

COMPARAÇÃO ENTRE OS MODELOS AMMI E GGE BIPLLOT PARA OS DADOS DE ENSAIOS MULTI-AMBIENTAIS

Kuang HONGYU^{1,2}
Fabiane de Lima SILVA²
Anderson Castro Soares de OLIVEIRA²
Danilo Augusto SARTI³
Lúcio Borges de ARAÚJO⁴
Carlos Tadeu dos Santos DIAS⁵

RESUMO: Interação genótipo por ambiente é uma questão extremamente importante no melhoramento genético de plantas e produção. A seleção e recomendação de genótipos superiores são dificultadas devido à constante ocorrência de interação, representa um grande desafio para os pesquisadores. Com isso, os ensaios multi-ambientais são essenciais para a seleção de linhagem eficaz e recomendação de cultivares. Existem vários modelos na literatura para análise de dados multi-ambientais, os modelos AMMI (*Additive Main effects and Multiplicative Interaction*) e GGE (*Genotype main effects + Genotype environment interaction*) biplot são os mais utilizados. O modelo AMMI combina a análise de variância e a análise de componentes principais, para ajustar, respectivamente, os efeitos principais de genótipo e ambiente e os efeitos da interação. O GGE biplot é um método baseado na análise de componentes principais para explorar os ensaios multi-ambientais e é permitido as visualizações dos gráficos de biplot as relações entre os ambientes de teste, genótipos e interação genótipo por ambiente. Os principais objetivos deste trabalho com a utilização dos modelos AMMI e GGE biplot: 1) investigação mega-ambiente para a compreensão do ambiente de destino; 2) avaliação de genótipos e ambientes dentro de cada mega-ambiente; 3) compreender as causas da interação genótipo por ambiente; e 4) Criar um novo método para a comparação dos modelos AMMI com GGE biplot.

- **PALAVRAS-CHAVE:** Interação genótipo por ambiente; modelo AMMI; GGE biplot; mega-ambiente; análise de componentes principais.

¹ Universidade de São Paulo - USP, Programa de Pós-graduação em Estatística e Experimentação Agronômica, Escola Superior de Agricultura "Luiz de Queiroz", ESALQ/USP, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: kuang_hongyu@hotmail.com

² Universidade Federal de Mato Grosso - UFMT, Departamento de Estatística. CEP: 78060-900, Cuiabá, MT, Brasil. E-mail: fabianezte@yahoo.com.br

³ Universidade de São Paulo - USP, Programa de Pós-graduação em Economia, Escola Superior de Agricultura "Luiz de Queiroz", ESALQ/USP, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: andersonco@gmail.com

⁴ Universidade Federal de Uberlândia - UFU, Faculdade de Matemática, CEP: 38408-100, Uberlândia, MG, Brasil. E-mail: araujob@gmail.com

⁵ Universidade de São Paulo - USP, Departamento de Ciências Exatas, CEP 13418-900, Piracicaba, SP, Brasil. E-mail: ctsdias@usp.br

1 Introdução

A resposta diferencial de genótipos em ambientes é frequente em estudos de experimentos multi-ambientes (MET) e é conhecido como interação genótipo por ambiente ($G \times E$), que reduz a correlação entre os valores fenotípicos e genotípicos e dificulta a seleção e recomendação de genótipos adaptados e estáveis (Camargo-Buitrago et al., 2011; Gauch, 2013). Dados de METs são conduzidos por vários anos para os principais produtos agrícolas no mundo, frequentemente resumidos em tabelas de duas entradas com genótipos nas linhas e ambientes (ou locais) nas colunas. Os METs são essenciais porque a presença da $G \times E$, ou seja, a mudança na performance relativa de genótipos por meio de diferentes ambientes, complica a avaliação de cultivar. Se não existisse a $G \times E$, um único cultivar prevaleceria no mundo inteiro e um único experimento bastaria para avaliação de cultivar (Gauch & Zobel, 1996; Rodrigues et al., 2014). A $G \times E$ ocorre em várias formas, com a forma mais extrema que consiste em interações cruzada, na classificação de genótipos mudanças entre ambientes, por exemplo, um genótipo responde de maneira diferente às variações ambientais e nesta situação, os melhores genótipos em um local podem não o ser em necessariamente os melhores em outros locais.

A $G \times E$ tem sido um foco de pesquisa entre biometristas e geneticistas quantitativos desde o início de 1900 (Yan & Kang, 2003). Com a noção de que a $G \times E$ é indesejável e/ou que confunde avaliação de genótipos, muito trabalho tem sido dedicada ao desenvolvimento de índices de estabilidade para quantificar e selecionar contra $G \times E$. Diversos métodos estatísticos destinados à avaliação da $G \times E$ estão disponíveis no sentido de entender melhor este efeito e a escolha do método mais adequado depende dos dados experimentais, os estudos de $G \times E$ vêm se destacando e ganhando grande aplicabilidade nas duas últimas décadas (Hongyu et al., 2014; Arciniegas-Alarcón et al., 2014).

No melhoramento de plantas, os principais objetivos de ensaios multi-ambientes (METs) são: (i) estudo de $G \times E$, (ii) avaliar a adaptabilidade e estabilidade genotípica, (iii) estabelecer relações entre os testes de ambientes, entre os genótipos, e entre genótipos e ambientes (ou locais) simultaneamente, e (iv) fazer previsões do valor da reprodução de genótipos que vai permitir fazer uma seleção precisa dos país para o próximo ciclo de reprodução. A presença da $G \times E$ complica este processo e normalmente é expresso como respostas inconsistentes de alguns genótipos em relação a outros, devido à mudança de classificação genotípica, ou como mudanças nas diferenças absolutas entre genótipos sem mudança classificação (Araújo et al., 2012; Gauch, 2013).

Há diversas teorias para a avaliação da adaptabilidade e estabilidade, as quais diferem quanto aos seus conceitos e procedimentos biométricos de estimação (Camargo-Buitrago et al., 2011; Silva & Benin, 2012). Dentre as metodologias mais recentes, têm-se O modelo AMMI (Gauch, 1992) é um método estatístico para compreender a estrutura de interações entre genótipos e ambientes e o modelo GGE biplot, proposto por Yan et al. (2000), que considera o efeito principal de genótipo mais a interação genótipo e ambiente. Ambas as análises, baseadas em gráficos biplot, representam uma matriz de dados.

A única diferença entre os modelos AMMI e GGE está na etapa inicial da análise, o modelo GGE analisa diretamente o efeito de $G + G \times E$, enquanto que a AMMI separa G da $G \times E$ e, na etapa final da análise com formação dos gráficos biplot para as duas

metodologias. No entanto, essa separação não é capaz de conferir superioridade à análise AMMI (Gauch, 2006; Yan et al., 2007). Comparando as metodologias AMMI e GGE Biplot na formação de mega-ambientes, Camargo-Buitrago et al. (2011) relataram 95,2% de coincidência nos resultados entre as metodologias. Assim, o modelo GGE biplot é mais indicado para a identificação de mega-ambientes, seleção de ambientes representativos e discriminativos e indicação de cultivares mais adaptadas e estáveis a ambientes específicos (Gauch et al., 2008; Yan, 2011). Por outro lado, a análise AMMI pode ser utilizada com eficiência na identificação de condições ambientais superiores para a exploração agrícola (seleção de locais de cultivo) e genótipos de superior desempenho médio (Gauch et al., 2008; Yan, 2011).

O objetivo deste trabalho foi utilizar a técnica da análise de GGE biplot para os dados multi-ambientes (MET): 1) investigação mega-ambiente para a compreensão do ambiente de destino; 2) avaliação de genótipos e ambientes dentro de cada mega-ambiente; 3) compreender as causas da interação genótipo por ambiente; e 4) Criar um novo método para a comparação dos modelos AMMI com GGE biplot.

2 Material e métodos

2.1 Biplot para tabela de dupla entrada

Foi utilizado neste trabalho o conjunto de dados MET sobre avaliação de produtividade de milho provenientes da empresa Criagene SK que fica no município de Casa Branca no estado de São Paulo. Os dados são relativos a experimentos com 15 genótipos de milho, em 4 locais distribuídos em (A1: Vazante 720m/MG; A2: município de Casa Branca 680m/SP; A3: Guaíra 520m/SP e A4: Vazante 650m/MG). Os ensaios foram conduzidos na safrinha de 2011. Têm-se ainda que em cada experimento foi utilizado um delineamento em blocos ao acaso, com 2 blocos em cada experimento.

O método biplot foi desenvolvido por Gabriel (1971) para representar graficamente resultados de análise de componentes principais ou de decomposição em valores singulares (DVS), no qual o valor de cada elemento de uma tabela de dupla entrada pode ser visualizado pelo produto de vetores e pelo cosseno do ângulo entre dois vetores (Yan & Kang, 2003). Quando duas matrizes apresentarem o mesmo número de linhas e colunas, é possível multiplicá-las. A nova matriz gerada da multiplicação das duas anteriores assume o mesmo número de linhas e colunas de ambas as matrizes.

Qualquer matriz de dados de duas entradas Z , com elementos z_{ij} , em que $i = 1, \dots, g$ linhas (genótipos) e $j = 1, \dots, e$ colunas (ambientes), que pode ser decomposto por DVS em p componentes principais (PC):

$$z_{ij} = \sum_{k=1}^p \lambda_k \alpha_{ik} \gamma_{jk} + \varepsilon_{ij}, \quad (1)$$

em que $p \leq \min(e, g-1)$. Todo PC é composto pela matriz de escores genotípicos α_{ik} , matriz de escores ambientais γ_{jk} , valor singular λ_k e ε_{ij} resíduo para o genótipo i no ambiente j que não é explicada pelo modelo. O modelo com algumas restrições: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ e com ortonormalidade nas escores α_{ik} , isto é, $\sum_{k=1}^g \alpha_{ik} \alpha_{ik'} = 1$ (se $k=k'$) e $\sum_{k=1}^g \alpha_{ik} \alpha_{ik'} = 0$ (se $k \neq k'$); com as restrições semelhantes para γ_{jk} (Yan, 2011).

Quando a matriz Z de posto p pode ser suficientemente aproximada por uma matriz de posto 2 (os primeiros dois componentes são os mais importantes, pois explicam maiores variações dos dados), isto é,

$$z_{ij} = \lambda_1 \alpha_{i1} \gamma_{j1} + \lambda_2 \alpha_{i2} \gamma_{j2} + \varepsilon_{ij}, \quad (2)$$

Pode ser apresentada graficamente em um biplot de dimensão 2 após um particionamento em valor singular apropriado.

$$z_{ij} = (\lambda_1^f \alpha_{i1})(\lambda_1^{1-f} \gamma_{j1}) + (\lambda_2^f \alpha_{i2})(\lambda_2^{1-f} \gamma_{j2}) + \varepsilon_{ij}, \quad (3)$$

em que $f = [0; 0,5; 1]$ é o fator de partição em valor singular (PVS).

O biplot é construído por meio da representação gráfica $\lambda_1^f \alpha_{i1}$ como abcissa, $\lambda_2^f \alpha_{i2}$ como ordenada para cada genótipo, e ao mesmo tempo traçando $\lambda_1^{1-f} \gamma_{j1}$ como abcissa e $\lambda_2^{1-f} \gamma_{j2}$ como ordenada para cada ambiente (Yan & Kang, 2003; Yan, 2011).

O expoente f é usado para redimensionar os escores de linhas e colunas para melhorar a interpretação visual do biplot para um propósito particular. No contexto de dados MET, os valores singulares são alocados inteiramente nos escores de genótipo (linha) se $f=1$, isto é “particionamento em valor singular centrada-genótipo” ou $PVS = 1$ (Yan, 2002), ou inteiramente aos escores de ambiente (coluna) se $f = 0$ (“particionamento em valor singular centrada-ambiente” ou $PVS = 2$); e $f = 0,5$ alocará as raízes quadradas dos valores singulares λ_k para ambos os escores de genótipo e ambiente (“particionamento em valor singular simétrico” ou $PVS = 3$).

Na análise biplot GGE, o partição em valor singular de genótipo-centrada e o ambiente-centrada são utilizados para avaliação de genótipos e ambiente de teste, respectivamente (Yan, 2011). Uma propriedade importante do biplot é que a aproximação de qualquer elemento da matriz original Z de posto 2 pode ser estimada visualmente por produto interno de vetores de genótipo e ambiente, respectivamente, e o cosseno do ângulo entre si. Isto é conhecido como a propriedade do produto interno do biplot (Yan & Holland, 2010; Yan, 2011).

2.2 Construção de GGE biplot

O modelo GGE biplot (Yan & Kang, 2003; Yan, 2011), que considera o efeito principal de genótipo mais a $G \times E$, são baseadas em gráficos biplot, representa graficamente uma matriz de dados. Este Biplot é construído nos dois primeiros componentes principais de uma Análise de Componentes Principais (ACP) utilizando Modelos de Regressões Locais (SREG). A primeira componente, quando se encontra altamente correlacionada com o efeito principal do genótipo, representa a proporção do rendimento que se deve somente às características do genótipo. A segunda componente representa a parte do rendimento devida a $G \times E$ (Yan & Holland, 2010; Yan, 2011).

Quando diferentes cultivares estão adaptadas as diferentes grupos de ambientes e a variação entre grupos é maior do que dentro do grupo, tem-se a formação de um mega-ambiente (Yan & Kang, 2003). A definição de mega-ambientes e a relação entre os ambientes auxiliam os melhoristas de plantas na identificação de genótipos que possuam adaptação ampla ou específica a determinados ambientes ou grupos de ambientes (Silva & Benin, 2012). Na análise GGE biplot, quando se estuda mega-ambientes, a média no gráfico não está relacionada à média geral, mas sim à média do mega-ambiente e esta

abordagem auxilia na identificação de genótipos que possuam adaptação ampla ou específica a determinados ambientes ou grupos de ambientes (Yan & Kang, 2003; Yan & Tinker, 2006).

Para um conjunto de dados MET, cada valor na tabela é a produtividade média de um genótipo em um ambiente (y_{ij}), que é a soma da média geral (μ), o efeito principal do genótipo (G_i) para o ambiente particular (E_j), e a interação específica (G x E) entre o genótipo e o ambiente (ϕ_{ij}), ignorando quaisquer erros aleatórios (Yan, 2011):

$$y_{ij} = \mu + G_i + E_j + \phi_{ij}, \quad (4)$$

O modelo GGE biplot não separa os efeitos do genótipo e da G x E, mantendo-os juntos em dois termos multiplicativos, que podem ser visualizado na equação (4) da metodologia SREG. Desde que apenas o G e G x E são pertinentes à avaliação genótipo, avaliação ambiente de teste, e delineamento de mega-ambiente; o efeito principal do ambiente E e a média geral devem ser removida de cada elemento para apenas manter G e G x E na tabela de dupla entrada (Yan, 2011):

$$y_{ij} - \mu - E_j = G_i + \phi_{ij}, \quad (5)$$

Os dados MET ambiente-centrado, após dimensionamento apropriado dos dados, são submetidas à DVS e análise de biplot (Yan et al., 2000; Yan, 2011).

Yan & Tinker (2006) propôs uma “relação de informação (IR)” para avaliar a adequação de um biplot em exibir os padrões de uma tabela de dupla entrada. Suponha que esta tabela de dupla entrada tem g genótipos e ambientes. O número máximo de PCs é requerido para representar completamente esta tabela é $k = \min(e, g-1)$. Se não há correlação entre os ambientes, todos os k PCs devem ser completamente independentes e a proporção da variação total explicada por cada PC deve ser exatamente $1/k$.

Quando existir alguma correlação entre os ambientes, a proporção da variação explicada pelos os primeiros PCs deve ser maior do que $1/k$, e a variação explicada por outros PCs deve ser inferior ou igual a $1/k$ (Yan & Tinker, 2006; Yan, 2011). A IR pode ser calculado para cada PC, que é a proporção da variação total explicada por cada PC multiplicado por k. A interpretação é a seguinte: um PC com $IR \geq 1$ contém padrões (associações entre ambientes), e um PC com $IR < 1$ não contém qualquer padrão ou informação. O biplot de dimensão 2 representa adequadamente os padrões nos dados, se apenas os dois primeiros PCs tem um $IR \geq 1$ (Yan & Tinker, 2006). Todas as análises deste trabalho foram feitas por meio de rotinas computacionais implementadas no software R (R Development Core Team, 2014).

2.3 Mega ambiente

Quando diferentes cultivares estão adaptadas as diferentes grupos de ambientes e a variação entre grupos é maior do que dentro do grupo, tem-se a formação de um “mega-ambiente” (Gauch & Zobel, 1997; Yan & Kang, 2003). A definição de mega-ambientes e a relação entre os ambientes auxiliam aos melhoristas de plantas na identificação de genótipos que possuam adaptação ampla ou específica a determinados ambientes ou grupos de ambientes (Silva & Benin, 2012). Na análise GGE biplot, quando se estuda mega-ambientes, a média no gráfico não está relacionada à média geral, mas sim à média

do mega-ambiente e esta abordagem auxilia na identificação de genótipos que possuam adaptação ampla ou específica a determinados ambientes ou grupos de ambientes (Yan & Kang, 2003; Yan & Tinker, 2006).

O objetivo da análise de mega-ambiente é tentar dividir uma região de corte alvo, em sub-regiões significativas de modo que $G \times E$ possa ser explorada (Yan & Kang, 2003; Yan, 2011). Quando um GGE biplot de dimensão 2 é julgado como uma aproximação suficiente dos dados (também é chamado como “Which-won-where”) é uma ferramenta eficaz para análise de mega-ambiente.

Este ponto de vista é constituído por um polígono irregular e um conjunto de linhas retas (iguais números de lados de polígono) que irradiam a partir da origem do biplot e intersectam cada um dos lados do polígono perpendicularmente. Os vértices do polígono são os marcadores de genótipos localizados mais longe da origem do biplot em todas as direções, de modo a que todos os genótipos estão contidos dentro do polígono (Yan, 2011). Uma linha que cruza perpendicularmente um lado do polígono representa ambientes hipotéticos, em que os dois genótipos que definem esse lado do polígono apresentaram produtividades boas; a posição relativa dos dois genótipos seria invertida em ambientes em lados opostos da linha ($G \times E$ cruzado). Assim, as linhas que se irradiam dividem o biplot em setores, para cada setor, existe um genótipo no vértice, que apresentou melhor desempenho para ambientes que se enquadram nesse setor (Yan & Kang, 2003; Yan, 2011).

2.4 Modelo AMMI

O modelo AMMI combina dois métodos na sua análise: análise de variância e a decomposição por valor singular em um único modelo, componentes aditivos para os efeitos principais de genótipos (g_i), ambientes (e_j) e componentes multiplicativos para os efeitos da interação $(ge)_{ij}$ para a resposta média do i -ésimo genótipo no j -ésimo ambiente sobre r blocos (repetições) foi adotado o modelo conforme a equação (Gauch, 1992; Dias & Krzanowski, 2006; Hongyu et al., 2014):

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

em que: Y_{ij} é a média do i -ésimo genótipo no j -ésimo ambiente, com $i=1, 2, \dots, g$ e $j=1, 2, \dots, e$; μ é a média geral; g_i e e_j são os efeitos do i -ésimo genótipo e j -ésimo ambiente, respectivamente; λ_k é o k -ésimo valor singular da matriz (GE) e com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$; γ_{ik} e α_{jk} são elementos dos k -ésimos valores singulares correspondentes ao i -ésimo genótipo e j -ésimo ambiente respectivamente; ρ_{ij} é o resíduo da interação $G \times E$; n é o número de eixos ou componentes principais retidos pelo modelo; r é o número de blocos e ε_{ij} é o erro médio experimental associado ao i -ésimo genótipo no j -ésimo ambiente, assumidos independentes, e $\varepsilon_{ij} \sim N(0, \frac{\sigma^2}{r})$.

Na primeira fase os efeitos principais, na parte aditiva (média geral, efeitos de genótipos e ambientes), são ajustados por uma análise de variância comum aplicada à matriz de média ($Y_{(g \times e)}$), resultando em um resíduo de não aditividade, isto é, na interação $G \times E$, dada por $(ge)_{ij}$, essa interação constitui a parte multiplicativa do modelo. Na segunda fase a interação é analisada pela decomposição por valores singulares (DVS) ou por análise de componentes principais da matriz de interações ($GE(g \times e) = [(ge)_{ij}]$) (Dias & Krzanowski, 2006).

A matriz GE é a matriz de interação entre os genótipos e os ambientes, (matriz de resíduo dos efeitos principais), em que cada elemento $(ge)_{ij}$ de **GE** são dados por (Hongyu et al., 2014):

$$(ge)_{ij} = Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_{..}$$

em que Y_{ij} é a média das repetições do genótipo i no ambiente j , com $i = 1, 2, \dots, g$ e $j = 1, 2, \dots, e$; \bar{Y}_i é a média do genótipo i ; \bar{Y}_j é a média do ambiente j e $\bar{Y}_{..}$ é a média geral do experimento.

Existem várias técnicas para atribuir os graus de liberdade a um modelo AMMI, um dos procedimentos usuais consiste em determinar os grau de liberdade associados a cada parcela da $SQ_{G \times E}$, ou seja, associada a λ_k^2 , relacionada a cada membro da família de modelos AMMI, obtém-se o quadrado médio (*QM*) correspondente a cada parcela (ou modelo), em seguida, é obtido um teste F avaliando-se a significância de cada componente em relação ao $QM_{\text{Erro médio}}$. Isso resulta num quadro de análise de variância semelhante ao tradicional, com desdobramento para fonte de variação da interação $G \times E$ (Gauch, 2013).

Dessa forma, a definição do número de eixos a serem retidos para explicar o padrão relacionado à interação leva em consideração a proporção da $SQ_{G \times E}$ acumulada até o n -ésimo eixo ($\sum_{k=1}^n \lambda_k^2 / SQ_{G \times E}$). O ponto de parada que determina a seleção do modelo (AMMI0, AMMI1, ..., ou AMMIF) baseia-se na significância do teste F para os sucessivos termos da interação. O método de Gollob (1968) é um dos mais utilizados para atribuir graus de liberdades a um modelo AMMI, a expressão do método é: $GL_{PC_k} = g + e - 1 - 2k$, com $k = 1, 2, \dots, p$, $p = \min(g-1, e-1)$, em que PC_k : o k -ésimo eixo de componente principal.

Na análise AMMI, a DVS é aplicada à matriz de interações GE (matriz de posto conhecido p) (Hongyu et al., 2014):

$$\widehat{GE}_{(g \times e)} = \sum_{k=1}^n \lambda_k \gamma_k \alpha'_k = \tilde{U} \tilde{S} \tilde{V}'$$

em que, $\tilde{U}_{(g \times n)}$ tem em suas colunas apenas os n primeiros vetores $\gamma_{(g \times 1)}$; $\tilde{S}_{(n)}$ é a matriz diagonal com os primeiros valores singulares: $\lambda_1, \lambda_2, \dots, \lambda_p$ (com $n < p$) e $\tilde{V}'_{(n \times e)}$ tem em suas primeiras linhas os n primeiros vetores $\alpha'_{(1 \times e)}$.

Assim, na análise AMMI tem-se a decomposição exata dada por $GE_{(g \times e)} = USV' = U \tilde{S} \tilde{V}' = GH'$ e a decomposição aproximada por n componentes $\widehat{GE}_{(g \times e)} = \tilde{U} \tilde{S} \tilde{V}' = \tilde{G} \tilde{H}' = \sum_{k=1}^n (\lambda_k^2 \gamma_k^2) (\gamma_k^2 \alpha'_k)$ sendo $\tilde{G}_{(g \times e)}$ efeitos de genótipos e $\tilde{H}'_{(n \times e)}$ efeitos de locais (Gauch, 1992; Gauch, 2013).

3 Resultados e discussão

3.1 GGE biplot para análise de dados MET

De acordo com relação de informação (IR) dos quatro componentes (Tabela 1), apenas os dois primeiros PCs contêm padrões ($IR_1 = 1,57 > 1$ e $IR_2 = 1,00 = 1$). Portanto, o biplot é considerado adequado para representar os padrões dos dados. O GGE biplot com

base neste conjunto de dados é apresentado na Figura 1, na abscissa do biplot apresentam-se os escores de PC1 e na ordenada os escores do PC2, dos genótipos e ambientes.

Tabela 1 - Valor singular, proporção explicada e relação da informação (IR) dos quatro componentes principais (PCs)

PC	Valor singular	Varição Explicado (%)	IR
1	20,99	39,35	1,57
2	16,60	24,61	1,00
3	15,63	21,82	0,87
4	12,62	14,22	0,57

Os 15 genótipos são rotulados como G1 a G15 e os quatro ambientes como A1 a A4. A linha reta traçada da origem do biplot para a colocação de um ambiente ou genótipo é chamado “vetor de ambiente” ou “vetor de genótipo”, de modo que a partir desses vetores, as interações específicas entre um genótipo e um ambiente (por exemplo, o desempenho de cada um dos genótipos em cada ambiente) possam ser visualizadas. A regra de interpretação é: 1) o desempenho de um genótipo de um ambiente é melhor do que a média, se o ângulo entre o seu e do ambiente é $< 90^\circ$; é pior do que a média se o ângulo $> 90^\circ$ (porque o cosseno de um ângulo obtuso é menor do que 0); e que está próximo da média se o ângulo é aproximadamente 90° (Yan & Tinker, 2006; Yan, 2011).

3.2 Análise mega-ambiente

A Figura 1 permite o agrupamento visual dos ambientes de teste com base em cruzamento $G \times E$ entre os melhores genótipos. O biplot foi construído no ambiente centrado (centralização = 2), (escala = 0) e (PVS = 2). Os vértices do polígono são formados pelos genótipos: G8, G9, G11, G4 e G2. Os quatros ambientes foram cortados em 3 grupos pelas linhas que saíram da origem do biplot, os grupos são formados por (i) A1; (ii) A4; (iii) A2 e A3.

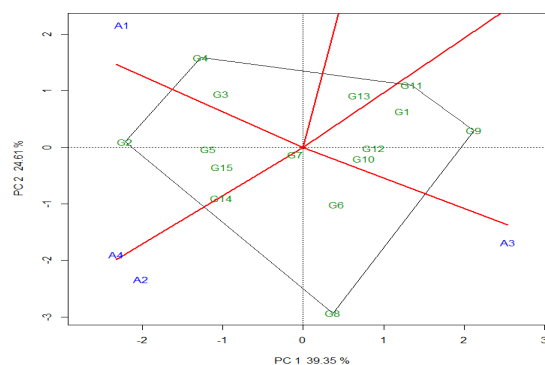


Figura 1 - O GGE Biplot (“Which-won-where”) para os dados de produtividade de milho (kg/ha), quais dos genótipos apresentaram melhor desempenho em que ambientes.

O genótipo G2 é o vértice do setor em que o ambiente A4 é colocado, portanto, é o genótipo que teve melhor desempenho neste ambiente; o G8 é o vértice no setor em que A2 e A3 são colocados, logo, é o genótipo mais adaptado nestes ambientes e é o mesmo caso do genótipo G4 no ambiente A1 (Figura 1). Nos setores de G9, G1, G11, G12 G10 e G13 que não contêm ambientes, significa que estes genótipos não eram produtivos em nenhum ambiente, ou seja, estes genótipos são os piores genótipos em relação à produtividade em alguns ou em todos os ambientes.

3.3 Avaliação de genótipos baseada em GGE biplot

Um “ideótipo” é uma forma ideal de planta ou genótipo para determinado ambiente e objetivo de cultivo, portanto, o ideótipo é um genótipo que apresenta alto desempenho médio e alta estabilidade por meio de um mega-ambiente. A visualização do GGE biplot “Média versus Estabilidade” é uma ferramenta eficaz para a avaliação de genótipos em ambos os aspectos (Yan et al., 2007; Yan, 2011). Na Figura 2, o pequeno círculo representa o “ambiente-média” e é definido pelas coordenadas médias de todos os ambientes de teste no biplot.

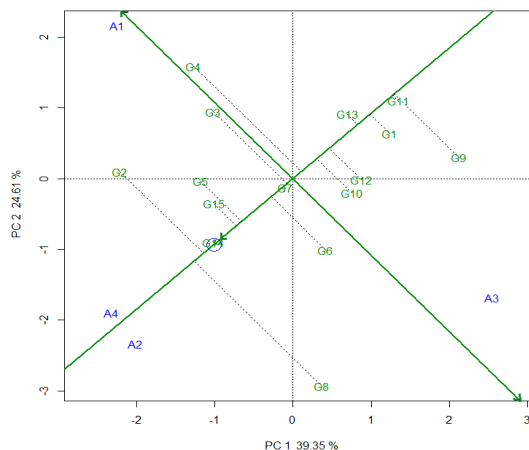


Figura 2 - O GGE Biplot (“Média versus Estabilidade”) com eixo do ambiente-média (EAM) para mostrar o desempenho médio e estabilidade dos genótipos.

A linha reta com uma única seta que passa pela origem do biplot e do ambiente-média é referido como o “eixo do ambiente-média” ou EAM. A seta aponta para um maior desempenho médio para os genótipos. A linha com duas setas que passa pela origem do biplot e é perpendicular ao EAM, com as setas apontam para a maior variabilidade de desempenho (menor estabilidade) em ambas as direções. Este biplot é baseado em (PVS = 1), ou seja, os valores singulares são totalmente compartimentados para as escores genotípicos (Yan, 2002; Yan, 2011).

Assim, os genótipos são classificados de acordo com sua produtividade média da seguinte forma: G8 > G2 > G14 > G15 > G5 >... > G3 > média geral > G4 > G10 >... > G1 > G11 > G9; O G8 foi altamente instável, pois tinha rendimento inferior ao esperado em ambiente A1, enquanto produziram relativamente bem em A2, A3 e A4; O G2 não foi estável, mas apresentou um desempenho bem superior em comparação aos outros genótipos em relação à produtividade, produzindo bem nos ambientes A1, A2 e A4, exceto no A3. O G7 foi estável e com produtividade pouco acima da média, já o genótipo G14 teve produtividade bem acima da média (terceiro melhor) e foi o mais estável (Figura 2). Note-se que, se o biplot explica apenas uma proporção da variação total, alguns genótipos aparentemente estáveis podem não ser verdadeiramente estáveis como as suas variações não podem ser completamente explicadas neste biplot.

Um ideótipo na Figura 2 pode ser um ponto (centro dos círculos concêntricos) sobre o EAM, no sentido positivo e tem um comprimento do vetor igual aos vetores mais longos dos genótipos sobre o lado positivo do EAM, isto é, mais alto desempenho médio (Figura 3). Portanto, os genótipos localizados mais perto do centro dos círculos concêntricos são mais desejáveis do que outros. Assim, G14 foi o ideótipo neste conjunto de dados, em seguida, os melhores foram os genótipos G15 e G5. Embora G8 e G2 tiveram rendimento médio superior, não foram estáveis, de acordo com Figura 4, o G12 foi mais desejável que o G8; os piores genótipos para recomendação foram G11 e G9. Os genótipos G13 e G11 são altamente estáveis (Figura 2 e 3), isso não significa que estes genótipos tiveram rendimentos bons, significa apenas que o desempenho relativo do G13 e G11 foram consistentes, mas estão longe de ser um genótipos ideais.

A Figura 3 ilustra um conceito importante em relação à estabilidade, o termo “estabilidade elevada” só tem sentido quando associado ao desempenho médio; o genótipo estável é desejado apenas quando apresenta alta performance média (Yan & Tinker, 2006; Yan, 2011).

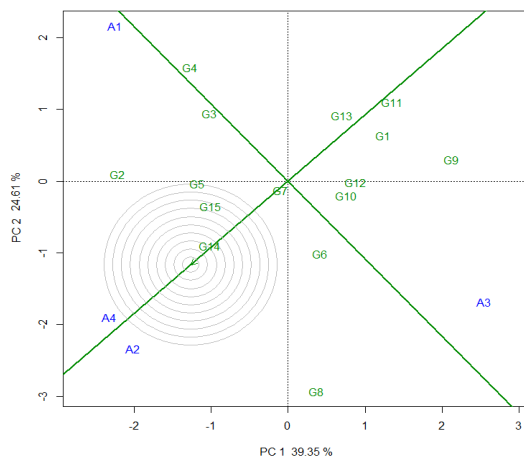


Figura 3 - O GGE Biplot com eixo do ambiente-média (EAM) para classificar os genótipos em relação ao ideótipo (no centro dos círculos concêntricos).

3.4 Avaliação do ambiente de teste

O objetivo da avaliação do ambiente de teste é identificar ambientes que podem ser utilizados para selecionar genótipos superiores de forma eficaz para um mega-ambiente. A seleção de um ambiente de teste deve apresentar maior discriminação dos genótipos e representatividade. A visualização no GGE biplot (Figura 4) foi projetada para esta finalidade.

Quando o biplot GGE baseia-se fora de escala (não padronizado), dados centrados no ambiente (Escala = 0), o comprimento do vetor de um ambiente é proporcional ao desvio padrão das médias de genótipos (SD), que é igual à raiz quadrada da variância fenotípica (σ_p) em ambiente de teste, o qual pode ser utilizado como uma medida do poder de discriminação do ambiente (YAN, 2011). Os ambientes de teste com vetores mais longos são mais discriminantes em relação aos genótipos. Aqueles ambientes com um vetor curto são menos discriminantes, o que significa que todos os genótipos tendem a executar de forma semelhante e pouca ou nenhuma informação sobre as diferenças genotípicas podem ser reveladas em tal ambiente, portanto, não devem ser utilizados como ambientes de teste. Um curto vetor também pode significar que o ambiente não é bem representada por PC1 e PC2 se o biplot não exibir adequadamente o G + GE dos dados. Todos os ambientes (A1, A2, A3 e A4) apresentaram vetores longos, significando que são mais discriminantes em relação aos genótipos (Figura 4).

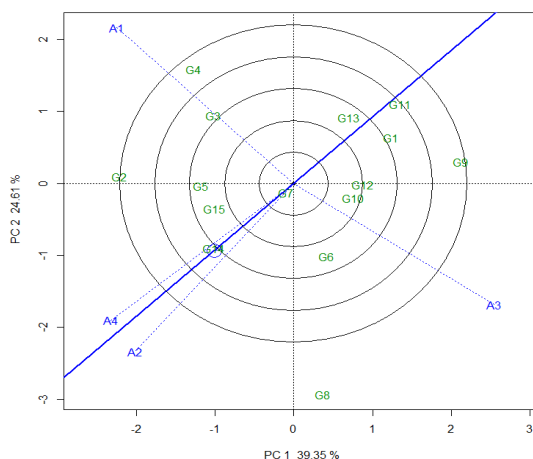


Figura 4 - O GGE biplot “discriminação e representatividade” para mostrar a capacidade discriminação e de representatividade dos ambientes de teste.

Uma segunda utilidade da Figura 4 é indicar ambientes de representatividade do ambiente teste. Os ambientes de teste que têm pequenos ângulos com EAM (eixo do ambiente-média), por exemplo, A4 e A2, são ambientes mais representativos do que aqueles que têm ângulos maiores, por exemplo, A1 e A3. Portanto, os ambientes de teste que são discriminantes e representativos (por exemplo, A4 e A2) são ambientes de teste ideal para a seleção de genótipos adaptados. Os ambientes de teste discriminante, mas não

representativos (por exemplo, A1 e A3) são úteis para a seleção de genótipos adaptados especificamente em mega-ambientes; ou para seleção de genótipos instáveis se o ambiente teste é um único mega-ambiente (Figura 4).

3.5 Análise AMMI para análise de dados MET

Os resultados da análise de variância conjunta relativa à produtividade de milho (*kg/ha*) mostraram que o efeito de genótipo, ambiente e o efeito da interação foram significativos ($p < 0,01$). Esses resultados indicam que os genótipos apresentaram comportamento diferenciado nos ambientes, o que justifica estudo mais aprofundado sobre o comportamento dos genótipos para identificar suas magnitudes de interação com os ambientes. O coeficiente de variação (CV%) do experimento mostrou-se de baixa magnitude (8,7199%), permitindo inferir que os experimentos foram bem conduzidos (Tabela 2).

Tabela 2 - Análise de variância conjunta dos experimentos para a produtividade de milho e decomposição das somas de quadrados da interação ($G \times E$)

Fonte de Variação	GL	SQ	QM	F	Valor-p
Blocos/Ambiente	4	1555492	388873	0,4721	0,756
Ambiente(E)	3	304154907	101384969	260,7149	4,846e-05***
Genótipo(G)	14	124525440	8894674	10,7984	2,712e-11***
Interação(G×E)	42	109478996	2606643	3,1645	3,370e-05***
PC1	16	68311642	4269478	5,18	0,0000***
PC2	14	26963397	1925957	2,32	0,0126*
PC3	12	14203957	1183663	1,44	0,1758 ^{NS}
Resíduo	56	46127517	823706	-	-
Total	300	695321348	-	-	-
Média Geral	10408,13				
CV(%)	8,7199				

NS: Não significativo; **Significativo a 0,01 de probabilidade; *** Significativo a 0,001 de probabilidade e * Significativo a 0,05 de probabilidade.

Na análise AMMI a soma de quadrados da interação ($SQ_{G \times E} = 109478996$) corresponde exatamente à soma dos autovalores ($\sum_{k=1}^n \lambda_k^2$) (Tabela 2), que pode estar inflacionado devido à presença de ruídos (variação não explicável) na variável resposta, em virtude disto, faz-se necessário realizar um ajuste da interação por decomposição em valores singulares, aplicada à matriz de interação GE , ou seja, pode ser decomposta em três componentes principais. Pela Tabela 2, o primeiro componente principal (PC1) foi significativo com ($p < 0,01$) pelo teste F de Gollob (1968), explicou 62,4% (68311642/109478996%) da variação da $SQ_{G \times E}$ e com 16 graus de liberdade. O PC2 foi significativo com ($p < 0,05$), explicou 24,6% (26963397/109478996%) da variação da $SQ_{G \times E}$ e com 14 graus de liberdade.

A não significância para o PC3 é desprezível e contém apenas ruído (variação aleatória não relacionada com o fenômeno da interação), que pode diminuir a eficiência da interpretação da estabilidade dos genótipos e ambientes na análise gráfica. Assim, a interpretação gráfica, considerando apenas as variações contidas nos dois eixos das componentes principais para avaliar a estabilidade dos genótipos e ambientes.

Os genótipos G2, G15, G14, G5, G3, G7 e G14 apresentaram produtividades acima da média em ordem decrescente e os mais estáveis foram G2, G9, G5, e G3, estes podem ser cultivados em todos os locais onde foram feitos os estudos (Figura 5). Entre estes, apenas os genótipos G2, G5 e G3 apresentaram produtividade acima da média geral e estão entre os dois grupos (produtivos e estáveis), indicando que esses genótipos associaram adaptabilidade com estabilidade. O G2 apresentou o melhor desempenho entre todos os genótipos, em relação de estabilidade e adaptabilidade (Figura 5 e Figura 6), portanto, pode ser considerado como ideótipo (maior Estabilidade e produtividade) e escolhido para recomendação pelos pesquisadores.

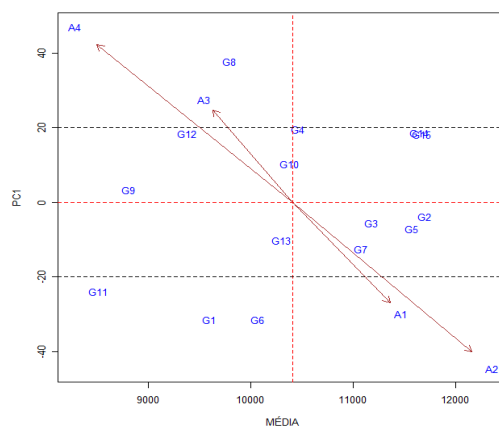


Figura 5 - Biplot AMMI1 para os dados de produtividade de milho (*kg/ha*) com quinze genótipos (G) e quatro ambientes (A).

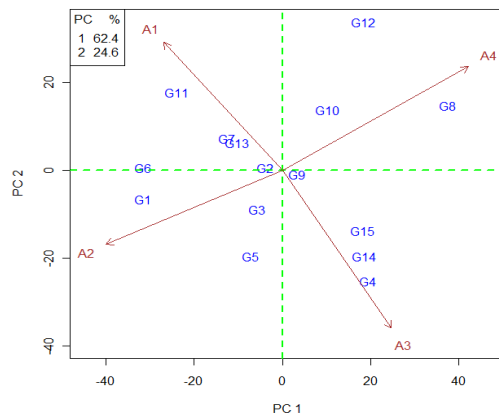


Figura 6 - Biplot para o modelo AMMI2 para os dados de produtividade de milho (*kg/ha*) para os quinze genótipos (G) e quatro ambientes (A).

Os ambientes mais favoráveis para a produtividade dos genótipos foram A2 (Casa Branca, SP - 680m) em seguida A1 (Vazante, MG - 720m), apresentaram maior produção entre todos e o ambiente menos favoráveis para produtividade foi A4 (Varante - MG 650m). E também há formação de dois grupos de ambientes: Grupo 1 formado pelos locais A4 e A3 e outro por A1 e A2, sendo que estes grupos apresentam características opostas (Figura 5).

Em relação às adaptações específicas dos genótipos em ambientes, verifica-se que os genótipos G4, G14 e G15 foram mais adaptados ao ambiente A3; G8 e G10 ao ambiente A4; G11, G7 e G13 ao ambiente A1, G1 e G3 ao ambiente A2 (Figura 6). Observa-se que os vetores relativos a esses genótipos apontam para uma direção comum, que pode ser orientada pelos ambientes (Gauch, 1992). Além disso, constata-se falta de adaptação dos genótipos G4, G14 e G15 ao ambiente A1, dos genótipos G8 e G12 ao ambiente A2, do genótipo G1 ao ambiente A4 e G11 ao A3 (Figura 6).

3.6 Comparação dos modelos AMMI com GGE biplot

Na Tabela 3 são apresentados os valores singulares dos quatro componentes principais das análises AMMI e GGE biplot. Para este conjunto de dados de MET, esses valores foram fundamentais para fazer a comparação entre os dois modelos.

Tabela 3 - Valores singulares dos quatro componentes principais (PCs) para os modelos AMMI e GGE biplot

Modelos	Valores singulares			
	1	2	3	4
AMMI	1561,95	981,32	712,24	$5,91 \times 10^{-6}$
GGE biplot	21	16,61	15,63	12,62

Calculando a correlação entre os valores singulares dos componentes principais dos dois modelos com $r = 0,9834$. Na comparação dos modelos existe uma forte correlação ($r > 0,95$) (Camargo-Buitrago et al., 2011) de PCs entre genótipo e a produtividade média em todos os ambientes. Portanto, os modelos AMMI e GGE biplot são muito semelhantes em relação à análise deste conjunto de dados MET.

Pela relação de informação ($IR \geq 1$) para os dois modelos que contêm os padrões nos componentes principais, o $IR_1=2,49$ no AMMI vs $IR_1 = 1,57$ do GGE biplot, logo, o PC1 do AMMI explicou maior proporção da variação dos dados do que o PC1 do GGE biplot.

Na análise AMMI, os dois primeiros componentes (PC1 e PC2) explicaram 87% da variação total da $G \times E$, já na análise GGE biplot, os dois primeiros componentes explicaram 63,96 % da variação total. O AMMI foi o melhor modelo para neste conjunto de dados, explicou maiores variações com os dois primeiros componentes, portanto, as interpretações dos gráficos de biplot são as mais confiáveis para representar os padrões nos dados sobre genótipos e ambientes.

Conclusões

A utilização conjunta dos modelos AMMI e GGE Biplot tornam mais eficazes a exploração da interação genótipo × ambiente e efetivamente resolveram os quatro principais objetivos para os dados multi-ambientes (MET).

Na comparação dos modelos para os dados MET provenientes da empresa Criagene SK, mostraram que os modelos são muito semelhantes em relação à análise de dados MET, pois são altamente correlacionados entre si. O AMMI foi o melhor e explicou a maior variação com os dois primeiros componentes (87%), portanto, as interpretações dos gráficos biplot são as mais confiáveis para representar os padrões nos dados sobre genótipos e ambientes.

HONGYU, K.; SILVA, F. S.; OLIVEIRA, A. C. S.; SARTI, D. A.; ARAÚJO, L. B.; DIAS, C. T. S. Comparison between AMMI models and GGE biplot for multi-environment trials. *Rev. Bras. Biom.*, São Paulo, v.33, n.2, p.139-155, 2015.

- **ABSTRACT:** *Genotype by environment interactions is an extremely important issue in plant breeding and production. The selection and recommendation of superior genotypes are hampered due to the constant occurrence of interaction, represents a major challenge for researchers. Thus, the multi-environment trials are essential for the selection of effective strain and recommendation of cultivars. There are several models in the literature for analysis of multi-environmental data, models AMMI (Additive Main effects and Multiplicative Interaction) and GGE (Genotype main effects + Genotype environment interaction) biplot are the most used. The AMMI model combines the analysis of variance and principal component analysis, to adjust, respectively, the main effects of genotype and environment and the effects of interaction. The GGE biplot is a method based on principal component analysis to explore the multi-environment trials and allowed the views of the biplot graphs relationships between test environments, genotypes and genotype by environment. The main objectives of this study with the use of models AMMI and GGE biplot: 1) mega-environment research for understanding the target environment; 2) evaluation of genotypes and environments within each mega-environment; 3) understand the causes of genotype by environment; and 4) Create a new method for comparing the AMMI models with GGE biplot.*
- **KEYWORDS:** *Genotype by environment interaction; AMMI model; biplot GGE; mega-environments; principal component analysis.*

Referências

ARAÚJO, M. F. C ; ARAÚJO, L. B. ; FARIA, P. N ; DIAS, C. T. S. Statistical test for genotype and environment contribution in the genotypes × environments interaction matrix. *Interciencia*, Caracas, v.37, n.7, p.515-519, 2012.

ARCINIEGAS-ALARCÓN, S.; DIAS, C. T. S.; GARCIA-PEÑA. Imputação múltipla livre de distribuição em tabelas incompletas de dupla entrada. *Pesquisa Agropecuária Brasileira*, v.49, n.9, p.683-691, 2014.

- CAMARGO-BUITRAGO, I.; INTIRE, E. Q. M.; GORDÓN-MENDOZA, R. Identificación de mega-ambientes para potenciar el uso de genótipos superiores de arroz em Panamá. *Pesquisa Agropecuária Brasileira*, v.46, n.9, p.1601-1069, 2011.
- DIAS, C.T.S.; KRZANOWSKI, W.J. Choosing components in the additive main effect and multiplicative interaction (AMMI) models. *Scientia Agricola*. v.63, n.2, p.169-175, 2006.
- GABRIEL, K.R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, v.58, n.3, p.453-467, 1971.
- GAUCH, H. G. *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*, Amsterdam, Elsevier, 1992.
- GAUCH, H. G. Statistical Analysis of Yield Trials by AMMI and GGE. *Crop Science*, v.46, n.4, p.1488-1500, 2006.
- GAUCH, H. G. A Simple Protocol for AMMI Analysis of Yield Trials. *Crop Science*, v.53, n.5, p.1860-1869, 2013.
- GAUCH, H. G.; ZOBEL, R. W. *AMMI analysis of yield trails*. In: KANG, M. S.; GAUCH, H. G. (Ed) *Genotype by environment interaction*. New York: CRC Press, p.416-428, 1996.
- GAUCH, H. G.; ZOBEL, R. W. Identifying mega-environments and targeting genotypes. *Crop Science*, v.37, n.2, p.311-326, 1997.
- GAUCH, H. G.; PIEPHO, H. P.; ANNICCHIARICO, P. Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Science*, v.48, n.3, p.866-889, 2008.
- GOLLOB, H.F. A statistical model which combines features of factor analitic and analysis of variance techniques. *Psychometrika*, Colorado Springs, v.33, n.1, p.73-115, 1968.
- HONGYU, K.; GARCIA-PENÑA, M.; ARAÚJO, L. B.; DIAS, C. T. S. Statistical analysis of yield trials by AMMI analysis of genotype x environment interaction. *Biometrical Letters*, Poznan, v.51, p.89-102, 2014.
- R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, Vienna, 2014.
- RODRIGUES, P. C.; MALOSETTI, M.; GAUCH, H. G.; VAN EEUWIJK F. A. A weighted AMMI algorithm to study genotype-by-environment interaction and QTL-by-environment interaction. *Crop Science*, v.54, n.4, p.1555-1570, 2014.
- SILVA, R. R.; BENIN, G. Análises Biplot: conceitos, interpretações e aplicações. *Ciência Rural*, Santa Maria, v.42, n.8, p.1404-1412, 2012.
- YAN, W. Singular-value partition for biplot analysis of multi-environment trial data. *Agronomy Journal*, v.94, n.5, p.990-996, 2002.
- YAN, W. GGE Biplot vs. AMMI Graphs for Genotype-by-Environment Data Analysis. *Journal of the India Society of Agricultural Statistics*, v.65, n.2, p.181-193, 2011.
- YAN, W.; KANG, M. S. *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists*. CRC Press, Boca Raton, FL, USA, 2003.271p.

YAN, W; TINKER, A. Biplot analysis of multi environment trial data: principles and applications. *Canadian Journal of Plant Science*, v.86, n.3, p.623-645, 2006.

YAN, W; HOLLAND, J. B. A heritability-adjusted GGE biplot for test environment evaluation. *Euphytica*. v.171, n.3, p.355-369, 2010.

YAN, W.; HUNT, L. A.; SHENG, Q.; SZLAVNICS, Z. Cultivar evaluation and mega-environment investigation based on GGE biplot. *Crop Science*, v.40, n. 3, p.597-605, 2000

YAN, W.; KANG, M.S.; MA, B.; WOODS, S.; CORNELIUS, P.L. GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Science*, v.47, n.2, p.643-653, 2007.

Recebido em 03.02.2015

Aprovado após revisão em 18.05.2015