lambda_k}^2} \right)^{-1} \boldsymbol{\Lambda}_k^\top \mathbf{M}_{k'} &= \left(\frac{\sigma_{\lambda_k}^2}{\sigma_{\lambda_k}^2 \boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k + \sigma_e^2} \right) \boldsymbol{\Lambda}_k^\top \mathbf{M}_{k'} = \left(\frac{\sigma_{\lambda_k}^2}{\sigma_{\lambda_k}^2 \boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k + \sigma_e^2} \frac{\boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k}{\boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k} \right) \boldsymbol{\Lambda}_k^\top \mathbf{M}_{k'} = \\ &= \left(\frac{\boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k \sigma_{\lambda_k}^2 + \sigma_e^2}{\sigma_e^2 + \boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k \sigma_{\lambda_k}^2} - \frac{\sigma_e^2}{\sigma_e^2 + \boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k \sigma_{\lambda_k}^2} \right) \frac{\boldsymbol{\Lambda}_k^\top \mathbf{M}_{k'}}{\boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k} = \left(1 - \frac{\sigma_e^2}{\sigma_e^2 + \boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k \sigma_{\lambda_k}^2} \right) \frac{\boldsymbol{\Lambda}_k^\top \mathbf{M}_{k'}}{\boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k}. \quad (2.76) \end{aligned}$$

Assim, o fator de encolhimento é dado por $1 - F^{-1}$, sendo $F^{-1} = \sigma_e^2 / (\sigma_e^2 + \boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k \sigma_{\lambda_k}^2)$, avaliando a importância do componente multiplicativo para o modelo. É nítida a semelhança com a forma algébrica do fator *shrinkage* para efeitos fixos e ambas as expressões equivalem a um teste *F* (análise de variância). Contudo, o método descrito por Cornelius e Crossa (1999) é aplicado sob a pressuposição de modelos para efeitos fixos, estando sujeito, por tanto, a várias limitações. Ao contrário, na inferência bayesiana os parâmetros são considerados variáveis aleatórias oferecendo grande flexibilidade à análise e ainda a possibilidade de incorporar informações além daquelas presentes nos dados experimentais. Além disso, o método MCMC não viola as restrições do modelo e não requer quaisquer procedimentos adicionais de correção no processo amostral iterativo.

Silva (2019) considerou a atribuição de priori de máxima entropia para as componentes de variância dos parâmetros singulares no modelo AMMI, constatando que sob essa hipótese o encolhimento das previsões para valores singulares associados a dimensões mais elevadas é ainda mais contundente. Como destacado na subseção 2.2.2, o método da máxima entropia pode ser utilizado para obter prioris quando a informação a ser incorporada toma forma de um vínculo em uma família de distribuições admissíveis (CATICHA, 2004). Além disso, a utilização da priori de máxima entropia pode imprimir maior eficiência ao algoritmo de amostragem, bem como ao processo de estimativa (SILVA et al., 2015).

2.7 Heterogeneidade de variância residual entre locais em análise de dados MET

Em modelos lineares usuais, pressuposições básicas, como a homogeneidade de variâncias, devem ser verificadas para que os resultados das análises sejam considerados válidos. Desvios de tais pressupostos podem comprometer a análise de variância (ANOVA). Efetuar a análise de variância sob heterocedasticidade dos erros faz com que ocorram perdas de eficiê-

cia na estimativa de tratamentos e na sensibilidade dos testes de significância, que serão proporcionais ao grau de severidade da heterogeneidade presente nos dados (COCHRAN, 1947; SCHEFFÉ, 1959; JOHNSON; WICHERN, 1998).

Em análise de dados METs, a maioria dos métodos estatísticos aplicados (a exemplo da análise conjunta de variância) assume variâncias idênticas dos erros residuais em todos os locais. Essa suposição, no entanto, pode não ser realista devido a diferentes circunstâncias, como a variabilidade da variância residual entre parcelas de local para local, variações genéticas diferentes entre anos/locais e fatores ambientais tais como o tipo de solo, temperatura, precipitações, dentre outros, fazendo com que alguns locais sejam mais parecidos do que outros (CASANOVES; MACCHIAVELLI; BALZARINI, 2005; CROSSA et al., 2006; HU; YAN; SHEN, 2013). Além disso, METs geralmente incluem vários genótipos e consequentemente o tamanho dos blocos devem ser grandes, sendo difícil garantir homogeneidade das parcelas no controle local (CASANOVES; MACCHIAVELLI; BALZARINI, 2007).

A aplicação de procedimentos paramétricos (concebidos sobre a premissa da homogeneidade) a dados heterogêneos pode resultar em avaliações imprecisas sobre o desempenho de genótipos (STROUP; BAENZIGER; MULITZE, 1994). Sabe-se que a heterogeneidade de variância entre os genótipos é afetada pela heterogeneidade de variância entre os ambientes e vice-versa (EDWARDS; JANNINK, 2006). Hu, Yan e Shen (2013) enfatizam os impactos da heterogeneidade da variância residual em testes de hipóteses considerando variedades em estágio avançado, em que o número de variedades é menor e os testes de hipóteses são de maior relevância. Para outros ensaios, especialmente os de avaliação de variedades de gerações anteriores (com número de variedades maior), o interesse principal não são os testes de significância, mas sim boas estimativas pontuais. Nesses casos, a heterogeneidade de variância do erro também traz implicações às análises. Assim, um exame detalhado sobre o comportamento do erro e que influência ele tem é recomendado para melhorar a avaliação de genótipos (HU; YAN; LI, 2014).

O trabalho inicial na modelagem de variâncias heterogêneas pode ser encontrado no melhoramento animal, enfatizando que ignorar essa fonte de variação reduz a resposta à seleção (HILL, 1984; FOULLEY; QUAAS, 1995). No contexto do melhoramento de plantas, modelos mistos têm mostrado utilidade (PIEPHO, 1997; SMITH; CULLIS; THOMPSON, 2001; PIEPHO et al., 2008; SMITH et al., 2015) e diferentes estruturas podem ser consideradas para acomodar variâncias específicas (SMITH; CULLIS; THOMPSON, 2005). Modelos mistos com

variâncias residuais heterogêneas permitem considerar eventuais diferenças na precisão em diferentes locais e melhorar a partição da variância total, nas variâncias atribuídas aos efeitos aleatórios incluídos na análise.

Nesse sentido, o método mais eficiente é a estimativa de todos os efeitos fixos e aleatórios em uma análise de estágio único com base em um modelo misto para dados replicados (SMITH; CULLIS; GILMOUR, 2001; WELHAM et al., 2010). Contudo, esse método exige uma grande carga computacional, podendo inviabilizar a análise para grandes conjuntos de dados METs, sobretudo se a estrutura de variância-covariância para a GEI for complexa. Nesses casos uma abordagem ponderada de dois estágios pode ser adequada (MÖHRING; PIEPHO, 2009; WELHAM et al., 2010; PIEPHO et al., 2012).

Versões em modelos mistos para o AMMI e o SREG foram propostas e levam a uma estrutura fatorial analítica para variância-covariância genética dos efeitos da GEI (AMMI) ou G+GE (SREG) (BURGUEÑO et al., 2008; BURGUEÑO et al., 2012). Essas abordagens oferecem vantagens estatísticas e biológicas em relação a métodos usualmente utilizados (que consideram efeitos como sendo fixos), dentre as quais se destaca a facilidade para acomodar a heterogeneidade de variância residual, além de fornecer as medidas de estabilidade e adaptabilidade em única abordagem (STEFANOVA; BUIRCHELL, 2010; CROSSA, 2012). Uma das críticas aos modelos mistos é a escolha de qual fator (ambiente ou genótipo) deve ser considerado aleatório (SMITH; CULLIS; THOMPSON, 2001; RESENDE, 2004; NUVUNGA et al., 2015). Algumas limitações também tem sido apontadas aos modelos FA como, por exemplo, a ocorrência de casos Heywood que tornam o modelo não identificável (SMITH; CULLIS; THOMPSON, 2001), bem como a dificuldade de construir intervalos de confiança exatos para os componentes de variância, uma vez que requerem suposições de normalidade assintótica (NUVUNGA et al., 2019).

No contexto dos modelos lineares-bilineares para efeitos fixos, aplicações do AMMI e GGE, geralmente, consideram variâncias residuais homogêneas, o que é bastante irrealista, já que a heterogeneidade de variâncias é geralmente encontrada em METs (SMITH; CULLIS; THOMPSON, 2005; CROSSA et al., 2006; EDWARDS; JANNINK, 2006; HU; YAN; LI, 2014). Todavia, existem procedimentos baseados em DVS que permitem lidar com a heterogeneidade, como o AMMI-ponderado descrito por Rodrigues et al. (2015). Na aplicação GGE biplot, o dimensionamento dos dados obtido escalando as médias das células pelos desvios padrão (ou erro padrão) fenotípicos e posterior ponderação pela raiz quadrada da herdabilidade-h,

em cada ambiente, é recomendado como um procedimento preliminar para remover o efeito da heterogeneidade nas análises gráficas (CROSSA; CORNELIUS, 1997; YAN; KANG, 2003; YAN, 2014). Mas estes procedimentos estão longe de ser um consenso pois essa análise ponderada atribui menos pesos para ambientes que têm um quadrado médio residual elevado. Além disso, na análise ponderada os pesos podem ser correlacionados com as respostas do rendimento no ambiente. Assim, pode ocorrer, por exemplo, que ambientes com rendimento elevado apresentem maior variância do erro e ambientes com baixos rendimentos apresentem variâncias de erro reduzidas, o que pode mascarar o verdadeiro desempenho de alguns genótipos em certos ambientes (CROSSA, 1990).

Outra opção vantajosa é a aplicação do método bayesiano, que tem sido reportada desde o inicio dos anos 90 no melhoramento animal (GIANOLA et al., 1992; SORENSEN; WAGEPETERSEN, 2003). Na ótica bayesiana, todas as quantidades desconhecidas na análise são consideradas aleatórias, evitando assim, a escolha de quais fatores devam ser considerados como aleatórios. Conforme já ressaltado, essa escolha não é concensual entre os pesquisadores na aplicação de modelos mistos (SMITH; CULLIS; THOMPSON, 2001). A integração da inferência bayesiana pelos melhoristas de plantas é um fenômeno mais recente como observado por Orellana (2012). Theobald, Talbot e Nabugoomu (2002), por exemplo, analisaram conjuntos de dados desbalanceados a partir de um modelo hierárquico bayesiano. O método bayesiano permite tratar modelos complexos com flexibilidade e estruturas de variâncias heterogêneas podem ser incorporadas sem maiores problemas (EDWARDS; JANNINK, 2006; COTES et al., 2006; BAUER et al., 2009; ORELLANA; EDWARDS; CARRIQUIRY, 2014; NUVUNGA et al., 2019).

Silva et al. (2019) modelaram variâncias heterogêneas entre os ambientes utilizando o modelo AMMI-Bayesiano. Para essa abordagem a distribuição condicional dos dados é $\mathbf{y}|\boldsymbol{\lambda}_k, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \mathbf{g}, \boldsymbol{\beta}, \sigma_{e_j}^2 \sim N(\boldsymbol{\mu}_{\mathbf{y}}, \mathbf{V})$, com $\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k$. Os parâmetros do modelo são como definidos na subseção (2.5), sendo que $\mathbf{V}_{n \times n}$ é uma matriz bloco diagonal composta por $\sigma_{e_j}^2$, com $j = 1, \dots, c$ (c é o número de ambientes), n_{e_j} o número de observações do ambiente j e $n = \sum_{j=1}^c n_{e_j}$. As densidades *a priori* são as mesmas atribuídas ao modelo (2.39) e para $\sigma_{e_j}^2$ são atribuidas prioris de jeffreys ($p(\sigma_{e_j}^2) \propto 1/\sigma_{e_j}^2$).

As densidades *a posteriori* condicionais completas também são desenvolvidas por Silva et al. (2019) e para os parâmetros lineares e bilineares são as seguintes:

i) Densidade condicional completa *a posteriori* para o vetor de efeitos principais de ambientes

$$\boldsymbol{\beta} | \dots \sim N \left[\left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 \right)^{-1} \mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{M}_\beta, \left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 \right)^{-1} \right] \quad (2.77)$$

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

ii) Densidade condicional completa *a posteriori* para o vetor de efeitos principais de genótipos

$$\mathbf{g} | \dots \sim N \left[\left(\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z} + \frac{1}{\sigma_g^2} \mathbf{I}_g \right)^{-1} \mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{M}_g, \left(\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z} + \frac{1}{\sigma_g^2} \mathbf{I}_g \right)^{-1} \right], \quad (2.78)$$

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

iii) Densidade condicional completa *a posteriori* para a variância genotípica

$$\sigma_g^2 | \dots \sim \text{Inv-}Esc - \chi^2 \left[n_g, \frac{\mathbf{g}^\top \mathbf{I}_g \mathbf{g}}{n_g} \right]. \quad (2.79)$$

iv) Densidades condicionais completas *a posteriori* para os parâmetros bilineares relacionados ao k -ésimo componente principal ($k = 1, \dots, t$)

$$\lambda_k | \dots \sim N^+ \left[\lambda_k - \left(\boldsymbol{\phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\phi}_k \right)^{-1} \boldsymbol{\phi}_k^\top \mathbf{V}^{-1} \mathbf{M}_{k'}, \left(\boldsymbol{\phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\phi}_k \right)^{-1} \right] \quad (2.80)$$

em que $\mathbf{M}_{k'} = \mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \mathbf{Z}g - \sum_{k' \neq k}^t \lambda_{k'} \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_{k'}) \mathbf{X}_2 \boldsymbol{\gamma}_{k'}$ e $\boldsymbol{\phi}_k = \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k$.

$$\boldsymbol{\alpha}_k | \dots \sim N \left[\left(\boldsymbol{\Lambda}_{\boldsymbol{\alpha}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\alpha}_k} \right)^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\alpha}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'}, \left(\boldsymbol{\Lambda}_{\boldsymbol{\alpha}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\alpha}_k} \right)^{-1} \right] \quad (2.81)$$

sendo $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}_k} = \lambda_k \text{diag}(\mathbf{X}_2 \boldsymbol{\gamma}_k) \mathbf{Z}$.

$$\boldsymbol{\gamma}_k | \dots \sim N \left[\left(\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \right)^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'}, \left(\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \right)^{-1} \right] \quad (2.82)$$

sendo $\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} = \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}_2$.

v) Para as componentes da variância $\sigma_{e_j}^2$, ($j = 1, \dots, c$), que compõe a diagonal de \mathbf{V} , as distribuições condicionais completas *a posteriori* podem ser representadas da seguinte forma:

$$P \left(\sigma_{e_j}^2 | \dots \right) \propto \left(\sigma_{e_j}^2 \right)^{\frac{n_{e_j}}{2}} \exp \left\{ -\frac{1}{2\sigma_{e_j}^2} (\mathbf{y}_{e_j} - \boldsymbol{\theta}_{e_j})^\top (\mathbf{y}_{e_j} - \boldsymbol{\theta}_{e_j}) \right\} \left(\sigma_{e_j}^2 \right)^{-1}. \quad (2.83)$$

Sendo essa distribuição proporcional a uma inversa-qui-quadrada-escalada denotada por:

$$\sigma_{e_j}^2 | \dots \sim \text{Inv-}E_{Sc} - \chi^2 \left[n_{e_j}, \frac{(\mathbf{y}_{e_j} - \boldsymbol{\theta}_{e_j})^\top (\mathbf{y}_{e_j} - \boldsymbol{\theta}_{e_j})}{n_{e_j}} \right] \quad (2.84)$$

em que \mathbf{y}_{e_j} e $\boldsymbol{\theta}_{e_j}$ são vetores de observações fenotípicas e de médias do ambiente j , respectivamente. Detalhes algébricos sobre a obtenção das densidades condicionais completas para todos os parâmetros são encontrados em Silva et al. (2019).

Para o modelo AMMI com homogeneidade de variâncias, Liu (2001) demonstrou que as densidades condicionais *a posteriori* para os vetores singulares, normalizados no subespaço amostral corrigido, são *VMF*. Mas sob a hipótese de heterogeneidade de variâncias entre ambientes, elas são gaussianas multivariadas como pode ser observado em (2.81) e (2.82). Isto também é relatado para versão bayesiana do modelo FA apresentada por Nuvunga et al. (2019), em que os escores fatoriais são amostrados diretamente utilizando as propriedades da decomposição espectral da matriz de variância-covariância genética.

Para exemplificação desse fato, considera-se o caso mais simples em que duas variáveis normais apresentam variâncias diferentes, ou seja, $X \sim N(m, \sigma_x^2)$, $Y \sim N(d, \sigma_y^2)$, e $\sigma_x^2 \neq \sigma_y^2$ (SILVA et al., 2019). Sendo assim, a densidade conjunta é uma distribuição normal bivariada com densidade dada por

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{(1-\rho^2)}} \exp \left\{ -\frac{1}{2} \left[\begin{array}{l} \left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \frac{x-\mu_X}{\sigma_X} \times \\ \times \frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \end{array} \right] \right\}. \quad (2.85)$$

Assumindo que as variáveis são independentes ($\rho = 0$), a equação (2.85) reduz-se a:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp \left\{ -\frac{1}{2} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\}. \quad (2.86)$$

Como ressaltado na equação (2.1) da subseção 2.1, a transformação de coordenadas retangulares para polares é obtida por $x = r \cos(\theta)$ e $y = r \sin(\theta)$ ($\theta \in [0, 2\pi]$), sendo as médias dadas por $m = t \sin(\mu)$ e $d = t \cos(\mu)$, em que t é a distância entre o ponto (m, d) à origem do sistema e μ o ângulo formado por esse ponto com o eixo das abscissas. Além disso, o jacobiano da transformação é igual a r (OLIVEIRA et al., 2015). Nessas termos, a distribuição conjunta para θ e r é dada por:

$$f_{\Theta,R}(\theta, r) = r f_{X,Y}(r \cos \theta, r \sin \theta) \quad (2.87)$$

$$f_{\Theta,R}(\theta, r) = \frac{r}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[\frac{(r\cos\theta - t\cos\mu)^2}{\sigma_x^2} + \frac{(r\sin\theta - t\sin\mu)^2}{\sigma_y^2} \right] \right\}. \quad (2.88)$$

Considerando $r = t = 1$, tem-se

$$f_{\Theta,R}(\theta, r) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[\frac{(\cos\theta - \cos\mu)^2}{\sigma_x^2} + \frac{(\sin\theta - \sin\mu)^2}{\sigma_y^2} \right] \right\} \quad (2.89)$$

que também pode ser reescrita como:

$$f_{\Theta,R}(\theta, r) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2\sigma_x^2\sigma_y^2} [\sigma_y^2(\cos\theta - \cos\mu)^2 + \sigma_x^2(\sin\theta - \sin\mu)^2] \right\}. \quad (2.90)$$

Assim, a densidade (2.90) não é *VMF* e, consequentemente, cada dimensão do vetor singular é ponderada pela variância ambiental. Além disso, para considerar o vetor no subespaço correto (com restrições de ortonormalidade), é necessário realizar uma transformação ortogonal como descrita na subseção 2.5.2.

A amostragem dos vetores nesse caso também é realizada no subespaço corrigido, com correto suporte, pela definição de variáveis auxiliares $\tilde{\alpha}_k = \mathbf{H}_k^\top \alpha_k$ e $\tilde{\gamma}_k = \mathbf{D}_k^\top \gamma_k$, em que \mathbf{H}_k e \mathbf{D}_k são matrizes de transformação linear ortogonal. A distribuição condicional completa *a posteriori* para o k -ésimo vetor singular genotípico, no subespaço corrigido, é:

$$\tilde{\alpha}_k | \dots \sim N(\mathbf{m}_{\alpha_k}, \mathbf{W}_{\alpha_k}) \quad (2.91)$$

em que $\mathbf{m}_{\alpha_k} = \Delta_{\alpha_k} \left(\Delta_{\alpha_k}^\top \Delta_{\alpha_k} \right)^{-1/2}$, $\mathbf{H}_k^\top \left[\left(\Lambda_{\alpha_k}^\top \mathbf{V}^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]$ e

$$\mathbf{W}_{\alpha_k} = \left\{ \left[\left(\Delta_{\alpha_k}^\top \Delta_{\alpha_k} \right)^{-1/2} \right]^\top \mathbf{H}_k^\top \left(\Lambda_{\alpha_k}^\top \mathbf{V}^{-1} \Lambda_{\alpha_k} \right) \mathbf{H}_k \bar{m}_{\alpha_k} \left(\Delta_{\alpha_k}^\top \Delta_{\alpha_k} \right)^{-1/2} \right\}^{-1}. \quad (2.92)$$

Sendo $\bar{m}_{\alpha_k} = \Delta_{\alpha_k}^\top \Delta_{\alpha_k}$ uma constante normalizadora.

Então o vetor $\tilde{\alpha}_k$ é amostrado no subespaço corrigido em \mathbb{R}^{r-s} , sendo obtido no subespaço correto em \mathbb{R}^r pela transformação inversa $\alpha_k = \mathbf{H}_k \tilde{\alpha}_k$.

De maneira análoga, a densidade condicional completa *a posteriori* para $\tilde{\gamma}_k$, no subespaço corrigido, é dada por:

$$\tilde{\gamma}_k | \dots \sim N(\mathbf{m}_{\gamma_k}, \mathbf{W}_{\gamma_k}) \quad (2.93)$$

em que $\mathbf{m}_{\gamma_k} = \Delta_{\gamma_k} \left(\Delta_{\gamma_k}^\top \Delta_{\gamma_k} \right)^{-1/2}$, $\mathbf{D}_k^\top \left[\left(\Lambda_{\gamma_k}^\top \mathbf{V}^{-1} \Lambda_{\gamma_k} \right)^{-1} \Lambda_{\gamma_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]$ e

$$\mathbf{W}_{\gamma_k} = \left\{ \left[\left(\Delta_{\gamma_k}^\top \Delta_{\gamma_k} \right)^{-1/2} \right]^\top \mathbf{D}_k^\top \left(\Lambda_{\gamma_k}^\top \mathbf{V}^{-1} \Lambda_{\gamma_k} \right) \mathbf{D}_k \bar{m}_{\gamma_k} \left(\Delta_{\gamma_k}^\top \Delta_{\gamma_k} \right)^{-1/2} \right\}^{-1}, \quad (2.94)$$

sendo $\bar{m}_{\gamma_k} = \Delta_{\gamma_k}^\top \Delta_{\gamma_k}$ uma constante normalizadora. Com a mesma argumentação utilizada para vetores singulares genotípico, o vetor γ_k é obtido no subespaço correto em \mathbb{R}^c pela transformação inversa $\gamma_k = \mathbf{D}_k \tilde{\gamma}_k$.

As distribuições condicionais completas para todos os parâmetros podem ser amostradas diretamente utilizando um amostrador de Gibbs e um algoritmo é descrito por Silva et al. (2019).

A modelagem bayesiana oferece várias vantagens quando comparadas aos métodos convencionais para efeitos fixos. Ressaltamos algumas delas no decorrer do texto como, por exemplo, a flexibilidade para lidar com conjuntos de dados desbalanceados e heterocedásticos e ainda para incorporar inferência aos parâmetros utilizando a distribuição posterior. Além disso, quando as distribuições posteriores estão disponíveis, os métodos bayesianos fornecem inferência válida sem contar com a pressuposição de normalidade assintótica. Isso representa uma vantagem, pois na teoria assintótica as propriedades dos estimadores de máxima verossimilhança são garantidas apenas para um número n grande de observações.

Uma desvantagem do método bayesiano, em relação à análise frequentista, seria a maior demanda por recursos e tempo computacionais devido a amostragem MCMC e demais análises. Essa situação, entretanto, pode ser amenizada pelo recente desenvolvimento tecnológico, bem como pelo trabalho de otimização de algoritmos. Além disso, essa desvantagem é compensada pela maior precisão no processo de inferência que geram mais informações para a tomada de decisões.

Nos capítulos que se seguem, compondo a segunda parte dessa tese, dois artigos de pesquisa são apresentados e tratam do desenvolvimento de duas versões do modelo GGE. Nesses estudos, desenvolvem-se novas especificações *a priori* para analisar GEI em MET visando superar problemas com os métodos atuais.

REFERÊNCIAS

- ALBERT, J. **Philosophical theories of probability**. New York: Springer Science & Business Media, 2007. 273 p.
- ANDERSON, T. W. **An introduction to multivariate statistical analysis**. 3. ed. New Jersey: Wiley, 2003. 752 p.
- ANDRIEU, C. et al. An introduction to MCMC for machine learning. **Machine Learning**, Springer, v. 50, n. 1-2, p. 5-43, 2003.
- ASFAW, A. et al. AMMI and SREG GGE biplot analysis for matching varieties onto soybean production environments in Ethiopia. **Scientific Research and Essays**, Academic Journals, v. 4, n. 11, p. 1322–1330, 2009.
- BAIO, G.; BERARDI, A.; HEATH, A. **Bayesian cost-effectiveness analysis with the R package BCEA**. New York: Springer, 2017. 176 p.
- BAKER, R. Tests for crossover genotype-environmental interactions. **Canadian Journal of Plant Science**, NRC Research Press, v. 68, n. 2, p. 405-410, 1988.
- BALESTRE, M. et al. Yield stability and adaptability of maize hybrids based on GGE biplot analysis characteristics. **Crop Breeding and Applied Biotechnology**, v. 9, n. 3, p. 219-228, 2009.
- BANERJEE, A. et al. Clustering on the unit hypersphere using von Mises-Fisher distributions. **Journal of Machine Learning Research**, v. 6, p. 1345-1382, 2005.
- BAUER, A. et al. Bayesian prediction of breeding values by accounting for genotype-by-environment interaction in self-pollinating crops. **Genetics Research**, Cambridge University Press, v. 91, n. 3, p. 193-207, 2009.
- BERGER, J. et al. The case for objective bayesian analysis. **Bayesian analysis**, International Society for Bayesian Analysis, v. 1, n. 3, p. 385-402, 2006.
- BERGER, J. O.; BERNARDO, J. M. On the development of the reference prior method. **Bayesian statistics**, v. 4, n. 4, p. 35-60, 1992.
- BERNARDO, J. M. Reference posterior distributions for Bayesian inference. **Journal of the Royal Statistical Society. Series B Methodological**, v. 41, p. 113-147, 1979.
- BERNARDO, J. M.; SMITH, A. F. **Bayesian theory**. New York: J. Wiley, 1994. v. 405. 597 p.
- BOLSTAD, W. M.; CURRAN, J. M. **Introduction to Bayesian statistics**. 3. ed. New Jersey: John Wiley & Sons, 2017. 620 p.
- BORÉM, A.; MIRANDA, G. V.; FRITSCHE-NETO. **Melhoramento de plantas**. Viçosa: Editora UFV, Universidade Federal de Viçosa, 2017. 543 p.
- BOX, G. E.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: A. Wesley, 1973. 588 p.
- BRADU, D.; GABRIEL, K. R. The biplot as a diagnostic tool for models of two-way tables. **Technometrics**, Taylor & Francis, v. 20, n. 1, p. 47-68, 1978.

- BURGUEÑO, J. et al. Genomic prediction of breeding values when modeling genotype \times environment interaction using pedigree and dense molecular markers. **Crop Science**, Wiley Online Library, v. 52, n. 2, p. 707-719, 2012.
- BURGUEÑO, J. et al. Modeling additive \times environment and additive \times additive \times environment using genetic covariances of relatives of wheat genotypes. **Crop science**, Wiley Online Library, v. 47, n. 1, p. 311-320, 2007.
- BURGUEÑO, J. et al. Using factor analytic models for joining environments and genotypes without crossover genotype \times environment interaction. **Crop Science**, Wiley Online Library, v. 48, n. 4, p. 1291-1305, 2008.
- CARGNIN, A. et al. Interação entre genótipos e ambientes e implicações em ganhos com seleção em trigo. **Pesquisa Agropecuária Brasileira**, SciELO Brasil, v. 41, n. 6, p. 987-993, 2006.
- CARVALHO, C. G. P. d. et al. Interação genótipo \times ambiente no desempenho produtivo da soja no Paraná. **Pesquisa Agropecuária Brasileira**, SciELO Brasil, v. 37, p. 989-1000, 2002.
- CASANOVES, F.; MACCHIAVELLI, R.; BALZARINI, M. Error variation in multi-environment peanut trials: Within-trial spatial correlation and between-trial heterogeneity. **Crop science**, Wiley Online Library, v. 45, n. 5, p. 1927-1933, 2005.
- CASANOVES, F.; MACCHIAVELLI, R.; BALZARINI, M. Models for multi-environment yield trials with fixed and random block effects and homogeneous and heterogeneous residual variances. **The Journal of Agriculture of the University of Puerto Rico**, v. 91, n. 3-4, p. 117-131, 2007.
- CASELLA, G.; BERGER, R. L. C. **Inferência estatística**. São Paulo: Cengage Learning, 2011. 588 p.
- CATICHA, A. Relative entropy and inductive inference. In: AMERICAN INSTITUTE OF PHYSICS. **AIP conference proceedings**. [S.l.], 2004. v. 707, n. 1, p. 75-96.
- CHEN, M. H.; SHAO, Q. M. Monte Carlo estimation of bayesian credible and HPD intervals. **Journal of Computational and Graphical Statistics**, Alexandria, v. 8, n. 1, p. 69-92, 1999.
- COCHRAN, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. **Biometrics**, JSTOR, v. 3, n. 1, p. 22-38, 1947.
- COOPER, M.; DELACY, I. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. **Theoretical and Applied Genetics**, Springer, v. 88, n. 5, p. 561-572, 1994.
- CORNELIUS, P.; CROSSA, J.; SEYEDSADR, M. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In: _____. **Genotype-by-environment interaction**. Boca Raton: CRC, 1996. p. 199-234.
- 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.; SEYEDSADR, M.; CROSSA, J. Using the shifted multiplicative model to search for “separability” in crop cultivar trials. **Theoretical and Applied Genetics**, v. 84, n. 1/2, p. 161-172, 1992.

- 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.
- COTES, J. M. et al. A bayesian approach for assessing the stability of genotypes. **Crop Science**, Wiley Online Library, v. 46, n. 6, p. 2654-2665, 2006.
- COVER, T. M.; THOMAS, J. A. **Elements of information theory**. 2. ed. New Jersey: John Wiley & Sons, 2006. 774 p.
- CROSSA, J. Statistical analyses of multilocation trials. **Advances in agronomy**, Elsevier, v. 44, p. 55-85, 1990.
- CROSSA, J. From genotype \times environment interaction to gene \times environment interaction. **Current Genomics**, Bentham Science Publishers, v. 13, n. 3, p. 225-244, 2012.
- CROSSA, J. et al. Modeling genotype \times environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. **Crop science**, Wiley Online Library, v. 46, n. 4, p. 1722-1733, 2006.
- CROSSA, J.; CORNELIUS, P. L. Sites regression and shifted multiplicative model clustering of cultivar trial sites under heterogeneity of error variances. **Crop Science**, Wiley Online Library, v. 37, n. 2, p. 406-415, 1997.
- CROSSA, J.; CORNELIUS, P. L.; YAN, W. Biplots of linear-bilinear models for studying crossover genotype \times environment interaction. **Crop Science**, Wiley Online Library, v. 42, n. 2, p. 619-633, 2002.
- CROSSA, J. et al. Bayesian estimation of the additive main effects and multiplicative interaction model. **Crop Science**, Wiley Online Library, v. 51, n. 4, p. 1458-1469, 2011.
- CROSSA, J.; VARGAS, M.; JOSHI, A. K. Linear, bilinear, and linear-bilinear fixed and mixed models for analyzing genotype \times environment interaction in plant breeding and agronomy. **Canadian Journal of Plant Science**, NRC Research Press, v. 90, n. 5, p. 561-574, 2010.
- CROSSA, J.; YANG, R.-C.; CORNELIUS, P. L. Studying crossover genotype \times environment interaction using linear-bilinear models and mixed models. **Journal of Agricultural, Biological, and Environmental Statistics**, Springer, v. 9, n. 3, p. 362-380, 2004.
- CRUZ, C.; REGAZZI, A.; CARNEIRO, P. C. **Modelos biométricos aplicados ao melhoramento genético**. 4. ed. Viçosa: UFV, 2012. 514 p.
- DENIS, J.-B.; GOWER, J. C. Asymptotic covariances for the parameters of biadditive models. **Utilitas Mathematica**, v. 46, p. 193-205, 1994.
- DENIS, J.-B.; GOWER, J. C. Asymptotic confidence regions for biadditive models: Interpreting genotype-environment interactions. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley Online Library, v. 45, n. 4, p. 479-493, 1996.
- DIAS, C. T. dos S.; KRZANOWSKI, W. J. Model selection and cross validation in additive main effect and multiplicative interaction models. **Crop Science**, v. 43, n. 3, p. 865-873, 2003.

- EDWARDS, J. W.; JANNINK, J.-L. Bayesian modeling of heterogeneous error and genotype × environment interaction variances. **Crop Science**, Wiley Online Library, v. 46, n. 2, p. 820-833, 2006.
- EFRON, B. Biased versus unbiased estimation. **Advances in Mathematics**, Elsevier, v. 16, n. 3, p. 259-277, 1975.
- EFRON, B.; MORRIS, C. Stein's paradox in statistics. **Scientific American**, JSTOR, v. 236, n. 5, p. 119-127, 1977.
- FARR, W. M.; MANDEL, I.; STEVENS, D. An efficient interpolation technique for jump proposals in reversible-jump Markov chain Monte Carlo calculations. **Royal Society open science**, The Royal Society Publishing, v. 2, n. 6, p. 1-12, 2015.
- FORKMAN, J.; PIEPHO, H.-P. Parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models. **Biometrics**, Wiley Online Library, v. 70, n. 3, p. 639-647, 2014.
- FOULLEY, J. L.; QUAAS, R. Heterogeneous variances in Gaussian linear mixed models. **Genetics Selection Evolution**, BioMed Central, v. 27, n. 3, p. 1-18, 1995.
- FRANK, S. A.; SMITH, D. E. Measurement invariance, entropy, and probability. **Entropy**, Molecular Diversity Preservation International, v. 12, n. 3, p. 289-303, 2010.
- GABRIEL, K. R. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, v. 58, n. 3, p. 453-467, 1971.
- GABRIEL, K. R. Le biplot-outil d'exploration de données multidimensionnelles. **Journal de la Société française de statistique**, v. 143, n. 3/4, p. 5-55, 2002.
- GAMERMAN, D.; LOPES, H. F. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. 2. ed. New York: Chapman & Hall, 2006. v. 68. 344 p.
- GAUCH, H. **Statistical analysis of regional yield trials**: AMMI analysis of factorial designs. Amsterdam: Elsevier Science Publishers, 1992. 278 p.
- GAUCH, H. G. Statistical analysis of yield trials by AMMI and GGE. **Crop Science**, v. 46, n. 4, p. 1488-1500, 2006.
- GAUCH, H. G. A simple protocol for AMMI analysis of yield trials. **Crop Science**, Wiley Online Library, v. 53, n. 5, p. 1860-1869, 2013.
- GAUCH, H. G.; PIEPHO, H.-P.; ANNICHIAIRICO, P. Statistical analysis of yield trials by AMMI and GGE: Further considerations. **Crop Science**, Wiley Online Library, v. 48, n. 3, p. 866-889, 2008.
- GAUCH, J. H. Model selection and validation for yield trials with interaction. **Biometrics**, JSTOR, v. 44, n. 3, p. 705-715, 1988.
- GAUCH, J. H. G.; ZOBEL, R. W. Identifying mega-environments and targeting genotypes. **Crop Science**, Wiley Online Library, v. 37, n. 2, p. 311–326, 1997.
- GELFAND, A. E.; SMITH, A. F. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, Alexandria, v. 85, n. 410, p. 398-409, 1990.

- GELMAN, A. et al. **Bayesian data analysis**. 2. ed. Boca Raton: CRC, 2003. 665 p.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. **IEEE Transactions on pattern analysis and machine intelligence**, v. 6, n. 6, p. 721-741, 1984.
- GHOSH, J. K.; DELAMPADY, M.; SAMANTA, T. **An introduction to Bayesian analysis**: theory and methods. New York: Springer Science & Business Media, 2007. 354 p.
- GIANOLA, D. et al. Estimation of heterogeneous variances using empirical bayes methods: theoretical considerations. **Journal of Dairy Science**, Elsevier, v. 75, n. 10, p. 2805-2823, 1992.
- GILLIES, D. **Philosophical theories of probability**. New York: Psychology Press, 2000. 238 p.
- GOLLOB, H. F. A statistical model which combines features of factor analytic and analysis of variance techniques. **Psychometrika**, Springer, v. 33, n. 1, p. 73-115, 1968.
- GREENACRE, M. J. **Biplots in practice**. [S.l.]: Fundacion BBVA, 2010.
- GREGORY, P. **Bayesian logical data analysis for the physical sciences**: A comparative approach with mathematica support . New York: Cambridge University Press, 2005. 488 p.
- GUILLY, S. et al. Analysis of multienvironment trials (MET) in the sugarcane breeding program of Réunion Island. **Euphytica**, Springer, v. 213, n. 9, p. 213, 2017.
- HADASCH, S.; FORKMAN, J.; PIEPHO, H.-P. Cross-validation in AMMI and GGE models: A comparison of methods. **Crop Science**, Wiley Online Library, v. 57, n. 1, p. 264-274, 2017.
- HARTIGAN, J. et al. Invariant prior distributions. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 35, n. 2, p. 836-845, 1964.
- HASTIE, D. **Towards automatic reversible jump Markov chain Monte Carlo**. 2005. 222 p. Tese (Doutorado em Filosofia) — University of Bristol, 2005.
- HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. **Operations Research**, INFORMS, v. 31, n. 6, p. 1109–1144, 1983.
- HENDERSON, C. R. **Applications of linear models in animal breeding**. Ontário: University of Guelp, 1984. 462 p.
- HILL, W. On selection among groups with heterogeneous variance. **Animal Science**, Cambridge University Press, v. 39, n. 3, p. 473-477, 1984.
- HOFF, P. D. **A first course in Bayesian statistical methods**. New York: Springer, 2009. v. 580. 270 p.
- HU, X.; YAN, S.; LI, S. The influence of error variance variation on analysis of genotype stability in multi-environment trials. **Field Crops Research**, Elsevier, v. 156, p. 84-90, 2014.
- HU, X.; YAN, S.; SHEN, K. Heterogeneity of error variance and its influence on genotype comparison in multi-location trials. **Field Crops Research**, Elsevier, v. 149, p. 322-328, 2013.

- HU, Z.; YANG, R.-C. Improved statistical inference for graphical description and interpretation of genotype \times environment interaction. **Crop Science**, Wiley Online Library, v. 53, n. 6, p. 2400-2410, 2013.
- JAMES, W.; STEIN, C. Estimation with quadratic loss. In: SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 1961, Berkeley. **Proceedings of the Fourth Berkeley Symposium on Mathematics and Statistics**. Berkeley: University of California Press, 1961. v. 1, p. 361-379.
- JAMMALAMADAKA, S. R.; SENGUPTA, A. **Topics in circular statistics**. New York: World Scientific, 2001. v. 5. 322 p.
- JARQUÍN, D. et al. A hierarchical bayesian estimation model for multienvironment plant breeding trials in successive years. **Crop Science**, Wiley Online Library, 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.
- JIANG, J. **Large sample techniques for statistics**. 2. ed. New York: Springer Science & Business Media, 2010. 628 p.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice hall Upper, 1998. 816 p.
- JOSSE, J. et al. Another look at bayesian analysis of AMMI models for genotype-environment data. **Journal of Agricultural, Biological, and Environmental Statistics**, Springer, v. 19, n. 2, p. 240-257, 2014.
- JÚNIOR, L. A. Y. B. et al. AMMI bayesian models to study stability and adaptability in maize. **Agronomy Journal**, Wiley Online Library, v. 110, n. 5, p. 1765-1776, 2018.
- KEMPTON, R. The use of biplots in interpreting variety by environment interactions. **The Journal of Agricultural Science**, Cambridge, v. 103, n. 1, p. 123-135, 1984.
- KOCH, K.-R. **Introduction to Bayesian statistics**. 2. ed. Berlin: Springer Science & Business Media, 2007. 257 p.
- KULKARNI, V. G. **Modeling and analysis of stochastic systems**. 2. ed. New York: CRC Press, 2011. 328 p.
- KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. **The annals of mathematical statistics**, JSTOR, v. 22, n. 1, p. 79-86, 1951.
- LAVORANTI, O. J.; DIAS, C. d. S.; KRAZNOWSKI, W. J. Phenotypic stability via AMMI model with bootstrap re-sampling. **Pesquisa Florestal Brasileira**, Colombo, n. 54, p. 45-52, 2007.

- LEY, C.; VERDEBOUT, T. **Modern directional statistics**. Boca Raton: CRC Press, 2017. 191 p.
- LIU, G. **Bayesian computations for general linear-bilinear models**. 2001. 134 p. Tese (Doutorado em Filosofia) — University of Kentucky, Lexington, 2001.
- LIU, J. S.; WONG, W. H.; KONG, A. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. **Biometrika**, Oxford University Press, v. 81, n. 1, p. 27-40, 1994.
- MAIA, M. C. C. et al. AMMI-bootstrap no estudo da interação genótipos por ambientes em soja. **Revista de Matemática e Estatística**, v. 24, n. 3, p. 7-24, 2006.
- MANDEL, J. The partitioning of interaction in analysis of variance. **Journal of Research of the National Bureau of Standards, Series B**, v. 73, p. 309-328, 1969.
- MANDEL, J. A new analysis of variance model for non-additive data. **Technometrics**, Taylor & Francis Group, v. 13, n. 1, p. 1-18, 1971.
- MARDIA, K. V. **Statistics of directional data**. London: Academic Press, 1972. 367 p.
- MARDIA, K. V.; JUPP, P. E. **Directional statistics**. Chichester: John Wiley & Sons, 2000. 429 p.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. London: Academic Press, 1979. 521 p.
- MARIN, J.-M.; ROBERT, C. P. **A Practical Approach to Computational Bayesian Statistics**. New York: Springer Science & Business Media, 2007. 264 p.
- MASH'AL, M.; HOSSEINI, R. K-means++ for mixtures of von Mises-Fisher distributions. In: IEEE. **2015 7th Conference on Information and Knowledge Technology (IKT)**. [S.l.], 2015. p. 1-6.
- MÖHRING, J.; PIEPHO, H.-P. Comparison of weighting in two-stage analysis of plant breeding trials. **Crop Science**, Wiley Online Library, v. 49, n. 6, p. 1977-1988, 2009.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. Tokyo: McGraw-Hill, 1974. 564 p.
- NUÑEZ-ANTONIO, G.; GUTIÉRREZ-PEÑA, E. A bayesian analysis of directional data using the projected normal distribution. **Journal of Applied Statistics**, Taylor & Francis, v. 32, n. 10, p. 995-1001, 2005.
- NUVUNGA, J. et al. Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. **Genetics and Molecular Research**, v. 14, n. 4, p. 14262-14278, 2015.
- NUVUNGA, J. J. et al. Bayesian factor analytic model: An approach in multiple environment trials. **PloS One**, Public Library of Science San Francisco, CA USA, v. 14, n. 8, p. e0220290, 2019.
- OAKLEY, H. et al. Joint modeling of additive and non-additive genetic line effects in single field trials. **Theoretical and Applied Genetics**, Springer, v. 113, n. 5, p. 809-819, 2006.

- OLIVEIRA, L. A. **Regiões de credibilidade para escores genotípicos e ambientais em modelo AMMI com efeitos aleatórios para genótipos.** 2014. 134 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) — Universidade Federal de Lavras, 2014.
- 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. et al. Performance of cowpea genotypes in the brazilian midwest using the bayesian additive main effects and multiplicative interaction model. **Agronomy Journal**, Wiley Online Library, v. 110, n. 1, p. 147-154, 2018.
- OLIVEIRA, L. A. et al. Bayesian GGE biplot models applied to maize multi-environments trials. **Genetics and Molecular Research**, v. 15, n. 2, p. 1-21, 2016.
- OMER, S.; SINGH, M. Comparing bayesian and frequentist approaches for GGE bi-plot analysis in multi-environment trials in sorghum. **European Journal of Experimental Biology**, v. 7, n. 6, p. 1-6, 2017.
- OOMS, J. C. L. **The Highest Posterior Density Posterior Prior for Bayesian Model Selection.** 2009. 22 p. Dissertação (Mestrado em Pesquisa: Metodologia e Estatística) — Universiteit Utrecht, Utrecht, 2009.
- ORELLANA, M.; EDWARDS, J.; CARRIQUY, A. Heterogeneous variances in multi-environment yield trials for corn hybrids. **Crop Science**, Wiley Online Library, v. 54, n. 3, p. 1048-1056, 2014.
- ORELLANA, M. A. **Bayesian prediction of crop performance modeling genotype by environment interaction with heterogeneous variances.** 2012. 81 p. Tese (Doutorado em Filosofia) — Iowa State University, Ames, 2012.
- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana.** Lisboa: Serviço de Educação e Bolsas; Fundação Calouste Gulbenkian, 2003. 446 p.
- PEREZ-ELIZALDE, S.; JARQUIN, D.; CROSSA, J. A general bayesian estimation method of linear-bilinear models applied to plant breeding trials with genotype \times environment interaction. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 17, n. 1, p. 15-37, 2012.
- PIEPHO, H. Analysis of genotype-by-environment interaction and phenotypic stability. **Genotype-by-environment interaction**, CRC Press Boca Raton, FL, p. 151-174, 1996.
- PIEPHO, H. et al. Blup for phenotypic selection in plant breeding and variety testing. **Euphytica**, Springer, v. 161, n. 1-2, p. 209-228, 2008.
- PIEPHO, H. P. Analyzing genotype-environment data by mixed models with multiplicative terms. **Biometrics**, v. 90, n. 3/4, p. 761-766, 1997.
- PIEPHO, H.-P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, Springer, v. 97, n. 1-2, p. 195-201, 1998.
- PIEPHO, H.-P. et al. A stage-wise approach for the analysis of multi-environment trials. **Biometrical Journal**, Wiley Online Library, v. 54, n. 6, p. 844-860, 2012.

- PIEPHO, H.-P.; MÖHRING, J. Selection in cultivar trials—is it ignorable? **Crop Science**, Wiley Online Library, v. 46, n. 1, p. 192-201, 2006.
- PRESS, S. J. **Subjective and objective Bayesian statistics: Principles, models, and applications**. 2. ed. New Jersey: John Wiley & Sons, 2003. v. 590. 591 p.
- 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**, Hayward, v. 7, n. 4, p. 493-497, Nov. 1992.
- RENCHER, A. C.; SCHAAALJE, G. B. **Linear models in statistics**. New York: John Wiley & Sons, 2008. 672 p.
- RESENDE, M. D. V. **Métodos estatísticos ótimos na análise de experimentos de campo**. Colombo: Embrapa Florestas, 2004. 57 p.
- RESENDE, M. D. V. de; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v. 37, n. 3, p. 182-194, 2007.
- ROAZZI, A.; O'BRIEN, D. P.; DIAS, M. d. G. B. Sobre o debate freqüentista versus probabilista:"sorte de tolo"torna-se uma explicação plausível. **Psicologia: Reflexão e Crítica**, SciELO Brasil, v. 16, n. 2, p. 389-402, 2003.
- ROBERT, C. **The Bayesian choice: from decision-theoretic foundations to computational implementation**. 2. ed. New York: Springer Science & Business Media, 2007. 617 p.
- ROBERT, C.; CASELLA, G. **Monte Carlo statistical methods**. 2. ed. New York: Springer Science & Business Media, 2004. 683 p.
- RODRIGUES, P. C. et al. A weighted AMMI algorithm to study genotype-by-environment interaction and qtl-by-environment interaction. **Crop Science**, Wiley Online Library, v. 54, n. 4, p. 1555-1570, 2015.
- ROMÃO, R. 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, 2009.
- SCHEFFÉ, H. **The Analysis of Variance**. New York: John Wiley and Sons, 1959. 478 p.
- SEYEDSADR, M.; CORNELIUS, P. L. Shifted multiplicative models for nonadditive two-way tables. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 21, n. 3, p. 807-832, 1992.
- SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, Nokia Bell Labs, v. 27, n. 3, p. 379-423, 1948.
- SILVA, C. P. **Justificativa bayesiana para efeito de encolhimento em modelos AMMI**. 2014. 83 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) — Universidade Federal de Lavras, Lavras, 2014.
- SILVA, C. P. **Uso do algoritmo reversible jump Markov chain Monte Carlo para seleção de termos multiplicativos do modelo AMMI**. 2019. 128 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) — Universidade Federal de Lavras, Lavras, 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**, Wiley Online Library, v. 59, n. 6, p. 2455–2472, 2019.
- SILVA, W. C. J.; DUARTE, J. B. Métodos estatísticos para estudo de adaptabilidade e estabilidade fenotípica em soja. **Pesquisa Agropecuária Brasileira**, SciELO Brasil, v. 41, n. 1, p. 23–30, 2006.
- SIMPSON, D. et al. Penalising model component complexity: A principled, practical approach to constructing priors. **Statistical Science**, Institute of Mathematical Statistics, v. 32, n. 1, p. 1-28, 2017.
- SINGH, M.; CECCARELLI, S.; GRANDO, S. Genotype × environment interaction of crossover type: detecting its presence and estimating the crossover point. **Theoretical and Applied Genetics**, Springer, v. 99, n. 6, p. 988-995, 1999.
- SMITH, A.; CULLIS, B.; GILMOUR, A. Applications: the analysis of crop variety evaluation data in Australia. **Australian & New Zealand Journal of Statistics**, Melbourne, v. 43, n. 2, p. 129-145, 2001.
- SMITH, A.; CULLIS, B.; THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. **Biometrics**, Wiley Online Library, v. 57, n. 4, p. 1138-1147, 2001.
- SMITH, A.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, Cambridge University Press, v. 143, n. 6, p. 449-462, 2005.
- SMITH, A. B. et al. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. **Theoretical and Applied Genetics**, Springer, v. 128, n. 1, p. 55-72, 2015.
- SORENSEN, D.; GIANOLA, D. **Likelihood, Bayesian, and MCMC methods in quantitative genetics**. New York: Springer Science & Business Media, 2002. 739 p.
- SORENSEN, D.; WAAGEPETERSEN, R. Normal linear models with genetically structured residual variance heterogeneity: a case study. **Genetics Research**, Cambridge University Press, v. 82, n. 3, p. 207-222, 2003.
- STEFANOVA, K. T.; BUIRCHELL, B. Multiplicative mixed models for genetic gain assessment in lupin breeding. **Crop Science**, Wiley Online Library, v. 50, n. 3, p. 880-891, 2010.
- STEIN, C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 3., 1956, Los Angeles. **Proceedings...** Los Angeles: University of California, 1956. v. 1, p. 197-206.
- STONE, J. V. **Bayes' rule**: A tutorial introduction to Bayesian analysis. [S.l.]: Sebtel Press, 2013.

- STROUP, W. W.; BAENZIGER, P. S.; MULITZE, D. K. Removing spatial variation from wheat yield trials: a comparison of methods. **Crop Science**, Wiley Online Library, v. 34, n. 1, p. 62-66, 1994.
- TEMPLEMAN, A.; XINGSI, L. Maximum entropy and constrained optimization. In: **Maximum Entropy and Bayesian Methods**. [S.l.]: Springer, 1989. p. 447-454.
- TEODORO, P. E. et al. Adaptability of cotton (*Gossypium hirsutum*) genotypes analysed using a bayesian AMMI model. **Crop and Pasture Science**, CSIRO, v. 70, n. 7, p. 615-621, 2019.
- THEOBALD, C. M.; TALBOT, M.; NABUGOOMU, F. A bayesian approach to regional and local-area prediction from crop variety trials. **Journal of Agricultural, Biological, and Environmental Statistics**, Springer, v. 7, n. 3, p. 403-419, 2002.
- TIERNEY, L. Markov chains for exploring posterior distributions. **The Annals of Statistics**, JSTOR, p. 1701-1728, 1994.
- 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, p. 201-219, 2000.
- VILLEGAS, C. et al. On the a priori distribution of the covariance matrix. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 40, n. 3, p. 1098-1099, 1969.
- WAAGEPETERSEN, R.; SORENSEN, D. A Tutorial on Reversible Jump MCMC with a View toward Applications in QTL-mapping. **International Statistical Review**, Wiley Online Library, v. 69, n. 1, p. 49-61, 2001.
- WELHAM, S. J. et al. A comparison of analysis methods for late-stage variety evaluation trials. **Australian & New Zealand Journal of Statistics**, Wiley Online Library, v. 52, n. 2, p. 125-149, 2010.
- XU, N.-y. et al. The application of GGE biplot analysis for evaluating test locations and mega-environment investigation of cotton regional trials. **Journal of Integrative Agriculture**, Elsevier, v. 13, n. 9, p. 1921–1933, 2014.
- YAN, W. GGE biplot — A windows application for graphical analysis of multienvironment trial data and other types of two-way data. **Agronomy Journal**, Wiley Online Library, v. 93, n. 5, p. 1111-1118, 2001.
- YAN, W. Singular-value partitioning in biplot analysis of multienvironment trial data. **Agronomy Journal**, Wiley Online Library, v. 94, n. 5, p. 990-996, 2002.
- YAN, W. Biplot analysis of incomplete two-way data. **Crop Science**, Wiley Online Library, v. 53, n. 1, p. 48-57, 2013.
- YAN, W. **Crop variety trials**: Data management and analysis. New York: John Wiley & Sons, 2014. 361 p.
- YAN, W. Mega-environment analysis and test location evaluation based on unbalanced multiyear data. **Crop Science**, Wiley Online Library, v. 55, n. 1, p. 113-122, 2015.
- YAN, W. Analysis and handling of $G \times E$ in a practical breeding program. **Crop Science**, Wiley Online Library, v. 56, n. 5, p. 2106-2118, 2016.

- YAN, W. et al. Two types of GGE biplots for analyzing multi-environment trial data. **Crop Science**, Wiley Online Library, v. 41, n. 3, p. 656-663, 2001.
- YAN, W. et al. Identifying essential test locations for oat breeding in eastern Canada. **Crop Science**, Wiley Online Library, v. 50, n. 2, p. 504-515, 2010.
- 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.
- YAN, W.; HOLLAND, J. B. A heritability-adjusted GGE biplot for test environment evaluation. **Euphytica**, Springer, v. 171, n. 3, p. 355-369, 2010.
- YAN, W.; HUNT, L. Interpretation of genotype × environment interaction for winter wheat yield in Ontario. **Crop Science**, Wiley Online Library, v. 41, n. 1, p. 19-25, 2001.
- YAN, W. et al. Cultivar evaluation and mega-environment investigation based on the GGE biplot. **Crop Science**, Wiley Online Library, v. 40, n. 3, p. 597–605, 2000.
- YAN, W.; KANG, M. S. **GGE biplot analysis:** A graphical tool for breeders, geneticists, and agronomists. Boca Raton: CRC Press, 2003. 286 p.
- YAN, W. et al. GGE biplot vs. AMMI analysis of genotype-by-environment data. **Crop Science**, Wiley Online Library, v. 47, n. 2, p. 643-653, 2007.
- YAN, W. et al. GGE biplot vs. AMMI graphs for genotype-by-environment data analysis. **Journal of the Indian Society of Agricultural Statistics**, Indian Society of Agricultural Statistics, v. 65, n. 2, p. 181-193, 2011.
- YAN, W.; RAJCAN, I. Biplot analysis of test sites and trait relations of soybean in Ontario. **Crop Science**, Wiley Online Library, v. 42, n. 1, p. 11-20, 2002.
- YAN, W.; TINKER, N. A. An integrated biplot analysis system for displaying, interpreting, and exploring genotype × environment interaction. **Crop Science**, Wiley Online Library, v. 45, n. 3, p. 1004-1016, 2005.
- YAN, W.; TINKER, N. A. Biplot analysis of multi-environment trial data: Principles and applications. **Canadian Journal of Plant Science**, NRC Research Press, v. 86, n. 3, p. 623-645, 2006.
- YANG, R.-C. Likelihood-based analysis of genotype–environment interactions. **Crop Science**, Wiley Online Library, v. 42, n. 5, p. 1434-1440, 2002.
- YANG, R.-C. Mixed-model analysis of crossover genotype–environment interactions. **Crop Science**, Wiley Online Library, v. 47, n. 3, p. 1051-1062, 2007.
- YANG, R.-C. et al. Biplot analysis of genotype × environment interaction: proceed with caution. **Crop Science**, v. 49, n. 5, p. 1564-1576, 2009.
- YASEEN, M. et al. **bayesammi:** Bayesian Estimation of the Additive Main Effects and Multiplicative Interaction Model. 2018. Disponível em: <<http://CRAN.R-project.org/package= bayesammi/>>. Acessado: 03 dez. 2018.

YASUTOMI, S.; TANAKA, T. Parameter estimation for von Mises-Fisher mixture model via gaussian distribution. In: IEEE. **Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific**. [S.l.], 2014. p. 1-5.

YOUNG, G. A.; SMITH, R. **Essentials of statistical inference**. New York: Cambridge University Press, 2005. v. 16. 238 p.

ZEFFA, D. M. et al. Genetic progress of seed yield and nitrogen use efficiency of brazilian carioca common bean cultivars using bayesian approaches. **Frontiers in Plant Science**, Frontiers, v. 11, p. 1-14, 2020.

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

SEGUNDA PARTE

ARTIGO 1

Shrinkage in the Bayesian analysis of the GGE model: a case study with simulation

Redigido conforme as normas da revista Plos One (versão revisada)

Shrinkage in the Bayesian analysis of the GGE model: a case study with simulation

Luciano Antonio de Oliveira¹, Carlos Pereira da Silva², Alessandra Querino da Silva¹,
Cristian Tiago Erazo Mendes², Joel Jorge Nuvunga³, Joel Augusto Muniz², Júlio Sílvio
de Sousa Bueno Filho², Marcio Balestre²

¹ Faculty of Exact Sciences and Technology (FACET), Federal University of Grande Dourados, Dourados, Mato Grosso do Sul, Brazil.

² Department of Statistics (DES), Federal University of Lavras, Lavras, Minas Gerais, Brazil.

³ Department of Agriculture, Chibuto College of Business and Entrepreneurship, Eduardo Mondlane University, Chibuto, Gaza, Mozambique.

*Corresponding author

E-mail: lucianoantonio@ufgd.edu.br (LAO)

Abstract

The linear-bilinear (or additive-multiplicative) model of the main effects of genotypes plus the genotype \times environment interaction — also called a Site Regression Model — has been widely used to analyze multi-environmental trials , and especially its graphical interpretation that considers only the first two main components. Many authors have noted the advantages from applying Bayesian inference in these models to replace the classic (fiducial or frequentist) approaches. One consequence of the Bayesian version in this case is the shrinkage of the posterior distribution for parameters that model the genotype \times environment interaction. This results in parsimonious models, eliminating parameters that would be present in a typical analysis of bilinear components. This work aims to extend the shrinkage method to estimators of the parameters that model the genotype + genotype \times environmental interaction effect that composes the multiplicative part of the model, using justification for prior maximum entropy. The results has showed that the Bayesian model with shrinkage effect is parsimonious, and estimates singular values with a greater contribution to the sum of squares of the genotype + genotype \times environmental interaction combinations, shrinking the smaller estimates zero. Hence, this method captured more patterns and discarded more noise than the typical Bayesian version, which assumed non-informative priors for multiplicative parameters.

Keywords: GGE model; genotype \times environment interaction; Bayesian method; shrinkage effect; bilinear parameters.

Introduction

In multi-environmental trials (MET), it is common to observe inconsistent genotype responses in different environments. This phenotypic instability in the face of environmental diversity is called genotype \times environment interaction (GEI) and introduces significant difficulty in selecting and widely recommendation of genotypes superior for the target environment [1–3].

Multiplicative (or linear-bilinear) models for fixed effects — such as the main additive effects and multiplicative interaction model (AMMI) and the site regression model (SREG), also known as the principal effects genotype model (G) plus GEI (G+GEI, or simply GGE) — are widely applicable in studying patterns of genotype

responses between environments [4,5]. As the bilinear parameters of these models are obtained by decomposition by singular values (DVS) of the GEI matrix (AMMI) or G + GEI (SREG), they offer a powerful tool to separate noise patterns, approaching high dimensional matrices by matrices with reduced dimensions [2,6,7]. Additionally, stability and adaptability can be directly interpreted by viewing biplots [4,8–12].

Bayesian inference can be used to overpower shortcomings of standard analyses of linear-bilinear models. Complex models with multivariate responses and heterogeneous variances can be dealt with proper and conjugate prior-likelihood specification [13,14]. Markov chain Monte Carlo methods (MCMC) made possible to sample from posterior distributions that would be unthinkable some decades ago. This is now feasible for broad class of models using standard statistical software like OpenBUGS, JAGS and Stan [15].

Among the many possibilities of summaries from the posterior introduced with MCMC we highlight [16–18]. For our study it is paramount to draw confidence regions for biplot representation of GEI from such complex models.

The Bayesian method as applied to multiplicative models has been proven useful in addressing fixed-effects models' limitations. The first works in this direction are by Viele and Srivivasan [19] and Liu [20], who conducted the sampling process in the adjustment of the AMMI model using a Markov chain Monte Carlo method (MCMC). Crossa et al. [21] and Perez-Elizalde et al. [22], in turn, showed how to incorporate inference to the AMMI-2 biplot through regions of credibility for scores that describe the effect of GEI. Literature also presents further development of this method [23–27].

Jarquin et al. [28] and Oliveira et al. [29] extended the Bayesian method to the SREG model using non-informative priors for all parameters and, similarly, revealed the method's flexibility in incorporating credible regions into the SREG-2 biplot. This is particularly important in the GGE biplot proposed by Yan et al. [30], which uses only the first two singular axes [31–35].

For AMMI analysis, Silva et al. [25] proposed a Bayesian approach analogous to the frequentist shrinkage method described by Cornelius et al. [36] and Cornelius and Crossa [37]. Such an approach favors obtaining more realistic estimates for bilinear parameters, which eliminates the need to use methods to determine the number of components to be retained in the model to explain the GEI. Based on this procedure, the estimates of the first singular values are more pronounced, and components related to

larger dimensions — which provide little contribution to explaining variations in the data) — are shrunk to zero.

Given the strong graphic appeal of the applications of the GGE biplot model and the first two main components' consequent role in this procedure, an exploration of the analytical properties of the shrinkage estimators for the parameters that describe the G + GEI effect using Bayesian inference would be highly relevant.

Thus, the current research aims to: i) illustrate how the effects of shrinkage can be extended to the GGE model using maximum entropy reasoning to derive prior distributions; ii) use cross-validation procedures to compare the models with and without shrinkage effects, and iii) show the implications and possible interpretations for the credibility regions incorporated in the biplot representation in a simulated scenario.

Materials and methods

Simulated data

To exemplify the method, a set of multi-environmental data was simulated with 20 genotypes (G1 to G20) evaluated in seven environments (E1 to E7) using a randomized complete block design with three replications. The main effects of genotypes (G), environments (E), and blocks (B) were simulated with Gaussian distributions $N(0, 4)$, $N(0, 4)$, and $N(0, 1)$, respectively.

Three distinct patterns were considered for the GEI: i) $\{G1, G2, G3, G4, G5\} \sim N(0, 4)$, with positive values in the $\{E1, E2, E3, E4\}$ subgroup and negative in $\{E5, E6, E7\}$; ii) $\{G6, G7, G8, G9, G10\} \sim N(0, 4)$, with negative values for $\{E1, E2, E3, E4\}$ and positive for $\{E5, E6, E7\}$, and iii) $\{G11, G12, \dots, G20\} \sim N(0, 1)$ for all environments. Genotypes are considered unstable in i) and ii) and stable in iii).

The simulated values were organized in a double-entry table and corrected for row and column effects, thus constituting the GEI interaction matrix. Each observation in the simulated set was obtained additively by the respective effects of genotypes, environments, blocks, and interactions, in addition to a general mean plus the simulated error of a normal $N(0, 6)$.

The GGE model

In the GGE model, the vector \mathbf{y} contains $n = v \times r$ phenotypic responses, where v is the number of genotypes repeated $r = b \times l$ in the combinations of blocks (b denotes the number of blocks) within locations (l denotes the number of locations). This is represented by:

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

where $\boldsymbol{\beta}_{rxl}$ indicates the vector of the effects of hierarchical blocks within locations. The terms λ_k , \mathbf{a}_k , and $\boldsymbol{\gamma}_k$ are, respectively, the singular value and the genotypic and environmental singular vectors related to the k th main component, in which $k=1, \dots, t$ and $t \{t = \min(v-1, l)\}$ is the rank of the matrix $\mathbf{GGE}_{(r \times c)}$. Subsequently, \mathbf{X}_1 , \mathbf{X}_2 , and \mathbf{Z} are design matrices, and the vector $\boldsymbol{\epsilon}_{n \times 1}$ contains the effects of experimental errors, with $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$; specifically, σ_e^2 is the residual variance, $\mathbf{0}_{n \times 1}$ the null vector, and \mathbf{I}_n the n order identity matrix. The sample data exhibits a normally multivariate conditional distribution $\mathbf{y} | \boldsymbol{\beta}, \lambda, \mathbf{a}, \boldsymbol{\gamma}, \sigma_e^2 \sim N(\boldsymbol{\theta}, \mathbf{I}_n \sigma_e^2)$, where $\boldsymbol{\theta} = \mathbf{X}_1 \boldsymbol{\beta} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z} \mathbf{a}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k$.

Assignment of prior information

After defining the model derived from the sampling process (likelihood specification), we assign prior distributions for all parameters in the model. This bring a huge methodological boost from Bayesian inference as we can quantify the degree of uncertainty (or belief) about each assumption or unobserved variables in the model using probabilistic reasoning [38,39].

The uncertainty regarding the model parameters was declared by means of the following prior densities: $\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \sigma_\beta^2 \sim N(\mathbf{0}, 10^8)$; $\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2 \sim N^+(0, \sigma_{\lambda_k}^2)$; $\mathbf{a}_k \sim$ spherical uniform in the corrected subspace; $\boldsymbol{\gamma}_k \sim$ spherical uniform in the corrected subspace; and $\sigma_e^2 \sim 1/\sigma_e^2$ [29]. Note that we need to specify hyperparameter values.

Those values chosen for $\boldsymbol{\beta}$, \mathbf{a}_k and $\boldsymbol{\gamma}_k$ reflect a vague knowledge about the parameters as is well justified by [40] and more recently by Zeng et al. [41]. For the

error variance σ_e^2 we used Jeffrey's prior specification, that reflects lack of knowledge about parameter, except its domain [40,42].

Special attention is paid to the prior density of λ_k , which is the positive normal subject to the order relationship $\lambda_1 \geq \dots \geq \lambda_t \geq 0$, with $\mu_{\lambda_k} = 0$ ($k = 1, \dots, t$). Two hypotheses regarding the scale parameters $\sigma_{\lambda_k}^2$ are considered:

- I. $\sigma_{\lambda_k}^2 = 10^8$, as noted by Oliveira et al. [29]. This model will be referred to as BGGE (Bayesian-GGE) model.
- II. The inverted gamma $\sigma_{\lambda_k}^2 \sim GI(a, b)$ choice is based on the concept of maximum entropy [43], with a degree of freedom $a = 1$ and scale parameter $b = 0$. This prior is the same as that obtained by Silva [44]. The model thus considered is the BGGE (Bayesian-GGE entropy) model. The S1 Appendix provides details on the maximum entropy principle and derivation of prior distributions used.

Complete posterior conditional density for parameters

From the Bayesian perspective all inference is made on posterior distribution $p(\boldsymbol{\theta} | \mathbf{y})$, that can be rewritten using Bayes Theorem as

$$p(\boldsymbol{\theta} | \mathbf{y}) = \frac{p(\mathbf{y}, \boldsymbol{\theta})}{p(\mathbf{y})} = \frac{p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \quad (2)$$

in which $p(\boldsymbol{\theta})$ is a prior density for $\boldsymbol{\theta}$, while $p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})$ is the joint density for data and parameters $(\mathbf{y}, \boldsymbol{\theta})$ and $p(\mathbf{y} | \boldsymbol{\theta})$ represents the sampling model (likelihood)[45]. Main inference is done on $p(\boldsymbol{\theta} | \mathbf{y})$ that is the posterior density given data and is proportional to the joint density distribution. Learning from data is then analogous to modifying $p(\boldsymbol{\theta})$ to $p(\boldsymbol{\theta} | \mathbf{y})$ using information from $p(\mathbf{y} | \boldsymbol{\theta})$. Likelihood function as well as joint posterior distribution derivation are described in S2 Appendix.

Full conditional posterior distributions are the same as presented by Oliveira et al. [29], exception when Hypothesis II holds, leading to a model with shrinkage.

Regarding the BGGE model, a hierarchical prior is assumed for λ_k at two levels, with the prior density of $\sigma_{\lambda_k}^2$ as an inverted-gamma obtained by the maximum

entropy method. Therefore, the complete conditional posterior distribution for λ_k under Hypothesis II (BGGE) is provided by

$$\lambda_k | \dots \sim N^+ \left[\left(\boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k + \frac{\sigma_e^2}{\sigma_{\lambda_k}^2} \right)^{-1} \boldsymbol{\Lambda}_k^\top \mathbf{M}_{k'}^\top, \left(\boldsymbol{\Lambda}_k^\top \boldsymbol{\Lambda}_k \right)^{-1} \sigma_e^2 \right] \quad (3)$$

where $\boldsymbol{\Lambda}_k = \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k$ and $\mathbf{M}_{k'} = \mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\mathbf{g} - \sum_{k' \neq k}^t \lambda_{k'} \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_{k'})\mathbf{X}_2\boldsymbol{\gamma}_{k'}$.

Subsequently, $\sigma_{\lambda_k}^2 | \dots \sim GI(a+1, \lambda_k^2 + 2b)$ has scale parameters $(\lambda_k^2 + 2b)$ and $(a+1)$ degrees of freedom. Hyperparameters for those prior distributions would be ($a=1$ and $b=0$), $\sigma_{\lambda_k}^2 | \dots \sim GI(2, \lambda_k^2)$ can be obtained. The S2 Appendix presents the further algebraic details.

The MCMC sampling and posterior inference

The sampling process was conducted using the Gibbs sampler, with an algorithm as described by Oliveira et al. [29]. To obtain the BGGE, hyperparameters $\sigma_{\lambda_k}^2$ ($k=1, \dots, t$) sampling was analogous to the scheme used by Silva et al.[25].

The singular vectors were sampled by defining the auxiliary variables with the correct support in the subspace corrected by orthogonal linear transformations, as per Viele and Srivivasan's [19] proposed method.

MCMC chains diagnostics as well as burn-in and thinning values used both Raftery and Lewis' [46] and Heidelberger and Welch's [47] criterion.

Marginal means, modes (MAP) and limits for highest posterior density (HPD, at 95% credibility) where then evaluated in the joint posterior sample. HPD were empirically constructed using Chen and Shao's [48] method. Bivariate regions (also with 95% credibility) have been derived to the biplot for the first two main components $(\lambda_1^{1/2}\boldsymbol{\alpha}_{i1}, \lambda_1^{1/2}\boldsymbol{\alpha}_{i2})$ and $(\lambda_1^{1/2}\boldsymbol{\gamma}_{i1}, \lambda_1^{1/2}\boldsymbol{\gamma}_{i2})$ with $i=1, \dots, v$ and $j=1, \dots, l$ using Hu and Yang's [49] method.

Prediction analysis and model selection

The cross-validation procedures consider three random loss scenarios (10%, 33%, and 50%) of genotypes, or 10-fold, 3-fold, and 2-fold, respectively. The models' predictive capacity was quantified by the mean of the sum of squares of the predicted residual errors (PRESS) and phenotypic correlation between the predicted value \hat{y}_{ij} and the observed y_{ij} (COR), as described in Nuvunga et al. [50].

The models were selected according to the number of bilinear components and with the following information criteria: the Bayesian information criterion, or BIC [51]; Akaike's [52] information criterion, or AIC; and the Akaike-Monte Carlo information criterion, or AICM [53].

The entire inference and simulation process was performed using R statistical software [54].

Results

Convergence of Markov chains and description of complete models

All variance components of the model had values for the dependency factor (I) less than five ($I<5$) [46]; and also passed the stationarity test according to the Heidelberger and Welch's criteria [47]. Linear parameters has even better convergence properties [55]. The S1 Fig presents graphs of the response trace and density of Markov chains for the complete models' residual variances, in which all components were retained. The random oscillations around a stable value indicate that the chains are independent samples of the joint posterior distribution.

Table 1 displays the posterior means for the singular values from the BGGE and BGGE models, adjusted according to the number of bilinear terms. This table also presents the frequentist estimates (GGE-fixed) obtained from the singular values decomposition (SVD) in the G + GEI matrix. Summaries of the joint posterior distribution illustrates shrinkage effect of Bayesian posterior estimates, and especially from the third singular value for the BGGE models. The first two principal components (PCs) of the GGE-fixed and BGGE models explain 90% and 95% of the variance in G + GEI, respectively. For the BGGE model same components explained nearly all its variability (>99,9%).

Table 1. Point estimates of singular values for the BGGE and BGGEE family of models as a function of the number of bilinear terms ($k=1, 2, \dots, 7$).

BGGE Model							
k	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
1	46.29						
2	46.49	16.17					
3	46.52	16.39	9.74				
4	46.53	16.44	9.84	5.23			
5	46.57	16.51	9.90	5.41	2.12		
6	46.61	16.48	9.92	5.43	2.14	0.87	
7	46.56	16.46	9.91	5.40	2.13	0.86	0.42

BGGEE Model							
k	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
1	46.07						
2	46.29	15.69					
3	46.32	15.66	<0.001				
4	46.25	15.69	<0.001	<0.001			
5	46.27	15.65	<0.001	<0.001	<0.001		
6	46.31	15.75	<0.001	<0.001	<0.001	<<0.001	
7	46.31	15.69	<0.001	<0.001	<0.001	<<0.001	<<0.001
GGE-fixed	47.04	17.72	11.8	8.85	6.79	4.53	2.70

Nota bene: A<<B means A<0.01B.

The histograms (Fig 1) illustrate the posterior distributions of the singular values (complete models) for BGGE and BGGEE. The distributions for the first singular values are approximately symmetric (Gaussian). The distributions shift more and more to the left, with decreasing mean singular values. This is clearer for BGGEE and for λ_3 , in which the mean, mode, and median are very close to zero, indicating that this and the other components do little to explain the variation of the interaction.

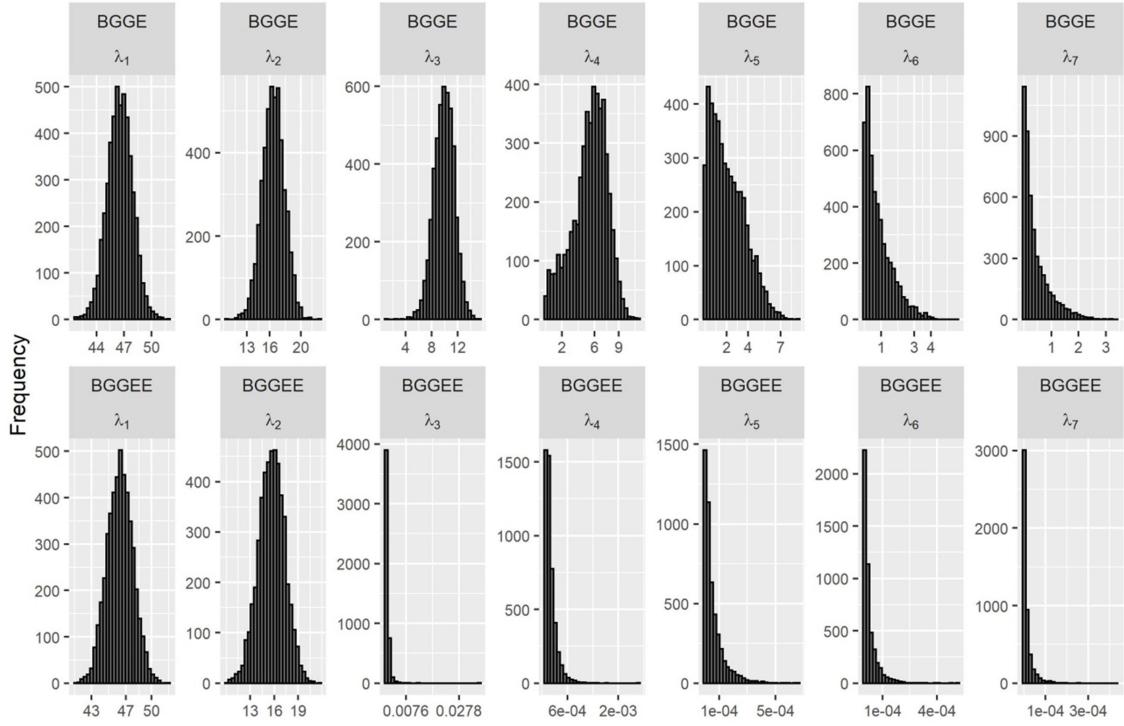


Fig 1. Histograms of the posterior marginal distributions of singular values, considering models of the complete dimensions for the proposed BGGE and BGGEE models.

Prediction analysis: Correlation and PRESS

Table 2 presents the analysis of cross-validation with correlations (COR) and PRESS in the proposed three scenarios (k -fold). The models are nearly indistinguishable for statistics calculated at the 10% level of random imbalances. However, the BGGEE model exhibits substantial advantages under a more severe unbalance. The S1 table notes the individual correlations for each k -fold ($k = 10, 3, 2$).

Table 2. Mean values of COR and PRESS for the BGGE and BGGEE models in three random, unbalanced scenarios.

BGGE Model			BGGEE Model	
Level (%)	COR	PRESS	COR	PRESS
10	0.78	9.13	0.78	8.82
33	0.58	14.46	0.64	13.34
50	0.46	31.11	0.71	10.69

Selection of models using information criteria

Fig 2 displays the values for the AIC, BIC, and AICM and σ_e^2 in the two models. The BGGE model has the lowest values for both BIC and AIC with two eigenvalues. Additionally, there is practically no difference between the models using more than two bilinear components greater than two for the AICM; the residual variance is stable thereafter, which implies that the BGGE-2 model is a parsimonious choice. The BGGE has a minimum of three (AICM), four (BIC), or five (AIC) components.

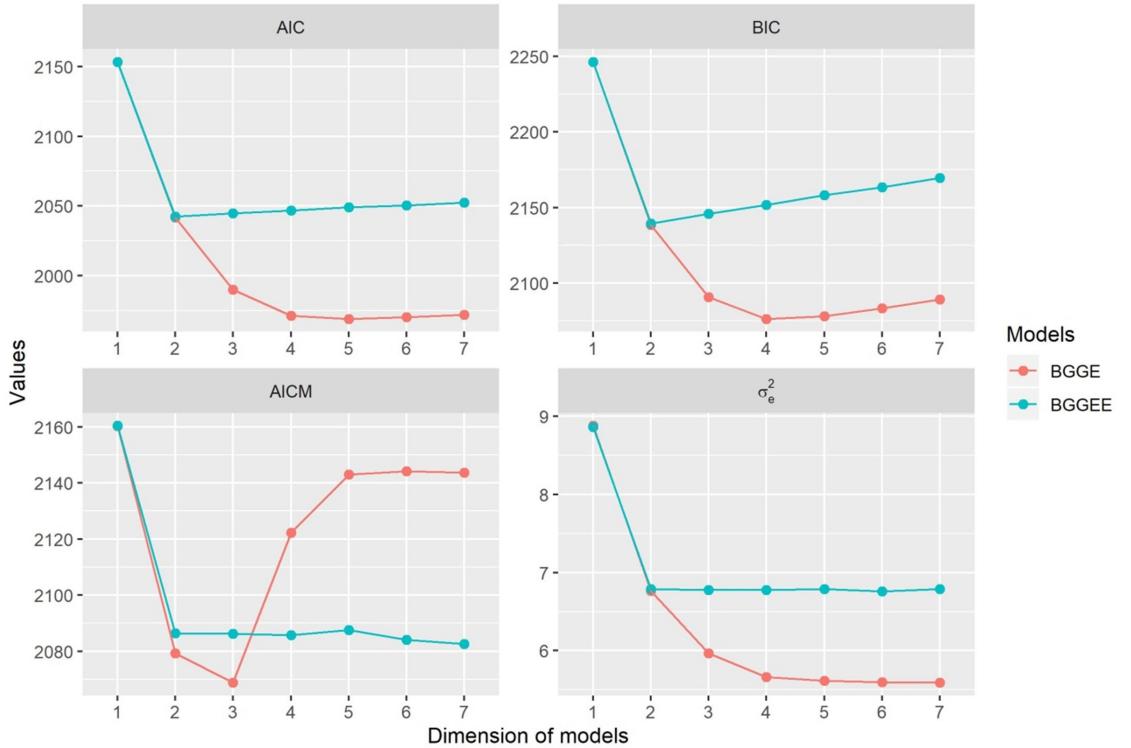


Fig 2. Results of the AIC, BIC and AICM and residual variance σ_e^2 , calculated for the BGGE- k and BGGEE- k family of models for $k = 1, \dots, 7$.

The information rate (IR) is typically included in GGE approaches that address fixed effects and quantifies each axis' contributions to explain the data [32]. This is also a criterion for separating noise patterns and size reduction. A value of $IR > 1$ indicates that the respective PC summarizes information for more than one variable; therefore, it displays patterns — or relationships. A value of $IR < 1$ reveals that previous PCs have already absorbed some information from the PC in question, and thus, its value can be considered noise [31,56].

The S2 Fig presents the posterior means and regions of HPD credibility for the information rates in the Bayesian models. Only the first PC stands out, in that $IR > 1$, indicating that the other PCs would not be informative. Regarding the frequentist GGE

model, IR values of 5.50 and 0.78 were obtained for PC1 and PC2, respectively. The IR values for subsequent PCs would be even lower due to the principal component analysis' properties.

Inferences from selected models

Inferences for the bilinear parameters and biplot analysis are considered only for the BGGE-3 and BGGE-2 models, selected using the AICM criterion. Table 3 illustrates the point and interval estimates of singular values and variance components for these models; no major discrepancies exist between the estimates for the first two singular values of the two models under analysis. The posterior mean of the BGGE model's residual variance is slightly lower as the HPD region does not capture the point estimate for the BGGE model's residual variance despite the intersection between the 95% credibility regions.

Table 3. Summaries of posterior distributions for singular values and variance components for the BGGE and BGGE models

Model	Parameter	Mean	Sd	HPD intervals (95%)	
				LL	UL
BGGE	λ_1	46.56	1.36	43.95	49.27
	λ_2	16.45	1.43	13.67	19.21
	λ_3	9.91	1.57	6.72	12.83
	σ_e^2	5.59	0.49	4.68	6.60
BGGEE	λ_1	46.31	1.50	43.34	49.17
	λ_2	15.69	1.64	12.52	18.92
	$\sigma_{\lambda_1}^2$	260.95	9584.96	132.68	6534.63
	$\sigma_{\lambda_2}^2$	60.68	1107.23	10.91	737.35
	σ_e^2	6.79	0.51	5.79	7.80

Sd, standard deviation; LL, lower limit; UL, upper limit.

Means and bivariate credibility regions based joint posterior distribution are incorporated into the biplot representation (Fig 3). This graph illustrates the "representativeness versus discrimination" graphic configuration [12]. Environmental vectors are included to allow for an interpretation of the angles between environments as well as their discriminative capacity. An additional average environment axis (AEA)

was drawn from the origin of the biplot passing through the environment's midpoint to indicate its representativeness relative to the target environment. It was observed that environments E4 and E5 were the most representative of the macro-environment in question. Environments E3, E6, and E7 are farther from the origin, with longer vectors and greater discriminative capacity.

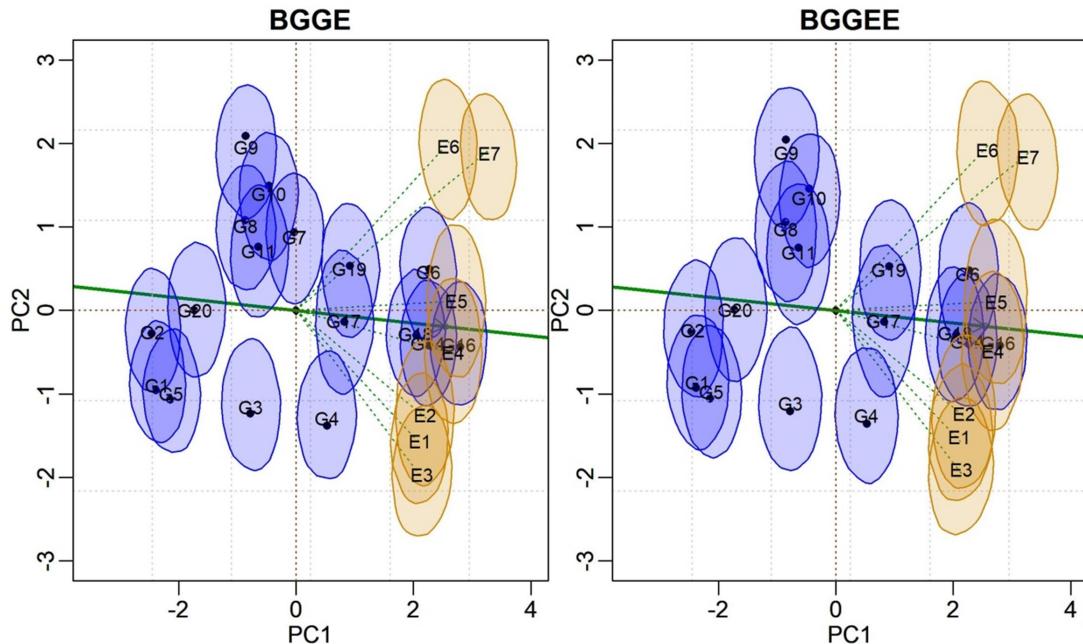


Fig 3. Biplots with bivariate regions at 95% credibility for the BGGE and BGGEE models' genotypic and environmental scores.

The patterns displayed in both biplots are similar, but the G7 in the BGGE biplot is not stable (the origin is not likely). This model allowed a crisper separation between some subgroups of genotypes and environments. The two environmental subgroups {E6, E7} and {E1, E2, E3} are clearly separable, as they are located in different quadrants. Environments E4 and E5 have credibility regions dispersed along the first and fourth quadrants.

Biplot visualization allows for the identification of genotypes with better performance (above general mean). Those are positioned to the right of PC1 (or the AEA). Those with responses below average are to the left. Genotypes with bivariate regions encompassing the biplot's origin constitute another subgroup in which yields do not statistically differ from the general mean, and the interaction effect is not relevant. Those genotypes were not plotted to simplify interpretations.

However, this global analysis is typically conducted in a preliminary sense. The recommendation is that evaluations of environments, as well as the selection and recommendations of genotypes be carried out within each mega-environment [1,31].

Two mega-environments may exist — MEGA1 = {E1, E2, E3, E4} and MEGA2 = {E5, E6, E7} — although overlaps occur. It should be noted that we use the term “mega-environment” for didactic purposes, as the proper environmental standard must be established by several years of experimentation. After mega-environment identification, further analyses were conducted for each, using BGGE and BGGE models.

The BGGE model

The configurations in Fig 4 is similar to that presented for the target environment (Fig 3). However, the average environment and ideal genotype were inserted, with their respective regions of bivariate credibility. This graphic form illustrates the “mean versus stability” pattern. Note that there is another axis through the origin but orthogonal to AEA. Genotypes that are farthest from the origin of this axis are those that contribute most to the interaction.

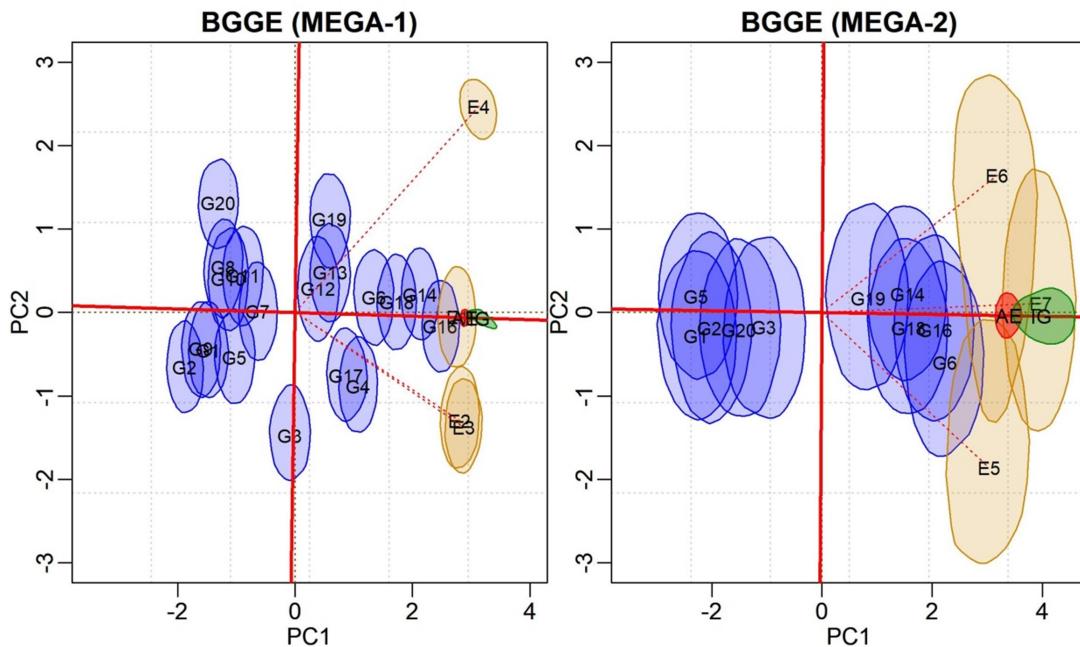


Fig 4. Biplot for the BGGE model with 95% credibility regions, including the genotypic and environmental scores for the average environment (red) and ideal genotype (green) for the two defined mega-environments.

From the biplots in Fig 4 we identify MEGA-1 as the most complex, as the E4 environment’s vector is nearly orthogonal to the {E2, E3} subgroup vector. E1 environment is the most representative this mega-environment, as its vector forms an acute angle with the AEA axis. Alternatively, the E4 environment has the greatest capacity for discriminating genotypes, although it is the least representative of this mega-environment. If this study’s objective was to select test sites, only one of the E2

or E3 environments would be selected, as they offer the same information. The G16 genotype would be the closest to the ideal for this mega-environment, followed by G14, with a region of overlapping credibility.

The MEGA-2 environment is simpler, despite a greater dispersion of its credible regions. For this mega-environment, E7 is the most representative and has the greatest discriminative capacity. It was not possible to highlight a single genotype with superior yield for MEGA-2, and a subgroup of genotypes {G6, G16, G14, G18, G19} have overlapping credibility regions for the distance to the ideal genotype .

The BGGEE model

Fig 5 depicts biplots for the mega-environments according to the BGGEE model. As the pattern in MEGA-1 is similar to that observed for BGGE model, the same conclusions follow. Regarding the MEGA-2, the second singular value has shrunk to near zero. This means only one main axis is needed to explain the G + GEI pattern. Consequently, this mega-environment is more homogeneous and the GEI has drastically decreased. It is again possible to observe a subgroup of genotypes with intersecting regions. Among these, the G6 and G16 genotypes exhibited the highest means.

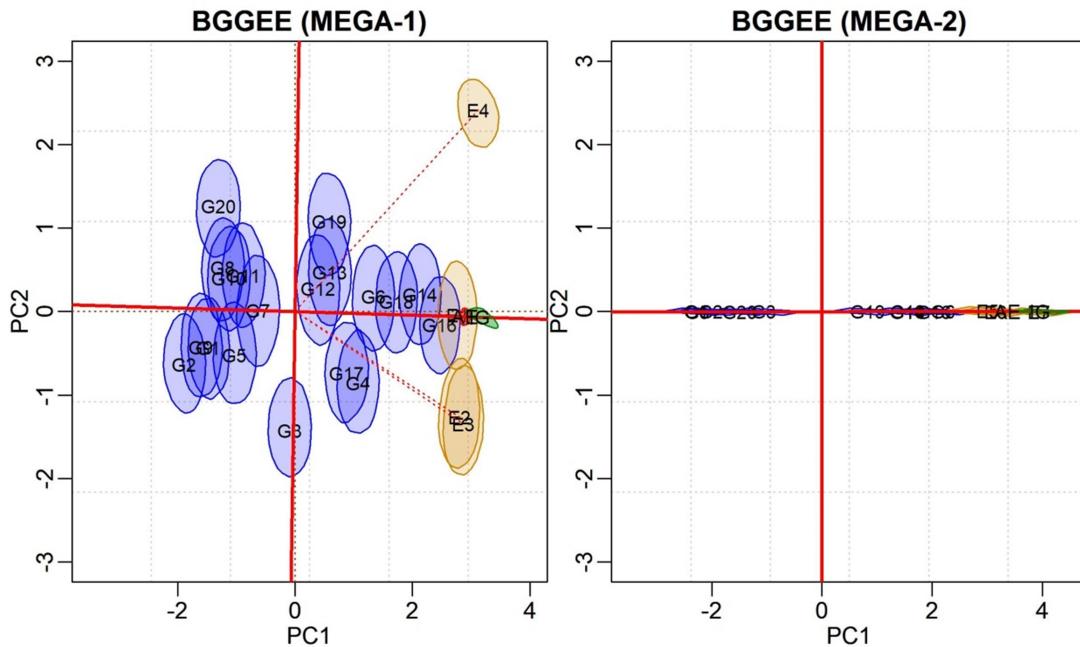


Fig 5. Biplot for the BGGEE model with 95% credibility regions, including the genotypic and environmental scores for the average environment (red) and ideal genotype (green) for to the two defined mega-environments.

Table 4 brings each environment's correlation with the average environment and its HPD regions. This correlation is approximated by the cosine of the angle formed

with the average environment. Note that no prior literature has presented credibility regions (or confidence limits) for this specific correlation as such an estimate would be difficult to deduce or approximate using fiducial or pure frequentist methods.

Table 4. Summaries for the posterior distribution of inner products for environments and average environment vectors.

Model	Mega	Env.	Mean	Median	Sd	LL	UL
BGGE	1	E1	0.99	0.99	0.006	0.98	1.00
		E2	0.91	0.92	0.029	0.86	0.97
		E3	0.91	0.91	0.028	0.85	0.96
		E4	0.77	0.77	0.025	0.72	0.82
	2	E5	0.85	0.85	0.076	0.73	1.00
		E6	0.87	0.87	0.072	0.75	1.00
		E7	0.99	0.99	0.019	0.95	1.00
BGGEE	1	E1	0.99	0.99	0.006	0.98	1.00
		E2	0.92	0.92	0.032	0.86	0.98
		E3	0.91	0.91	0.031	0.85	0.97
		E4	0.78	0.78	0.027	0.73	0.83
	2	E5	0.99	1.00	<<0.001	0.99	1.00
		E6	0.99	1.00	<<0.001	0.99	1.00
		E7	1.00	1.00	<<0.001	0.99	1.00

Sd, standard deviation; **LL**, lower limit; **UL**, upper limit; **Env.**, environment; **Mega**, mega-environments; A<<B means A<0.01B.

In Fig 6 are depicted posterior means of the distances from each genotype to the ideal, and respective HPD regions for each mega-environment, for both methods. These graphs display the distances to the ideal genotype inferred for G16 and G14 in the first mega-environment, and there is no relevant difference. G16 has marginally the smallest distance to the ideal among all genotypes. In the second mega-environment, G6 and G16 are the closest to the ideal genotype, although their credibility regions overlap with {G14, G18, G19} subgroup for the two models in question.

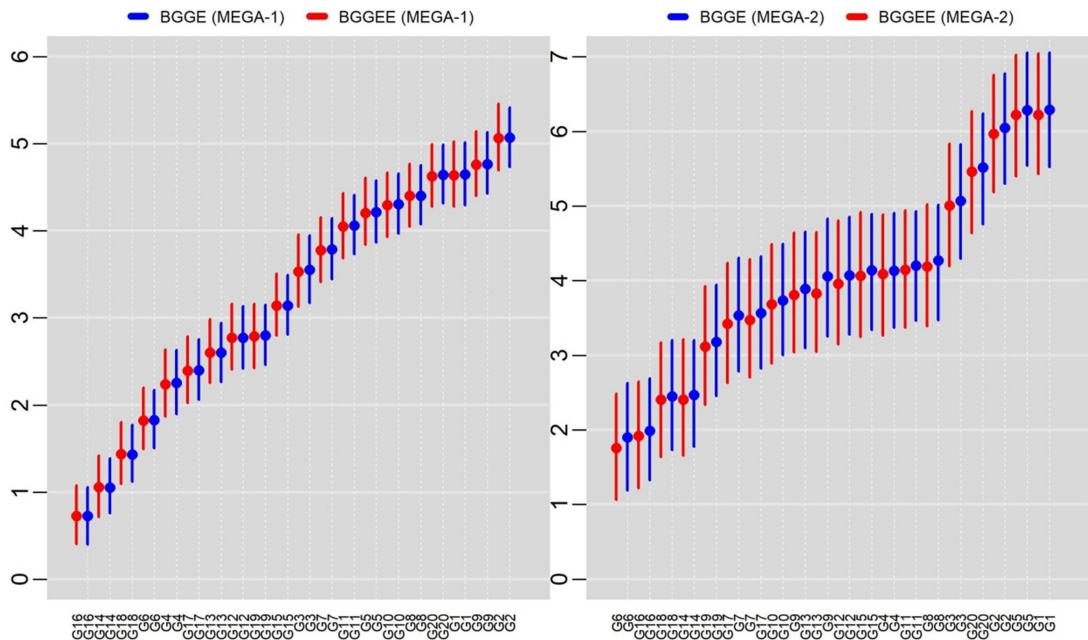


Fig 6. Mean scores and HPD (95% credibility) regions for distances of the genotypes to the ideal genotype in each mega-environment using the BGGE and BGGEE models.

The S3-S5 Figs illustrates the “who won where” configuration as a polygon with vertices (genotypes) furthest from the biplot’s origin. In this representation, orthogonal lines to the sides of these polygons pass through the origin, such that the genotypes represented at the vertices have higher yields in environments in the same sector[30].

Discussion

Bayesian approaches to the SREG model as implemented by Jarquin et al. [28] and Oliveira et al. [29] use objective *a priori* distributions for bilinear parameters; such approaches also produce estimates that not only have shrunk from the ordinary least-squares solutions, but are also similar to the best linear unbiased predictor [24,25].

The BGGEE is a Bayesian justification for the shrinkage method described by Cornelius et al. [36]. Although this method does not ensure unbiased estimators, it benefits from greater precision and restricts itself to the parametric space (a common feature for REML end Bayesian estimators [57,58]). Cornelius and Crossa [37] observed that the shrinkage method results in better models compared to those adjusted by F-tests, which are as good as — or superior to — those obtained by random models in general, and by best linear unbiased predictors in particular.

The shrinkage method described by Cornelius and Crossa [37] and Cornelius et al. [36] is based only on restricting the minimum mean-squared error (MSE). However, the authors note that this is restricted to balanced, homogeneous scenarios. The two-stage procedure changes the model specifications and involves *ad hoc* correction of the estimators.

The method presented here is sufficiently flexible to deal with missing data and heterogeneous variances, as well as to incorporate other information into the analysis (in the form of *a priori* distributions). The MCMC sampling method preserves the order and relationship of singular values and does not violate the model's assumptions.

In the simulated example, marked shrinkage occurred in the estimates of singular values in the BGGEE model relative to those in the BGGE model (Table 1). The first two axes explain nearly all the variation. This was also verified for the Bayesian version of AMMI as described by Silva et al. [25].

Although the information criteria are not entirely consistent, we determined the BGGEE-2 model is ideal; in this model, the AICM values (Fig 2) based on the second component stabilize and the singular values shrunk to zero (Table 1). Our choice for the AICM criterion follows Silva's [44] argument that the criterion results in more parsimonious, robust models given the different prior distributions. If the aim of the study is to infer various parameters of interest — such as genotypic predictions, environmental effects, or biplots, among others — the full model or that with two bilinear terms are equivalent. In other words, additional analyses and model selection would be unnecessary.

The information rate (IR) has been commonly used in applications to quantify each axis' contribution and explain the variability of G + GEI [32]. Only the first axis would be informative for our simulated data and considered hypotheses (S2 Fig).

More expressive shrinkage effects were not observed for the first two components of BGGEE relative to the DVS solutions to the GGE fixed model; specifically, these are the truly important dimensions in our analysis.

As a result of what was previously discussed, the prediction regarding the BGGEE model was consistently superior to that of the BGGE, especially with severe imbalance (Table 2). This highlights the BGGEE's superior ability to capture patterns and discard noise.

One major criticism of conventional biplot methods involves their descriptive, non-inferential nature. We use a flexible parametric procedure for biplot inferences

based on a posterior joint distribution to address a primary deficiency in previous methods [59–62]. Additionally, the biplot can incorporate such configurations as the “who won where” and “average environment axis” pattern as well as the regions of credibility for the average environment and ideal genotype.

Proposed method is flexible in associating uncertainty to other parameters of interest, such as the correlations between environments — approximated by the cosine of the angles — and distances from genotypes to the ideal genotype. In the cases, point estimates have always been presented in standard biplot analyses [56]. Similarly, inferences regarding virtually any other summaries can be incorporated into the GGE biplot configurations from the joint posterior samples.

A common criticism of Bayesian methods is their sensitivity to specifications of prior distributions, which can lead to significantly different results for the same likelihood. Eliciting prior distributions is very different than deriving them from maximum entropy principle [43,63]. We argue that this virtually solves the subjectivity problem. It is questionable if this is a desirable feature, but that is another discussion.

Using Jeffrey’s prior according to Silva et al. [25] results in scaled, inverse chi-squared conditional densities with one (1) degree of freedom, and consequently, improper marginal posterior distributions. To address this problem they used the extended prior specifications from Ter Braak et al. [64]. As a result, marginal posterior distributions may be bimodal. Inappropriate or bimodal posteriors did not occur when using the maximum entropy principle for prior specification. This was also by Silva [44] in the context of the AMMI model.

Bayesian inference using maximum entropy principle to specify prior distributions results in a more flexible SREG model. This model specification has a significant advantage over fixed-effects methods and its restrictive assumptions. In this sense, the method presented here is promising.

Conclusions and future research

Bayesian version of GGE using maximum entropy principle to prior specification made possible to make inferences marginal the model dimension. With this specification the full model and models with just two bilinear terms are almost indistinguishable, however, model selection is not a required step. This prior also avoids sampling problems observed with Jeffrey’s priors for variance components of bilinear

terms, resulting in proper and uni-modal marginal posterior distributions. With maximum entropy priors captured more patterns and discarded more noise than observed by other priors used in literature.

Bayesian GGE was very flexible in assigning genotypic and environmental scores in biplot. True posterior approximations for credibility regions for scores and also for the ideal genotype derived from the method were not yet presented for usual biplot techniques and are not easily available from frequentist or fiducial techniques.

Despite its great flexibility and applicability, this implementation is relatively difficult without suitable software, which is a work in progress. Other important practical developments that need to be tackled are generalizations for heterogeneous variance and non-continuous response variables. Object of future investigation.

References

1. Gauch HG, Zobel RW. Identifying Mega-Environments and Targeting Genotypes. *Crop Sci.* 1997;37(2):311–26. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.2135/cropsci1997.0011183X003700020002x>
2. Gauch HG. A Simple Protocol for AMMI Analysis of Yield Trials. *Crop Sci.* 2013;53(5):1860–9. Available from: <http://doi.wiley.com/10.2135/cropsci2013.04.0241>
3. Shahriari Z, Heidari B, Dadkhodaie A. Dissection of genotype × environment interactions for mucilage and seed yield in *Plantago* species: Application of AMMI and GGE biplot analyses. Chen Z-H, editor. *PLoS One*. 2018;13(5):e0196095. Available from: <https://dx.plos.org/10.1371/journal.pone.0196095>
4. Gauch Jr HG. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Amsterdam: Elsevier Science Publishers; 1992.
5. Crossa J, Cornelius PL. Sites Regression and Shifted Multiplicative Model Clustering of Cultivar Trial Sites under Heterogeneity of Error Variances. *Crop Sci.* 1997;37(2):406–15. Available from:

- <https://onlinelibrary.wiley.com/doi/abs/10.2135/cropsci1997.0011183X003700020017x>
6. Cornelius PL, Seyedsadr MS. Estimation of general linear-bilinear models for two-way tables. *J Stat Comput Simul.* 1997;58(4):287–322. Available from: <http://www.tandfonline.com/doi/abs/10.1080/00949659708811837>
 7. Piepho HP. Robustness of statistical tests for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trials. *Theor Appl Genet.* 1995;90(3–4):438–43. Available from: <http://link.springer.com/10.1007/BF00221987>
 8. Gabriel KR. The biplot graphic display of matrices with application to principal component analysis. *Biometrika.* 1971;58(3):453–67. Available from: <https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/58.3.453>
 9. Bradu D, Gabriel KR. The Biplot as a Diagnostic Tool for Models of Two-Way Tables. *Technometrics.* 1978;20(1):47–68. Available from: <http://www.tandfonline.com/doi/abs/10.1080/00401706.1978.10489617>
 10. Kempton RA. The use of biplots in interpreting variety by environment interactions [Internet]. Vol. 103, *The Journal of Agricultural Science.* 1984. p. 123–35. Available from: https://www.cambridge.org/core/product/identifier/S0021859600043392/type/journal_article
 11. Cooper M, DeLacy IH. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theor Appl Genet.* 1994;88(5):561–72. Available from: <http://link.springer.com/10.1007/BF01240919>
 12. Yan W. GGEbiplot-A Windows Application for Graphical Analysis of Multienvironment Trial Data and Other Types of Two-Way Data. *Agron J.* 2001;93(5):1111–8. Available from: <http://doi.wiley.com/10.2134/agronj2001.9351111x>

13. Ma X, Chen S, Chen F. Correlated random-effects bivariate poisson lognormal model to study single-vehicle and multivehicle crashes. *J Transp Eng.* 2016;142(11):4016049.
14. Zeng Q, Wen H, Huang H, Pei X, Wong SC. A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. *Accid Anal Prev.* 2017;99:184–91. Available from: <http://dx.doi.org/10.1016/j.aap.2016.11.018>
15. Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Stat Sci.* 2017;32(1):1–28. Available from: <http://projecteuclid.org/euclid.ss/1491465621>
16. Theobald CM, Talbot M, Nabugoomu F. A Bayesian approach to regional and local-area prediction from crop variety trials. *J Agric Biol Environ Stat.* 2002;7(3):403–19.
17. Edwards JW, Jannink J-L. Bayesian Modeling of Heterogeneous Error and Genotype × Environment Interaction Variances. *Crop Sci.* 2006;
18. Orellana M, Edwards J, Carriquiry A. Heterogeneous variances in Multi-Environment yield trials for corn Hybrids. *Crop Sci.* 2014;
19. Viele K, Srinivasan C. Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback–Leibler Information. *J Stat Plan Inference.* 2000;84(1–2):201–19. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0378375899001512>
20. Liu G. Bayesian computations for general linear-bilinear models. University of Kentucky; 2001.
21. Crossa J, Perez-Elizalde S, Jarquin D, Cotes JM, Viele K, Liu G, et al. Bayesian Estimation of the Additive Main Effects and Multiplicative Interaction Model. *Crop Sci.* 2011;51(4):1458–69. Available from: <http://doi.wiley.com/10.2135/cropsci2010.06.0343>

22. 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. *J Agric Biol Environ Stat.* 2012;17(1):15–37. Available from: <http://link.springer.com/10.1007/s13253-011-0063-9>
23. Josse J, van Eeuwijk F, Piepho H-P, Denis J-B. Another Look at Bayesian Analysis of AMMI Models for Genotype-Environment Data. *J Agric Biol Environ Stat.* 2014; Available from: <http://link.springer.com/10.1007/s13253-014-0168-z>
24. Antonio de Oliveira L, Pereira da Silva C, Nuvunga JJ, Da Silva AQ, Balestre M. Credible Intervals for Scores in the AMMI with Random Effects for Genotype. *Crop Sci.* 2015;55(2):465–76. Available from: <http://doi.wiley.com/10.2135/cropsci2014.05.0369>
25. da Silva CP, de Oliveira LA, Nuvunga JJ, Pamplona AKA, Balestre M. A Bayesian Shrinkage Approach for AMMI Models. Hsiao CK, editor. *PLoS One.* 2015;10(7):e0131414. Available from: <https://dx.plos.org/10.1371/journal.pone.0131414>
26. da Silva CP, de Oliveira LA, Nuvunga JJ, Pamplona AKA, Balestre M. Heterogeneity of Variances in the Bayesian AMMI Model for Multienvironment Trial Studies. *Crop Sci.* 2019;59(6):2455–72. Available from: <http://doi.wiley.com/10.2135/cropsci2018.10.0641>
27. Romão RF, Nuvunga JJ, Silva CP, Oliveira LA, Mendes CTE, Balestre M. Research Article Predictive ability of AMMI and factorial analytical models in the study of unbalanced multi-environment data. *Genet Mol Res.* 2019;18(3). Available from: http://www.funpecrp.com.br/gmr/articles/year2019/vol18-3/pdf/gmr18176_-_predictive-ability-ammi-and-factorial-analytical-models-study-unbalanced-multi-environment.pdf
28. Jarquín D, Pérez-Elizalde S, Burgueño J, Crossa J. A Hierarchical Bayesian Estimation Model for Multienvironment Plant Breeding Trials in Successive Years. *Crop Sci.* 2016;56(5):2260–76. Available from: <http://doi.wiley.com/10.2135/cropsci2015.08.0475>

29. de Oliveira LA, da Silva CP, Nuvunga JJ, da Silva AQ, Balestre M. Bayesian GGE biplot models applied to maize multi-environments trials. *Genet Mol Res.* 2016;15(2):1–21. Available from: <http://www.funpecrp.com.br/gmr/year2016/vol15-2/pdf/gmr8612.pdf>
30. Yan W, Hunt L a., Sheng Q, Szlavnics Z. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. *Crop Sci.* 2000;40(3):597–605. Available from: <http://doi.wiley.com/10.2135/cropsci2000.403597x>
31. Yan W, Kang MS. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists.* 2002.
32. Yan W, Tinker NA. Biplot analysis of multi-environment trial data: Principles and applications. *Can J Plant Sci.* 2006;86(3):623–45. Available from: <http://www.nrcresearchpress.com/doi/10.4141/P05-169>
33. Yan W, Kang MS, Ma B, Woods S, Cornelius PL. GGE Biplot vs. AMMI Analysis of Genotype-by-Environment Data. *Crop Sci.* 2007;47(2):643–53. Available from: <http://doi.wiley.com/10.2135/cropsci2006.06.0374>
34. Balestre M, Souza JC, Pinho RGV, Oliveira RL, Paes JMV. Yield stability and adaptability of maize hybrids based on GGE biplot analysis characteristics. *Crop Breed Appl Biotechnol.* 2009;9(3):219–28. Available from: <http://www.sbmp.org.br/cbab/siscbab/modules/news/article.php?storyid=625>
35. Yan W, Holland JB. A heritability-adjusted GGE biplot for test environment evaluation. *Euphytica.* 2010;171(3):355–69. Available from: <http://link.springer.com/10.1007/s10681-009-0030-5>
36. Cornelius PL, Crossa J, Seyedsadr MS. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. *Genotype by Environment Interaction.* 1996.
37. Cornelius PL, Crossa J. Prediction Assessment of Shrinkage Estimators of Multiplicative Models for Multi-Environment Cultivar Trials. *Crop Sci.*

- 1999;39(4):998–1009. Available from:
<https://onlinelibrary.wiley.com/doi/abs/10.2135/cropsci1999.0011183X003900040007x>
38. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis, Third Edition (Texts in Statistical Science). Book. 2013.
39. Zeng Q, Wen H, Wong SC, Huang H, Guo Q, Pei X. Spatial joint analysis for zonal daytime and nighttime crash frequencies using a Bayesian bivariate conditional autoregressive model. *J Transp Saf Secur.* 2020;12(4):566–85.
Available from: <https://doi.org/10.1080/19439962.2018.1516259>
40. Bernardo JM, Smith AFM. Bayesian theory. Vol. 405. John Wiley & Sons; 2009.
41. Zeng Q, Guo Q, Wong SC, Wen H, Huang H, Pei X. Jointly modeling area-level crash rates by severity: a Bayesian multivariate random-parameters spatio-temporal Tobit regression. *Transp A Transp Sci.* 2019;15(2):1867–84.
42. Gelman A. Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis.* 2006.
43. Jaynes ET. Information Theory and Statistical Mechanics. *Phys Rev.* 1957;106(4):620–30. Available from:
<https://link.aps.org/doi/10.1103/PhysRev.106.620>
44. Silva CP da. Uso do algoritmo Reversible Jump Markov Chain Monte Carlo para seleção de termos multiplicativos do modelo AMMI. 2019; Available from:
<http://repositorio.ufla.br/jspui/handle/1/33974>
45. Box GEP, Tiao GC. Bayesian inference in statistical analysis. Vol. 40. John Wiley & Sons; 2011.
46. Raftery AE, Lewis S. How Many Iterations in the Gibbs Sampler? *Bayesian Statistics 4* (Bernardo, JM, Berger, JO, Dawid, AP and Smith, AFM, eds). Oxford: Oxford University Press; 1992. p. 763–73.
47. Heidelberger P, Welch PD. Simulation Run Length Control in the Presence of an

- Initial Transient. *Oper Res.* 1983;31(6):1109–44. Available from:
<http://pubsonline.informs.org/doi/abs/10.1287/opre.31.6.1109>
48. Chen M-H, Shao Q-M. Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *J Comput Graph Stat.* 1999;8(1):69–92. Available from:
<http://www.tandfonline.com/doi/abs/10.1080/10618600.1999.10474802>
49. Hu Z, Yang R-C. Improved Statistical Inference for Graphical Description and Interpretation of Genotype × Environment Interaction. *Crop Sci.* 2013;53(6):2400–10. Available from:
<http://doi.wiley.com/10.2135/cropsci2013.04.0218>
50. Nuvunga JJ, da Silva CP, de Oliveira LA, de Lima RR, Balestre M. Bayesian factor analytic model: An approach in multiple environment trials. Shi L, editor. *PLoS One.* 2019;14(8):e0220290. Available from:
<https://dx.plos.org/10.1371/journal.pone.0220290>
51. Schwarz G. Estimating the Dimension of a Model. *Ann Stat.* 1978;6(2):461–4.
Available from: <http://projecteuclid.org/euclid-aos/1176344136>
52. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 1974;19(6):716–23. Available from:
<http://ieeexplore.ieee.org/document/1100705/>
53. Raftery AE, Newton MA, Satagopan JM, Krivitsky PN. Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity. *Bayesian Stat.* 2007;
54. R Core Development Team. R: A language and environment for statistical computing. Vienna, Austria. 2019.
55. Cowles MK, Carlin BP. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *J Am Stat Assoc.* 1996;91(434):883–904. Available from:
<http://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476956>
56. Yan W. Crop Variety Trials: Data Management and Analysis [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2014. Available from:

- <http://doi.wiley.com/10.1002/9781118688571>
57. Smith A, Cullis B, Thompson R. Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. *Biometrics*. 2001;57(4):1138–47. Available from: <http://doi.wiley.com/10.1111/j.0006-341X.2001.01138.x>
58. Resende MDV de, Duarte JB. Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesqui Agropecuária Trop (Agricultural Res Trop)*. 2007;37(3):182–94.
59. Denis J-B, Gower JC. Asymptotic covariances for the parameters of biadditive models. *Util Math*. 1994;46:193–205.
60. Denis J-B, Gower JC. Asymptotic Confidence Regions for Biadditive Models: Interpreting Genotype- Environment Interactions. *Appl Stat*. 1996;45(4):479. Available from: <https://www.jstor.org/stable/10.2307/2986069?origin=crossref>
61. Yang R-C, Crossa J, Cornelius PL, Burgueño J. Biplot Analysis of Genotype × Environment Interaction: Proceed with Caution. *Crop Sci*. 2009;49(5):1564–76. Available from: <http://doi.wiley.com/10.2135/cropsci2008.11.0665>
62. Yan W, Glover KD, Kang MS. Comment on “Biplot Analysis of Genotype × Environment Interaction: Proceed with Caution,” by R.-C. Yang, J. Crossa, P.L. Cornelius, and J. Burgueño in Crop Science 2009 49:1564-1576. *Crop Sci*. 2010;50(4):1121–3. Available from: <http://doi.wiley.com/10.2135/cropsci2010.01.0001e>
63. Jaynes ET. Probability Theory : The Logic Of Science By E . T . Jaynes. College University. 2003.
64. ter Braak CJF, Boer MP, Bink MCAM. Extending Xu’s Bayesian Model for Estimating Polygenic Effects Using Markers of the Entire Genome. *Genetics*. 2005;170(3):1435–8. Available from: <http://www.genetics.org/lookup/doi/10.1534/genetics.105.040469>

Supporting information

S1 Appendix. Maximum entropy prior

S2 Appendix. Posterior distribution for variance components

S1 Table. Individual COR and PRESS

S1 Fig. Graph of the trace for variance components

S2 Fig. Credibility region for the information rate (IR)

S3 Fig. Default “who won where” to the target environment

S4 Fig. Default “who won where” for mega-environments second BGGE

S5 Fig. Default “who won where” for mega-environments second BGGEE

S1 Appendix

A priori derivation of the maximum entropy for $\sigma_{\lambda_k}^2$

Shannon entropy [1,2] for a continuous variable y is given by

$$H(y) = - \int_S p(y) \log(p(y)) dy \quad (1)$$

In our framework, to derive maximum entropy density for $\sigma_{\lambda_k}^2 \in (0, \infty)$, the only constraint assumed is a de que $E[\sigma_{\lambda_k}^2] = \vartheta$ ($0 < \vartheta < \infty$), and of course, density normalization for $p(\sigma_{\lambda_k}^2)$. To make for easier algebraic identities we assumed $\tau_{\lambda_k} = 1/\sigma_{\lambda_k}^2$, that implies the following restrictions:

$$\int_0^\infty \tau_{\lambda_k} p(\tau_{\lambda_k}) d\tau_{\lambda_k} = \vartheta \Rightarrow \zeta \left[\int_0^\infty \tau_{\lambda_k} p(\tau_{\lambda_k}) d\tau_{\lambda_k} - \vartheta \right] \quad \text{and} \quad (2)$$

$$\int_0^\infty p(\tau_{\lambda_k}) d\tau_{\lambda_k} = 1 \Rightarrow \zeta_0 \left[\int_0^\infty p(\tau_{\lambda_k}) d\tau_{\lambda_k} - 1 \right] \quad (3)$$

Using Lagrange multipliers for the restrictions and maximizing over parameter space (taking first derivatives equal to zero) yields maximum entropy density distribution for $p(\tau_{\lambda_k})$ [2]:

$$p(\tau_{\lambda_k}) = \frac{e^{-\zeta \tau_{\lambda_k}}}{\int_0^\infty e^{-\zeta \tau_{\lambda_k}} d\tau_{\lambda_k}}. \quad (4)$$

Note that

$$\int_0^\infty e^{-\zeta \tau_{\lambda_k}} d\tau_{\lambda_k} = \int_0^\infty (\zeta/\zeta) e^{-\zeta \tau_{\lambda_k}} d\tau_{\lambda_k} = (1/\zeta) \int_0^\infty \zeta e^{-\zeta \tau_{\lambda_k}} d\tau_{\lambda_k} = 1/\zeta; \zeta > 0.$$

Then,

$$p(\tau_{\lambda_k}) = e^{-\zeta \tau_{\lambda_k}} / (1/\zeta) = \zeta e^{-\zeta \tau_{\lambda_k}}.$$

The density of $\sigma_{\lambda_k}^2$, using linear transformation techniques [3] is given by:

$$p_{\sigma_{\lambda_k}^2}(\sigma_{\lambda_k}^2) = \left| \frac{\partial}{\partial \sigma_{\lambda_k}^2} g^{-1}(\sigma_{\lambda_k}^2) \right| p_{\tau_{\lambda_k}}(g^{-1}(\sigma_{\lambda_k}^2)) I_{D(\sigma_{\lambda_k}^2)}, \quad (5)$$

observing that $\sigma_{\lambda_k}^2 = g(\tau_{\lambda_k}) = 1/\tau_{\lambda_k} \Rightarrow (\sigma_{\lambda_k}^2)^{-1} = \tau_{\lambda_k}$. Solving the derivative within the modulus of equation (5) results in

$$\frac{\partial}{\partial \sigma_{\lambda_k}^2} g^{-1}(\sigma_{\lambda_k}^2) = \frac{\partial}{\partial \sigma_{\lambda_k}^2} \left(\frac{1}{\sigma_{\lambda_k}^2} \right) = -\frac{1}{(\sigma_{\lambda_k}^2)^2}. \quad (6)$$

Using (6) in (5) yields the maximum entropy density for $\sigma_{\lambda_k}^2$:

$$p(\sigma_{\lambda_k}^2) = \frac{\varsigma}{\Gamma(1)} (\sigma_{\lambda_k}^2)^{-(1+1)} e^{-\varsigma/\sigma_{\lambda_k}^2}$$

where $\Gamma(\cdot)$ is the gamma function. Therefore, $\sigma_{\lambda_k}^2 \sim Inv-Gamma(a, b)$, with $a = 1$ and $b = \varsigma$.

References

- 1 Templeman AB, Xingsi LA. A maximum entropy approach to constrained non-linear programming. *Engineering Optimization*. 1987; 12(3): 191–205.
- 2 Cover TM, Thomas JA. Elements of information theory. 2^a ed. New Jersey: John Wiley & Sons; 1999.
- 3 Mood AM, Graybill FA, Boes, DC. Introduction to the theory of statistics. 3^a ed. Tokyo: McGraw-Hill; 1974.

S2 Appendix

Posterior distribution for variance components of $(\sigma_{\lambda_k}^2)$ singular values

Initially, the likelihood function is expressed as

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

The joint posterior distribution is then obtained by applying Bayes' theorem connecting *a priori* information with the likelihood function:

$$\begin{aligned} p(\Phi | \mathbf{y}) &= p(\mathbf{y} | \boldsymbol{\theta}, \sigma_e^2) \propto p(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \sigma_\beta^2) p(\sigma_e^2 | v, u) \times \\ &\quad \times \prod_{k=1}^t p(\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2) p(\sigma_{\lambda_k}^2) p(\boldsymbol{\alpha}_k) p(\boldsymbol{\gamma}_k) \end{aligned} \quad (2)$$

where $\Phi = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \sigma_e^2, \boldsymbol{\sigma}_\lambda)$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_t)$, $\boldsymbol{\sigma}_\lambda = (\sigma_{\lambda_1}^2, \dots, \sigma_{\lambda_t}^2)$ and

$t = \min(g-1, e)$.

Using prior distribution from maximum entropy reasoning $(\sigma_{\lambda_k}^2 \sim GI(a, b))$, full conditional posterior density for $\sigma_{\lambda_k}^2$ is:

$$p(\sigma_{\lambda_k}^2 | \dots) \propto (\sigma_{\lambda_k}^2)^{-1} \exp \left\{ -\frac{1}{2\sigma_{\lambda_k}^2} \lambda_k^2 \right\} \times (\sigma_{\lambda_k}^2)^{-a-1} \exp \left\{ -\frac{b}{\sigma_{\lambda_k}^2} \right\}, \quad (3)$$

$$\text{or } p(\sigma_{\lambda_k}^2 | \dots) \propto (\sigma_{\lambda_k}^2)^{-a-1-1} \exp \left\{ -\frac{1}{2\sigma_{\lambda_k}^2} (\lambda_k^2 + 2b) \right\} \text{ and} \quad (4)$$

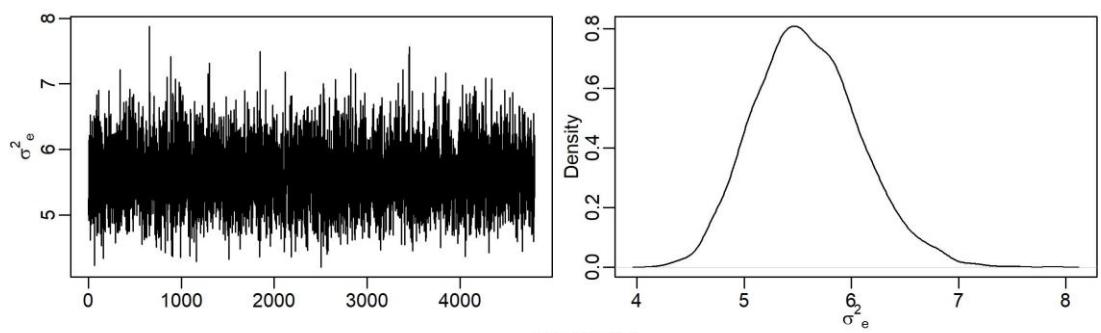
$$p(\sigma_{\lambda_k}^2 | \dots) \propto (\sigma_{\lambda_k}^2)^{-(a+1)-1} \exp \left\{ -\frac{1}{2\sigma_{\lambda_k}^2} (\lambda_k^2 + 2b) \right\}. \quad (5)$$

That equation (5) represents an inverse gamma density with scale parameters $(\lambda_k^2 + 2b)$ and with $(a+1)$ degrees of freedom, or $\sigma_{\lambda_k}^2 | \dots \sim GI(a+1, \lambda_k^2 + 2b)$.

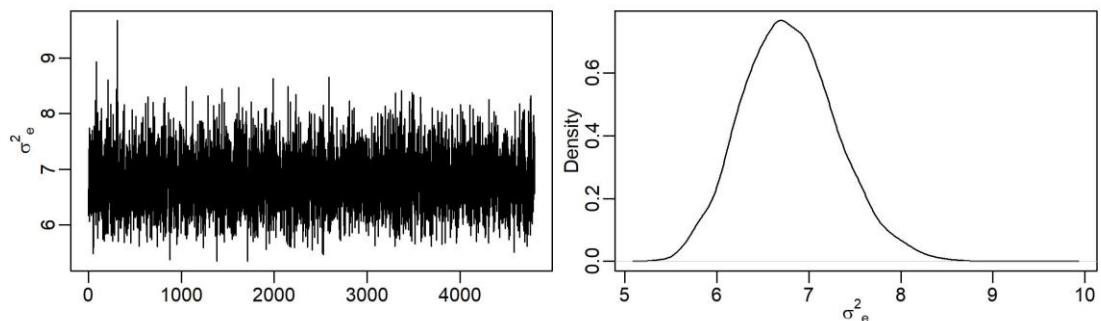
S1Table. Individual COR and PRESS values calculated for the three random unbalanced scenarios (RUS) with simulated genotypes.

Level (%)	IMB.	COR	PRESS	Level (%)	IMB.	COR	PRESS
10	1	0.91	6.37	10	1	0.86	7.27
10	2	0.89	5.20	10	2	0.90	5.12
10	3	0.72	7.00	10	3	0.72	7.08
10	4	0.72	6.03	10	4	0.72	5.97
10	5	0.87	9.00	10	5	0.90	8.24
10	6	0.89	8.58	10	6	0.90	7.65
10	7	0.61	20.82	10	7	0.65	19.66
10	8	0.80	9.51	10	8	0.80	8.87
10	9	0.90	5.37	10	9	0.91	5.09
10	10	0.49	13.37	10	10	0.49	13.22
33	1	0.44	15.64	33	1	0.53	14.42
34	2	0.68	13.30	34	2	0.73	12.37
35	3	0.61	14.43	35	3	0.66	13.24
50	1	0.67	15.12	50	1	0.68	11.09
50	2	0.24	47.07	50	2	0.74	10.29

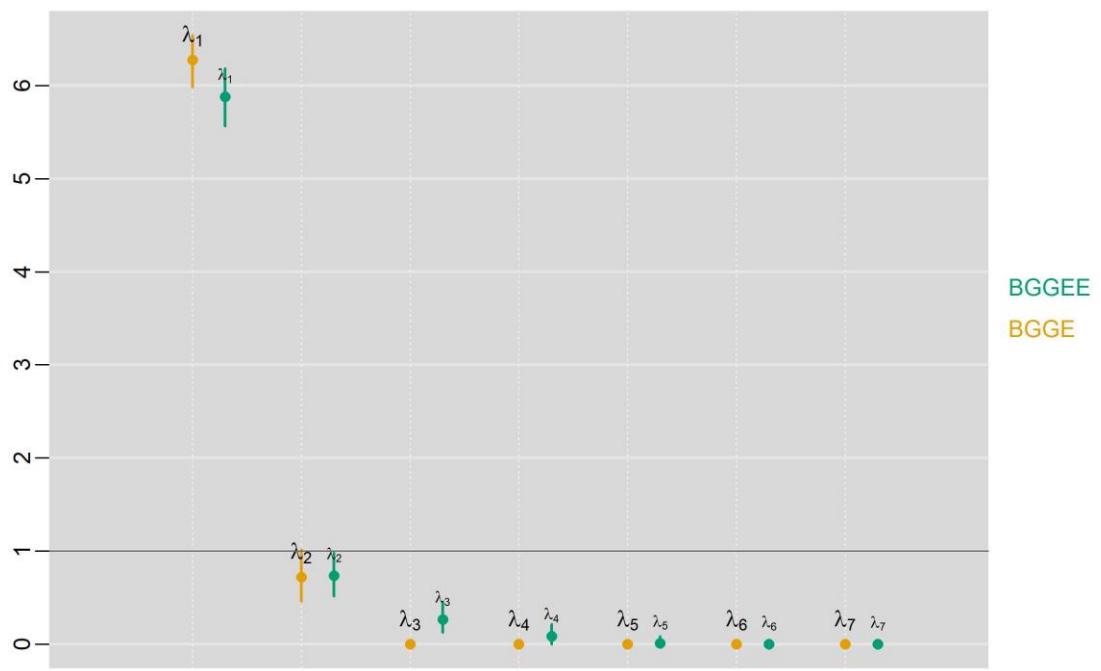
BGGE



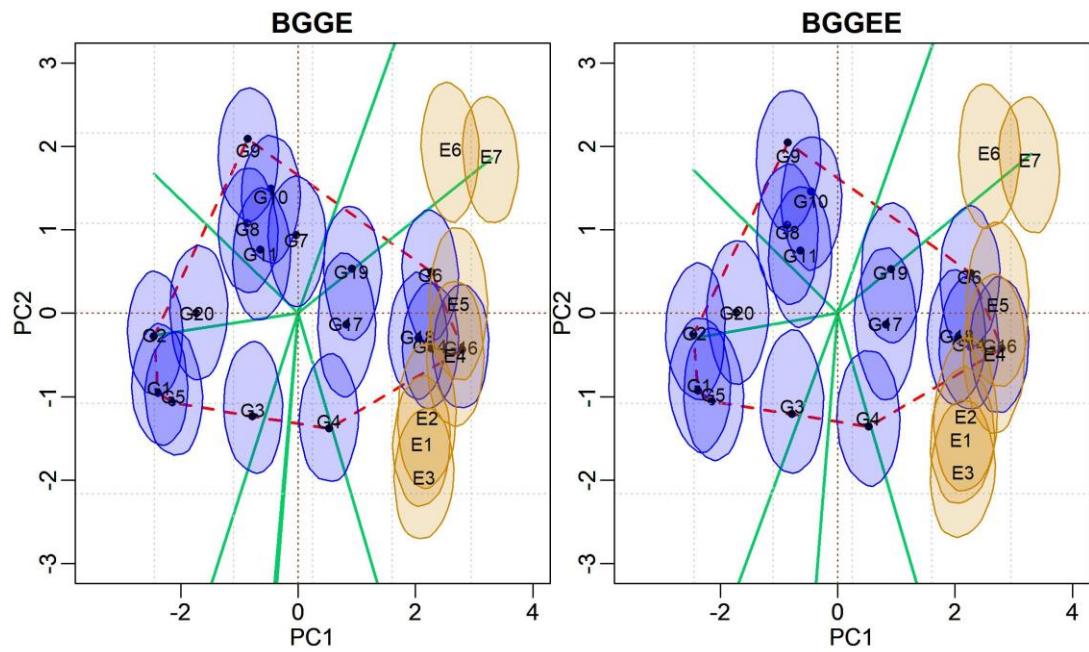
BGGEE



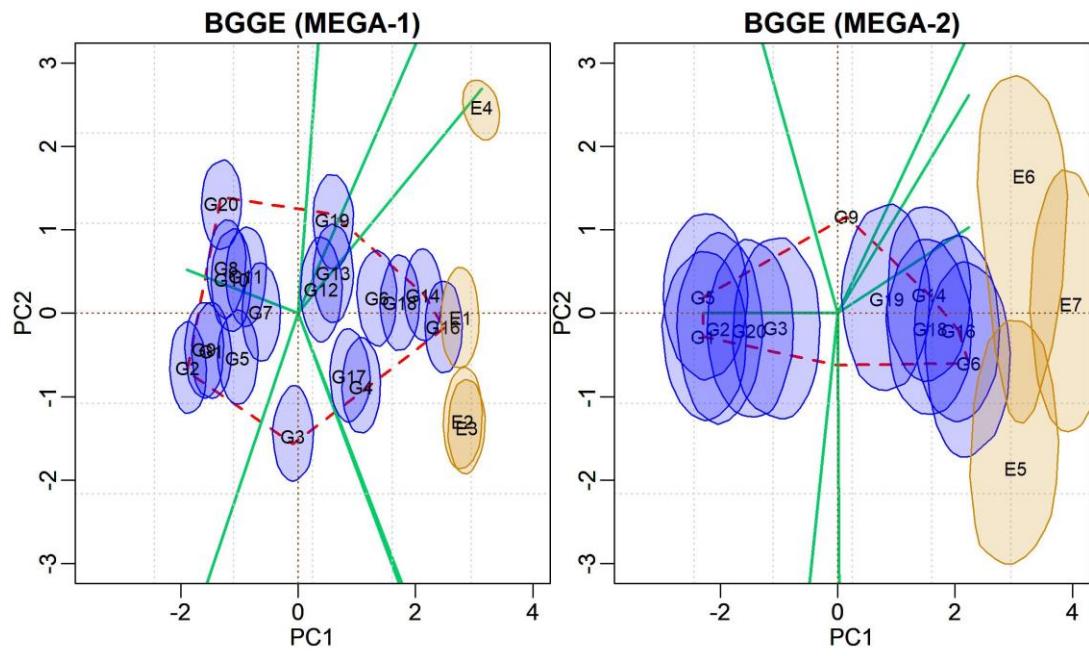
S1 Fig. Response trace and densities graph for the residual variance in the BGGE and BGGEE models.



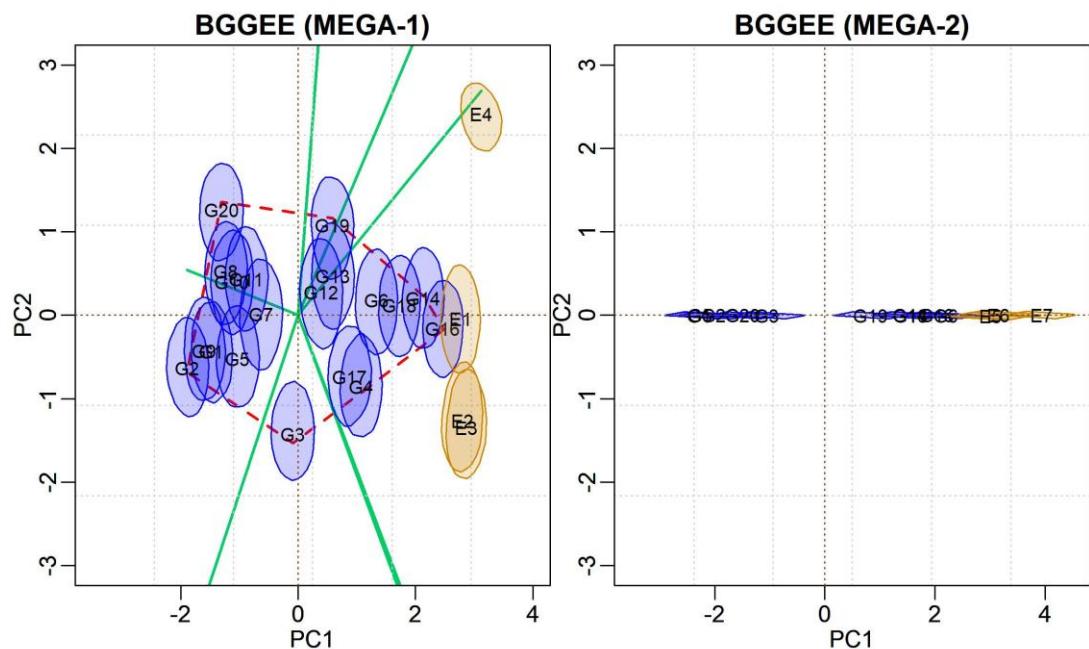
S2 Fig. The posterior means and HPD regions at 95% credibility for the information rate (IR), referring to the BGGE and BGGEE models' main components.



S3 Fig. The GGE biplot and bivariate regions at 95% credibility for the genotypic and environmental scores, showing the “who won where” graphic formulation for the target environment according to the BGGE and BGGEE models.



S4 Fig. The GGE biplot and bivariate regions at 95% credibility for the genotypic and environmental scores, showing the “who won where” graphic formulation for the two mega-environments defined according to the BGGE model.



S5 Fig. The GGE biplot and bivariate regions at 95% credibility for the genotypic and environmental scores, showing the “who won where” graphic formulation for the two mega-environments defined according to the BGGEE model.

ARTIGO 2**Modelagem bayesiana GGE para ensaios multiambientais com variâncias heterogêneas**

O artigo será traduzido e submetido à uma revista na área de estatística com enfase em análise de dados multiambientais (versão preliminar)

Modelagem bayesiana GGE para ensaios multiambientais com variâncias heterogêneas

Resumo: Modelos multiplicativos (ou lineares-bilineares) são úteis para analisar respostas de cultivares em ambientes nas fases finais de programas de melhoramento de plantas, oferecendo varias possibilidades para estudar a interação genótipos × ambientes (GEI) que, geralmente, dificulta a seleção de genótipos superiores. Dentre esses modelos, o de efeitos principais de genótipos (G) mais GEI (GGE), merece especial destaque devido à ampla aplicabilidade por pesquisadores e melhoristas na análise de dados provenientes de ensaios multiambientais. Contudo, a maioria das aplicações do GGE admite que a variância residual seja a mesma para todos os locais de teste — o que pode não ser realista. O principal objetivo desse trabalho foi conduzir uma análise GGE considerando heterogeneidade de variâncias residuais nos locais utilizando a inferência bayesiana e avaliar suas implicações no inter-relacionamento dos genótipos e dos ambientes. O método proposto foi exemplificado em cenários controlados e também com dados reais. Os resultados mostraram a grande flexibilidade do método bayesiano para modelar a heterogeneidade residual e incorporar inferência ao biplot por meio de regiões de credibilidade aos escores que explicam o efeito G+GEI. Constatou-se que negligenciar a heterocedasticidade nos dados pode comprometer os resultados da análise de estabilidade e adaptabilidade, propiciando interpretações equivocadas. Implicações dessa suposição também foram observadas na avaliação ambiental, principalmente para os dados reais em que o padrão se mostrou mais complexo. Diferentes níveis de correlações entre os ambientes foram observados e aspectos importantes foram afetados, como a representatividade do ambiente alvo e capacidade em discriminar genótipos que são fundamentais na seleção e descarte de genótipos e ambientes de teste. Para dados reais, modelar a heterogeneidade teve maior influência na subdivisão do ambiente alvo em subgrupos com efeitos semelhantes para GEI. Nesse sentido, a análise considerando diferentes precisões experimentais se mostrou robusta, captando as estruturas de variâncias subjacentes nos dados sendo, portanto, mais adequada do que a sua versão que assume variância única para todos os locais.

Palavras chave: Modelos multiplicativos, Análise GGE, inferência bayesiana, heterogeneidade de variâncias.

Introdução

Ensaio multiambientes (MET) são conduzidos no decorrer de vários anos e usados para estimar o rendimento de cultivares em um conjunto de locais (ou ambientes) de interesse, visando a seleção e recomendação de genótipos superiores. A resposta diferencial das cultivares entre ambientes, referida como interação genótipo × ambiente (GEI), dificulta o trabalho do melhorista de plantas. A GEI pode ser do tipo cruzada (COI), que leva a não correlação entre as respostas dos genótipos de um ambiente a outro, ocasionando mudanças no ranqueamento e, por isso, também é referida como GEI do tipo qualitativa. Por outro lado, a GEI é dita não cruzada, ou quantitativa, quando ocorrem alterações de escalas nos caracteres, caracterizadas pela

heterogeneidade de variância genética dentro dos diferentes ambientes (BAKER, 1988; BURGUEÑO et al., 2008; CROSSA, 2012; VARGAS et al., 2015; HERNÁNDEZ et al., 2019). Na presença de COI, avaliações e decisões não podem ser estabelecidas de forma generalizada. Nesse sentido, diferentes metodologias foram desenvolvidas para analisar dados METs visando identificar genótipos de ampla adaptação ou aqueles que possuam adaptação específica, no sentido de aproveitar os efeitos positivos da GEI para recomendações regionalizadas.

Modelos multiplicativos (ou lineares-bilineares) são úteis para analisar dados organizados em uma tabela de dupla entrada onde dois fatores, com interação, estejam sendo avaliados e adquiriram grande popularidade no estudo da GEI (GAUCH, 1988; CORNELIUS; SEYEDSADR, 1997). Dentre esses modelos, o de efeitos principais de genótipos (G) mais GEI, referido na literatura como GGE, e o de efeitos principais aditivos e interação multiplicativa (AMMI) merecem especial destaque devido à ampla aplicabilidade por pesquisadores e melhoristas na análise de dados MET (CORNELIUS; CROSSA; SEYEDSADR, 1996; GAUCH, 1988; GAUCH; ZOBEL, 1997; YAN; TINKER, 2006; GAUCH, 2006; YAN et al., 2007; GAUCH; PIEPHO; ANNICHIAIRO, 2008; BALESTRE et al., 2009; YAN, 2014).

O modelo GGE, bem como outros modelos lineares-bilineares, combinam em uma única abordagem o estudo da adaptabilidade e estabilidade oferecendo mais possibilidades para analisar a GEI do que métodos baseados em regressão simples ou múltipla, de genótipos em ambientes, como, por exemplo, aqueles sugeridos por Finlay e Wilkinson (1963) e Eberhart e Russell (1966). Além disso, métodos de seleção de termos bilineares possibilitam a obtenção de modelos mais parcimoniosos, já que a matriz da GEI (AMMI) ou de G+GEI (GGE) pode ser descrita pela decomposição por valores singulares (DVS) e aproximada por outra de dimensão inferior, mas ainda descrevendo o padrão da matriz original (KRZANOWSKI, 1987; GAUCH, 1988; DIAS; KRZANOWSKI, 2003; FORKMAN; PIEPHO, 2014).

Ao permitir a simplificação de dados em poucas dimensões, padrões de respostas de genótipos em ambientes podem ser visualizados graficamente utilizando biplots (GABRIEL, 1971; BRADU; GABRIEL, 1978; KEMPTON, 1984; GAUCH, 1992; COOPER; DELACY, 1994; YAN et al., 2001). Essa representação gráfica de escores genotípicos e ambientais, em relação aos primeiros eixos principais, tornou-se um procedimento padrão na aplicação de modelos multiplicativos e sua interpretação tem sido utilizada para subsidiar a tomada de importantes decisões em relação aos dados

MET analisados. O modelo GGE biplot, uma variante do GGE, quando se utiliza apenas os dois primeiros eixos singulares, tornou-se um popular método para avaliar respostas de cultivares entre ambientes (YAN et al., 2000; YAN; TINKER, 2006; GUILLY et al., 2017).

Apesar das reconhecidas vantagens, os modelos lineares-bilineares convencionais (admitindo efeitos fixos para os parâmetros), como o GGE, possuem limitações, dentre as quais se destaca a dificuldade em lidar com conjuntos de dados desbalanceados e com heterogeneidade de variâncias. O método padrão de estimação dos modelos fixos AMMI e GGE não leva em consideração a heterocedasticidade nos dados. Embora existam métodos baseados na DVS para esse fim (YAN; KANG, 2003; YAN, 2014; RODRIGES et al., 2015). Tais procedimentos, entretanto, são suscetíveis a críticas, pois são ajustados em dois estágios e permitem perda de informação (ROMÃO et al., 2019; GOGEL; SMITH; CULLIS, 2018; HERNANDEZ et al., 2019).

A suposição de variância única em ambientes é restritiva e tem sido colocada em cheque em diversos trabalhos (SINGH; CECCARELLI; GRANDO, 1999; CROSSA et al., 2006; HU; SPILKE, 2011; HU; YAN; SHEN, 2013; SMITH et al., 2015). Por outro lado, modelar a heterogeneidade de variâncias nos dados permite obter melhores estimativas para efeitos genotípicos, uma vez que as estimativas obtidas para as componentes de variância são mais precisas (SMITH; CULLIS; THOMPSON, 2001; HU; YAN; SHEN, 2013).

Modelos mistos lineares oferecem maior flexibilidade quando comparados à modelagem de efeitos fixos e essa classe de modelos tem sido bem desenvolvida (PIEPHO, 1997; SMITH; CULLIS; THOMPSON, 2005; CROSSA et al., 2006; OKEY et al., 2006; BURGUEÑO et al., 2008; PIEPHO et al., 2008; STEFANOVA; BUIRCHELL, 2010; SMITH; CULLIS, 2018; ALVARADO et al., 2020). Modelos mistos são versáteis para lidar com a heterogeneidade de variâncias, dados incompletos e com a correlação espacial dentro e entre ambientes (CROSSA, 2012).

Versões em modelos mistos lineares-bilineares para o AMMI e GGE foram propostas e estão disponíveis em softwares amigáveis, o que as tornaram, até certo ponto, populares (PIEPHO, 1997; 1998; SMITH; CULLIS; THOMPSON, 2001, PIEPHO; MOHRING, 2006). Tais abordagens são decorrentes de se assumir uma estrutura fatorial analítica para matriz de variância-covariância genética e, por isso, são referidos como Modelos Fatoriais Analíticos (ou simplesmente FA). Mas, esses modelos exigem alta demanda computacional e, eventualmente, apresentam problemas

de não identificabilidade na estimativa de parâmetros (NUVUNGA et al., 2015). Não está claro também como regiões de confiança paramétricas, tal como as construídas em biplots, sob o modelo de efeitos fixos, podem ser estendidas para um modelo de efeitos mistos, como destacado por Crossa et al. (2011). Alguns desses problemas foram contornados por Nuvunga et al. (2019) que aplicaram o FA utilizando a metodologia bayesiana.

A integração da inferência bayesiana ao melhoramento de plantas é um evento mais recente. A modelagem bayesiana oferece vantagens práticas e ampla flexibilidade para modelar o efeito da GEI, permitindo utilizar estruturas de variância-covariância complexas para explorar ou acomodar melhor as interações (COTES et al., 2006; EDWARDS; JANNINK, 2006; ORELLANA et al., 2014). Os primeiros trabalhos utilizando a metodologia bayesiana aplicada ao modelo AMMI foram propostos por Viele e Srinivasan (2000) e Liu (2001) que mostraram como o processo amostral pode ser conduzido, corretamente, utilizando métodos de monte Carlo via cadeia de Markov (MCMC), principalmente para os parâmetros que descrevem a GEI, cujo suporte da distribuição *a posteriori* não é trivial. Crossa et al. (2011) e Perez-Elizalde, Jarquin e Crossa (2012), por sua vez, demonstraram empiricamente a flexibilidade do modelo bayesiano para incorporar incerteza ao biplot AMMI2, por meio de regiões de credibilidade bivariadas construídas para os escores genotípicos e ambientais, bem como a utilizar informações de experimentos anteriores incorporados por meio de prioris.

Essa proposta representou um considerável impacto visto que a análise biplot padrão tem um caráter mais descritivo e mesmo os métodos de inferência frequentistas desenvolvidos e aplicados, a esta representação gráfica, têm sido criticados na literatura devido à adoção de suposições restritivas quanto à distribuição de probabilidade dos escores individuais ou ainda pelos processos problemáticos de reamostragem nas linhas e colunas da matriz de interação (DENIS; GOWER; 1994; 1996; MAIA et al., 2006; LAVORANTI; DIAS; KRAZNOWSKI, 2007; YANG et al., 2009; YAN; GLOVER; KANGYAN, 2010). Outras contribuições ao método foram publicadas recentemente (JOSSE et al., 2014; OLIVEIRA et al., 2015; SILVA et al., 2015; JARQUIN et al., 2016; OLIVEIRA et al., 2016; ROMÃO et al., 2019; SILVA et al., 2019).

A maioria dos estudos realizados envolvendo a aplicação da inferência bayesiana a modelos multiplicativos tem se restringido ao modelo AMMI e pesquisas e aplicações com a versão bayesiana do GGE ainda são escassas na literatura. Jarquin et al. (2016),

Oliveira et al. (2016) e Omer e Singh (2017) são exceções e abordaram aspectos relacionados ao modelo GGE-bayesiano. Contudo, prioris específicas não foram assumidas explicitamente para componentes de variâncias residuais nos locais. Implicações e importância de tais pressuposições na modelagem de efeitos principais de genótipos e efeito da GEI, bem como detalhes de suas implementações, na análise AMMI bayesiana, foram consideradas por Silva et al. (2019).

O modelo GGE possui um forte apelo gráfico e apresenta uma explicação biológica plausível para a utilização dos dois primeiros eixos principais na análise biplot (YAN et al., 2000). Ressalta-se assim, que a modelagem da heterogeneidade de variâncias traria importantes contribuições a essa análise. Com isso, os valores singulares poderiam ser estimados de forma mais precisa e os genótipos seriam avaliados com precisões diferentes (em cada ambiente), o que poderia aumentar a capacidade de recuperar informações sobre diferenças experimentais entre os ensaios no biplot.

Sendo assim, os principais objetivos desse artigo foram: i) considerar a suposição de heterogeneidade da variância residual entre locais na análise GGE utilizando a inferência bayesiana e ii) verificar as implicações dessa suposição no inter-relacionamento dos genótipos e dos ambientes em representações biplot. O método foi exemplificado a partir de dados simulados (homocedástico e heterocedástico), a fim de avaliar o modelo em um cenário controlado, e com dados de MET reais para avaliar seu comportamento em um cenário não-controlado.

3 Material e Métodos

Dados simulados

Os efeitos ambientais foram inicialmente simulados proporcionais à densidade normal para quantis derivados de desvios ambientais igualmente espaçados entre -4,5 a 4,5. Os efeitos são produzidos como $5 \times p_j - 1$, em que p_j é a densidade normal para o j -ésimo quantil. Para o efeito da GEI foram simulados três padrões distintos: i) genótipos estáveis (inferiores {G1 a G4}, medianos {G5 a G12} e superiores {G13 a G16}); ii) genótipos instáveis com reforço ambiental positivo (G17 e G18) e iii) genótipos instáveis com reforço ambiental negativo (G19 e G20). A matriz com os valores assim simulados foi corrigida para média de linhas e colunas. Os efeitos

principais de genótipos foram obtidos por meio de uma distribuição gaussiana $N(0, 3)$ e ordenados, em ordem crescente, de G1 a G20.

Os padrões médios de resposta de cada um dos subgrupos (em relação ao padrão GEI simulado) destacados acima, em cada ambiente, somados a uma constante comum ($\mu = 10$), podem ser observados na Figura 1.

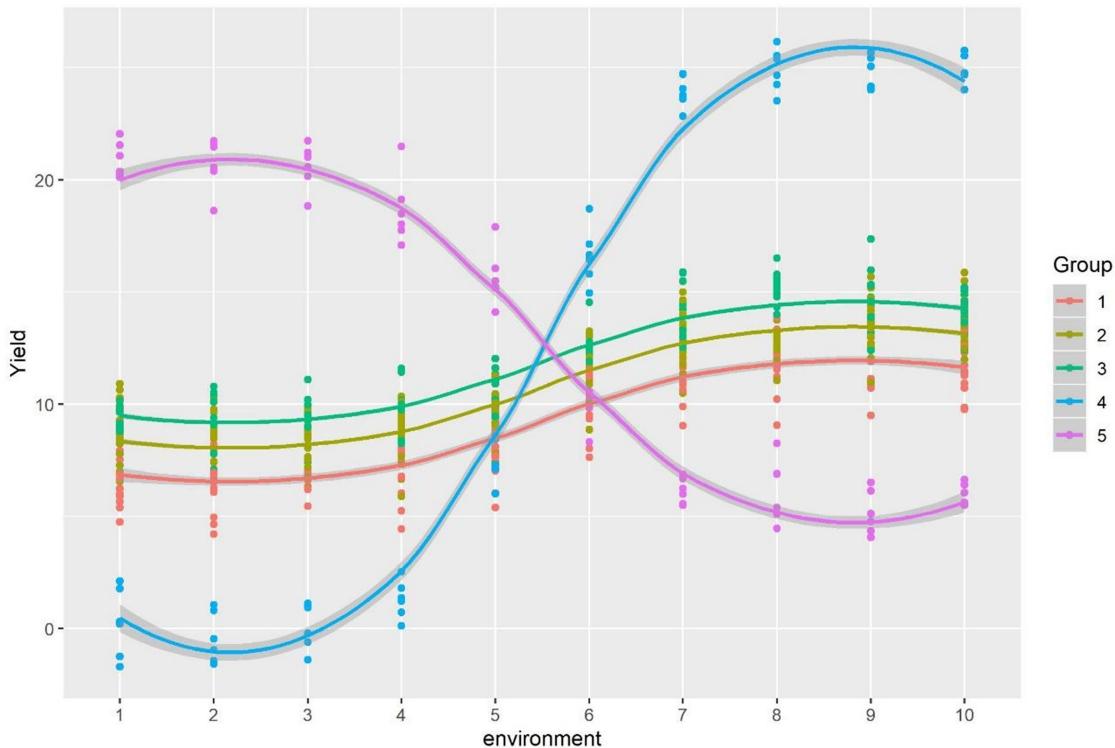


Figure 1. Rendimentos médios de subgrupos de genótipos, relativo ao padrão GEI simulado, em cada ambiente.

Para emular a variância homogênea considerou-se $\sigma_e^2 = 1$. Para emular a variância heterogênea, a diagonal da matriz de covariância foi a seguinte: $diag \{V\} = [10,5; 0,5; 3,0; 2,0; 3,0; 5,0; 9,0; 9,0; 15,0; 17,0]$. Todos os dados foram simulados a partir de distribuições normais multivariadas.

Dados Reais

O conjunto de dados reais é proveniente de uma rede de ensaios de valor de cultivo e uso coordenado pela Embrapa Milho e Sorgo com 36 genótipos de sorgo, dos quais 33 são híbridos experimentais desenvolvidos pelo programa de melhoramento de sorgo biomassa da Embrapa Milho e Sorgo e os 3 restantes são cultivares testemunhas ou controles. Os genótipos foram avaliados na safra agrícola 2014/2015 em 10 locais

distintos no Brasil. Características geográficas e aspectos de cultivo destes locais são apresentadas na Tabela 1.

Tabela 1. Características geográficas dos locais em que os ensaios de sorgo foram realizados na safra agrícola 2014/2015.

Ambientes	LAT	LONG	ALT	DP	DC
E1 - Dourados-MS	22°13'S	54°48'O	430m	11/11/14	13/05/15
E2 - Dracena-SP	21°28'S	51°31'O	421m	15/11/14	28/04/15
E3 - Goiânia-GO	16°40'S	49°15'O	823m	18/12/14	23/05/15
E4 - Guaíra-SP	24°04'S	54°15'O	220m	26/11/14	14/04/15
E5 - Lavras-MG	21°14'S	45°00'O	919m	22/11/14	12/05/15
E6-Nova Porterinha-MG	15°48'S	43°18'O	85m	22/11/14	20/05/15
E7 - Pelotas-RS	31° 46'S	52°20'O	7m	06/12/14	25/03/15
E8 - Sete Lagoas-MG	19°27'S	44°14'O	761m	06/11/14	12/05/15
E9 - Sinop-MT	11° 50'S	55°38'O	384m	04/12/14	11/05/15
E10 - Uberlândia-MG	18°55'S	48°16'O	863m	06/12/14	24/05/15

LAT, latitude; LONG, longitude; ALT, altitude; DP, data de plantio; DC, Data de colheita.

Os experimentos foram dispostos em um delineamento em látice triplo 6×6 . As parcelas foram constituídas por quatro linhas de 5,0 m de comprimento, espaçadas de 0,7 m uma da outra, considerando apenas as duas linhas centrais como úteis. A variável analisada foi a produção de biomassa seca total da parcela (DMY) convertida em toneladas por hectare (ton/ha). Os tratos culturais são descritos detalhadamente em Delgado et al. (2019).

Modelo GGE

O modelo GGE, em notação vetorial, é dado por:

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

em que:

$\mathbf{y}_{n \times 1}$ é o vetor de observações fenotípicas com $n = r \times l$, sendo r o número genótipos repetidos $l = c \times b$, sendo c o número locais e b o número de blocos;

$\boldsymbol{\beta}_{l \times 1}$ é o vetor de efeitos de blocos hierarquizados dentro de ambientes;

λ_k é o k -ésimo valor singular, com $k = 1, \dots, t$ sendo $t \{t = \min(r-1, c)\}$ o posto da matriz de dupla entrada contendo os erros de não aditividade aos efeitos principais de ambientes;

α_k e γ_k são os k -ésimos vetores singulares, genotípico e ambiental, respectivamente, relacionados ao k -ésimo componente principal.

Além disso, \mathbf{X}_1 , \mathbf{X}_2 e \mathbf{Z} são matrizes de delineamento, associadas a β , α_k e γ_k , respectivamente, e $\epsilon_{n \times 1}$ é o vetor contendo os erros aleatórios em que $\epsilon \sim N(\mathbf{0}, \mathbf{V})$, sendo $\mathbf{0}$ o vetor nulo e \mathbf{V} uma matriz bloco diagonal de dimensão $n \times n$ composta por $\sigma_{e_j}^2$, com $j = 1, \dots, c$ em que n_j é número de observações do ambiente j e $n = \sum_j^c n_j$. A parte bilinear do modelo está sujeita às restrições de ordem $\lambda_k \geq \lambda_{k+1}$ e de ortonormalidade dos vetores singulares ($\alpha_k^\top \alpha_k = \gamma_k^\top \gamma_k = 1$ e $\alpha_k^\top \alpha_{k'} = \gamma_k^\top \gamma_{k'} = 0$; $k \neq k'$) para $k = 1, \dots, t$. A distribuição amostral de \mathbf{y} , condicionada no vetor paramétrico $\boldsymbol{\theta} = (\beta, \lambda_k, \alpha_k, \gamma_k, \mathbf{V})$, é normal multivariada denotada por $\mathbf{y} | \boldsymbol{\theta} \sim N(\boldsymbol{\mu}_y, \mathbf{V})$, com

$$\boldsymbol{\mu}_y = \mathbf{X}_1 \beta + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z} \alpha_k) \mathbf{X}_2 \gamma_k.$$

Distribuições *a priori* para os parâmetros do modelo

As distribuições *a priori* assumidas para os parâmetros β , λ_k , α_k e γ_k do modelo são as mesmas atribuídas por Oliveira et al. (2016) sob homogeneidade de variâncias:

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

$$\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2 \sim N^+(\mu_{\lambda_k}, \sigma_{\lambda_k}^2), \quad \mu_{\lambda_k} = 0 \text{ e } \sigma_{\lambda_k}^2 = 10^8;$$

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

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

Para a componente de variância residual associada ao ambiente j , com $j = (1, \dots, c)$, é atribuída uma priori de Jeffreys, ou seja, $p(\sigma_{e_j}^2) \propto 1/\sigma_{e_j}^2$.

Distribuições condicionais completas *a posteriori*

A função de verossimilhança para $\boldsymbol{\theta}$ segundo o modelo (1) é dada por

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

Conectando verossimilhança com as informações *a priori* (via teorema de Bayes) obtém-se a distribuição conjunta *a posteriori* que pode ser escrita como:

$$P(\boldsymbol{\theta} | \mathbf{y}) \propto L(\boldsymbol{\theta} | \mathbf{y}) p(\boldsymbol{\beta} | \boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma_{\boldsymbol{\beta}}^2) \left\{ \prod_{k=1}^p P(\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2) P(\mathbf{a}_k) P(\boldsymbol{\gamma}_k) \right\} \prod_{j=1}^c P(\sigma_{e_j}^2). \quad (3)$$

As distribuições condicionais completas *a posteriori* para os parâmetros do modelo (1) são obtidas por manipulações algébricas a partir da distribuição conjunta *a posteriori* (3), sob as hipóteses consideradas *a priori*. São elas:

i. Distribuição condicional completa *a posteriori* para $\boldsymbol{\beta}$

$$\boldsymbol{\beta} | \dots \sim N \left[\left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 + \mathbf{I} \frac{1}{\sigma_{\beta}^2} \right)^{-1} \mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{M}, \left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 + \mathbf{I} \frac{1}{\sigma_{\beta}^2} \right)^{-1} \right] \quad (4)$$

sendo $\mathbf{M} = \mathbf{y} - \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k$ e $\sigma_{\beta}^2 = 10^8$.

ii. Distribuição condicional completa *a posteriori* para λ_k

$$\lambda_k | \dots \sim N^+ \left[\left(\boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k + \frac{1}{\sigma_{\lambda_k}^2} \right)^{-1} \boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \mathbf{M}_k, \left(\boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k + \frac{1}{\sigma_{\lambda_k}^2} \right)^{-1} \right] \quad (5)$$

sendo $\mathbf{M}_k = \mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \sum_{k' \neq k}^t \lambda_{k'} \text{diag}(\mathbf{Z}\mathbf{a}_{k'}) \mathbf{X}_2 \boldsymbol{\gamma}_{k'}$, $\boldsymbol{\Phi}_k = \text{diag}(\mathbf{Z}\mathbf{a}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k$ e $\sigma_{\lambda}^2 = 10^8$.

iii. Distribuição condicional completa *a posteriori* para \mathbf{a}_k

$$\mathbf{a}_k | \dots \sim N \left[\left(\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \right)^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_k, \left(\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \right)^{-1} \right] \quad (6)$$

sendo $\boldsymbol{\Lambda}_{\mathbf{a}_k} = \lambda_k \text{diag}(\mathbf{X}_2 \boldsymbol{\gamma}_k) \mathbf{Z}$.

iv. Distribuição condicional completa *a posteriori* para $\boldsymbol{\gamma}_k$

$$\boldsymbol{\gamma}_k | \dots \sim N \left[\left(\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \right)^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \mathbf{M}_k, \left(\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \right)^{-1} \right] \quad (7)$$

sendo $\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} = \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k) \mathbf{X}_2$.

v. Distribuição condicional completa *a posteriori* para $\sigma_{e_j}^2$

$$\sigma_{e_j}^2 | \dots \sim Inv-Esc - \chi^2 \left[n_j, \frac{(\mathbf{y}_j - \boldsymbol{\mu}_{\mathbf{y}_j})^\top (\mathbf{y}_j - \boldsymbol{\mu}_{\mathbf{y}_j})}{n_j} \right] \quad (8)$$

em que \mathbf{y}_j , $\boldsymbol{\mu}_{\mathbf{y}_j}$ e n_j são o vetor de respostas fenotípicas, vetor de médias e o número de observações referentes ao ambiente j , respectivamente.

Devido a restrição de ortogonalidade ($\mathbf{a}_k^\top \mathbf{a}_{k'} = \boldsymbol{\gamma}_k^\top \boldsymbol{\gamma}_{k'} = 0$ para $k \neq k'$), os vetores \mathbf{a}_k e $\boldsymbol{\gamma}_k$ devem ser distribuídos somente em um subespaço restrito da esfera unitária em \mathbb{R}^p (para $p = r$ ou $p = c$, respectivamente). Eles devem ser ortogonais a $s(0 \leq s \leq t-1)$ direções indicadas por vetores na dimensão p . A amostragem dos vetores singulares é realizada no subespaço corrigido ($p-s$), com correto suporte, pela definição de variáveis auxiliares $\tilde{\mathbf{a}}_k = \mathbf{H}_k^\top \mathbf{a}_k$ e $\tilde{\boldsymbol{\gamma}}_k = \mathbf{D}_k^\top \boldsymbol{\gamma}_k$, em que \mathbf{H}_k e \mathbf{D}_k são matrizes de transformação linear ortogonal. As distribuições condicionais completas *a posteriori* para os vetores singulares no subespaço corrigido são dadas por:

$$\tilde{\mathbf{a}}_k | \dots \sim N(\mathbf{m}_{\mathbf{a}_k}, \mathbf{W}_{\mathbf{a}_k}) \quad (9)$$

com $\mathbf{m}_{\mathbf{a}} = \Delta_{\mathbf{a}_k} \left(\Delta_{\mathbf{a}_k}^\top \Delta_{\mathbf{a}_k} \right)^{-\frac{1}{2}}$, $\Delta_{\mathbf{a}_k} = \mathbf{H}_k^\top \left(\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k} \right)^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_k$ e

$$\mathbf{W}_{\mathbf{a}_k} = \left\{ \left[\left(\Delta_{\mathbf{a}_k}^\top \Delta_{\mathbf{a}_k} \right)^{-\frac{1}{2}} \right]^\top \mathbf{H}_k^\top \left(\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k} \right) \mathbf{H}_k \bar{m}_{\mathbf{a}_k} \left(\Delta_{\mathbf{a}_k}^\top \Delta_{\mathbf{a}_k} \right)^{-\frac{1}{2}} \right\}^{-1},$$

sendo $\bar{m}_{\mathbf{a}_k} = \Delta_{\mathbf{a}_k}^\top \Delta_{\mathbf{a}_k}$, em que $\left(\Delta_{\mathbf{a}_k}^\top \Delta_{\mathbf{a}_k} \right)^{1/2}$ é uma constante normalizadora.

Então o vetor $\tilde{\mathbf{a}}_k$ é amostrado no subespaço corrigido em \mathbb{R}^{r-s} , sendo obtido no subespaço correto em \mathbb{R}^r pela transformação inversa $\mathbf{a}_k = \mathbf{H}_k \tilde{\mathbf{a}}_k$.

De maneira análoga, a densidade condicional completa *a posteriori* para $\tilde{\boldsymbol{\gamma}}_k$, no subespaço corrigido, é dada por:

$$\tilde{\boldsymbol{\gamma}}_k | \dots \sim N(\mathbf{m}_{\boldsymbol{\gamma}_k}, \mathbf{W}_{\boldsymbol{\gamma}_k}) \quad (10)$$

em que $m_{\boldsymbol{\gamma}_k} = \Delta_{\boldsymbol{\gamma}_k} \left(\Delta_{\boldsymbol{\gamma}_k}^\top \Delta_{\boldsymbol{\gamma}_k} \right)^{-\frac{1}{2}}$, com $\Delta_{\boldsymbol{\gamma}_k} = \mathbf{D}_k^\top \left(\Lambda_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \Lambda_{\boldsymbol{\gamma}_k} \right)^{-1} \Lambda_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \mathbf{M}_k$ e

$$\mathbf{W}_{\boldsymbol{\gamma}_k} = \left\{ \left[\left(\Delta_{\boldsymbol{\gamma}_k}^\top \Delta_{\boldsymbol{\gamma}_k} \right)^{-\frac{1}{2}} \right]^\top \mathbf{D}_k^\top \left(\Lambda_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \Lambda_{\boldsymbol{\gamma}_k} \right) \mathbf{D}_k \bar{m}_{\boldsymbol{\gamma}_k} \left(\Delta_{\boldsymbol{\gamma}_k}^\top \Delta_{\boldsymbol{\gamma}_k} \right)^{-\frac{1}{2}} \right\}^{-1}.$$

sendo $\bar{m}_{\gamma_k} = \Delta_{\gamma_k}^\top \Delta_{\gamma_k}$.

Os detalhes algébricos, de como as condicionais foram obtidas, estão apresentados no Apêndice S1.

As hipóteses e distribuições condicionais *a posteriori* descritas acima se referem a modelagem bayesiana do GGE sob heterogeneidade de variâncias residuais nos locais e este modelo será referido no decorrer do texto como BGGEH.

Para efeito de comparações o modelo GGE que pressupõe variância única em todos os locais (BGGE), como descrito em Oliveira et al. (2016), também foi ajustado.

Para essa análise a distribuição amostral é $\mathbf{y} | \boldsymbol{\theta} \sim N(\boldsymbol{\mu}_y, \mathbf{I}\sigma_e^2)$, sendo \mathbf{I} uma matriz identidade de ordem n . As densidades *a priori* atribuídas aos parâmetros são basicamente as mesmas utilizadas para o caso heterogêneo. A exceção, como já ressaltado, está na hipótese sobre a estrutura de variância. No BGGE considera-se uma variância residual σ_e^2 comum em que *a priori* é $p(\sigma_e^2) \propto 1/\sigma_e^2$. Ao assumir esse pressuposto, as densidades *a posteriori* condicionais são diferentes e descritas, em detalhes, por Oliveira et al. (2016).

Amostragem e inferência para parâmetros lineares e bilineares

Como todas as distribuições condicionais completas são conhecidas o processo de amostral foi realizado de forma direta utilizando um amostrador de Gibbs que pertence à classe dos métodos de Monte Carlo via Cadeia de Markov (MCMC). O algoritmo para a amostragem é análogo ao descrito por Silva et al. (2019), no contexto do AMMI com heterogeneidade, sendo exemplificado no Apêndice S2.

Para todas as cadeias de Markov, valores foram descartados a fim de eliminar o efeito inicial (*burn-in*). Além disso, analisou-se a função de auto-correlação das cadeias adotando-se intervalos (*thinning*) fixos para seleção de uma amostra aproximadamente não-correlacionada para os parâmetros. Essas correções foram baseadas em amostras de treinamento com base no critério de Raftery e Lewis (1992).

A convergência das cadeias MCMC foi monitorada pelos métodos de Heidel-Berger e Welch (1983) e Raftery e Lewis (1992).

As estimativas pontuais e intervalares para β , λ_k , e $\sigma_{e_j}^2$ foram obtidas pela média, máximo *a posteriori* (MAP) e mediana dos valores MCMC simulados. Intervalos de máxima densidade *a posteriori* (HPD), ao nível de 95% de credibilidade,

foram construídos utilizando o método de Chen e Shao (1999). Estimativas pontuais para α_k e γ_k , foram obtidas pelo método utilizado por Liu (2001).

Regiões de credibilidade bivariadas (a 95%) para escores genotípicos ($\lambda_1^{1/2}\alpha_{i1}, \lambda_1^{1/2}\alpha_{i2}$) e ambientais ($\lambda_1^{1/2}\gamma_{j1}, \lambda_1^{1/2}\gamma_{j2}$), com $i = 1, \dots, r$ e $j = 1, \dots, c$, foram incorporadas ao biplot utilizando-se o método de Hu e Yang (2013).

Seleção de modelos utilizando critérios de informação

A seleção de modelo, de acordo com o número de componentes bilineares retidos, foi realizada pelos seguintes critérios: critério de informação bayesiano (BIC), critério de informação de Akaike (AIC) e pelo critério de informação de Akaike Monte Carlo (AICM). As estatísticas utilizadas são as seguintes:

- I. $BIC = -2 \ln(L(\hat{\theta}_p)) + p \ln(n)$, em que $L(\hat{\theta}_p)$ é a máxima verossimilhança do modelo estimado, p o número de parâmetros e n número de observações na cadeia (Schwarz, 1978);
- II. $AIC = -2 \ln(L(\hat{\theta}_p)) + 2p$, em que p é o número de parâmetros a serem estimados no modelo (Akaike, 1974);
- III. $AICM = -2(s_l^2 - \bar{l})$, sendo \bar{l} a média *a posteriori* da log-verossimilhança marginal e s_l^2 a variância *a posteriori* da log-verossimilhança marginal (Raftery et al., 2007).

Todo o processo de inferência e simulação foi realizado utilizando-se o software estatístico R (R Core Team, 2019).

Resultados

Para todos os modelos analisados, os testes de convergências indicaram boas propriedades com fator de dependência (I) sempre menor que cinco ($I < 5$) (RAFTERY; LEWIS, 1992). Além disso, as cadeias para todos os parâmetros passaram pelo teste de estacionariedade, o que é indício de convergência (HEIDELBERGER; WELCH, 1983).

Gráficos dos traços das cadeias das componentes da variância residual para os dados reais são apresentadas no Apêndice S3 (Figuras S1 a S4). Nesses gráficos é possível observar oscilações aleatórias de valores em torno de um valor constante (estacionariedade) o que indica que as cadeias convergiram para distribuição

estacionária, ou seja, para a distribuição marginal *a posteriori*. Esse padrão foi observado para todos os parâmetros em todos os ajustes (dados não mostrados).

Análise dos dados simulados

Na Tabela 2 são apresentados resumos pontuais e intervalares das componentes da variância residual para os dados homocedásticos simulados ($\sigma_e^2 = 1$). Intervalos de credibilidade das componentes se sobrepõem indicando não existir diferenças expressivas entre os valores. A estimativa pontual da variância comum (BGGE) é incluída em todas as regiões HPD das componentes de variâncias estimadas no modelo BGGEH. Ressalta-se que o modelo que considera variância específica em cada ambiente contempla bem a estrutura homocedástica dos dados.

Tabela 2. Resumos pontuais e intervalares das densidades *a posteriori* das componentes da variância residual dos modelos BGGEH e BGGE para dados simulados considerando o cenário homocedástico.

Modelo	PAR	MAP	MED	ME	DP	HPD (a 95% de credibilidade)	
						LI	LS
BGGE	σ_e^2	1,02	1,04	1,04	0,07	0,91	1,17
	$\sigma_{e_1}^2$	1,03	1,05	1,08	0,23	0,68	1,55
	$\sigma_{e_2}^2$	0,92	0,98	1,01	0,22	0,62	1,46
	$\sigma_{e_3}^2$	0,75	0,79	0,82	0,18	0,50	1,20
	$\sigma_{e_4}^2$	1,03	1,06	1,09	0,24	0,67	1,58
	$\sigma_{e_5}^2$	1,01	1,06	1,08	0,23	0,68	1,53
	$\sigma_{e_6}^2$	1,03	1,10	1,13	0,25	0,71	1,64
	$\sigma_{e_7}^2$	1,12	1,15	1,19	0,25	0,73	1,68
	$\sigma_{e_8}^2$	0,88	0,91	0,94	0,21	0,56	1,37
	$\sigma_{e_9}^2$	0,98	1,09	1,12	0,24	0,70	1,60
	$\sigma_{e_{10}}^2$	0,65	0,70	0,73	0,18	0,42	1,08

PAR, parâmetro; **MAP**, máximo *a posteriori*; **MED**, mediana; **ME**, média; **DP**, desvio padrão; **LI**, limite inferior; **LS**, limite superior.

Estimativas *a posteriori* das componentes de variância para dados heterocedásticos são apresentados na Tabela 3. As estimativas de maior e menor magnitude foram observadas para os ambientes E10 e E2, respectivamente (valores paramétricos simulados), sendo a razão E10/E2 igual a 34. Considerando estimativas pontuais, o maior valor foi observado para E9 e o menor para E2. A razão entre a maior e menor delas foi igual a 33,61, 32,08 e 30,83 para MAP, mediana e média, respectivamente. Estimativas pontuais para as componentes de variância são próximas

dos valores paramétricos simulados e contempladas, nas respectivas regiões HPD, pelo modelo BGGEH.

Tabela 3. Resumos pontuais e intervalares das densidades *a posteriori* das componentes da variância residual dos modelos BGGEH e BGGE para dados simulados considerando o cenário heterocedástico.

PAR	VS	MODELO	HPD (a 95% de credibilidade)					
			MAP	MED	ME	DP	LI	LS
σ_e^2	-	BGGE	7,53	7,62	7,64	0,54	6,62	8,69
$\sigma_{e_1}^2$	10,50	BGGEH	10,41	10,75	11,02	2,22	7,19	15,35
$\sigma_{e_2}^2$	0,50		0,57	0,64	0,68	0,55	0,35	1,09
$\sigma_{e_3}^2$	3,00		2,20	2,36	2,44	1,01	1,52	3,49
$\sigma_{e_4}^2$	2,00		1,98	2,17	2,24	0,60	1,37	3,25
$\sigma_{e_5}^2$	3,00		3,21	3,37	3,46	0,81	2,19	4,89
$\sigma_{e_6}^2$	5,00		5,81	6,10	6,26	1,58	4,01	8,90
$\sigma_{e_7}^2$	9,00		12,86	13,78	14,15	3,01	8,98	20,30
$\sigma_{e_8}^2$	9,00		12,24	12,79	13,16	2,99	7,61	18,99
$\sigma_{e_9}^2$	15,00		19,29	20,56	21,10	4,48	13,01	29,77
$\sigma_{e_{10}}^2$	17,00		15,22	16,61	17,08	3,81	10,16	24,54

PAR, parâmetro; **VS**, valor paramétrico simulado; **MAP**, máximo *a posteriori*; **MED**, mediana; **ME**, média; **DP**, desvio padrão; **LI**, limite inferior; **LS**, limite superior.

Estimativas pontuais da variância comum (BGGE) estão incluídas apenas nos intervalos relacionados aos ambientes E1, E6 e E8. Para as demais componentes, não existem sequer sobreposições entre suas HPDs com a HPD da variância comum. Além disso, nenhum dos valores simulados está contido na região de credibilidade para a variância comum. Esses resultados indicam que o modelo BGGEH também foi eficiente em capturar a estrutura de heterogeneidade presente e que a suposição de uma variância comum não seria condizente.

Resumos pontuais e intervalares para valores singulares são apresentadas nas Tabelas S1 e S2 (Apêndice S4). Para dados homogêneos (Tabela S1), BGGE e BGGEH possuem estimativas pontuais bem próximas para os três primeiros valores singulares. Os dois primeiros componentes principais explicam praticamente toda variabilidade de G+GEI em ambos os modelos. A soma de quadrado total de G+GEI (SQ_{GGE}) para BGGEH (3420,82) é ligeiramente menor que aquela obtida pelo BGGE (3432,53). Nota-se, novamente, a semelhança entre BGGE e BGGEH na análise de dados homocedásticos.

Para o cenário heterocedástico (Tabela S2) ocorre encolhimento mais acentuado para as estimativas do segundo e terceiro valores singulares no BGGEH (em relação ao BGGE). O primeiro componente principal em BGGEH explica 96,64% da SQ_{GGE} sendo que o percentual explicado pelos dois primeiros eixos é de 99,06%. No modelo BGGE, o primeiro PC explica 86,14% e os dois primeiros componentes explicam 97,45% da variabilidade total da G+GEI. A soma de quadrado total recuperada pelo modelo BGGE é maior (3537,98) do que aquela obtida pelo BGGEH (2983,55) e essa diferença é mais acentuada quando comparada à análise com dados homogêneos. Isto pode ser um indicativo de que a interação esteja sendo superestimada no BGGE.

Na Figura 2, são apresentados os biplots juntamente com regiões de credibilidade bivariadas para escores genotípicos e ambientais. Em cada biplot foi inserido um sistema de eixos composto pelo eixo do ambiente médio (AEA) e sua ordenada (AEC). O ambiente médio (AE) e o genótipo ideal (IG), definido para ter o rendimento máximo entre os ambientes, também foram representados, com as respectivas regiões de credibilidade. O sentido da semirreta na AEA (da origem para o IG) indica maiores valores de rendimento médio para genótipos. Vetores ambientais foram inseridos e auxiliaram na visualização de correlações entre os ambientes, sendo as normas desses vetores uma medida da capacidade do ambiente em discriminar genótipos. Quanto menor o ângulo entre um dado ambiente em relação ao ambiente médio, mais representativo ele será do ambiente-alvo. Genótipos, por sua vez, são comparados (no ambiente médio) em função da sua distância euclidiana ao genótipo ideal. Esses conceitos, em termos pontuais, são descritos por Yan e Kang (2003) e Yan (2014). Apenas genótipos e ambientes cujas regiões bivariadas não incluem a origem são representados para simplificar as interpretações.

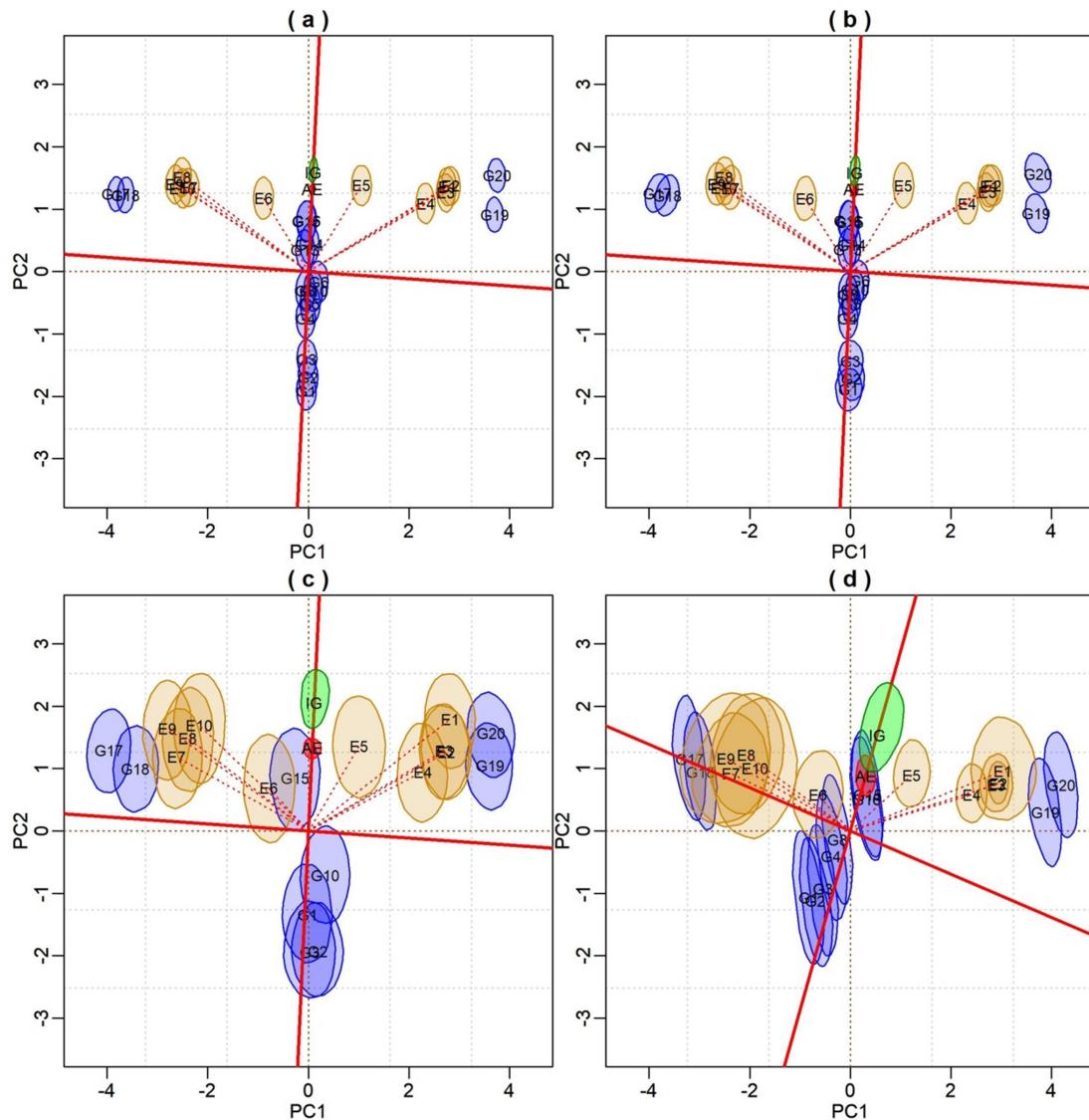


Figura 2. Biplot GGE com regiões bivariadas, a 95% de credibilidade, para escores genotípicos (G), ambientais (E), ambiente médio (AE) e genótipo ideal (IG) em cenário homocedástico: (a) BGGE e (b) BGGEH e em cenário heterocedástico: (c) BGGE e (d) BGGEH.

Os biplots para o cenário homocedástico (Figuras 2: a e b) possuem praticamente o mesmo padrão. As regiões de credibilidade para escores ambientais, genótipo ideal, ambiente médio e escores genotípicos mais afastados da origem apresentam amplitudes ligeiramente maiores para o BGGEH (Figura 2b). Não obstante às diferenças de amplitudes, esses biplots fornecem, essencialmente, as mesmas inferências no que se refere à estabilidade e adaptabilidade.

Biplots exibidos na Figura 2 (c e d) ilustram o cenário heterocedástico. No modelo BGGE (Figura 2c), o genótipo mais próximo ao ideal foi o G15, podendo ser considerado estável, devido a sua região de credibilidade ser interceptada pelo eixo AEA e apresenta rendimento superior à média geral, já que a região de credibilidade

bivariada está acima da AEC (valores menores ou iguais a média geral não são prováveis). Os ambientes E5 e E6 são os mais representativos do ambiente médio. Aqueles ambientes com vetores mais longos têm maior capacidade discriminativa de genótipos.

Para o modelo BGGEH (Figura 2d), E5 e E6 também são os ambientes mais representativos, enquanto G15 e G16 são os genótipos com menor distância ao ideal, mas suas regiões de credibilidade são interceptadas pelo eixo AEC, o que indica que a valores menores ou igual a média geral dos rendimentos são prováveis. Ambos os modelos destacaram G17, G18, G19 e G20 como os genótipos com contribuições importantes para a GEI.

Contudo, diferenças expressivas são visualizadas entre as duas representações. Para BGGEH observam-se diferentes amplitudes para as regiões de credibilidade ambientais, que refletem as magnitudes das componentes de variância residual em cada ambiente, indicando que a comparação de genótipos envolve diferentes precisões. Nota-se ainda que os ambientes relacionados a componentes de maior magnitude (à esquerda da AEA) tiveram maior efeito de encolhimento que aqueles com menores variações (á direita). O E1 possui a componente de maior magnitude dentre os ambientes à esquerda da AEA, o que é facilmente visualizado no biplot pela maior amplitude de sua região bivariada, bem como pelo encolhimento mais acentuado do seu vetor médio em comparação ao biplot BGGE.

Regiões bivariadas para os escores genotípicos do BGGEH apresentam maior variabilidade no sentido do eixo maior (PC2). Isso pode ser devido ao fato do primeiro componente explicar uma grande porcentagem da variação da SQ_{GGE} (cerca de 97%) e, portanto, apresentar maior precisão. A magnitude do valor singular está diretamente associada a menor dispersão dos escores em torno da média. O segundo valor singular (Tabela S2) explica apenas 2% da variabilidade de G+GEI e isso faz com que escores genotípicos no segundo PC tenham maior dispersão quando comparado ao primeiro, bem como aos escores plotados no biplot BGGE (em que PC2 explica 11,32% da SQ_{GGE}).

No biplot BGGEH, o genótipo G10 não é representado, já que a região de credibilidade para os escores englobou a origem. Contudo, mais genótipos são representados (saíram da origem). Observa-se também rotação mais acentuada do sistema de coordenadas adicional para o modelo que modela heterogeneidade, de forma que o eixo AEC intercepta todas as regiões presentes no segundo quadrante (em relação

ao sistema PC1 e PC2), indicando que para os respectivos genótipos valores nulos e negativos possuem probabilidade diferente de zero de serem observados — quando avaliado no ambiente médio — contrastando, portanto, com o modelo BGGE. Nota-se ainda que regiões de credibilidade para ambiente médio e genótipo ideal têm maiores amplitudes no biplot BGGEH e os vetores são mais encolhidos mostrando que modelar as diferentes precisões, nos ambientes, afeta a análise em um ambiente médio.

No que se refere a estabilidade é possível notar que todos os genótipos simulados, para não influenciar expressivamente a GEI, foram identificados em ambos os biplots. Contudo é possível notar que todos os genótipos estáveis simulados respondem positivamente à melhoria ambiental (Figura 1) e isso, de forma geral, parece ter sido captado melhor no BGGEH biplot. Outro ponto a ser ressaltado é que, embora todos os genótipos importantes para o padrão da interação tenham sido corretamente identificados, em ambas as modelagens, nota-se encolhimento em direção a origem para esses genótipos no BGGEH (em relação a BGGE). Como já ressaltado, isso pode sugerir que a interação esteja sendo superestimada na análise BGGE.

Biplots exibindo o padrão “quem venceu onde” são apresentados nas Figuras S5 e S6 (Apêndice S5). Nessa versão biplot, um polígono é incorporado a representação gráfica, em que os vértices são as médias de escores genotípicos mais afastadas da origem. Semirretas traçadas da origem e ortogonais aos lados desse polígono dividem o biplot em setores, sendo que o genótipo do vértice tem maior rendimento nos ambientes que se encontram no mesmo setor. Esses biplots (assim como aqueles apresentados na Figura 2) sinalizam que uma análise baseada no ambiente médio não seria plausível.

Na Figura S5 (cenário homocedástico), o padrão apresentado é o mesmo para ambos os biplots. Para o cenário heterocedástico (Figura S6), observam-se diferenças nas posições de algumas regiões de credibilidade em relação aos setores, principalmente para os genótipos mais próximos à origem do biplot. No BGGEH, observa-se o agrupamento mais nítido dos ambientes E1, E2, E3 e E4, bem como maior distinção desse subgrupo com E5, quando comparado com o padrão exibido pelo BGGE. Agrupamento mais nítido também é visualizado no BGGEH para E7, E8, E9 e E10, bem como maior proximidade com E6.

Os aspectos ressaltados acima indicam que a suposição sobre a estrutura da variância residual, no modelo GGE, acarreta diferenças em termos inferenciais para o conjunto de dados heterocedásticos simulados. Por outro lado, não houve fortes implicações em termos de mega-ambientes ao se comparar os dois modelos, o que é de

grande importância quando se geram os polígonos no biplot. Aqui foi simulado um forte efeito para a GEI, mas o padrão foi explicado basicamente pelo primeiro componente principal e isso pode ter favorecido os desempenhos semelhantes entre os modelos, no que se refere à análise de adaptabilidade. Sob uma interação mais complexa, e menos evidente, os modelos poderiam levar a diferenças mais impactantes na identificação de subgrupos separáveis de ambientes.

Análise dos dados reais

Estimativas pontuais e intervalares para componentes da variância residual (BGGEH) e variância residual comum (BGGE) são apresentadas na Tabela 4. A maior variabilidade é observada no ambiente E1 e a menor em E7, sendo que a razão média entre a maior e menor delas foi de 11,14. Como observado, a estimativa da variância comum, obtida a partir do modelo BGGE, está incluída em intervalos de credibilidade de apenas três componentes, que são aquelas relacionadas aos ambientes E2, E5 e E10. Para os demais ambientes, intervalos HPD para as componentes não se sobrepõem ao da variância comum.

Tabela 4. Estimativas obtidas por modelo misto linear e resumos pontuais e intervalares das densidades *a posteriori* para componentes de variâncias residual dos modelos BGGEH e BGGE aplicados ao conjunto MET real de sorgo para a produção de biomassa seca total em ton/ha.

PAR	IMLM	MODELO	HPD (a 95% de credibilidade)					
			MAP	MED	ME	DP	LI	LS
σ_e^2	-	BGGE	141,17	141,90	142,19	7,42	128,69	157,60
$\sigma_{e_1}^2$	372,59		420,46	434,53	440,75	70,84	304,47	578,74
$\sigma_{e_2}^2$	123,84		122,15	126,80	129,20	22,56	87,34	172,53
$\sigma_{e_3}^2$	88,92		81,33	83,78	85,70	15,54	58,47	117,88
$\sigma_{e_4}^2$	303,02		262,76	266,35	270,37	40,80	197,38	351,28
$\sigma_{e_5}^2$	121,65	BGGEH	118,24	121,75	123,89	20,91	86,89	166,49
$\sigma_{e_6}^2$	60,28		53,55	55,20	56,24	9,61	39,17	74,90
$\sigma_{e_7}^2$	45,37		37,74	39,91	40,54	6,94	28,71	54,12
$\sigma_{e_8}^2$	89,51		85,23	85,51	87,00	15,11	59,21	116,42
$\sigma_{e_9}^2$	67,70		59,41	63,21	64,3	11,11	44,52	86,60
$\sigma_{e_{10}}^2$	150,96		126,42	131,57	133,93	21,12	97,14	176,78

PAR, parâmetro; **IMLM**, modelo linear misto aplicado individualmente em cada local; **MAP**, máximo *a posteriori*; **MED**, mediana; **ME**, média; **DP**, desvio padrão; **LI** limite inferior; **LS**, limite superior.

Para fins de comparação, estimativas da variância residual, em cada local, também foram obtidas por ajustes individuais com modelo linear misto (IMLM), utilizando a biblioteca lme4 (BATES et al., 2015) implementada no software R, e estão

incluídas nos intervalos bayesianos estimados para os respectivos locais, segundo o BGGEH (Tabela 4), indicando semelhanças entre BGGEH e a análise IMLM. A razão entre as estimativas (maior/menor) obtidas nas análises individuais foi de 8,21. Observa-se ainda que, em 9 dos 10 ambientes, o modelo BGGEH apresentou estimativa da componente da variância do erro menor, em relação a análise IMLM. Embora não se queira comparar rigorosamente a análise conjunta com a análise individual (em cada ambiente), esse resultado sugere um melhor ajuste para BGGEH.

Resumos *a posteriori* para valores singulares são apresentados na Tabela 5. Os dois primeiros componentes principais explicaram 93,09% e 91,41% da SQ_{GGE} para BGGE e BGGEH, respectivamente. Contudo, a estimativa pontual *a posteriori* do percentual explicado pelo primeiro componente do BGGEH foi ligeiramente maior. Notam-se também menores magnitudes e encolhimentos mais expressivos para estimativas pontuais *a posteriori* dos três primeiros valores singulares no modelo BGGEH e a partir daí observa-se comportamento contrário, com efeito de encolhimento cada vez mais acentuado para BGGE. A SQ_{GGE} para BGGEH, a exemplo do cenário simulado, foi menor que a do BGGE.

Tabela 5. Resumos pontuais e intervalares *a posteriori* para quantidades relacionadas aos valores singulares obtidos pelos modelos BGGE e BGGEH, aplicados ao conjunto de dados MET real.

λ_k	Modelo BGGE						
	HPD (a 95% de credibilidade)						
	PER	PEA	SQA	ME	DP	LI	LS
λ_1	80,46	80,46	39653,31	199,13	7,43	184,88	213,46
λ_2	12,63	93,09	45879,47	78,91	7,89	63,61	93,85
λ_3	5,18	98,27	48431,52	50,52	9,78	31,60	68,15
λ_4	1,47	99,73	49154,34	26,88	13,49	0,01	47,39
λ_5	0,22	99,95	49260,53	10,30	8,29	<0,01	26,66
λ_6	0,04	99,99	49279,91	4,40	4,62	<<0,01	13,95
λ_7	0,01	100	49284,1	2,05	2,63	<<0,01	7,68
λ_8	0	100	49285,05	0,97	1,60	<<0,01	3,99
λ_9	0	100	49285,26	0,46	0,85	<<0,01	2,30
λ_{10}	0	100	49285,31	0,23	0,51	<<0,01	1,14
Modelo BGGEH							
HPD (a 95% de credibilidade)							
	PER	PEA	SQA	ME	DP	LI	LS
λ_1	83,9	83,90	36980,16	192,30	11,29	171,42	215,65

λ_2	7,51	91,41	40291,16	57,54	8,98	40,78	75,57
λ_3	3,88	95,30	42002,37	41,37	7,50	26,80	55,67
λ_4	2,08	97,38	42920,4	30,30	6,03	18,83	42,21
λ_5	1,22	98,60	43459,46	23,22	5,36	12,93	33,88
λ_6	0,74	99,34	43786,83	18,09	4,78	8,88	27,54
λ_7	0,41	99,75	43965,09	13,35	4,45	4,31	22,00
λ_8	0,17	99,92	44038,43	8,56	3,78	0,93	15,52
λ_9	0,06	99,98	44066,53	5,301	3,19	<0,01	10,92
λ_{10}	0,02	1	44075,58	3,02	2,38	<<0,01	7,55

PER, percentual relativo da soma de quadrados de G+GEI pela componente k ; **PEA**, percentual da soma de quadrados de G+GEI acumulado; **SQA**, soma de quadrados acumulado em relação aos valores singulares; **ME**, média; **DP**, desvio padrão; **LI**, limite inferior; **LS**, limite superior.

Nota: A<<B significa que A<0,01B.

Resultados da aplicação dos critérios de informação para seleção de modelos são apresentados na Figura 3. Modelos BGGEH, de forma geral, obtiveram valores expressivamente menores em ambos os critérios. Em relação aos modelos vencedores (em cada versão GGE) e com dois termos bilineares, o BGGEH também obteve desempenhos notavelmente melhores (com destaque ao AICM). A obtenção de um modelo que se aproxime bem dos dados reais é um aspecto importante para validar inferências na análise de estabilidade do genótipo. Esses resultados sugerem que o modelo com heterogeneidade de variância residual entre locais se ajustou melhor aos dados e se mostrou mais adequado do que sua versão homogênea.

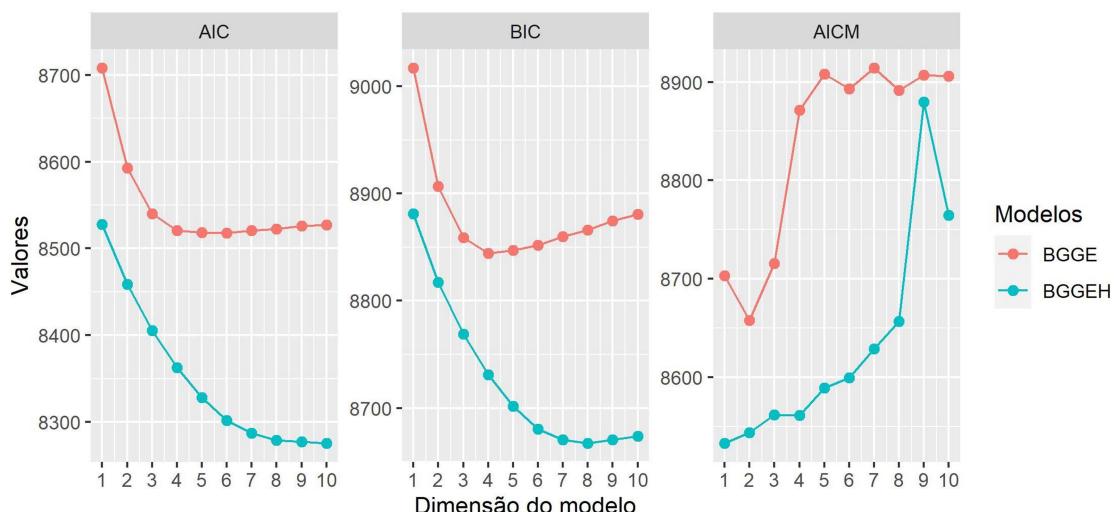


Figura 3. Resultados dos critérios de informação utilizados para seleção de modelos.

Nos biplots apresentados na Figura 4 é possível avaliar genótipos e ambientes com base no ambiente médio. As regiões de credibilidade para o ambiente médio e genótipo ideal permitem a avaliação gráfica da “representatividade versus discriminação” em relação a ambientes, bem como da “média versus estabilidade” em relação a genótipos, de forma semelhante ao que foi realizado para dados simulados. Genótipos cujas regiões englobaram a origem também não foram representados para simplificar as interpretações.

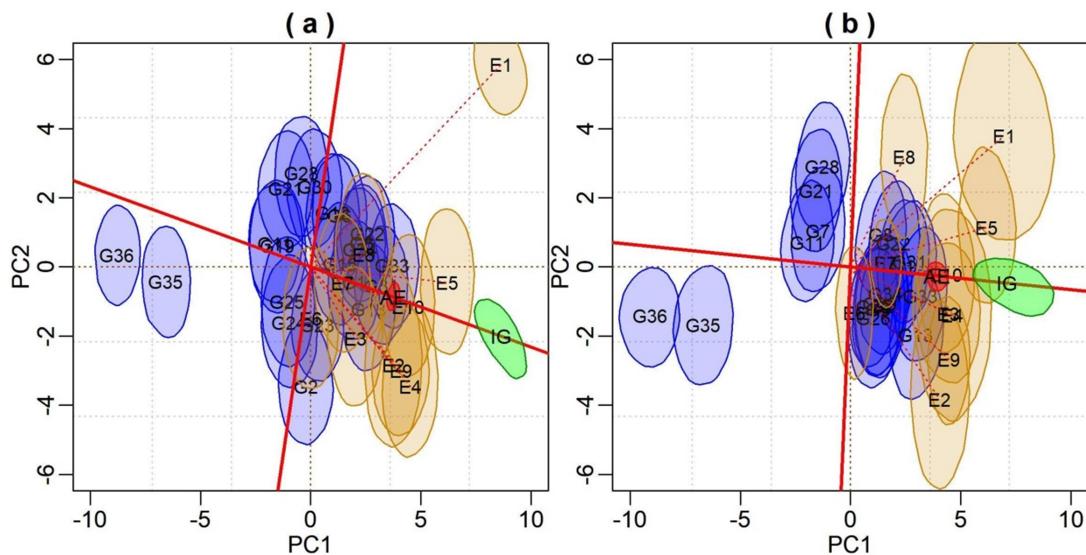


Figura 4. Biplots BGGE e BGGEH juntamente com regiões bivariadas, a 95% de credibilidade, para escores genotípicos (G), escores ambientais (E), ambiente médio (AE) e genótipo ideal (IG). Foi adicionado em cada biplot o sistema ortogonal AEA e AEC.

As regiões de credibilidade ambientais para o BGGEH traduzem as variações específicas observadas em cada ambiente por meio de amplitudes heterogêneas com diferentes precisões. Observa-se que E1 é encolhido em direção à interação média em BGGEH, mas se destaca como o ambiente com maior capacidade de discriminação em BGGE, sendo considerado como o ambiente mais importante para a interação. O oposto ocorre com E8, que tem contribuição aparentemente pequena para a interação no BGGE e devido à variância, relativamente pequena, aparece mais distante do eixo médio no BGGEH. A região de credibilidade para o ambiente E6 engloba a origem, sinalizando que esse ambiente seria aquele com menor capacidade para discriminar genótipos (sob as duas hipóteses), ou ainda, o que apresenta menor associação com os demais ambientes. Mas isso pode ser afirmado com maior precisão para o BGGEH.

Além desses, poderiam ser ressaltados outros aspectos em relação à comparação do biplot BGGE e BGGEH. É possível observar que o ambiente médio e o genótipo

ideal estão mais afastados da origem em BGGE que também apresenta rotação mais acentuada do sistema de eixo adicional ($AEA \times AEC$). Outro aspecto visível é a maior distância entre o ambiente médio e o genótipo ideal na análise BGGE. Além disso, E1 apresenta um comportamento extremo aos demais ambientes no biplot BGGE, o que influencia a análise com base em um ambiente médio. Isso é contrastante com a análise BGGEH, em que E1 compõe um subgrupo com E5 e E8, sendo a maior variabilidade de E1 modelada em termos de maior amplitude de sua região de credibilidade.

Constata-se também que alguns padrões são mantidos e outros sofrem alterações mais expressivas nas representações gráficas. Como o caso do G2, que é um dos genótipos mais importante para a interação no modelo BGGE, mas não é destacado no BGGEH. Regiões genotípicas interceptadas pela AEA sinalizam que os respectivos genótipos não tem contribuição importante para a GEI. Nesse sentido, observa-se um maior número de genótipos não estáveis no biplot BGGE (considerando a avaliação no ambiente médio) devido a maior rotação do sistema adicional de eixos. Isso sugere que a hipótese sobre a estrutura de variância em modelo GGE pode influenciar a avaliação de estabilidade e adaptabilidade de genótipos em ensaios MET.

Na Figura 5 são apresentados os biplots exibindo o padrão “quem venceu onde” para ambos os modelos. Essa configuração gráfica geralmente é utilizada para a identificação de mega-ambientes. Como visualizado, o biplot para o modelo heterogêneo (BGGEH) apresenta uma separação mais nítida entre subgrupos de ambientes baseado na propriedade do produto interno, o que também pode ser visualizado na Figura S7 (Apêndice S5). Embora a maioria das regiões de credibilidade esteja dispersa por mais de um setor e ainda apresente diferentes níveis de sobreposições — o que referimos por sobreposições complexas. Isso é um indicativo de que ignorar a presença de heterocedasticidade nos dados pode comprometer as interpretações, sugerindo equivocadamente uma simplificação da estrutura da GEI.

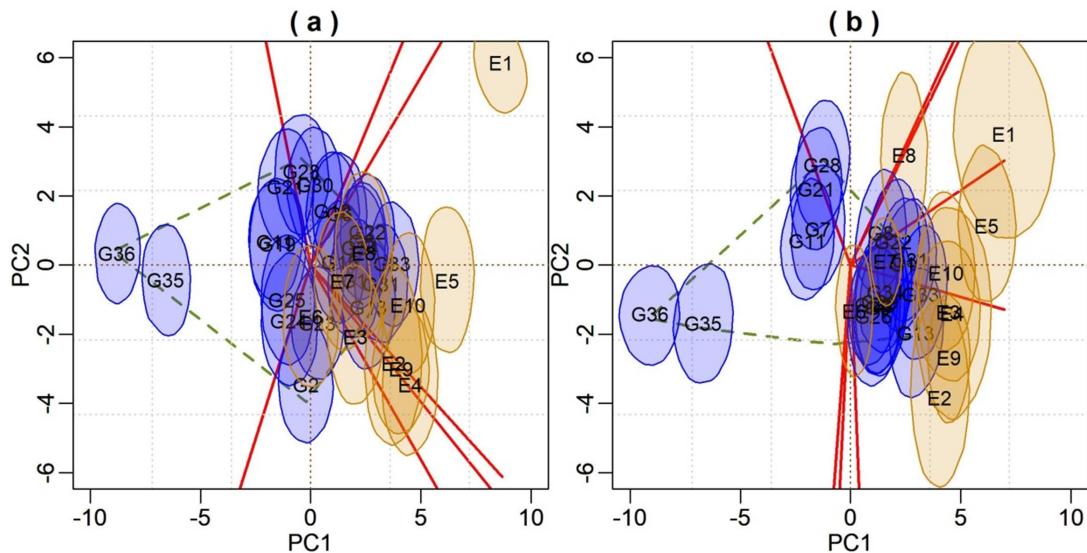


Figura 5. Biplots GGE com regiões bivariadas, a 95% de credibilidade, para escores genotípicos e ambientais exibindo “o padrão quem venceu onde”; (a) biplot BGGE e (b) biplot BGGEH.

Outras diferenças podem ser apontadas com relação aos modelos, como mudança em relação aos genótipos que compõem os vértices do polígono do lado direito do PC2 (1º e 4º quadrantes). Uma melhor visualização do padrão de ambientes e genótipos pode ser obtida em biplots pontuais, construídos com as médias *a posteriori* (Figura 6). Ao retirar as regiões de credibilidade, observam-se mais nitidamente as diferenças de rotações de escores médios que podem afetar as interpretações e possíveis inferências.

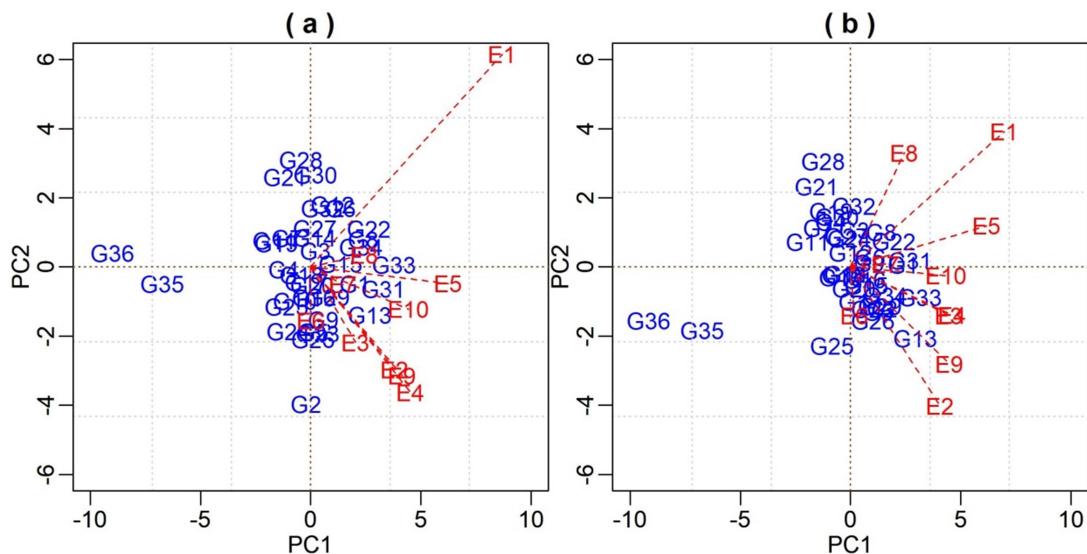


Figura 6. Biplots BGGE (a) e BGGEH (b) construídos com as médias *a posteriori* de escores genotípicos (G) e ambientais (E).

No biplot BGGEH, observa-se que uma avaliação baseada no ambiente médio seria menos indicada. Percebe-se também adaptabilidade de G21, G28 e G32 ao ambiente E8. Além disso, o vetor ambiental E2 é o segundo com a maior norma. Essas interpretações são discrepantes daquelas possíveis de serem realizadas no biplot BGGE. Diferenças em relação a interpretações de acordo com as posições de escores genotípicos também são observadas e podem levar a mudanças na classificação genotípica em relação à estabilidade.

Biplots GGE para modelos de efeitos fixos também são apresentados em materiais suplementares (Apêndice S5) para efeito de comparações. Na Figura S8a, o GGE biplot, obtido sem qualquer procedimento de dimensionamento preliminar para a matriz de G+GEI (escala ou ponderação), exibe padrão semelhante àquele observado para o modelo BGGE. Por outro lado, o GGE biplot (S8b) foi obtido a partir da padronização das médias das células pelo desvio padrão fenotípico em cada coluna, bem como pela ponderação pela a raiz da quadrada herdabilidade em cada ambiente. O biplot a partir dos dados “corrigidos” exibe um padrão mais semelhante aquele apresentado ao BGGEH. Yan (2014) ressalta a necessidade desse procedimento preliminar (escala, ponderação) para lidar com heterogeneidade na análise biplot, sendo esse mais um argumento em favor do método aqui proposto.

Discussão

Heterogeneidade de variância residual nos locais ocorrem rotineiramente em ensaios MET e pode acarretar estimativas ineficientes, caso não seja contabilizada na análise (SMITH; CULLIS; THOMPSON, 2001; CROSSA et al., 2006). Além disso, a presença de GEI acentuada nos dados é indicativa de que existe diferença de variância residual entre os ambientes (CROSSA; CORNELIUS, 1997; EDWARDS; JANNINK, 2006; CROSSA, 2012). Esse fato foi constatado aqui pelos resultados dos critérios de informação aplicados aos dados reais e mostram que métodos para modelar variâncias específicas em ambientes seriam mais adequados. Quando há heterogeneidade de variância residual entre ambientes, pesos diferentes devem ser assumidos em procedimentos de estimação dos parâmetros (WELHAM et al., 2010; RODRIGUES et al., 2015).

Implicações da suposição de variâncias residuais heterogêneas (ou homogêneas) podem ser observadas nas regiões de credibilidade bivariadas nos biplots, com mudanças rotacionais do sistema adicional de eixos e diferentes precisões, refletidas em

distintas amplitudes dos intervalos para os escores — principalmente ambientais (Figuras 2, 4, 5, S1, S2 e S3). Observa-se no biplot BGGEH, que em ambiente com maior variância haverá maior encolhimento das interações em relação à média geral e nos ambientes com menor variância a interação será mais preservada. Esse fato ficou evidente nos ambientes E1 e E8, dos dados reais (Figuras 4, 5 e S3). Isso pode ser observado também nos biplots referentes aos dados heterocedásticos simulados, com notável diferença entre os modelos, destacando a precisão desigual nos ambientes testados (Figura 2). Tais constatações nos levam a crer que a estimativa no modelo BGGE pode estar superestimando a interação e poderia levar a acreditar que o modelo com variâncias homogêneas é superior (maior variância de interação), mas o modelo heterogêneo claramente faz mais sentido.

Mudanças expressivas na configuração do padrão biplot “quem venceu onde” também podem ser visualizadas (Figuras S1 e S2). Diferenças de posicionamentos de alguns escores ambientais e genotípicos, bem como a rotação do sistema de eixos, afetam diretamente a interpretação da adaptabilidade e estabilidade. Alterações gráficas semelhantes (em função da estrutura de variâncias) foram observadas por Rodrigues et al. (2015), com o W-AMMI, por Silva et al. (2019), que modelaram a heterogeneidade de variâncias no modelo AMMI-Bayesiano e por Nuvunga et al (2015) que compararam o padrão do modelo GGE e o FA-SREG2.

Nossa abordagem é uma extensão da modelagem de Silva et al. (2019) ao modelo GGE com o objetivo de verificar o impacto das inferências sobre os parâmetros bilineares, bem como na representação gráfica GGE biplot. Trabalhos abordando o modelo GGE sob o enfoque bayesiano (OLIVEIRA et al., 2016; JARQUÍN et al., 2016; OMER; SINGH, 2017) não consideraram o cenário heterocedástico entre locais. AMMI e GGE são os mais populares modelos utilizados em análise MET e foram intensamente discutidos na literatura (YAN et al., 2007; GAUCH, 2006; GAUCH, 2013). Claro que a utilização de um ou outro método pode depender dos objetivos das análises. Contudo, na avaliação MET, o GGE oferece a flexibilidade de uma avaliação gráfica direta. Para o AMMI, as conclusões, em termos de recomendações específicas, devem levar em conta o rendimento nominal do genótipo no ambiente (GAUCH; ZOBEL, 1997), qual não pode ser acessado diretamente no biplot com as propriedades de produto interno.

A questão da heterogeneidade da variância residual em ambientes também não é levada em consideração na maioria das aplicações utilizando o procedimento padrão de estimativa de modelos lineares-bilineares de feitos fixos como AMMI e GGE. Embora

exista a possibilidade de tratar a heterocedasticidade nos dados a partir de algoritmos baseados em DVS, como o AMMI ponderado (W-AMMI) proposto por Rodrigues et al. (2015). Esses autores recomendam o W-AMMI como alternativa a análises MET realizadas em dois estágios, em que vetores de médias genotípicas e erros padrão são estimados em ensaios individuais (em um primeiro estágio) com a posterior análise conjunta por modelos mistos ponderados no segundo estágio (PIEPHO et al., 2012).

No GGE biplot, Yan (2014) recomenda a utilização de métodos de dimensionamento de dados como passo preliminar à obtenção da representação gráfica. Nesse sentido, as médias de células, em tabelas de dupla entrada, podem ser escaladas pelo desvio padrão-SD (ou erro padrão-ED) e ponderadas pela raiz quadrada da herdabilidade-h (ou pelo desvio padrão genotípico) em cada ambiente. Biplots referentes aos dados reais sem essas correções, bem como para médias SD-escaladas e h-ponderadas são apresentados em material suplementar (Figura S4). Como argumenta Yan (2014), a essência da escala e ponderação de dados é dar a diferentes ambientes, diferentes pesos na análise conjunta, sendo que esses pesos são expressos como o comprimento dos vetores ambientais no biplot GGE.

Mas estes procedimentos estão longe de ser um consenso, pois essa análise ponderada atribui menos pesos para ambientes que têm um quadrado médio residual mais elevado. Uma desvantagem da análise ponderada é que os pesos podem ser correlacionados com as respostas do rendimento no ambiente. Assim, pode ocorrer, por exemplo, que ambientes com rendimento elevado apresentem maior variância do erro e ambientes com baixos rendimentos apresentem variâncias de erro reduzidas, o que pode mascarar o verdadeiro desempenho de alguns genótipos em certos ambientes (CROSSA, 1990; SILVA et al., 2019). Além disso, as equações apresentadas para modelos multiplicativos fixos são válidas apenas para situações ideais, ou seja, quando dados completos e balanceados estão disponíveis. Em um cenário não balanceado, componentes de variância e herdabilidade devem ser estimados a partir de métodos adequados como, por exemplo, os modelos mistos (LITTELL, 1996; SPILKE; PIEPHO; HU, 2005; KELLY et al., 2007). Embora existam procedimentos de imputação de dados faltantes com base na DVS (YAN, 2013).

Modelos mistos se tornaram populares para análise MET, em especial a classe dos modelos mistos multiplicativos (MMM), também referidos como modelo Fatorial Analítico (FA). Eles são apontados como os melhores, dentre a classe de modelos mistos, para analisar ensaios de cultivares (KELLY et al., 2007). Modelos FA têm sido

considerados superiores a modelos de efeitos fixos em vários aspectos, incluindo as análises biplots regulares, em especial ao GGE biplot. Eles se destacam pela ampla flexibilidade para manipular dados ausentes, heterogeneidade entre ambientes, genótipos e correlações dentro dos ambientes, além de poder acomodar a dependência espacial entre ambientes (SMITH; CULLIS; THOMPSON, 2001; BEECK et al., 2010; CULLIS et al. 2010; CROSSA, 2012). Não obstante a essas vantagens, MMM são exigentes computacionalmente, além de, eventualmente, estimarem parâmetros fora do espaço paramétrico — o que é referido como casos Heyhood. Embora existam várias alternativas para contornar estas limitações, incluindo o algoritmo de matrizes esparsas ou *Average Information* modificado de Thompson et al (2003) e o PEX-EM (DIFFEY et al., 2017), a sua modelagem continua sendo um desafio.

O método bayesiano, por outro lado, tem se mostrado promissor na estimação de parâmetros relacionados à estabilidade e adaptabilidade genotípica em dados MET. Nuvunga et al. (2019) mostrou que sob o enfoque bayesiano é possível considerar modelos de dimensão completa na análise FA. Além disso, procedimentos de rotação de eixos não são necessários. Outra vantagem a ser destacada é que estimativas não são obtidas fora do espaço paramétrico. O método bayesiano também permite flexibilidade para incorporar estruturas mais complexas para as matrizes de variância-covariância (EDWARDS; JANNINK, 2006; ORELLANA et al., 2014). Em nossa abordagem, componentes de variância são consideradas de forma direta em análise de estagio único apenas estabelecendo hipóteses e informações *a priori*. Além disso, todas as dimensões do GGE podem ser estimadas sem maiores problemas. Essa também é uma limitação da abordagem ponderada proposta por Rodrigues et al. (2015). Eles reconhecem que, para o W-AMMI, o número de componentes deve ser decidido antes da execução do algoritmo.

Vale ressaltar que nossa análise gráfica foi baseada em propriedades inferenciais o que, geralmente, não é contemplado em análise de efeitos fixos e não são representadas, usualmente, em modelos FA. A versão bayesiana do FA-SERG (BFA) de Nuvunga et al. (2019), similar a nossa proposta, permite a incorporação de regiões de credibilidade bivariadas no biplot, garante indentificabilidade no processo de estimação e elimina a necessidade de rotacionalidade de cargas fatoriais, o que não é realizado diretamente na análise usual FA. Contudo, os próprios autores recomendam cautela quanto à interpretação da representação gráfica (biplot BFA) que deve ser realizada separadamente para genótipos e ambientes, pois suas respostas não possuem a mesma

escala e também não contam com as propriedades do produto interno. A nossa proposta, por sua vez, fornece um método direto e de fácil interpretação dos parâmetros. A incorporação de incerteza sobre os escores em biplots, no contexto frequentista, por sua vez, tem sido controversa na literatura (YANG et. al., 2009; YAN et. al., 2010; CROSSA et. al., 2011).

Diferenças entre os biplots BGGE e BGGEH, no cenário controlado heterocedástico, foram observadas e podem afetar as decisões em relação a descartes de genótipos, bem como de ambientes que apresentem alta colinearidade (Figura 2: c e d). Para dados com homogeneidade de variâncias, observaram-se menores diferenças em relação às amplitudes das regiões de credibilidade bivariadas e os biplots apresentaram praticamente o mesmo padrão, comportando também as mesmas inferências possíveis (Figura 2: a e b). Para os conjuntos de dados aqui simulados, esses resultados sugerem que o BGGEH é um método mais robusto de análise, já que contemplou bem ambas as estruturas de variâncias subjacentes.

Para todos os critérios de seleção de modelos utilizados, o BGGEH ajustou-se melhor aos dados. Como salientado por Piepho (1998), a utilidade de qualquer medida de estabilidade depende basicamente de quão bem o modelo se aproxima aos dados. Isso significa que a escolha da classe de modelos candidatos e a estimativa da componente de variância têm implicações diretas para a medida de estabilidade (HU; YAN; LI, 2014). Nesse sentido, o modelo BGGEH seria melhor alternativa para representar estabilidade de genótipos em ensaios METs do que a sua versão que considera homogeneidade de variâncias nos locais.

Os impactos da modelagem da heterogeneidade residual na estimativa pontual e teste de hipóteses de efeitos de genótipos já são reconhecidos e apresentados em vários trabalhos (CASANOVES; MACCHIAVELLI; BALZARINI, 2005; HU; SPILKE, 2011; HU; YAN; SHEN, 2013), sendo que a classificação da estabilidade do rendimento de genótipos é baseada nas estimativas dos parâmetros do modelo utilizado. Sendo assim, uma estimativa confiável dos parâmetros do modelo também é um pré-requisito importante para inferências válidas a respeito da análise de estabilidade do genótipo (HU; YAN; LI, 2014). Aqui mostramos que assumir homogeneidade de variâncias em situação de claro desvio dessa pressuposição, longe de ser uma prática inofensiva, pode acarretar interpretações equivocadas e influenciar a avaliação da estabilidade de genótipos em ensaios MET. Isso foi marcante principalmente na análise de dados reais.

A matriz G+GEI simulada com baixa dimensão e com subgrupos contrastantes em relação ao efeito da GEI pode ter influenciado alguma semelhança entre os biplots BGGE e BGGEH, já que os dois primeiros componentes principais explicaram praticamente todo o padrão existente nos dados. Particularmente, o padrão mais interessante foi explicado basicamente pelo primeiro PC. No entanto, esse exemplo simples foi suficiente para mostrar as principais diferenças entre os modelos, no que se refere a análise biplot e informações recuperada pelos valores singulares.

Para os dados reais observou-se um padrão mais complexo, embora os dois primeiros PCs também tenham captado grande parte da variação. Para este exemplo específico, observa-se mais nitidamente o efeito de encolhimento nas estimativas dos três primeiros valores singulares do BGGEH (Tabela 6), mas a partir daí observa-se comportamento contrário, e um menor decaimento de valores do BGGEH em relação ao BGGE. Para conjuntos com interações mais complexas seria esperado que mais eixos fossem necessários para explicar o padrão MET no modelo BGGEH, assim como relatado por Silva et al. (2019) para o modelo BAMMIH. Pelos aspectos aqui observados, acreditamos que o método é promissor e aponta novas perspectivas para análise de dados multiambientais.

Referências

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716-723, 1974.
- ALVARADO et al. META-R: A software to analyze data from multi-environment plant breeding trials. **The Crop Journal**, v. 8, n.5, p. 745-756, 2020.
- BAKER, R. J. Tests for crossover genotype-environmental interactions. **Canadian Journal of Plant Science**, v. 68, n. 2, p. 405-410, 1988.
- BALESTRE, M. et al. Yield stability and adaptability of maize hybrids based on GGE biplot analysis characteristics. **Crop Breeding and Applied Biotechnology**, v. 9, n. 3, p. 219-228, 2009.
- BATES, D. et al. Fitting Linear Mixed-Effects Models Using lme4. **Journal of Statistical Software**, v.67, n.1, p. 1-48, 2015.
- BEECK, C. P. et al. Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. **Genome**, v. 53, n.11, p. 992-1001, 2010.
- BRADU, D.; GABRIEL, K. R. The biplot as a diagnostic tool for models of two-way tables. **Technometrics**, v. 20, n. 1, p. 47–68, 1978.

- BURGUEÑO, J. et al. Using factor analytic models for joining environments and genotypes without crossover genotype \times environment interaction. **Crop Science**, v. 48, p. 1291-1305, 2008.
- CASANOVES, F.; MACCHIAVELLI, R.; BALZARINI, M. Error variation in multienvironment peanut trials: Within-trial spatial correlation and between-trial heterogeneity. **Crop Science**, v. 45, n.5, p. 1927-1933, 2005.
- CHEN, M. H.; SHAO, Q. M. Monte Carlo estimation of Bayesian credible and HPD intervals. **Journal of Computational and Graphical Statistics**, Alexandria, v. 8, n. 1, p. 69-92, 1999.
- COOPER, M.; DELACY, I. H. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. **Theoretical and Applied Genetics**, v. 88, n. 5, p. 561-572, 1994.
- CORNELIUS, P. L.; CROSSA, J.; SEYEDSADR, M. S. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In: KANG, M. S.; GAUCH, H. G. (Org.). **Genotype-by-environment interaction**. Boca Raton: CRC, 1996. p. 199-234.
- CORNELIUS, P. L.; SEYEDSADR, M. S. Estimation of general linear-bilinear models for two-way tables. **Journal of Statistical Computation and Simulation**, v. 58, n. 4, p. 287-322, 1997.
- COTES, J. M. et al. A Bayesian approach for assessing the stability of genotypes. **Crop Science**, v. 46, n. 6, p. 2654-2665, 2006.
- CROSSA, J. Statistical analyses of multilocation trials. **Advances in Agronomy**, v. 44, p. 55-85, 1990.
- 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.
- CROSSA, J. et al. Modeling genotype \times environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. **Crop Science**, v. 46, n. 4, p. 1722-1733, 2006.
- CROSSA, J. et al. Bayesian estimation of the additive main effects and multiplicative interaction model. **Crop Science**, v. 51, n. 4, p. 1458-1469, 2011.
- CROSSA, J. From genotype \times environment interaction to gene \times environment interaction. **Current Genomics**, v. 13, n. 3, p. 225-244, 2012.
- CULLIS, B. R. et al. Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. **Genome**, v. 53, n.11, p. 1002-1016, 2010.

- DELGADO, I. D. et al. Genotype by environment interaction and adaptability of photoperiod-sensitive biomass sorghum hybrids. **Bragantia**, v. 78, n. 4, p. 509-521, 2019.
- DENIS, J. B.; GOWER, J. C. Asymptotic covariances for parameters of biadditive models. **Utilitas Mathematica**, v. 46, p. 193-205, 1994.
- DENIS, J.-B.; GOWER, J. C. Asymptotic confidence regions for biadditive models: Interpreting genotype-environment interactions. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, v. 45, n. 4, p. 479-493, 1996.
- DIAS, C. T. S.; KRZANOWSKI, W. J. Model selection and cross validation in additive main effect and multiplicative interaction models. **Crop Science**, v. 43, n. 3, p. 865-873, 2003.
- DIFFEY, S. M., et al. A new REML (parameter expanded) EM algorithm for linear mixed models. **Australian & New Zealand Journal of Statistics**, v.59, n. 4, p. 433-448, 2017.
- EBERHART, S. A.; RUSSELL, W. A. Stability parameters for comparing varieties. **Crop Science**, v. 6, n.1, p. 36-40, 1966.
- EDWARDS, J. W.; JANNINK, J. L. Bayesian modeling of heterogeneous error and genotype environment interaction variances. **Crop Science**, v. 46, n.2, p. 820-833, 2006.
- FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, v. 14, p., n.6, 742-754, 1963.
- FORKMAN, J.; PIEPHO, H. P. Parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models. **Biometrics**, v. 70, n. 3, p. 639-647, 2014.
- GABRIEL, K. R. The biplot graphic display of matrices with application to principal components analysis. **Biometrika**, v. 58, n. 3, p. 453-467, 1971.
- GAUCH, H. G. Model selections and validation for yield trials with interactions. **Biometrics**, v. 44, p. 705-715, 1988.
- GAUCH, H. G. **Statistical analysis of regional yield trials: AMMI analysis of factorial designs**. New York: Elsevier, 1992. 278 p.
- GAUCH, H. G. Statistical analysis of yield trials by AMMI and GGE. **Crop Science**, v.46, n.4, p.1488-1500, 2006.
- GAUCH, H. G. A Simple Protocol for AMMI Analysis of Yield Trials. **Crop Science**, v.53, n. 5, p.1860-1869, 2013.
- GAUCH, H. G.; PIEPHO, H. P.; ANNICHIAIRICO, P. Statistical analysis of yield trials by AMMI and GGE: further considerations. **Crop Science**, v.48, n3, p.866-889, 2008.

- GAUCH, H. G.; ZOBEL, R. W. Identifying mega-environments and targeting genotypes. **Crop Science**, v.37, n.2, p. 311-326, 1997.
- GOGEL, B.; SMITH, A.; CULLIS, B. Comparison of a one-and two-stage mixed model analysis of Australia's National Variety Trial Southern Region wheat data. **Euphytica**, v. 214, n. 2, p. 44, 2018.
- GUILLY, S. et al. Analysis of multienvironment trials (MET) in the sugarcane breeding program of Réunion Island. **Euphytica**, v. 213, n. 9, p. 213, 2017.
- 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.
- HERNÁNDEZ, M. V. et al. Modeling Genotype \times Environment Interaction Using a Factor Analytic Model of On-Farm Wheat Trials in the Yaqui Valley of Mexico. **Agronomy Journal**, v. 111, n. 6, p. 2647-2657, 2019.
- HU, X.; SPILKE, J. Variance–covariance structure and its influence on variety assessment in regional crop trials. **Field Crops Research**, v. 120, n. 1, p. 1-8, 2011.
- HU, X.; YAN, S.; LI, S. The influence of error variance variation on analysis of genotype stability in multi-environment trials. **Field Crops Research**, v.156, p. 84-90, 2014.
- HU, X.; YAN, S.; SHEN, K. Heterogeneity of error variance and its influence on genotype comparison in multi-location trials. **Field Crops Research**, v.149, 322-328, 2013.
- 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.
- JOSSE, J. et al. Another look at Bayesian analysis of AMMI models for genotype-environment data. **Journal of Agricultural, Biological and Environmental Statistics**, v. 19, n. 2, p. 240-257, 2014.
- KEMPTON, R. A. The use of biplots in interpreting variety by environment interactions. **Journal of Agricultural Science**, v. 103, n. 1, p. 123-135, 1984.
- KELLY, A. M. et al. The accuracy of varietal selection using factor analytic models for multienvironment plant breeding trials. **Crop Science**, v. 47, n. 3, p. 1063-1070, 2007.
- KRZANOWSKI, W. Cross-validation in principal component analysis. **Biometrics**, v. 43, n. 3, p. 575-584, 1987.
- LAVORANTI, O. J.; DIAS, C. T. S.; KRAZNOWSKI, W. J. Phenotypic stability via AMMI model with bootstrap re-sampling. **Pesquisa Florestal Brasileira**, n. 54, p. 45-52, 2007.

- LIU, G. **Bayesian computations for general linear-bilinear models.** 2001. 169 p. Thesis (Doctor of Philosophy) - University of Kentucky, Lexington, 2001.
- LITTELL, R. C. Analysis of unbalanced mixed model data: Traditional ANOVA versus contemporary methods. **Conference on Applied Statistics in Agriculture**, 1996. Disponível em: <<https://newprairiepress.org/cgi/viewcontent.cgi?article=1318&context=agstatconference>>. Acesso em: 02 de janeiro de 2021.
- MAIA, M. C. C. et al. AMMI-Bootstrap no estudo da interação genótipos por ambientes em soja. **Revista de Matemática e Estatística**, v. 24, n. 3, p. 7-24, 2006.
- NUVUNGA, J. J. et al. Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. **Genetics and Molecular Research**, v. 14, n.4, p. 14262-14278, 2015.
- NUVUNGA, J. J. et al. Bayesian factor analytic model: An approach in multiple environment trials. **PloS One**, v. 14, n. 8, p. e0220290, 2019.
- OAKLEY, H. et al. Joint modeling of additive and non-additive genetic line effects in single field trials. **Theoretical and Applied Genetics**, v. 113, n. 5, p. 809-819, 2006.
- 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 multi-environments trials. **Genetics and Molecular Research**, v. 15, n. 2, p. 1-21, 2016.
- OMER, S. O.; SINGH, M. Comparing Bayesian and Frequentist Approaches for GGE Bi-plot Analysis in Multi-Environment Trials in Sorghum. **European Journal of Experimental Biology**, v. 7, n. 6, p. 40, 2017.
- ORELLANA, M.; EDWARDS, J.; CARRIQUIY, A. Heterogeneous variances in multi-environment yield trials for corn hybrids. **Crop Science**, v. 54, n. 3, p. 1048-1056, 2014.
- PEREZ-ELIZALDE, S.; JARQUIN, D.; CROSSA, J. A general Bayesian estimation method of linear-bilinear models applied to plant breeding trials with genotype \times environment interaction. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 17, n. 1, p. 15-37, 2012.
- PIEPHO, H. P. Analyzing genotype-environment data by mixed models with multiplicative terms. **Biometrics**, v. 53, n. 2, p. 761-766, 1997.
- PIEPHO, H. P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical Applied Genetics**, v. 97, p. 195-201, 1998.
- PIEPHO, H. P.; MÖHRING, J. Selection in Cultivar Trials—Is It Ignorable?. **Crop Science**, v. 46, n. 1, p. 192-201, 2006.
- PIEPHO, H. P et al. Blup for phenotypic selection in plant breeding and variety testing. **Euphytica**, v. 161, n. 1-2, p. 209-228, 2008.

- PIEPHO, H. P. et al. A stage-wise approach for the analysis of multi-environment trials. **Biometrical Journal**, v. 54, n. 6, p. 844-860, 2012.
- R CORE TEAM. **R**: A Language and Environment for Statistical Computing. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org/>>.
- RAFTERY, A. E.; LEWIS, S. How many iterations in the Gibbs sampler? In: BERNARDO, J. M. et al. (Ed.). **Bayesian Statistics**. Oxford: Oxford University, 1992. p. 763-773.
- 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.
- RODRIGUES, P.C. et al. A Weighted AMMI Algorithm to Study Genotype-by-Environment Interaction and QTL-by-Environment Interaction. **Crop Science**, v.54, n. 4, p.1555-1570, 2015.
- 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.
- SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v. 6, n. 2, p. 461- 464, 1978.
- 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.
- SINGH, M.; CECCARELLI, S.; GRANDO, S. Genotype x environment interaction of crossover type: detecting its presence and estimating the crossover point. **Theoretical and Applied Genetics**, v. 99, n. 6, p. 988-995, 1999.
- SMITH, A. B.; CULLIS, B. R. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. **Euphytica**, v. 214, n.8, p. 143, 2018.
- SMITH, A.; CULLIS, B.; THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend, **Biometrics**, v. 57, p. 1138-1147, 2001.
- SMITH, A.; CULLIS, B.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **Journal of Agricultural Science**, v. 143, n. 2, p. 449-462, 2005.
- SMITH, A. B. et al. Factor analytic mixed models for the provision of grower information fromnational crop variety testing programs. **Theoretical and Applied Genetics**, v. 128, n. 1, p. 55-72, 2015.
- SPILKE, J.; PIEPHO, H. P.; HU, X. Analysis of unbalanced data by mixed linear models using the MIXED procedure of the SAS system. **Journal of Agronomy and Crop Science**, v.191, n.1, p.47-54, 2005.

- STEFANOVA, K. T.; BUIRCHELL, B. Multiplicative mixed models for genetic gain assessment in lupin breeding. **Crop Science**, v. 50, n. 3, p. 880-891, 2010.
- VARGAS, M. et al. Analysis and interpretation of interactions in agricultural research. **Agronomy Journal**, v. 107, n. 2, p. 748-762, 2015.
- 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.
- WELHAM, S. J. et al. A comparison of analysis methods for late-stage variety evaluation trials. **Australian & New Zealand Journal of Statistics**, v. 52, n. 2, p. 125-149, 2010.
- YAN, W. Biplot analysis of incomplete two-way data. **Crop Science**, v. 53, n. 1, p. 48-57, 2013.
- 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.
- YAN, W. et al. Two types of GGE biplot for analyzing multi-environment trial data. **Crop Science**, v. 41, p. 656-663, 2001.
- YAN, W. et al. GGE biplot vs. AMMI analysis of genotype-by-environment data. **Crop Science**, v. 47, n. 2, p. 643-653, 2007.
- 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, PL Cornelius, and J. Burgueño in 2009, v. 49, n. 5: 1564-1576. **Crop Science**, v. 50, n. 4, p. 1121-1123, 2010.
- YAN, W.; KANG, M.S. **GGE Biplot Analysis: A graphical tool for breeders, geneticists, and agronomists**. Boca Raton: CRC Press, 2003. 286 p.
- YAN, W.; TINKER, A. Biplot analysis of multi environment trial data: principles and applications. **Canadian Journal of Plant Science**, v. 86, n. 3, p. 623-645, 2006.
- YANG, R. C. et al. J. Biplot analysis of genotype environment interaction: proceed with caution. **Crop Science**. v. 49, n. 5, p. 1564-1576, 2009.

Materiais suplementares

Apêndice S1. Distribuições condicionais completas *a posteriori*

Apêndice S2. Exemplificação do algoritmo amostral

Apêndice S3. Gráficos de traços e densidades das componentes da variância residual para dados reais

Apêndice S4. Resumos pontuais e intervalares *a posteriori* dos valores singulares para dados simulados

Apêndice S5. Representações biplots para dados simulados e dados reais

Apêndice S1

Distribuições condicionais completas *a posteriori*

Distribuição *a posteriori* para β condicionada nos demais parâmetros

$$\begin{aligned} P(\beta | \dots) &\propto \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_y)^\top \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right\} P(\beta) \\ P(\beta | \dots) &\propto \exp\left\{-\frac{1}{2}(\mathbf{M} - \mathbf{X}_1\beta)^\top \mathbf{V}^{-1}(\mathbf{M} - \mathbf{X}_1\beta)\right\} \exp\left\{-\frac{1}{2\sigma_\beta^2}\beta^\top \beta\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\left(\mathbf{M} - \mathbf{X}_1\beta\right)^\top \mathbf{V}^{-1}\left(\mathbf{M} - \mathbf{X}_1\beta\right) + \beta^\top \mathbf{I} \frac{1}{\sigma_\beta^2} \beta\right]\right\} \end{aligned}$$

em que $\mathbf{M} = \mathbf{y} - \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\alpha_k) \mathbf{X}_2 \gamma_k$.

Resolvendo a expressão quadrática dentro dos colchetes, tem-se:

$$* \quad \mathbf{M}^\top \mathbf{V}^{-1} \mathbf{M} - \mathbf{M}^\top \mathbf{V}^{-1} \mathbf{X}_1 \beta - \beta^\top \mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{M} + \beta^\top \left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 + \mathbf{I} \frac{1}{\sigma_\beta^2} \right) \beta,$$

então

$$\begin{aligned} P(\beta | \dots) &\propto \exp\left\{-\frac{1}{2}\left[-2\beta^\top \mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{M} + \beta^\top \left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 + \mathbf{I} \frac{1}{\sigma_\beta^2} \right) \beta\right]\right\} \\ &\propto \exp\left\{\begin{aligned} &-\frac{1}{2}\left[\beta - \left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 + \mathbf{I} \frac{1}{\sigma_\beta^2} \right)^{-1} \mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{M}\right]^\top \left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 + \mathbf{I} \frac{1}{\sigma_\beta^2} \right)^{-1} \times \\ &\times \left[\beta - \left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 + \mathbf{I} \frac{1}{\sigma_\beta^2} \right)^{-1} \mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{M} \right] \end{aligned}\right\} \end{aligned}$$

ou seja, $\beta | \dots \sim N\left[\left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 + \mathbf{I} \frac{1}{\sigma_\beta^2} \right)^{-1} \mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{M}, \left(\mathbf{X}_1^\top \mathbf{V}^{-1} \mathbf{X}_1 + \mathbf{I} \frac{1}{\sigma_\beta^2} \right)^{-1}\right]$

assumindo $\sigma_\beta^2 = 10^8$.

Para os parâmetros bilineares (considerando-se $k = 1, \dots, t$) as distribuições condicionais *a posteriori* são dadas por:

i) Distribuição *a posteriori* para λ_k condicionada nos demais parâmetros

$$\begin{aligned}
P(\lambda_k | \dots) &\propto \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_y)^\top \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) \right\} P(\lambda_k) \\
P(\lambda_k | \dots) &\propto \exp \left\{ -\frac{1}{2} (\mathbf{M}_{k'} - \lambda_k \boldsymbol{\Phi}_k)^\top \mathbf{V}^{-1} (\mathbf{M}_{k'} - \lambda_k \boldsymbol{\Phi}_k) \right\} \times \exp \left\{ -\frac{1}{\sigma_{\lambda_k}^2} \lambda_k^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\mathbf{M}_{k'}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} - \lambda_k \mathbf{M}_{k'}^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k - \lambda_k \boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \mathbf{M}_{k'} - \lambda_k^2 \boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k + \frac{1}{\sigma_{\lambda_k}^2} \lambda_k^2 \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[-2\lambda_k \boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \mathbf{M}_{k'} - \lambda_k^2 \left(\boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k + \frac{1}{\sigma_{\lambda_k}^2} \right) \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\begin{array}{l} \left[\lambda_k - \left(\boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k + \frac{1}{\sigma_{\lambda_k}^2} \right)^{-1} \boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top \left(\boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k + \frac{1}{\sigma_{\lambda_k}^2} \right) \times \\ \times \left[\lambda_k - \left(\boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k + \frac{1}{\sigma_{\lambda_k}^2} \right)^{-1} \boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right] \end{array} \right] \right\}.
\end{aligned}$$

Então

$$\lambda_k | \dots \sim N^+ \left[\left(\boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k + \frac{1}{\sigma_{\lambda_k}^2} \right)^{-1} \boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \mathbf{M}_{k'}, \quad \left(\boldsymbol{\Phi}_k^\top \mathbf{V}^{-1} \boldsymbol{\Phi}_k + \frac{1}{\sigma_{\lambda_k}^2} \right)^{-1} \right],$$

sendo $\mathbf{M}_{k'} = \mathbf{y} - \mathbf{X}_1 \boldsymbol{\beta} - \sum_{k' \neq k}^t \lambda_{k'} \text{diag}(\mathbf{Z} \mathbf{a}_{k'}) \mathbf{X}_2 \boldsymbol{\gamma}_{k'}$, $\boldsymbol{\Phi}_k = \text{diag}(\mathbf{Z} \mathbf{a}_k) \mathbf{X}_2 \boldsymbol{\gamma}_k$ e $\sigma_{\lambda_k}^2 = 10^8$.

ii) Distribuição *a posteriori* para \mathbf{a}_k condicionada nos demais parâmetros

$$\begin{aligned}
P(\mathbf{a}_k | \dots) &\propto \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_y)^\top \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) \right\} P(\mathbf{a}_k) \\
P(\mathbf{a}_k | \dots) &\propto \exp \left\{ -\frac{1}{2} (\mathbf{M}_{k'} - \boldsymbol{\Lambda}_{\mathbf{a}_k} \mathbf{a}_k)^\top \mathbf{V}^{-1} (\mathbf{M}_{k'} - \boldsymbol{\Lambda}_{\mathbf{a}_k} \mathbf{a}_k) \right\}
\end{aligned}$$

sendo $\boldsymbol{\Lambda}_{\mathbf{a}_k} = \lambda_k \text{diag}(\mathbf{X}_2 \boldsymbol{\gamma}_k) \mathbf{Z}$. Trabalhando a forma quadrática na exponencial, tem-se:

$$* \quad (\mathbf{M}_{k'} - \boldsymbol{\Lambda}_{\mathbf{a}_k} \mathbf{a}_k)^\top \mathbf{V}^{-1} (\mathbf{M}_{k'} - \boldsymbol{\Lambda}_{\mathbf{a}_k} \mathbf{a}_k) = \mathbf{M}_{k'}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} - \mathbf{M}_{k'}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \mathbf{a}_k - \mathbf{a}_k^\top \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} + \\ + \mathbf{a}_k^\top \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \mathbf{a}_k.$$

Então

$$P(\mathbf{a}_k | ...) \propto \exp \left\{ -\frac{1}{2} \left[-2 \mathbf{a}_k^\top \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} + \mathbf{a}_k^\top \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \mathbf{a}_k \right] \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left[\mathbf{a}_k - (\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k})^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top \left(\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \right) \times \right. \\ \left. \times \left[\mathbf{a}_k - (\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k})^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right] \right\},$$

$$\text{ou seja } \mathbf{a}_k | ... \sim N \left[(\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k})^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'}, (\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k})^{-1} \right].$$

iii) Distribuição *a posteriori* para $\boldsymbol{\gamma}_k$ condicionada nos demais parâmetros

$$P(\boldsymbol{\gamma}_k | ...) \propto \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_y)^\top \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) \right\} P(\boldsymbol{\gamma}_k)$$

$$P(\boldsymbol{\gamma}_k | ...) \propto \exp \left\{ -\frac{1}{2} (\mathbf{M}_{k'} - \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \boldsymbol{\gamma}_k)^\top \mathbf{V}^{-1} (\mathbf{M}_{k'} - \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \boldsymbol{\gamma}_k) \right\}$$

Desenvolvendo algebraicamente a forma quadrática tem-se

$$* \quad (\mathbf{M}_{k'} - \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \boldsymbol{\gamma}_k)^\top \mathbf{V}^{-1} (\mathbf{M}_{k'} - \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \boldsymbol{\gamma}_k) = \mathbf{M}_{k'}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} - \mathbf{M}_{k'}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \boldsymbol{\gamma}_k - \boldsymbol{\gamma}_k^\top \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} + \\ + \boldsymbol{\gamma}_k^\top \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \boldsymbol{\gamma}_k$$

em que $\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} = \lambda_k \text{diag}(\mathbf{Z}\mathbf{a}_k) \mathbf{X}_2$. Então podemos reescrever $P(\boldsymbol{\gamma}_k | ...)$ como

$$P(\boldsymbol{\gamma}_k | ...) \propto \exp \left\{ -\frac{1}{2} \left[-2 \boldsymbol{\gamma}_k^\top \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} + \boldsymbol{\gamma}_k^\top \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \boldsymbol{\gamma}_k \right] \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\gamma}_k - (\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k})^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top \left(\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k} \right) \times \right. \\ \left. \times \left[\boldsymbol{\gamma}_k - (\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k})^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right] \right\},$$

$$\text{de forma que } \boldsymbol{\gamma}_k | ... \sim N \left[(\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k})^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'}, (\boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\boldsymbol{\gamma}_k})^{-1} \right].$$

Distribuição *a posteriori* condicionada nos demais parâmetros para $\sigma_{e_j}^2$

Conforme especificado, \mathbf{V} é uma matriz diagonal composta pelas componentes $\sigma_{e_j}^2$, em que $j = 1, \dots, c$, denota o j -ésimo ambiente. A distribuição *a posteriori* para $\sigma_{e_j}^2$, condicionada nos demais parâmetros do modelo, pode ser representada por:

$$P(\sigma_{e_j}^2 | \dots) \propto (\sigma_{e_j}^2)^{-\frac{n_{e_j}}{2}} \exp \left\{ -\frac{1}{2\sigma_{e_j}^2} (\mathbf{y}_j - \boldsymbol{\mu}_{e_j})^\top (\mathbf{y}_j - \boldsymbol{\mu}_{e_j}) \right\} (\sigma_{e_j}^2)^{-1}$$

que é a equação de uma distribuição qui-quadrado invertida com parâmetro de escala $(\mathbf{y}_j - \boldsymbol{\mu}_{e_j})^\top (\mathbf{y}_j - \boldsymbol{\mu}_{e_j}) / n_{e_j}$ e graus de liberdade n_{e_j} , sendo denotada por

$$\sigma_{e_j}^2 | \dots \sim \text{Inv-Esc} - \chi^2 \left[n_{e_j}, \frac{(\mathbf{y}_j - \boldsymbol{\mu}_{e_j})^\top (\mathbf{y}_j - \boldsymbol{\mu}_{e_j})}{n_{e_j}} \right].$$

Um dos pontos fundamentais no processo de amostragem é que as distribuições condicionais *a posteriori* para os vetores singulares devem ser obtidas no subespaço corrigido de dimensão $(p-s)$ sendo $(0 \leq s \leq k-1)$ em que $(k=1, \dots, t)$ e $p=r$ (para genótipos) ou $p=c$ (para ambientes). Aqui consideramos, sem perda de generalidade, que $r > c$, e, portanto, $c = t$.

A distribuição condicional para o vetor singular genotípico já foi obtida em \mathbb{R}^r e é dada por:

$$\mathbf{a}_k | \dots \sim N \left[\left(\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \right)^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'}, \left(\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \right)^{-1} \right],$$

sendo sua equação proporcional a

$$\propto \exp \left\{ -\frac{1}{2} \left[\mathbf{a}_k - \left(\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \right)^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top \left(\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \right) \times \left[\mathbf{a}_k - \left(\boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k} \right)^{-1} \boldsymbol{\Lambda}_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right] \right\}.$$

Desenvolvendo a forma quadrática entre chaves obtém-se a seguinte expressão:

$$\begin{aligned} & \mathbf{a}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \mathbf{a}_k - \mathbf{a}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} - \left[(\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top \\ & (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \mathbf{a}_k + \left[(\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k} (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \end{aligned}$$

O vetor singular \mathbf{a}_k é distribuído em um subespaço restrito na dimensão r , pois deve ser ortogonal aos demais s vetores $\mathbf{a}_{k'}$ ($k' < k$). Para cada k ($k=1, \dots, t$), a amostragem é realizada por uma variável auxiliar $\tilde{\mathbf{a}}_k = \mathbf{H}_k^\top \mathbf{a}_k$ no subespaço $(r - s)$ em que \mathbf{H}_k é uma matriz tal que suas colunas formam uma base ortonormal. Desde que \mathbf{H}_k é ortogonal tem-se que $\mathbf{a}_k = \mathbf{H}_k \mathbf{H}_k^\top \mathbf{a}_k$ e $\tilde{\mathbf{a}}_k^\top = (\mathbf{H}_k^\top \mathbf{a}_k)^\top$, de forma que:

$$\begin{aligned} \mathbf{a}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \mathbf{a}_k &= \mathbf{a}_k^\top \mathbf{H}_k \mathbf{H}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \mathbf{H}_k \mathbf{H}_k^\top \mathbf{a}_k = (\mathbf{H}_k^\top \mathbf{a}_k)^\top \mathbf{H}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \times \\ &\times \mathbf{H}_k \mathbf{H}_k^\top \mathbf{a}_k = \tilde{\mathbf{a}}_k^\top \mathbf{H}_k^\top \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k} \mathbf{H}_k \tilde{\mathbf{a}}_k. \end{aligned}$$

A expressão $\mathbf{a}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'}$ pode ser reescrita como:

$$\begin{aligned} \mathbf{a}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} &= \mathbf{a}_k^\top \mathbf{H}_k \mathbf{H}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \times \\ &\times \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} = \tilde{\mathbf{a}}_k^\top \mathbf{H}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \mathbf{H}_k \mathbf{H}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'}. \end{aligned}$$

Além disso, $\left[(\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \mathbf{a}_k$, também pode ser reescrita da seguinte forma:

$$\begin{aligned} \left[(\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \mathbf{a}_k &= \left[(\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top \times \\ &\times \mathbf{H}_k \mathbf{H}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \mathbf{H}_k \mathbf{H}_k^\top \mathbf{a}_k = \left[(\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top \times \\ &\times \mathbf{H}_k \mathbf{H}_k^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) \mathbf{H}_k \tilde{\mathbf{a}}_k. \end{aligned}$$

Por fim, podemos reescrever

$$\left[(\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'} \right]^\top (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k}) (\Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \Lambda_{\mathbf{a}_k})^{-1} \Lambda_{\mathbf{a}_k}^\top \mathbf{V}^{-1} \mathbf{M}_{k'}$$

como:

$$\left[\left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top V^{-1} M_{k'} \right]^\top \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right) \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top V^{-1} M_{k'} =$$

$$\left[\left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top V^{-1} M_{k'} \right]^\top H_k H_k^\top \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right) H_k H_k^\top \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top V^{-1} M_{k'}$$

e desta forma, tem-se:

$$\tilde{\alpha}_k^\top H_k^\top \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right) H_k \tilde{\alpha}_k - \tilde{\alpha}_k^\top H_k^\top \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right) \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top V^{-1} M_{k'} -$$

$$- \left[\left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top V^{-1} M_{k'} \right]^\top \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right) H_k \tilde{\alpha}_k +$$

$$+ \left[\left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top V^{-1} M_{k'} \right]^\top H_k H_k^\top \Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} H_k H_k^\top \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top V^{-1} M_{k'}$$

Assim, a distribuição *a posteriori* condicionada nos demais parâmetros para o vetor $\tilde{\alpha}_k$ no subespaço corrigido ($r - s$) com $0 \leq s \leq t - 1$ é dada por:

$$\tilde{\alpha}_k | \dots \sim N(\mathbf{m}_{\alpha_k}, \mathbf{W}_{\alpha_k})$$

$$\text{com } \mathbf{m}_{\alpha_k} = \Delta_{\alpha_k} \left(\Delta_{\alpha_k}^\top \Delta_{\alpha_k} \right)^{-\frac{1}{2}}, \quad \Delta_{\alpha_k} = \left[H_k^\top \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right)^{-1} \Lambda_{\alpha_k}^\top V^{-1} M_{k'} \right] \text{ e}$$

$$\mathbf{W}_{\alpha_k} = \left\{ \left[\left(\Delta_{\alpha_k}^\top \Delta_{\alpha_k} \right)^{-\frac{1}{2}} \right]^\top H_k^\top \left(\Lambda_{\alpha_k}^\top V^{-1} \Lambda_{\alpha_k} \right) H_k \overline{m}_{\alpha_k} \left(\Delta_{\alpha_k}^\top \Delta_{\alpha_k} \right)^{-\frac{1}{2}} \right\}^{-1},$$

sendo $\overline{m}_{\alpha_k} = \Delta_{\alpha_k}^\top \Delta_{\alpha_k}$, em que $\left(\Delta_{\alpha_k}^\top \Delta_{\alpha_k} \right)^{1/2}$ é uma constante normalizadora.

Desta forma, é possível amostrar $\tilde{\alpha}_k$ em \mathbb{R}^{r-s} e depois obter o vetor α_k no subespaço correto $S_{(r-s)}^r$ em \mathbb{R}^r por meio da transformação inversa $\alpha_k = H_k \tilde{\alpha}_k$.

Com a mesma argumentação obtém-se a condicional *a posteriori* para $\tilde{\gamma}_k$, dado os demais parâmetros, no subespaço corrigido ($c - s$), sendo dada por:

$$\tilde{\gamma}_k | \dots \sim N(\mathbf{m}_{\gamma_k}, \mathbf{W}_{\gamma_k})$$

em que $\mathbf{m}_{\gamma_k} = \Delta_{\gamma_k} \left(\Delta_{\gamma_k}^\top \Delta_{\gamma_k} \right)^{-\frac{1}{2}}$, com $\Delta_{\gamma_k} = \left[D_k^\top \left(\Lambda_{\gamma_k}^\top V^{-1} \Lambda_{\gamma_k} \right)^{-1} \Lambda_{\gamma_k}^\top V^{-1} M_{k'} \right]$ sendo D_k a

respectiva matriz de transformação ortogonal. Por sua vez,

$$\mathbf{W}_{\gamma_k} = \left\{ \left[\left(\Delta_{\gamma_k}^\top \Delta_{\gamma_k} \right)^{-\frac{1}{2}} \right]^\top D_k^\top \left(\Lambda_{\gamma_k}^\top V^{-1} \Lambda_{\gamma_k} \right) D_k \overline{m}_{\gamma_k} \left(\Delta_{\gamma_k}^\top \Delta_{\gamma_k} \right)^{-\frac{1}{2}} \right\}^{-1} \text{ e } \overline{m}_{\gamma_k} = \Delta_{\gamma_k}^\top \Delta_{\gamma_k}.$$

Apêndice S2

Exemplificação do algoritmo amostral

O processo de amostragem iterativa foi conduzido a partir dos seguintes passos:

A1 - Atribuir valores iniciais aos parâmetros: $\Phi^0 = [\beta^0, \lambda^0, \alpha^0, \gamma^0, (\sigma_{e_j}^2)^0], j = 1, \dots, c$.

A2- A partir desses valores iniciais, a l -ésima iteração pode ser obtida da seguinte forma:

- a) Gerar $\beta^l | \lambda^{l-1}, \alpha^{l-1}, \gamma^{l-1}, (\sigma_{e_j}^2)^{l-1}$ a partir da condicional *a posteriori* (equação 4);
- b) Gerar a l -ésima observação dos parâmetros bilineares por meio da sequência d1), d2) e d3) abaixo, para $k = 1, 2, \dots, t$:
- d1) Gerar $\lambda_k^l | \beta^l, \alpha^{l-1}, \gamma^{l-1}, (\sigma_{e_j}^2)^{l-1}$ a partir da condicional *a posteriori* (equação 5);
- d2) Gerar $(\alpha_k)^l | \beta^l, \lambda_k^l, \gamma_k^{l-1}, (\sigma_{e_j}^2)^{l-1}$:

 - d2-i) Gerar $(\tilde{\alpha}_k)^l$ da condicional *a posteriori* (equação 9);
 - d2-ii) obter $(\alpha_k)^l = \mathbf{H}_k (\tilde{\alpha}_k)^l$.

- d3) Gerar $(\gamma_k)^l | \beta^l, \lambda_k^l, \alpha_k^l, (\sigma_{e_j}^2)^{l-1}$:

 - d3-i) Gerar $(\tilde{\gamma}_k)^l$ da condicional *a posteriori* (equação 10);
 - d3-ii) obter $(\gamma_k)^l = \mathbf{D}_k (\tilde{\gamma}_k)^l$.

- e) Gerar $(\sigma_{e_j}^2)^l | \beta^l, \lambda^l, \alpha^l, \gamma^l$ com $j = 1, \dots, c$ (c é o número de ambientes ou locais) a partir da condicional *a posteriori* (equação 8).

Retornar a A2 mudando contador para $l + 1$ e repetir essa sequência de passos iterativamente até a convergência.

Apêndice S3

Gráficos de traços e densidades das componentes da variância residual para dados reais

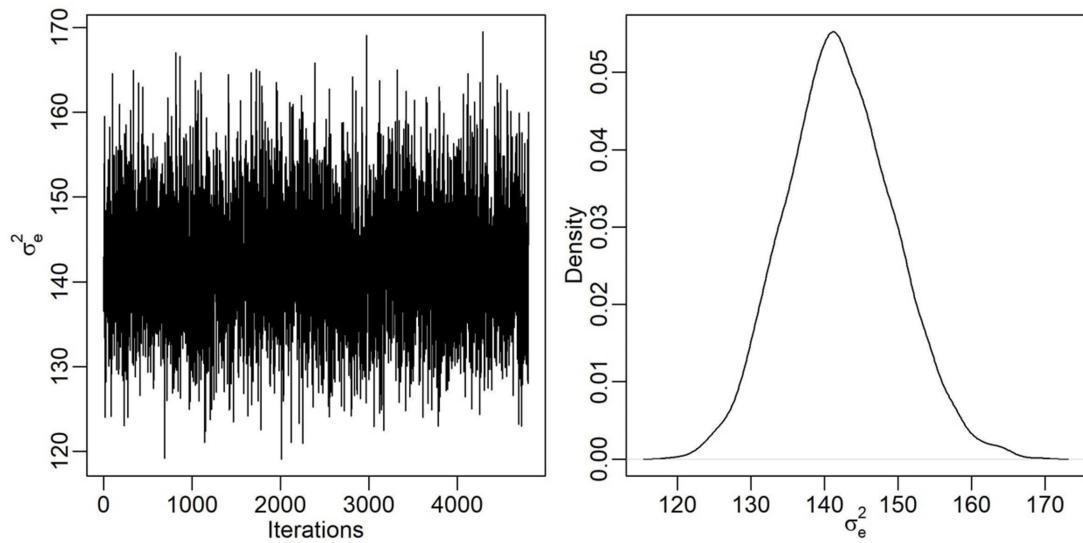


Figura S1. Gráfico de traços e densidade da distribuição *a posteriori* da variância residual para o modelo BGGE.

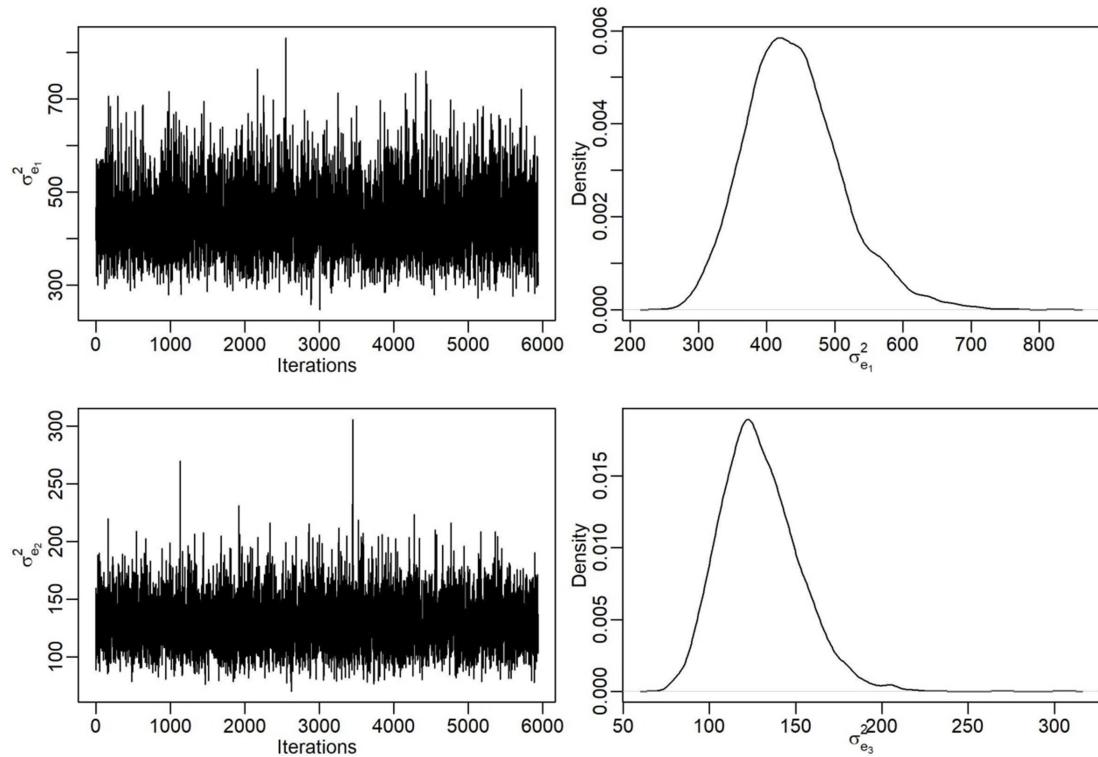


Figura S2. Gráficos de traços e densidades das distribuições *a posteriori* das componentes da variância residual relacionadas aos ambientes E1 e E2 para o modelo BGGEH.

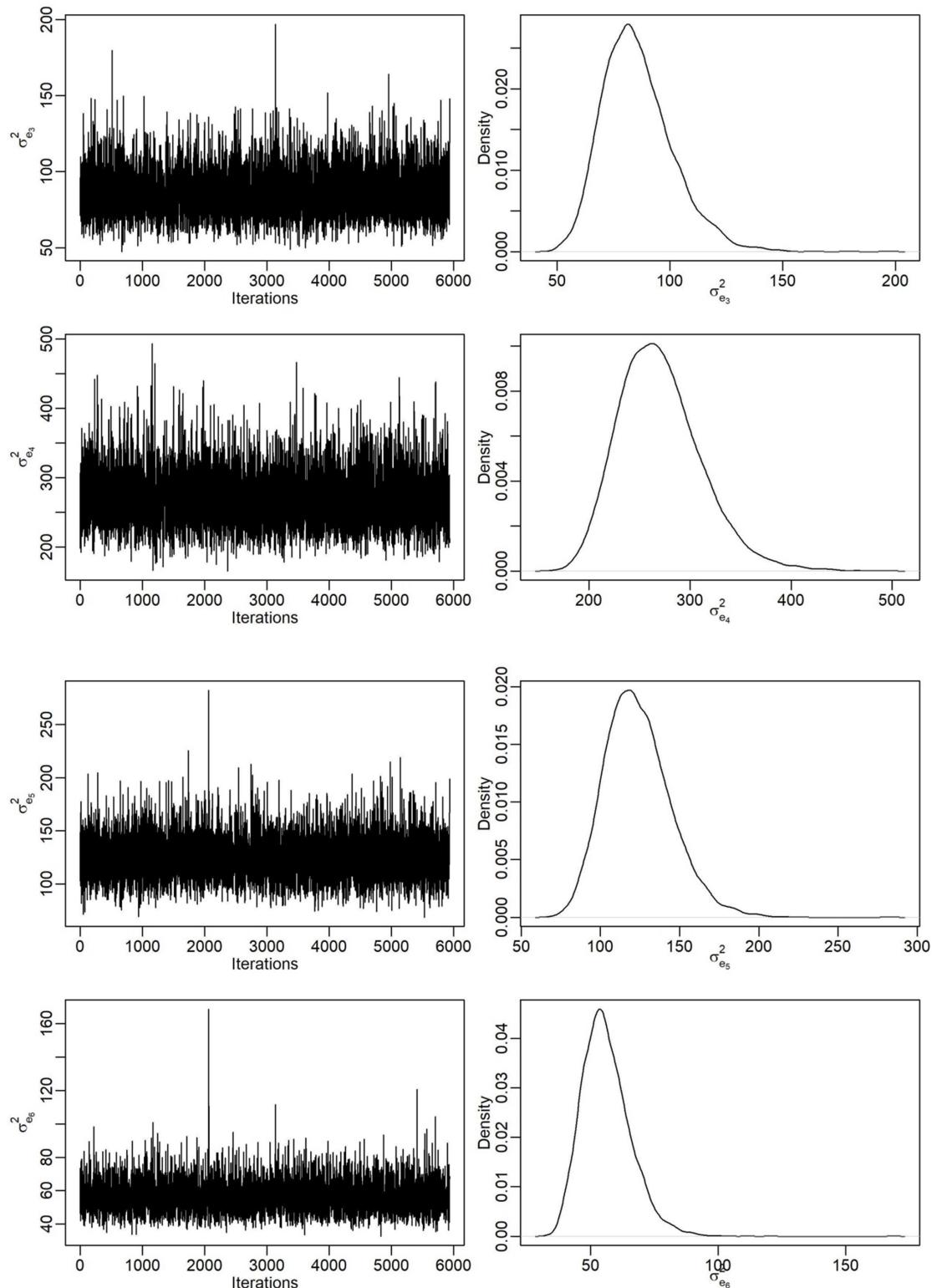


Figura S3. Gráficos de traços e densidades das distribuições *a posteriori* das componentes da variância residual relacionadas aos ambientes de E3 a E6 para o modelo BGGEH.

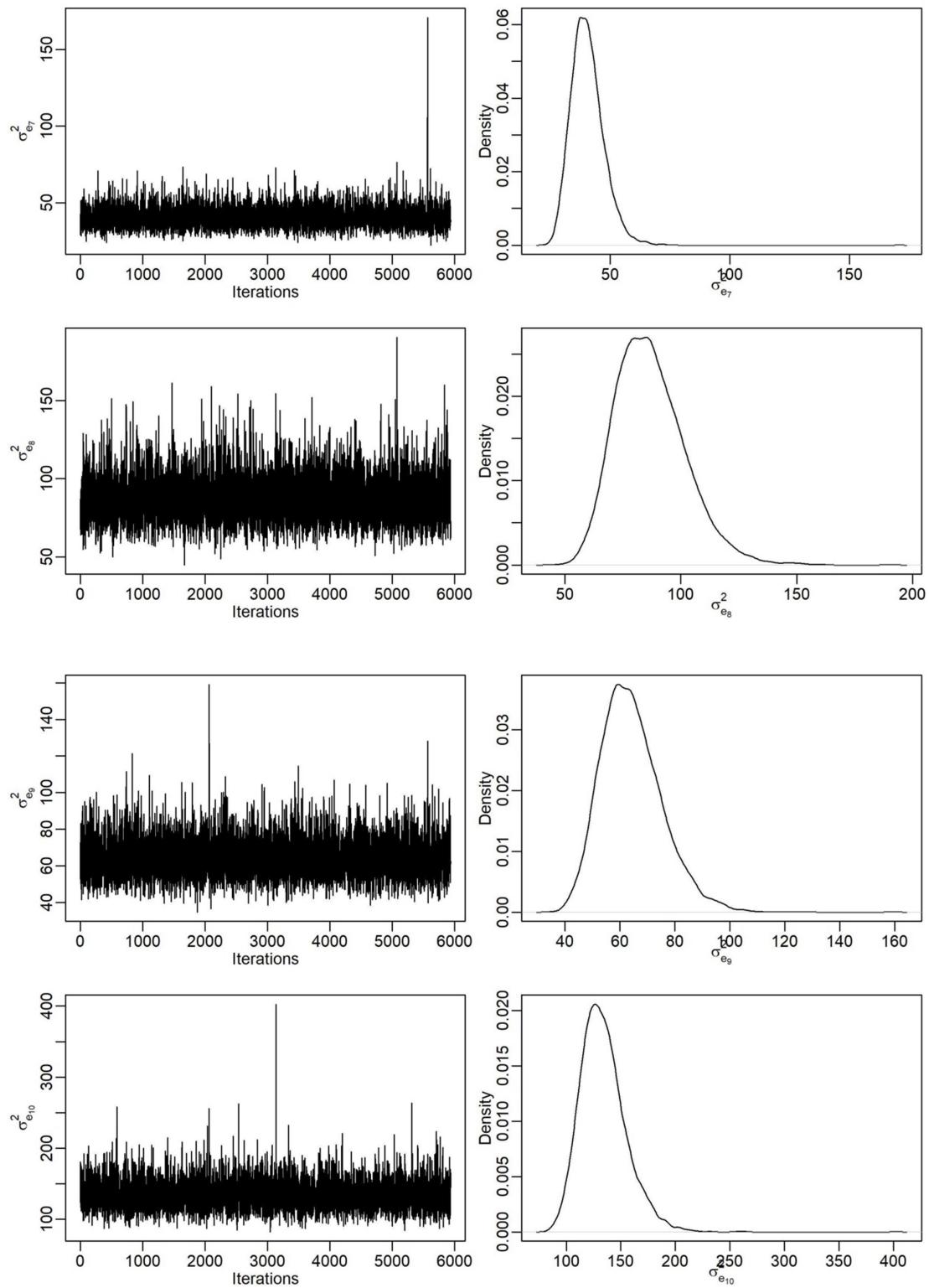


Figura S4. Gráficos de traços e densidades das distribuições *a posteriori* das componentes da variância residual relacionadas aos ambientes de E7 a E10 para o modelo BGGEH.

Apêndice S4

Resumos pontuais e intervalares *a posteriori* dos valores singulares para dados simulados

Tabela S1. Resumos pontuais e intervalares das distribuições *a posteriori* para valores singulares dos modelos BGGE e BGGEH, considerando dados simulados com homogeneidade de variância.

BGGE							
PAR	PER	PEA	SQA	ME	DP	HPD a 95% de credibilidade	
						LI	LS
λ_1	90,58	90,58	3109,33	55,76	0,59	54,61	56,91
λ_2	9,35	99,93	3429,00	17,91	0,60	16,72	19,08
λ_3	0,06	99,99	3432,12	1,46	0,93	<0,01	3,11
λ_4	0,01	100	3432,45	0,57	0,50	<0,01	1,57
λ_5	0	100	3432,51	0,25	0,28	<<0,01	0,84
λ_6	0	100	3432,53	0,12	0,16	<<0,01	0,47
λ_7	0	100	3432,53	0,06	0,10	<<0,01	0,26
λ_8	0	100	3432,53	0,03	0,06	<<0,01	0,13
λ_9	0	100	3432,53	0,01	0,03	<<0,01	0,06
λ_{10}	0	100	3432,53	0,01	0,02	<<0,01	0,03

BGGEH							
PAR	PER	PEA	SQA	ME	DP	HPD a 95% de credibilidade	
						LI	LS
λ_1	90,64	90,64	3100,63	55,68	0,98	53,64	57,48
λ_2	9,16	99,80	3413,84	17,70	0,93	15,95	19,57
λ_3	0,14	99,94	3418,80	2,23	0,67	0,99	3,61
λ_4	0,04	99,98	3420,14	1,16	0,49	0,22	2,11
λ_5	0,01	99,99	3420,59	0,67	0,36	<0,01	1,28
λ_6	0	100	3420,75	0,40	0,27	<<0,01	0,89
λ_7	0	100	3420,80	0,23	0,20	<<0,01	0,62
λ_8	0	100	3420,81	0,13	0,14	<<0,01	0,40
λ_9	0	100	3420,82	0,07	0,09	<<0,01	0,24
λ_{10}	0	100	3420,82	0,03	0,05	<<0,01	0,15

PAR, parâmetro; **PER**, percentual relativo da soma de quadrados de G+GEI explicada por cada componente; **PEA**, percentual da soma de quadrados de G+GEI acumulado; **SQA**, soma de quadrados acumulada em função dos valores singulares; **ME**, média *a posteriori*; **DP**, desvio padrão; **LI**, limite inferior; **LS**, limite superior.

Nota: A<<B significa que A<0,01B.

Tabela S2. Resumos pontuais e intervalares das distribuições *a posteriori* para valores singulares dos modelos BGGE e BGGEH considerando dados simulados com heterogeneidade de variância.

BGGE							
PAR	PER	PEA	SQA	ME	DP	HPD a 95% de credibilidade	
						LI	LS
λ_1	86,14	86,14	3047,53	55,20	1,54	52,34	58,32
λ_2	11,32	97,46	3447,90	20,01	1,64	16,73	23,16
λ_3	2,37	99,83	3531,69	9,15	2,35	4,36	13,65
λ_4	0,15	99,98	3536,86	2,27	1,67	<0,01	5,49
λ_5	0,01	99,99	3537,74	0,94	0,93	<<0,01	2,77
λ_6	0	100	3537,92	0,43	0,53	<<0,01	1,54
λ_7	0	100	3537,97	0,21	0,30	<<0,01	0,85
λ_8	0	100	3537,98	0,10	0,18	<<0,01	0,45
λ_9	0	100	3537,98	0,05	0,11	<<0,01	0,23
λ_{10}	0	100	3537,98	0,02	0,06	<<0,01	0,13

BGGEH							
PAR	PER	PEA	SQA	ME	DP	HPD a 95% de credibilidade	
						LI	LS
λ_1	96,64	96,64	2883,28	53,70	1,55	50,77	56,83
λ_2	2,42	99,06	2955,58	8,50	1,98	4,85	12,48
λ_3	0,74	99,80	2977,68	4,70	1,74	1,25	8,04
λ_4	0,14	99,94	2981,91	2,05	1,11	0,01	4,06
λ_5	0,04	99,98	2983,08	1,09	0,73	<0,01	2,44
λ_6	0,02	100	2983,42	0,59	0,49	<0,01	1,52
λ_7	0	100	2983,52	0,31	0,31	<<0,01	0,95
λ_8	0	100	2983,55	0,16	0,19	<<0,01	0,56
λ_9	0	100	2983,55	0,08	0,12	<<0,01	0,34
λ_{10}	0	100	2983,55	0,04	0,07	<<0,01	0,18

PAR, parâmetro; **PER**, percentual relativo da soma de quadrados de G+GEI explicada por cada componente; **PEA**, percentual da soma de quadrados de G+GEI acumulado; **SQA**, soma de quadrados acumulada em função dos valores singulares; **ME**, média *a posteriori*; **DP**, desvio padrão; **LI**, limite inferior; **LS**, limite superior.

Nota: A<<B significa que A<0,01B.

Apêndice S5

Representações biplots para dados simulados e dados reais

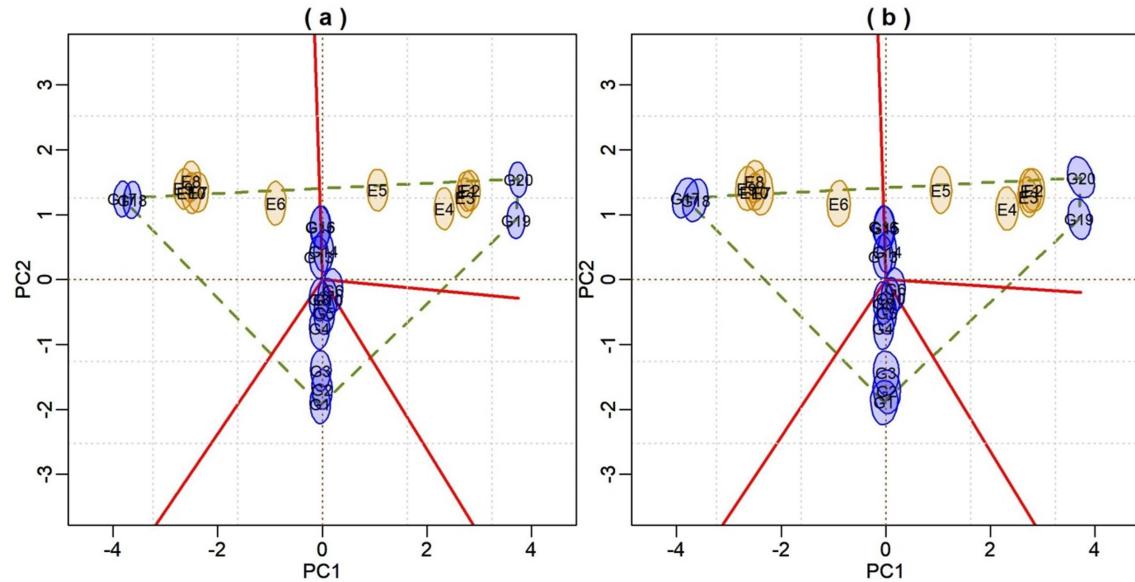


Figura S5. Biplots BGGE (a) e BGGEH (b) exibindo o padrão “quem venceu onde” para dados simulados homocedásticos e regiões bivariadas, a 95% de credibilidade, para escores genotípicos (G) e ambientais (E).

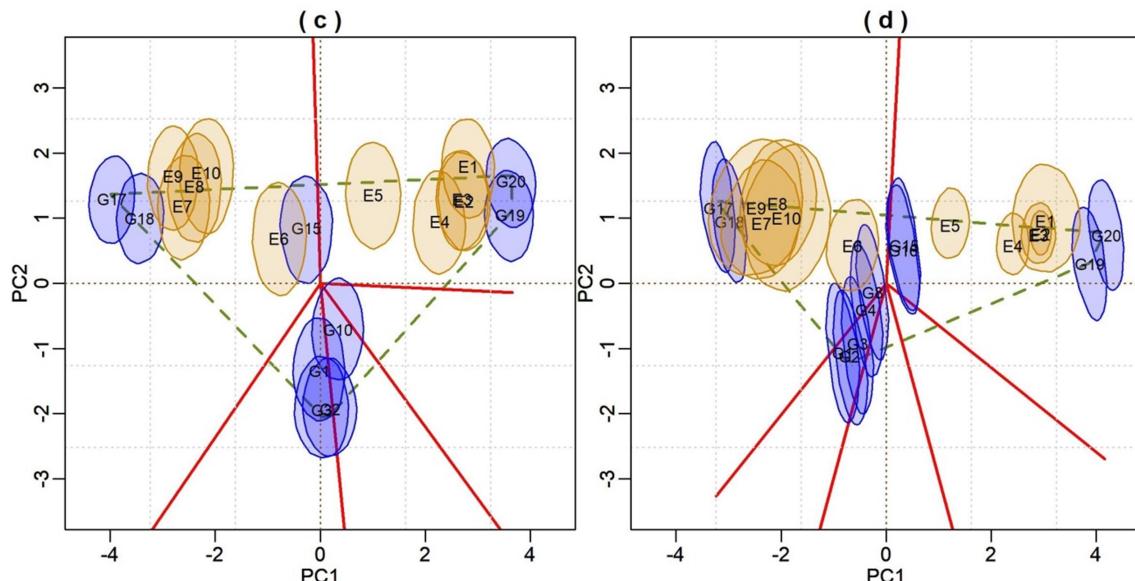


Figura S6. Biplots BGGE (a) e BGGEH (b) exibindo o padrão “quem venceu onde” para dados simulados heterocedásticos e regiões bivariadas, a 95% de credibilidade, para escores genotípicos (G) e ambientais (E).

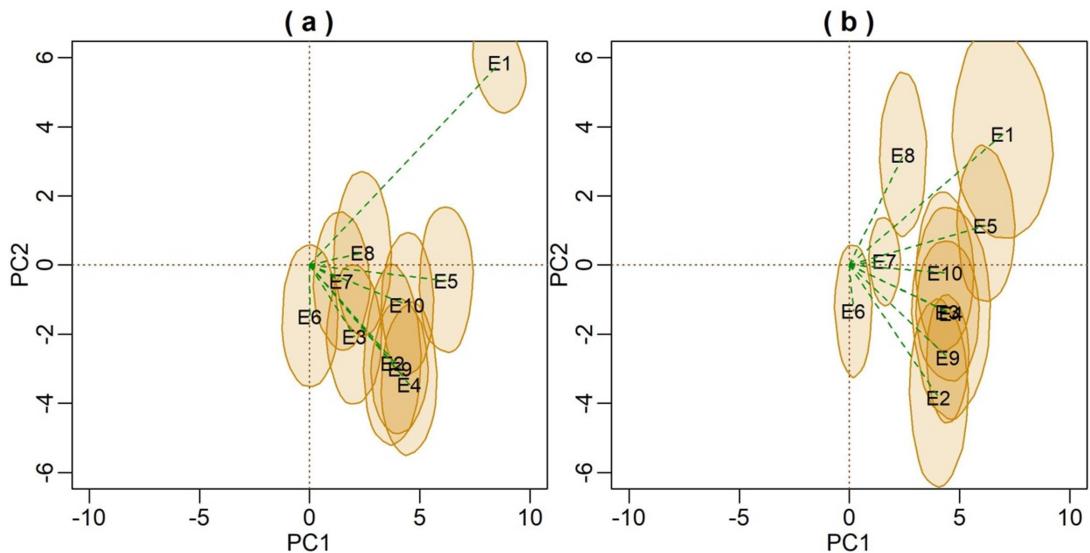


Figura S7. Biplot GGE para dados reais com regiões bivariadas, a 95% de credibilidade, para escores ambientais (E) segundo os modelos BGGE (a) e BGGEH (b).

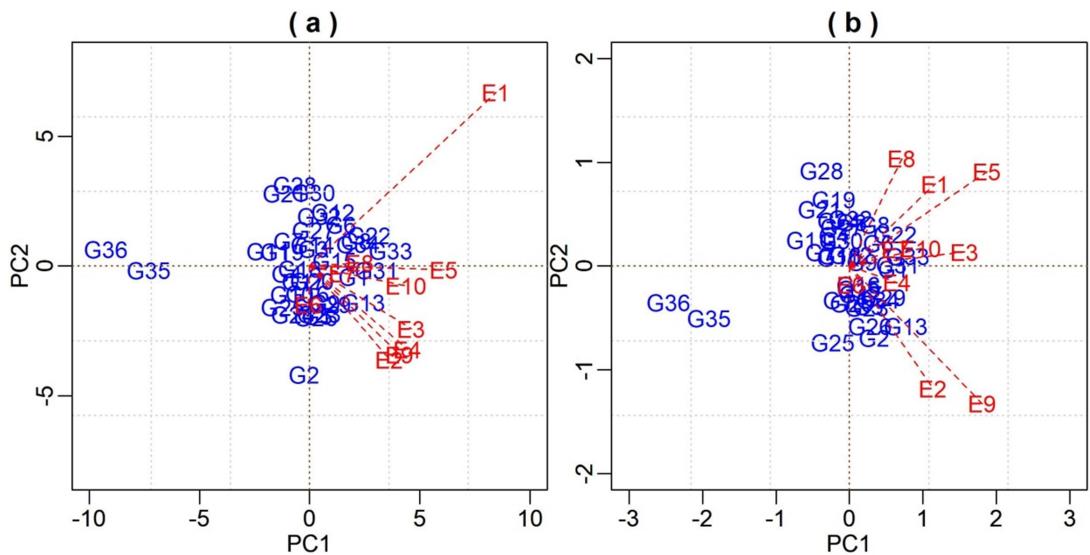


Figura S8. Biplots GGE para efeitos fixos: (a) biplot GGE com fator de escala ($s=1$) e ponderação pela herdabilidade ($h=1$); (b) biplot corrigido preliminarmente em que as médias das células estão na escala do desvio padrão e são ponderadas pela raiz quadrada da herdabilidade em cada ambiente.

CONCLUSÃO

A utilização do critério da máxima entropia flexibilizou a análise GGE, evitando decidir sobre a dimensão do modelo, bem como problemas de amostragem, já que distribuições *a posteriori* marginais impróprias ou bimodais não ocorreram. Além disso, inferências para quantidades de interesse relacionadas ao efeito da G+GEI no biplot foram incorporadas de forma direta e inédita.

O modelo BGGEH identificou diferentes precisões experimentais no biplot, captando bem a heterocedasticidade presente nos dados. As suposições sobre a estrutura da variância residual foram importantes na análise GGE, afetando a avaliação de adaptabilidade e estabilidade na interpretação biplot. O BGGEH também se ajustou melhor aos dados mostrando ser uma melhor alternativa do que a versão GGE homoscedástica.

A utilização de modelos multiplicativos, utilizando inferência bayesiana, oferece amplas perspectivas, mas carece da implementação em software com interface amigável para os pesquisadores, o que é um trabalho em andamento. Outros desenvolvimentos práticos, que precisam ser tratados, são considerar encolhimento e heterogenidade simultaneamente e fazer generalizações dessa modelagem para variáveis de resposta não contínuas, que serão objeto de investigação futura.