



**ANDRÉ GRADOWSKI DE FIGUEIREDO**

**APLICAÇÃO DE MODELOS MISTOS NA  
AVALIAÇÃO DE ESTABILIDADE E  
ADAPTABILIDADE EM MILHO UTILIZANDO  
DADOS DESBALANCEADOS**

**LAVRAS - MG**

**2014**

**ANDRÉ GRADOWSKI DE FIGUEIREDO**

**APLICAÇÃO DE MODELOS MISTOS NA AVALIAÇÃO DE  
ESTABILIDADE E ADAPTABILIDADE EM MILHO UTILIZANDO  
DADOS DESBALANCEADOS**

Trabalho de conclusão de curso apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação Profissional em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

Orientador

Dr. Renzo Garcia Von Pinho

Coorientadores

Dr. Marcio Balestre

Dr. Heyder Dinis Silva

**LAVRAS - MG**

2014

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

Figueiredo, André Gradowski de.

Aplicação de modelos mistos na avaliação de estabilidade e adaptabilidade em milho utilizando dados desbalanceados / André Gradowski de Figueiredo. – Lavras : UFLA, 2014.

67 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2014.

Orientador: Renzo Garcia Von Pinho.

Bibliografia.

1. SREG. 2. Elipse de confiança. 3. Matrizes (Matemática). I. Universidade Federal de Lavras. II. Título.

CDD – 633.1523

**ANDRÉ GRADOWSKI DE FIGUEIREDO**

**APLICAÇÃO DE MODELOS MISTOS NA AVALIAÇÃO DE  
ESTABILIDADE E ADAPTABILIDADE EM MILHO UTILIZANDO  
DADOS DESBALANCEADOS**

Trabalho de conclusão de curso apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação Profissional em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

APROVADA em 24 de março de 2014.

Dr. Márcio Balestre UFLA

Dr. Heyder Diniz Silva UFLA

Dr. Guilherme Ferraudo UFLA

Dr. Renzo Garcia Von Pinho  
Orientador

**LAVRAS - MG**

**2014**

Meu pai Andézio Gonçalves de Figueiredo “*in memoriam*”;

Minha esposa Rosana Noronha Gradowski;

**DEDICO**

## AGRADECIMENTOS

A Deus, acima de tudo pela força, benção, oportunidades e auxílio em todos os momentos da minha vida.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Biologia (DBI) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES pela oportunidade do Mestrado Profissional.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio concedido ao Mestrado Profissional em Genética e Melhoramento de Plantas.

À Monsanto do Brasil LTDA pelo incentivo e suporte para a realização deste mestrado, em especial para Dra Jennifer Ralston, MSc. Fernando Koshima, Malu Maes e Dra. Vivian Lavagnoli.

Aos meus orientadores, Professor Dr. Renzo Garcia Von Pinho, Professor Dr. Márcio Balestre e Dr. Heyder Diniz Silva pela orientação, ajuda e dedicação.

Ao Professor Dr. Magno Antônio Patto Ramalho pelos ensinamentos compartilhados, pelo exemplo de dedicação aos seus alunos, de humildade e de estudo.

Aos membros da banca, pela disposição e contribuição.

Aos todos os professores da pós-graduação do curso de genética e melhoramento pelos ensinamentos.

Aos todos os monitores, em especial a Msc. Bruna Line Carvalho o MSc. Paulo Prado, MSc. Samuel Fernandes, MSc. Kaio Olimpio.

Aos meus familiares pelo apoio e incentivo, em especial minha mãe Diva Gradowski, minha esposa Rosana Noronha de Figueiredo, meus filhos Tomas e Rafaela e minha querida amiga Cremilda dos Anjos.

Aos colegas do programa de pós-graduação por tornarem mais agradáveis os momentos de estudo.

A todos os amigos do programa de pós graduação, especialmente Fabiano Ferreira, Felipe Arashida, Juliano Araujo, Daniela Augustinho e Jurandir Segundo.

Aos funcionários do Departamento de Biologia e da Monsanto do Brasil, pela disposição em ajudar.

Ao GEN – grupo de estudo em Genética e Melhoramento, pela disponibilização de informações técnicas.

E a todos que de algum modo contribuíram para a realização deste curso e deste trabalho, meus sinceros agradecimentos.

Muito obrigado !

*“Para conquistar coisas importantes, devemos não apenas agir mas também  
sonhar, não apenas planejar mas também acreditar”.*

*Anatole France*

## RESUMO GERAL

O objetivo desse trabalho foi tratar a abordagem multiambientes (MET) sob a ótica de modelos mistos associados à análise de fatores para avaliação de estabilidade e adaptabilidade em programas de melhoramento que normalmente apresentam desbalanceamento de dados. Além disto, comparar os ambientes de safra e de safrinha através deste estudo de estabilidade e adaptabilidade. Para isso, 28 híbridos foram avaliados em 35 locais distribuídos em quatro épocas/anos (Safra 2010, Safrinha 2011, Safra 2011 e Safrinha 2012). Alguns desses híbridos foram avaliados nas primeiras épocas e não avaliados nas posteriores ou vice-versa; desse modo, o conjunto de dados utilizados nesse trabalho simula a dinâmica de um programa de melhoramento com descarte e inclusão de genótipos ao longo dos anos. Foi observado no biplot de escores e cargas fatoriais que os locais foram mais similares dentro de épocas do que entre épocas, sugerindo que um mesmo local pode se comportar de maneira muito divergente de ano para ano. Esse fato questiona a elaboração de mega-ambientes em dados de apenas um ano. Observou-se empiricamente que modelos fatoriais analíticos têm interpretação direta com a análise GGE biplot, dado que o primeiro escore fatorial apresentou um ajuste perfeito ( $r^2 = 0.99$ ) com os E-BLUPs dos genótipos. Dada a pressuposição de normalidade para os escores fatoriais foi possível a construção de elipses de confiança e posterior comparação direta dos genótipos no biplot. Observou-se que é possível realizar análise de estabilidade e adaptabilidade em experimentos desbalanceados com descarte e inclusão de genótipos ao longo dos anos. Essa abordagem permitiu verificar determinadas tendências de um programa de melhoramento, comparando diretamente híbridos desenvolvidos para ambientes de safra e safrinha. A interpretação do biplot é direta e intuitiva e com as mesmas propriedades da análise GGE biplot ou AMMI se for o caso.

Palavras-chave: SREG. Elipse de confiança. MET. Matrizes não estruturadas (UN).

## GENERAL ABSTRACT

The objective of this work was to treat the multi-environment approach (MET) under the point of view of mixed models associated with the factor analysis to evaluate the stability and adaptability in breeding programs which would normally present unbalanced data. In addition, compare the environments of the harvest and interim-harvest through this stability and adaptability study. In order to do this, 28 hybrids were evaluated in 35 locations distributed in four periods/years (2010 Harvest, 2011 Interim-harvest, 2011 Harvest and 2012 Interim-harvest). Some of these hybrids were evaluated in the first periods and not evaluated in those subsequent periods or vice-versa; thus, the set of data used in this work simulates the dynamics of a breeding program with discarding and inclusion of genotypes over the years. On the score and factorial loads biplot, it was observed that the locations were more similar inside periods than in between periods, suggesting that a same location may behave in a very divergent manner from year to year. This fact questions the elaboration of mega-environments in data from a single year. It was empirically observed that the analytical factorial models present direct interpretation with the GGE biplot analysis, given that the first factorial score presented a perfect adjustment ( $r^2 = 0.99$ ) with the genotype E-BLUPs. Given the normality presupposition for the factorial scores, it was possible to construct confidence ellipses and posterior direct comparison of genotypes in the biplot. It was observed that it was possible to perform stability and adaptability analysis in unbalanced experiments with discarding and inclusion of genotypes over the years. This approach allowed the verification of certain tendencies of a breeding program, directly comparing hybrids developed for harvest and interim-harvest. The interpretation of the biplot is direct and intuitive, and presents the same properties of the GGE biplot analysis or the AMMI, if that is the case.

Keywords: SREG. Confidence ellipse. MET. Unstructured matrixes (UN).

## SUMÁRIO

PRIMEIRA PARTE.....	10
1 INTRODUÇÃO.....	10
2 REFERENCIAL TEÓRICO.....	15
2.1 Interação genótipos por ambientes.....	15
2.2 Análise de fatores (FA).....	19
2.3 Análise de fatores sob modelos multiplicativos mistos (FAMM).....	22
2.4 Estrutura Fator Analítico.....	25
REFERÊNCIAS.....	30
SEGUNDA PARTE - ARTIGO.....	34
ARTIGO 1 Aplicação de modelos mistos na avaliação de estabilidade e adaptabilidade em milho utilizando dados desbalanceados: uma interpretação GGE Biplot em modelos fatoriais analíticos.....	34
1 INTRODUÇÃO.....	36
2 MATERIAL E MÉTODOS.....	40
2.1 Material genético e locais.....	40
2.2 Análise MET em modelos mistos com estruturas UN.....	42
2.3 Análise GGE e aproximação fatorial analítico.....	45
3 RESULTADOS.....	48
4 DISCUSSÕES.....	58
REFERÊNCIAS.....	64

## **PRIMEIRA PARTE**

### **1 INTRODUÇÃO**

No Brasil, o milho é cultivado em duas épocas distintas, na primeira época também chamada de safra, cuja semeadura se estende do final de setembro até meados de dezembro; e na segunda época, também conhecida como safrinha, com semeadura de janeiro a abril. A safrinha é bastante recente tendo seu primeiro registro na década de 1980, contudo, ultrapassou em importância a área cultivada no verão, sendo que na safra 2013/2014 as estimativas para a safrinha apontaram para uma participação de 57,8% do total cultivado no Brasil (COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB, 2014).

As condições edafoclimáticas do cultivo de safrinha e de verão variam bastante, entretanto a grande maioria dos cultivares plantados na safra são também recomendados para a safrinha. Isto se deve a duas razões principais. A primeira devido aos programas de melhoramento de milho na safrinha serem recentes e menores que na safra. E a segunda razão se deve pelo fato de muitos cultivares oriundos dos programas de melhoramento de safra apresentarem bom desempenho e adaptação ao cultivo de safrinha, e por isso serem recomendados também para esta época de semeadura.

Existem atualmente 467 cultivares de milho disponíveis no Brasil (CRUZ; PEREIRA FILHO; QUEIROZ, 2014), no entanto o plantio é concentrado em um pequeno número de cultivares. O sucesso de um cultivar no mercado está associado ao elevado desempenho produtivo, a características agrônômicas desejáveis e sua elevada estabilidade. Apesar disto a recomendação de novos cultivares baseia-se basicamente na produtividade média nos ambientes avaliados (FRITSCHÉ-NETO et al., 2010).

Dentre os principais desafios para o melhorista de milho está a escolha de genótipos que associem alta produtividade com estabilidade. E neste sentido a discriminação de ambientes em grupos ou mega-ambientes e a estimação da adaptação trazem complexidade aos programas de melhoramento além de elevar os custos. O genótipo ideal seria aquele que preservasse a máxima produção dentro de um mega-ambiente ou entre mega-ambientes.

Para identificação desse genótipo ideal, torna-se necessária a quantificação da interação genótipo por ambientes (GE). Nesse sentido, diversas metodologias complementares ou não à ANOVA foram propostas com objetivo de decompor essa interação GE e identificar genótipos estáveis. Dentre essas abordagens, destacam-se as abordagens AMMI (additive and multiplicative models interaction) e GGE Biplot (CROSSA, 1990; GABRIEL, 1971; GAUCH; ZOBEL, 1997; YAN et al., 2000). Essas análises exploratórias têm se mostrado extremamente eficientes na detecção de padrões de interação GE, identificação de mega-ambientes, genótipos ideais, adaptação específica etc. As vantagens e desvantagens de ambas as análises já foram discutidas de forma profunda em Gauch, Piepho e Annicchiarico (2008) e Yan et al. (2007). No entanto, enquanto seus defensores focam a crítica nas particularidades de cada análise, um aspecto extremamente importante é deixado de lado, que é: como aplicar esses métodos frente ao dinamismo de um programa de melhoramento, principalmente quando falamos da cultura do milho?

Não é novidade que a cada ano agrícola, diversos genótipos são descartados e tantos outros são incorporados ao sistema de avaliação de híbridos. Nesse sentido, um programa de melhoramento é, necessariamente, desbalanceado. Além disso, às vezes o melhorista deseja comparar genótipos que já foram descartados com os atuais, ou até mesmo avaliar o progresso de seu programa de forma eficaz. Sob essas condições, tanto a abordagem AMMI tradicional quanto a GGE-biplot são falhas, pois dependem da decomposição de

valor singular de uma matriz de co-variâncias que, por sua vez, depende de matrizes balanceadas de GE ou da interação genótipo+genótipo por ambiente (G+GE). Logo, à medida que se introduz ou se retira híbridos ao longo de um programa de melhoramento, ou até mesmo quando se obtém perdas de materiais em determinados ambientes, dentro de um mesmo ano agrícola, essa matriz retangular de GE ou G+GE torna-se extremamente desbalanceada, dificultando a aplicação do método de componentes principais. Outro aspecto relevante é que o AMMI e GGE com interpretação baseada em ANOVA reivindicam para uma fina descrição da interação em termos de soma de quadrados e inferência, aditividade, homogeneidade das variâncias e independência de resíduos. Assim, esse tipo de análise não explora de forma eficaz a particularidade dos dados que podem apresentar heterogeneidade de variâncias e dependência.

Nesse sentido, Piepho (1998) e Smith, Cullis e Thompson (2001) propuseram a análise MET sob a ótica de modelos mistos multivariados utilizando estrutura fatorial analítica (FA) considerando ambientes ou genótipos como efeitos aleatórios respectivamente. A vantagem dessa abordagem relaciona-se à capacidade de o modelo lidar com dados altamente desbalanceados, heterogeneidade de variâncias e covariâncias residuais e genotípicas. Nesse sentido, essa técnica torna-se naturalmente atrativa para lidar com dados de um programa de melhoramento.

Ao comparar estruturas de covariâncias em modelos mistos aplicados em análises de ensaios de múltiplos ambientes (MET), Kelly et al. (2007) questionam que: embora bem estudada, a técnica ainda era pouco aplicada nos programas de melhoramento da Austrália. No Brasil, o mesmo cenário é observado, ou seja, embora a abordagem já seja bem conhecida (REZENDE; THOMPSON, 2004), pouco uso tem-se observado, talvez por falta de softwares gratuitos ou até mesmo pela sua complexidade e alta demanda computacional.

Para tentar diminuir essas limitações presentes no algoritmo inicial proposto por Smith, Cullis e Thompson (2001) e Thompson et al. (2003) propuseram uma alteração do algoritmo AI (Average Information), no que eles chamaram de estimação de rank reduzido, com objetivo de evitar casos de Heyhood, comumente obtidos no trabalho original de Smith, Cullis e Thompson (2001); ou seja, quando as variâncias convergem para estimativas iguais a zero ou negativas em alguns casos. No entanto, percebe-se que a aplicação do método AI não é tão simples, tão pouco sua adaptação - como pode ser observado nos doze passos descritos no trabalho de Thompson et al. (2003) e, embora utilizassem a estrutura fator analítico, o modelo, ainda sim, é menos informativo do que um modelo baseado em matrizes não estruturadas (UN), que estima todas as variâncias e covariâncias residuais e genotípicas.

Os modelos FA buscam reduzir o número de parâmetros referentes às estimativas de covariâncias genotípicas UN, contudo, sua aplicação é complicada e não há garantias de convergência. Por outro lado, sua versão mais completa (modelos UN) pode ser aplicada facilmente utilizando o algoritmo EM descrito por Dempster, Laird e Rubin (1977), e que apresenta um comportamento mais estável, pois restringe as estimativas dos componentes no espaço do parâmetro. Além disso, esse tipo de estrutura é normalmente mais informativa do que suas versões com menos parâmetros, além de ser mais tratável em casos de matrizes de posto incompleto (THOMPSON et al., 2003).

Estruturas UN exigem a estimação de uma grande quantidade de parâmetros que crescem de forma abrupta quando se tem um grande número de ambientes. Também, as estimativas não são diretamente interpretáveis em termos de estabilidade e adaptabilidade.

Nesse contexto, diversos trabalhos foram realizados comparando essas duas estruturas mais tradicionais (FA(2) e UN) com várias outras, tais como (FA( $\neq$  2), simetria composta, autorregressiva de primeira ordem, estrutura

diagonal, etc.). Contudo, a recomendação de uma ou outra estrutura pode depender do tipo de dados que se tem em mãos (KELLY et al., 2007; MEYER, 2009; PIEPHO, 1998). Embora todas essas estruturas sejam válidas e a recomendação varie de acordo com o tipo de dados, a FA(2) tornou-se padrão quando se busca análise MET em modelos mistos variados (KELLY et al., 2007).

Uma alternativa à ideia de FA seria a aplicação de análise de componentes principais (PCA) nas matrizes de  $G + GE$  em modelos mistos multivariados sob estruturas UN. Porém, as covariâncias representadas na decomposição de SVD não são estimativas REML, ou seja, as componentes de dispersão estimadas para decomposição de valor singular nos métodos AMMI e GGE convergem - exceto para as ocorrências de homogeneidade de variâncias e dados balanceados -, na maioria dos casos, em quadrados médios de componentes viciados. Nesse sentido, o objetivo desse trabalho foi avaliar híbridos de um programa de melhoramento de milho com modelos multivariados mistos em estrutura UN construindo o biplot baseado na análise de fatores com estimativas REML de covariâncias e comparando esses resultados e interpretações com GGE biplot tradicional.

## **2 REFERENCIAL TEÓRICO**

### **2.1 Interação genótipos por ambientes**

Em programas de melhoramento de plantas existe a busca constante pela seleção e recomendação dos melhores genótipos ou cultivares no que se refere à produtividade, estabilidade de produção e ampla adaptabilidade aos mais variados ambientes. Neste processo, a interação entre genótipos e ambiente GxE representa um dos maiores desafios aos melhoristas (CARVALHO et al., 2005).

Essa interação pode ser percebida pelas respostas diferenciadas de genótipos quando avaliados em diferentes ambientes (ano, local, época de semeadura, entre outros), ou seja, quando os genótipos não possuem performances semelhantes em todos os ambientes de teste (CROSSA, 1990). A interação G x E representa o principal motivo por que diversos experimentos devem ser conduzidos em vários ambientes, os denominados ensaios multiambientais (MET – multienvironment trials) (GAUCH, 1992).

De acordo com Allard e Bradshaw (1964), as variações ambientais que contribuem para a interação com os genótipos podem ser classificadas em dois

tipos: previsíveis e não previsíveis. O primeiro tipo abrange aquelas características do clima cuja variação é previsível, tais como local, solo, comprimento do dia, insolação e também as variações ambientais determinadas pela ação do homem, como, época e densidade de semeadura, níveis de adubação e outros tratos culturais. As variações imprevisíveis são as oscilações no clima que não podem ser previstas, tais como volume e distribuição das chuvas, variações da temperatura e outros.

A resposta dos genótipos frente a esse tipo de variação pode ser classificada como do tipo simples ou complexa. Como exemplo podemos

verificar na figura 1 que os genótipos apresentam comportamentos semelhantes nos dois ambientes (E1 e E2). Caracterizando, portanto, um caso específico onde não ocorre interação, pois a diferença de resposta deve-se exclusivamente à diferença entre ambientes e a recomendação do melhor genótipo é a mesma para ambos ambientes.

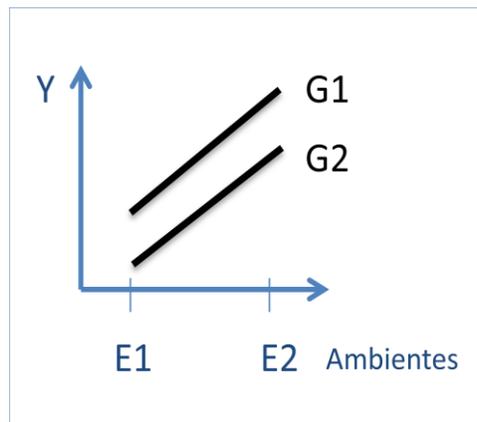


Figura 1 Comportamento de dois genótipos (G1 e G2) em duas condições ambientais (E1 e E2) com ausência de interação

Outro exemplo pode ser observado na figura 2, onde a resposta dos dois genótipos (G1 e G2) é diferente nos dois ambientes, ou seja, G1 tem resposta mais acentuada à melhoria do ambiente. Neste caso, porém, ocorre interação com ambientes, mas que não ocasiona grandes problemas como a seleção, pois a classificação dos genótipos nos dois ambientes não é alterada e, por esta razão, é denominada de interação simples.

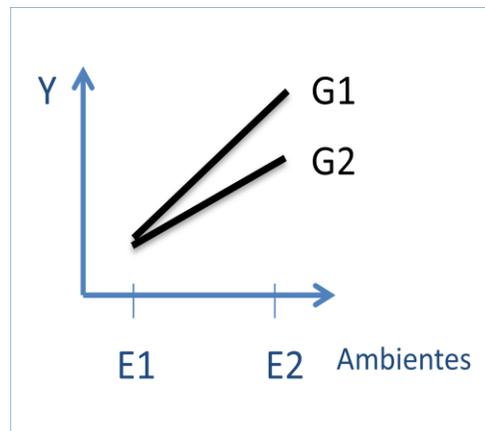


Figura 2 Comportamento de dois genótipos (G1 e G2) em duas condições ambientais (E1 e E2) com interação simples

Na figura 3 (a e b) observa-se uma inversão de comportamento dos genótipos nos dois ambientes. O G1 foi superior no E1 e inferior no E2 (Figura 3a). Esta corresponde a uma situação de interação complexa (cruzada), onde normalmente, existe um genótipo mais adaptado para cada ambiente.

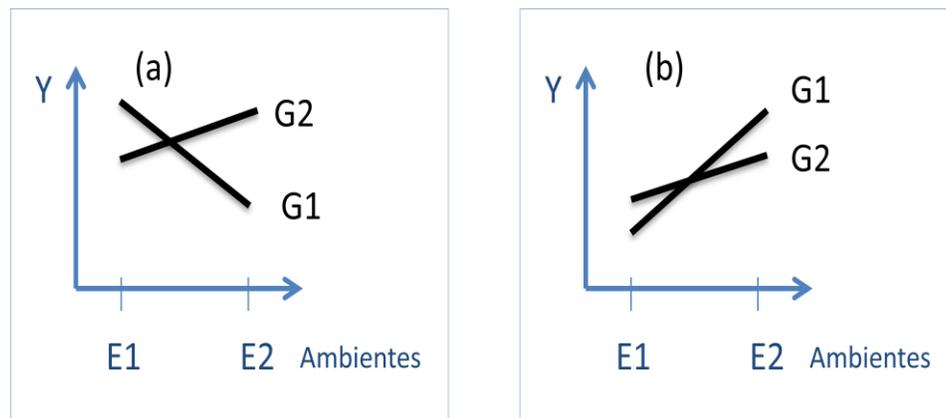


Figura 3 Comportamento de dois genótipos (G1 e G2) em duas condições ambientais (E1 e E2) com interação cruzada ou qualitativa

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 genótípicas de cada material, que conferem maior ou menor adaptabilidade e estabilidade de produção.

Quando se consideram vários genótipos avaliados em vários ambientes, a combinação dos processos descritos nas figuras 1, 2 e 3 formam um emaranhado de situações difíceis de serem interpretadas, exigindo métodos adequados de análise da interação G x E.

Existe uma concordância entre os melhoristas de plantas de que a interação G x E tem um importante significado para a obtenção de variedades superiores. Isso se deve ao fato de que a existência da mesma produz uma barreira de dificuldades 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 foram 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).

Para atenuar a interação genótipos x ambientes algumas medidas podem ser tomadas, tais como: a) identificar uma cultivar específica para cada ambiente, b) identificar cultivares com maior estabilidade fenotípica e c) realizar o zoneamento ecológico ou estratificação ambiental (RAMALHO et al., 2012).

Considerando o item b diversos métodos foram desenvolvidos com o objetivo de quantificá-la (CROSSA, 1990). Contudo, Duarte e Vencovsky (1999) afirmaram que alguns métodos baseados em regressão tendem a explicar a variação devido à interação numa única dimensão, simplificando os modelos de resposta, quando, na verdade, a interação pode ser bastante complexa. Além

disso, esses métodos, de forma geral, não permitem identificar se interações específicas de genótipos com ambientes são positivas ou negativas, dificultando a capitalização dos efeitos positivos da interação. Crossa (1990) propõe que a aplicação de métodos multivariados pode ser útil para explorar melhor as informações contidas nos dados. O autor recomenda, principalmente, o método AMMI que vem sendo muito aplicado nos últimos anos.

A principal crítica a esses métodos se refere ao seu caráter apenas descritivo da interação, onde a inferência pode ser prejudicada se alguns pressupostos da ANOVA não forem atendidos. Ao passo que a aditividade está intimamente relacionada à soma de quadrados, a inferência depende exclusivamente de outras suposições como homogeneidade de variâncias e resíduos não correlacionados. Nesse sentido, a descrição AMMI da interação poder ser prejudicada quanto à sua relação com a variância genética ou mesmo a significância do teste dos efeitos de tratamentos. Outro problema surge com dados desbalanceados que limitam a aplicação da técnica de componentes principais.

## **2.2 Análise de fatores (FA)**

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 JÚNIOR et al., 2005). De maneira geral, a análise de fatores aborda o problema de analisar a estrutura das interrelações (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  $z$  (Equação 1) um vetor de variáveis aleatórias, com matriz de covariância  $\Sigma$ , pode-se representar o modelo fatorial como:

$$Z = \mu + \Lambda f + \delta \quad (1)$$

Em que :

$\mu$  : representa o vetor de médias;

$\Lambda$  : matriz  $q \times m$  de cargas fatoriais;

$f$  : vetor  $m \times 1$  de fatores comuns;

$\delta$  : o vetor  $q \times 1$  de fatores específicos;

Na forma mais comum de análise fatorial, as colunas de  $\Lambda$  são ortogonais, ou seja,  $\gamma_i' \gamma_j = 0$  para  $i \neq j$ , em que  $\gamma_i$  é a  $i$ -ésima coluna de  $\Lambda$ . Daí que os elementos de  $f$  são não correlacionados. Além disso, os fatores

comuns são assumidos ter variância unitária, isto é,  $Var(f) = I$ . As colunas  $\gamma_i$  são determinadas como os autovetores correspondentes de  $\Sigma$ , escalado pela raiz quadrada dos respectivos autovalores. No entanto,  $\Lambda$  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)  $\delta_i$  são assumidos como distribuídos de forma independente com variâncias heterogêneas  $\psi_i$ , sendo os vetores  $f$  e  $\delta$  não correlacionados. Assim, a matriz de covariância de  $z$  (Equação 2) sob o modelo de FA pode ser dada por:

$$Var(z) = \Sigma_{FA} = \Lambda\Lambda' + \Psi \quad (2)$$

Em que:

$\Psi = diag(\psi_i)$  é uma matriz diagonal de variâncias específicas. Isto implica que todas as covariâncias entre os níveis de  $z$  são devidas aos fatores comuns, enquanto que os fatores específicos explicam a variação adicional de elementos individuais de  $z$  (MEYER, 2009). Para  $m$  fatores comuns, descreve-se  $q(q+1)/2$  elementos de  $\Sigma_{FA}$  através de  $p = q + mq - m(m-1)/2$  parâmetros, que consistem em  $q$  variâncias específicas  $\psi_i$  e  $m(2q - m + 1)/2$  elementos de  $\Lambda$  e os restantes  $m(m-1)/2$  elementos determinados por restrições de ortogonalidade.

Para valores pequenos de  $m$ , 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 desestruturado, o número de fatores comuns  $q(q+1)/2$  que pode ser definido é restrito.

Se todas as variâncias específicas  $\psi_i$  são diferentes de zero, o número mínimo de características para as quais é imposta uma estrutura FA para redução do 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  $z$ . 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,  $\Sigma = \sigma^2 \mathbf{1}\mathbf{1}' + \psi \mathbf{I}$ , que é um modelo de FA com um único fator comum e  $\Lambda = \sigma \mathbf{1}$  (onde  $\mathbf{1}$  refere-se a um vetor com todos os elementos igual a um) e variâncias específicas iguais  $\psi$  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.

### 2.3 Análise de fatores sob modelos multiplicativos mistos (FAMM)

A análise de grupos de experimentos ou ensaios 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.

No entanto, frequentemente essas suposições são violadas e, não obstante, inferências são realizadas em torno da estimativa de interação. Uma alternativa seria o relaxamento dessas suposições e adoção de modelos que abarquem heterogeneidade de variâncias, correlação residual e níveis de desbalanceamento.

A modelagem de efeitos da interação G x E para ensaios multiambientais (METs) dentro de uma estrutura de modelo misto é uma prática pouco comum em muitos programas de melhoramento de plantas (KELLY et al., 2007). Abaixo será apresentada a análise de modelos mistos associado à análise de fatores.

O modelo misto tradicional (Equação 3) pode ser representado por:

$$y = X\beta + Zu + e \quad (3)$$

Onde se assume que:  $u \sim N(0, \Sigma)$  e  $e \sim N(0, R)$

Em que:

$y$  : vetor de observações;

$b$  : vetor dos efeitos fixos, com matriz de incidência

$u$  : vetor dos efeitos aleatórios, com matriz de incidência  $Z$ ;

$e$  : vetor de erros aleatórios.

O modelo fator de analítico (FA) é uma forma parcimoniosa usada para aproximar a forma totalmente desestruturada da matriz de variância-covariância genética ( $\Sigma$ ) 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:

$$y = Xb + Z[Lf + \delta] + e, \text{ com } u = [Lf + \delta], \text{ em que:}$$

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

$f$  é o vetor de escores fatoriais para os genótipos nos ambientes;

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

Sob esse modelo, a matriz de covariância genética (Equação 4) é dada por:

$$\Sigma = \Lambda\Lambda' + \psi \quad (4)$$

Em que:

$\Lambda\Lambda' = VD_\alpha V'$ , onde  $D_\alpha$  é a matriz diagonal dos  $m$  autovalores e  $V$  é a matriz dos autovetores. Na técnica FAMM, a estrutura de covariância (Equação 5) pode ser simplificada assumindo:

$$\Sigma = \Lambda_p\Lambda_p' + \Psi \quad (5)$$

Em que:

$\Lambda_p$ : é a matriz das cargas dos fatores;

$\Psi$ : é a matriz diagonal de variâncias específicas

Dentro da abordagem de modelos mistos pode-se estimar os autovalores e autovetores diretamente sem a necessidade de se estimar  $\Sigma$  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  $(\Lambda \otimes I_g)f$  é singular, isto conduz à estimação sob posto reduzido, portanto restrições devem ser impostas aos parâmetros do modelo fator analítico (RESENDE, 2007).

Uma característica fundamental do modelo de FA para os dados MET é a generalidade da estrutura de variância associado para efeitos G x E, seja no ambiente ou na dimensão do genótipo. Contudo, o modelo que irá proporcionar

o melhor ajuste para os dados, deriva de uma matriz não estruturada (SMITH; CULLIS; THOMPSON, 2005).

Smith, Cullis e Thompson (2001) utilizaram o modelo FA motivados pela abordagem da genética quantitativa para interação G x E tratando a interação como a ausência de correlação genética entre ambientes. Falconer e Mackay (1996) afirmam que o conceito de correlação genética pode ser aplicado à solução de alguns problemas relacionados com a interação genótipo e ambientes. Assim, um caráter medido em dois ambientes deve ser considerado não como um caráter, mas como dois. Logo, se a correlação genética entre eles é elevada, o desempenho em dois ambientes diferentes representa quase o mesmo caráter. Por outro lado, se essa correlação for baixa ou mesmo negativa, então significa a presença do efeito da interação genótipos por ambientes.

Apesar das recomendações de Piepho (1997, 1998) e Smith, Cullis e Thompson (2001), os modelos de FA não são amplamente utilizados na Austrália para a análise regular de dados MET (KELLY et al., 2007). No Brasil a situação não é diferente; seja pela complexidade técnica da análise ou pela alta demanda computacional. Como regra geral, os programas de melhoramento ainda analisam seus materiais utilizando dados de apenas um ano, ou, quando é possível avaliar em épocas diferentes, realiza-se um peneiramento de genótipos de modo a balancear o experimento e tornar os tratamentos comparáveis - descartando, assim, informações que poderiam ser relevantes.

#### **2.4 Estrutura Fator Analítico**

Na análise de experimentos multiambientes (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  $u_{ij}$ . O modelo de análise é postulado em

termos de efeitos genotípicos não observáveis em diferentes ambientes (Equação 6):

$$u_{ij} = \sum_{r=1}^k \lambda_{jr} f_{ir} + \delta_{ij}, \quad (6)$$

Em que:

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

$\lambda_{jr}$  : carregamento do fator  $r$  no ambiente  $j$ ;

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

$\delta_{ij}$  : 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). O modelo de fator analítico assume dependência em um conjunto de fatores hipotéticos aleatórios  $f_r^{(sx1)}$ , ( $r = 1 \dots k < s$ ). Em notação vetorial (Equação 7), o modelo de análise de fatores para estes efeitos em diferentes ambientes pode ser dado por:

$$u_s = (\lambda_1 \otimes I_g) f_1 + \dots + (\lambda_k \otimes I_g) f_k + \delta \quad (7)$$

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 (Equação 8), o modelo pode ser descrito como:

$$u_s = (\Lambda \otimes I_g) f + \delta \quad (8)$$

Em que :

$$\Lambda^{(s \times k)} = [\lambda_1 \dots \lambda_k]$$

$$f^{(g \times 1)} = (f_1', f_2' \dots f_k')$$

A distribuição conjunta (Equação 9) de  $f$  e  $\delta$  pode ser dada por:

$$\begin{pmatrix} f \\ \delta \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_k \otimes I_g & 0 \\ 0 & \Psi \otimes I_g \end{pmatrix} \right), \quad (9)$$

Em que:

$$\Psi = \text{diag}(\psi_1 \dots \psi_p)$$

$\psi_i$  variância específica para o  $i$ -ésimo ensaio.

A matriz de variância para efeitos de genótipos nos ambientes (Equação 10) é dada por:

$$\text{var}(u_s) = (\Lambda \otimes I_g) \text{var}(f) (\Lambda' \otimes I_g)' + \text{var}(\delta) = (\Lambda \Lambda' + \Psi) \otimes I_g \quad (10)$$

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

$$\sigma_j^2 = \sum_{r=1}^k \lambda_{jr}^2 + \psi_j : \text{variância genotípica em ambiente } j;$$

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

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

genotípica entre os ambientes  $j$  e  $j'$

A equação para  $u_s$  tem a forma de uma regressão (aleatória) em  $k$  covariáveis ambiental  $\lambda_1 \dots \lambda_k$ , na qual todas as regressões passam pela origem. Nessa suposição temos um modelo semelhante ao SREG. No entanto, pode-se utilizar o intercepto (não zero) separado para cada genótipo. Isto é equivalente ao modelo AMMI com efeitos de genótipos principais,  $u_g$  e um modelo fator analítico  $k$  para interação  $G \times E$ . Realizando essa separação, a expressão de  $u_g$  torna-se:

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

O modelo pode ser escrito como:

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

Para o modelo de fator analítico, os BLUPs referentes aos escores  $f$  e resíduos  $\delta$  podem ser obtidos em termos de  $u_g$  como:

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

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

O modelo FA é um modelo 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. Nesse contexto, surge a analogia FAMM também chamada AMMI aleatórios (RESENDE, 2007). Também o modelo FA pode ser descrito como um SREG se não houver separação dos efeitos de genótipos e ambientes.

## REFERÊNCIAS

ALLARD, R. W.; BRADSHAW, A. D. Implications of genotype-environmental interactions in applied plant breeding. **Crop Science**, Madison, v. 4, n. 5, p. 503-508, Sept. 1964.

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.

CARVALHO, H. W. L. de et al. Adaptabilidade e estabilidade de cultivares de milho no Nordeste brasileiro. **Pesquisa Agropecuária Brasileira**, Brasília, v. 40, n. 5, p. 471-477, 2005.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Séries históricas**. Disponível em: <<http://www.conab.gov.br/conteudos.php?a=1252&t=2>>. Acesso em: 22 jan. 2014.

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

CRUZ, J. C.; PEREIRA FILHO, I. A.; QUEIROZ, L. R. **Milho: cultivares para 2013/2014**. Sete Lagoas: EMBRAPA Milho e Sorgo, 2014. Disponível em: <<http://www.cnpms.embrapa.br/milho/cultivares/index.php>>. Acesso em: 16 abr. 2014.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. F. Maximum likelihood from incomplete data with EM algorithm. **Journal of Royal Statistic Society B**, London, v. 39, p. 1-38, 1977.

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

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

FRITSCHÉ-NETO, R. et al. Análise de fatores e SREG GGE biplot para a estratificação da interação genótipos  $\times$  ambientes em milho. **Ciência Rural**, Santa Maria, v. 40, n. 5, p. 1043-1048, maio 2010.

GABRIEL, K. R. The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, London, v. 58, p. 453-467, 1971.

GAUCH, H. G. **Statistical analysis of regional yield trials: AMMI analysis of factorial designs**. New York: Elsevier, 1992. 278 p.

GAUCH, H. G.; PIEPHO, H. P.; ANNICCHIARICO, P. Statistical analysis of yield trials by AMMI and GGE: further considerations. **Crop Science**, Madison, v. 48, n. 3, p. 866-889, Mar. 2008.

GAUCH, H. G.; ZOBEL, R. W. Identifying mega-environments and targeting genotypes. **Crop Science**, Madison, v. 37, p. 311-326, Mar. 1997.

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

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.

MEYER, K. Factor-analytic models for genotype  $\times$  environment type problems and structured covariance matrices. **Genetics Selection Evolution**, Les Ulis, v. 41, n. 21, 2009. Disponível em: <<http://www.gsejournal.org/content/41/1/21>>. Acesso em: 10 jan. 2014.

PIEPHO, H. P. Analyzing genotype-environment data by mixed models with multiplicative terms. **Biometrics**, Washington, v. 53, n. 4, p. 761-767, 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.

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. **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; 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.

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. 6, p. 449-462, Dec. 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. 4, p. 1138-1147, Dec. 2001.

THOMPSON, R. et al. A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. **Biometrics**, Washington, v. 45, n. 4, p. 445-459, 2003.

YAN, W. et al. GGE biplot vs. AMMI analysis of genotype-by-environment data. **Crop Science**, Madison, v. 47, n. 2, p. 643-655, Mar. 2007.

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, May/June 2000.

**SEGUNDA PARTE - ARTIGO**

**ARTIGO 1 Aplicação de modelos mistos na avaliação de estabilidade e adaptabilidade em milho utilizando dados desbalanceados: uma interpretação GGE Biplot em modelos fatoriais analíticos**

Andre Gradowski de Figueiredo <sup>1</sup>

Heyder Diniz Silva<sup>2</sup>

Marcio Balestre<sup>3</sup>

Renzo Garcia Von Pinho<sup>4</sup>

**Artigo formatado de acordo com a NBR 6022 (ABNT, 2003).**

---

<sup>1</sup> Monsanto do Brasil LTDA. Av. Nações Unidas, 12901, Torre Norte – Brooklin, Cep 04578-910. São Paulo, São Paulo, Brasil.

<sup>2</sup> Monsanto do Brasil LTDA. C.P 3037. Uberlândia, Minas Gerais, Brasil.

<sup>3</sup> Departamento de Agricultura, Universidade Federal de Lavras. C.P 3037. Lavras. Minas Gerais, Brasil.

<sup>4</sup> Departamento de Ciências Exatas, Universidade Federal de Lavras, C.P 3037. Lavras, Minas Gerais, Brasil

## RESUMO

O objetivo desse trabalho foi tratar a abordagem multiambientes (MET) sob a ótica de modelos mistos associados à análise de fatores para avaliação de estabilidades e adaptabilidade em híbridos. Para isso, 28 híbridos foram avaliados em 35 ambientes distribuídos em quatro épocas/anos (Safrinha 2010, Safra 2010, Safra 2011 e Safrinha 2011). Alguns desses híbridos foram avaliados nas primeiras épocas e não avaliados nas posteriores ou vice-versa; desse modo, o conjunto de dados utilizados nesse trabalho simulam a dinâmica de um programa de melhoramento com descarte e inclusão de genótipos ao longo dos anos. Foi observado no biplot de escores e cargas fatoriais que os locais foram mais similares dentro de épocas do que entre épocas, sugerindo que um mesmo local pode se comportar de maneira muito divergente de ano para ano. Esse fato questiona a elaboração de mega-ambientes em dados de apenas um ano. Observou-se empiricamente que modelos fatoriais analíticos têm interpretação direta com a análise GGE biplot, dado que o primeiro escore fatorial apresentou um ajuste perfeito ( $r^2 = 0.99$ ) com os E-BLUPs dos genótipos. Dada a pressuposição de normalidade para os escores fatoriais foi possível a construção de elipses de confiança e posterior comparação direta dos genótipos no biplot. Observou-se que é possível realizar análise de estabilidade e adaptabilidade em experimentos desbalanceados com descarte e inclusão de genótipos ao longo dos anos. Essa abordagem permitiu verificar determinadas tendências de um programa de melhoramento, comparando diretamente híbridos desenvolvidos para ambientes de safra e safrinha. A interpretação do biplot é direta e intuitiva e com as mesmas propriedades da análise GGE biplot ou AMMI se for o caso.

Palavras-chave: SREG. Elipse de confiança. MET. Matrizes não estruturadas (UN).

## 1 INTRODUÇÃO

O principal desafio do melhorista é a escolha de genótipos com alta produtividade e estabilidade. Além disto, a discriminação de ambientes em grupos ou mega-ambientes e a estimação da adaptação dificultam o trabalho do melhorista na recomendação do melhor genótipo. Assim, o genótipo ideal seria aquele que preservasse a máxima produção dentro de um mega-ambiente ou entre mega-ambientes.

Para identificação desse genótipo ideal, torna-se necessária a quantificação da interação genótipo por ambientes (GE). Nesse sentido, diversas metodologias complementares ou não à ANOVA foram propostas com objetivo de decompor essa interação GE e identificar genótipos estáveis. Dentre essas abordagens, destacam-se as abordagens AMMI (additive and multiplicative models interaction) e GGE Biplot (CROSSA, 1990; GAUCH; ZOBEL, 1997; YAN et al., 2000). Essas análises descritivas têm se mostrado extremamente eficientes na descoberta de padrões de interação GE, identificação de mega-ambientes, genótipos ideais, adaptação específica etc. As vantagens e desvantagens de ambas as análises já foram discutidas de forma profunda em Gauch, Piepho e Annicchiarico (2008) e Yan et al. (2007). No entanto, enquanto seus defensores focam a crítica nas particularidades de cada análise, um aspecto extremamente importante é deixado de lado, que é: como aplicar esses métodos frente ao dinamismo de um programa de melhoramento, principalmente quando falamos da cultura do milho?

Não é novidade que a cada ano agrícola, diversos genótipos são descartados e tantos outros são incorporados ao sistema de avaliação de

híbridos. Nesse sentido, um programa de melhoramento é, necessariamente, desbalanceado. Além disso, às vezes o melhorista deseja comparar genótipos que já foram descartados com os atuais, ou até mesmo avaliar o progresso de seu programa de forma eficaz. Sob essas condições, tanto a abordagem AMMI tradicional quanto a GGE biplot são falhas, pois dependem da decomposição de valor singular de uma matriz de covariâncias que, por sua vez, depende de matrizes com dados balanceados de GE ou G+GE. Logo, à medida que se introduz ou se retira híbridos ao longo de um programa de melhoramento, ou até mesmo quando se obtém perdas de materiais em determinados locais, dentro de um mesmo ano agrícola, essa matriz retangular de GE ou G+GE torna-se extremamente desbalanceada, dificultando a aplicação do método de componentes principais. Outro aspecto relevante é que o AMMI e GGE com interpretação baseada em ANOVA reivindicam - para uma fina descrição da interação em termos de soma de quadrados e inferência-, aditividade, homogeneidade das variâncias e independência de resíduos. Assim, esse tipo de análise não explora de forma eficaz a particularidade dos dados que podem apresentar heterogeneidade de variâncias e dependência.

Nesse sentido, Piepho (1998) e Smith, Cullis e Thompson (2001) propuseram a análise MET sob a ótica de modelos mistos multivariados utilizando estrutura fatorial analítico (FA) considerando locais ou genótipos como efeitos aleatórios respectivamente. A vantagem dessa abordagem relaciona-se à capacidade do modelo lidar com dados altamente desbalanceados, heterogeneidade de variâncias e covariâncias

residuais e genotípicas. Nesse sentido, essa técnica torna-se naturalmente atrativa para lidar com dados de um programa de melhoramento.

Ao comparar estruturas de covariâncias em modelos mistos aplicados em análises de ensaios de múltiplos ambientes (MET), Kelly et al. (2007) questionam que: embora bem estudada, a técnica ainda era pouco aplicada nos programas de melhoramento da Austrália. No Brasil, o mesmo cenário é observado, ou seja, embora a abordagem já seja bem conhecida (REZENDE; THOMPSON, 2004), pouco uso tem-se observado, talvez por falta de softwares gratuitos ou até mesmo pela sua complexidade e alta demanda computacional.

Para tentar diminuir essas limitações presentes no algoritmo inicial proposto por Smith, Cullis e Thompson (2001) e Thompson et al. (2003) propuseram uma alteração do algoritmo AI (Average Information), no que eles chamaram de estimação de rank reduzido, com objetivo de evitar casos de Heyhood, comumente obtidos no trabalho original de Smith, Cullis e Thompson (2001); ou seja, quando as variâncias convergem para estimativas iguais a zero ou negativas em alguns casos. No entanto, percebe-se que a aplicação do método AI não é tão simples, tão pouco sua adaptação - como pode ser observado nos doze passos descritos no trabalho de Thompson et al. (2003) -, e, embora utilizassem a estrutura fator analítico, o modelo, ainda sim, é menos informativo do que um modelo baseado em matrizes não estruturadas (UN), que estima todas as variâncias e covariâncias residuais e genotípicas.

Os modelos FA buscam reduzir o número de parâmetros referentes às estimativas de covariâncias genotípicas UN, contudo, sua aplicação é complicada e não há garantias de convergência. Por outro

lado, sua versão mais completa (modelos UN) pode ser aplicada facilmente utilizando o algoritmo EM - que apresenta um comportamento mais estável, pois restringe as estimativas dos componentes no espaço do parâmetro. Além disso, esse tipo de estrutura é normalmente mais informativo do que suas versões com menos parâmetros, além de ser mais tratável em casos de matrizes de posto incompleto (THOMPSON et al., 2003).

Estruturas UN exigem a estimação de uma grande quantidade de parâmetros que crescem de forma abrupta quando se tem um grande número de locais. Também, as estimativas não são diretamente interpretáveis em termos de estabilidade e adaptabilidade.

Nesse contexto, diversos trabalhos foram realizados comparando essas duas estruturas mais tradicionais (FA(2) e UN) com várias outras, tais como (FA( $\neq 2$ ), simetria composta, autorregressiva de primeira ordem, estrutura diagonal, etc). Contudo, a recomendação de uma ou outra estrutura pode depender do tipo de dados que se tem em mãos (KELLY et al., 2007; MEYER, 2009; PIEPHO, 1998). Embora todas essas estruturas sejam válidas e a recomendação varie de acordo com o tipo de dados, a FA(2) tornou-se padrão quando se busca análise MET em modelos mistos variados (KELLY et al., 2007).

Uma alternativa à ideia de FA seria a aplicação de análise de PCA nas matrizes de  $G + GE$  em modelos mistos multivariados sob estruturas UN. Porém, as covariâncias representadas na decomposição de SVD não são estimativas REML, ou seja, as componentes de dispersão estimadas para decomposição de valor singular nos métodos AMMI e GGE convergem - exceto para as ocorrências de homogeneidade de variâncias

e dados balanceados -, na maioria dos casos, em quadrados médios de componentes viciados. Nesse sentido, o objetivo desse trabalho foi avaliar híbridos de um programa de melhoramento de milho com modelos multivariados mistos em estrutura UN construindo o biplot baseado na análise de fatores com estimativas REML de covariâncias e comparando esses resultados e interpretações com GGE biplot tradicional.

## **2 MATERIAL E MÉTODOS**

### **2.1 Material genético e locais**

Vinte e oito híbridos foram avaliados em 35 locais distribuídos em diversas regiões do Brasil (Tabela 1). Esses locais foram divididos em 4 épocas distintas. Duas safras em 2010 e 2011 e duas safrinhas 2011 e 2012. Dentro de cada época alguns híbridos foram avaliados, mas em outras épocas não, caracterizando assim um experimento desbalanceado (Figura 1). Esses locais foram distribuídos nas regiões Sul, Sudeste e Centro-Oeste do Brasil, onde está concentrada grande parte da produção nacional de milho. A semeadura foi mecanizada, em sistema de plantio direto sem revolvimento do solo e as datas de semeadura representam as épocas de plantio predominantes em cada região. O estande final desejado foi de 70.000 plantas por hectare nos cultivos de safra e de 60.000 plantas por hectare nos cultivos de safrinha. O controle de ervas daninha e pragas foi realizado por meio da aplicação de herbicida pré-emergente, pós-emergente e inseticidas para que a cultura de milho pudesse desenvolver sem interferência das mesmas.

Tabela 1 Distribuição de ensaios, nos anos agrícolas 2010/11 e 2011/12, Brasil

Época	Locais	Locais	UF	Época	Locais	Locais	UF
Safrinha 2010	Abelardo Luz	ABE	SC	Safrinha 2011	Abelardo Luz	ABE	SC
	Araguari	ARA	MG		Araguari	ARA	MG
	Boa Esperanca	BOA	MG		Castro	CAS	PR
	Carazinho	CAR	RS		Guarapuava	COL	PR
	Castro	CAS	PR		Costa Rica	COS	MS
	Guarapuava	COL	PR		Coxilha	COX	RS
	Coxilha	COX	RS		Madre de Deus	MAD	MG
	Madre de Deus	MAD	MG		Paraguacu	PRG	MG
	Maua Da Serra	MAU	PR		Ponta Grossa	PGR	PR
	Ponta Grossa	PGR	PR		Uberlandia	UBE	MG
Safrinha 2011	Rio Verde	RVE	PR	Vargem Grande	VGR	SP	
	Uberlandia	UBE	MG	Jataí	CAI	GO	
	Campo Novo	CNO	SC	Campo Novo	CNO	MT	
	Campo Verde	CVE	MT	Campo Verde	CVE	MT	
	Montividiu	MON	GO	Dourados	DOU	MS	
	Nova Mutum	NMU	MT	Rondonopolis	RDB	MT	
	Rondonopolis	RBD	MT	Sorriso	SOR	MT	
	Sorriso	SOR	MT				

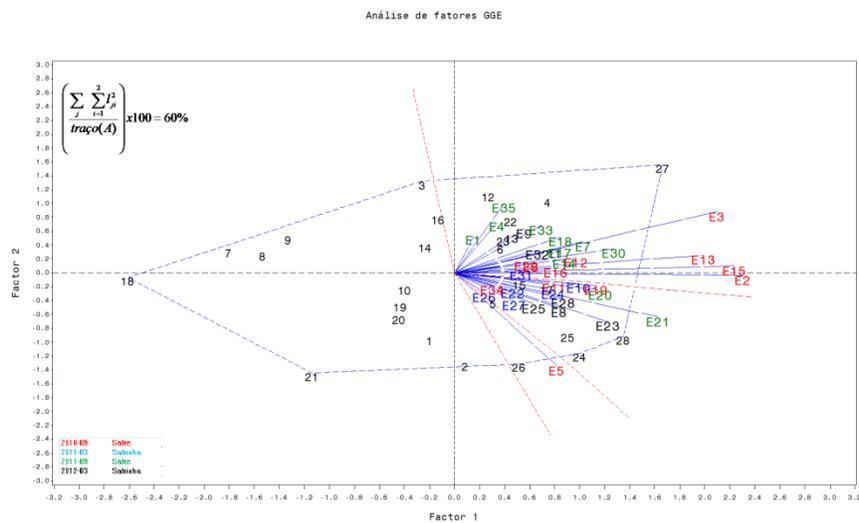


Figura 1 Cargas e escores fatoriais referentes a análise multiambiental considerando diferentes locais, épocas de cultivo e anos

Para implantação dos experimentos, utilizou-se o delineamento em blocos casualizados (DBC), com duas repetições. A unidade experimental foi constituída de quatro fileiras de cinco metros de comprimento, espaçadas de cinquenta centímetros, onde foi determinado o estande final de colheita. A umidade dos grãos foi determinada no momento da colheita e a produtividade corrigida a 14% .

## 2.2 Análise MET em modelos mistos com estruturas UN

A análise MET foi conduzida utilizando modelos mistos multivariados (Equação 1) considerando que cada local possui variâncias específicas e dados possivelmente correlacionados com os demais locais. Foi então assumido o modelo linear misto descrito por:

$$y = X\beta + Za + e \quad (1)$$

O vetor  $y$  corresponde às observações de parcelas em cada ambiente. A matriz  $X$  é a matriz de incidência dos efeitos fixos correspondentes aos blocos. A matriz  $Z$  (Equação 2) é a matriz de incidência dos efeitos aleatórios correspondentes aos híbridos dentro de locais. Os vetores  $\beta$  e  $a$  correspondem às estimativas dos efeitos fixos e aleatórios respectivamente e  $e$  é o erro aleatório. Essa representação pode ser expandida da seguinte forma:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_1 & O & \dots & \dots & \dots & O \\ O & X_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & O \\ O & \dots & \dots & \dots & O & X_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} Z_1 & O & \dots & \dots & \dots & O \\ O & Z_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & O \\ O & \dots & \dots & \dots & O & Z_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad (2)$$

Onde cada subscripto corresponde aos subvetores e submatrizes das observações e delineamento em cada local. Note que o comprimento das matrizes e vetores pode variar entre locais. Assumiu-se que o vetor  $y$  é multivariado normal com  $E(y) = X\beta$  e  $\text{cov}(y) = ZGZ^T + \Sigma$ , e  $G = I_t \otimes A$ , onde  $A$  é a matriz de (co) variâncias genóticas entre os híbridos nos diferentes locais e  $\Sigma = P(I_{lkr} \otimes V)P^T$ , onde  $V$  é a matriz de covariâncias residuais e  $P$  é a matriz que relaciona os resíduos entre os locais  $k$ , sendo identidade para o caso balanceado. Essa matriz garante que  $\Sigma$  seja uma matriz quadrática de mesma dimensão dos dados independente do grau de desbalanceamento.

Utilizando o algoritmo EM descrito por Dempster, Laird e Rubin (1977) a solução REML para os elementos das matrizes  $A$  e  $V$  pode ser dada por (Equação 3):

$$\tilde{\sigma}_{a_{ij}} = \left[ a_i^T a_j + \text{tr}(C_{ij}^{-1}) \right] / t \quad (3)$$

Com:

$$\tilde{\sigma}_{a_{ij}} = \begin{cases} \sigma_{a_k}^2 & \text{se } i = j \\ \sigma_{a_{ij}} & \text{se contrário} \end{cases}$$

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 (Equação 4):

$$\tilde{\sigma}_{e_{ij}} = \left\{ e_i^T e_j + tr \left( [WC^{-1}W^t]_{ij} \right) \right\} / n^* \quad (4)$$

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

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

Essa abordagem é típica de matrizes não estruturadas (UN), ou seja, todos os componentes da variância  $\{ \sigma_{e_k}^2 \sigma_{a_k}^2 \}$  e covariâncias  $\{ \sigma_{e_{ij}} \sigma_{a_{ij}} \}$  são estimados simultaneamente com os efeitos fixos e aleatórios (E-BLUPs).

### 2.3 Análise GGE e aproximação fatorial analítico

Uma vez que os valores genotípicos em cada local são estimados conjuntamente na análise MET (modelo 1), é evidente que seus efeitos estão confundidos com a interação GE. Uma forma de separá-los é realizando a decomposição de valor singular da matriz retangular ( $t \times k$ ) e utilizando componentes principais para sua representação. Nesse caso teríamos uma análise similar ao modelo SREG ou GGE biplot descrita em Yan et al. (2000).

Uma alternativa seria a aplicação direta de análise de fatores na matriz das estimativas REML das variâncias de ( $A$ ). Nesse caso teríamos que (Equações 5,6):

$$\hat{a} = Lf + \delta = (Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1}(y - X\beta) \quad (5)$$

Sendo

$$G = I \otimes [LL' + \Psi] \quad (6)$$

Onde  $f \sim N(0, I)$  e  $\delta \sim N(0, \Psi)$ , sendo  $f$  os escores fatoriais,  $\delta$  é o fator específico,  $L$  é a matriz de cargas fatoriais e  $\Psi$  a matriz das variâncias específicas.

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

$L'\Psi^{-1}L = \Delta$ , onde  $\Delta$  é uma matriz diagonal. A chave para estimação de  $L$  e  $\Psi$  é a obtenção da matriz  $\Psi^{-1/2}(A - \Psi)\Psi^{-1/2}$ . Assim,

utilizando o Algoritmo EM obtém-se que:  $\hat{L} = \Psi^{1/2} \hat{P}$ , sendo que  $\hat{P}$  provém da decomposição espectral de  $\Psi^{-1/2}(A - \Psi)\Psi^{-1/2}$  e  $\hat{\Psi} = \text{diag}(A - \hat{L}\hat{L}')$ , onde temos um processo iterativo que continua até a convergência das matrizes  $L$  e  $\Psi$ . Assim, as cargas fatoriais referentes a cada ambiente podem ser descritas por  $L$ . O próximo passo consiste na obtenção dos escores fatoriais referentes aos efeitos de  $\hat{a}$ . Utilizando a notação de Meyer (2009), podemos substituir a equação 5 dentro do modelo 1 e teremos (Equação 7):

$$y = X\beta + Z[(L \otimes I)f + \delta] + e \quad (7)$$

Logo, a distribuição condicional conjunta utilizada na solução EM - onde  $\hat{a}$  era tomado como informação perdida - pode ser dada por (Equação 8):

$$p(y, f, \delta) = p(f)p(\delta | \Psi)p(y | \theta, \Sigma) \quad (8)$$

Onde  $\theta = \{f, \delta, \beta\}$

A matriz das equações de modelos mistos com a solução para  $\theta$  pode então ser obtida por:

$$\begin{bmatrix} \beta \\ f \\ \delta \end{bmatrix} = \begin{bmatrix} X'\Sigma^{-1}X & X'\Sigma^{-1}W & X'\Sigma^{-1}Z \\ \Omega'\Sigma^{-1}X & \Omega'\Sigma^{-1}\Omega + I & \Omega'\Sigma^{-1}Z \\ Z'\Sigma^{-1}X & Z'\Sigma^{-1}\Omega & Z'\Sigma^{-1}Z + \Psi^{-1} \otimes I \end{bmatrix} \begin{bmatrix} X'\Sigma^{-1}y \\ \Omega'\Sigma^{-1}y \\ Z'\Sigma^{-1}y \end{bmatrix}$$

Onde os escores fatoriais (Equações 9,10) são estimados por:

$$\hat{f} = (\Omega' \Sigma^{-1} \Omega + I)^{-1} \Omega' \Sigma^{-1} (y - X \beta - Z \delta) = [L' (LL' + \Psi)^{-1} \otimes I] \hat{a} \quad (9)$$

$$\hat{\delta} = (Z' \Sigma^{-1} Z + I \otimes \Psi^{-1})^{-1} Z' \Sigma^{-1} (y - X \beta - \Omega f) = [\Psi (LL' + \Psi)^{-1} \otimes I] \hat{a} \quad (10)$$

Dado que  $\hat{a} = (j \otimes I_t) a_g + a_{ge}$ , envolve o efeito de genótipo mais o efeito de interações GE, o método tradicional desenvolvido por Smith trabalha com fatorial analítico em  $a_{ge}$  no denominado FAMM (REZENDE; THOMPSON, 2004). Contudo, a aplicação nesse estudo foi em  $\hat{a}$ . Sendo assim,  $a_g$  pode ser facilmente obtida marginalizando a matriz retangular dos  $\hat{a}$ 's; porém o resultado não marginalizado é uma análise similar ao GGE-biplot.

O ajuste da análise de fatores foi estimado pelo somatório das comunalidades (Equação 11) dada pela expressão:

$$Adj = \frac{\sum_j \sum_{i=1}^2 l_{ji}^2}{\text{traço}(A)} \quad (11)$$

Todas as análises foram realizadas por meio de um código elaborado sob a plataforma SAS-IML.

Dada à pressuposição de normalidade de  $f$  e  $\delta$ , a estimativa do erro padrão foi dada por (Equações 12,13):

$$Ep(\hat{f}) = \left[ \text{trace}(\Omega' \Sigma^{-1} \Omega + I)^{-1} \right]^{1/2} \quad (12)$$

$$Ep(\hat{\delta}) = \left[ \text{trace}(Z' \Sigma^{-1} Z + I \otimes \Psi^{-1})^{-1} \right]^{1/2} \quad (13)$$

Que corresponde à inversa da informação de Fisher na abordagem EM.

Com base no vetor  $\hat{f}$  e seu erro padrão as elipses de confiança foram construídas utilizando a library (car) no software R.

### 3 RESULTADOS

Dado a estrutura dos dados deste trabalho, a análise em modelos AMMI ou GGE biplot teria aplicação prejudicada uma vez que a matriz de genótipos em locais foi altamente desbalanceada (Tabela 2 e 3). Além disso, observou-se que a utilização da análise de variância seria prejudicada uma vez que seus pressupostos foram claramente violados, ou seja, não houve independência de resíduos e a heterogeneidade da variância foi extremamente elevada. Como exemplo, podemos citar a variância residual de Vitória da Colônia que foi de  $\sigma_e^2 = 0.096$ , ao passo que a variância de Costa Rica foi de  $\sigma_e^2 = 2.474$ . Além disso, os resíduos foram correlacionados com valor mínimo de -0.53 e valor máximo de 0.51. Esses resultados reforçam que a abordagem de ANAVA na ideia de modelos aditivos e multiplicativos seria prejudicada do ponto de vista da inferência.

Tabela 2 Valores genotípicos preditos para 28 híbridos avaliados em diferentes locais e épocas. As caselas em amarelo se referem os materiais não avaliados em determinado local/época

HÍBRIDO	SAFRA 2010												SAFRINHA2011					
	ABE	COX	CAR	CAS	RVE	MAU	ARA	BOA	MAD	UBE	PGR	COL	MON	CNO	CVE	NMU	SOR	RBR
1	13.1	11.1	10.5	6.9	10.8	5.9	10.3	11.8	15.3	13.9	7.2	6.8	9.6	9.2	10.8	5.2	9.0	9.8
2	14.7	12.9	13.1	6.4	13.0	7.0	13.1	13.7	13.6	16.6	6.8	6.5	11.7	8.4	11.1	4.8	10.1	9.4
3	13.0	14.1	10.6	7.3	11.0	6.6	12.9	14.3	13.5	16.1	7.3	6.2	9.0	6.9	7.1	3.7	9.3	10.4
4	15.7	16.7	7.6	7.3	11.9	7.6	13.1	11.8	12.8	14.4	7.5	6.8	13.4	8.6	10.1	5.4	8.6	10.8
5	15.2	14.3	12.7	7.5	10.9	6.7	12.5	14.6	14.3	16.7	6.8	6.3	10.5	10.6	10.6	6.2	10.2	9.9
6	14.7	13.8	6.5	7.8	11.3	6.9	13.1	14.6	13.3	16.3	6.9	6.0	10.9	9.1	9.4	5.7	8.8	9.9
7	9.6	11.6	9.2	6.4	9.4	5.5	9.0	10.4	12.4	13.5	6.4	6.3	8.1	6.5	6.7	4.3	9.2	9.4
8	9.9	11.6	11.6	6.6	9.7	6.2	11.5	11.1	13.1	14.6	6.9	6.1	7.5	7.9	7.6	4.6	8.8	8.0
9	11.0	11.5	5.2	6.7	11.0	5.6	9.5	10.5	13.7	14.6	5.9	5.8	8.8	7.9	8.2	5.2	8.4	8.6
10	16.8	14.2	14.5	6.8	10.5	6.3	10.0	13.4	13.9	16.7	7.4	7.0	8.7	7.9	8.4	4.5	9.4	10.0
11	16.0	15.9	15.1	7.9	11.8	7.0	11.7	14.5	15.1	16.4	8.0	7.4	10.6	10.3	10.8	6.0	9.9	9.5
12	14.1	16.1	12.4	7.2	11.1	7.5	13.8	13.1	16.1	16.5	7.5	6.0	9.6	9.0	10.2	4.0	8.8	10.6
13	16.6	16.7	8.5	7.6	12.9	7.6	12.9	14.6	16.4	15.0	8.5	7.0	10.7	8.0	10.8	5.3	9.4	10.2
14	13.3	14.3	12.5	7.9	10.7	6.1	11.4	13.9	14.2	13.5	6.9	6.9	8.0	8.9	9.2	5.2	9.0	9.1
15	14.2	13.6	7.8	7.7	12.3	6.4	12.6	15.5	12.6	13.5	6.2	6.4	10.6	8.9	9.6	4.8	9.5	9.2
16	14.5	15.9	16.9	7.9	11.4	6.5	12.5	12.5	13.5	15.7	8.0	6.6	9.3	8.8	9.0	4.4	8.6	11.4
17	16.3	17.2	16.4	8.3	11.1	7.2	12.7	13.8	14.9	16.4	7.5	7.9	9.8	8.5	8.3	5.6	10.1	11.7
18	3.4	3.8	4.7	5.7	9.7	2.5	2.9	3.8	12.9	12.2	5.8	5.8	9.0	7.9	7.9	4.7	7.6	9.8
19	14.0	13.6	12.1	7.3	11.0	5.8	10.4	13.9	15.3	14.7	6.7	6.7	8.8	8.4	9.9	5.2	9.4	9.6
20	14.4	13.4	12.0	6.7	11.9	6.5	11.3	11.3	14.3	15.0	6.7	6.3	9.5	9.1	9.9	5.2	9.1	10.2
21	10.9	10.9	12.9	6.3	10.8	5.2	11.2	10.4	12.1	12.9	6.1	6.2	10.0	8.7	8.9	5.1	9.5	9.1
22	10.6	13.3	10.6	7.9	12.1	6.2	13.7	14.1	15.4	16.7	7.4	5.9	11.0	8.5	9.0	4.2	8.8	10.5
23	12.8	13.3	9.7	7.3	10.8	6.7	11.8	14.1	15.4	15.7	7.2	6.5	10.5	9.3	9.7	5.1	10.4	10.3
24	13.9	13.2	12.4	8.3	13.0	6.9	13.5	15.3	14.8	17.7	8.4	6.9	12.0	9.5	10.0	5.8	10.3	11.0
25	13.3	12.5	14.7	8.1	13.6	7.0	13.8	15.3	15.6	19.8	7.4	7.1	11.1	8.4	9.7	5.3	10.4	10.9
26	17.2	15.1	13.2	7.8	11.2	6.6	11.7	16.2	14.4	13.9	7.5	7.7	9.8	8.9	9.3	5.7	10.9	10.1
27	17.2	18.5	5.0	9.0	13.2	9.0	16.2	18.7	16.2	16.3	8.9	6.9	11.6	8.1	10.2	4.8	9.6	10.7
28	15.7	15.6	14.9	7.5	11.7	8.4	15.3	14.3	17.1	18.5	8.1	7.1	11.5	9.9	11.1	6.1	10.8	11.2

Tabela 3 Valores genotípicos preditos para 28 híbridos avaliados em diferentes locais e épocas. As caselas em amarelo se referem os materiais não avaliados em determinado local/época

HÍBRIDO	SAFRA2011										SAFRINHA2012							
	ABE	COX	CAS	ARA	MAD	PAR	UBE	COS	PGR	VGR	COL	DOU	CAI	CNO	CVE	SOR	RBR	
1	10.9	8.9	15.0	12.9	11.3	12.3	12.9	14.2	15.1	9.5	14.1	10.7	7.3	10.1	11.3	9.9	8.7	
2	10.9	8.7	13.8	12.4	9.5	12.1	14.5	11.4	15.7	9.7	11.7	11.2	8.3	12.6	10.0	10.5	8.5	
3	12.0	9.2	14.5	13.3	10.8	11.9	15.0	9.5	15.0	9.8	14.6	9.3	8.7	9.4	10.0	8.4	9.3	
4	13.0	10.3	17.1	14.8	12.2	13.3	14.7	11.8	18.0	12.9	16.4	10.9	9.8	11.9	10.3	9.4	10.3	
5	11.1	8.6	15.3	14.1	12.6	12.5	12.7	12.7	14.3	12.1	14.1	10.7	10.2	10.0	10.6	9.2	9.5	
6	12.1	10.1	15.9	15.3	11.5	11.6	15.3	11.1	16.1	11.8	14.4	10.3	9.5	12.7	10.7	10.3	10.9	
7	11.7	8.9	13.3	11.1	10.4	10.9	12.4	6.7	12.6	9.3	13.7	8.1	7.9	7.4	8.9	7.3	8.3	
8	11.8	8.6	14.3	10.9	10.2	10.6	12.1	9.4	12.6	9.5	13.2	8.1	7.0	8.0	8.1	7.6	8.0	
9	12.7	9.8	14.6	11.9	9.1	10.7	13.2	8.4	13.5	10.2	14.4	8.5	8.6	10.4	7.7	10.0	8.7	
10	12.7	8.9	14.4	13.4	10.6	11.7	14.7	11.6	14.3	8.5	15.8	10.0	6.8	9.8	9.8	8.9	8.6	
11	13.3	9.8	17.3	13.4	12.4	13.9	14.8	13.3	15.5	11.4	16.8	10.6	8.8	12.1	9.2	9.6	10.0	
12	12.5	11.6	15.4	13.8	11.9	14.0	11.9	12.3	16.1	12.3	16.7	9.6	10.2	8.6	8.8	8.1	9.1	
13	12.3	9.7	15.4	12.8	9.3	12.5	12.9	12.2	15.0	11.6	15.6	9.7	7.1	9.7	9.5	10.6	10.1	
14	12.1	9.6	14.8	12.3	10.6	12.8	14.2	10.9	14.7	9.6	16.0	9.7	9.3	9.5	10.0	9.3	9.1	
15	11.2	8.9	15.5	13.3	12.3	14.1	16.3	10.8	16.7	10.4	15.1	11.2	11.8	11.7	11.0	10.5	10.2	
16	12.8	9.0	15.0	13.9	8.8	12.9	14.5	11.3	15.1	10.2	16.7	9.3	10.0	8.1	10.3	7.9	9.2	
17	13.1	10.1	15.6	13.7	10.1	12.0	17.2	12.9	16.3	9.0	16.7	11.0	8.1	13.2	10.4	9.9	9.3	
18	12.1	8.6	14.9	12.0	10.0	10.7	12.0	8.6	12.5	9.0	13.8	8.3	7.4	9.8	9.4	8.1	8.3	
19	11.6	10.1	13.2	13.0	10.7	11.8	12.4	10.9	14.1	10.1	14.7	10.2	7.3	10.1	10.5	9.6	8.7	
20	11.1	9.6	13.6	13.3	10.1	11.4	12.7	11.8	13.1	10.8	14.1	10.6	8.7	9.2	10.0	10.0	8.3	
21	10.1	6.8	12.4	12.0	8.6	10.6	12.2	10.7	13.9	9.5	13.2	10.2	6.9	10.2	9.9	7.3	8.6	
22	12.7	10.1	17.0	14.9	13.9	13.6	14.5	12.8	17.5	12.0	15.2	10.4	9.5	12.0	10.8	9.1	9.2	
23	11.5	10.0	16.2	13.5	13.1	13.9	13.8	12.3	15.8	11.6	16.1	10.5	9.4	12.0	9.5	8.1	11.1	
24	11.5	7.1	17.5	14.4	11.9	12.5	16.5	14.6	15.6	10.8	12.9	11.1	8.8	12.3	12.1	11.0	10.1	
25	12.3	9.0	16.5	13.6	10.4	11.6	19.0	13.6	16.4	8.2	11.0	11.5	8.7	15.8	11.5	13.4	8.2	
26	10.7	8.0	14.5	13.3	12.0	12.7	15.5	13.4	14.8	9.5	16.0	11.8	7.0	11.7	11.6	9.4	10.6	
27	12.8	11.3	18.5	14.8	13.6	14.9	16.6	12.9	18.0	13.0	16.4	10.6	10.7	11.7	10.9	11.6	12.0	
28	11.3	9.6	16.0	13.9	11.9	11.9	13.4	16.0	15.7	11.7	13.1	11.4	7.3	12.5	10.4	10.1	9.2	

As predições dos valores genotípicos dos híbridos não avaliados podem ser observadas na tabela 1. As caselas em amarelo correspondem à

predição dos genótipos não avaliados em determinado local/época. É possível observar que o híbrido 27 apresentou alto desempenho na safra 2011 e safrinha 2012. De maneira análoga, seu desempenho nos locais em que não fora avaliado também foi alto; exceto para ambiente Carazinho, onde esse híbrido foi predito com baixo desempenho. Esse resultado indica que esse ambiente destoa claramente dos demais. Como exemplo, podemos citar o híbrido 16 que foi vencedor nesse ambiente, apresentando um valor genotípico de  $16.914 \text{ ton}\cdot\text{ha}^{-1}$ , porém, na análise geral, esse genótipo demonstrou um desempenho bem próximo da média geral, ocupando a 16ª posição no ranking.

A correlação genética entre locais variou de -0.33 a 0.89, indicando a presença de interação genótipos por ambientes. Os locais com correlação mais negativa (-0.33) foi Abelardo (safra-2011) e Sorriso (safrinha -2011), ambos localizados em regiões geográficas bastante distintas (regiões sul e centro-oeste do Brasil). Por outro lado, os locais mais similares foram Mauá da Serra e Coxilha (Safra 2010) localizados em estados diferentes, porém na mesma região geográfica.

O resumo da matriz de covariâncias genéticas pode ser realizado por meio de um biplot de escores e cargas fatoriais. A figura 2 representa as cargas fatoriais de locais referentes à matriz *A* e os escores fatoriais obtidos via modelo fator analítico.

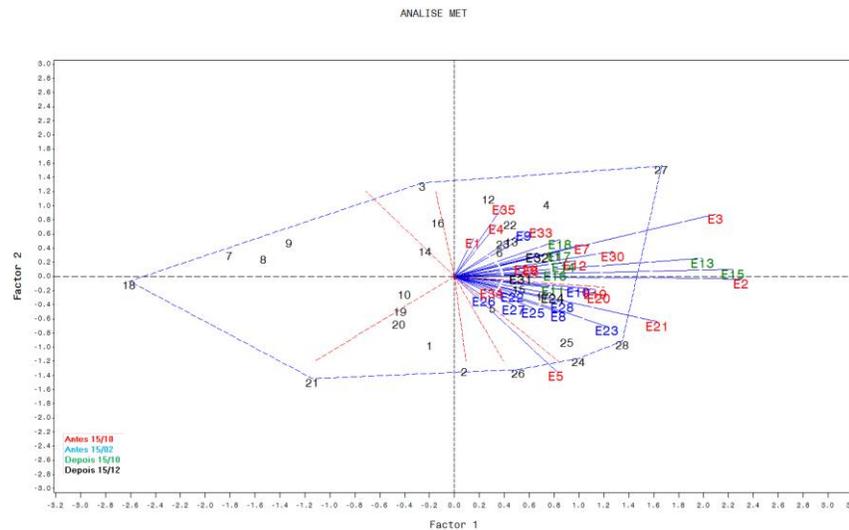


Figura 2 Cargas e escores fatoriais referentes a análise multiambientes considerando diferentes épocas de semeadura

Podemos observar claramente três possíveis mega-ambientes delimitados pelos locais 27,28 e 26. O biplot capturou aproximadamente 60% da variação de  $\hat{a} = (j \otimes I_t)a_g + a_{ge}$ . Com base nesse biplot é possível observar um padrão que dificilmente encontraríamos diretamente na matriz  $A$ , devido à sua dimensão de 35x35. Por exemplo, foi possível observar que locais dentro de um ano agrícola tendem a apresentar respostas mais similares do que aquelas obtidas em anos ou épocas diferentes. Nesse sentido, nota-se uma clara distinção dos locais que compõem o ano de 2011 (verde) em relação aos locais de 2010 (vermelho) - também em relação aos anos 2011 safrinha (azul) e 2012 safrinha (preto). Esse resultado sugere que mesmo avaliando um genótipo em um mesmo local, mas em anos diferentes, resultados bastante

divergentes podem ser encontrados. Esse efeito foi bem evidenciado nos locais 1 e 2 que pertencem ao mesmo ambiente Abelardo (ABE), porém apresentaram uma baixa repetibilidade de um ano para outro (Figura 1, Tabela 4).

Tabela 4 Correlação média entre e dentro de safras obtidas através de 35 locais

	Safra 2010	Safrinha2011	Safra 2011	Safrinha2012
Safra 2010	0.54	-	-	-
Safrinha2011	0.37	0.40	-	-
Safra 2011	0.36	0.26	0.45	-
Safrinha2012	0.36	0.38	0.35	0.41

Outro resultado interessante foi que nenhum híbrido obteve uma estabilidade geral, ou seja, nenhum híbrido ficou próximo à coordenada (FA(1),0), o que sugere a utilização de programas de melhoramento com características bem específicas para cultivo de safra e cultivo de safrinha. Isso pode ser confirmado nessa mesma figura, onde os locais de safra (vermelho e verde) ficaram mais distribuídos no segundo quadrante ao passo que os locais de safrinha (azul e preto) ficaram mais distribuídos no quarto quadrante.

Na figura 2 podemos observar a disposição dos locais com base na sua época de semeadura. Ao contrário da figura 1, pouco ou nenhum padrão pôde ser bem estabelecido. As duas épocas de quebra de safras foram determinadas utilizando o critério de semeadura antes ou depois de 15 de outubro para locais de safra (vermelho e verde), e antes ou depois de 15 de fevereiro para locais “safrinha” (preto e azul). A única separação

clara nesse biplot se refere aos locais “safra” e “safrinha”, porém dentro de cada uma dessas épocas de cultivo nenhum padrão claro foi observado; sugerindo que a interação dos genótipos com os ambientes deve-se muito mais aos anos ou safras do que épocas de semeadura dentro de safras.

A análise separada de locais de safra e safrinha ajudou a discriminar de forma mais pronunciada o efeito de ano agrícola. Na figura 3 temos a abordagem MET apenas para os locais de safra (2010 e 2011). Dado que o biplot nesse caso apresenta maior proporção de explicação, a separação entre a safra 2010 e 2011 ficou mais nítida. Novamente, observa-se que um mesmo local tem uma forte variação de resposta em diferentes ambientes e, neste caso, a estratificação geográfica ou mesmo a determinação de mega-ambientes com análise de um único ano pode ser enganosa para esse conjunto de dados.

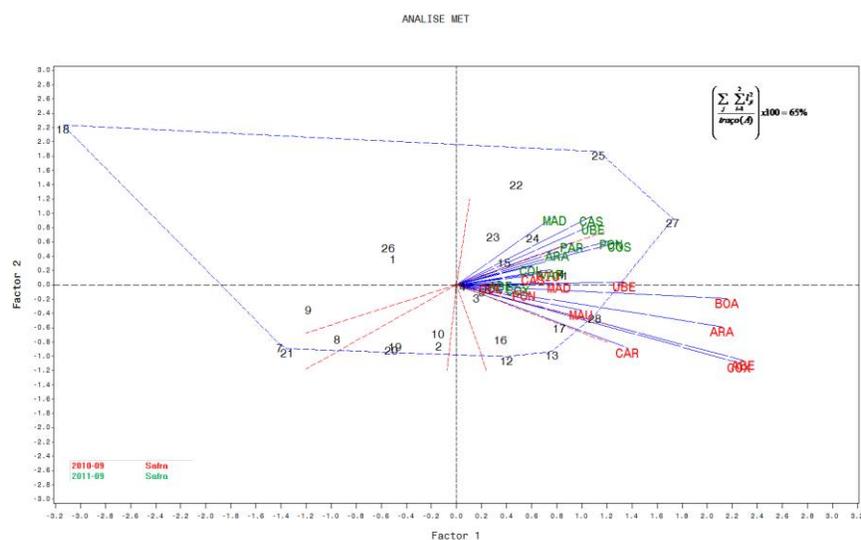


Figura 3 Cargas e escores fatoriais referentes a análise multiambientes dentro do cultivo de safras (2010 e 2011)

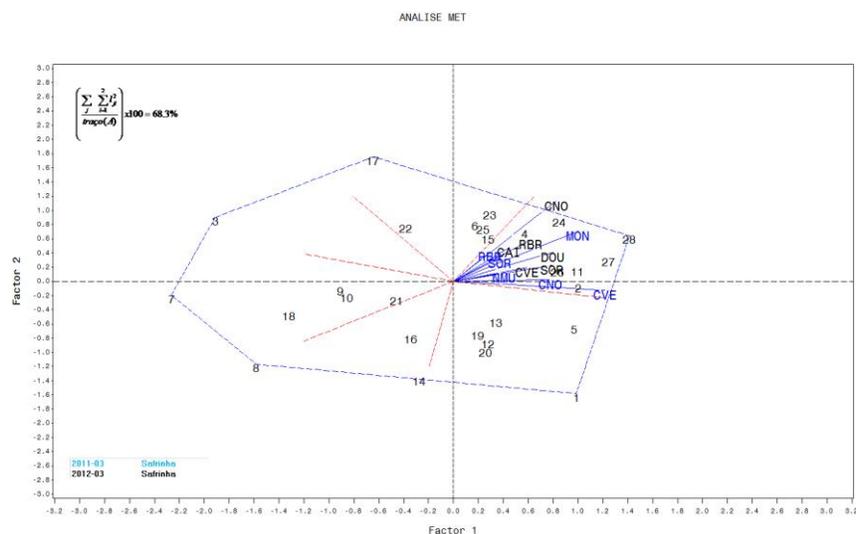


Figura 4 Cargas e escores fatoriais referentes a análise multiambientes dentro de época de cultivo safrinhas (2011 e 2012).

Resultado semelhante foi observado para as análises em safrinha, como pode ser verificado na figura 4 onde, dois grupos distintos foram observados, embora a figura utilizada para a separação desses grupos apresentou forte sobreposição, o que sugere que a correlação entre essas duas épocas são similares, ou seja, os ambientes de safrinha devem ter pouca variação ao longo dos anos (Tabela 4).

A decomposição da matriz correspondente a  $\hat{a} = (j \otimes I_t) a_g + a_{ge}$  em  $Lf + \delta$  produz, em tese, uma interpretação similar à análise GGE biplot, ou seja, o primeiro componente principal captura boa parte do efeito principal de genótipos e o segundo a interação GE. Isto pode ser comprovado na figura 5, onde a porção  $(j \otimes I_t) a_g$  foi praticamente igual ao primeiro escore fatorial da análise  $FA(2)$ . Podemos observar que a

escala de valores foi equivalente entre os E-BLUPS e o primeiro escore fatorial dado o ajuste da regressão que foi de praticamente 1 ( $r^2=0.989$ ). A comparação pode ser estendida quando aplicamos a análise de decomposição por valor singular na matriz de correspondente ao vetor  $\hat{a}$  (Figura 6). Nessa figura fica evidente a equivalência das análises GGE biplot e fatorial quanto à interpretação dos resultados. Pequenas distorções foram observadas principalmente pelo fato de a análise de componentes principais ser, em princípio, descritiva e não paramétrica, enquanto que a análise de fatores utiliza pressuposições de normalidade. Esse resultado pode ter contribuído para que as cargas e escores obtidos na análise MET fossem capazes de discriminar mais os ambientes e obter escala de valores mais próximos dos EBLUPS do que o primeiro componente principal do GGE biplot.

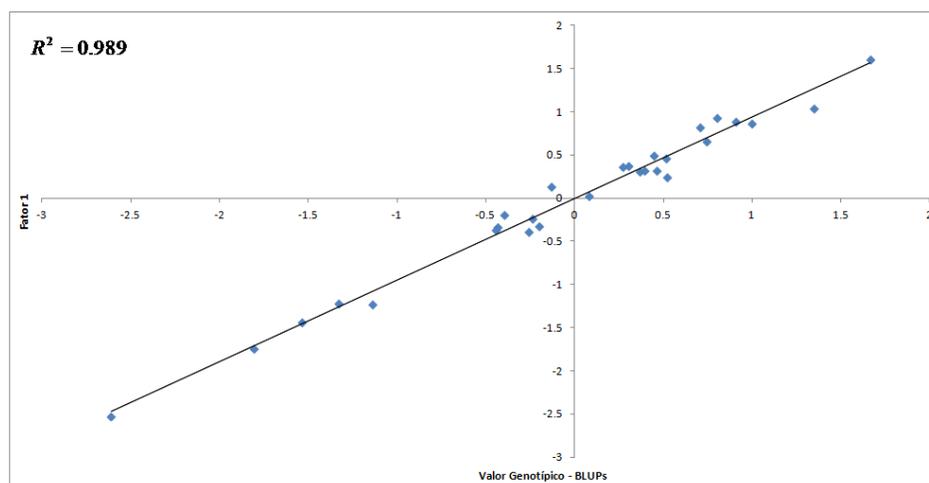


Figura 5 Ajuste do primeiro fator da análise FA(2) em relação aos E-BLUPS dos híbridos obtidos na análise multiambiental

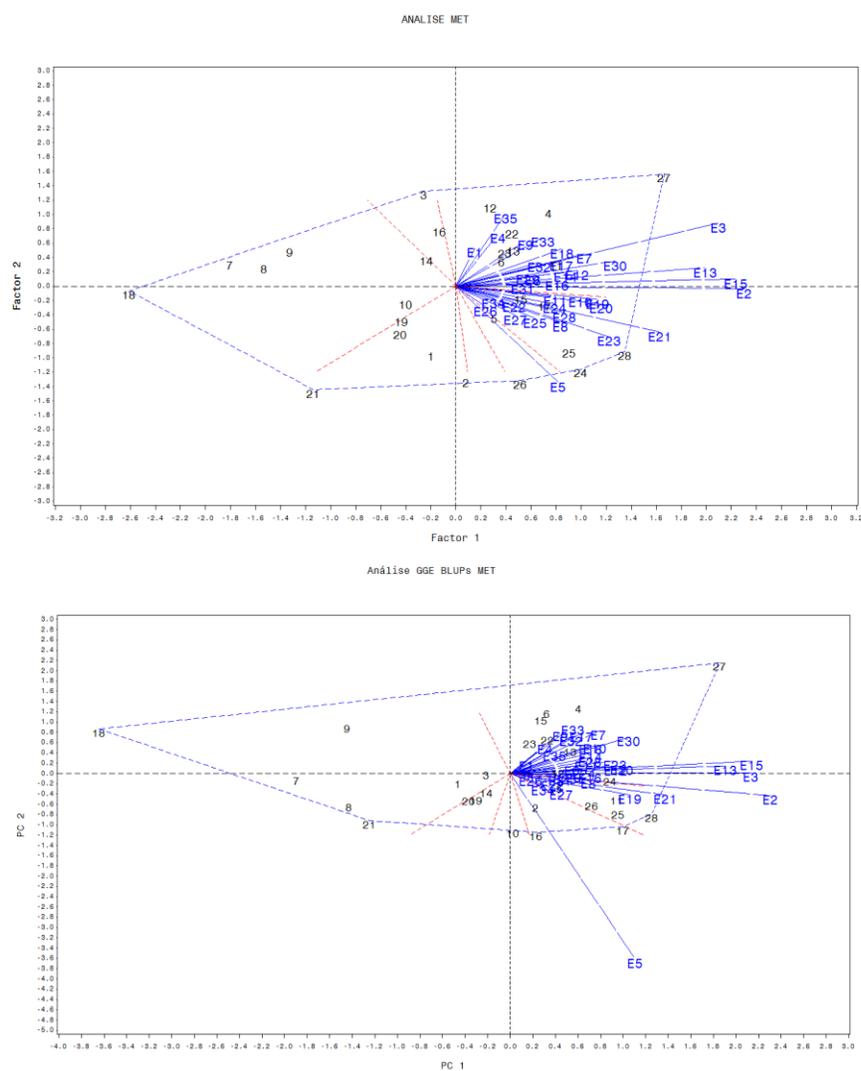


Figura 6 Comparação do biplot da análise de fatores da matriz de covariância genotípica de híbridos em relação à decomposição de valor singular da matriz retangular dos E-BLUPs de genótipos obtido em cada local

Outro resultado interessante foi a diminuição do efeito de escala nos efeitos de genótipos e ênfase nos locais no gráfico de fatores. Em

outras palavras, o biplot dos escores e cargas apresentaram menor dispersão nos genótipos e maior discriminação dos locais respectivamente, enquanto a análise de componentes principais apresentou o efeito contrário.

Dado que a comparação direta entre a análise multiambiental na presença de desbalanceamento e o GGE biplot clássico é inviável de ser realizada, aplicaram-se então a comparação de ambas as abordagens apenas nos locais balanceados (Safra 2011 e Safrinha 2012). Nesse cenário, a análise de fatores foi capaz de explicar maior proporção da interação GE, ou seja, 63.6% vs 55.4% (gráfico não apresentado). Esse resultado demonstra que mesmo em condições de balanceamento, a aplicação de modelos mistos multiambiental proporciona maior explicação da variância de interação além de fornecer uma interpretação similar ao GGE-Biplot.

Na figura 7 podemos observar as elipses de confiança ( $\alpha = 0.05$  e  $0.10$ ) para alguns escores fatoriais referentes a alguns genótipos representativos.

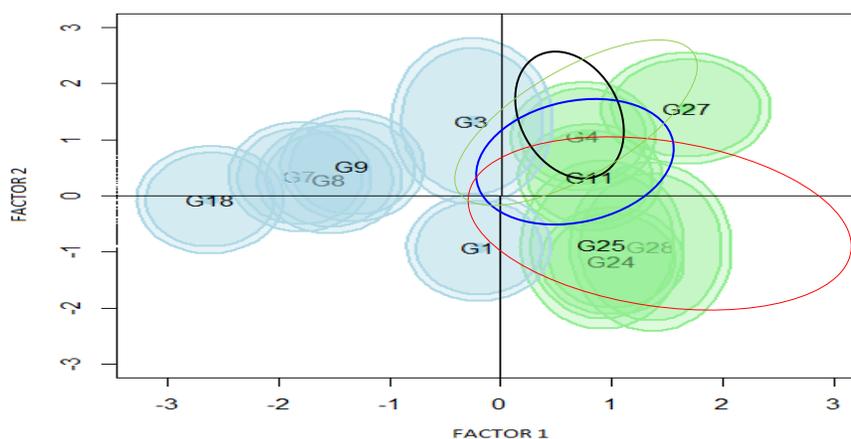


Figura 7 Elipses de confiança para os escores fatoriais de alguns genótipos considerando um intervalo de 90 e 95%

Nessa figura 7 observamos que não há sobreposição de intervalos entre os melhores genótipos e os de menor desempenho. Nota-se também que o genótipo 27 é não estável para ambientes safra e safrinha, pois sua elipse não abrangeu a origem referente ao eixo Fator 2. Outro resultado interessante foi que os híbridos não avaliados na safra 2011 e safrinha 2012 (3, 25 e 28) apresentaram uma elipse mais dispersa quando comparadas à dos híbridos não avaliados na safra 2010 e safrinha 2011 (híbrido 27, por exemplo). Esse resultado sugere que os experimentos avaliados nas duas primeiras safras foram menos precisos.

#### 4 DISCUSSÕES

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) 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.

Considerando a heterogeneidade das variâncias, nesse trabalho foi demonstrado que os resíduos foram bem diferentes em magnitude, moderadamente correlacionados e que os dados desbalanceados seguem a dinâmica de um programa de melhoramento de milho com descarte e inclusão de genótipos ao longo dos anos. Nesse sentido, a aplicação de métodos convencionais baseados em ANAVA seria dificultada para a quantificação da estabilidade e adaptabilidade considerando a relação dessa descrição AMMI ou GGE com a soma de quadrados ou quadrados médios. Além disto, o padrão da decomposição de GE em contrastes de médias seria influenciado pela covariância residual entre locais, apresentando, dessa forma, maior erro de estimação. Nesse sentido, as abordagens AMMI e GGE seriam muito mais descritivas de GE ou G+GE do que informativas sobre a variância desses efeitos (ou quadrado médio de GE ou G+GE para um modelo fixo).

A analogia entre o modelo fator analítico e os métodos baseados em ANAVA como AMMI e GGE-biplot já foram reportadas por outros autores (BURGUEÑO et al., 2008; REZENDE; THOMPSON, 2004; SMITH; THOMPSON; CULLIS, 2001). Crossa et al. (2013) demonstraram uma clara associação entre modelos SREG e análise de fatores (CROSSA; VARGAS; JOSHI, 2010). Nosso trabalho reforça de

maneira empírica essa associação, que pode ser observada nas figuras 2, 5 e 6. Sendo assim, podemos estender todas as interpretações do método GGE em nossos biplots. Nesse sentido, a análise é útil não apenas no julgamento de estabilidade e adaptabilidade, mas tem interpretação direta quanto à identificação de mega-ambientes, genótipo ideal e estratificação ambiental.

Nossos resultados demonstraram que em ambientes tropicais e subtropicais que representam boa parte do território brasileiro, um mesmo local dificilmente pode ser considerado como o mesmo ambiente ao longo dos anos. Nesse sentido, foi mais fácil encontrar associação positiva entre locais diferentes dentro de um mesmo ano agrícola do que associação positiva de um mesmo local ao longo dos anos (Figuras 1 e 3 e Tabela 2 e 3). Nesse sentido, a estratificação ambiental em dados de apenas um ano agrícola pode ser enganosa. Assim, a determinação de mega-ambientes deve ser obtida somente com dados de vários anos; o que é inviável de se alcançar de forma clara nos métodos convencionais já que diferentes genótipos são avaliados ao longo dos anos. Nesse trabalho, observamos que a análise MET, via modelos mistos, pode ser uma poderosa ferramenta nesse sentido.

Análises do tipo MET, via modelos mistos, têm sido amplamente difundidas graças aos trabalhos iniciais de Piepho (1998) e Smith, Cullis e Thompson (2001). Segundo Burgueño et al. (2011), o método FA apresenta alta capacidade preditiva em experimentos com desbalanceamento simulados. Embora não aplicamos validação cruzada no conjunto de dados, os resultados de predição dos híbridos desbalanceados foram coerentes. Por exemplo, o híbrido 27 apresentou

ótimo desempenho nos locais de Safra e foi considerado o campeão naquele mega-ambiente, mesmo que tenha sido avaliado apenas em um dos dois anos. Situação semelhante foi observada no híbrido 28 que foi o vencedor na safrinha no ano em que foi avaliado e também no ano não avaliado. Novos estudos estão sendo realizados utilizando outro conjunto de dados com diferentes níveis de desbalanceamentos e os resultados preliminares demonstram que desempenho do método é satisfatório com correlações medianas ( $r \Rightarrow 0.6$ ) mesmo apresentando perdas de mais de 50% de parcelas em cada ambiente (Nuvunga Master degree thesis - 2013- dados ainda não publicados).

A aplicação do método MET tem se restringido a estruturas disponíveis em alguns softwares tais como o ASREML. A opção, nesse estudo, pela estrutura UN deve-se principalmente pela lógica de que as demais estruturas tentam apenas reproduzi-la - mesmo que de maneira parcimoniosa - e que as vantagens e desvantagens da aplicação de estruturas conhecidas dependem do conjunto de dados em mãos (KELLY et al., 2007; PIEPHO, 1998). Nesse sentido, devido ao pequeno número de genótipo e locais, optamos por trabalhar diretamente com matrizes completas de covariâncias. Porém, em alguns casos, pode-se ter que buscar estruturas mais parcimoniosas como a abordagem FA(2) disponível no software ASREML (GILMOR et al., 2005). No entanto, vale destacar que segundo Kelly et al. (2007), a abordagem de matrizes não estruturadas ainda não foi implementada de forma eficaz nesse programa.

Outro resultado interessante foi a exploração da pressuposição de normalidade na estimação do erro das cargas fatoriais. Essa abordagem

permitiu a construção de elipses de confiança no biplot. Devido a isto, podemos imaginar que essa ferramenta pode se tornar padrão dado à discussão em torno da incerteza contida nos padrões do biplots que são altamente descritivos e passíveis de críticas quanto à inferência ou em relação aos métodos não paramétricos utilizados para acessá-la (YAN, 2010; YANG et al., 2009).

Embora com um número limitado de genótipos e locais, nesse trabalho observou-se que é possível realizar análise de estabilidade e adaptabilidade em experimentos desbalanceados que apresentem uma estrutura dinâmica de descarte e inclusão de genótipos ao longo dos anos. Essa abordagem permitiu verificar determinadas tendências de um programa de melhoramento, comparando diretamente híbridos desenvolvidos para ambientes de safra e safrinha. A interpretação dos biplot é direta e intuitiva e com as mesmas propriedades da análise GGE biplot ou AMMI se for o caso. Assim, a aplicação dessa técnica permite que o melhorista tenha uma visão mais geral do seu programa melhoramento avaliando não apenas o progresso genético, estabilidade, adaptabilidade, mas também a própria ideia de estratificação ambiental.

**Application of mixed models in the evaluation of stability and adaptability  
in corn using unbalanced data: a GGE biplot interpretation in analytical  
factorial models**

**ABSTRACT**

The objective of this work was to treat the multi-environment approach (MET) under the point of view of mixed models associated with the factor analysis to evaluate the stability and adaptability in hybrids. In order to do this, 28 hybrids were evaluated in 35 locations distributed in four periods/years (2010 Harvest, 2011 Interim-harvest, 2011 Harvest and 2012 Interim-harvest). Some of these hybrids were evaluated in the first periods and not evaluated in those subsequent periods or vice-versa; thus, the set of data used in this work simulates the dynamics of a breeding program with discarding and inclusion of genotypes over the years. On the score and factorial loads biplot, it was observed that the locations were more similar inside periods than in between periods, suggesting that a same location may behave in a very divergent manner from year to year. This fact questions the elaboration of mega-environments in data from a single year. It was empirically observed that the analytical factorial models present direct interpretation with the GGE biplot analysis, given that the first factorial score presented a perfect adjustment ( $r^2 = 0.99$ ) with the genotype E-BLUPs. Given the normality presupposition for the factorial scores, it was possible to construct confidence ellipses and posterior direct comparison of genotypes in the biplot. It was observed that it was possible to perform stability and adaptability analysis in unbalanced experiments with discarding and inclusion of genotypes over the years. This approach allowed the verification of certain tendencies of a breeding program, directly comparing hybrids developed for harvest and interim-harvest. The interpretation of the biplot is direct and intuitive, and presents the same properties of the GGE biplot analysis or the AMMI, if that is the case.

Keywords: SREG. Confidence ellipse. MET. Unstructured matrixes (UN).

## REFERÊNCIAS

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

BURGUEÑO, J. et al. Using factor analytic models for joining environments and genotypes without crossover genotype  $\times$  environment interaction. **Crop Science**, Madison, v. 48, n. 4, p. 1291-1305, July 2008.

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

CROSSA, J. et al. Evaluation and interpretation of interactions. **Agronomy Journal**, Madison, v. 105, n. 5, p. 1-12, May 2013.

CROSSA, J. et al. Modeling genotype  $\times$  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, July 2006.

CROSSA, J.; VARGAS, M.; JOSHI, A. K. Linear, bilinear, and linear-bilinear fixed and mixed models for analyzing genotype  $\times$  environment interaction in plant breeding and agronomy. **Canadian Journal and Plant Science**, Ottawa, v. 90, n. 5, p. 561-574, 2010.

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

GAUCH, H. G.; PIEPHO, H. P.; ANNICCHIARICO, P. Statistical analysis of yield trials by AMMI and GGE: further considerations. **Crop Science**, Madison, v. 48, n. 3, p. 866-889, Mar. 2008.

GAUCH, H. G.; ZOBEL, R. W. Identifying mega-environments and targeting genotypes. **Crop Science**, Madison, v. 37, p. 311-326, Mar. 1997.

GILMOUR, A. R. et al. **ASReml user guide**. Release 2.0. Hemel Hempstead: HSN, 2005. Disponível em:  
<<https://www.vsni.co.uk/resources/documentation/asreml-user-guide>>.  
Acesso em: 10 nov. 2013.

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.

MEYER, K. Factor-analytic models for genotype  $\times$  environment type problems and structured covariance matrices. **Genetics Selection Evolution**, Les Ulis, v. 41, n. 21, p. 1-11, 2009.

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.

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.

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. 6, p. 449-462, Dec. 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. 4, p. 1138-1147, Dec. 2001.

THOMPSON, R. et al. A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. **Biometrics**, Washington, v. 45, n. 4, p. 445-459, 2003.

YAN, W. Comment on Biplot analysis of genotype  $\times$  environment interaction: proceed with caution. **Crop Science**, Madison, v. 50, n. 4, p. 1121-1123, July 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, May/June 2000.

YAN, W. et al. GGE biplot vs. AMMI analysis of genotype-by-environment data. **Crop Science**, Madison, v. 47, n. 2, p. 643-655, Mar. 2007.

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