JOEL JORGE NUVUNGA

ANÁLISE DE FATORES PARA ENSAIOS MULTIAMBIENTES SOB DIFERENTES NÍVEIS DE DESBALANCEAMENTO USANDO MODELOS MISTOS

LAVRAS-MG

2014

JOEL JORGE NUVUNGA

ANÁLISE DE FATORES PARA ENSAIOS MULTIAMBIENTES SOB DIFERENTES NÍVEIS DE DESBALANCEAMENTO USANDO MODELOS MISTOS

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 obtenção do título de Mestre.

Orientador
Dr. Renato Ribeiro de Lima

Coorientador
Dr. Marcio Balestre

LAVRAS-MG

2014

Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e Serviços da Biblioteca Universitária da UFLA

Nuvunga, Joel Jorge.
Análise de fatores para ensaios multiambientes sob diferentes níveis de desbalanceamento usando modelos mistos / Joel Jorge Nuvunga. – Lavras : UFLA, 2014.
81 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2014.
Orientador: Renato Ribeiro de Lima.
Bibliografia.

1. Interação genótipo-ambiente. 2. Variância não estruturada. 3. Adaptabilidade. 4. Estabilidade. 5. Fator analítico. I. Universidade Federal de Lavras. II. Título.

CDD – 519.535

JOEL JORGE NUVUNGA

ANÁLISE DE FATORES PARA ENSAIOS MULTIAMBIENTES SOB DIFERENTES NÍVEIS DE DESBALANCEAMENTO USANDO MODELOS MISTOS

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 obtenção do título de Mestre.

Aprovada em 20 de Fevereiro de 2014.

Dr. Júlio Sílvio de Souza Bueno Filho	UFLA
Dr. José Aírton Rodrigues Nunes	UFLA
Dr. Marcio Balestre	UFLA

Dr. Renato Ribeiro de Lima
Orientador

LAVRAS-MG

2014

Aos meus pais,

Jorge Nuvunga (In memoriam) e

Tahate Cossa,

que me ensinaram a importância dos estudos

e em todos os momentos de dificuldade,

sempre me aconselharam.

Aos meus irmãos exemplos de perseverança,

solidariedade e pela companhia constante, amizade,

paciência e amor.

DEDICO

AGRADECIMENTOS

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX), pela oportunidade concedida para a realização do mestrado;

Aos meus orientadores, o Prof. Dr. Renato Ribeiro de Lima e Márcio Balastre, por ajudarem nos meus primeiros passos no Mestrado, pelo conhecimento compartilhado, confiança no meu trabalho e apoio;

Ao Professor Doutor Carvalho Carlos Ecole, pelo apoio incondicional para esta conquista e ao Doutor Manuel Amane pelo incentivo para continuar com os estudos;

Aos Profs. Drs. Júlio Sílvio de Sousa Bueno Filho, Daniel Ferreira Furtado, José Airton Rodrigues Nunes, João Domingos Scalon, serei eternamente agradecido pela paciência, pelos ensinamentos e pela valiosa colaboração. Aos professores do Departamento de Ciências Exatas da DEX/UFLA, obrigada pela amizade e contribuição na minha formação;

Aos meus colegas do Mestrado, pelo constante apoio e amizade, a todos vocês que fizeram parte deste meu aprendizado e de uma forma muito especial. Agradeço a: Luciano Oliveira, Carlos Pereira, Andrezza Kellen, Fernando Ribeiro, Carlos Muianga, Rafael Lemos, e Adriano Carvalho;

Aos meus irmãos; Rita, Elisar, Marta, Alfredo, Matias, Jorge, Rute, Aida, Lúcia e Maria, pela amizade e companheirismo de toda vida;

A todos moçambicanos em Lavras com os quais compartilhei os melhores momentos e, em especial aos amigos Mateus Come e Chadreque Nhanengue, pelo convívio e paciência nos dois anos do Mestrado;

A Joaquim Uate, Edmundo Caetano, Bartolomeu Tanguene e Gilda Aparecida, pela amizade e convivência, durante minha estadia em Lavras;

Ao Momade Álvaro, Noimilto Mindo, Bacar, Lídia e a todos que colaboraram direta e indiretamente para esta conquista;

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Ministério de Ciência e Tecnologia de Moçambique (MCT) pelo apoio financeiro.

RESUMO

Em ensaios de múltiplos ambientes é comum a presença de dados desbalanceados, heterogeneidade de variâncias e covariâncias de resíduos, que podem dificultar o trabalho de seleção do melhorista. Além disso, a declaração de que um genótipo é estável, pode muitas vezes causar dúvidas. Com o objetivo de avaliar as alternativas para o estudo da interação G×E sob diferentes níveis de desbalanceamento, para este trabalho foram testados três níveis de desbalanceamento em um conjunto de dados reais, adotando-se um modelo misto com variância não estruturada (UN) e a validação cruzada para validar genótipos estáveis. Foram considerados dados provenientes de ensaios multiambientes com 55 híbridos de milho, avaliados nos anos 2005 e 2006. As análises foram feitas em dois estágios: no primeiro, os componentes de variância foram estimados pelo método da máxima verossimilhança restrita adotando o modelo mito, via algoritmo EM, enquanto que no segundo estágio aplicou-se a análise FA (fator analítica) com objetivo de obter escores fatorais e a posição relativa de cada genótipo no biplot. Realizaram-se desbalanceamentos aleatórios nos dados, considerando níveis de 10%, 30% e 50% de parcelas perdidas e, em seguida, os escores foram reestimados utilizando o modelo FA. Os resultados mostraram que a análise FA é robusta na análise de dados multiambientes (MET) sob diferentes níveis de perdas aleatórias nas parcelas, o que inclui os casos em que nem todos os genótipos são testados em todos os ambientes. Desbalanceamentos de 10%, 30% e 50% apresentaram valores médios da correlação de 0,7; 0,6 e 0,56. De maneira geral, os genótipos considerados estáveis no biplot apresentaram menor erro quadrático de predição e menores elipses preditivas. Assim, os resultados permitem inferir que a soma de quadrados dos erros de predição PRESS poderia ser utilizada como alternativa para avaliar o desempenho de genótipos considerados estáveis no biplot. Esse resultado se confirmou pela amplitude das elipses de predição, que foram menores nesses genótipos. Verificou-se que a análise de fatores usando modelo misto é robusta sob os diferentes níveis de desbalanceamento, com valores de correlação variando de médio a alto, dependendo do nível de perda estabelecido. Assim, não há dúvidas quanto ao potencial desse tipo de análise para avaliação da estabilidade no melhoramento de plantas.

Palavras-chave: Interação GxE. Variância não estruturada. Adaptabilidade. Estabilidade. Fator analítico.

ABSTRACT

It is common the presence of unbalanced data, and heterogeneity of residuals variances and covariances, which may become the work of plant breeders more difficult, mainly when it was considered multi-environment trials. Furthermore, the affirmation that a genotype is stable, under these conditions, may not be correct. However, aiming to evaluate the alternatives to study the genotype × environment interaction (G×E), under different unbalanced levels, it was carried out this study in which were assumed three different unbalanced levels on a real dataset, and it was adopted a mixed model with unstructured variance to analyse and to validate stable genotypes by using cross-validation. It was considered data from multi environment trials with 55 maize hybrids, assessed between 2005 and 2006. analyses were carried out in two stages: (i) the components of variance were estimated by considering restricted maximum likelihood method, using EM algorithm; and (ii) it was applied the factor analytic multiplicative mixed (FA) aiming to obtain factorial scores and relative position of each genotype in a biplot. Different unbalanced conditions were randomly performed by considering 10%, 30% and 50% of missed experimental units. Thus, the scores were estimated in different unbalanced conditions by using the FA-based analysis.. The results indicated that the FA-based analysis is robust to analyse data from multi environment trials (MET), under different levels of unbalancing, including cases in which not all genotypes are evaluated in all environments. Unbalancing of 10%, 30% and 50% showed correlation average of 0.7, 0.6, and 0.56, respectively. In general, genotypes which were considered stable in the biplot presented the lowest prediction square error and the smallest predictive ellipses. With these results, it is inferred that the Residuals The prediction error sum of squares (PRESS) could be an alternative method to evaluate the performance of genotypes considered stable in the biplot,what it was confirmed by the amplitude of the predictive ellipses. Furthermore, the factor analytic multiplicative mixed model analysis is robust under different unbalanced levels, with values of correlation raging from medium to high, depending on the established level of losses. Therefore, this type of analysis is proper and has potential to use in the assessing stability in programs of plant breeding.

Keywords: Interaction genotype × environment. Unstructured variance. Adaptability. Stability. Factor analytic multiplicative mixed model.

SUMÁRIO

1 INTRODUÇÃO

A identificação de genótipos com alta produtividade e estabilidade de produção e ampla adaptabilidade aos mais variados ambientes é um dos principais objetivos dos programas do melhoramento genético. Entretanto, essa seleção é afetada pela interação G x E. Existem diferentes metodologias destinadas à avaliação da interação G x E, em que a escolha de um método depende dos dados experimentais, especialmente do número de ambientes disponíveis, da precisão requerida e do tipo de informação desejada. Além disso, existem métodos alternativos e complementares que podem ser utilizados conjuntamente (CRUZ; REGAZZI; CARNEIRO, 2004).

Uma das metodologias utilizadas na avaliação da interação G x E é baseada em análise multiplicativa, que explora a resposta dos genótipos em ambientes específicos, descrevendo a interação G x E de uma forma mais criteriosa (RESENDE, 2004). A vantagem dos métodos multiplicativos reside na possibilidade de agrupamento de ambientes e genótipos semelhantes, permitindo também a identificação dos genótipos com maior potencial em cada subgrupo de ambientes, por meio do gráfico biplot.

Nos métodos multiplicativos, os dados MET (multiambientes) são frequentemente analisados em duas etapas: na primeira, os efeitos dos genótipos são estimados separadamente para cada ensaio e, na segunda são combinados para formar os dados para uma análise geral. A abordagem de dois estágios é uma aproximação à análise conjunta dos dados brutos de todos os ensaios. Se existe uma heterogeneidade de variância do erro entre os ensaios e ou repetição desigual nos ensaios, essa aproximação pode ser ruim. Uma alternativa é o uso do modelo misto, com efeitos principais de genótipos e ambientes (pelo menos um dos quais é aleatório) e interação G x E aleatória (PATTERSON et al., 1977). Essa interação é geralmente assumida como um conjunto de efeitos aleatórios independentes com variâncias constantes. Porém, a validade dessas suposições é questionável. Segundo Smith, Cullis e Thompson (2001), muitos autores, incluindo Patterson e Nabugoomu (1992), reconhecem a possibilidade

da existência da heterogeneidade de variância. Nesse contexto, modelos que contemplem essa heterogeneidade de variância para interação G x E e relaxamento da suposição de independência podem ser necessários.

Dentre os métodos propostos destacam-se o uso de dois modelos mistos multiplicativos: AMMI - *additive main effects and multiplicative interactions* e FAMM - *factor analytic multiplicative mixed models*. Dentre esses modelos, o que vem sendo mais utilizado na análise de dados de MET são os modelos mistos multiplicativos de fator analítico (que consideram aleatórios os efeitos dos genótipos e interação G x E). Esses modelos foram propostos por Piepho (1997) e, mais tarde, foram designados FAMM (ou simplesmente FA) por Resende e Thompson (2004). Os modelos FA foram propostos em detrimento aos AMMI devido ao fato desses últimos apresentarem pelo menos cinco grandes limitações: consideram os efeitos de genótipo e de G x E como fixos; são adequados apenas para dados balanceados; não consideram a variação espacial dentro dos ensaios; não consideram a heterogeneidade de variância entre ensaios e não consideram os diferentes números de repetições nos ensaios. No entanto, estas são características geralmente encontradas em experimentos de campo. Por essas razões, o FAMM com efeitos aleatórios de genótipo e de G x E, é conceitualmente e funcionalmente superior ao AMMI.

Kelly et al. (2007), Piepho (1998) e Smith, Cullis e Thompson (2001, 2005), mostraram a superioridade dos modelos FAMM, no estudo da interação G x E. Contudo, os estudos propostos por estes autores limitaram-se a comparar modelos e estrutura da matriz de variâncias e covariâncias genéticas, na presença de heterogeneidade de variâncias. Apesar de terem demonstrado que esses modelos são adequados para estudo da interação G x E na presença de desbalanceamento dos dados (nem todos os genótipos cultivados em todos locais), nenhum desses estudos avaliou a robustez do modelo FA na presença de alto índice de desbalanceamento (por perda de parcelas, genótipo ou bloco). Recentemente Crossa et al. (2011b) verificara a robustez dos modelos FA na presença de desbalanceamentos, sem, contudo testar diferentes níveis de perda.

Apesar do grande atrativo dessa técnica no melhoramento de plantas, uma das dificuldades encontradas por pesquisadores na adoção dos modelos FA refere-se a sua implementação computacional, pois os pacotes disponíveis, não exploram o modelo de regressão em que assenta o modelo FA (SMITH; CULLIS; THOMPSON, 2001). Por conseguinte, as equações do modelo misto são relativamente densas, reduzindo seriamente a velocidade computacional das análises para conjuntos de dados com um grande número de ambientes ou quando se ajusta a variância de modelos fator analíticos com vários fatores (THOMPSON et al., 2003). Visando melhorar a estabilidade computacional, Thompson et al. (2003) sugeriram a aplicação de matrizes esparsas na estrutura FA, porém, sua implementação também é computacionalmente intensiva como pode ser observado nos doze passos propostos pelos autores. O outro problema prático com o modelo FA é a ocorrência frequente dos casos Heywood, onde alguns parâmetros da estrutura FA tornam-se nulos ou negativos, o que pode prejudicar a análise (SILVA; DUTKOWSKI, 2006; SMITH; CULLIS; THOMPSON, 2001; THOMPSON et al., 2003). Nesse sentido, uma das formas de confirmar as estabilidades de genótipos descritas em biplots em seria por meio de validação cruzada (LAVORANTI, 2003; YANG et al., 2009) utilizando a estatística da soma de quadrados dos erros de predição (PRESS). Uma vez que nessa abordagem o desbalanceamento não destrói a estrutura de interação como verificado em análises bootstrap de AMMI ou GGE (LAVORANTI, 2003; YAN, 2010; YANG et al., 2009), a precisão das elipses de confiança obtidas na validação cruzada tem interpretação genética direta em termos de estabilidade, ou seja, quanto menos sensível é o desempenho de genótipo em relação a sua perda em ambientes contrastantes, mas estável podemos considerar esse genótipo.

1.1 Objetivo Geral

Avaliar o desempenho da análise MET (multiambientes) no estudo da interação G x E sob os diferentes níveis de desbalanceamento (por perda de parcelas) usando modelo misto multivariado.

1.2 Objetivos Específicos

a) Aplicar o modelo misto multivariado com o propósito de analisar a estrutura da matriz de variâncias e covariâncias das e interações genótipo x ambiente na presença de dados balanceados e desbalanceados;

b) Aplicar a estrutura fator analítico (FA) como forma de avaliar a estabilidade e adaptabilidade dos genótipos;

c) Determinar regiões de confiança de predição dos escores genotípicos (blup´s) nos diferentes níveis de desbalanceamento.

2 REVISÃO DE LITERATURA

Nesta seção, é apresentada uma revisão de literatura, que visa a abordar os conceitos, básicos sobre a análise de fatores, modelos mistos, interação genótipo por ambiente, métodos de estudo do genótipo por ambiente e métodos de validação cruzada.

2.1 Modelos mistos multivariados (MMM)

O modelo misto multivariado é uma extensão do modelo linear multivariado. Isto significa que o modelo pode ser estimado adicionando um componente aleatório, assumindo que cada um dos elementos de **Y** tem uma correlação sistemática com a parte linear do modelo.

A análise simultânea de vários caracteres visando estimar a estrutura de covariância ou correlação e também a predição de valores genéticos para fins de seleção é realizada de maneira eficiente pelo procedimento REML/BLUP (multitrait) multivariado ou pela análise multivariada não estruturada. Nesse caso, o modelo multivariado é especificado de forma a contemplar a covariância ambiental existente entre os caracteres (RESENDE, 2002, 2007).

Os modelos multivariados destinam-se à avaliação de indivíduos, simultaneamente para dois ou mais caracteres e apresentam grande relevância no contexto de seleção envolvendo agregados genotípicos.

A combinação de técnicas de análise multivariada com os modelos mistos é importante para a análise de múltiplos caracteres, múltiplos experimentos e, em alguns casos, medidas repetidas. Dentre as técnicas multivariadas, a análise de fatores tem se destacado, se mostrando muito eficiente na análise de dados MET quando associada aos modelos mistos.

A técnica de análise de fatores associada ao modelo misto é designada FAMM (*factor analytic mixed multiplicative mixed*) que é mais indicada para análise de múltiplos experimentos. A análise de fatores enfatiza a atribuição da covariância entre variáveis a fatores comuns. Isto é relevante quando as

variáveis referem-se a ambientes ou experimentos e todos os ambientes são alvos da análise e não apenas aqueles que mais contribuem para a variação total. Por outro lado, a covariância ou correlação entre ambientes, atribuídas a fatores comuns considera a similaridade e dissimilaridade entre ambientes, o que é uma propriedade interessante nesse contexto (RESENDE; THOMPSON, 2004).

Definição do modelo

O modelo para uma análise multivariada se assemelha a empilhar modelos univariados para cada um dos caracteres (MRODE; THOMPSON, 2005). Por exemplo, considere uma análise multivariada para dois caracteres, com o modelo para cada característica é dada em (1), isto é, para um caráter (ambiente 1):

$$y_1 = X_1 b_1 + Z_1 u_1 + e_1 : \quad (1)$$

E para o ambiente 2:

$$y_2 = X_2 b_2 + Z_2 u_2 + e_2 \quad (2)$$

em que:

$\mathbf{y_i}$ é vetor de observações para o caractere i, $\mathbf{b_i}$ é vetor de efeitos fixos para do ambiente i, $\boldsymbol{u}_i$ = vetor dos efeitos aleatórios de genótipo para o ambiente i, $\boldsymbol{e}_i$ vetor de efeitos residuais aleatórios para o ambiente i, e $\mathbf{X_i}$ e $\mathbf{Z_i}$ são matrizes de incidência relativas para os efeitos fixos e efeitos aleatórios do genótipo, respectivamente, para o ambiente i.

Se os genótipos são ordenados dentro de cada ambiente, o modelo de análise multivariada para os dois ambientes pode ser escrito como:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (3)$$

É assumido que:

$$Var\begin{bmatrix} u_1 \\ u_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} I\sigma_{11} & I\sigma_{12} & 0 & 0 \\ I\sigma_{21} & I\sigma_{22} & 0 & 0 \\ 0 & 0 & R_{11} & R_{12} \\ 0 & 0 & R_{21} & R_{22} \end{bmatrix}, \quad (4)$$

em que: $\mathbf{I}\sigma_{ij}$ são elementos de G-matriz de variâncias e covariâncias genéticas, σ_{11} = variância genética aditiva para efeitos diretos para o ambiente 1; σ_{12} = σ_{21} = covariância genética aditiva entre os dois ambientes, σ_{22} = variância genética aditiva para efeitos diretos para o ambiente 2; $\boldsymbol{I}$ é a matriz identidade e, $\boldsymbol{R}$ = matriz de variância e covariância para os efeitos residuais.

As equações do modelo misto multivariado (MMM) são da mesma forma como as do caso univariado, e estas são os seguintes:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}\begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}, \quad (5)$$

em que:

$$X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}; \quad Z = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}; \; b = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} \text{ e } \; u = \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \end{bmatrix}.$$

Escrevendo as equações para cada um dos ambientes no modelo separadamente, o MME

torna-se:

$$\begin{bmatrix} X_1'R_{11}X_1 & X_1'R_{12}X_2 & X_1'R_{11}Z_1 & X_1'R_{12}Z_2 \\ X_2'R_{12}X_1 & X_2'R_{22}X_2 & X_2'R_{12}Z_1 & X_2'R_{22}Z_2 \\ Z_1'R_{11}X_1 & Z_1'R_{12}X_2 & Z_1'R_{11}Z_1 + I\sigma_{11} & Z_1'R_{12}Z_2 + I\sigma_{12} \\ X_2'R_{21}X_1 & Z_2'R_{22}X_2 & Z_1'R_{21}Z_1 + I\sigma_{21} & Z_2'R_{22}Z_2 + I\sigma_{22} \end{bmatrix}\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \begin{bmatrix} X_1'R_{11}y_1 + X_1'R_{11}y_2 \\ X_2'R_{12}y_1 + X_2'R_{22}y_2 \\ Z_1'R_{11}y_1 + Z_1'R_{12}y_2 \\ X_2'R_{21}y_1 + Z_2'R_{22}y_2 \end{bmatrix} \quad (6)$$

E a solução dada por:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \begin{bmatrix} X_1'R_{11}X_1 & X_1'R_{12}X_2 & X_1'R_{11}Z_1 & X_1'R_{12}Z_2 \\ X_2'R_{12}X_1 & X_2'R_{22}X_2 & X_2'R_{12}Z_1 & X_2'R_{22}Z_2 \\ Z_1'R_{11}X_1 & Z_1'R_{12}X_2 & Z_1'R_{11}Z_1 + I\sigma_{11} & Z_1'R_{12}Z_2 + I\sigma_{12} \\ X_2'R_{21}X_1 & Z_2'R_{22}X_2 & Z_1'R_{21}Z_1 + I\sigma_{21} & Z_2'R_{22}Z_2 + I\sigma_{22} \end{bmatrix}^{-1}\begin{bmatrix} X_1'R_{11}y_1 + X_1'R_{11}y_2 \\ X_2'R_{12}y_1 + X_2'R_{22}y_2 \\ Z_1'R_{11}y_1 + Z_1'R_{12}y_2 \\ X_2'R_{21}y_1 + Z_2'R_{22}y_2 \end{bmatrix} \quad (7)$$

Deve-se notar que, se $\mathbf{R_{12}}$, $\mathbf{R_{21}}$, e $\sigma_{12} = \sigma_{21}$ são ajustados para zero, as matrizes nas equações acima reduzem ao habitual modelo em que se realizam análises de um único ambiente (modelo univariado) já que os dois ambientes tornam-se não correlacionados (MRODE; THOMPSON, 2005).

2.2 Análise de fatores (AF)

A análise de fatores ou análise fatorial é um nome genérico dado a uma classe de métodos estatísticos multivariados cujo propósito principal é definir a estrutura subjacente e explicar o comportamento de um número relativamente grande de variáveis observadas, em termos de um número relativamente pequeno de variáveis latentes ou fatores em uma matriz de dados (HAIR JUNIOR et al., 2005). Em termos gerais, a análise de fatores aborda o problema de analisar a estrutura das inter-relações (correlações) entre um grande número de variáveis, definindo um conjunto de dimensões latentes comuns, chamado de fatores. Com a análise fatorial, o pesquisador pode primeiro identificar as dimensões separadas da estrutura, e então determinar o grau em que cada variável é explicada por cada dimensão. Uma vez que essas dimensões e a explicação de cada variável estejam determinadas, os principais objetivos da análise fatorial são conseguidos, isto é, a redução ou resumo de dados e o estudo da variação em uma quantidade de variáveis originais usando um número menor de fatores (JOHNSON; WICHERN, 2007).

Os fatores podem ser não correlacionados (fatores ortogonais) ou correlacionados (fatores oblíquos). As variáveis são agrupadas por meio de suas correlações, ou seja, aquelas pertencentes a um mesmo grupo serão fortemente correlacionadas entre si, mas pouco correlacionadas com as variáveis de outro grupo. Cada grupo de variáveis representará um fator (JOHNSON; WICHERN, 2007).

Seja $\mathbf{Z}$ um vetor de variáveis aleatórias, com matriz de covariância $\mathbf{\Sigma}$, pode-se representar o modelo fatorial como:

$$Z=\mu+\mathbf{\Lambda f}+\boldsymbol{\delta}, \quad (8)$$

em que

$\boldsymbol{\mu}$: representa o vetor de médias;

$\boldsymbol{\Lambda}$: matriz qxm de cargas fatoriais;

$\mathbf{f}$: vetor $mx1$ de fatores comuns;

$\boldsymbol{\delta}$: o vetor $qx1$ de variâncias específicas;

Na forma mais comum de análise fatorial, as colunas de $\boldsymbol{\Lambda}$ são ortogonais, ou seja, $\gamma_i^{'}\gamma_j = 0$ para $i \neq j$, em que γ_i é a i-ésima coluna de $\boldsymbol{\Lambda}$. Daí que os elementos de $\mathbf{f}$ são não correlacionados. Além disso, os fatores comuns são assumidos ter variância unitária, isto é, $Var(\mathbf{f}) = I$. As colunas γ_i são determinadas como os autovetores correspondentes de $\boldsymbol{\Sigma}$, escalado pela raiz quadrada dos respectivos autovalores. No entanto, $\boldsymbol{\Lambda}$ não é único e é frequentemente alvo de uma transformação ortogonal para se obter cargas fatoriais interpretáveis, ao invés daqueles derivados a partir dos autovetores. Finalmente, os fatores específicos (erros) δ_i são assumidos como distribuídos de forma independente com variâncias heterogêneas ψ_i, sendo os vetores $\mathbf{f}$ e $\boldsymbol{\delta}$ não correlacionados. Isso dá a matriz de covariância de $\mathbf{Z}$ sob o modelo de FA:

$$Var(\mathbf{Z}) = \Sigma_{\text{FA}} = \Lambda\Lambda' + \Psi, \quad (9)$$

em que

$\boldsymbol{\Psi} = \mathbf{diag}(\boldsymbol{\psi}_\mathbf{i})$ é uma matriz diagonal de variâncias específicas. Isto implica que todas as covariâncias entre os níveis de $\mathbf{Z}$ são devidos aos fatores comuns, enquanto que os fatores específicos explicam a variação adicional de elementos individuais de $\mathbf{Z}$ (MEYER, 2009). Para m fatores comuns, este descreve $q(q+1)/2$ elementos de Σ_{FA} por meio de $p = q + mq - m(m-1)/2$ parâmetros, que consistem em q variâncias específicas ψ_i e $m(2q-m+1)/2$

elementos de $\mathbf{\Lambda}$ e os restantes $m(m-1)/2$ elementos determinados por restrições de ortogonalidade.

Para valores pequenos de $\mathbf{m}$, um modelo FA oferece uma maneira parcimoniosa de modelar as covariâncias entre um considerável número de variáveis. Como p não pode exceder o número de parâmetros no caso não estruturado, o número de fatores comuns $q(q+1)/2$ que podem ser definidos é restrito.

Se todas as variâncias específicas ψ_i são diferentes de zero, o número mínimo de características para as quais é imposta uma estrutura FA para redução no número de parâmetros é q = 4. A estrutura FA para a variância de $\mathbf{Z}$ é mais apropriada se todos os caracteres q envolvidos são correlacionados de forma relativamente uniforme. Nesse caso, um pequeno número de fatores, é geralmente suficiente para modelar as covariâncias entre os elementos de $\mathbf{Z}$. O modelo FA inclui muitas estruturas de covariância corriqueiramente utilizadas para modelar problemas de interação G x E em casos especiais. O mais simples cenário é a estrutura de simetria composta, ou seja, $\mathbf{\Sigma} = \sigma^2 11' + \psi I$, que é um modelo de FA com um único fator comum e $\mathbf{\Lambda} = \sigma 1$ (onde 1 refere-se a um vetor com todos os elementos igual à um) e variâncias específica iguais ψ para todas as variáveis (MEYER, 2009). Jennrich e Schluchte (1986) propuseram uma estrutura FA como opção para modelar as covariâncias de dados entre medidas repetidas e exemplos típicos em que tal modelo seja adequado são aquelas em que as mesmas medidas são tomadas em diferentes circunstâncias (como diferentes locais para interação G X E).

2.3 Efeitos de genótipos fixos ou aleatórios

A classificação dos efeitos de genótipo em fixo ou aleatório interfere diretamente na definição do modelo e, consequentemente, na utilização de diferentes funções para ranquear os genótipos. Embora a distinção entre as duas

abordagens (fixos vs aleatório) possa parecer sutil e até mesmo semântica, eles levam à diferentes modelos lineares e, portanto, diferentes funções dos dados que são utilizados para classificar os genótipos. Isto resulta em diferentes propriedades dos critérios de classificação entre as abordagens- aleatórias e fixas (WHITE; HODGE, 1989).

Se genótipos são tomados como fixos as suas médias serão estimadas usando os melhores estimadores lineares não viesados (BLUEs) baseados em mínimos quadrados generalizados e caso sejam considerados aleatórios serão usados melhores preditores lineares não viesados (BLUPs) (FISCHER et al., 2009; HENDERSON, 1984; SEARLE; CASELLA; MCCULLOCH, 1992).

A suposição de que efeitos de genótipos sejam aleatórios tem sido debatida na literatura. Um argumento, contra, essa suposição é que os genótipos geralmente não são uma amostra aleatória de uma população definida, já que, genótipos em teste são o resultado de um processo de seleção. Embora, na maioria dos casos isto seja verdade pode-se considerar que os genótipos em teste são uma amostra aleatória de alguma população hipotética de genótipos que poderiam ter surgido como um resultado do processo de seleção levando aos genótipos atualmente disponíveis (PIEPHO; MÖHRING, 2006, 2010).

Se os efeitos de genótipos são tomados como aleatórios ou como fixos depende da finalidade da análise (SMITH; CULLIS; THOMPSON, 2001, 2005) e sobre a forma como os genótipos foram gerados. Se o interesse for na estimativa da média de genótipos, são tomados como fixo. Se o foco está em predizer o valor genético potencial dos genótipos em futuros experimentos, estes podem ser considerados como aleatórios de uma população base (FISCHER et al., 2009; HENDERSON, 1984; RESENDE, 2007).

No melhoramento de plantas a predição de valores genéticos é de interesse, mas devido à seleção, e assegurar condições ideais a população-base como, cruzamentos ao acaso, equilíbrio de ligação e falta de endogamia, ela não existe (FISCHER et al., 2009; PIEPHO et al., 2008). Até agora, os melhoristas de plantas, muitas vezes vem tratando genótipos como um efeito fixo, ignorando todas as covariâncias entre os genótipos oriundos de descendências ou processo de avaliação. Assumindo genótipos como efeitos aleatórios, é possível obter

predições dos genótipos e dos efeitos da interação aleatória GxE. Além da separação de efeitos genéticos em aditivos e não aditivos (PIEPHO; MÖHRING, 2010). A desvantagem de tomar o efeito genético como aleatório é a exigência de se estimar um componente de variância. Se há pouca informação para estimar o componente de variância, tanto a estimativa de componentes de variância e os BLUPs são incertos. Assim, Searle, Casella e McCulloch (1992) propuseram a considerar os efeitos como aleatórios, se o número de genótipos for grande. Eeuwijk (1995) sugeriu ter pelo menos dez graus de liberdade para estimar os componentes de variância.

2.4 Predição dos efeitos aleatórios (G e GxE)

A predição de uma observação futura é um problema que tem sido extensivamente estudado.

Os valores genéticos são variáveis aleatórias não observáveis, preditas a partir dos valores fenotípicos observáveis, comumente usados nos programas de melhoramento de plantas. A sua predição, que pode ser feita de forma pontual ou intervalar, deve ser precisa e acurada, pois aumentam os ganhos pretendidos, diminuindo as possibilidades de erro na seleção (PINTO JÚNIOR, 2004). A predição pontual fornece os valores genéticos preditos, ao passo que a intervalar inclui os intervalos de confiança dos valores e dos ganhos genéticos, propiciando uma recomendação mais segura dos indivíduos envolvidos e, portanto, deve ser preferencial (RESENDE, 2002).

Os valores genéticos preditos, entretanto, não são iguais aos valores genéticos verdadeiros dos indivíduos. Conforme Vleck, Pollak e Oltenacu (1987), a proximidade entre esses dois pode ser avaliada com base na estatística denominada acurácia, a qual se refere à correlação entre os valores genéticos preditos e verdadeiros dos indivíduos.

Resende (2002) argumenta que o sucesso do melhoramento genético depende da adoção de procedimentos de seleção acurados, e que a estruturação dos mesmos baseia-se na estimação dos componentes de variação e predição dos

valores genéticos visando à avaliação genética dos candidatos a seleção. O procedimento ótimo de predição de valores genéticos e seleção. usado no melhoramento de espécies é o BLUP (*Best Linear Unbiesed Prediction*) para dados balanceados e desbalanceados. O BLUP ajusta os dados para efeitos ambientais identificáveis e simultaneamente prediz os valores genéticos dos indivíduos. Pois os BLUPs são calculados com base na verdadeira forma para a matriz de variância e covariâncias genética.

A seleção é geralmente exercida em vários caracteres. No melhoramento animal, a fim de evitar o viés devido à seleção, é comum realizar as análises utilizando modelo misto multicaracter (multitrait mixed model) (HENDERSON; QUAAS, 1976; MRODE, 1996; PIEPHO et al., 2008). Neste contexto, para a predição dos efeitos aleatórios, têm-se usado o BLUP Multicaracter (multivariado) por apresentar vantagem quando os caracteres são altamente correlacionados. Entretanto, essa abordagem apresenta a desvantagem de poder tornar as equações do modelo misto muito extensas.

No melhoramento vegetal a abordagem multivariada tem sido utilizada com culturas perenes, sendo muito raro no melhoramento de culturas anuais (PIEPHO et al., 2008). Simeão et al. (2002) utilizaram BLUP multivariado considerando ambientes como caracteres diferentes em erva-mate (*Ilex paraguariensis*). O BLUP multivariado considera adequadamente a questão da interação G x E e heterogeneidade de variâncias, permitindo também explorar as diferentes herdabilidades entre os ambientes.

Embora o modelo misto multivariado seja o procedimento mais recomendando para lidar com heterogeneidade de variâncias e interação G x E, uma possível heterogeneidade de variância entre blocos dentro de locais não é levada em consideração. Este fato pode conduzir à seleção de maior número de indivíduos nos blocos, mais variáveis fenotipicamente, o que é incorreto quando na verdade a herdabilidade nesses blocos não é superior (RESENDE, 2007).

Na prática, os componentes de variância devem ser estimados com a maior precisão possível, empregando-se o procedimento padrão no contexto dos modelos lineares mistos, que é o da máxima verossimilhança restrita (REML),

conforme Searle, Casella e McCulloch (1992). Tal procedimento permite a seleção de indivíduos com os maiores valores genéticos, independentemente de sua procedência, sendo esta a estratégia mais plausível em termos seletivos, em detrimento da seleção de procedências (RESENDE, 2007).

O impacto da escolha do modelo na predição dos efeitos G x E tem sido considerado por Crossa et al. (2006), Kelly et al. (2007) e Piepho (1998), onde as técnicas de validação cruzada em cinco conjuntos de dados MET foram utilizados para comparar BLUPs baseado em uma gama de modelos, em termos da sua precisão preditiva para "preencher" as células na tabela G x E. Os modelos considerados incluem fator analítico (FA), e modelos de variância não estruturadas (UN). Kelly et al. (2007) e Piepho (1998) concluíram que a precisão de previsão BLUPs a partir dos modelos de FA foi superior à do modelo uniforme, mas os resultados também parecem indicar que eles são geralmente inferiores à dos modelos de variância não-estruturada. Note-se que para o modelo FA no Piepho (1998), uma variância comum foi assumida pela falta ajuste, enquanto Smith, Cullis e Thompson (2001) permitiu uma separação (a chamada especificação) de variância para cada ensaio.

2.5 Interação Genótipo x Ambiente (G x E)

Os experimentos multi-locais ou multiambientes (MET) são um tipo especial de experimentos, muito usados em melhoramento genético de plantas, nos quais alguns genótipos são avaliados em diferentes locais. Nesses estudos é comum encontrar-se uma resposta diferenciada na resposta dos genótipos aos diferentes ambientes, que recebe o nome de interação genótipo x ambientes ou G x E.

Na presença da interação, os resultados das avaliações podem variar de um ambiente para o outro, ocasionando mudanças na posição relativa dos genótipos ou mesmo na magnitude das suas diferenças.

Para Santos (2009) e Vencovsky e Barriga (1992) é muito importante o conhecimento da interação G x E, seja do tipo genótipos x locais ou genótipos x

anos ou outros, pois estes orientam o planejamento e adoção de estratégias do melhoramento e recomendação de cultivares, além de ser determinante na estabilidade fenotípica dos genótipos para uma região.

O conhecimento da interação G x E é de extrema importância nos programas de melhoramento, pois o seu conhecimento permite a seleção de genótipos com ampla adaptação ou específica, escolher o local da seleção e determinar o número ideal de ambientes e genótipos para seleção (FOX; CROSSA; ROMAGOSA, 1997; SANTOS, 2009).

a) A interação G x E

O caráter de um indivíduo é o conjunto de informações biológicas que o identifica. As diferentes manifestações de um dado caráter definem o fenótipo (F). O fenótipo por sua vez, é influenciado pelo genótipo (G), que é a constituição genética de um indivíduo, e pelo ambiente (E) que pode ser definido como o conjunto das condições que afetam o crescimento e desenvolvimento do organismo (RAMALHO et al., 2012).

O F é função do G, do E e da interação G x E. Esse último componente ocorre devido à diferenciação do comportamento dos genótipos nos vários ambientes de cultivo.

No processo de avaliação e desenvolvimento de cultivares, o conhecimento da interação G x E é de grande importância para a seleção e/ou indicação dos cultivares para os diferentes ambientes de cultivo.

A existência ou não da interação G x E está representada nas Figuras 1, 2 e 3, onde estão exemplificadas quatro situações de respostas das cultivares as condições ambientais. Na Figura 1 os genótipos apresentam desempenhos relativos semelhantes nos dois ambientes (E1 e E2). Portanto, não há interação e a recomendação do melhor genótipo é a mesma para os dois ambientes.

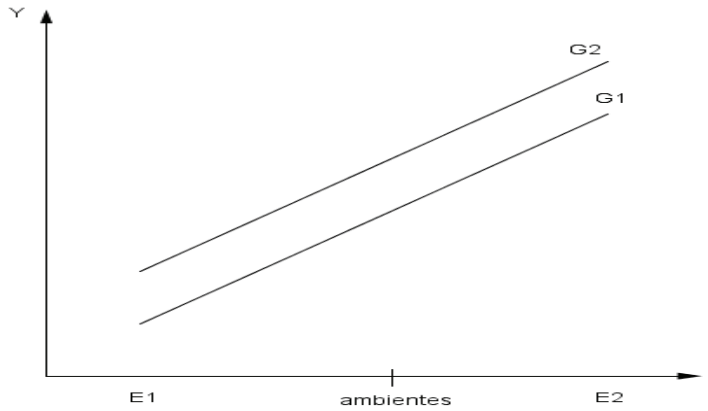


Figura 1 Comportamento de dois genótipos (*G*1 e *G*2) em duas condições ambientais (*E*1 e *E*2) com ausência de interação

Na Figura 2 o desempenho relativo dos dois genótipos (G1 e G2) é diferente nos dois ambientes, pois o G1 tem resposta mais acentuada à melhoria do ambiente, considerando-se o E2 melhor do que El. Neste caso ocorre interação. No entanto, não é um grande problema, porque a classificação dos genótipos nos dois ambientes não é alterada e, por esta razão, é denominada de interação simples. Os dois genótipos poderão ser recomendados para os dois ambientes ou será recomendado somente o melhor genótipo no caso de a diferença ser suficientemente grande para tal.

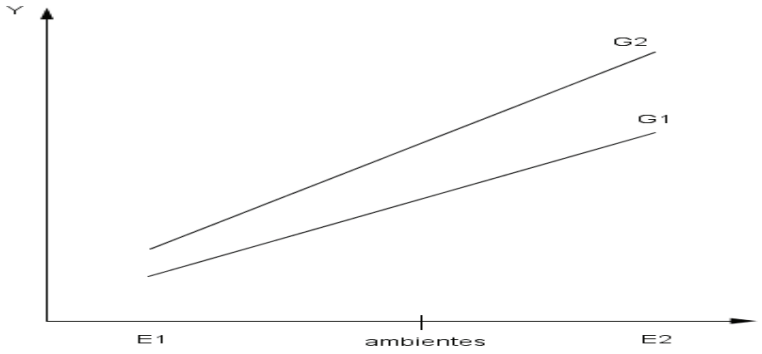


Figura 2 Comportamento de dois genótipos (G1 e G2) em duas condições ambientais (E1 e E2) com interação simples ou quantitativa

Na Figura 3 (a e b) observa-se uma inversão de comportamento das cultivares nos dois ambientes. O G1 foi superior no El e inferior no E2 (figura 3a). Esta corresponde a uma situação de interação complexa (cruzada ou

qualitativa), onde normalmente, existe um genótipo mais adaptado para cada ambiente específico. Nessa Figura 3 pode-se observar que o G1 é mais indicado para o El e o G2 para o E2.

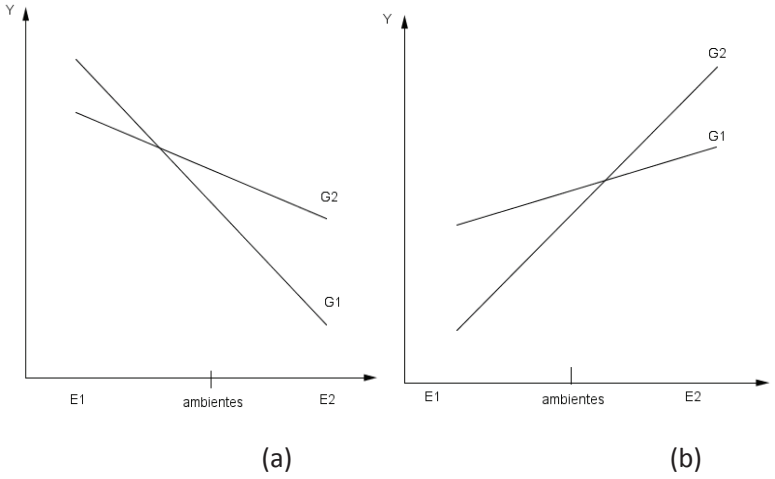


Figura 3 Comportamento de dois genótipos (G1 e G2) em duas condições ambientais (E1 e E2) com interação cruzada ou qualitativa

As respostas diferenciadas dos genótipos às variações ambientais tais como, oscilação de temperatura, altitude, ocorrência de doenças, tipo e fertilidade do solo, entre outras, são atribuídas às diferentes constituições genotípicas de cada material, que conferem maior ou menor adaptabilidade e estabilidade de produção.

Quando se consideram vários genótipos avaliados em vários ambientes, a combinação de situações como as das Figuras 1, 2 e 3 formam um emaranhado de situações, difícil de ser interpretado, exigindo métodos adequados de análise da interação G x E.

Existe uma concordância geral entre melhoristas de plantas de que a interação G x E tem um importante significado para a obtenção de variedades superiores. Porque a existência da mesma produz uma barreira de dificuldades aos melhoristas na identificação de genótipos superiores, tanto no processo de

seleção, quanto no processo de recomendação de cultivares. Essa interação indica que o comportamento dos genótipos nos experimentos depende principalmente das condições ambientais a que são submetidos. Assim, a resposta obtida de um genótipo, em comparação a outro, é variável, sendo que essas variações se apresentam devido à mudança de ambientes (ARAÚJO; DIAS, 2006).

b) Adaptabilidade e estabilidade

A presença da interação G x E interfere de forma intensa nos programas de melhoramento, pois em uma situação ideal as cultivares deveriam possuir adaptabilidade a vários ambientes e terem boa estabilidade. Porém, o fator interação faz com que, na maioria das vezes, as cultivares sejam indicadas a ambientes específicos, por possuírem maior adaptabilidade em algumas condições ambientais (CAMPBELL; JONES, 2005). O termo adaptabilidade refere-se à capacidade dos genótipos responderem de forma positiva ao estímulo do ambiente, enquanto a estabilidade refere-se à capacidade dos genótipos desempenharem um comportamento previsível em função do estímulo do ambiente (CRUZ; REGAZZI; CARNEIRO, 2004).

A condução de experimentos em vários locais é necessária para a quantificação da interação G x E. Os estudos dos parâmetros de adaptabilidade e estabilidade fenotípica dos genótipos têm sido de grande contribuição nesse aspecto, pois fornecem informações sobre o comportamento de cada genótipo em várias condições ambientais (CRUZ; REGAZZI; CARNEIRO, 2004).

Diferentes metodologias para avaliar a adaptabilidade e a estabilidade têm sido desenvolvidas e/ou aprimoradas. Tais procedimentos se baseiam em análises de variância, regressão linear, regressão não linear, análise multivariada e estatística não paramétrica (CROSSA, 1990).

Na prática, os programas de melhoramento genético envolvem, pelo menos, três etapas: escolha dos parentais que darão origem à população base;

seleção das progênies superiores dessa população; e sua avaliação em um grande número de ambientes.

Quando se avaliam materiais geneticamente distintos, em uma série de ambientes, o componente interação G x E aparece, normalmente, afetando o ganho com a seleção (MAIA et al., 2009).

Adaptações específicas de genótipos a ambientes, de acordo com Gauch e Zobel (1996), podem fazer a diferença entre uma boa e uma excelente cultivar. Pela mesma razão, sob o ponto de vista de recursos genéticos, a exploração dessa interação é interessante para manter a variabilidade genética da espécie. Contudo, para que seja possível tirar proveito desses efeitos positivos, de acordo com Duarte e Zimmermam (1995), é preciso se dispor de metodologias estatísticas adequadas para se estimar e explorar a interação, permitindo, assim, recomendações regionalizadas.

Resende (2004) salienta que um modelo multivariado, considerando-se todos os locais simultaneamente, é adequado para a seleção, tendo como alvo a produtividade média ao longo de todos os ambientes. No entanto, para o mesmo autor, um modelo mais completo pode permitir inferências adicionais, tais como: seleção de genótipos específicos para cada local; seleção de genótipos estáveis por meio dos locais; seleção de genótipos responsivos (com alta adaptabilidade) à melhoria do ambiente; e seleção pelos três atributos (produtividade, estabilidade e adaptabilidade). Simultaneamente, esse tipo de seleção pode ser realizado pelo método da média harmônica da performance relativa dos valores genéticos (MHPRVG), que classifica os efeitos genotípicos como aleatórios e, portanto, fornece estabilidade e adaptabilidade genotípica e não fenotípica.

Existem atualmente várias metodologias de análise de adaptabilidade e estabilidade citadas na literatura. Cargnelutti Filho et al. (2009) classificaram essas metodologias em:

1) as que são baseadas em análise de variância e dão informação sobre a estabilidade dos genótipos avaliados;

2) aquelas que usam a regressão linear e informam sobre a adaptabilidade e a estabilidade dos genótipos;

3) as que se baseiam na regressão bissegmentada não-linear e linear;

4) as de estatísticas não-paramétricas, e

5) as que analisam os efeitos principais aditivos e a interação multiplicativa (AMMI).

Todas elas dão informações importantes. No entanto, são limitadas quando se tem dados desbalanceados, delineamentos experimentais não ortogonais e heterogeneidade de variâncias entre os locais onde são conduzidos os ensaios (RESENDE, 2004).

d) Métodos estatísticos para estudo da interação G x E

A existência de interação G x E têm sido reconhecidas há muito tempo de acordo com Freeman e Perkins (1971), sendo a referência mais antiga feita por Fisher e Mackenzie em 1923. Desde então, muitos trabalhos têm sido feitos para análises estatísticas da interação genótipos x ambientes, seja por estatísticos, agrônomos, melhoristas e geneticistas (ARAÚJO; DIAS, 2006).

A análise de variância conjunta é o método mais comum para identificar a existência de interação G x E a partir de ensaios MET. Se a interação G x E for significativa, um ou mais dos vários métodos paramedir estabilidade de genótipos pode ser usado para identificar genótipos estáveis.

Existem vários métodos para a análise da interação G x E, os quais podem ser classificados em quatro grupos: a análise de componentes de variância, análise de estabilidade, métodos multivariados e métodos qualitativos. A seguir serão descritos alguns desses métodos.

i. Análise de variância convencional

A análise conjunta de experimentos é de grande interesse, em especial, para os melhoristas, porque as estimativas de componentes de

variância em experimentos conduzidos em único ambiente costumam ser superestimadas, pois o fator ambiente costuma influenciar nesses casos. Desta forma vários autores vêm destacando a importância do estudo do componente da interação G x E (CROSSA, 1990).

Para avaliar a importância e a magnitude das interações podem ser utilizados métodos de análise de variância. Uma das formas seria a análise de variância conjunta em blocos casualizados.

A análise de variância conjunta dos dados observados (y_{ijk}), que pode ser o rendimento do genótipo i no ambiente j no bloco k, é executada considerando-se o modelo estatístico:

$$y_{ijk} = \mu + b_{k(j)} + g_i + e_j + (ge)_{ij} + \varepsilon_{ijk}\,, \qquad (10)$$

em que

μ : é uma constante inerente a cada observação;

$b_{k(j)}$: é o efeito do k-ésimo bloco dentro j-ésimo ambiente;

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_{ijk} : é erro experimental associado ao i-ésimo genótipo, no j-ésimo ambiente e no k-ésimo bloco, com $\varepsilon_{ijk} \sim N(0;\sigma^2)$.

A interação não aditiva, conforme definido em (16) implica que o valor esperado do i-ésimo genótipo no ambiente j (Y_{ij}) depende 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 nos ambientes devem ser homogêneas para testar diferenças genotípicas. Se as variâncias dos erros são heterogêneas, essa análise está sujeita a críticas, como a

de que o teste F dos quadrados médios de G x E contra as variações de erro apresenta viés para resultados significativos (CROSSA, 1990).

Um teste correto para a significância é realizado ponderando-se cada genótipo com o inverso da sua variância residual estimada. Essa análise ponderada atribui menos pesos para ambientes que têm um quadrado médio residual elevado. Uma desvantagem da análise ponderada é que os pesos podem ser correlacionados com as respostas do rendimento no ambiente. Assim, pode ocorrer, por exemplo, que ambientes com rendimento elevado apresentem maior variância do erro e ambientes com baixos rendimentos apresentem variâncias de erro reduzidas, o que pode mascarar o verdadeiro desempenho de alguns genótipos em certos ambientes (CROSSA, 1990).

Uma das principais deficiências da análise de variância conjunta de ensaios multi-locais é que ela não explora qualquer estrutura subjacente dentro da observação não-aditiva G x E (CROSSA, 1990).

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

A análise de variância dos ensaios multi-locais é útil para estimar componentes de variância relacionadas com diferentes fontes de variação, incluindo genótipos e G x E.

Em geral, a metodologia de componentes de variância é importante em ensaios multi-locais, desde erros na mensuração do desempenho produtivo de um genótipo que surgem em grande parte da interação G x E. Portanto, o conhecimento da magnitude dessa interação é necessário para: (a) obter estimativas eficientes dos efeitos genotípicos e (b) determinar recurso ideal à alocar, como é o número de parcelas e os locais a serem incluídos em estudos futuros.

ii. Metodologia AMMI *(additive main effects and multiplicative interaction)*

O método AMMI surge com a finalidade de estudar detalhadamente as interações (G x E) por meio da decomposição ortogonal da soma de quadrados das interações, fato que o torna vantajoso se comparado aos métodos tradicionais. Além disso, esse método apresenta uma boa capacidade preditiva.

A análise AMMI é uma combinação de métodos univariados (análise de variância) com métodos multivariados (análise de componentes principais e decomposição por valores singulares). Nesse modelo, por meio de uma análise gráfica, em biplot, busca-se identificar, simultaneamente, padrões de interação para genótipos e ambientes.

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 (ZOBEL; WRIGHT; GAUCH, 1988).

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

$$y_{ij} = \mu + g_i + e_j + \sum_{k=1}^{n} \lambda_n \gamma_{ik} \alpha_{jk} + \delta_{ij} + e_{ij};\ i = 1,2,...G\ e\ j = 1,2,...E\ , \qquad (11)$$

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

$$\sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk} + \delta_{ij}\ , \qquad (12)$$

em que

y_{ij} : é a média da produção do genótipo i no ambiente j,

μ : é uma constante inerente a cada observação;

g_i : é o efeito do genótipo i;

e_j é o efeito do ambiente j,

λ_n : é o n-ésimo valor singular de ge (escalar);

Logo, γ_{ik} e α_{jk} são os elementos relacionados ao genótipo i e ao ambiente j dos vetores singulares γ_k e α'_k, respectivamente.

O índice k (k= 1,2,..., n); em que:

$$p = min\ \{G-1, E-1\}, \qquad (13)$$

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 DVS (Decomposição de Valores Singulares); deixando-se um resíduo adicional denotado por ρ_{ij}. Para n=p não se tem mais a aproximação e sim uma decomposição exata da matriz, implicando em ρ_{ij} nulo.

Sob as restrições de identificabilidade:

$$\Sigma_i g_i = \Sigma_j e_j = \Sigma_i (ge)_{ij} = \Sigma_j (ge)_{ij} = 0, \qquad (14)$$

além da média geral (μ) e do erro experimental médio (e_{ij}), os demais termos do modelo resultam da chamada decomposição por valores singulares (DVS) da matriz de interações:

$$GE_{(gxa)} = \left[(g\hat{e})_{ij} \right].$$

A matriz de interações é obtida como resíduo do ajuste aos efeitos principais por ANOVA, aplicada à matriz de médias $Y_{(gxa)} = \left[Y_{ij} \right]$, $\gamma_{k(gx1)}$ e $\alpha'_{k(1xa)}$ são os respectivos vetores singulares (vetor coluna e vetor linha) associados a λ_k (DUARTE; VENCOVSKY, 1999; PIEPHO, 1995).

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

$$y = \underbrace{\mu + g_i + e_j}_{\text{aditiva}} + \underbrace{\sum_{k=1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}}_{(ge)_{ij}:\text{multiplicativa}} + \varepsilon_{ij} \qquad (15)$$

Sob o ponto de vista da análise de componentes principais, além dos termos já definidos anteriormente, tem-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) $\Rightarrow \lambda_k^2$ 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 .

Note-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 G x E, 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 $SQ_{G \times E}$, 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: $\sum_{k=n+1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}$.

Como em ACP (Análise de Componentes Principais), estes 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 $\sum_{k=n+1}^{p} \lambda_k \gamma_{ik} \alpha_{jk}$, respectivamente (DUARTE; VENCOVSKY, 1999).

iii. Análise de fatores sob modelos multiplicativos mistos (FAMM)

A análise de grupos de experimentos ou de experimentos conduzidos em múltiplos ambientes (MET) tem sido tradicionalmente baseada em modelos simples, os quais assumem homogeneidade de variância residual entre os experimentos, independência de erros dentro de ensaio, efeitos da interação G x E como um grupo de efeitos aleatórios independentes.

A modelagem de efeitos da interação G x E para ensaios multi-ambientes (METs) dentro de uma estrutura de modelo misto é agora uma prática comum em muitos programas de melhoramento de plantas.

O modelo misto tradicional é dado por:

$$y = Xb + Zu + e, \qquad (16)$$

em que

y : vetor de observações;

b : vetor dos efeitos fixos, com matriz de incidência X;

u : vetor dos efeitos aleatórios, com matriz de incidência Z, $\mathbf{u} \sim \mathbf{N(0, \Sigma)}$

e : vetor de erros aleatórios, $\mathbf{e} \sim \mathbf{N(0, R)}$.

O modelo fator de analítico (FA) é uma forma parcimoniosa usada para aproximar a forma totalmente não estruturada da matriz de variância-covariância genética ($\mathbf{\Sigma}$) no modelo de dados MET (KELLY et al., 2007).

Uma extensão dos modelos mistos para incorporar a análise de fatores (modelo misto fator analítico) (FAMM) pode ser escrito como:

$$y=Xb+Z[Lf+\delta]+e \quad , \qquad (17)$$

com $u=[Lf+\delta]$,

em que

$\mathbf{L} = \mathbf{\Lambda} \otimes \mathbf{I_g}$ a matriz de cargas fatoriais;

$\mathbf{f}$ é o vetor de escores fatoriais para os indivíduos nos ambientes;

$\boldsymbol{\delta}$ é o vetor de erros representando a falta de ajuste do modelo fatorial.

Sob esse modelo, a matriz de covariância genética é dada por

$$\Sigma=\Lambda\Lambda'+\psi, \qquad (18)$$

em que

$$\Lambda\Lambda'=VD_{\alpha}V', \qquad (19)$$

$\mathbf{D}_{\alpha}$ é a matriz diagonal dos m autovalores e $\mathbf{V}$ é a matriz dos autovetores. Escolhendo-se V e D_{α} referentes apenas à dimensão m esse modelo misto é reduzido e ajusta somente os m fatores. Na técnica FAMM, a estrutura de covariância é simplificada para

$$\Sigma=\Lambda_p\Lambda_p^{'}+\Psi, \qquad (20)$$

em que:2

$\mathbf{\Lambda}_p$: é a matriz dos carregamentos dos fatores nas variáveis;

$\mathbf{\Psi}$: é a matriz diagonal de variâncias específicas $Var(\delta_i)$ (RESENDE; THOMPSON, 2004).

A metodologia de modelos mistos padrão pode ser usada para estimar autovalores e autovetores diretamente sem a necessidade de se estimar $\mathbf{\Sigma}$ completa. A principal diferença para o modelo multivariado misto tradicional refere-se ao fato de que os parâmetros a serem estimados fazem parte da matriz de incidência dos efeitos genéticos aleatórios. Como a distribuição de $[\mathbf{\Lambda} \otimes \mathbf{I_g}]\mathbf{f}$ é singular, isto conduz à estimação sob posto reduzido, restrições devem ser

impostas aos parâmetros do modelo fator analítico (RESENDE, 2007). Uma maior aplicação dos modelos fator analíticos mistos é na análise de experimentos multi-ambientes no estudo da interação G x E (já discutido em 2.6), e torna-se melhor nessa análise por reunir em um só método os procedimentos de análise multivariada, análise de adaptabilidade e estabilidade e modelos mistos.

Uma característica fundamental do modelo de FA para os dados MET é a capacidade de generalização da estrutura de variância associado para efeitos G x E, seja no ambiente ou na dimensão do genótipo. 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 (SMITH; CULLIS; THOMPSON, 2005).

Smith, Cullis e Thompson (2001) utilizam o modelo FA neste contexto em que a análise foi motivada pela abordagem da genética quantitativa para interação G x E, como explicado no Falconer e Mackay (1996). Falconer e Mackay (1996) em Smith, Cullis e Thompson (2005) afirmam que

> o conceito de correlação genética pode ser aplicado à solução de alguns problemas relacionados com a interação genótipo e ambiente [...] um caráter medido em dois ambientes diferentes deve ser considerado não como um personagem, mas como dois [...] Se a correlação genética entre eles é elevada, o desempenho em dois ambientes diferentes representa quase o mesmo carácter [...] Se for baixa, então os caracteres são, em grande medida diferente.

Assim, Smith, Cullis e Thompson (2001) utilizam um modelo de FA para aproximar uma matriz não-estruturada para a dimensão do ambiente de ($\boldsymbol{\Sigma}$)(isto é, a matriz de variâncias e covariâncias entre ambientes) (SMITH; CULLIS; THOMPSON, 2005). Kelly et al. (2007), utilizando a abordagem de Smith, Cullis e Thompson (2001, 2005), demonstraram que o modelo FA é geralmente o melhor modelo para o ajuste de uma série de conjuntos de dados em estudos iniciais de um programa de melhoramento. Além disso, demonstram a superioridade do modelo de FA em conseguir o objetivo mais comum de MET, nomeadamente a seleção de genótipos superiores, por meio do uso dos

melhores preditores lineares não viesados (BLUPs) de efeitos de genótipo em cada ambiente, considerados individualmente ou como uma média ponderada entre ambientes.

Os modelos FAMM propiciam uma abordagem realística completa para análise de dados de múltiplos experimentos.

Apesar de as recomendações de Piepho (1997, 1998) e Smith, Cullis e Thompson (2001), os modelos de FA não são amplamente utilizados fora da Austrália para a análise regular de dados MET (KELLY et al., 2007).

2.6 Estrutura Fator Analítica

Um método associado à avaliação de vários tratamentos ou genótipos e vários ambientes é dado em (10).

O efeito da constante é fixo, o efeito do ambiente pode ser considerado fixo ou aleatório e os demais efeitos são considerados como aleatórios. Um modelo referente aos efeitos aleatórios em cada ambiente pode ser representado por:

$$Y_{ijk} = \mu + g_{ij} + e_j + \varepsilon_{ijk} \qquad (21)$$

em que:

μ : é uma constante inerente a cada observação;

g_{ij} : é o efeito do i-ésimo genótipo;

e_j : é o efeito do j-ésimo ambiente;

ε_{ijk} : é erro experimental associado ao i-ésimo genótipo, no j-ésimo ambiente e no k-ésimo bloco, com $\varepsilon_{ijk} \sim N(0;\sigma^2)$.

Na análise de experimentos multi-ambientes (MET), o uso da análise de fatores pode propiciar uma classe de estruturas para a matriz de variância e covariância G_0 , associada aos efeitos g_{ij}. O modelo de análise é postulado em termos de efeitos genotípicos não observáveis em diferentes ambientes:

$$g_{ij} = \sum_{r=1}^{k} \lambda_{jr} f_{ir} + \delta_{ij}, \qquad (22)$$

em que

g_{ij} : efeito do genótipo i no ambiente j;

λ_{jr} : carregamento do fator r no ambiente j;

f_{ir} : escore para o genótipo i no fator r;

δ_{ij} : erro representando a falta de ajuste do modelo.

O modelo FA é apresentado com base em Resende e Thompson (2004) e Smith, Cullis e Thompson (2001, 2005). Aplicado a G genótipos e E ambientes, o modelo de fator analítico postula dependência em um conjunto de fatores hipotéticos aleatórios $f_r^{(gx1)}, (r = 1...k < s)$. Em notação vetorial, o modelo de análise de fatores para estes efeitos em diferentes ambientes é:

$$u_s = (\lambda_1 \otimes I_g) f_1 + ... + (\lambda_k \otimes I_g) f_k + \delta \qquad (23)$$

Onde:

$\lambda_r^{(sx1)}$: Cargas ou pesos dos fatores nos ambientes;

$\delta^{(gs\ x1)}$: vetor de resíduos ou a falta de ajuste para o modelo (também chamado de vetor de fatores específico).

De um modo compacto, o modelo é:

$$u_s = (\lambda_1 \otimes I_g) f + \delta \qquad (24)$$

em que :

$\Lambda^{(sxk)} = [\lambda_1 ... \lambda_k]$

$f^{(gkx1)} = (f_1^{'}, f_2^{'} ... f_k^{'})'$

A distribuição conjunta de f e δ é dada por:

$$\begin{bmatrix} f \\ \delta \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_k \otimes I_g & 0 \\ 0 & \Psi \otimes I_g \end{bmatrix}\right),$$

em que:

$$\Psi = diag(\psi_1 ... \psi_p) \ ;$$

ψ_i variância específica para o i-ésimo ensaio.

A matriz de variância para efeitos de genótipos nos ambientes é dada por

$$\text{var}(u_s) = (\Lambda \otimes I_g)\,\text{var}(f)(\Lambda' \otimes I_g + \text{var}(\delta) = (\Lambda\Lambda' + \Psi) \otimes I_g \qquad (25)$$

O modelo para efeitos de genótipos em cada ambiente conduz a um modelo de G em que:

$\sigma_{gij} = \sum_{i=1}^{k} \lambda_{jr}^2 + \psi_j$: variância genotípica em ambiente j;

$\sigma_{gij} = \sum_{i=1}^{k} \lambda_{jr}\lambda_{j'r}$: covariância genotípica entre ambientes j e j ';

$\rho_{gij} = \sum_{i=1}^{k} \lambda_{jr}\lambda_{j'r} / [(\sum_{i=1}^{k} \lambda_{jr}^2 + \psi_j)(\sum_{i=1}^{k} \lambda_{j'r}^2 + \psi_{j'})]^{1/2}$: correlação genotípica entre os ambientes j e j '

A equação (24) para u_s tem a forma de uma regressão (aleatória) em k covariáveis ambiental $\lambda_1 ... \lambda_k$, na qual todas as regressões passam pela origem. Pode ser mais apropriado para permitir que o intercepto (não-zero) separado para cada genótipo. Isto é equivalente ao modelo com efeitos de genótipos principais, u_g e um modelo fator analítico k para interação G x E. Em seguida, a expressão de u_g torna-se:

$$u_g = (1_s \otimes I_g)g + g = ge\,(1_s \otimes I_g)g + (\Lambda \otimes I_g)f + \delta \qquad (26)$$

.

Vetor g tem média zero e variância $\delta_g^2 I$ ou $\delta_g^2 \mathbf{A}$, onde $\mathbf{A}$ é uma matriz de correlação genética ou de parantesco. O modelo pode ser escrito como:

$$u_g = \left(\sigma_g 1_s \otimes I_g\right) f_0 + \left(\Lambda \otimes I_g\right) f + \delta = \left(\Lambda_g \otimes I_g\right) f_g + \delta \quad (27)$$

em que:

$$\Lambda_g^{s(k+1)} = [\sigma_g 1_s \Lambda]; f_0 = g / \sigma_g; f_g' = (f'_0 f') \quad (28)$$

Assim estimação BLUEs dos efeitos fixos é dada por:

$$\hat{b} = (X'V^{-1}X)^{-1} X'V^{-1}y \quad (29)$$

em que $V = ZR^{-1}Z' + \Sigma^{-1}$ partindo de (10).

Para o modelo fator-analítico, os BLUPs dos escores dos **f** e resíduos **δ** para cada ambiente podem **ser** obtidos em termos de u_g como:

$$\tilde{f} = [\Lambda'(\Lambda\Lambda' + \Psi)^{-1} \otimes I_g]\tilde{u}_g \quad (30)$$

$$\tilde{\delta} = [\Psi(\Lambda\Lambda' + \Psi)^{-1} \otimes I_g]\tilde{u}_g \quad (31)$$

Assim, o modelo com efeitos principais de genótipos e um modelo de fator analítico de ordem k para interações G x E é um caso especial de um modelo fator analítico de ordem (k +1) efeitos de genótipos de análise em cada ambiente, em que as primeiras cargas são restringidas a ser iguais. A característica que distingue as equações para g, dos problemas de padrão e de regressão aleatória multivariada é que ambas as co-variáveis e os coeficientes de regressão são desconhecidos e, por conseguinte, deve ser calculado a partir dos dados. O modelo então é multiplicativo de coeficientes genotípicos e ambientais (conhecido como cargas e escores fatoriais, respectivamente). Aqui reside a analogia com modelos AMMI. No entanto, uma diferença fundamental é que o modelo multiplicativo na equação para g_s acomoda efeitos aleatórios, enquanto AMMI é um modelo de efeitos fixos. Modelos FAMM são também chamados AMMI aleatórios (RESENDE, 2007).

2. 7 Seleção dos modelos FAMM

O objetivo do modelo fator-analítico para efeitos G × E é explicar as covariâncias genéticas entre os E ambientes em termos de um número muito menor de k fatores (desconhecido) $f_1,...,f_k$.

Segundo Resende e Thompson (2004) e Smith, Cullis e Thompson (2001) a adequação dos modelos FAMM de várias ordens k pode ser formalmente testado, uma vez que são ajustados via abordagem de modelos mistos. O modelo com k fatores, denotada FAK, é hierárquico dentro do modelo com k+1 fatores. Modelos, incluindo o efeito principal do genótipo (g) são intermediários entre os modelos de análise de fator de ordem k (FAK) e de ordem FAK +1. Modelo FA1+g é intermédio entre os modelos FA1 e FA2. Testes de razão de máxima verossimilhança restrita (REMLLRT) podem ser utilizados para a comparação de tais modelos. Outras abordagens para testar o ajuste de modelos de fatores analíticos envolvem comparações com a matriz de covariância não estruturada, o qual é muito difícil de obter, com um grande número de ambientes (MARDIA et al., 1988).

2.8 Algoritmos utilizados na estimação de componentes de variância em modelos FAMM utilizando REML

No modelo proposto por Smith, Cullis e Thompson (2001), dado em (28), para calcular as estimativas dos efeitos fixos e aleatórios, exigem-se estimativas dos parâmetros $\mathbf{\Sigma}$ e $\mathbf{R}$. Em termos do modelo fator-analítico, os parâmetros de variância associados $\mathbf{\Sigma}$ são $\mathbf{\Lambda}$ e $\mathbf{\Psi}$. As estimativas dos componentes de variância, são obtidas utilizando o método de REML (PATTERSON; THOMPSON, 1971). Smith, Cullis e Thompson (2001) usaram um algoritmo de escores conhecido como o algoritmo Informação Média (AI) (GILMOUR; THOMPSON; CULLIS, 1995) para a obtenção dos componentes de variância FA. Este é um algoritmo de escores de Fisher modificado, no qual

a matriz de informação esperada é substituída por uma média aproximada das matrizes de informação observadas e esperadas. O software mais utilizado para estimação dos parâmetros de variância, via máxima verossimilhança restrita (REML), desses modelos é o pacote ASReml (GILMOUR et al., 2002).

Em termos de componentes de variância FA, a implementação original no pacote ASReml em l baseou-se no algoritmo proposto por Smith, Cullis e Thompson (2001), que não explora o modelo de regressão em que acomoda o modelo FA. Por conseguinte, as equações do modelo misto são relativamente densas, reduzindo seriamente a velocidade computacional das análises para conjuntos de dados com um grande número de ambientes ou quando se ajusta a variância de modelos fator analíticos com vários fatores (THOMPSON et al., 2003).

O outro problema prático com o modelo FA é a ocorrência frequente dos casos Heywood (SMITH; CULLIS; THOMPSON, 2001). Nestes casos, uma ou mais variância específicas tendem a zero, o que implica que a matriz de variâncias para os efeitos de interação genótipo x ambiente é de posto incompleto (doravante denominado de posto reduzido (variância) do modelo). Esse problema ocorre às vezes em aplicações multivariados e é difícil garantir que as estimativas REML dos parâmetros de variância dos modelos de variância complexos, tais como o modelo de variância desestruturada, permaneçam dentro do espaço paramétrico. No caso desestruturado pode haver uma vantagem na montagem de um modelo de variância que envolve uma matriz que não é de posto completo, por meio da decomposição de Cholesky. Isto é equivalente ao modelo de variância de posto reduzido para os genótipos em cada ambiente (THOMPSON et al., 2003).

Para resolver os problemas encontrados na implementação dos modelos com estrutura FA, Thompson et al. (2003) propuseram o uso do algoritmo AI modificado para a estimativas REML de posto reduzido (RR) ou os componentes de variância FA.

2.8.1 Alternativas de estimação computacional do modelo FA

Teoricamente, um modelo com estrutura de matriz e covariância não estruturada (UN) $\boldsymbol{\Sigma}$ seria o modelo de variância mais completo para encaixar os efeitos de n procedências em cada um dos q ensaios considerados, pois está trata os vários locais como se fossem diferentes caracteres. No entanto, o número de parâmetros a ser estimado na matriz UN é $q(q+1)/2$ e assim o processo de estimação pode se tornar instável com o aumento de q devido a uma super parametrização do modelo (SILVA et al., 2009; SMITH; CULLIS; THOMPSON, 2001; THOMPSON et al., 2003). Esse modelo contempla tanto a heterogeneidade de variâncias quanto a covariância entre locais. No entanto, essa modelagem é a mais complexa possível e, com grande número de ambientes, é impraticável devido à necessidade de estimação de um grande número de parâmetros e a dificuldade de convergência da análise (RESENDE, 2007).

Silva et al. (2009) não verificaram a convergência do modelo quando usaram a matriz de covariância UN completa para análise univariada de cada caractere. No entanto, na busca de modelos parcimoniosos para modelar os efeitos de u_g fizeram uma análise conjunta (multivariada) de todos os caracteres usando um modelo multiplicativo associado com a análise fatorial com uma aproximação à forma UN onde verificaram a convergência do modelo usando o algoritmo AI (de informação-média). Apesar de garantir a convergência usando o procedimento proposto, o algoritmo AI pode conduzir a um modelo FA que não é de posto completo o que impõe a restrição de que os elementos da matriz de variância específicas estejam dentro de espaço paramétrico, que pode levar a problemas de convergência. Os mesmos autores verificaram que o algoritmo proposto por Thompson et al. (2003) solucionou esses problemas, ajustando-se diretamente a estrutura FA sem necessidade de aproximar a forma da matriz UN.

A abordagem do processo de estimação no modelo fator analítico descrita em Smith, Cullis e Thompson (2001) são computacionalmente intensivos. Um algoritmo alternativo que utiliza métodos de matrizes esparsas é

dado em Thompson et al. (2003). Este algoritmo foi proposto para reduzir o tempo de computação. Ele também acomoda casos em que algumas (ou todas) as variâncias específicas precisam ser condicionadas a assumirem o valor zero, conduzindo assim a uma estrutura de variância que não seja de posto completo. Segundo Smith, Cullis e Thompson (2002), as pesquisas em modelos fator-analítico deveriam focar em alternativas para o algoritmo AI, em particular, no EM (DEMPSTER; LAIRD; RUBIN, 1977) e método de esperança-maximização com parâmetros estendidos (PX-EM) (LIU; RUBIN; WU, 1998). Contudo, as mesmas continuam sendo conduzidas usando esse algoritmo.

2.9 Dados faltantes (*missing data*)

Segundo McKnight et al. (2007, p. 2), "de um modo geral, o termo dados faltantes significa que está faltando algum tipo de informação sobre o fenômeno em que estamos interessados". Normalmente, são observações que deveriam ter sido feitas, mas não foram por algum motivo. Quando isso acontece, a capacidade de entender a natureza do fenômeno pode ser reduzida e o impacto nos resultados dos estudos nem sempre são conhecidos, tornando-se difícil extrair um conhecimento útil a partir dos dados analisados (MCKNIGHT et al., 2007; VERONEZE; FRANÇA; ZUBEN, 2011).

Litle e Rubin (2002) distinguem três tipos de padrões de dados faltantes: falta informativa ou faltantes não ao acaso (MNAR- *missing or missing not at random*), faltantes ao acaso (MAR- *missing at random*) e faltantes completamente ao acaso (MCAR- *missing completely at random*).

MCAR- nesta situação, as observações faltantes não são diferentes das não faltantes em termos da análise realizada. Neste caso, os faltantes surgiram de maneira aleatória e, portanto, o único problema gerado pelos dados faltantes é a perda de poder da análise a ser realizada;

MAR- neste caso, os dados faltantes dependem das variáveis preenchidas e, portanto, podem ser totalmente explicadas pelas demais variáveis

presentes no banco de dados. Logo, ao realizar o tratamento dos dados faltantes de forma que sejam consideradas as informações que "causam" os faltantes, é possível realizar uma análise não viesada. Neste, os dados faltantes são causados por alguma variável observada, disponível para análise e correlacionada com a variável que possui dados faltantes (GRAHAM et al., 1995).

MNAR- nesta situação os faltantes são gerados de forma não mensurável, ou seja, eles dependem de eventos que o pesquisador não consegue observar e controlar. Este é o caso mais grave, em que para tratamento dos dados faltantes, em alguns casos, são necessárias técnicas mais complicadas.

Os dados a partir de um indivíduo podem ser subdivididos em dados observados e ausentes. Se um padrão de dados perdidos depende dos dados observados, mas não sobre os dados em falta, o padrão de dados em falta é MAR. Se depender de dados observados e perdidos é informativo. Se for independente, tanto dos dados observados e não observados, é MCAR. MCAR e, com a premissa adicional de seperabilidade, o padrão MAR é ignorada se REML é usado (FISCHER et al., 2009; VERBEKE; MOLENBERGHS, 2000).

Durante a seleção, os genótipos recém-criados são adicionados, enquanto genótipos selecionados são descartados. Portanto, os dados de melhoramento de plantas são quase sempre selecionados e desbalanceados. Isto resulta em dados faltantes o que complica a análise, por exemplo, na estimação da Heredabilidade (FISCHER et al., 2009; PIEPHO; MÖHRING, 2007).

No melhoramento de plantas, o padrão de dados em falta muitas vezes é informativo, devido à falta de informação para decisões de seleção ou falta de informações de pedigree. Os melhoristas costumam usar informações de pedigree durante concepção dos seus experimentos. É comum que os genótipos da mesma linhagem sejam testados no mesmo ensaio, muitas vezes lado a lado. Se os testes de um conjunto de genótipos não foram realizados em cada local, a informação pedigree influencia o padrão de dados faltantes. Piepho e Möhring (2006) mostraram que os dados em falta, devido à seleção, podem ser ignorados, se todos os dados utilizados para a seleção estão disponíveis e são incluídos na análise.

2.10 Técnica da elipse de confiança

Segundo Schofield e Breach (1972), elipse de confiança é uma forma conveniente de expressar graficamente a incerteza posicional de um ponto, e sendo absoluta, fornece a medida de incerteza relativa do ponto analisado em relação ao ponto fixo em estudo.

Esta técnica do gráfico da elipse de confiança é mais utilizada para verificar a compatibilidade entre os laboratórios, e é baseada do método de Youden (CHUI et al., 2004). O planejamento experimental para a construção da elipse de confiança prevê a distribuição de um par de amostras semelhantes, não necessariamente de concentrações iguais, porém de concentrações próximas. A elipse é construída para cada eixo simulado e é representado por um ponto. As retas que passam pelas médias dos escores, em x (resultados relativos a uma das simulações) e em y (resultados relativos a outro escore fatorial), dividem o diagrama em quadrantes. Pontos encontrados nos quadrantes; superior direito e inferior esquerdo representam os escores que podem estar incorrendo em erros sistemáticos. Na prática, quando somente erros aleatórios estão presentes, os pontos devem estar distribuídos de modo uniforme em todos os quadrantes. Se os pontos se encontrarem mais concentrados nos quadrantes superior direito e inferior esquerdo, isto é interpretado como evidência de ocorrência de erros sistemáticos, ou seja, os escores tendem a obter valores altos ou baixos, em ambas as amostras do par.

A elipse de confiança é traçada de tal modo que qualquer ponto tem a mesma probabilidade de estar dentro da elipse e, em geral, é estabelecido o grau de 95% de confiança. Geralmente os pontos se situam dentro de uma elipse, cujo eixo maior faz um ângulo de aproximadamente 45^0 com o eixo da horizontal. Portanto a inclinação maior da elipse está próxima de +1 e a do eixo menor, de -1. A dispersão dos pontos ao longo do eixo maior está associada aos erros sistemáticos, enquanto que ao longo do eixo menor está associada aos erros aleatórios (CHUI et al., 2004).

No caso em que os erros aleatórios podem ser considerados iguais, a elipse estará posicionada no gráfico com seu eixo maior a 45° em relação ao eixo das abcissas. A dispersão em torno do eixo menor da elipse representa apenas os erros aleatórios, enquanto que a dispersão ao longo do eixo maior representa os erros sistemáticos. Quando os erros aleatórios são ambos pequenos, mas não necessariamente iguais em relação aos erros sistemáticos, a elipse de confiança apresentar-se-á orientada com seu eixo maior a aproximadamente 45°, em relação ao eixo das abcissas, porém, com uma forma mais alongada. Se os erros aleatórios das duas amostras forem bem diferentes, e o erro sistemático de uma delas se aproximar do erro aleatório, a elipse de confiança poderá ter seu eixo maior entre 30° e 90°, em relação ao eixo das abcissas. Dependendo dos valores atribuídos aos erros sistemáticos e aos erros aleatórios, o eixo maior pode até apresentar-se na horizontal, ou seja, a 0° com relação ao eixo das abcissas (CHUI et al., 2004).

2.11 Elipses de confiança para predição

Várias são as vantagens estatísticas e biológicas dos modelos AMMI e SREG (Sites Regression Analysis) mistos, como a capacidade de incorporar informações e flexibilidade para lidar com dados desbalanceados, sem a necessidade de imputação dos dados em falta e heterogeneidade de variância na análise de MET. No entanto, eles apresentam uma limitação pois não está claro como regiões de confiança assintóticas paramétricas, construídas para modelos de efeitos fixos (GOWER; DENIS, 1996), podem ser estendidas para modelos de efeitos mistos (CROSSA et al., 2011a). Além da teoria assintótica, regiões de confiança para os parâmetros de interação do modelo AMMI tem sido propostas utilizando procedimentos bootstrap (LAVORANTI, 2003; YANG et al., 2009) e inferência Bayesiana (CROSSA et al., 2011a). Entretanto, na literatura sobre a análise fatorial não encontramos nada formal escrito sobre inferência para escores fatorais. Crossa (2012) reconhece ser difícil propor intervalos de confiança para os escores fatoriais. Neste contexto as elipses de confiança para

predição podem ser úteis para representar as regiões de confiança dos escores fatoriais.

Uma elipse de confiança para predição é uma região de confiança para predizer uma nova observação na população. Também mostra onde uma porcentagem especificada dos dados deverá ficar.

Seja $\bar{y}$ e S a média e a matriz de covariâncias de uma amostra aleatória de tamanho n de uma distribuição normal bivariada com média $\bar{y}$ e ${}_2\Sigma_2$.

Considerando ${}_2y_1$ como uma variável aleatória bivariada para uma nova observação e observando que a variável $(y-\bar{y}) \sim N_2(0,(1+\frac{1}{n})\Sigma)$ é independente de S, tem-se que uma elipse de confiança a $100(1-\alpha)\%$ para predição é dada pela equação:

$$(y-\bar{y})'\Sigma^{-1}(y-\bar{y}) = \frac{2(n-1)(n+1)}{(n-2)n} F_{2,n-2}(1-\alpha) \qquad (35)$$

Segundo Dias (2012), a família de elipses gerada por diferentes valores críticos F tem um centro comum, que é a média amostral, e eixos maior e menor comuns. Graficamente as elipses indicam a correlação entre as variáveis. Quando os eixos das variáveis são padronizados (dividindo as variáveis pelos seus respectivos desvios padrão), a razão dos dois comprimentos dos eixos (em distâncias Euclidianas) reflete a magnitude da correlação entre as duas variáveis.

2.12 Validação cruzada

A validação cruzada é uma técnica para avaliar a capacidade de generalização de um modelo a partir de um conjunto de dados. Esta técnica é amplamente empregada em problemas onde o objetivo da modelagem é a predição. Busca-se então estimar o quão acurado é este modelo na prática, ou seja, o seu desempenho para um novo conjunto de dados.

O conceito central das técnicas de validação cruzada é o particionamento do conjunto de dados em subconjuntos mutuamente exclusivos, e posteriormente, utilizar alguns desses subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento) e outros subconjuntos (dados de validação ou de teste) é empregado na validação do modelo.

Diversas formas de realizar a validação cruzada foram sugeridas, sendo as três mais utilizadas o método: *holdout*, *k-fold* e *leave-one-out.*

Para todos os métodos de particionamento, citados acima e apresentados a seguir, a acurácia final do modelo estimado é obtido por:

$$Ac_f = \frac{1}{v}\sum_{i=1}^{v} \varepsilon_{y_i,\hat{y}_i} = \frac{1}{v}\sum_{i=1}^{v} (y_i - \hat{y}_i) \qquad (32)$$

onde v é o número de dados de validação e $\varepsilon_{y_i,\hat{y}_i}$ é o resíduo dado pela diferença entre o valor real da saída i e o valor predito. Com isso, é possível inferir de forma quantitativa a capacidade de generalização do modelo.

a) Método *k-fold*

O método de validação cruzada denominado *k-fold* consiste em dividir o conjunto total de dados em k subconjuntos mutuamente exclusivos do mesmo tamanho. Destes k subconjuntos, um subconjunto é retido para ser utilizado na validação do modelo e os e os *k-1* restantes são utilizados para estimação dos parâmetros e calcula-se a acurácia do modelo. O processo de validação cruzada é, então, repetido k vezes, de modo que cada um dos k subconjuntos sejam utilizados exatamente uma vez como dado de teste para validação do modelo.

Ao final das k iterações, calcula-se a acurácia sobre os erros encontrados, por meio da equação descrita anteriormente, obtendo assim uma medida mais confiável sobre a capacidade do modelo de representar o processo gerador dos dados.

b) Método leave-one-out

O método *leave-one-out* é um caso específico do *k-fold*, com *k* igual ao número total de dados. Nesta abordagem são calculados N erros, um para cada dado.

Apesar de apresentar uma investigação completa sobre a variação do modelo em relação aos dados utilizados, este método possui um alto custo computacional, sendo indicado para situações onde poucos dados estão disponíveis.

Dias e Krzanowski (2003) propuseram métodos baseados em procedimento *leave-one-out* completo, que otimiza o processo de validação cruzada. Este modelo deseja predizer os elementos x_{ij} da matriz **X** por meio do modelo:

$$x_{ij} = \sum_{k=1}^{n} d_k u_{ik} v_{jk} + \varepsilon_{ij} \qquad (33)$$

Em que d_k é raiz quadrada dos autovalores da matriz $\mathbf{XX'}$, a i-ésima coluna $\mathbf{v_{ik}} = (\mathbf{v_{i1}}, .., \mathbf{v_{ik}})$ da matriz $\mathbf{V_{pxp}}$ e o autovetor correspondente ao i-ésimo maior autovalor d_k^2 de $\mathbf{X'X}$ e a j-ésima coluna $(\mathbf{u_{j1}}, ...\mathbf{u_{nj}})'$ da matriz $\mathbf{U_{nxp}}$ e o autovetor correspondente ao i-ésimo maior autovalor d_k^2 de $\mathbf{XX'}$, ε_{ij} é o ruído.

Os métodos apresentados em Gabriel (2002) e Krzanowski (1987) no qual prediz-se o valor $\hat{x}_{ij}^n$ de $x_{ij} (i = 1, ..., g; j = 1, ..., e)$ para cada possível escolha de n (o número de componentes), e a medida de discrepância entre o valor atual e predito como:

$$PRESS(n) = \sum_{i=1}^{g} \sum_{j=1}^{e} (x_{ij}^n - x_{ij})^2 \qquad (34)$$

3 MATERIAL E MÉTODOS

Nesta seção são apresentados os dados utilizados para realização do estudo, bem como a descrição completa da metodologia utilizada na realização das análises.

3.1 Material

Os dados utilizados foram descritos por Machado et al. (2008). Os experimentos foram conduzidos em nove ambientes, caracterizando assim, experimentos multiambientes, no ano agrícola de 2005/06, em propriedades de agricultores e estações experimentais (Tabela 1). Foram avaliados 55 híbridos de milho (codificados como: G1, G2, G3,...,G55) no delineamento de blocos casualizados com três repetições e a parcela experimental era constituída de duas linhas de 3 m de comprimento, mantendo uma população, após o desbaste, de 55.000 plantas por hectare. O caráter avaliado foi da produtividade de espigas despalhadas (t ha^{-1}), corrigida para 13% de umidade.

Tabela 1 Características dos ambientes de condução dos experimentos

Ambiente	Município	Latitude	Longitude	DMS[1] ($t.ha^{-1}$)	CV[2] (%)	Produtividade média ($t.ha^{-1}$)
Área Experimental/DBI (E1)	Lavras, MG	21°13´S	44°58´W	5,139	14,1	10,803
Área Experimental Geneze (E2)	Guarda-Mor, MG	17°34´S	47°08´W	2,844	13,5	6,212
Área Experimental Bionacional (E3)	Barreiras, BA	12°08´S	45°00´W	1,851	12,0	4,549
Área Experimental Prezzotto (E4)	Jussara, GO	23°35´S	52°28´W	2,195	12,5	5,152
Fazenda Vitorinha (E5)	Lavras, MG	21°12´S	44°58´W	4,423	20,8	6,246

Tabela 1, conclusão

Ambiente	Município	Latitude	Longitude	DMS[1] ($t.ha^{-1}$)	CV[2] (%)	Produtividade média ($t.ha^{-1}$)
Área Experimental Coopadão (E6)	São Gotardo, MG	19°18´S	46°03´W	3,107	11,3	8,085
Fazenda Faepe (E7)	Ijaci, MG	21°09´S	44°56´W	4,551	10,1	13,192
Fazenda Faepe (E8)	Ijaci, MG	21°10´S	44°56´W	3,248	10,7	8,896
Fazenda Mato Dentro (E9)	Lavras, MG	21°10´S	45°03´W	4,309	14,6	8,737

[1] DMS=diferença mínima significativa (Tukey 5%); [2] coeficiente de variação ambiental.
Fonte: Machado et al. (2008)

Estimação e predição

A abordagem (estimação em duas fases) usada foi a mesma adotada por Smith, Cullis e Thompson (2005). Porém, considerando o algoritmo EM no processo de estimação e assumindo uma matriz de variância e covariâncias e residual não estruturada aproximado ao modelo FA via análise de fatores.

A seguir são apresentadas as descrições detalhadas de cada estágio de estimação.

3.2 Métodos

Análise do primeiro estágio

A análise MET foi realizada considerando o seguinte modelo (16):

$$y=Xb+Zu+e,$$

em que: $\mathbf{y}$ é o vetor de observações de parcelas referente cada ambiente, $\mathbf{b}$ e $\mathbf{u}$ são os vetores de efeitos fixos (blocos) e aleatórios (genótipos) respectivamente, $\mathbf{e}$ o vetor aleatório de erros e $\mathbf{X}$ e $\mathbf{Z}$ as matrizes de incidência o para efeitos

fixos e aleatórios. Para esse conjunto de dados assumiu-se que $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$ e $\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Genericamente, essa equação pode ser expandida da seguinte forma:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} X_1 & O & \dots & \dots & \dots & O \\ O & X_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & O \\ O & \dots & \dots & \dots & O & X_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{bmatrix} + \begin{bmatrix} Z_1 & O & \dots & \dots & \dots & O \\ O & Z_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & O \\ O & \dots & \dots & \dots & O & Z_n \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix}$$

Onde cada subscrito corresponde aos subvetores e submatrizes das observações e delineamento em cada ambiente

Dada a matriz de equações de modelos mistos (MEMM) abaixo:

$$C = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + \Sigma^{-1} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

Tomando o logaritmo natural e derivando a equação (38) em relação à $\mathbf{b}$ e $\mathbf{u}$ e considerando $\boldsymbol{\Sigma}$ e $\mathbf{R}$ conhecidas, têm-se as equações dos modelos mistos multivariada (MMM) dadas por:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + \Sigma^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \qquad (39)$$

As soluções de $\hat{b}$ e $\hat{u}$ são dados por:

$$\begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + \Sigma^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \qquad (40)\ .$$

Com algumas manipulações e assumindo $V = ZR^{-1}Z' + \Sigma^{-1}$ os estimadores dos efeitos fixos e aleatórios são dados por:

$$\hat{b} = (X'V^{-1}X)^{-1}\, X'V^{-1}y \qquad (41)$$

$$\hat{u} = Z'\Sigma V^{-1}(y - Xb) \quad . \tag{42}$$

Fazendo: $C = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + \Sigma^{-1} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$

Utilizando o algoritmo EM descrito por Dempster, Laird e Rubin (1977) a solução REML para os elementos das matrizes **Σ** e **R** podem ser dada por:

$$\hat{\sigma}_{u_{ij}} = \left[u_i^T u_j + tr\left(C_{ij}^{-1}\right) \right] \Big/ t \tag{43}$$

com:

$$\tilde{\sigma}_{u_{ij}} = \begin{cases} \sigma_{u_k}^2 \ se\ i = j \\ \sigma_{u_{ij}} \ se\ contrário \end{cases}$$

A matriz C_{ij}^{-1} corresponde a submatrizes *i j* da inversa C^{-1} da matriz das equações de modelos mistos. O estimador da variância residual pode ser dado por:

$$\tilde{\sigma}_{e_{ij}} = \left\{ e_i^T e_j + tr\left(\left[KC^{-1}K^t \right]_{ij} \right) \right\} \Big/ n^* \tag{44}$$

$$\tilde{\sigma}_{eij} = \begin{cases} \sigma_{e_k}^2 \ se\ i = j \\ \sigma_{e_{ij}} \ se\ contrário \end{cases}$$

Onde *K={X,Z}* e o traço dependem da submatriz relacionada à *i* e *j* , sendo n^* o comprimento do vetor {*j,i*}.

Essa abordagem estima todos os parâmetros de dispersão de uma matriz de (co) variâncias não estruturadas (UN), ou seja, todas as variâncias $\{\sigma_{e_k}^2, \sigma_{u_k}^2\}$ e covariâncias $\{\sigma_{e_{ij}}, \sigma_{u_{ij}}\}$ são estimadas simultaneamente com os efeitos fixos (EBLUEs) e aleatórios (EBLUPs). Essa abordagem permite a análise MET com dados faltantes e heterogeneidade de variâncias. No entanto, as estabilidades e adaptabilidades dos genótipos não são obtidas de forma direta. Sendo assim, a

análise de fatores na verossimilhança restrita pode ser realizada em um passo posterior.

Análise de Fatores na Verossilhança restrita (Análise do estágio 2)

As análises estatísticas foram realizadas pelo procedimento REML/BLUP, em que os componentes de variância são estimados pela máxima verossimilhança restrita (REML) via algoritmo EM, sendo os valores genotípicos preditos pela melhor predição linear não viciada.

Dada a matriz de variância genética (**Σ**) e os BLUP dos genótipos do estágio 1 e assumindo **Σ** como uma estrutura de FA ou seja (**LL**`+**Ψ**), e que os BLUPs possam ser representados por fatores comuns na forma (**u = Lf + δ**), um modelo equivalente a (17) é reescrito para acomodar a análise de fatores, denominado de modelo FAMM, o qual é dado por:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}[\mathbf{Lf} + \boldsymbol{\delta}] + \mathbf{e}$$

$$\mathbf{f} \sim N(\mathbf{0}, \mathbf{I}),\ \boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Psi}) \text{ e } \mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$$

$$\text{Com } \mathbf{u} = \mathbf{Lf} + \boldsymbol{\delta},$$

em que o **f** e **δ** representam os vetores dos escores fatoriais (BLUP´s) e variância específica respectivamente, **L** a matriz de cargas fatoriais, **X** e **Z** as matrizes de delineamento.

Assim, a solução matriz das equações de modelos mistos re parametrizadas (**W = ZL**) pode ser dada por:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}W & X'R^{-1}Z \\ W'R^{-1}X & W'R^{-1}W + I & W'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}W & Z'R^{-1}Z + \Psi^{-1} \otimes I \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{f} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ W'R^{-1}y \\ ZR^{-1}\acute{y} \end{bmatrix} \quad (45).$$

Resolvendo (45) em relação $\hat{b}$, $\hat{f}$ e $\hat{\delta}$ tem-se:

$$\begin{bmatrix} \hat{b} \\ \hat{f} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}W & X'R^{-1}Z \\ W'R^{-1}X & W'R^{-1}W + I & W'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}W & Z'R^{-1}Z + \Psi^{-1} \otimes I \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}y \\ W'R^{-1}y \\ ZR^{-1}\acute{y} \end{bmatrix} \quad (46).$$

Assumindo a matriz de covariância genética modelada pela estrutura FA ($\boldsymbol{\Sigma}$=LL´+Ψ) a solução dos efeitos fixos e aleatórios (os estimadores dos escores fatoriais e da variância especificas (BLUP´s de $\mathbf{f}$ e $\boldsymbol{\delta}$)), segundo Meyer (2009) são obtidos de forma equivalente por:

$$\hat{\mathbf{b}} = (\mathbf{X'\Sigma^{-1}X})^{-1}\mathbf{X'\Sigma^{-1}(y - Wf - Z\delta)} \quad (47)$$

$$\hat{\mathbf{f}} = (\mathbf{W'\Sigma^{-1}W + I})^{-1}\mathbf{W'\Sigma^{-1}(y - Xb - Z\delta)} \quad (48)$$

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z'\Sigma^{-1}Z + \Psi^{-1}\ddot{A}I})^{-1}\mathbf{Z'\Sigma^{-1}(y - Xb - Wf)} \quad (49)$$

A vantagem de usar este procedimento está na garantia da convergência do modelo dentro do espaço do parâmetro evitando a ocorrência dos casos Heywood.

Para o ajuste de modelos de análise de fatorial foram testados uma, duas e três cargas fatoriais (assemelhando-se a teste de modelos FA (1), FA (2) e FA (3) respetivamente), a fim de determinar quantos fatores são necessários para explicar a variabilidade genética e predição de genótipos. Para verificar a qualidade de ajuste do número de fatores do modelo FA foi usado o teste de razão de verossimilhança modificado de Bartlett (JOHNSON; WICHERN, 2007).

$$\lambda = n \ln \frac{(|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|)}{|\mathbf{R}|} \sim \chi^2_n \quad (50)$$

Em que:

$\boldsymbol{\Sigma} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$: é a "matriz de variâncias e covariâncias" (positiva definida) do modelo fatorial ajustado;

$\mathbf{R}$: é a matriz de variáveis padronizadas.

Rotação de cargas

Quando o número de cargas é maior que um (k> 1), a matriz de cargas não é única. Esta não singularidade requer medidas corretivas durante a escolha de modelos FA-k, pois o modelo de variância é denominado não identificável. Estas medidas corretivas incluem a restrição de identificabilidade ou racionamento das cargas fatoriais.

Para garantir a unicidade de nas escolhas de **L** utilizou-se a seguinte restrição:

$L^t\Psi^{-1}L = \Delta$, onde $\mathbf{\Delta}$ é uma matriz diagonal. A chave para estimação de **L** e $\mathbf{\Psi}$ é a obtenção da matriz $\Psi^{-1/2}(A-\Psi)\Psi^{-1/2}$. Assim, utilizando o Algoritmo EM obtém-se que: $\hat{L} = \Psi^{1/2}\hat{\mathrm{P}}$, sendo que $\hat{\mathrm{P}}$ provém da decomposição espectral de $\Psi^{-1/2}(A-\Psi)\Psi^{-1/2}$ e $\hat{\Psi} = diag(A-\hat{L}\hat{L}^t)$, onde temos um processo iterativo que continua até a convergência das matrizes **L** e $\mathbf{\Psi}$. Assim, as cargas fatoriais referentes a cada ambiente podem ser descritas por **L**. O próximo passo consiste na obtenção dos escores fatoriais referente aos efeitos de $\hat{\mathbf{u}}$.

Validação do modelo

O desbalanceamento nos dados foi feito de forma aleatória separando a população de treino e de validação. O conjunto de dados original contém 1485 observações, que foram submetidas a perdas aleatórias sob os diferentes níveis de (10%, 30% e 50%). O processo foi repetido mil vezes para cada nível, totalizando três mil tipos diferentes de perdas. No primeiro caso (10%), foram retirados 145 elementos da matriz de dados; no segundo caso (30%), foram retirados 446 elementos e finalmente, no terceiro caso (50%), foram retirados 743 elementos a cada desbalanceamento.

Em cada nível de desbalanceamento considerado o modelo foi aproximado a uma análise de fatores para ter uma medida de estabilidade e adaptabilidade genotípica (análise do segundo estágio).

A capacidade preditiva dos modelos foi medida por meio do cálculo da PRESS (*Residuals The prediction error sum of squares*) e da correlação entre os valores preditos e observados em cada vez e em todas 1000 vezes que se simulou a perda aleatória dos genótipos no nível de desbalanceamento considerado, ou seja, a expressão da PRESS pode ser dada por:

$$PRESS(n) = \frac{1}{m}\sum_{i=1}^{g}\sum_{j=1}^{e}(f_{ij}^{n} - f_{ij})^2 . \qquad (53)$$

$$correlação = \frac{\sum(f_1^n - \overline{f}_1)(f_2^n - \overline{f}_2)}{\sqrt{\text{var}(f_1)}\sqrt{\text{var}(f_2)}}, \qquad (54)$$

em que:

f_{ij}^{n} é o valor predito do escore do genótipo i no ambiente j a cada simulação

f_{ij} é o valor do escore do genótipo i no ambiente j

m é o número de vezes em que o genótipo foi perdido

$\text{var}(f_1)$ é a variância do escore fatorial 1

$\text{var}(f_2)$ é a variância do escore fatorial 2

$\overline{f}_1 e \overline{f}_2$ são as médias dos escores fatorais 1 e 2 respetivamente

Elipses de confiança para predição

Nesse estudo é apresentada uma proposta de validação da estabilidade da análise FA utilizando intervalos de confiança assintóticos para os escores fatoriais por meio de uma elipse de confiança para predição dos genótipos desbalanceados.

Assim é possível comparar a classificação de estável com base na previsibilidade do comportamento dos genótipos sob perdas.

Os resultados preditos para os escores fatorais foram estatisticamente avaliados, usando a técnica da elipse de confiança de predição com 95% de confiança. Nesse sentido, quanto menor a elipse maior é capacidade preditiva desse genótipo, ou seja, menor os desvios da média em torno da média de desempenho. A representação das elipses foi feita usando o pacote CAR do R (R CORE TEAM, 2013).

Todas as análises deste trabalho foram realizadas utilizando o PROC IML do programa SAS 9.3 (STATISTICAL ANALYSIS SYSTEM INSTITUTE - SAS INSTITUTE, 2013) e R (R CORE TEAM, 2013).

4 RESULTADOS E DISCUSSÃO

A apresentação dos resultados e discussão foi feita, de forma separada, apresentando-se os resultados e seguidos da discussão. Primeiramente, o estudo do diagnóstico do modelo proposto com dados balanceados, em seguida são apresentados os resultados do diagnóstico do modelo sob diferentes níveis de desbalanceamento, validação cruzada e por fim apresentar-se-á as regiões de confiança para os escores fatoriais. A discussão foi apresentada para o modelo balanceado, estrutura da matriz das variâncias e covariância, estrutura dos erros, e para a validação cruzada.

4.1 Resultados

Dos modelos FA propostos, o que mais explicou a variação genotípica total (85%) foi o modelo com duas cargas fatoriais (similar a um fator analítico de ordem 2) (FA2). A Tabela 2 apresenta os resultados das cargas fatoriais rotacionados pelo método Varimax. Observa-se que o primeiro fator explicou quase 70% da variação total e o segundo 15%. As variâncias específicas foram baixas para todos ambientes considerados. Os altos valores de variância comum mostram que os dois fatores explicaram grande porcentagem da variância de cada ambiente e que, o modelo FA2 foi o que melhor se ajustou ao conjunto de dados. O modelo conseguiu explicar quase todos ambientes com a exceção dos ambientes E5, E6, E7 e E8. Nessa mesma tabela pode-se notar que o ambiente E1 apresenta uma correlação elevada com o Fator 1 ao passo que o Fator 2 é explicado pelos ambientes E2, E4, E7 e E9.

Tabela 2 Carregamentos estimados (na escala de correlações) para os dados balanceados ajustados a partir do Modelo FA2

Ambiente	Fator 1	Fator2	Variância Comum	Variância específica
E1	0,997	-0,030	0,995	0,000
E2	0,152	0,697	0,982	0,000
E3	0,190	0,423	0,966	0,170

Tabela 2, conclusão

Ambiente	Fator 1	Fator2	Variância Comum	Variância específica
E4	0,570	0,710	0,993	0,035
E5	0,275	0,453	0,806	0,348
E6	0,511	0,611	0,907	0,438
E7	0,253	0,765	0,974	0,402
E8	0,032	0,623	0,997	0,350
E9	0,521	0,800	0,998	0,000
Variância Total	70%	85%		

As matrizes de variância-covariância genotípicas e residuais não estruturadas (UN) são apresentadas na Tabela 3.

Tabela 3 Matriz de Variância-covariância genotípica e residual (em vermelho) aplicado ao conjunto de dados do milho em 9 locais

Amb[1]	1	2	3	4	5	6	7	8	9
1	2,78$^{(2,33)}$	0,45	-0,12	-0,06	0,18	-0,06	-0,20	0,19	-0,06
2	0,17	0,60$^{(0,68)}$	-0,08	-0,07	-0,06	-0,08	0,20	-0,08	0,13
3	0,14	0,29	0,22$^{(0,29)}$	-0,01	0,01	0,01	-0,08	-0,03	0,05
4	0,48	0,22	0,10	0,28$^{(0,41)}$	0,01	0,06	-0,04	0,09	0,04
5	0,29	0,21	0,17	0,20	0,47$^{(1,69)}$	-0,13	-0,13	-0,07	-0,15
6	0,88	0,42	0,25	0,40	0,40	1,17$^{(0,83)}$	-0,05	0,10	-0,12
7	0,45	0,37	0,19	0,47	0,36	1,05	1,48$^{(1,79)}$	0,25	-0,11
8	0,01	0,45	0,10	0,13	0,10	0,19	0,17	0,62$^{(0,94)}$	0,11
9	0,44	0,28	0,08	027	0,17	0,45	0,53	0,30	0,34$^{(1,16)}$

[1]Amb=Ambiente

Nessa tabela observa-se claramente a heterogeneidade de variâncias, tanto genética quanto residuais, justificando assim o relaxamento dessa restrição no modelo. Nota-se também a presença de heterogeneidade das covariâncias genéticas entre os locais (estas covariâncias representam a variância genotípica mais a variância da interação entre pares de locais). A variância residual foi

maior para os ambientes 1, 5 e 7, o que sugere a influência da média desses ambientes na magnitude dessas variâncias (Tabela 1).

Na Figura 4 é apresentado o gráfico dos escores fatoriais do modelo FA2 ajustado aos dados balanceados. Esse biplot apresentou características muito similares ao GGE biplot proposto por Yan et al. (2000), sob médias fenotípicas (Figura 5). A análise GGE captou 64% da variação genotípica total nos dois primeiros componentes; valor este muito abaixo dos 85% explicados pelo modelo FA (2). Comparando as duas metodologias (Figura 4 e 5) nota-se que houve apenas uma ligeira troca no ranqueamento dos genótipos e nos ambientes considerando o primeiro eixo. A adaptabilidade e estabilidade para maior parte dos genótipos sugerida pela análise GGE biplot e FA foram semelhantes para alguns ambientes.

Na Figura 5 é apresentado o gráfico dos escores fatoriais (análogo ao GGE-biplot) do modelo FA2 ajustado aos dados balanceados.

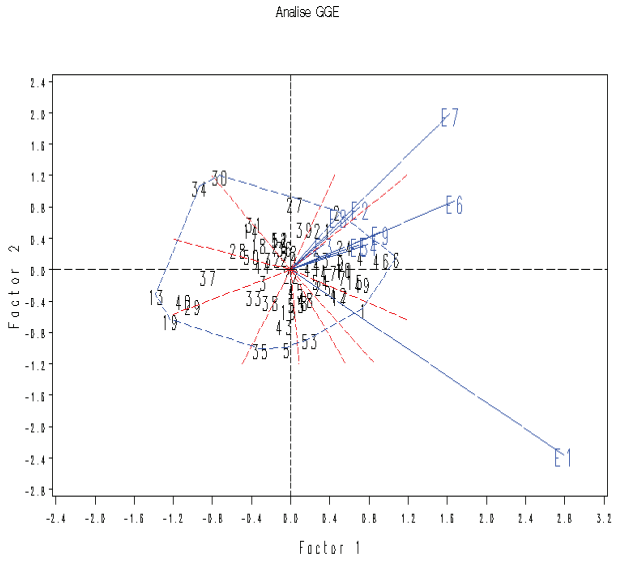


Figura 4 GGE-biplot para dados de produtividade de grãos (t/ha), em milho, com 55 genótipos (G) e nove ambientes (E)

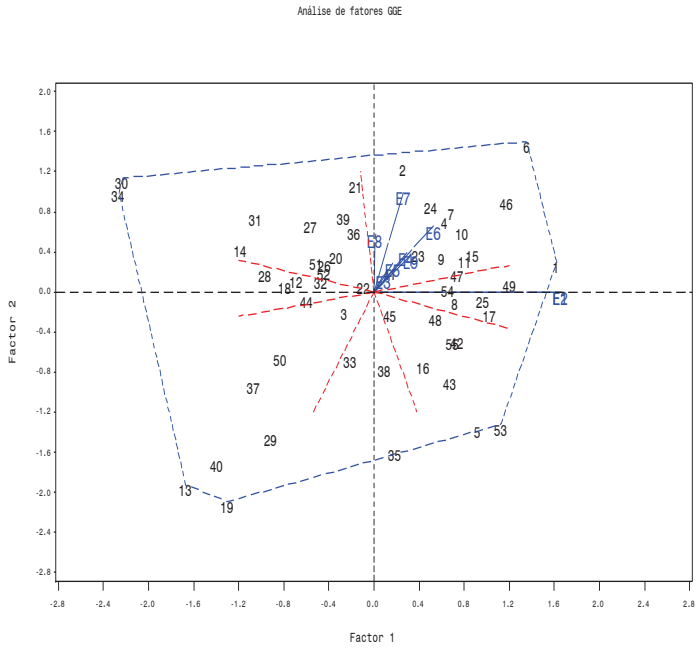


Figura 5 Biplot para dados de produtividade de grãos (t/ha), em milho, com 55 genótipos (G) e nove ambientes (E) considerando o modelo FA

A relação entre os modelos FA2 e AMMI2 foi demonstrada por Smith, Cullis e Thompson (2002) e entre os modelos FA2 e SREG2 por Burgueño et al. (2008). O modelo FA adotado neste trabalho é muito similar ao GGE2 em que o efeito de G está confundido coma interação GxE. A comparação gráfica do biplot resultante destes modelos é muito clara pelas propriedades do modelo FA2 sendo, pois muito similar à do modelo GGE biplot de Crossa et al. (2010), Stefanova e Buirchell (2010) e Yan et al. (2007). No GGE Biplot os genótipos são avaliados quanto à adaptabilidade a partir de estimativas aproximadas dadas pelos escores de Componente Principal 1 (que também relaciona-se com a parte simples da interação G×E) e no modelo FA essa relação pode ser obtida por meio do escore fatorial 1, pois não se observam cargas fatoriais negativas de ambientes e existe uma alta correlação dos escores fatoriais com os BLUPs (0.90). Da mesma forma que a estabilidade de um genótipo na análise GGE pode ser descrita por meio do Componente Principal 2 (que relaciona-se com a Parte

Complexa da interação G×E), assim poder-se-ia interpretar o escore fatorial 2 na análise FA. Assim, genótipos produtivos e estáveis deverão possuir escores elevados para Fator 1, porém valores próximos de zero para Fator 2, ou seja, são genótipos não específicos para grupos de ambientes.

Escores baixos indicam genótipos e/ou ambientes que contribuem pouco ou quase nada para a interação G×E (ou não são explicados pelas cargas de ambientes), sendo, portanto, estáveis. Dessa forma, os genótipos considerados estáveis foram o G12, G18, G22, G25, G32, G44, G45 e G49, G54. Tais genótipos podem ser recomendados amplamente desde que tenham médias elevadas como observada no genótipo G49. Os ambientes que menos contribuíram para interação foram o E3, E5, E6 e E8.

Os genótipos mais afastados da origem são os que mais contribuíram para a interação (ou resposta específica), sendo estes G5, G13, G6, G19, G29, G30, G34, G36, G37, G40, G46, G53.

4.1.1 Diagnósticos do modelo sob diferentes níveis de desbalanceamento e validação cruzada.

Para validação dos modelos foi usada a correlação e a PRESS como parâmetros de avaliação. A correlação serviu para avaliar a acurácia do modelo e a PRESS serviu como base para indicar a distância relativa do genótipo no biplot.

Os resultados de validação cruzada demostraram que é possível predizer o desempenho de híbridos utilizando modelos FA. Na Tabela 4, verifica-se que a correlação foi de moderada magnitude (0,56-0,70) para todos os níveis de desbalanceamento, com magnitude e inversamente proporcional ao nível de perda utilizado. Nessa tabela observa-se que as modas das correlações variaram de 0,64 a 0.9 e a mediana de 0,71 a 0,56 com desvio padrão variando entre 0,13 e 0,21.

Tabela 4 Estatísticas descritivas das correlações entre os valores observados e preditos

Nível de Desbalanceamento	Média	Erro padrão	Mediana	Moda	Desvio padrão	Mínimo	Máximo
10%	0,70	0,01	0,71	0,90	0,21	0,12	0,96
30%	0,60	0,05	0,58	0,79	0,14	0,23	0,91
50%	0,56	0,04	0,56	0,64	0,13	0,03	0,86

Dado que as perdas em parcelas foram aleatórias, a porcentagem de híbridos desbalanceados no conjunto de dados também variou a cada ciclo. Por exemplo, considerando um nível de 10% de perdas de parcelas, o número total de híbridos com dados perdidos variou de 25 a 38, porém não houve uma clara diferença nas correlações desses resultados (Tabela 5).

Quanto ao desbalanceamento a 30%, o número de híbridos com dados faltantes variou de 48 a 55 híbridos e a 50% praticamente todos os híbridos sofreram perdas. Independentemente do nível de desbalanceamento aplicado, a grandeza dos valores de correlação ficou acima de 0,5. Também é importante destacar que em todos os procedimentos os híbridos foram faltantes pelo menos uma vez em cada local.

Tabela 5 Estatística descritiva das correlações entre os valores observados e preditos em relação ao número de híbridos desbalanceados no nível de 10% de perdas nas parcelas.

Número de híbrido desbalanceados	Média	Mínimo	Máximo	Variância
25	0,82	0,75	0,91	0,05
26	0,69	0,53	0,89	0,03
27	0,66	0,53	0,89	0,03
28	0,70	0,34	0,94	0,04
29	0,64	0,23	0,97	0,05

Tabela 5, conclusão

Número de híbrido desbalanceados	Média	Mínimo	Máximo	Variância
30	0,71	0,05	0,95	0,05
31	0,71	0,12	0,97	0,06
32	0,68	0,02	0,96	0,05
33	0,71	0,21	0,96	0,05
34	0,71	0,25	0,97	0,04
35	0,65	0,33	0,96	0,04
36	0,70	0,39	0,97	0,03
37	0,58	0,10	0,88	0,13
38	0,82	0,39	0,76	0,06

Tabela 6 Estatística descritiva das correlações entre os valores observados e preditos em relação ao número de híbridos desbalanceados no nível de 30% de perda nas parcelas

Número de híbrido desbalanceados	Média	Mínimo	Máximo	Variância
48	0,71	0,31	0,86	0,01
49	0,63	0,25	0,78	0,02
50	0,60	0,25	0,89	0,03
51	0,59	0,29	0,90	0,02
52	0,58	0,27	0,90	0,02
53	0,60	0,23	0,90	0,02
54	0,60	0,23	0,90	0,02
55	0,59	0,27	0,91	0,02

Tabela 7 Estatística descritiva das correlações entre os valores observados e preditos em relação ao número de híbridos desbalanceados no nível de 50% de perda nas parcelas

Número de híbridos desbalanceados	Média	Mínimo	Máximo	Variância
53	0,540	0,344	0,661	0,011
54	0,550	0,230	0,830	0,017
55	0,560	0,030	0,851	0,018

Outro parâmetro usado para validação foi o da estatística PRESS e sua variância. A Tabela 8 apresenta os resultados da estatística PRESS para os genótipos considerados estáveis e não estáveis de acordo com a Figura 5.

Tabela 8 Validações cruzada o desbalanceamento sob diferentes níveis de perda de parcelas nos ambientes

Nível de desbalanceamento	Genótipos			
	Estáveis		Não Estáveis	
	PRESS	Var(PRESS)	PRESS	Var(PRESS)
10%	0,19	0,06	0,50	0,56
30%	0,31	0,06	0,77	0,82
50%	0,32	0,07	0,78	0,98

Observa-se na Tabela 8 que a estatística PRESS difere para os dois grupos de genótipos consideráveis estáveis pela análise do biplot. De maneira geral, a PRESS foi de magnitude baixa demonstrando que o modelo consegue predizer os BLUPs dos genótipos desbalanceados. Verificando as variâncias da PRESS, observa-se que os genótipos mais estáveis tiveram maior precisão de escores fatoriais sob desbalanceamento em relação aos não estáveis. Esse

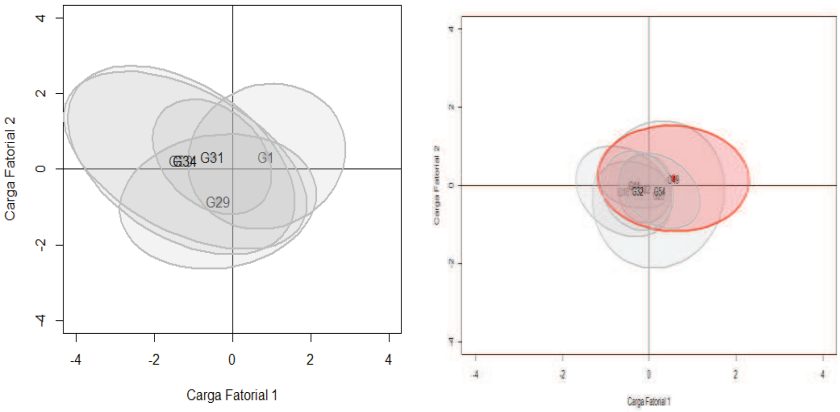
resultado sugere que o segundo escore fatorial pode ser utilizado para descrever genótipos estáveis na análise FA com G+GE.

A PRESS a 10% foi a mais baixa comparada a dos outros níveis de desbalanceamento para os dois grupos de genótipos, embora a dispersão tenha sido praticamente a mesma. A capacidade preditiva do modelo para níveis de perdas de 30% e 50% de parcelas nos ambientes pode ser considerada a mesma nos dois grupos. Contudo, observou-se que nos híbridos não estáveis a dispersão a 50% foi maior. Porém, de forma geral, esses resultados Indicam que mesmo em níveis de desbalanceamento entre 30% e 50% de perda de parcelas nos ambientes, o modelo FA continua sendo eficiente em predizer os valores genotípicos de híbridos perdidos.

4.2 Regiões de confiança para a predição dos escores

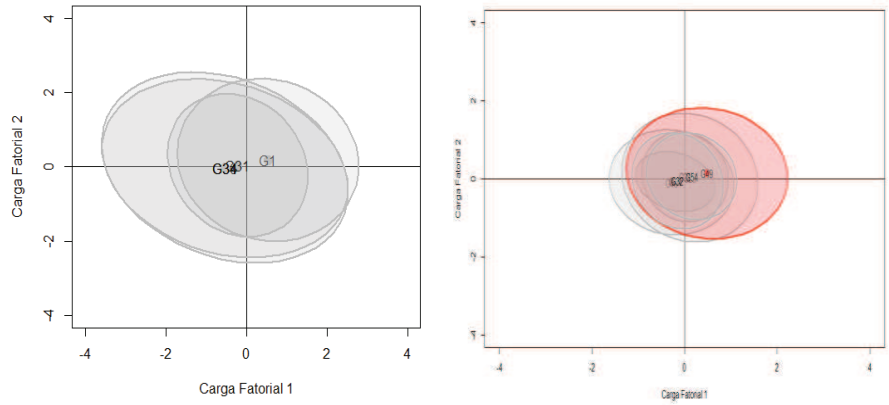
A representação gráfica (Figura 5) permite identificar os genótipos que apresentam adaptabilidade e estabilidade aos diversos ambientes, ou seja, genótipos estáveis e não estáveis que contribuem para a interação. Contudo, a representação gráfica é apenas descritiva, já que não se considerou nenhuma incerteza com relação aos escores ou cargas fatoriais. Com base nessa limitação, este trabalho propôs o uso de regiões empíricas geradas pela predição dos escores fatoriais como parâmetro para medir e validar a estabilidade de um dado genótipo. O objetivo foi incorporar a incerteza e facilitar a interpretação dos resultados de predição dos escores dos genótipos faltantes em uma análise de dados MET.

Nas Figuras 6, 7 e 8 são apresentadas as elipses de confiança para a predição dos escores genotípicos, com níveis de 10%, 30% e 50% de desbalanceamento, respectivamente.



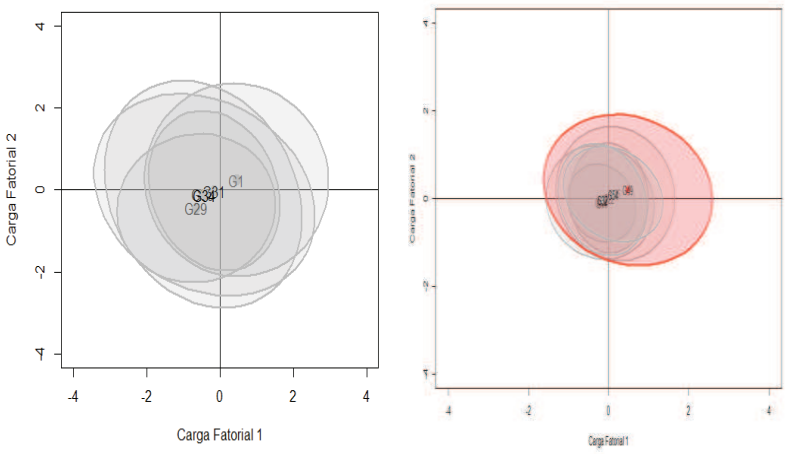
(a) (b)

Figura 6 Regiões de confiança 95% para predição dos escores fatorais (BLUP) dos genótipos não estáveis (a) e estáveis (b), para dados de produção considerando desbalanceamento de 10%



(a) (b)

Figura 7 Regiões de confiança 95% para predição dos escores fatorais (BLUP) dos genótipos não estáveis (a) e estáveis (b), para dados de produção considerando desbalanceamento de 30%



(a) (b)

Figura 8 Regiões de confiança 95% para predição dos escores fatorais (BLUP) dos genótipos não estáveis (a) e estáveis (b) e para dados de produção considerando desbalanceamento de 50%

Verifica-se por meio dessas Figuras que os genótipos estáveis (a) apresentaram elipses mais concentradas em torno do ponto médio, ao passo que os genótipos não estáveis (b) apresentaram regiões de confiança mais dispersas.

A exceção à regra foi observada no genótipo G49 (destacado a vermelho) que no biplot foi considerado como estável e produtivo, mas apresentou comportamento pouco previsível na validação cruzada, ou seja, se a contribuição desse genótipo para interação for nula, sua predição seria pouco dispersa, na segunda carga fatorial. Porém, de modo geral, a identificação de genótipos estáveis no biplot de escores fatoriais corrobora com sua capacidade preditiva via validação cruzada sugerindo que essa técnica pode ser utilizada como medida de incerteza na estabilidade captada no biplot.

4.2 Discussão

A dinâmica de um programa de melhoramento de milho exige do melhorista certa flexibilidade em lidar com a introdução e o descarte de materiais. Nesse foco, Crossa et al. (2006), Piepho (1998), Smith, Cullis e Gilmour (2001) e Smith, Cullis e Thompson (2001, 2005) trataram de buscar uma aproximação entre a análise multicaracterística e as análises multiambientes considerando cada ambiente como uma variável. Essa aproximação permitiu lidar com dados desbalanceados e correlacionados; além da heterogeneidade das variâncias.

Nossos resultados mostraram que a aplicação de modelos mistos em dados de múltiplos ambientes pode ir muito além do estudo da interação genótipos por ambientes. Na verdade, esse tipo de modelagem se encaixa dentro da dinâmica real de um programa, onde os dados são, por natureza, desbalanceados.

No que concerne à capacidade preditiva e ajuste, a análise FA2 foi satisfatória sob desbalanceamento simulado, e a estrutura UN foi minimamente afetada nos seus componentes mesmo com 50% de desbalanceamento. Embora nesse trabalho foram tratados o modelos FA não como concorrente de um modelo UN, mas como uma técnica complementar para análise de biplot, (ou como análise de fatores clássicas na verossimilhança restrita) alguns resultados têm sugerido que pouca diferença preditiva é observada quando diferentes estruturas são modeladas (BURGUEÑO et al., 2007, 2008; CROSSA et al., 2011b; KELLY et al., 2007). Segundo esses autores, modelos com cargas fatoriais maior 2 podem (ou não) aumentar a capacidade de predição, mas certamente iram aumentar a complexidade do modelo e terão um maior critério de informação. Os resultados obtidos por Burgueño et al. (2007, 2008) mostraram que os modelos FA com mais de dois fatores melhoram estimativas de variâncias e covariâncias, mas isso não se refletiu nos EBLUPs de genótipos (BLUPs). Crossa et al. (2006) mostraram que o uso de uma estrutura de G x E fator analítica utilizando de dois a nove fatores apenas influenciaram

ligeiramente os BLUPs, porém não observaram alteração na classificação dos genótipos.

4.2.1 Estrutura da matriz de variâncias e covariâncias

Neste trabalho, foi usada a forma clássica de análise de fatores como alternativa a estimação dos modelos fator analíticos, onde o ajuste do modelo ocorreu em duas fases, sendo a primeira via modelo misto multivariado com matriz de variâncias e covariâncias não estruturada e a segunda fase pela a aproximação do modelo FA com objetivo de visualizar graficamente a relação dos genótipos com os ambientes; além de realizar o estudo da adaptabilidade e estabilidade.

Porém, de maneira muito difundida na literatura, sugere-se a modelagem direta da matriz UN via FA, o que nem sempre é a melhor saída para um conjunto específico de dados e, muitas vezes, transforma-se apenas em uma solução computacional, cujos valores em FA não têm significado direto em termos de variância genética, pois capta apenas parte da sua magnitude ou, quando não, refletem medidas fora do espaço paramétrico (vide aplicação de matrizes esparsas em Thompson et al. (2003).

A vantagem do ajuste do modelo em dois estágios via EM é justificada pela garantia de convergência no espaço do parâmetro evitando casos Heywood, ganho na convergência e na redução da demanda computacional para um conjunto de dados mais simples, e não requer a seleção de modelos ocorrendo diretamente com estimação de parâmetros como proposto por Kelly et al. (2007), Piepho (1998) e Resende (2007). Assim, em casos onde o número de genótipos é elevado e se tem boas estimativas dos componentes de (co) variâncias a estrutura UN pode apresentar grande aumento na predição (BALESTRE et al., 2012).

A desvantagem de estruturas de variâncias G x E mais complexas é a exigência da estimação de um número maior de parâmetros de dispersão ou covariância. Para um número maior de ambientes e poucos genótipos, o aumento

do número de estimativas de componentes de variância necessários pode resultar em problemas de convergência, perda de eficiência e aumento da demanda computacional (MÖHRING; MELCHINGER; PIEPHO, 2010; WELHAM et al., 2010). Além disso, a má estimativa dos componentes de (co) variâncias pode, com efeito, piorar enormemente a predição em estruturas UN em relação a modelos diagonal (BALESTRE et al., 2012).

Outros ganhos práticos no uso do modelo adotado neste trabalho estão no melhor entendimento da interação G × E, por meio do uso dos vetores ambientais para sintetizar o padrão G x E graficamente, semelhante à metodologia Biplot de Kempton (1984).

4.2.2 Estruturas de erro

No que concerne à análise tradicional MET onde GE é inserida diretamente em um modelo linear misto, frequentemente se assume que a variância residual é comum para todas as observações, porém, como observado nesse trabalho e em diversos outros é importante à inclusão dos efeitos dos resíduos heterogêneos (KELLY et al., 2007; RESENDE, 2007; RONNEGARD et al., 2010; SMITH; CULLIS; THOMPSON, 2001).

Nossos resultados demonstram que houve não apenas heterogeneidade de variâncias, mas também covariância residual, que de certa forma, pode auxiliar o melhorista no agrupamento de ambientes dado que esse valor remete ao ranqueamento das interações dos genótipos nas parcelas, sendo que ambientes com alta correlação residual podem ser muito mais similares em estrutura do que aqueles que apresentam apenas o ranqueamento de genótipos similar.

Segundo Kelly et al. (2007) uma prática comum adotada pelos melhoristas de plantas é a avaliação independente dos resultados de cada ensaio. Isto é equivalente a predição do modelo de variância diagonal, o que permite a heterogeneidade de variância, mas nenhuma correlação no desempenho dos genótipos em todos os ensaios, tampouco a covariância de resíduos. O modelo

adoptado neste trabalho associado à FA é conceitualmente superior a essa abordagem, uma vez que capta, além da heterogeneidade de variância residual e suas covariâncias, uma estrutura de covariância mais complexa relacionada ao efeito genético, resultando em maior precisão da predição, tanto para estudos individuais quanto por meio de ensaios em rede.

4.2.3 Diagnósticos do modelo sob diferentes níveis de desbalanceamento e validação cruzada.

Os valores das correlações obtidos nesse trabalho foram de moderados a altos, dependendo da medida de posição adotada. A diferença nessas medidas revela o caráter assimétrico das correlações, sendo que os valores de maior densidade (moda) na distribuição empírica variaram de 0.9 a 0.64. Esses resultados mostram a robustez da predição de interação G x E quando esta é modelada por meio de modelos mistos em estruturas de heterogeneidade de variâncias. Por exemplo, a diminuição de correlações entre valores preditos e observados foi de 4% para níveis de 30% e 50%. Esse resultado sugere que aumentando a porcentagem de híbridos perdidos nos locais em relação à perda de parcelas não interfere substancialmente na predição. Embora esses resultados sejam animadores, deve-se salientar que os ambientes, genótipos e níveis de desbalanceamento considerados neste estudo são restritivos em comparação com análises.

A incerteza sobre o ranqueamento de genótipos em biplots quanto à estabilidade é motivo de controversa na literatura (CROSSA et al., 2010; YAN et al., 2009; YANG et al., 2009). Várias ferramentas de análise têm sido adotadas, tais como bootstrap, análise bayesiana e construção de intervalos assintóticos (CROSSA et al., 2011a; GOWER; DENIS, 1996; LAVORANTI, 2003). Nesse trabalho, sugeriu-se aplicar a capacidade preditiva do modelo FA via validação cruzada como forma de validar a posição de um genótipo no biplot. Essa abordagem não destrói a estrutura GE como sugerida por Yang et al. (2009) e permite a predição de dados faltantes. De maneira geral, os resultados

obtidos via biplot foram concordantes com os obtidos via validação cruzada, demonstrado que genótipos estáveis possuem comportamento previsível e que essa previsibilidade pôde ser verificada pela menor soma de quadrados do erro de predição (PRESS) dos genótipos estáveis em relação aos considerados não estáveis e também pela menor amplitude de elipses de predição. . Assim, a análise de fatores em modelos mistos permite não apenas lidar com dados desbalanceados e heterogeneidade de variâncias, mas também permite uma clara interpretação gráfica do Biplot semelhante ao SREG2 e fornece ao melhorista uma forma de validação da estabilidade via validação cruzada uma vez que a análise tolera altos níveis de desbalanceamento.

5 CONCLUSÃO

Assim, nossos resultados permitem inferir que a PRESS poderia ser utilizada como alternativa para avaliar a o desempenho de genótipos considerados estáveis no biplot. Esse resultado se confirmou pela amplitude das elipses de predição que foram menores nesses genótipos. Verificou-se que a análise de fatores usando modelo misto é robusta sob diferentes níveis de desbalanceamento dos dados, com valores de correlação variando de médio a alto, dependendo do nível de perda estabelecido. Assim, não há dúvidas quanto ao potencial desse tipo de análise para avaliação da estabilidade no melhoramento de plantas.

REFERÊNCIAS

ARAÚJO, L. B.; DIAS, C. T. S. Métodos de correção de autovalores e regressão isotônica nos modelos AMMI. **Revista de Matemática e Estatística**, Jaboticabal, v. 24, n. 2, p. 71-89, 2006.

BALESTRE, M. et al. Bayesian mapping of multiple traits in maize: the importance of pleiotropic effects in studying the inheritance of quantitative traits. **Theoretical and Applied Genetics**, Berlin, v. 6, n. 3, p. 1-15, Aug. 2012.

BURGUEÑO, J. et al. Modeling genotype x environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. **Crop Science**, Madison, v. 46, n. 4, p. 1722-1733, 2007.

BURGUEÑO, J. et al. Using factor analytic models for joining environments and genotypes without crossover genotype × environment interaction. **Crop Science**, Madison, v. 48, p. 1291-1305, July/Aug. 2008.

CAMPBELL, B. T.; JONES, M. A. Assessment of genotype x environment interactions for yield and fiber quality in cotton performance trials. **Euphytica**, Wageningen, v. 144, n. 1, p. 69-78, 2005.

CARGNELUTTI FILHO, A. et al. Associação entre métodos de adaptabilidade e estabilidade em milho. **Ciência Rural**, Santa Maria, v. 39, n. 2, p. 340-347, 2009.

CHUI, Q. et al. O papel dos programas interlaboratoriais para a qualidade dos resultados analíticos. **Química Nova**, São Paulo, v. 27, n. 6, p. 993-1003, 2004.

CROSSA, J. From genotype × environment interaction to gene × environment interaction. **Current Genomics**, London, v. 13, n. 3, p. 225-244, 2012.

CROSSA, J. Statistical analyses of multilocation trials. **Advances in Agronomy**, San Diego, v. 44, n. 1, p. 55-85, Jan. 1990.

CROSSA, J. et al. Bayesian estimation of the additive main effects and multiplicative interaction model. **Crop Science**, Madison, v. 51, p. 1458-1469, 2011a.

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. et al. Prediction assessment of linear mixed models for multienvironment trials. **Crop Science**, Madison, v. 51, p. 944-954, 2011b.

CROSSA, J. et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, Austin, v. 186, p. 1-12, Oct. 2010.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa, MG: UFV, 2004. 480 p.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood estimation from incomplete data via the EM algorithm: with discussion. **Journal of the Royal Statistical Society Series B**, London, v. 39, p. 1-38, 1977.

DIAS, C. T. S. **Análise multivariada**. Piracicaba: ESALQ, 2012. 14 p. Disponível em: <http://www.esalq.usp.br/departamentos/lce/tadeu/aula7.pdf>. Acesso em: 20 dez. 2013.

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". Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60 p. (Série Monografias, 9).

DUARTE, J. B.; ZIMMERMANN, M. J. O. Correlation among yield stability parameters in common bean. **Crop Science**, Madison, v. 35, n. 3, p. 905-912, 1995.

EEUWIJK, F. A. van. Linear and bilinear modeis for the analysis of multi-environment trials: an inventory ofmodels. **Euphytica**, Wageningen, v. 84, n. 1, p. 1-7, 1995.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4th ed. Edinburgh: Longman, 1996. 464 p.

FISCHER, S. et al. Impact of genetic divergence on the ratio of variance due to specific vs. general combining ability in winter triticale. **Crop Science**, Madison, v. 49, n. 6, p. 2119-2122, Nov. 2009.

FOX, P. N.; CROSSA, J.; ROMAGOSA, I. Multi-environment testing and genoty pe environment. In: KEMPTON, R. A.; FOX, P. N. (Ed.). **Statistical methods for plant variety evaluation**. New York: Chapman & Hall, 1997. p. 117-138.

FREEMAN, G. H.; PERKINS, J. M. Environmental and genotype-environmental components of variability: VIII., relationships between genotypes grown in different environments and measures of these environments. **Heredity**, Cary, v. 27, p. 15-23, 1971.

GABRIEL, K. R. Le biplot-outil d'exploration de données multidimensionelles. **Journal de la Société Française de Statistique**, Paris, v. 143, n. 3/4, p. 5-55, 2002.

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, 2006. p. 85-86.

GILMOUR, A. R. et al. **ASReml user guide**. Release 1.0. Hemel Hempstead: VSN International, 2002. 372 p.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B. R. Al, an efficient algorithm for REML estimation in lin-ear mixed models. **Biometrics**, Washington, v. 51, p. 1440-1450, 1995.

GOWER, J. C.; DENIS, J. B. Asymptotic confidence regions for biadditive models: interpreting genotype-environment interactions. **Journal of the Royal Statistical Society**, London, v. 45, n. 4, p. 479-493, 1996.

GRAHAM, J. W. et al. **Analysis with missing data in prevention research:** the science of prevention: methodological advances from alcohol and substance abuse research. Washington: Elsevier, 1995. 366 p.

HAIR JÚNIOR, J. F. et al. **Análise multivariada de dados**. Porto Alegre: Bookman, 2005. 539 p.

HENDERSON, C. R. **Applications of linear models in animal breeding**. Guelph: University Guelph, 1984. 439 p.

HENDERSON, C. R.; QUAAS, R. L. Multiple trait evaluation using relatives' records. **Journal of Animal Science**, Champaign, v. 43, p. 1188-1197, 1976.

JENNRICH, R. L.; SCHLUCHTER, M. D. Unbalanced repeated-measures models with structured covariance matrices. **Biometrics**, Washington, v. 42, p. 805-820, 1986.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. Englewood Cliffs: Prentice-Hall, 2007. 767 p.

KELLY, A. M. et al. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. **Crop Science**, Madison, v. 47, n. 3, p. 1063-1070, 2007.

KEMPTON, R. A. The use of biplots in interpreting variety by environment interactions. **Journal of Agricultural Science**, Cambridge, v. 103, p. 123-135, 1984.

KRZANOWSKI, W. J. Cross-validation in principal component analysis. **Biometrics**, Washington, v. 43, p. 575-584, 1987.

LAVORANTI, O. J. **Estabilidade e Adaptabilidade fenotípica através da reamostragem "bootstrap" no modelo AMMI**. 2003. 166 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2003.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical analysis with missing data**. 2nd ed. Hoboken: J. Wiley, 2002. 408 p.

LIU, C. H.; RUBIN, D. B.; WU, Y. N. Parameter expansion to accelerate EM-the PX-EM algorithm. **Biometrika**, London, v. 85, n. 2, p. 755-770, June 1998.

MACHADO, J. C. et al. Estabilidade de produção de híbridos simples e duplos de milho oriundos de um mesmo conjunto gênico. **Bragantia**, Campinas, v. 67, n. 3, p. 627-631, 2008.

MAIA, M. C. C. et al. Seleção simultânea para produção, adaptabilidade e estabilidade genotípicas em clones de cajueiro, via modelos mistos. **Pesquisa Agropecuária Tropical**, Goiânia, v. 39, n. 1, p. 43-50, 2009.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. London: Academic, 1988. 512 p.

MCKNIGHT, P. E. et al. **Missing data:** a gentle introduction. New York: The Guilford, 2007. 251 p.

MEYER, K. Factor-analytic models for genotype × environment type problems and structured covariance matrices. **Genetics Selection and Evolution**, Paris, v. 41, n. 21, 2009. Disponível em: <http://www,gsejournal,org/content/41/1/21>. Acesso em: 10 dez. 2013.

MÖHRING, J.; MELCHINGER, A. E.; PIEPHO, H. P. REML-based diallel analysis in plant breeding. **Crop Science**, Madison, v. 51, p. 470-478, Mar. 2010.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. Wallingford: CAB International, 1996. 208 p.

MRODE, R. A.; THOMPSON, R. **Linear modelsfor the prediction of animal breeding values**. 2nd ed. Edinburgh: CABI, 2005. 360 p.

PATTERSON, H. D. et al. Variability of yields of cereal varieties in U,K, trials. **Journal of Agricultural Science**, Cambridge, v. 89, n. 1, p. 239-245, 1977.

PATTERSON, H. D.; NABUGOOMU, F. REML and the analysis of series of crop variety trials. In: INTERNATIONAL BIOMETRIC CONFERENCE, 16., 1992, Hamilton. **Proceedings…** Hamilton: IBC, 1992. p. 77-93.

PATTERSON, H. D.; THOMPSON, R. Recovery of interblock information when block sizes are unequal. **Biometrika**, London, v. 31, p. 100-109, 1971.

PIEPHO, H. P. Analyzing genotype-environment data by mixed models with multiplicative terms. **Biometrics**, Washington, v. 53, n. 2, p. 761-767, June 1997.

PIEPHO, H. P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, Berlin, v. 97, n. 1/2, p. 195-201, July 1998.

PIEPHO, H. P. Robustness of statistical test for multiplicative terms in additive main effects and multiplicative interaction model for cultivar trial. **Theoretical and Applied Genetics**, New York, v. 90, n. 3/4, p. 438-443, Mar. 1995.

PIEPHO, H. P. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, Wageningen, v. 161, n. 1/2, p. 209-228, May 2008.

PIEPHO, H. P.; MÖHRING, J. Generation means analysis using mixed models. **Crop Science**, Madison, v. 50, n. 5, p. 1674-1680, 2010.

PIEPHO, H. P.; MÖHRING, J. Selection in cultivar trials: is it ignorable? **Crop Science**, Madison, v. 46, n. 1, p. 192-201, Jan. 2006.

PINTO JÚNIOR, J. E. **REML/BLUP para a análise de múltiplos experimentos, no melhoramento genético de Eucalyptus grandis W. Hill ex Maiden**. 2004. 113 f. Tese (Doutorado em Agronomia) - Universidade Federal do Paraná, Curitiba, 2004.

R CORE TEAM. **R:** a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <http://www.R-project.org/>. Acesso em: 10 nov. 2013.

RAMALHO, M. A. P. et al. **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. Lavras: UFLA, 2012. 522 p.

RESENDE, M. D. V. de. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: EMBRAPA Informação Tecnológica; Colombo: EMBRAPA Florestas, 2002. 975 p.

RESENDE, M. D. V. de. **Matemática e estatística na análise de experimentos e no melhoramento de plantas**. Brasília: EMBRAPA Informação Tecnológica; Colombo: EMBRAPA Florestas, 2007. 559 p.

RESENDE, M. D. V. de. **Métodos estatísticos ótimos na análise de experimentos de campo**. Colombo: EMBRAPA Florestas, 2004. 65 p. (Documentos, 100).

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

RONNEGARD, L. et al. Genetic heterogeneity of residual variance: estimation of variance components using double hierarchical generalized linear models. **Genetics Selection and Evolution**, Paris, v. 42, n. 1, p. 8-18, 2010.

SANTOS, V. B. **Avaliação de linhagens genotípicas do arroz em terras altas via modelos mistos**. 2009. 153 p. Tese (Doutorado em Fitotecnia) - Universidade Federal de Lavras, Lavras, 2009.

SCHOFIELD, W.; BREACH, M. **Engineering surveying**. 6th ed. Burlington: Elsevier, 1972. 615 p.

SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance components**. New York: J. Wiley, 1992. 528 p.

SILVA, J. C. e; DUTKOWSKI, G. Genotype by environment interaction for growth of Eucalyptus globulus in Australia. **Tree Genetics & Genomes**, Heidelberg, v. 2, n. 2, p. 61-75, Apr. 2006.

SILVA, J. C. e et al. Genetic parameters for growth, wood density and pulp yield in Eucalyptus globulus. **Tree Genetics & Genomes**, Heidelberg, v. 5, n. 2, p. 291-305, Apr. 2009.

SIMEÃO, R. M. et al. Avaliação genética em erva-mate pelo procedimento BLUP individual multivariado sob interação genótipo x ambiente. **Pesquisa Agropecuária Brasileira**, Brasília, v. 37, n. 11, p. 1589-1596, nov. 2002.

SMITH, A. B.; CULLIS, B. R.; GILMOUR, A. The analysis of crop variety evaluation data in Australia. **Australian and New Zealand Journal of Statistics**, Melbourne, v. 43, n. 2, p. 129-145, June 2001.

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **Journal of Agricultural Science**, Cambridge, v. 143, n. 2, p. 449-462, June 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**, Washington, v. 57, n. 6, p. 1138-1147, Dec. 2001.

SMITH, A. B.; CULLIS, E. R.; THOMPSON, R. Exploring variety-environment data using random effect AMMI models with adjustment for spatial field trends: 1., theory. In: KANG, M. S. (Ed.). **Quantitative genetics, genomics and plant breeding**. Cary: Oxford University, 2002. p. 323-336.

STATISTICAL ANALYSIS SYSTEM INSTITUTE. **SAS**. Version 9.3. Cary, 2013. Software.

STEFANOVA, K.; BUIRCHELL, B. Multiplicative mixed models for genetic gain assessment in lupin breeding. **Crop Science**, Madison, v. 50, p. 880-891, May/June 2010.

THOMPSON, R. et al. A sparse implementation of the Average Information algorithm for factor analytic and reduced rank variance models. **Australian and New Zealand Journal of Statistics**, Melbourne, v. 45, n. 4, p. 445-459, 2003.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 496 p.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. Berlin: Springer, 2000. 569 p.

VERONEZE, R.; FRANÇA, F. O.; ZUBEN, F. J. von. Assessing the performance of a swarm-based biclustering technique for data imputation. In: IEEE CONGRESS ON EVOLUTIONARY COMPUTATION, 35., 2011, New Orleans. **Proceedings…** New Orleans: IEEE, 2011. p. 386-393.

VLECK, L. D. V.; POLLAK, E. J.; OLTENACU, E. A. B. **Genetics for the animal sciences**. New York: W. H. Freeman, 1987. 391 p.

WELHAM, S. J. et al. A comparison of analysis methods for late-stage variety evaluation trials. **Australian & New Zealand Journal of Statistics**, Oxford, v. 52, n. 2, p. 125-149, 2010.

WHITE, T. L.; HODGE, G. R. **Predicting breeding values with application in forest tree improvement**. Dordrecht: Kluwer Academic, 1989. 369 p.

YAN, W. Comment on "Biplot Analysis of Genotype × Environment Interaction: Proceed with Caution," by R.-C. Yang, J. Crossa, P.L. Cornelius, and J. Burgueño in Crop Science. **Crop Science**, Madison, v. 50, p. 1121-1123, 2010.

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

YAN, W. et al. GGE biplot vs. AMMI analysis of genotype-by-environment data. **Crop Science**, Madison, v. 47, p. 643-655, 2007.

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

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