ROGÉRIO FERNANDES ROMÃO

ROBUSTEZ NA CAPACIDADE PREDITIVA DOS MODELOS AMMI E FATORIAIS ANALÍTICOS NO ESTUDO DE DADOS MULTI-AMBIENTAIS DESBALANCEADOS

LAVRAS - MG
2017

ROGÉRIO FERNANDES ROMÃO

ROBUSTEZ NA CAPACIDADE PREDITIVA DOS MODELOS AMMI E FATORIAIS ANALÍTICOS NO ESTUDO DE DADOS MULTI-AMBIENTAIS DESBALANCEADOS

Dissertação 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 Mestre.

Orientador
Dr. Márcio Balestre

LAVRAS - MG
2017

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

Romão, Rogério Fernandes.
Robustez na capacidade preditiva dos modelos AMMI e Fatoriais analíticos no estudo de dados multi-ambientais desbalanceados / Rogério Fernandes Romão. - 2017.
68 p.

Orientador(a): Márcio Balestre.
.
Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2017.
Bibliografia.

1. Dados faltantes. 2. Mega ambientes. 3. Interação GE. I. Balestre, Márcio . . II. Título.

ROGÉRIO FERNANDES ROMÃO

ROBUSTEZ DA CAPACIDADE PREDITIVA DOS MODELOS AMMI E FATORIAIS ANALÍTICOS NO ESTUDO DE DADOS MULTI-AMBIENTAIS DESBALANCEADOS

Dissertação 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 Mestre.

APROVADO em 16 de Fevereiro de 2017.

Prof. Dr. Thelma Sáfadi
Prof. Dr. José Airton Rodrigues Nunes

_____

Prof. Dr. Márcio Balestre
Orientador

LAVRAS - MG
2017

À minha família,
pelo apoio incondicional, força, incentivo e amizade sem igual.
DEDICO

AGRADECIMENTOS

A Deus, por iluminar meu caminho e me dar forças para seguir sempre em frente.

Aos meus pais e aos meus irmãos, pela educação, conforto familiar e apoio nos meus estudos.

A minha esposa e aos meus filhos, pela paciência, incentivo e carinho. Sem vocês, tudo seria muito mais difícil.

Ao meu orientador, Prof. Dr. Márcio Balestre, por sua dedicação, paciência, sensibilidade e apoio ao longo desta jornada.

Aos amigos e colegas, em especial ao Joel Nuvunga, Carlos Pereira da Silva e Luciano de Oliveira, pela amizade, ajuda e disposição em colaborar.

Aos amigos e conterrâneos, em particular ao Carlos Balate, pelo convívio, amizade e companhia.

Ao CNPq, pela concessão de bolsa e oportunidade de continuar com os estudos.

Aos docentes e funcionários do Programa de Estatística e Experimentação Agropecuária (UFLA), pelo apoio e contribuição no meu crescimento.

A todos muito obrigado.

RESUMO

O presente trabalho teve por objetivo verificar a robustez na capacidade preditiva dos modelos AMMI utilizando diversas abordagens bayesianas e frequentistas, e Fatorial Analítico (FA) no estudo de dados multi-ambientais (MET) desbalanceados, usando dados simulados. Para verificar a ficiência destes métodos foram feitos desbalanceamentos aleatórios nos dados, com níveis de 10%, 33% e 50% de perda. Para avaliar a capacidade preditiva de dados faltantes nos modelos propostos, foram utilizadas a estatística PRESS (*prediction error sum square*) e a correlação entre o valor predito e observado usando a validação cruzada. Os resultados mostraram que em termos preditivos, ao nível de 10% de desbalanceamento, os modelos AMMI Bayesiano com heterogeneidade de variâncias (AMMIB-D) e modelos AMMI via algoritmo EM para efeitos aleatórios de genótipo e fixo de ambiente (EM-AMMI M) foram superiores seguidos dos modelos AMMI Bayesiano com homogeneidade de variâncias (AMMIB-I) e FA2. A 30% de perda dos dados, o modelo AMMIB-I foi superior, seguidos dos modelos EM-AMMI M, AMMIB-D, modelos AMMI via algoritmo EM para efeitos fixos de ambiente e genótipo (EM-AMMI F) e FA2. A 50% de perda dos dados, os modelos AMMIB-I e AMMIB-D foram superiores, seguido do modelo FA2. Com isso pode-se concluir que os os modelos AMMI seja frequentista ou bayesiano e Fatorial Analítico foram robustos no estudo de dados MET com altos níveis de perda de genótipos nos ambientes.

Palavras-chave: Dados faltantes, mega ambientes, interação GE, validação cruzada.

ABSTRACT

The present work aimed to verify the robutness of the AMMI predictive ability through using several Bayesian and Frequentist approaches, and Analytical Factor (FA) in the study of unbalanced multi-environmental data (MET), using simulated data. To verify the efficiency of these methods, random unbalanced was performed using 10%, 33% and 50% of loss. To evaluate the predictive ability of the missing data in proposed models, the PRESS statistics (*prediction error sum square*) and the correlations between the observed and predicted values were used, through cross-validation methods. The results showed that in predictive terms, at the level of 10% of unbalance the Bayesian AMMI models with variance heterogeneity (AMMIB-D) and AMMI models through EM algorithm for random effects of genotype and fixed environment (EM-AMMI M) Were superior followed by Bayesian AMMI models with homogeneity of variances (AMMIB-I) and FA2. At 30% of data loss, the AMMIB-I was superior, followed by EM-AMMIB-D models, AMMI models through EM algorithm for fixed effects of environment and genotype (EM-AMMI F) and FA2. At 50% data loss, the AMMIB-I and AMMIB-D models were superior, followed by the FA2 model. It can be concluded that the AMMI models are frequentist or Bayesian and Factorial Analytical were robust in the study of MET data with high levels of loss of genotypes in the environments.

Keywords: Missing data, mega environments, GE interaction, cross validation.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Identificar e recomendar genótipos (G) com alta produtividade e ampla adaptabilidade, aos mais variados ambientes (E), tem se configurado como um dos principais objetivos dos programas do melhoramento genético. Entretanto, o trabalho do melhorista é frequentemente dificultado pela presença da interação entre genótipo e ambiente (GE), que influencia o comportamento dos genótipos em função das condições ambientais.

Existem na literatura diferentes metodologias destinadas à avaliação da interação GE, que variam de métodos uni e/ou multivariados e que dependem de aspectos como: dados experimentais, número de ambientes, precisão requerida e do tipo de informação desejada. A análise da interação GE, no contexto do melhoramento de plantas, tem sido alvo de muitos estudos, e vários métodos têm sido propostos, contribuindo para a eficiência das análises e para seleção dos genótipos sob diferentes condições ambientais.

Dentre os métodos que foram desenvolvidos para modelar, de forma eficiente, o efeito da interação GE podemos destacar os modelos lineares-bilineares de efeitos fixos. Dentre os modelos dessa classe, destacam-se os de efeitos principais aditivos e interação multiplicativa (*additive main effects and multiplicative interactions* -AMMI) e o de efeito principal de genótipo mais interação (GGE ou SREG2-*site regression*), que são os mais populares e com ampla aplicabilidade (CORNELIUS et al. 1996; CROSSA, 2002, 2012). Nesses modelos, os padrões de resposta de genótipos e ambientes podem ser visualizados graficamente usando biplots (GABRIEL, 1971), que permitem agrupamento de ambientes e genótipos semelhantes quanto ao efeito da interação, bem como a identificação gráfica do genótipo com maior potencial em cada subgrupo de ambientes.

Esses métodos de análise apresentam sérias limitações, como é o caso da falta de flexibilidade para tratar dados desbalanceados e com heterogeneidade de variâncias (CROSSA et al, 2011; SMITH et al. 2001; KELLY et al. 2007; OLIVEIRA et al. 2015). Essas limitações levam a busca por métodos alternativos. Dos métodos propostos, os que vêm se destacando são aqueles baseados em modelos mistos, como o Fatorial Analítico (FA) de Smith et al.

(2001), AMMI ponderado (*weighted* AMMI) de Rodrigues et al. (2014) e o EM-AMMI de Gauch e Zobel (1990), e os métodos baseados em inferência bayesiana como o de Edwards e Jannik (2006), os modelos lineares bayesianos (tais como AMMI-Bayesiano; Crossa et al. 2011; Oliveira et al. 2015 e SRGE/GGE Bayesiano , Oliveira et al. 2016) e os duplos hierárquicos generalizados de Lee e Nelder (2006).

Dentre esses métodos, o que vem se destacando é o modelo Fatorial Analítico (FA), que considera como aleatórios os efeitos dos genótipos e interação GE, além de lidar com dados heterocedásticos e com erros correlacionados, e ser flexível para lidar com dados faltantes sem requerer passos adicionais em seu procedimento para imputação de valores perdidos.

Existem relatos também da utilização de métodos de imputação de valores faltantes associados aos modelos bilineares tradicionais, como a utilização de métodos não paramétricos, Bootstrap e validação cruzada (ARCINIEGAS-ALARCÓN e DIAS, 2009). Esses métodos não são triviais quando aplicados aos modelos baseados na abordagen AMMI, fazendo com que os modelos FA sejam preferidos para lidar com esse problema (KELLY et al. 2007; BURGUEÑO et al. 2007).

Devido às suas propriedades, os métodos baseados em inferência bayesiana podem ser úteis para lidar com casos em que se tenham dados faltantes, pois é sabido que durante o processo de seleção no melhoramento de plantas, os genótipos recém-criados são adicionados, enquanto que os genótipos selecionados são descartados, além de que vários genótipos não são testados em todos ambientes, resultando em dados desbalanceados, que geralmente complicam as análises (FISCHER et al. 2009; PIEPHO and MOHRING, 2007).

Apesar de conhecidas as vantagens dos métodos referidos ao lidar com dados desbalanceados, cada um deles apresenta seus méritos e deméritos (OLIVEIRA et al. 20015, NUVUNGA et al. 2015, CROSSA et al. 2006). Os modelos de efeitos fixos, como já ressaltado, são baseados em ANAVA e geralmente requerem imputação dos dados faltantes e o modelo FA é baseado em modelos mistos, que pode levar a qualidade de predições diferentes,

porém faltam estudos que fazem este tipo de comparação. Bergamo et al. (2008) e Arciniegas-Alarcón e Días (2009) mostraram a eficiência dos métodos de imputação -AMMI e Nuvunga et al. (2015) mostraram a robustez do modelo FA para lidar com dados faltantes.

Embora os referidos autores tenham testado a robustez dos modelos citados, não existe relato na literatura de trabalhos comparando a eficiência dos mesmos na predição de dados faltantes, apesar de vários autores defenderem a robustez do modelo FA em lidar com dados faltantes e com heterogeneidade de variâncias.

Dessa forma, neste trabalho objetivou-se verificar a robustez na capacidade preditiva dos modelos AMMI via algoritmo EM para efeitos fixos de ambiente e genótipo (EM-AMMI F) e efeitos aleatórios de genótipo e fixo de ambiente (EM-AMMI M), AMMI Bayesiano com homogeneidade de variâncias (AMMIB-I), AMMI Bayesiano com heterogeneidade de variâncias (AMMIB-D) e Fatotial Analítico (FA) no estudo de dados multi-ambientais (MET) desbalanceados, usando dados simulados.

2 REFERENCIAL TEÓRICO

Nesta seção são apresentados os aspetos teóricos relacionados ao estudo da interação GE, análise de variância convencional, métodos de análise multivariada, modelos EM-AMMI, AMMIB e FA.

2.1 O estudo da interação Genótipo por Ambiente (GE)

Para estudar a interação GE precisamos de vários genótipos sendo testados em vários ambientes e, geralmente, por vários anos. Em geral, a interação GE é estudada mais detalhadamente durante os ensaios finais dos programas de melhoramento. Respostas diferenciais dos genótipos desenvolvendo-se em diferentes ambientes mostraram uma flutuação significativa na produtividade em relação aos outros. Essas mudanças são influenciadas por diferentes condições ambientais e são referidas como interação GE.

O estudo da interação GE tem várias implicações em um programa de melhoramento e na etapa de avaliação de genótipos. Assim, deve-se buscar alternativas para amenizar o seu efeito, com destaque para a identificação de genótipos de comportamento previsível (estabilidade) e responsivos à melhoria do ambiente (adaptabilidade).

Inúmeros métodos de análise de estabilidade e adaptabilidade, baseados em diferentes princípios, já foram descritos. Os mais utilizados são baseados em regressões, com utilização de um (EBERHART e RUSSEL, 1966) ou dois segmentos de reta (CRUZ et al. 1989). Como nem todos os dados se ajustam a modelos lineares, outra possibilidade é a utilização de métodos não paramétricos (CRUZ e CARNEIRO, 2003), como os de Lin e Binns (1988), modificado por Carneiro (1998), o de Annicchiarico (1992) e o da análise da interação multiplicativa dos efeitos principais aditivos (AMMI).

O método de Lin e Binns (1988), modificado por Carneiro (1998), identifica os genótipos mais estáveis, por meio de um único parâmetro de estabilidade e adaptabilidade, e contempla os desvios em relação à produtividade máxima obtida em cada ambiente, além de possibilitar o detalhamento dessa informação para ambientes favoráveis e desfavoráveis. O método de Annic-

chiarico (1992) avalia a estabilidade por meio do risco associado em relação à adoção das cultivares.

Das metodologias já propostas para o estudo da interação GE, os modelos estatísticos multiplicativos são muito úteis para estudar padrões de performance de genótipos por ambientes e fazer predições a respeito da performance média de genótipos à ambientes específicos (GAUCH, 1988; GAUCH e ZOBEL, 1996). Dentre esses modelos, os que vem ganhando espaço são os modelos que combinam o estudo da adaptabilidade e estabilidade conhecidos como modelos linear-bilineares, como o *Site Regression* (SREG) (CORNELIUS e CROSSA, 1996), os modelos de efeitos aditivo principais e de interação multiplicativa (AMMI) (GAUCH, 1988; GAUCH e ZOBEL, 1997), e o Fator Analítico - *factor analytic multiplicative mixed models* (FA). As vantagens e desvantagens de cada um desses métodos podem ser encontradas em trabalhos de revisão de Smith et al. (2005), Yan et al. (2007) e Crossa et al. (2012).

Entre esses modelos, o modelo AMMI evidenciou-se no estudo de estabilidade e adaptabilidade, que utiliza componentes principais e estuda a interação em um modelo multiplicativo (GAUCH e ZOBEL, 1996). Apesar das vantagens, esse método apresenta o problema de não poder lidar com dados faltantes e com heterogeneidade de variâncias. Como forma de contornar esse problema várias propostas de modificação do modelo AMMI têm sido desenvolvidas, como os modelos EM-AMMI, W-AMMI e AMMI baseados em imputação (GAUCH e ZOBEL, 1990, RODRIGUES et al. 2014, ARCINIEGAS-ALARCÓN e DIAS, 2009). Dentre trabalhos usando essas propostas, temos a destacar os trabalhos de:

• Calinski et al. (1992), que detectaram que o uso de estimativas AMMI baseadas em mínimos quadrados alternados pode ser uma boa solução na cultura do trigo em matrizes de dimensão 10×28 e 15×12 (cultivares × locais);

• Bergamo et al. (2008), que propõem um método de imputação múltipla livre de distribuição (IMLD) aplicado especificamente à matriz de interação GE com genótipos de *Eucalyptus grandis* em uma matriz de dimensão

20×7 (cultivares × locais) com base na decomposição por valores singulares, sem usar o modelo AMMI e sem pressuposições estruturais ou distribucionais sobre os conjuntos de dados;

• Rodrigues et al. (2014), que apresentam um algoritmo generalizado para estimação em modelo AMMI ponderado (W-AMMI) com heterogeneidade de variâncias;

• Paderewski e Rodrigues (2014), que apresentam um pacote para imputação de valores faltantes no EM-AMMI.

Apesar dessa extensiva lista de métodos aplicados ao modelo AMMI para tratar de dados desbalanceados, esses têm sido preteridos no estudo da interação GE por métodos baseados em modelos mistos como fator analítico - FA. Crossa (2012) destaca que o modelo FA lida com a matriz de variâncias e covariâncias e que, dependendo do objetivo do pesquisador, este pode ser visto como FA-AMMI se o ajuste do genótipo (G) for feito de forma separada com GE e FA-SREG se o ajuste for feito $G + GE$.

Os modelos FA são amplamente aceitos por apresentarem as seguintes vantagens: consideram os efeitos de genótipo e de GE como aleatórios; são adequados não apenas para dados balanceados; consideram a variação espacial dentro dos ensaios; consideram a heterogeneidade de variâncias entre ensaios, consideram os diferentes números de repetições nos ensaios e não requerem imputação de valores faltantes. No entanto, essas são características geralmente encontradas em experimentos de campo. Kelly et al. (2007), Piepho (1998) e Smith, Cullis e Thompson (2001, 2005) mostraram a superioridade dos modelos FA no estudo da interação GE.

2.2 Análise de variância convencional

Considere um ensaio em que a produção de g genótipos é medida em e ambientes, cada um com r repetições. O método clássico para analisar a variação de rendimento total contida nas ger observações é a análise de variância (FISHER, 1918 - 1925).

Após a remoção do efeito do bloco ao combinar os dados, as variações nas observações são divididas em duas partes:

i. Efeito aditivo principal para genótipos g e ambientes e;

ii. Efeitos não aditivos devido a interação ge.

A análise de variância dos dados observados (y_{ij}), é executada considerando-se o modelo estatístico:

$$y_{ij} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ij} \ , \tag{1}$$

em que:
y_{ij} é a resposta média do genótipo i no ambiente j;
μ é a constante, média geral das observações;
g_i é o efeito do i-ésimo genótipo;
e_j é o efeito do j-ésimo ambiente;
$(ge)_{ij}$ é o efeito da interação do i-ésimo genótipo com o j-ésimo ambiente;
ε_{ij} é o erro experimental médio associado à observação y_{ij}, assumido como independente $\varepsilon_{ij} \sim N(0, \sigma^2)$.

A interação não aditiva, conforme definido em (1), implica que o valor esperado do i-ésimo genótipo no ambiente j (Y_{ij}) dependa não apenas dos níveis de g separadamente, mas também na combinação particular de níveis de g e e (CROSSA, 1990).

A principal limitação dessa análise é que as variâncias dos erros dos ambientes devem ser homogêneas para testar diferenças genotípicas. Uma forma mais adequada para o teste de significância é realizada ponderando-se cada genótipo com o inverso da sua variância residual estimada (COCHRAN e COX, 1957). Essa análise ponderada atribui menos pesos para ambientes que têm um quadrado médio residual elevado. A desvantagem da análise ponderada é que os pesos podem ser correlacionados com as respostas do rendimento no ambiente (ambientes com rendimento elevado apresentam maior variância do erro e ambientes com baixos rendimentos apresentam variâncias de erro reduzidas), o que pode mascarar o verdadeiro desempenho de alguns genótipos em certos ambientes (CROSSA, 1990).

Uma das principais deficiências da análise de variância conjunta de ensaios multiambientes é que ela não explora qualquer estrutura subjacente dentro da observação não aditiva (GE).

Com a análise de variância não se consegue determinar o padrão de resposta de genótipos e ambientes. As informações contidas nos $(g-1)(e-1)$ graus de liberdade são perdidas, principalmente se não for feita uma análise mais aprofundada.

A análise de variância dos ensaios multiambientes (MET) é útil para estimar componentes de variância relacionadas com diferentes fontes de variação, incluindo genótipos e GE. Em geral, a metodologia de componentes de variância é importante em ensaios MET, desde a estimação dos erros até a mensuração do desempenho produtivo de um genótipo, que surge em grande parte da interação GE. Portanto, o conhecimento do tamanho dessa interação é necessário para:

i. Obter estimativas eficientes dos efeitos genotípicos;

ii. Determinar recurso ideal de alocações, que é o número de parcelas e os locais a serem incluídos em estudos futuros.

2.3 Métodos de análise multivariada

De acordo com Crossa (1990), a análise multivariada tem três propósitos principais:

i. Eliminar o ruído do padrão de dados (ou seja, distinguir a variação sistemática da não sistemática);

ii. Resumir os dados;

iii. Revelar uma estrutura nos dados.

Em contraste com os métodos estatísticos clássicos, o propósito da análise multivariada é elucidar a estrutura interna dos dados a partir dos quais as hipóteses podem ser geradas e posteriormente testadas por métodos estatísticos (CROSSA, 1990).

Segundo Cruz e Regazzi (1994), embora as técnicas de análise multivariada tenham sido desenvolvidas para solucionar problemas em áreas específicas, ela pode ser aplicada em diversas áreas da ciência, pois os métodos multivariados são escolhidos com base nos objetivos da pesquisa.

Dois grupos de técnicas multivariadas têm sido usados para elucidar a estrutura interna da interação GE:

1. Métodos para a distinção entre grupos

a) Análise de agrupamento;

b) Análise discriminante.

2. Métodos para o estudo da estrutura de covariâncias ou correlação entre variáveis

c) Componentes principais (ACP);

d) Análise de fatores (AF).

Segundo Baker et al. (1988), a técnica (*c*) é a que tem maior aplicação em genética e melhoramento, visto que ela tem por finalidade abordar aspectos como a geração, a seleção e a interpretação dos componentes investigados.

As técnicas de ordenação, como análise de componentes principais e análise de fatores, assumindo que os dados são contínuos, procuram representar o genótipo e relações do ambiente tão fielmente quanto possível em um pequeno espaço dimensional reduzido. Em um gráfico representam-se conjuntamente os ambientes similares e os genótipos semelhantes entre si, sendo que quanto mais diferente, mais afastados se apresentam. A ordenação é eficaz para visualizar as relações e reduzir o ruído (GAUCH e ZOBEL, 1988).

Técnicas de classificação como análise de agrupamento e análise discriminante procuram descontinuidades nos dados. Esses métodos agrupam entidades similares e são eficazes para resumir a redundância nos dados (CROSSA, 1990). O método (*a*) permite a formação de grupos (não conhecidos previamente) por meio de técnicas de agrupamento aplicados sobre medidas de dissimilaridade entre fenótipos. O método (*b*) tem como maior aplicação a descriminação ou alocação de um conjunto de genótipos em grupos ou populações previamente conhecidos, usando para isso um certo número de caracteres avaliados.

Dentre as técnicas multivariadas, as mais usadas nos programas de melhoramento para o estudo da interação GE são: AMMI e FAMM.

2.4 Metodologia AMMI

A análise AMMI é uma combinação de métodos univariados (análise de variância) com métodos multivariados (análise de componentes principais e decomposição de valores singulares). Esta combina em um único modelo, componentes aditivos para os efeitos principais de genótipos g_i e de ambientes e_j, e componentes multiplicativos $(ge)_{ij}$ para os efeitos da interação (GAUCH e ZOBEL, 1988).

Assim, a resposta média de um genótipo i num ambiente j é dada por:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} + \varepsilon_{ij} \ , \tag{2}$$

com $(ge)_{ij}$ modelado por:

$$\sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} \ , \tag{3}$$

em que:
Y_{ij} é a resposta média do i-ésimo genótipo no j-ésimo ambiente;
μ é a constante, média geral das observações;
g_i é o efeito do i-ésimo genótipo;
e_j é o efeito do j-ésimo ambiente;
λ_k é o k-ésimo valor singular de GE (escalar);
γ_{ik} e α_{jk} são os elementos relacionados ao i-ésimo genótipo e ao j-ésimo ambiente dos vetores γ_k e α'_k, respectivamente.

O índice k (k = 1, 2,..., p) em que: $p = min\{g\text{-}1,\ e\text{-}1\}$, é o posto de GE, tomado até n no somatório ($n < p$), determina uma aproximação de mínimos quadrados para a matriz GE pelos n primeiros termos da decomposição por valores singulares (DVS); deixando-se um resíduo adicional denotado por ρ_{ij}. Para $n = p$ não se tem mais a aproximação, mas sim uma decomposição exata da matriz, implicando em ρ_{ij} nulo. Sob as restrições

de identificabilidade:

$$\Sigma_i g_i = \Sigma_j e_j = \Sigma_i (ge)_{ij} = \Sigma_j (ge)_{ij} = 0 \; . \tag{4}$$

Além da média geral (μ) e do erro experimental médio (ε_{ij}), os demais termos do modelo resultam na chamada DVS da matriz de interações $GE_{(g \times e)}$. A matriz de interações é obtida como resíduo do ajuste aos efeitos principais, por ANAVA, aplicada à matriz de médias, $Y_{(g \times e)} = [Y_{ij}]$. $\gamma_{k(g \times 1)}$ e $\alpha'_{k(1 \times e)}$ são os respectivos vetores singulares (vetor coluna e vetor linha) associados a λ_k (DUARTE e VENKOVSKY, 1999).

Para ilustrar os componentes aditivos e multiplicativos no modelo, pode-se escrevê-los ainda da seguinte forma:

$$Y_{ij} = (A + e_{ij}) + Z + \varepsilon_{ij} \; , \tag{5}$$

onde:

$$A = \mu + g_i + e_j \; , \tag{6}$$

em que:
$A + e_{ij} = \mu + g_i + e_j$ - parte aditiva; e
$Z = \sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk}$ - parte multiplicativa.

Sob o ponto de vista da análise de componentes principais, além dos termos já definidos anteriormente, têm-se ainda as seguintes correspondências:

- λ_k é a raiz quadrada do k-ésimo autovalor das matrizes $(GE)(GE)'$ e $(GE)'(GE)$ (de iguais autovalores não nulos) $\Rightarrow$ λ_k^2 é o k-ésimo autovalor;

- γ_{ik} é o i-ésimo elemento (relacionado ao genótipo i) do k-ésimo autovetor de $(GE)(GE)'$ associado a λ_k^2 e α_{jk} é o j-ésimo elemento (relacionado ao ambiente j) do k-ésimo autovetor de $(GE)'(GE)$ associado a λ_k^2.

Nota-se que o termo $(ge)_{ij}$ (interação no modelo tradicional) é agora descrito como uma soma de p parcelas, cada uma resultante da multiplicação

de λ_k, expresso na mesma unidade de Y_{ij}, por um efeito genotípico (γ_{ik}) e um efeito ambiental (α_{jk}), ambos adimensionais. O termo λ_k traz uma informação relativa à variação devida à interação GE, na k-ésima parcela, de forma que a soma das p parcelas recompõem toda a variação ($SQ_{GXE} = \sum_{k=1}^{p} \lambda_k^2$). Os efeitos γ_{ik} e α_{jk} representam pesos para o genótipo i e para o ambiente j, naquela parcela da interação λ_k^2.

Entretanto, pela abordagem AMMI não se busca recuperar toda a SQ_{GXE}, mas apenas a parcela mais fortemente determinada por genótipos e ambientes (linhas e colunas da matriz GE), ou seja o padrão (parte determinística ou sistemática). Assim, a interação do genótipo i com o ambiente j é descrita por: $\sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk}$, descartando-se o resíduo adicional ρ_{ij} dado por $\sum_{k=n+1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}$. Aqui, como em ACP, esses eixos captam, sucessivamente, porções cada vez menores da variação presente na matriz GE ($\lambda_1^2 \geq \lambda_2^2 \geq ... \geq \lambda_p^2$). Por isso, o método AMMI é visto como um procedimento capaz de separar padrão e ruído na análise da $SQ_{GXE} = \sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk}$ e $\sum_{p}^{k=n+1} \lambda_k \gamma_{ik} \alpha_{jk}$, respectivamente (WEBER et al. 1996).

Dependendo do número de termos multiplicativos que tenham sido incluídos, o modelo AMMI pode ser chamado de AMMI0, AMMI1, AMMI2, etc.

2.4.1 EM-AMMI

Em ensaios desbalanceados, um esquema iterativo construído com base no procedimento do algoritmo EM-AMMI é usado para realizar imputação de dados para análise AMMI, por meio do algoritmo EM. Os parâmetros aditivos são inicialmente definidos pelo cálculo da média geral, médias de genótipos e médias dos ambientes obtidos a partir dos dados observados.

Os resíduos das células são inicializados com o valor da resposta média das células observadas mais a média geral, subtraídos as médias dos genótipos e dos ambientes, porém, as interações para as posições em falta são inicialmente definidos como zero. Os parâmetros multiplicativos iniciais são obtidos a partir da DVS dessa matriz de resíduos, e os valores ausentes

são preenchidos pelas estimativas AMMI adequadas. Em iterações subsequentes, o procedimento habitual AMMI é aplicado à matriz completa e os valores ausentes são atualizados pelas estimativas AMMI correspondentes. O processo iterativo é interrompido quando não se verificarem mudanças significativas em sucessivas iterações.

Dependendo do número de termos multiplicativos utilizado, o método de imputação pode ser referido como o EM-AMMI0, EM-AMMI1, etc (GAUCH e ZOBEL, 1990). Os estudos de Calinski et al. (1992), Piepho (1995), Arciniegas-Alarcon e Dias (2009) e Paderewski e Rodrigues (2014) mostraram que os melhores resultados para imputação em modelos AMMI são obtidos incluindo no máximo uma componente multiplicativa.

É importante ressaltar que o modelo de análise nem sempre será o mesmo que o modelo de imputação. Na análise de experimentos desbalanceados, é recomendável escolher o número de componentes multiplicativos do modelo AMMI somente a partir da informação observada e fazer a estimação clássica dos parâmetros com base nas matrizes completadas por imputação.

No entanto, o modelo AMMI não irá ser apreciado como um modelo de análise, mas será avaliado apenas como um modelo de imputação. Arciniegas-Alarcón et al. (2011) observaram que os erros associados aos modelos de imputação AMMI aumentam à medida que o número de componentes multiplicativos aumenta.

2.4.1.1 Algoritmo EM-AMMI

O método EM-AMMI completa o conjunto de dados faltantes de acordo com os efeitos principais (G e E) e a iteração (GE). O algoritmo EM pode ser representado de acordo com o fluxograma apresentado na Figura 1 e funciona da seguinte forma (PADEREWSKI e RODRIGUES, 2014; GAUCH e ZOBEL, 1990):

1. O pesquisador pode fornecer valores iniciais para as células com valores ausentes (com o argumento da função EM-AMMI). Caso contrário, os valores iniciais são calculados com o aumento da média geral por efeitos principais de linhas e efeitos principais colunas. Dessa forma, a matriz de

observações é pré-preenchida;

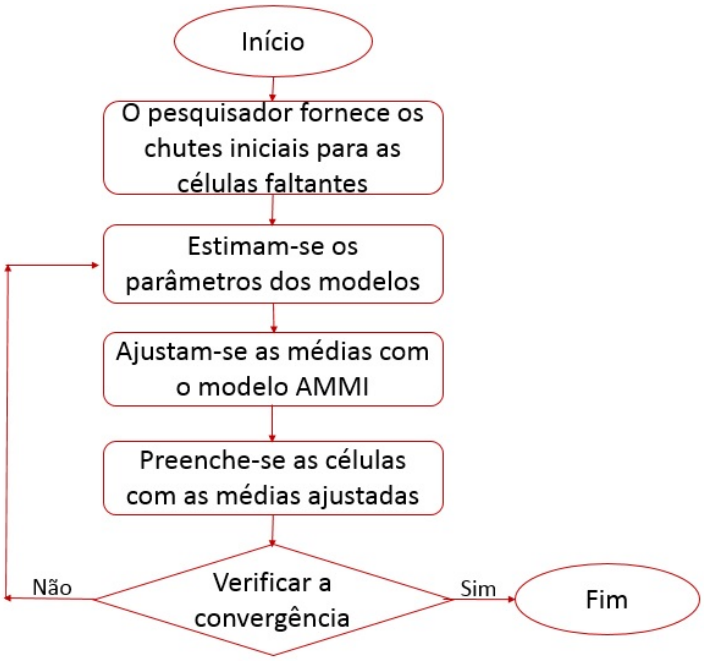
2. Os parâmetros do modelo AMMI são estimados;

3. As médias ajustadas são calculadas com base no modelo AMMI com n componentes principais;

4. As células com valores omissos são preenchidas com as médias ajustadas;

5. Se a alteração máxima nestes valores (a distância de Chebyshev entre valores ausentes estimados em duas etapas iterativas sucessivas) é maior do que a precisão assumida, os passos 2 a 5 são repetidos. Caso contrário, o algoritmo para.

Figura 1 – Algoritmo EM



2.4.2 Modelo AMMI-Bayesiano (AMMIB)

O modelo bayesiano permite que o pesquisador incorpore seu conhecimento prévio a respeito da distribuição do parâmetro θ a ser estimado. Visto

que θ pode assumir diferentes graus de incerteza, estes podem ser representados através de modelos probabilísticos para θ, formalmente denominados como distribuição a priori (Pinho, 2006).

A distribuição a priori para a quantidade de interesse θ deve representar, probabilisticamente, o conhecimento que se tem sobre θ antes da realização do experimento. A combinação de verossimilhança de cada um dos possíveis valores de θ, L(θ,y), e a priori p(θ) leva à distribuição a posteriori p(θ|y).

Cotes et al. (2006) mostraram como incorporar informações a priori visando a melhor estimação dos parâmetros, argumentando que a metodologia Bayesiana representa uma opção flexível para modelar a interação GE.

Para superar as limitações do modelo AMMI (poder lidar com dados desbalanceados e com heterogeneidade de variâncias), Viele e Srinivasan (2000), propuseram o método bayesiano para estimação de parâmetros no modelo AMMI. Utilizando a abordagem bayesiana é possível obter amostras de distribuições de probabilidade dos parâmetros do modelo, utilizando métodos Monte Carlo com cadeias de Markov, tais como o amostrador de Gibbs, Metropolis-Hastings, entre outros.

Outras metodologias bayesianas para o modelo AMMI foram realizadas por Crossa et al. (2011) e Liu (2001), tendo nessas metodologias sido utilizado o modelo AMMI com as mesmas restrições do AMMI convencional.

Nessa dissertação, o modelo AMMI será abordado em uma perspectiva bayesiana considerando um modelo aleatório para genótipos. Para tanto, será atribuída priori não informativa para o efeito de ambientes e priori informativa para o efeito de genótipos. Sendo o modelo dado por:

$$y = X_1\beta + Zg + \sum_{k=1}^{t} \lambda_k diag(Z\alpha_k) X_2\gamma_k + \varepsilon \ , \tag{7}$$

em que:

- y é o vetor de observações com dimensão $n \times 1$;
- X_1 é a matriz de delineamento com dimensão $n \times e$ associada à β;
- β é o vetor de efeitos ambientais (locais e blocos);

- Z é uma matriz de delineamento com dimensão $n \times g$ associada com o vetor g de efeitos genotípicos com dimensão $g \times 1$;
- λ_k é o k-ésimo valor singular;
- α_k é o k-ésimo vetor singular referente ao genótipo;
- γ_k é o k-ésimo vetor singular referente ao ambiente;
- X_2 é a matriz de delineamento com dimensão $n \times 1$ associada à γ_k;
- ε é o vetor da distribuição de probabilidade residual.

2.5 Estrutura Fator Analítico (FA)

Em modelo misto, a componente aleatória g, pode ser modelada pelo modelo FA, que expressa o efeito aleatório do genótipo i no ambiente j como uma função linear de variáveis latentes f_{ik} com coeficientes τ_{jk}, para $k = 1, 2, ..., t$, além de um δ_{ij} residual que é modelado como $g_{ij} = \sum_{k=1}^{t} {}_{ik}\tau_{jk} + \delta_{ij}$. Nessa expressão, τ_{jk} representa a carga fatorial do j-ésimo ambiente no fator latente k, f_{ik} é o escore do genótipo no fator latente k e δ_{ij} é o termo residual. Na notação matricial a equação anterior é expressa como:

$$\boldsymbol{g} = (\boldsymbol{\tau}_1 \otimes \mathbf{I}_g)\mathbf{f}_1 + (\boldsymbol{\tau}_2 \otimes \mathbf{I}_g)\mathbf{f}_2 + ... + (\boldsymbol{\tau}_k \otimes \mathbf{I}_g)\mathbf{f}_k + \boldsymbol{\delta} \ , \tag{8}$$

em que:
o vetor $(\boldsymbol{\tau}_k \otimes \mathbf{I}_g)\mathbf{f}_k$ e o vetor $\boldsymbol{\delta}$ são da ordem $ge \times 1$.

Outra forma de apresentar essa equação é dada por:

$$\boldsymbol{g} = (\boldsymbol{\Gamma} \otimes \mathbf{I}_g)\mathbf{f} + \boldsymbol{\delta} \ , \tag{9}$$

em que:

$$\mathbf{\Gamma} = \begin{bmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1k} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \tau_{p1} & \tau_{p2} & \cdots & \tau_{kp} \end{bmatrix} \tag{10}$$

é uma matriz de ordem $p \times k$ com a k-ésima coluna contendo as cargas de ambiente para o k-ésimo fator latente. Desde que se assume que os genótipos não são relacionados com os efeitos aleatórios, $\mathbf{f}$ e $\boldsymbol{\delta}$ são independentes e têm uma distribuição normal conjunta com um vetor de média zero e variâncias $V(\mathbf{f}) = \mathbf{I}_k \otimes \mathbf{I}_g = \mathbf{I}_{kg}$ e $V(\boldsymbol{\delta}) = \mathbf{\Psi} \otimes \mathbf{I}_g$ (de ordem $pg \times pg$), respectivamente, em que $\mathbf{\Psi}$ é uma matriz diagonal $(\sigma^2_{\eta_1}, \sigma^2_{\eta_2}, ..., \sigma^2_{\eta_k})$ de ordem $p \times p$.

Logo, a matriz de variâncias e covariâncias do efeito aleatório g, o qual separa os componentes ambientais e genotípicos, é dada por:
$g = (\mathbf{\Gamma} \otimes \mathbf{I}_g)V(\mathbf{f})(\mathbf{\Gamma}' \otimes \mathbf{I}_g) + V(\boldsymbol{\delta}) = (\mathbf{\Gamma} \otimes \mathbf{I}_g)(I_k \otimes \mathbf{I}_g)(\mathbf{\Gamma}' \otimes \mathbf{I}_g) + (\mathbf{\Psi} \otimes \mathbf{I}_g)$

$$g = (\mathbf{\Gamma\Gamma}' + \mathbf{\Psi}) \otimes \mathbf{I}_g = FA(k) \otimes \mathbf{I}_g \ , \tag{11}$$

em que os elementos da matriz $\mathbf{\Gamma\Gamma}'$,

$$\mathbf{\Gamma\Gamma}' = \begin{bmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1k} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \tau_{p1} & \tau_{p2} & \cdots & \tau_{kp} \end{bmatrix} \begin{bmatrix} \tau_{11} & \tau_{21} & \cdots & \tau_{k1} \\ \tau_{12} & \tau_{22} & \cdots & \tau_{k2} \\ \vdots & \vdots & \cdots & \vdots \\ \tau_{1p} & \tau_{2p} & \cdots & \tau_{kp} \end{bmatrix} , \tag{12}$$

são estimativas da variância genética dentro do j-ésimo ambiente (ou seja, elementos da diagonal de σ^2_{gj}) e as estimativas das covariâncias genéticas entre j-ésimo e o j'-ésimo (ou seja, elementos fora da diagonal).

Assim, o modelo FA pode ser interpretado como uma regressão linear de G e GE em covariáveis latentes ambientais (cargas ambientais $\boldsymbol{\tau}_{jk}$), com cada genótipo tendo uma inclinação separada (escores genotípicos, $\mathbf{f}_{ik}$), mas um intercepto comum (se efeitos principais de genótipos não são distintos da GE). Os declives de genótipos medem a sensibilidade dos genótipos a

fatores ambientais hipotéticos representados pelas cargas de cada ambiente.

2.6 Análise de fatores sob modelos multiplicativos mistos

Como o seu nome indica, um modelo multiplicativo envolve o produto de: (i) um componente devido ao genótipo e (ii) um componente devido ao ambiente em que o genótipo é cultivado.

Modelos multiplicativos mais gerais permitem o encaixe da soma de vários termos multiplicativos, ao invés de apenas um termo multiplicativo, como na análise de estabilidade para a descrição de performance genotípicas em diferentes ambientes.

A análise de grupos de experimentos ou de experimentos conduzidos em dados MET tem sido tradicionalmente baseada em modelos simples, os quais assumem homogeneidade de variância residual entre os ensaios, independência de erros dentro de ensaio, efeitos da interação GE como um grupo de efeitos aleatórios independentes. A análise de dados de grupos de experimentos por meio de modelos realísticos é um problema estatístico complexo que demanda extensões ao modelo linear misto padrão.

No contexto dos modelos mistos, Piepho (1997) apresentou um modelo misto multiplicativo de FA com efeitos aleatórios de genótipo e de GE, o qual é conceitualmente e funcionalmente melhor que o AMMI (RESENDE e THOMPSON, 2004).

No mesmo contexto, Smith et al. (2001), apresentou uma classe geral de modelos FA que abrangem a abordagem de Piepho (1998), e inclui erros ambientais para cada ensaio. Esta classe geral de modelos propicia uma abordagem realística completa para análise de dados MET.

A seguir é apresentada uma extensão dos modelos mistos para incorporar a análise de fatores segundo Meyer (2009).

Modelo misto tradicional:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon} \; . \tag{13}$$

Modelo misto fator analítico (FA):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}[(\boldsymbol{\Gamma} \otimes \mathbf{I_g})\mathbf{f} + \boldsymbol{\delta}] + \boldsymbol{\varepsilon} \ , \tag{14}$$

$$\mathbf{g} = [(\Gamma \otimes \mathbf{I}_g)\mathbf{f} + \boldsymbol{\delta}] \ , \tag{15}$$

em que:
f- é o vetor de escores fatoriais para os indivíduos nos fatores;
$\boldsymbol{\delta}$- é o vetor de erros representando a falta de ajuste do modelo fatorial;
$\boldsymbol{\Gamma}$- é a matriz dos carregamentos dos fatores nas variáveis.

Sob esse modelo, a matriz de covariâncias genética é dada por $\boldsymbol{\Sigma} = \boldsymbol{\Gamma\Gamma}' + \boldsymbol{\Psi}$, em que $\boldsymbol{\Gamma\Gamma}' = \mathbf{V}\mathbf{D}_\alpha\mathbf{V}'$, $\mathbf{D}_\alpha$ é a matriz diagonal dos m autovalores e $\boldsymbol{V}$ é a matriz dos autovetores. Escolhendo-se $\boldsymbol{V}$ e $\boldsymbol{D_\alpha}$ referentes apenas à dimensão p, esse modelo misto é reduzido e ajusta somente os p fatores. Na técnica FA, a estrutura de covariâncias é simplificada para $\boldsymbol{\Sigma} = \boldsymbol{\Gamma_p\Gamma_p} + \boldsymbol{\Psi}'$, em que $\boldsymbol{\Psi}$- é a matriz diagonal de variâncias específicas $Var(\boldsymbol{\delta}_i)$.

A metodologia de modelos mistos padrão pode ser usada para estimar autovalores e autovetores diretamente sem a necessidade de se estimar $\boldsymbol{\Sigma}$ completa. A principal diferença para o modelo multivariado misto tradicional refere-se ao fato de os parâmetros a serem estimados fazerem parte da matriz de incidência dos efeitos genéticos aleatórios. Como $(\boldsymbol{\Gamma} \otimes \mathbf{I}_g)\mathbf{f}$ é singular, isso conduz à estimação sob posto reduzido, então restrições devem ser impostas aos parâmetros do modelo FA (MEYER, 2009; SMITH et al. 2001; RESENDE, 2007). Uma maior aplicação dos modelos FA é na análise de dados MET no estudo da interação GE, e tornaram-se populares por reunir em um só método a análise de adaptabilidade e estabilidade.

Uma característica fundamental do modelo de FA para os dados MET é a generalidade da estrutura de variância associada aos efeitos GE. O modelo de variância mais geral, e por conseguinte, o modelo que irá proporcionar o melhor ajuste (no sentido de probabilidade) para os dados, é uma matriz não-estruturada. Isto pode ser difícil de ajustar a partir de uma perspectiva

computacional, em particular para grandes estruturas (m grande para uma matriz na dimensão genótipo ou t grande para dimensão do ambiente). O modelo FA com termos multiplicativos suficientes tem sido encontrado para proporcionar uma parcimoniosa aproximação para a forma não estruturada e é geralmente mais robusto computacionalmente (SMITH et al. 2005).

Assim, Smith et al. (2001) utilizam um modelo de FA para aproximar uma matriz de covariâncias não estruturada para a dimensão do ambiente de $var(g)$, isto é, a matriz de variâncias e covariâncias entre ambientes.

2.7 Conceitos básicos de imputação de dados

A perda de dados é um problema frequente durante a realização de experimentos bem como durante a coleta de dados. Em estudos de melhoramento genético, muitas vezes os melhoristas se deparam com falta de genótipos em alguns ambientes, gerando conjunto de observações incompletas, dificultando deste modo o uso de técnicas multivariadas.

A presença de observações em falta dificulta a análise estatística destes dados, na medida em que quanto mais observações se perdem, mais será o aumento do viés nas estimativas dos parâmetros, causando desta maneira falsas inferências. Os métodos de imputação, também conhecidos como métodos de substituição, foram desenvolvidos com o propósito de resolver o problema de observações faltantes durante a análise dos dados.

A imputação de dados é uma técnica que permite substituir ou preencher os dados faltantes ou ausentes por meio de valores estimados determinados a partir de um conjunto de dados específicos. Rubin (1976) desenvolveu algumas técnicas de imputação simples:

i) imputação pela média dos dados disponíveis;

ii) imputação pela vizinhança do dado faltante;

iii) imputação pela regressão linear;

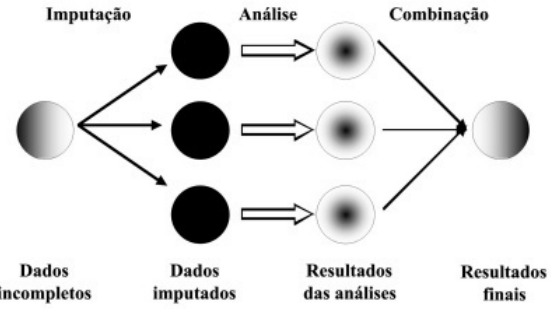
iv) imputação por meio da máxima verossimilhança.

Essas técnicas têm sido bastante usadas devido a sua facilidade de implementação, mas as estimativas produzidas por elas subestimam a variabilidade da variável imputada, aumentando o viés no desvio padrão dos

dados.

Rubin (1987) desenvolveu novos métodos que forneceram melhores estimativas (com menos viés) quando comparado com os métodos da imputação simples, a imputação múltipla. Desde a sua publicação, a imputação múltipla se tornou uma abordagem referencial devido as suas propriedades estatísticas. Ela ocorre quando para cada dado ausente são imputados vários valores, gerando igual número de bancos de dados completos. Em seguida, cada conjunto de dados é analisado separadamente usando técnicas estatísticas, gerando igual número de resultados. Esses são, por sua vez, combinados obtendo-se o resultado final, que é a estimativa pontual de um parâmetro que é obtido através da média das múltiplas imputações e o seu erro padrão é obtido através da variância das múltiplas imputações (Figura 2).

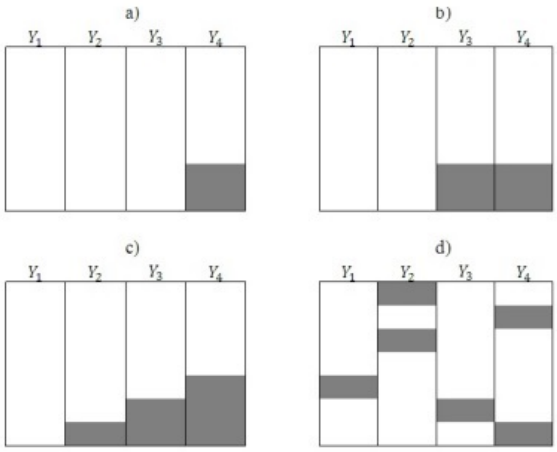
Figura 2 – Esquema de imputação múltipla
(fonte: http://www.scielosp.org/pdf/rbepid/v13n4/05.pdf)



2.8 Padrões de dados ausentes

Verificando os padrões de comportamento dos dados ausentes, pode-se identificar a forma com que as unidades ausentes estão distribuídas em um vetor de matriz de dados, descrevendo a localização dos valores em falta (DIAS et al. 2014). A Figura 3 ilustra algumas formas de se representar unidades ausentes em um conjunto de dados, definindo deste modo um certo padrão.

Figura 3 – Alguns padrões de comportamento de dados (fonte: www.teses.usp.br/teses/disponiveis/11/11134/.../Maria$_J$oseane$_C$ruz$_d$a$_S$ilva.pdf)



a) Padrão univariado (*univariate poattern*) - apresenta uma falta de dados isoladamente em uma variável, o que é comum em estudos experimentais;

b) Padrão de não resposta (*unit nonresponse pattern*) - geralmente ocorre em pesquisas realizadas por meio de questionários como o censo. Quando alguns itens não são respondidos ou preenchidos causam valores em falta para questionários;

c) Padrão monótono (*monotone pattern*) - geralmente ocorre quando os indivíduos participantes da pesquisa em algum momento não podem continuar no estudo devido a alguns fatores. Esse tipo de padrão de dados é característico de experimentos longitudinais, ou seja, quando as variáveis são medidas ao longo do tempo;

d) Padrão geral (*general pattern*) - geralmente ocorre quando se tem dispersão arbitrária de unidades ausentes por toda a matriz de dados. Aparentemente é aleatório, porém pode ou não existir uma relação entre a falta de valores de uma variável e a tendência da falta de dados referente a outra variável medida.

3 MATERIAL E MÉTODOS

Nesta seção, são apresentados os materiais experimentais utilizados para a realização do estudo e os métodos de análises utilizados.

3.1 Dados simulados

Para a realização desse estudo foi simulado um conjunto de dados com 7 ambientes (E) e 20 genótipos (G), dos quais: i) 5 genótipos com interação positiva com todos os ambientes instáveis; ii) 5 genótipos com interação negativa com todos os ambientes instáveis; iii) 10 genótipos com interação positiva e negativa com todos os ambientes estáveis (para ilustrar o padrão de resposta, foi produzido o mapa de calor que pode ser visto na figura 1 do Apendeci). Assumindo um modelo estatístico com distribuição normal de $\mu = 15$ e variância heterogênea, usando um delineamento em blocos completos ao acaso com 3 repetições, em que:

- $\varepsilon \sim N(0, R) \quad R = diag(\sigma^2_{e_1}, ..., \sigma^2_{e_k})$;
- y$\sim N(\mu, V) \quad V = ZGZ^{'} + R$;

3.2 Métodos

A análise dos dados foi feita usando os modelos EM-AMMI F e EM-AMMI M (*Additive main effects and multiplicative interactions- expectation maximization*) descrito por Gauch e Zobel (1990), AMMI Bayesiano com homogeneidade de variância - AMMIB-I descrito por Oliveira et al. (2015), Liu (2001) e Crossa et al. (2011), AMMI Bayesiano com heterogeneidade de variâncias - AMMIB-D descrito por Silva et al. (2016) e FA (*factor analytic multiplicative mixed models*) descrito por Smith et al. (2001).

Para verificar a eficiência desses métodos foram feitos desbalanceamentos aleatórios nos dados, que consistiu na perda total do genótipo no ambiente, e em seguida, foram efetuadas as análises.

No conjunto de dados simulados foram feitos desbalanceamentos nos dados, com níveis de 10%, 33% e 50% de perda, para avaliar a capacidade

preditiva de dados faltantes nos modelos propostos.

As análises foram feitas usando o *software* R. Os resultados obtidos pelos diferentes métodos foram comparados entre si. Para avaliar a capacidade preditiva dos modelos foi usada a estatística PRESS (*prediction error sum square*) e realizada a correlação entre o valor predito e observado usando a validação cruzada.

3.2.1 Modelo EM-AMMI

O ajuste do EM-AMMI foi feito de acordo com o algoritmo proposto por Paderewski e Rodrigues (2014), que envolve 5 etapas (como mostrado na Figura 1), em que: i) O pesquisador fornece os valores iniciais para as células faltantes; ii) Estimam-se os parâmetros dos modelos; iii) Ajustam-se as médias com o modelo AMMI-n; iv) Preenche-se as células com as médias ajustadas; v) Se a convergência for alcançada, estabelecida a distância de Chebyshev o algoritmo para, caso contrário repetem-se os passos ii) a v).

Uma aplicação adequada do algoritmo EM para o EM-AMMI funciona da seguinte maneira. Primeiro, calcula-se as médias de todas as células com valores presentes. Em seguida, inicializam-se os parâmetros aditivos do EM-AMMI calculando as médias não ponderadas dos genótipos pela média do ambiente e média geral. Depois, inicializam-se os resíduos de interação, para as células com dados presentes (ou seja, a interação é igual a média da célula, subtraída as médias do genótipo e do ambiente, adicionada a média geral), e imputa-se interações residuais zero para células com dados em falta.

Desta maneira, a matriz de interação não tem células não especificadas, de modo que se calcula o eixo do componente principal (PC) perfeitamente para parâmetros multiplicativos do EM-AMMI, continuando para o maior número de eixos PC conforme desejado. Em seguida, reestima-se e revê-se cada célula em falta com o atual modelo EM-AMMI. Posto isso, encaixa-se EM-AMMI a esses dados corrigidos, o tratamento de valores imputados, o mesmo faz-se para os dados reais.

Este processo é repetido de forma iterativa até que haja convergência, isto é, até que os valores imputados para as células ausentes tenham aceita-

velmente pequenas alterações. Após a convergência, o modelo EM-AMMI ajusta as células imputadas perfeitamente com um resíduo zero (no limite da precisão numérica). Daí o algoritmo EM ajusta o modelo, ignorando as células ausentes no sentido de que eles recebem valores imputados que se adequem no modelo perfeitamente.

3.2.2 Modelo Fatorial Analítico

O modelo de dados MET que considera uma série de m genótipos (não necessariamente em todos os ensaios) medidos em cada um dos vários p ambientes pode ser representado pela seguinte expressão:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zg} + \boldsymbol{\varepsilon} \ , \tag{16}$$

em que:
$\boldsymbol{y}$ é o vetor de observações;
$\boldsymbol{\beta}$ vetor dos efeitos fixos, com a matriz de incidência $\boldsymbol{X}$;
$\boldsymbol{g}$ vetor dos efeitos aleatorios, com a matriz de incidência $\boldsymbol{Z}$;
$\boldsymbol{\varepsilon}$ vetor de resíduos do modelo.

Assumimos que a distribuição conjunta de $\boldsymbol{g}$ e $\boldsymbol{\varepsilon}$ é:

$$\begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{g} \end{pmatrix} \sim N \left(\boldsymbol{0}, \ \begin{pmatrix} \boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G} \end{pmatrix} \right) \ , \tag{17}$$

sendo $\boldsymbol{R}$ e $\boldsymbol{G}$ matrizes de variâncias e covariâncias $(p \times p)$ de resíduos do modelo e dos efeitos genéticos aditivos, respectivamente, e $\boldsymbol{I}$ é a matriz identidade $(m \times m)$. A distribuição de $\boldsymbol{y}$ é dada por $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{V})$, em que $\boldsymbol{V} = \boldsymbol{ZGZ}' + \boldsymbol{R}$.

Os efeitos aleatórios do vetor $\boldsymbol{g}'$ são praticamente os vetores de efeitos genéticos para m variedades em p ambientes. Pode ser considerado como uma matriz bidimensional de efeitos, ordenada como variedades dentro de ambientes.

A estrutura de variâncias associada é assumida separada e é apresentada como um produto direto do $\boldsymbol{G}_e$ e $\boldsymbol{G}_v$, ou seja, $\boldsymbol{V}(\boldsymbol{g} = \boldsymbol{G}_e \otimes \boldsymbol{G}_v)$, em que

$\boldsymbol{G}_e$ e $\boldsymbol{G}_v$, são as matrizes simétricas para ambientes e genótipos. Supõe-se também que $\boldsymbol{G}_v = \boldsymbol{I}_m$, embora uma matriz de parentesco poderia ser construída e usada como uma estrutura alternativa (OAKEY et al. 2006). A matriz de variâncias genética da $\boldsymbol{G}_e = \{\sigma_{jj}\}$ tem elementos diagonais que são as variâncias genéticas para ambientes individuais e os elementos fora da diagonal são as covariâncias genéticas entre pares de ambientes. Nós formulamos esta matriz como uma estrutura de variâncias FA, $\boldsymbol{\Gamma\Gamma}^{'} + \boldsymbol{\Psi}$.

Os melhores estimadores lineares não viesados (BLUE) para os efeitos fixos e os melhores preditores lineares não viesados (BLUPS) para os efeitos aleatórios na Eq. (16) são, respectivamente:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{'}\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}\boldsymbol{V}^{-1}\boldsymbol{y} \tag{18}$$

e

$$\hat{\boldsymbol{g}} = \boldsymbol{GZV}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}). \tag{19}$$

Uma referência mais precisa com as equações acima serão BLUES e BLUPs empíricos, devido ao fato de os parâmetros em $\boldsymbol{V}$ serem desconhecidos e substituídos pelas suas estimativas REML. Os procedimentos para os cálculos dos parâmetros $\boldsymbol{\Gamma}$ e $\boldsymbol{\Psi}$ do modelo FA e, consequentemente, para os cálculos dos BLUPs dos escores genotípicos $\hat{\boldsymbol{f}}$ podem ser encontrados em Smith et al. (2001).

3.2.2.1 Estimação

Os efeitos fixos (BLUEs) e os efeitos aleatórios (BLUPs) são estimados por:
$\hat{\boldsymbol{g}}$=$(\boldsymbol{G}_e \otimes \boldsymbol{I}_m)\boldsymbol{Z}^{'}\boldsymbol{Py}$,
em que:
$\boldsymbol{V} = Var(y) = \boldsymbol{Z}_0\boldsymbol{G}_0\boldsymbol{Z}_0^{'} + \boldsymbol{Z}(\boldsymbol{G}_e \otimes \boldsymbol{I}_m)\boldsymbol{Z}^{'} + \boldsymbol{R}$,
e
$\boldsymbol{P} = \boldsymbol{V}^{-1} + \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^{'}\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}\boldsymbol{V}^{-1}$.

No modelo FA, BLUPs dos escores genotípicos $\hat{\boldsymbol{f}}$ e resíduais $\hat{\boldsymbol{\delta}}$ são obtidos como função de $\hat{\boldsymbol{g}}$ como:

$\hat{\boldsymbol{f}} = Var(\boldsymbol{f}[\boldsymbol{Z}(\boldsymbol{\Gamma} \otimes \boldsymbol{I}_m)]' \boldsymbol{P}\boldsymbol{y} = [\boldsymbol{\Gamma}'(\boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})^{-1} \otimes \boldsymbol{I}_m]\hat{\boldsymbol{g}}$,

e

$\hat{\boldsymbol{\delta}} = [\boldsymbol{\Psi}(\boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})^{-1} \otimes \boldsymbol{I}_m]\hat{\boldsymbol{g}}$.

O cálculo da porcentagem de variância genética explicada por k fatores, tanto para ambientes individuais (denotado v_j) e geral (denotado $\bar{v}$), é dado por:

$v_j = \frac{\sum_{j=1}^{k} \hat{\lambda}_{rj}^2}{\sum_{j=1}^{k} \hat{\lambda}_{rj}^2 + \hat{\psi}} \times 100$ e $\bar{v} = \frac{tr(\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Gamma}}')}{tr(\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Gamma}}' + \hat{\boldsymbol{\Psi}})} \times 100$.

O modelo FA pode ser escolhido com base na porcentagem global e na distribuição de valores individuais em cada ambiente, uma vez que é desejável que o modelo escolhido tenha alguns ambientes com valores baixos e muitos ambientes com valores elevados.

3.2.2.2 Ajuste dos modelos FA

O modelo FA foi ajustado usando o pacote ASReml-R (BUTLER et al. 2009) dentro do R (R CORE TEAM, 2016). A variância e parâmetros do modelo misto da Equação (16) foram estimados usando máxima verossimilhança restrita (REML).

No modelo FA, os parâmetros de variância são as cargas e as variâncias específicas, e as estimativas REML destes denotados por $\hat{\lambda_{rj}}$ e $\hat{\psi}$ $(r = 1, .., k; j = 1, ..., p)$. Nota-se que quando $K > 1$, a matriz de cargas não é única e para sua estimativa são necessárias restrições. O algoritmo no ASReml-R (Butler et a.l 2009) corrige todos os $k(k-1)/2$ elementos do triângulo superior de $\boldsymbol{\Gamma}$ a zero. Uma vez que uma estimativa $\boldsymbol{\Gamma}$ foi obtida, a matriz pode ser rotacionada como desejado para fins de interpretação.

Obtidas as estimativas de todos os componentes de variância, obtemos estimadores empíricos dos efeitos fixos (EBLUES) e melhores preditores dos efeitos aleatórios (EBLUPs). Em termos do modelo FA denotamos os EBLUPs dos escores fatoriais e resíduos da regressão genética por $\hat{f}_{rj}$ e $r\hat{\delta}_{jr}$ $(r = 1, ..., k; j = 1, ..., p)$.

O processo de ajuste do modelo inicia-se com o ajuste de um modelo

FA1, e em seguida, prossegue-se para modelos de ordem superior, conforme necessário. Uma ordem apropriada pode ser determinada usando o teste de razão de máxima verossimilhança restrita (LRTREML) que compara sequências de modelos FA base.

3.2.3 Modelo AMMI-Bayesiano com homogeneidade de variâncias (AMMIB-I)

O modelo AMMIB usado foi apresentado por Liu (2001) e discutido por Oliveira et al. (2015). Esses autores afirmam que, para uma análise bayesiana do modelo AMMI, é necessário especificar as distribuições a priori para os parâmetros, as condicionais completas e o processo de amostragem.

3.2.3.1 Distribuições a priori para os parâmetros do modelo

Considerando o modelo:

$$y = X_1\beta + Zg + \sum_{k=1}^{t} \lambda_k diag(Z\alpha_k)X_2\gamma_k + \varepsilon \ , \tag{20}$$

as distribuições a priori para o mesmo são:

- $\boldsymbol{\beta}|\mu_\beta, \sigma^2_\beta \sim N(\mu_\beta, I\sigma^2_\beta)$, considerando variância infinita ($\sigma^2_\beta \to \infty$), têm-se $\boldsymbol{\beta} \sim$ constante;
- $\mathrm{g}|\mu_g, \sigma^2_g \sim N(0, I\sigma^2_g)$;
- $\lambda_k|\mu_{\lambda_k}, \sigma^2_{\lambda_k} \sim N^+(\mu_{\lambda_k}, \sigma^2_{\lambda_k})$, considerando $\sigma^2_{\lambda_k} \to \infty$, têm-se $\lambda_k|\mu_{\lambda_k}, \sigma^2_{\lambda_k} \sim$ constante;
- α_k e $\gamma_k \sim$ uniforme esférica no subespaço corrigido;
- $\sigma^2_g \sim inv{-}\chi^2(v_g, S^2_g)$ e considerando o valor zero para o grau de liberdade e parâmetro de escala a densidade a priori reduz-se a $(\sigma^2_g)^{-1}$;
- $\sigma^2_\varepsilon \sim inv - \chi^2(v_e, S^2_e)$, de forma análoga à considerada para variância genotípica, tem-se $(\sigma^2_\varepsilon)^{-1}$.

Para os vetores β e g são atribuídas distribuições a priori normais multivariadas. Para o vetor β considera-se variância infinita, o que equivale assumir um conhecimento vago a respeito dos parâmetros que compõem este vetor (indicado pela constante). O efeito aleatório para genótipos é ob-

tido atribuindo-se uma priori hierárquica para g em que σ_g^2 tem distribuição proporcional a $(\sigma_g^2)^{-1}$, resultado da atribuição do valor zero aos hiperparâmetros (v_g, S_g^2) para a densidade qui-quadrado escalada invertida. Devido ao fato da priori atribuída a σ_g^2 ser não informativa, a incerteza em relação estimação de g é determinada basicamente pela função de verossimilhança, ou seja, a partir dos dados experimentais.

Para os vetores singulares α_k e γ_k, são atribuídas distribuições esféricas uniformes no subespaço corrigido (VIELLE e SRNIVASAN, 2000). Estes vetores são distribuídos em um espaço restrito em R^p, sendo ortogonais a $t-1$ vetores no espaço de dimensão p ($p = r$ ou $p = c$, respectivamente). A distribuição uniforme esférica é uma distribuição não informativa. Para os valores singulares são atribuídas distribuições a priori normais truncadas, pelo fato dessas variáveis assumirem somente valores positivos e $\lambda_k \geq \lambda_{k+1}$.

3.2.3.2 Distribuições condicionais completas a posteriori

Combinando as informações referentes aos dados (função de verossimilhança) com as densidades à priori, como estabelecidas acima, através do teorema Bayes, a densidade da distribuição conjunta a posteriori pode ser encontrada. Consideremos inicialmente a função de verossimilhança:

$$L(y|\theta) = \frac{1}{(2\pi)^{\frac{n}{2}}|I\sigma_e^2|^{\frac{1}{2}}} exp\{-\frac{1}{2\sigma_e^2}(y-\theta)'(y-\theta)\} \; . \tag{21}$$

em que:
$\theta = X_1\beta + Zg + \sum_{k=1}^{t} \lambda_k diag(Z\alpha_k)X_2\gamma_k$

Aplicando o teorema de Bayes obtém-se a distribuição conjunta a posteriori para os parâmetros:

$$p(\theta|y) \propto p(y|\theta,\sigma_e^2)p(g|\mu_g,\sigma_g^2)p(\beta|\mu_\beta,\sigma_\beta^2)p(\sigma_g^2|v_g,S_g^2)p(\sigma_e^2|v_e,S_e^2)\times \tag{22}$$

$\times \prod_{k=1}^{t} p(\lambda_k|\mu_{\lambda_k},\sigma_{\lambda_k}^2)p(\alpha_k)p(\gamma_k)$,
em que:
$\theta = (\alpha, \gamma, \lambda, g, \beta, \sigma_g^2, \sigma_e^2)$.

A distribuição conjunta a posteriori pode ainda ser reescrita como:

$$p(\theta|y) \propto (\sigma_e^2)^{-\frac{n}{2}} exp\{-\frac{1}{2\sigma_e^2}(y-\theta)^{'}(y-\theta)\}(\sigma_e^2)^{-\frac{n_g}{2}} exp\{-\frac{1}{2\sigma_g^2}g^{'}g\}(\sigma_g^2)^{-1}(\sigma_e^2)^{-1}. \tag{23}$$

As distribuições condicionais completas para os parâmetros são obtidas da distribuição conjunta à posteriori. Por meio de manipulações algébricas, completando quadrados e ainda observando termos que são constantes em relação ao parâmetro considerado (estes termos podem ser absorvidos pela constante de normalização), obtêm-se:
$p(\beta|outros) \propto exp\{-\frac{1}{2\sigma_e^2}(\beta-(X^{'}X)^{-1}(X^{'}\Psi)^{'}X^{'}X(\beta-(X^{'}X)^{-1})X^{'}\Psi)\}$

$$\beta|... \sim N[(X^{'}X)^{-1}X^{'}(y-Zg-\Theta),(X^{'}X)^{-1}\sigma_e^2] \ , \tag{24}$$

em que:

- $\Psi = y - Zg - \sum_{k=1}^{t} \lambda_k diag(Z\alpha_k)X\gamma_k$ e $\Theta = \sum_{k=1}^{t} \lambda_k diag(Z\alpha_k)X\gamma_k$.
- p(g|outros)$\propto exp\{-\frac{1}{2\sigma_e^2}[(g-(Z^{'}Z+I\frac{\sigma_e^2}{\sigma_g^2})^{-1}Z^{'}\Gamma)^{'}(Z^{'}Z+I\frac{\sigma_e^2}{\sigma_g^2})$
$(g-(Z^{'}Z+I\frac{\sigma_e^2}{\sigma_g^2})^{-1}Z^{'}\Gamma))]\}$.

Considerando $\Gamma = y - X\beta - \sum_{k=1}^{t} \lambda_k diag(Z\alpha_k)X\gamma_k$ e desenvolvendo a expressão dentro dos colchetes tem-se:

$$g|... \sim N[(Z^{'}Z+I\frac{\sigma_e^2}{\sigma_g^2})^{-1}Z^{'}(y-X\beta-\Theta),(Z^{'}Z+I\frac{\sigma_e^2}{\sigma_g^2})^{-1}\sigma_e^2] \ , \tag{25}$$

$p(\sigma_e^2|outros) \propto (\sigma_e^2)^{-(\frac{n}{2}+1)} \cdot exp\{-\frac{1}{2\sigma_e^2}(y-\theta)^{'}(y-\theta)\}$

$$\sigma_e^2|... \sim inv-\chi^2[n,(y-\theta)^{'}(y-\theta)] \ , \tag{26}$$

$p(\sigma_g^2|outros) \propto (\sigma_g^2)^{-(\frac{n_g}{2}+1)} \cdot exp\{-\frac{1}{2\sigma_g^2}g^{'}g\}$

$$\sigma_g^2|... \sim inv - \chi^2[n_g, g^{'}g] \ , \tag{27}$$

$p(\lambda_k|outros) \propto exp\{-\frac{1}{2\sigma_e^2}[(\lambda_k - (\phi^{'}\phi)^{-1}\phi^{'}\Delta)(\phi^{'}\phi)(\lambda_k - (\phi^{'}\phi)^{-1}\phi^{'}\Delta)]\}$,
em que:
$\Delta = y - X\beta - Zg - \sum_{k\neq k'}^{t} \lambda_{k'} diag(Z\alpha_{k'})X\gamma_{k'}$ e $\phi = diag(Z\alpha_k)X\gamma_k$.
temos então:

$$\lambda_k|... \sim N^{+}[(\phi^{'}\phi)^{-1}\phi^{'}\Delta), (\phi^{'}\phi)^{-1}\sigma_e^2] \ , \tag{28}$$

em que:
$p(\alpha|outros) \propto exp\{-\frac{1}{2\sigma_e^2}(y - X\beta - Zg - \Lambda\alpha_k - E)^{'}(y - X\beta - Zg - \Lambda\alpha_k - E)\}$
com:
$\Lambda = diag(X\gamma_k)Z$ e $E = \sum_{k\neq k'}^{t} \lambda_{k'} diag(X\gamma_{k'})Z\alpha_{k'}$
$p(\alpha_k|outros) \propto exp\{\frac{\lambda_k}{\sigma_e^2}[\alpha_k^{'}\Lambda^{'}(y - X\beta - Zg)]\}$
$p(\alpha_k|outros) \propto exp\{k\alpha_k^{'}\Lambda^{'}(y - X\beta - Zg)]\}$.

Assim, α_k possui distribuição proporcional a von-Mises Fisher com parâmetros $k = \frac{\lambda_k}{\sigma_e^2}$ e μ_{α_k} e pode ser representado por:

$$\alpha_k|... \sim VMF[k, \mu_{\alpha_k}] \ , \tag{29}$$

sendo:
$\mu_{\alpha_k} = \Lambda^{'}(y - X\beta - Zg)$.

A densidade condicional a posteriori para γ_k é dada por:

$\gamma_k|... \sim VMF[k, \mu_{\gamma_k}]$,

com cálculos análogos aos feitos para α_k, em que $\mu_{\gamma_k} = \Omega^{'}(y - X\beta - Zg)$ e $\Omega = diag(Z\alpha_k)X$.

3.2.4 Modelo AMMI com heterogeneidade de variâncias (AMMIB-D)

Para esse modelo foram assumidas as mesmas priores assumidas para o modelo AMMIB com homogeneidade de variâncias, com a diferença de que neste modelo assumiram-se variâncias heterogêneas, ao qual foi atribuída uma priori inversa qui-quadrada escalada invertida, para cada parâmetro de variância.

Considerando as informações disponíveis, a verossimilhança é dada por:

$$L(y|\theta) = \frac{1}{(2\pi)^{\frac{n}{2}}|V|^{\frac{n}{2}}} exp\{-\frac{1}{2}(y-\theta)^{'}V^{-1}(y-\theta)\} \ , \tag{30}$$

em que:
$\theta = X_1\beta + Zg + \sum_{k=1}^{p} \lambda_k diag(Z\alpha_k)X_2\gamma_k$.

■ Distribuições condicionais completas a posteriori para os parâmetros
$P(\Theta|y) \propto L(\theta|y)P(\beta|\mu_\beta, \sigma^2_\beta)P(g|\mu_g, \sigma^2_g)P(V)P(\sigma^2_g|v_g, S^2_g)\times$

$$\times \prod_{k=1}^{p} P(\lambda_k|\mu_{_k}, \sigma^2_{\lambda_k})P(\alpha_k)P(\gamma_k) \ . \tag{31}$$

■ Distribuições condicionais completas a posteriori para $\boldsymbol{\beta}$

$$P(\beta|...) \propto esp\{-\frac{1}{2}(y-\theta)^{'}V^{-1}(y-\theta)\} \ , \tag{32}$$

$A_1 = y - Zg\sum \lambda_k diag(Z\alpha_k)X_2\gamma_k$,
$(A_1 - X_1\beta)^{'}V^{-1}(A_1 - X_1\beta) = A_1^{'}V^{-1}A_1 - A_1^{'}V^{-1}X_1\beta - \beta^{'}X_1^{'}V^{-1}A_1 +$
$\beta^{'}X_1^{'}V^{-1}X_1\beta$,
$\propto exp\{-\frac{1}{2}(-2\beta^{'}X_1^{'}V^{-1}A_1 + \beta^{'}X_1^{'}V^{-1}X_1\beta)$,
$\propto exp\{-\frac{1}{2}(-2\beta - (X_1^{'}V^{-1}X_1)^{-1}X_1^{'}V^{-1}A_1)^{'}(X_1^{'}V^{-1}X_1)(\beta - (X_1^{'}V^{-1}X_1)^{-1}X_1^{'}V^{-1}A_1)$,

$$\beta|... \sim N[(X_1^{'}V^{-1}X_1)^{-1}X_1^{'}V^{-1}A_1; (X_1^{'}V^{-1}X_1)] \ . \tag{33}$$

■ Distribuições condicionais completas a posteriori para $\boldsymbol{g}$

$$P(g|...) \propto exp\{-\frac{1}{2}(y-\theta)^{'}V^{-1}(y-\theta)\}exp\{-\frac{1}{2\sigma_g^2}g^{'}I_g g\} \ , \qquad (34)$$

$A_2 = y - X_1\beta - \sum \lambda_k diag(Z\alpha_k)X_2\gamma_k$,
$P(g|...) \propto exp\{-\frac{1}{2}(y-\theta)^{'}V^{-1}(y-\theta)\}exp\{-\frac{1}{2\sigma_g^2}g^{'}I_g g\}$,
$P(g|...) \propto exp\{-\frac{1}{2}[(A_2 - Zg)^{'}V^{-1}(A_2 - Zg) + \frac{1}{\sigma_g^2}(g^{'}I_g g)]\}$,
$P(g|...) \propto exp\{-\frac{1}{2}[A_2^{'}V^{-1}A_2 - A_2^{'}V^{-1}Zg - g^{'}Z^{'}V^{-1}A_2 + g^{'}Z^{'}V^{-1}Zg + g^{'}\frac{1}{\sigma_g^2}I_g g]\}$,
$\propto exp\{-\frac{1}{2}[-2g^{'}Z^{'}V^{-1}A_2 + g^{'}(Z^{'}V^{-1}Z + \frac{1}{\sigma_g^2}I_g)g]\}$,
$\propto exp\{-\frac{1}{2}[[g-(Z^{'}V^{-1}Z+\frac{1}{\sigma_g^2}I_g)^{-1}]Z^{'}V^{-1}A_2]^{'}(Z^{'}V^{-1}Z+\frac{1}{\sigma_g^2}I_g)[g-(Z^{'}V^{-1}Z+$
$\frac{1}{\sigma_g^2}I_g)^{-1}Z^{'}V^{-1}A_2]]\}$,

$$g|... \sim N[(Z^{'}V^{-1}Z + \frac{1}{\sigma_g^2}I_g)^{-1}Z^{'}V^{-1}A_2; (Z^{'}V^{-1}Z + \frac{1}{\sigma_g^2}I_g)^{-1}] \ . \qquad (35)$$

■ Distribuições condicionais completas a posteriori para $\boldsymbol{\sigma}_g^2$

$$P(\sigma_g^2|...) \propto (\sigma_g^2)^{-\frac{n_g}{2}} exp\{-\frac{1}{2\sigma_g^2}g^{'}I_g g\}(\sigma_g^2)^{-1} \ , \qquad (36)$$

$\propto (\sigma_g^2)^{-(\frac{n_g}{2}+1)}exp\{-\frac{1}{2\sigma_g^2}g^{'}I_g g\}$,

$$\sigma|... \sim Inv - Esc - \chi^2[n_g, g^{'}I_g g] \ . \qquad (37)$$

■ Distribuições condicionais completas a posteriori para $\boldsymbol{\lambda}_k$

$$P(\lambda_k|...) \propto exp\{-\frac{1}{2}(y-\theta)^{'}V^{-1}(y-\theta)\} \ , \qquad (38)$$

$A_4 = y - X_1\beta - Zg - \sum_{k^{'}\neq K}^{p} \lambda_{k^{'}} diag(Z\alpha_{k^{'}})X_2\gamma_{k^{'}}$,
$\phi = diag(Z\alpha_k)X_2\gamma_k$,
$P(\lambda_k|...) \propto exp\{-\frac{1}{2}(A_4 - \lambda_k\phi)^{'}V^{-1}(A_4 - \lambda_k\phi)\}$,
$\propto exp\{-\frac{1}{2}(A_4^{'}V^{-1}A_4 - \lambda_k A_4^{'}V^{-1}\phi - \lambda_k\phi^{'}V^{-1}A_4 - \lambda_k^2\phi^{'}V^{-1}\phi)\}$,

$\propto exp\{-\frac{1}{2}(-2\lambda_k\phi^{'}V^{-1}A_4 - \lambda_k^2\phi^{'}V^{-1}\phi)\}$,
$\propto exp\{-\frac{1}{2}([\lambda_k-(\phi^{'}V^{-1}\phi)^{-1}\phi V^{-1}A_4]^{'}(\phi^{'}V^{-1}\phi)[\lambda_k-(\phi^{'}V^{-1}\phi)^{-1}\phi V^{-1}A_4])\}$,

$$\lambda_k|... \sim N^{+}([\lambda_k - (\phi^{'}V^{-1}\phi)^{-1}\phi^{'}V^{-1}A_4]; (\phi^{'}V^{-1}\phi)^{-1}) \ . \tag{39}$$

■ Distribuições condicionais completas a posteriori para $\boldsymbol{\alpha}_k$

$$P(\alpha_k|...) \propto \{-\frac{1}{2}(y-\theta)^{'}V^{-1}(y-\theta)\} \ , \tag{40}$$

$\Delta_1 = \lambda_k diag(X_1\gamma_k)Z$,
$\propto exp\{-\frac{1}{2}(A_4 - \Delta_1\alpha_k)^{'}V^{-1}(A_4 - \Delta_1\alpha_k)\}$,
$(A_4 - \Delta_1\alpha_k)^{'}V^{-1}(A_4 - \Delta_1\alpha_k) = A_4^{'}V^{-1}A_4 - A_4^{'}V^{-1}\Delta_1\alpha_k - \alpha_k^{'}\Delta_1^{'}V^{-1}A_4 +$
$\alpha_k^{'}\Delta_1^{'}V^{-1}\Delta_1\alpha_k$,
$\propto exp\{-\frac{1}{2}[-2\alpha_k^{'}\Delta_1^{'}V^{-1}A_4 + \alpha_k^{'}\Delta_1^{'}V^{-1}\Delta_1\alpha_k]\}$,
$\propto \{-\frac{1}{2}[\alpha_k-(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]^{'}(\Delta_1^{'}V^{-1}\Delta_1)[\alpha_k-(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V_4^{-1}]\}$,

$$\alpha|... \sim N([(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]; (\Delta_1^{'}V^{-1}\Delta_1)^{-1}) \ . \tag{41}$$

■ Distribuições condicionais completas a posteriori para $\boldsymbol{\gamma}_k$

$$P(\gamma_k|...) \propto exp\{-\frac{1}{2}(y-\theta)^{'}V^{-1}(y-\theta)\} \ , \tag{42}$$

$\Delta_2 = \lambda_k diag(Z\alpha_k)X_2$,
$\propto exp\{-\frac{1}{2}(A_4 - \Delta_2\gamma_k)^{'}V^{-1}(A_4 - \Delta_2\gamma_k)\}$,
$(A_4 - \Delta_2\gamma_k)^{'}V^{-1}(A_4 - \Delta_2\gamma_k) = A_4^{'}V^{-1}A_4 - A_4^{'}V^{-1}\Delta_2\gamma_k - \gamma_k^{'}\Delta_2^{'}V^{-1}A_k +$
$\gamma_k^{'}\Delta_2^{'}V^{-1}\Delta_2\gamma_k$,
$\propto exp\{-\frac{1}{2}[-2\gamma_k^{'}\Delta_2^{'}V^{-1}A_4 + \gamma_k^{'}\Delta_2^{'}V^{-1}\Delta_2\gamma_k]\}$,
$\propto exp\{-\frac{1}{2}[\gamma_k-(\Delta_2^{'}V^{-1}\Delta_2)^{-1}\Delta_2^{'}V_4^{-1}]^{'}(\Delta_2^{'}V^{-1}\Delta_2)[\alpha_k-(\Delta_2^{'}V^{-1}\Delta_2)^{-1}\Delta_2^{'}V^{-1}A_4]\}$,

$$\gamma_k|... \sim N([(\Delta_2^{'}V^{-1}\Delta_2)^{-1}\Delta_2^{'}V^{-1}A_4]; (\Delta_2^{'}V^{-1}\Delta_2)^{-1}) \ . \tag{43}$$

■ Distribuições condicionais completas a posteriori para $\boldsymbol{V}(\sigma_{e_k}^2)$

A matriz **V** apresenta uma estrutura diagonal, não sendo possível amostrar diretamente, então procurou-se amostrar cada elemento da matriz.
$P(\sigma^2_{e_k}|...) \propto (\sigma^2_e)^{-\frac{n_e}{2}} exp\{-\frac{1}{2\sigma^2_e}(y_e-\theta_e)^{'}(y_e-\theta_e)\}(\sigma^2_e)^{-1}$,
$P(\sigma^2_{e_k}|...) \propto (\sigma^2_e)^{-(\frac{n_e}{2}+1)} exp\{-\frac{1}{2\sigma^2_e}(y_e-\theta_e)^{'}(y_e-\theta_e)\}$,

$$\sigma^2_{e_k}|... \sim Inv-Esc-\chi^2[n_e,(y_e-\theta_e)^{'}(y_e-\theta_e)] \ . \tag{44}$$

Amostra no espaço correto para os vetores singulares:

$$\alpha_k|... \sim N([(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4];(\Delta_1^{'}V^{-1}\Delta_1)^{-1}) \ , \tag{45}$$

$$\propto exp\{-\tfrac{1}{2}[\alpha_k-(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]^{'}(\Delta_1^{'}V^{-1}\Delta_1)[\alpha_k-(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]\} \ ,$$

Desenvolvendo o produto temos:
$\alpha_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)\alpha_k - \alpha_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]-$
$-[(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]^{'}(\Delta_1^{'}V^{-1}\Delta_1)\alpha_k+$
$+[(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]^{'}(\Delta_1^{'}V^{-1}\Delta_1)(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4$.
1)
$\alpha_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)\alpha_k = \alpha_k^{'}H_kH_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)H_kH_k^{'}\alpha_k =$
$= (H_k^{'}\alpha_k)^{'}H_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)H_kH_k^{'}\alpha_k = \widetilde{\alpha}_k^{'}H_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)H_k\widetilde{\alpha}$,
sendo:
$\widetilde{\alpha}_k^{'} = (H_k^{'}\alpha_k)^{'}$.
2)
$\alpha_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4 =$
$\alpha_k^{'}H_kH_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4 =$
$\widetilde{\alpha}_k^{'}H_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4$.
3)
$[(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]^{'}(\Delta_1^{'}V^{-1}\Delta_1)\alpha_k =$
$[(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]^{'}H_kH_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)H_kH_k^{'}\alpha_k =$
$[(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]^{'}H_kH_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)H_k\widetilde{\alpha}_k$.
4)
$[(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]^{'}(\Delta_1^{'}V^{-1}\Delta_1)(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4 =$
$[(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4]^{'}H_kH_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)H_kH_k^{'}(\Delta_1^{'}V^{-1}\Delta_1)^{-1}\Delta_1^{'}V^{-1}A_4$.

Dessa forma temos:
$\widetilde{\alpha}_k^{'} H_k^{'}(\Delta_1^{'} V^{-1} \Delta_1) H_k \widetilde{\alpha}_k^{'} - \widetilde{\alpha}_k^{'} H_k^{'}(\Delta_1^{'} V^{-1} \Delta_1)(\Delta_1^{'} V^{-1} \Delta_1)^{-1} \Delta_1^{'} V^{-1} \Delta_1 -$
$-[(\Delta_1^{'} V^{-1} \Delta_1)^{-1} \Delta_1^{'} V^{-1} A_4]^{'} H_k H_k^{'}(\Delta_1^{'} V^{-1} \Delta_1) H_k \widetilde{\alpha}_k +$
$+[(\Delta_1^{'} V^{-1} \Delta_1)^{-1} \Delta_1^{'} V^{-1} A_4]^{'} H_k H_k^{'}(\Delta_1^{'} V^{-1} \Delta_1) H_k H_k^{'}(\Delta_1^{'} V^{-1} \Delta_1)^{-1} \Delta_1^{'} V^{-1} A_4$.

Assim, temos que a condicional a posteriori para $\widetilde{\alpha}_k$, dados os demais parâmetros no subespaço corrigido, é dada por:

$$\widetilde{\alpha}_k | ... \sim N([H_k^{'}(\Delta_1^{'} V^{-1} \Delta_1)^{-1} \Delta_1^{'} V^{-1} A_4]; [H_k^{'}(\Delta_1^{'} V^{-1} \Delta_1) H_k]^{-1}) \ . \quad (46)$$

O processo de amostragem foi conduzido tal como em Oliveira et al. (2015), exceto para os vetores singulares que foram amostrados a partir de uma normal multivariada e não Von Miss- Fisher, apresentada por Liu (2001) e Crossa et al. (2011).

3.3 Amostragem a partir das distribuições condicionais completas a posteriori dos parâmetros

O processo de amostragem será feito como descrito em Oliveira et al. (2015), para os dados balanceados e desbalanceados.

3.4 Validação cruzada

Durante o processo de validação, foi usado o esquema proposto por de Los Campos et al. (2009) e aplicado em estudos MET por BURGUEÑO et al. (2012), considerando todos os modelos adaptados nesse estudo (Tabela 1).

A validação cruzada (*k-fold*) usada por esses autores consiste em dividir aleatoriamente n observações em k subconjuntos $(S_1, S_2, ..., S_k)$ não sobrepostos. A validação cruzada é, então, aplicada a cada partição dos dados, isto é, $k-1$ grupos tomados como o conjunto de treino e o grupo restante como o conjunto de validação. Neste trabalho, foi usado $k = 10$, ou seja um esquema de validação cruzada 10 vezes.

Tabela 1: Modelos lineares usados para comparar a predição dos genótipos ausentes nos ensaios MET.

Modelo	Efeito fixo				Efeito aleatório			
EM-AMMI F	E	G	GXE	-	-	-	-	erro
EM-AMMI M	E	-	-	-	-	G	GXE	erro
FA	E	-	-	-	-	G	GXE	erro
AMMIB-I	-	-	-	-	E	G	GXE	erro
AMMIB-D	-	-	-	-	E	G	GXE	erro

em que:

EM-AMMI F: AMMI via EM com efeitos fixos;

EM-AMMI M: AMMI via EM com efeitos mistos;

AMMIB-I: AMMI-Bayesiano com homogeneidade de variâncias;

AMMIB-D: AMMI-Bayesiano com heterogeneidade de variâncias.

4 RESULTADOS

Para ilustrar a capacidade preditiva dos diferentes métodos usados no estudo de dados MET, foram simulados dados com sete ambientes e 20 genótipos. A Tabela 2 apresenta os valores simulados para as variâncias residuais de cada ambiente, bem como os valores estimados por cada modelo. Nessa tabela verificou-se que o modelo AMMI de efeito fixo-AMMIF (EM-AMMI F ou EM-AMMI M) superestimaram a variância média, e os modelos FA estimaram a variância muito próxima ao valor real, sendo o modelo FA2 que chegou mais próximo. Observando os valores estimados dentro de cada ambiente, por cada modelo, nota-se que, ao assumir a variância comum em alguns casos, a variância do ambiente é superestimada e em outros subestimada, os modelos FA tiveram estimativas próximas dos valores simulados, porém com algumas divergências em alguns ambientes.

Tabela 2: Valores simulados e média geométrica da variância residual considerando os diferentes modelos.

Modelo						
Ambiente	Valor simulado	AMMIB-I	AMMIB-D	EM-AMMI F*	FA2	FA7
1	0,2188	3,9643	0,4111	5,56	0,201	0,202
2	0,985	3,9643	1,8807	5,56	1,721	1,722
3	2,036	3,9643	1,7070	5,56	1,552	1,555
4	5,012	3,9643	4,3815	5,56	4,416	4,212
5	7,509	3,9643	6,7934	5,56	7,016	6,678
6	9,641	3,9643	9,1451	5,56	9,405	9,289
7	16,55	3,9643	13,0065	5,56	13,962	14,038
Média G.	3,081	3,9643	3,3437	5,560	2,999	2,959
logLik					-573,9	-607,9

*As estimativas do EM-AMMI F e EM-AMMI M foram as mesmas

Para comparar a capacidade preditiva dos modelos na predição dos dados faltantes, realizaram-se desbalanceamentos de 10%, 33% e 50% de perdas de genótipos no ambiente. Dentre os modelos comparados, têm-se os modelos EM-AMMI F, EM-AMMI M ajustados em dois estágios que consiste em fazer a predição dos valores fenotípicos por modelos mis-

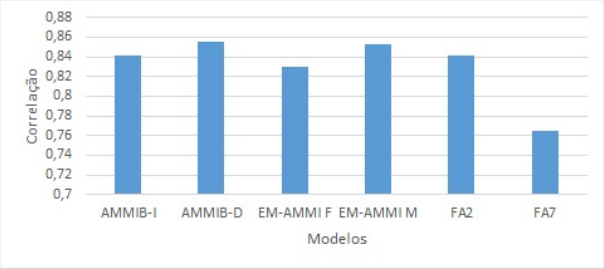
tos (REML/BLUP) e imputação pelo algoritmo EM; modelos FA2, FA7, AMMIB-I e AMMIB-D. Vale ressaltar que, embora seja difícil ajustar um modelo FA completo (no Asreml), devido ao custo computacional e problemas de convergência, não foram observadas dificuldades para desbalanceamentos de 10 e 33%. Entretanto, para desbalanceamento de 50% a convergência do modelo FA7 não foi verificada, tendo o modelo falhado em convergir no terceiro eixo (a sáida do erro de convergência pode ser visto no Apendice), e a opção pelo FA2 deve-se a relatos na literatura que é o melhor modelo para predição de dados faltantes (KELLY, 2007).

A Figura 4 mostra o comportamento dos resultados de validação cruzada para todos modelos em competição. Os resultados demonstraram que com perdas de 10% dos genótipos no ambiente as correlações, entre o valor predito e o observado, ficaram acima dos 0.65, demonstrando a robustez dos modelos na predição dos valores em falta com este nível de perda. Dentro os *folds* pode-se verificar perdas e ganhos marginais de cada modelo em relação ao outro.

Figura 4 – Gráfico de barras da correlação proveniente da validação cruzada considerando desbalanceamento de 10% de perda de genótipos nos ambientes usando o modelo FA, EM-AMMI F, EM-AMMI M, AMMIB-I e AMMIB-D.

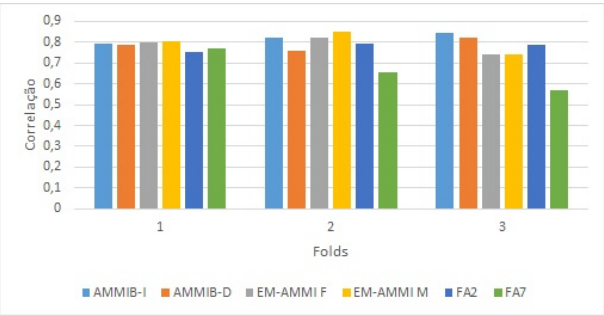
A Figura 5 mostra o desempenho médio de cada modelo com 10% de perdas, sendo que o modelo com menor capacidade preditiva é o FA7 seguido do EM-AMMI F, e de forma geral os modelos AMMIB-I e FA2 tiveram a mesma capacidade preditiva. Os modelos AMMIB-D e EM-AMMI M apresentaram maior capacidade preditiva dentre todos os modelos.

Figura 5 – Gráfico de barras da correlação média proveniente da validação cruzada considerando desbalanceamento de 10% de perda de genótipos nos ambientes usando o modelo FA, EM-AMMI F, EM-AMMI M, AMMIB-I e AMMIB-D.



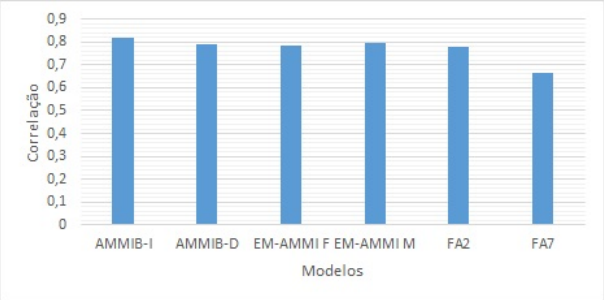
As Figuras 6 e 7 mostram o desempenho dos modelos com 33% de perdas nos genótipos. Na Figura 6, verifica-se que a correlação foi de moderada magnitude (0.57-0.84) para todos os *folds*. É possível verificar que o modelo bayesiano tem uma capacidade preditiva maior que os modelos FA. Também pode-se verificar que o modelo EM-AMMI F foram superiores aos modelos FA. Com 33% de perdas, o modelo EM-AMMI M foi superior ao modelo AMMIB-D (Figura 7). Vale destacar que os modelos que assumem homogeneidade variância, nesse nível de perda, foram superiores aos modelos que assumem heterogeneidade que pode ser causada pelo tipo de genótipo perdido se estável/adaptado ou não (pois é evidente a presença de heterogeneidade nos dados Tabela 2).

Figura 6 – Gráfico de barras da correlação proveniente da validação cruzada considerando desbalanceamento de 33% de perda de genótipos nos ambientes usando o modelo FA, EM-AMMI F, EM-AMMI M, AMMIB-I e AMMIB-D.



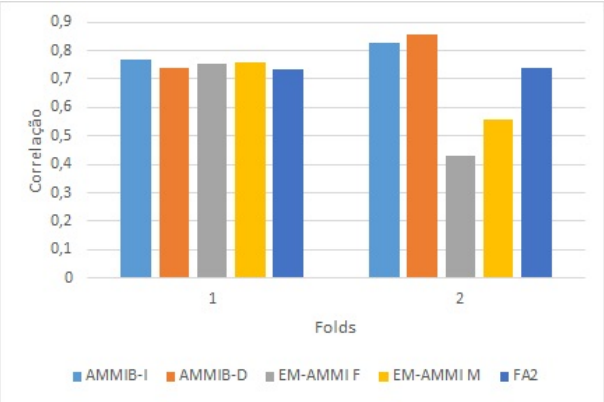
Quanto ao desempenho médio para configuração de 33% de perdas, o modelo AMMI-I foi o que apresentou uma melhor predição em relação aos demais modelos, e o modelo FA7 foi o que apresentou menor capacidade. Contudo a predição dos modelos pode ser considerada boa pois esteve acima de 0,6 que pode ser tido como uma boa predição.

Figura 7 – Gráfico de barras da correlação média proveniente da validação cruzada considerando desbalanceamento de 33% de perda de genótipos nos ambientes usando o modelo FA, EM-AMMI F, EM-AMMI M, AMMIB-I e AMMIB-D.



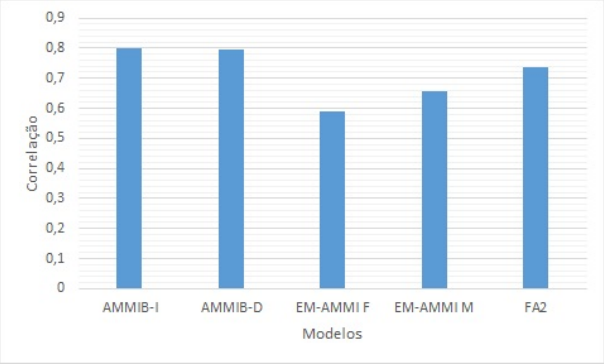
Como já relatado acima, não foi possível ajustar o modelo FA7 com 50% de perdas de genótipos no ambiente (ver Apendice - saída do erro de convergência). Para este nível a comparação em relação aos modelos FA foi com o FA2, verificou-se no *fold* 1 (Figura 8) que não houve grandes discrepâncias entre a predição dos modelos.

Figura 8 – Gráfico de barras da correlação proveniente da validação cruzada considerando desbalanceamento de 50% de perda de genótipos nos ambientes usando o modelo FA, EM-AMMI F, EM-AMMI M, AMMIB-I e AMMIB-D.



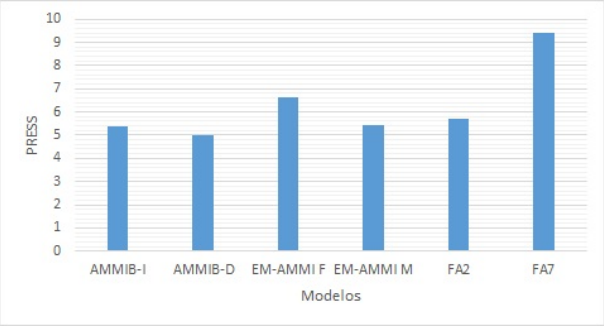
Comparando o desempenho médio dos modelos (Figura 9), os modelos baseados em efeitos fixos, na sua construção, (EM-AMMI F e EM-AMMI M) tiveram o pior desempenho e os modelos bayesianos o melhor. A predição do modelo FA2 perdeu em cerca de 6% para os modelos bayesianos e foi superior em 15% em relação ao modelo EM-AMMI F, indicando que em altos níveis de perda os modelos baseados em efeitos fixos seriam piores.

Figura 9 – Gráfico de barras da correlação média proveniente da validação cruzada considerando desbalanceamento de 50% de perda de genótipos nos ambientes usando o modelo FA, EM-AMMI F, EM-AMMI M, AMMIB-I e AMMIB-D.



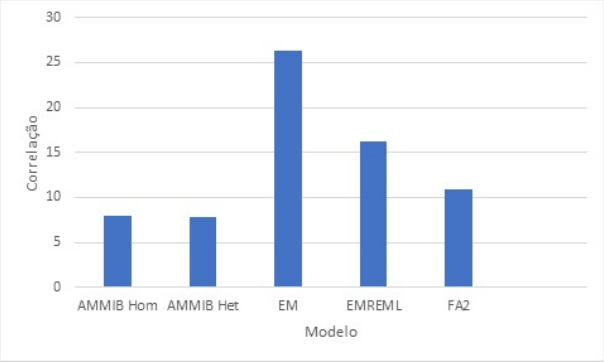
As Figuras 10 e 11 mostram a PRESS, a 10% e 33% de perdas, respectivamente, e indicando que os modelos bayesianos seriam os preferidos em relação aos demais por apresentarem os valores mais baixos.

Figura 10 – Gráfico de barras da PRESS média proveniente da validação cruzada considerando o nível de desbalanceamento de 10% de perda de genótipos nos ambientes usando o modelo FA, EM-AMMI F, EM-AMMI M, AMMIB-I e AMMIB-D.



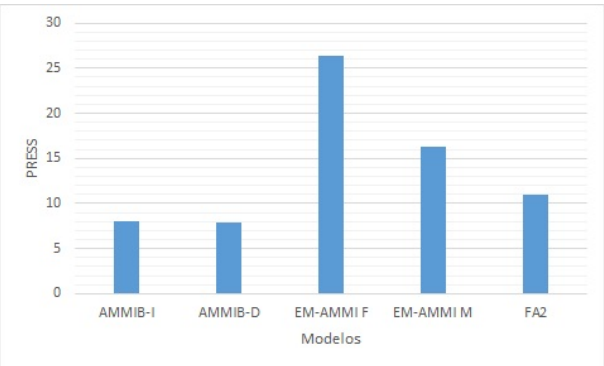
A 10% e a 33% de perdas seriam escolhidos os modelos AMMIB-I e AMMIB-D respectivamente. Independentemente, nos dois níveis de perdas, o modelo FA7 seria o preterido, pois apresentou o valor mais elevado da PRESS e também não teve uma correlação boa entre o valor predito e observado, em comparação com os demais modelos. Em todos níveis de perda considerados, o modelo AMMIB teve a menor PRESS, portanto o melhor modelo.

Figura 11 – Gráfico de barras da PRESS média proveniente da validação cruzada considerando o nível de desbalanceamento de 33% de perda de genótipos nos ambientes usando o modelo FA, EM-AMMI F, EM-AMMI M, AMMIB-I e AMMIB-D.



A 50% a PRESS dos modelos bayesianos foi a mais baixa e quase a mesma. Considerando esse nível de perda a escolha dos modelos seria em ordem bayesianos FA2, EM-AMMI M e o pior modelo seria EM-AMMI F.

Figura 12 – Gráfico de barras da PRESS média proveniente da validação cruzada considerando o nível de desbalanceamento de 50% de perda de genótipos nos ambientes usando o modelo FA, EM-AMMI F, EM-AMMI M, AMMIB-I e AMMIB-D.



5 DISCUSSÃO

O estudo de dados provenientes de ensaio multiambientes tem assumido grande importância em programa de melhoramentos de plantas. O modelo AMMI tem se destacado devido à ampla aplicabilidade e desempenho na análise da interação GE. Outra opção cada vez mais presente na literatura é a utilização de modelos FA que permitem contornar dificuldades presentes em abordagens tradicionais. Esse estudo teve como intenção comparar as principais abordagens relacionadas ao AMMI (EM-AMMI M de RAJU, 2002; EM-AMMI F de Gauch e Zobel (1990) e os Modelos AMMIB-I de Oliveira et al. (2015); Liu (2001) e AMMIB-D de Silva et al. (2016 - não publicado)) e o modelo FA como proposto por Smith et al. (2001). Na literatura o modelo FA2 vem sendo popularizado como o melhor método para análise de ensaios MET com desbalanceamentos dados (KELLY et al. 2007; BURGUEÑO et al. 2008; NUVUNGA et al. 2015).

Como relatado por Smith et al. (2015) e Beeck et al. (2010), o problema maior que poderia se encontrar na comparação desses métodos seria com ajuste dos modelos FA, pois esses são sensíveis a chutes inicias e apresentam grandes problemas de convergência. Como estratégia de comparação nesse estudo, buscou-se comparar os modelos completos, mas observando a limitação dos modelos FA. Os modelos FA aqui considerados são aqueles sem efeitos principais de genótipos. Os modelos FA1 a FA7 convergiram para todas as simulações com 10 e 33% de perdas, desde que valores iniciais sensíveis fossem dados para os parâmetros FA.

O modelo FA3 falhou em convergir com 50% de perdas, de forma que o ajuste dos modelos de ordem superior, a esse nível de desbalanceamento, não foi tentado. As dificuldades computacionais podem ter sido causadas por problemas com o algoritmo de ajuste do modelo (THOMPSON et al. 2003). A experiência na montagem destes modelos indica que a convergência é geralmente conseguida se o usuário progride através de uma sequência sensível de modelos, começando com um modelo diagonal e continuando por dimensões crescentes do modelo FA. Smith et al. (2015), Beeck et al. (2010) relatam que o ajuste de modelos FA parte do modelo FA1, e em

seguida se testando aqueles de ordem superior, utilizando-se critérios de informação para escolha da ordem apropriada. Contudo, se for identificada falha de convergência no modelo de ordem k, o de ordem $k+1$ não é possível de ser ajustado.

Embora este estudo não tenha como objetivo a comparação do ajuste dos modelos, ou verificar o melhor modelo na estimação dos componentes de variância, nosso estudo foi de verificarmos essa precisão preditiva quando os componentes de variância são estimados a partir do modelo, em vez de se supor conhecidos. Dos modelos em competição foi possível verificar que aqueles que assumem heterogeneidade de variância, como FA e AMMIB-D, tiveram melhores ajustes e boa estimação dos componentes de variância (Tabela 2). Esse fato mostra que assumir homogeneidade de variância como pressuposto não é razoável e pode comprometer os resultados da análise de dados MET (CROSSA et al. 1990; EDWARDS E JANNNIK, 2006; ORELLANA, 2012 e ORELLANA et al. 2014).

Os modelos foram comparados com base na sua capacidade preditiva calculada a correlação entre os valores fenotípicos preditivos e observados; os modelos foram ajustados ao conjunto de treinamento e seus valores preditos foram correlacionados aos valores observados no conjunto de validação. Além disso, não relatamos a probabilidade de mudança de classificação produzidas pelo desempenho predito e pelo desempenho observado como em Burgueño et al. (2012). Pois, como relatado por Kelly et al. (2007) o mais importante, do ponto de vista do programa de melhoramento, é a precisão preditiva de cada modelo de análise, pois isso afeta diretamente os ganhos obtidos no processo de seleção. Isso poderia justificar, em trabalhos futuros, incluir na comparação de modelos a probabilidade de classificação dos genótipos com a mudança do modelo de análise na predição de genótipos faltantes.

Constatamos, nesse estudo, que os modelos, de forma geral, são robustos para analisar os dados de ensaios MET que se tenha perdas de genótipos, ou que não sejam testados todos genótipos em todos ambientes de avaliação.

Quanto à acurácia preditiva observou-se que os modelos Bayesianos, de

forma geral, foram os melhores quando comparados ao demais. Embora na literatura alguns autores (KELLY et al. 2007, BURGUEÑO et al. 2008,2012 e NUVUNGA et al. 2015) defendam que o modelo FA2 seja o melhor na predição de genótipos faltantes ou análise de ensaios MET com dados desbalanceados, vale destacar que os estudos conduzidos por esses autores não incluem modelos bayesianos.

Demostrou-se aqui a robustez dos modelos bayesianos na capacidade preditiva com altos níveis de perda, fato que ainda não havia sido testado em estudos presentes na literatura. As magnitudes de herdabilidades, estimadas no presente estudo, foram, em geral, moderadas (isto é, entre 0.69, e 0.89). A precisão preditiva melhorada se traduz diretamente na herdabilidade e, portanto, no ganho genético obtido no programa de melhoramento genético.

Embora tenha se verificado em termos de acurácia preditiva que os modelos bayesianos tenham sido os melhores em relação aos demais, inclusive o modelo EM-AMMI M tenha sido ligeiramente superior aos FA com 33% de perda, vale destacar o outro parâmetro de comparação a PRESS que nos indica o quão modelo está ajustado (KELLY et al. 2007; DIAS e KRAZANOSKI, 2003).

Comparando os modelos com homogeneidade variância, o modelo AMMIB-I teve menor valor e, portanto, o melhor, indicando que os modelos que assumem efeitos aleatórios são sempre melhores que os de efeito fixo ou misto (BURGUEÑO et al. 2008). Em termos de PRESS os modelos FA foram sempre melhores que os modelos AMMI de efeitos fixos, como também já foi verificado por Kelly et al. (2007).

A metodologia bayesiana resultou melhorias em termos de PRESS de 43 a 47% sobre uma análise FA7, e de 6% para o modelo AMMI-EM de efeito fixo. A partir de estudos empíricos sob validação cruzada, Burgueño et al. (2011) relataram que para dados de culturas sem interação cruzada GE, ou seja, sem reclassificação de genótipos, os modelos FA não melhoraram nem perderam a capacidade de predição (correlação entre desempenho observado e predito de genótipos) quando comparados com um modelo misto simples com variâncias homogêneas e covariâncias independentes entre pares

de locais).

Como foi observado, os modelos bayesianos, com níveis de perda de 50%, apresentaram capacidade preditiva maior em relação aos modelos FA. Esse ganho pode ser visto como marginal, pois dentro dos $k-fold$ os modelos FA conservaram a previsibilidade. Embora os modelos bayesianos tenham tido capacidade preditiva média superior, eles apresentam uma desvantagem em relação aos demais modelos quanto ao custo computacional. Certamente, o grande avanço tecnológico vivido nos dias atuais aliado a trabalhos que visam implementar algoritmos mais eficientes tendem a amenizar essa desvantagem.

Outra questão que nos chamou a atenção, nessa abordagem, foi a de considerar ou não heterogeneidade de variâncias para o modelo AMMI bayesiano. Usando diferentes estruturas de variância para validação cruzada e análises de simulação de ensaios MET de híbridos de milho, So e Edwards (2011), por exemplo, não encontraram alguma melhora substancial ao incluir estruturas heterogêneas de (co)variância genética sobre modelos mais simples. Como podemos observar o ganho na predição ao considerar o modelo AMMIB-I de variâncias em relação ao AMMIB-D foi praticamente nulo. Vale ressaltar que, com níveis de 10% de desbalanceamento, os modelos AMMI de efeito fixo, que assumem variâncias comuns, foram superiores ao modelo FA7 e com predição quase igual ao modelo FA2, tido como melhor na literatura.

Aos 10% e 33% de perdas, o modelo EM-AMMI M mostrou ter uma boa capacidade preditiva, que pode ter sido influênciado pelo algorítmo EM, que segundo Paderewski (2014) este algoritmo é muito bom na inputação de dados faltantes, o que pode ter contribuído na qualidade do ajuste.

Uma observação a se ter em conta, em relação ao fato dos modelos FA terem sido superados na capacidade preditiva pelos demais, foi a fraca ligação genética entre os ambientes que pode ter sido ocasionada pela natureza dos dados usados na simulação, e ao menor número de ambientes e genótipos quando comparados com os usados por Kelly et al. (2007) e Smith et al. (2015) que mostraram a superioridade desses modelos.

6 CONCLUSÃO

Os modelos AMMI seja fixo ou bayesianos e FA são robustos no estudo de dados MET com altos níveis de perda de genotipos no ambiente.

Em termos preditivos, ao nível de 10% de desbalanceamento, os modelos AMMIB-D e EM-AMMI M foram superiores seguidos dos modelos AMMIB-I e FA2. A 30% de perda dos dados, o modelo AMMIB-I foi superior, seguidos dos modelos EM-AMMI M, AMMIB-D, EM-AMMI F e FA2. A 50% de perda dos dados, os modelos AMMIB-I e AMMIB-D foram superiores, seguido do modelo FA2.

Na estimação de parâmetros simulados, os modelos FA foram melhores que os modelos AMMI.

REFERÊNCIAS

ANNICCHIARICO, P. Cultivar adaptation and recommendation from alfafa trials in Northern Italy. **Journal of Genetics and Plant Breeding**, v.46, p.269-278, 1992.

ARCINIEGAS-ALARCON, S. and DIAS, C. T. S. Análise AMMI com dados imputados em experimentos de interação genótipo x ambiente de algodão. **Pesq. agropec. bras.**[online]. Vol.44, n.11, pp.1391-1397. ISSN 0100-204X. 2009. http://dx.doi.org/10.1590/S0100-204X2009001100004. . Acesso em 15/06/2016.

ARCINIEGAS-ALARCÓN, S.; GARCÍA-PEÑA, M.; DIAS, C. T. S. Data imputation in trials with genotype x environment interaction. **Interciencia**. Caracas, v.36, p. 444-449, 2011.

BAKER, J.F.; STEWART, T.S.; LONG, C.R. et al. Multiple regression and principal componentes analysis of puberty and growth in cattle. **Journal of Animal Science**, v.66, n.9, p.2147-2158, 1988.

BEECK, C. P., COWLING, W. A., SMITH, A. B., and CULLIS, B. R. Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. Genome 53, 992-1001. 2010.

BERGAMO, G. C.; DIAS, C. T. DOS S.; KRZANOWSKI, W. J. Distribution-free multiple imputation in an interaction in an interaction matrix through singular value decomposition. **Scientia Agricola**, Piracicaba, v.65, n.4, p.422-427, 2008.

BURGUEÑO, J., J. CROSSA, P.L. CORNELIUS, R. TRETHOWAN, G. MCLAREN, and KRISHNAMACHARI, A. Modeling additive × environment and additive × additive × environment using genetic covariances of relatives of wheat genotypes. **Crop Sci.**. 47:311-320. doi:10.2135/cropsci2006.09.0564. 2007.

BURGUEÑO, J. et al. Using factor analytic models for joining environments and genotypes without crossover genotype

$\times environmentinteraction$.**Crop Science**, $Madison, v.48, p.1291-1305, 2008$.

BURGUEÑO, J.; DE LOS CAMPOS, G.; WEIGEL, K. and CROSSA, J. Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. **Crop Sci.** 52:707-719. doi: 10.2135/cropsci2011.06.0299. 2012.

BUTLER, D. G; CULLIS, B. R; GILMOUR, A. R; GOGEL, B J. ASReml-R reference manual, release 3. 160 technical report, Queensland Department of Primary Industries. 2009.

CALINSKI, T.; CZAJKA, S.; DENIS, J.B.; KACZMAREK. Z. EM. Algorithms applied to estimation of missing data in series of variety trials. **Biuletyn Oceny Odmian**, Poznan, v.24-25, p.7-31, 1992.

DE LOS CAMPOS, G., H. NAYA, D. GIANOLA, J. CROSSA, A. LEGARRA, E. MANFREDI, K. WEIGEL, AND COTES, J.M.. Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**, p.182:375-385. doi:10.1534/genetics.109.101501. 2009.

CARNEIRO, P.C.S. Novas metodologias de análise da adaptabilidade e estabilidade de comportamento. **Universidade Federal de Viçosa**, Viçosa. p. 168. 1998.

COCHRAN, W. G.; COX, G. M. Experimental designs. ed. 2. **John Wiley and Sons**, New York. p.611. 1957.

CORNELIUS, P. L.; CROSSA, J.; SEYEDSADR, M. S. Statistical test and estimator of multiplicative model for genotype-by-environment interaction. In: KANG, M.S. e GAUCH Jr, H. G. (Ed.). **Genotype-by-Environment Interaction**. New York: Boca raton, p. 199-234. 1996.

Cotes, J. M., Crossa, J., Sanches, A., Cornelius, P. L. A Bayesian approach for assessing the stability of genotypes. **Crop Sci**., 46:2654-2665, 2006

CROSSA, J. Statistical analyses of multilocation trials. Adv. Agron., 44:

55-85, 1990.

CROSSA. J., P.L. CORNELIUS and W. YAN. Biplots of linear models for studying crossover genotype × environment interaction. **Crop Sci.**, 42: p.619-633. 2002.

CROSSA. J. et al. Modeling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. **Crop Science**, Madison, v. 46, p.1722-1733, 2006.

CROSSA, J., P. PÉREZ, G. DE LOS CAMPOS, G. MAHUKU, S. DREISIGACKER, and MAGOROKOSHO, C. Genomic selection and prediction in plant breeding. **J. Crop Improv**. 25: p.239-261. doi:10.1080/15427528.2011.558767. 2011.

CROSSA, J. From genotype × environmente interaction to gene × environment interaction. **Current Genomics**, London, v. 13, p.220-244. 2012.

CRUZ, C.D.; TORRES, R.A. de; VENCOVSKY, R. An alternative approach to the stability analysis proposed by Silva and Barreto. **Revista Brasileira de Genética**, v.12, p.567-580, 1989.

CRUZ, C.D.; CARNEIRO, P.C.S. Modelos biométricos aplicados ao melhoramento genético. **Editora UFV**, Viçosa. p.579. 2003.

CRUZ, C.D.; REGAZZI, A.J. Modelos biométricos aplicados ao melhoramento genético. 1.ed. Viçosa, MG: UFV, **Imprensa Universitária**, 390p. 1994.

DIAS, C. T. dos S.; HONGYU, K.; de Araújo, L. B.; da Silva, M. J. C.; Peña, M. G.; Araújo, M. F. C. Rodrigues, P. C.; Faria, P. N.; Alarcón, S. A.. A Metodologia AMMI: Com Aplicação ao Melhoramento Genético. **Universidade de São Paulo**, p.80-84. 2014.

DIAS, C. T. S.; KRZANOWSKI, W. J. Model selection and cross-validation in additive main effect and multiplicative interaction

(AMMI) models. Crop Science, Madison, v. 43, p. 865-873, 2003.

DUARTE, J.B.; VENCOVSKY, R. Interação genótipo x ambiente: uma introdução à analise "AMMI". **: Sociedade Brasileira de Genética**, Ribeirão Preto. p.60, 1999. (Série Monografias, 9).

EBERHART, S. A.; e RUSSELL, W. A. Stability parameters for comparing varieties. **Crop Science**, Madison, v. 6, p. 36 - 40. 1966.

EDWARDS J.D., JANNINK J.L. Bayesian modeling of heterogeneous error and genotype × environment interaction variances. **Crop Sci**. 46:820-833. doi:10.2135/cropsci2005.0164. 2006.

FISCHER, J.; PETERSON, G.D.; GARDNER, T. A.; GORDON, L. J.; FAZEY, I.;ELMQVIST, T.; FELTON, A.; FOLKE, C.; DOVERS, S. Integrating Resilience Thinking and Optimisation for Conservation. **Trends in Ecology Evolution**. v. 24, n. 10, p.549-554. 2009.

GABRIEL. K. R. The biplot graphic display of matrices with application to principal component analysis. **University**, Jerusalem. p.453-466. 1971.

GAUCH, H. G.; ZOBEL, R. W. Predictive and postdictive success of statistical analysis of yield trials. **Theoretical and Applied Genetics**, Berlin, v. 76, n. 1, p.1-10, 1988.

GAUCH, H.G.; ZOBEL, R.W. Imputing missing yield trial data. **Theoretical and Applied Genetics**, New York. v.79, p.753-761. 1990.

GAUCH, H.G.; ZOBEL, R.W. AMMI analysis of yield trials. In: KANG, M.S.; GAUCH, H.G. (Ed.). Genotype by environment interaction. **Boca Raton: CRC Press**, p.85-122. 1996.

GAUCH, H.G.; ZOBEL, R.W. Identifying megaenvironments and targeting genotypes. Crop Science, DOI: 10.2135/cropsci1997.0011183X003700020002x.v.37, p.311-326, 1997.

GAUCH, H. G.; ZOBEL, R. W. AMMI analysis of yield trials. In: KANG, M. S.; GAUCH, H. G. (Ed.). Genotype by environment interaction. **Boca Raton: CRC Press**, v. 4, p.85. 2006.

Kelly, A.M., A.B. Smith, J.A. Eccleston, and B.R. Cullis. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. **Crop Sci.** 47:1063-1070. 2007.

LEE, Y. and NELDER, J.A. Double hierarchical generalized linear models. **Journal of the Royal Statistical Society**: Series C (Applied Statistics). p. 1-29, 2006.

LIN, C.S.; BINNS, M.R. A superiority measure of cultivar performance for cultivar x location data. **Canadian Journal of Plant Science**, v.68, p.193-198, 1988.

LIU, G. **Bayesian computations for general linear-bilinear models**. 2001. 150 p. Thesis (Doctor of Philosophy) - University of Kentucky, Lexington, 2001.

NUVUNGA, J.J. ; OLIVEIRA, L.A. ; PAMPLONA, A.K.A. ; SILVA, C.P. ; LIMA, R.R. ; BALESTRE, M. . Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. **Genetics and Molecular Research**, v. 14, p. 14262-14278, 2015.

OAKEY H., Verbyla A., Pitchford W., Cullis B., Kuchel H. Joint modeling of additive and non-additive genetic line effects in single field trials. **Theor. Appl. Genet**. 113: 809-819. 2006

OLIVEIRA, L. A., SILVA, C. P., NUVUNGA, J. J., SILVA A. Q., BALESTRE, M. Bayesian GGE biplot models applied to maize multi-environments trials. **Crop Science** doi: 10.4238/gmr.15028612. 2016.

OLIVEIRA, L. A., SILVA, C. P., NUVUNGA, J. J., SILVA A. Q., BALESTRE, M. Credible intervals for genotypic and environmental scores in the AMMI model with random effects for genotype. **Genet Mol Res.** doi: 10.2135/ cropsci2014.05.0369. 2015.

ORELLANA, M. A.; EDWARDS, J.; CARRIQUIRY, A. Heterogeneous variances in multi-environment yield trials for corn hybrids. **Crop Science**, Madison, v. 54, p. 1048-1056, 2014.

ORELLANA, M. A. Bayesian prediction of crop performance modeling genotype by environment interaction with heterogeneous variances. 2012. Paper 12740. Disponível em: <http://lib.dr.iastate.edu/etd/12740>. Acesso em: 10 out. 2016.

PADEREWSKI, J. AND RODRIGUES, P. C. The usefulness of EM-AMMI to study the influence of missing data pattern and application to Polish post-registration winter wheat data. **Astralian Jornal of Crop Science**. AJCS 8(4):640-645. 2014.

PIEPHO, H.P. Methods for estimating missing genotype-location combinations in multilocation trials - an empirical comparison. Informatik, **Biometrie und Epidemiologie in Medizin und Biologie**. Stuttgart, vol.26, n.4, p.335-349. 1995.

PIEPHO, H.-P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. Theoretical and Applied Genetics 97, 195-201. 1998.

PIEPHO, H.P., AND MÖHRING J.. Computing heritability and selection response from unbalanced plant breeding trials. **Genetics** 177:1881-1888. doi:10.1534/genetics.107.074229. 2007.

PINHO, E. M. Estimação Bayesiana para Medidas de Desempenho de Testes Diagnósticos. Universidade Federal de São Carlos, 2006.

R CORE TEAM. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2016.

RESENDE, M. D. V. de; THOMPSON, R. Factor analytic multiplicative mixed models in the analysis of multiple experiments. **Revista de Matemática e Estatística**, v. 22, n. 2, p.1-22. 2004.

RESENDE, M. D. V. Matemática e estatística na análise de experimentos e no melhoramento de plantas. Brasilia: EMBRAPA Informação Tecnologica; **Colombo: Embrapa Florestas**. p.555-559. 2007.

RODRIGUES, P. C.; MALOSETTI, M.; GAUCH, JR, H. G.; and VAN

EEUWIJK, F. A. A Weighted AMMI Algorithm to Study Genotype-by-Environment Interaction and QTL-by-Environment Interaction. **Crop Sci.** 54:1555-1570. doi: 10.2135/cropsci2013.07.0462. 2014.

RUBIN, D.B. Inference and Missing Data. **Biometrika**, 63, 581-592. 1976.

RUBIN, D. B. Multiple Imputation for Nonresponse in Surveys. **Wiley**, New York. P. 1-8. 1987.

SILVA, C. P., OLIVEIRA, L. A., NUVUNGA, J. J., PAMPLONA, A. K. A., BALESTRE, M. Heterogeneity of variance in AMMI-Bayesian model in the study of multi-environment data, **Crop Sciencs**. 2016.

SMITH AB, CULLIS BR, THOMPSON R: The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **J Agric Sci**, 143:449-462, 2005.

SMITH, A. B., CULLIS, B. R., THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. **Biometrics**, 57, 1138-1147, 2001.

VIELE, K.; SRINIVASAN, C. Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback-Leibler information. **Journal of Statistical Planning and Inference**, Amsterdam, v. 84, n. 1/2, p. 201-219, 2000.

WEBER, W. E,; WRICKE, G.; WESTERMANN, T. Selection of genotypes and prediction of performance y analyzing genotype-by-environment interactions. In KANG, M.S. GAUCCH, H. G., ed. Geotype by environment interaction. Boca Raton: CRC Press, Cap. 13, p. 353-371. 1996.

YAN, W.; HUNT, L.A.; SHENG, Q.L.; SZLAVNICS, Z. Cultivar evaluation and mega-environment investigation based on the GGE Biplot. **Crop Science**, v.40, p.597-605. 2000.

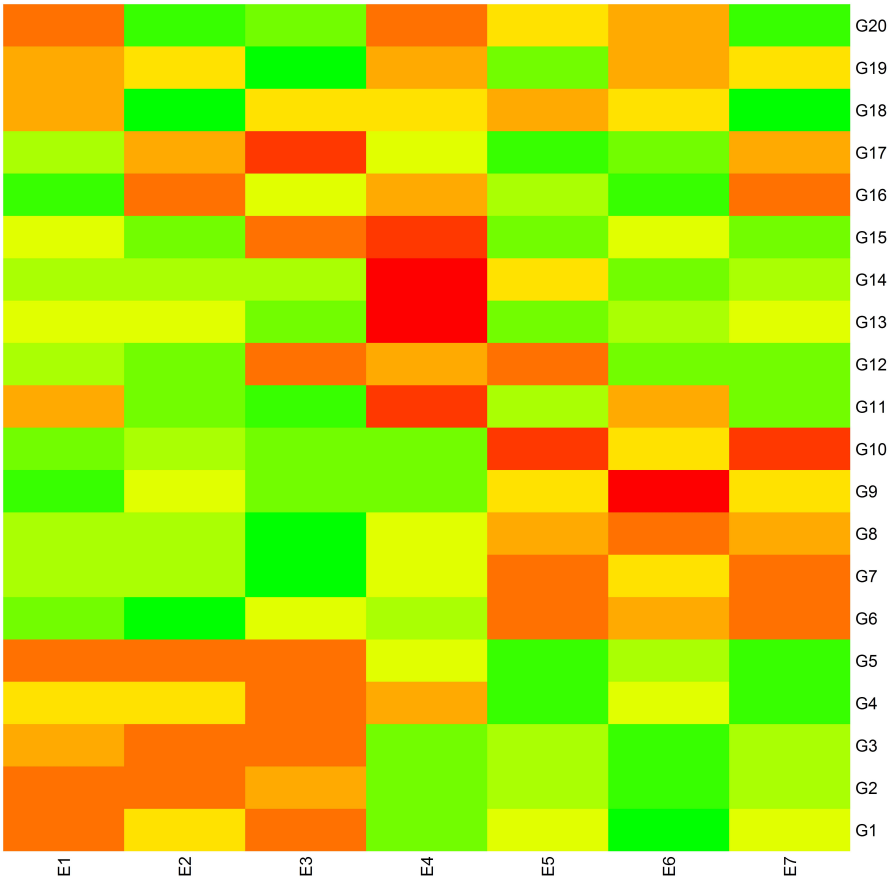
YANG, R. et al. Biplot analysis of genotype x environment interaction:

proceed with caution. **Crop Science**, Madison, v. 49, p.1560-1576. 2009.

ZOBEL, R. WRIGHT, M J., GAUCH H. G.: Statistical analysis of a yield trial. **Agronomy Journal**, Madson, v.80, p.388-393, 1988.

APENDICE

Figura 1 - Mapa de calor



Saida do erro de convergência

LogLik	S2	DF	wall	cpu
⋮	⋮	⋮	⋮	
-312.6945	1.0000	203	19:54:26	0.0
-321.3655	1.0000	203	19:54:26	0.0 (15 restrained)
-313.8110	1.0000	203	19:54:26	0.0
-322.4961	1.0000	203	19:54:26	0.0 (15 restrained)
-314.4015	1.0000	203	19:54:26	0.0
-316.8335	1.0000	203	19:54:26	0.0
-326.8192	1.0000	203	19:54:26	0.0
-319.4371	1.0000	203	19:54:26	0.0
-314.8019	1.0000	203	19:54:26	0.0
-320.7562	1.0000	203	19:54:26	0.0
-334.8077	1.0000	203	19:54:26	0.0
-350.1607	1.0000	203	19:54:26	0.0

Logliklihood decreased to -372.32 - trying again with reduced updates

-347.2707	1.0000	203	19:54:26	0.0
-345.4351	1.0000	203	19:54:26	0.0
-339.4198	1.0000	203	19:54:26	0.0
-355.2745	1.0000	203	19:54:26	0.0
-337.9524	1.0000	203	19:54:26	0.0
-330.9841	1.0000	203	19:54:26	0.0

Terminating with nfault = 3

Finished on: Sun Feb 05 20:31:04 2017

Convergence failed