

Gráficos *biplot* e *joint plot* para o estudo da interação tripla

Lúcio Borges de Araújo^{I*} Mirian Fernandes Carvalho Araújo^I Carlos Tadeu dos Santos Dias^{II}

Resumo: *Este trabalho tem os seguintes objetivos: propor uma sistemática para o estudo e a interpretação da estabilidade e adaptabilidade fenotípica, através do modelo de Tucker3; avaliar o uso do gráfico joint plot; comparar os resultados obtidos pela metodologia de três entradas com os resultados obtidos pelo modelo AMMI de duas entradas e análise biplot. Os dados a serem utilizados são relativos a 13 genótipos de feijão que foram conduzidos em 9 experimentos. Os resultados indicaram que o gráfico joint plot, facilita o entendimento da interação tripla e traz ao pesquisador informações mais reais sobre a interação tripla, do que a modelagem AMMI de duas entradas; o genótipo 6 é o que menos contribui para a interação e os genótipos 12, 9 e 5 são os que mais contribuem para a interação.*

Palavras-chave: *Interação genótipos×ambiente ×anos; Modelo Tucker3; Biplot; Estabilidade*

1 INTRODUÇÃO

Os experimentos multi-ambientais (MET) são conduzidos através de vários anos para os principais produtos agrícolas no mundo, constituindo um passo caro, mas essencial para a liberação de um novo genótipo de um produto agrícola e, conseqüentemente, a recomendação de cultivar. Por isto, melhoria nos métodos usados para análise de dados é de interesse à comunidade de melhoristas.

Para a descrição da resposta média de genótipos em ambientes e para o estudo e interpretação interação genótipo × ambiente (GE), em METs de experimentos agrícolas, duas classes de modelos são comumente utilizadas: modelos lineares e modelos lineares-bilineares.

Considere agora o caso em que os METs são avaliados através de vários anos (ou seja, genótipos × locais × anos), em que os dados podem ser organizados em arranjo de três entradas que, neste caso, cada entrada se refere a genótipos, locais e anos.

Em alguns casos o investigador pode estar interessado em saber se existe uma estrutura comum encoberta pelos locais com relação aos anos e como os vários genótipos respondem através da estrutura formada por ambientes e anos. Um procedimento para ganhar uma compreensão clara em arranjo GLA de três entradas é determinar uma estrutura dimensional menor, expressado em componentes principais, para a interação genótipos × locais × anos e então estudar as relações entre estes componentes. Para os dados organizados em arranjo de três entradas existem alguns modelos para analisá-los, como por exemplo, os modelo propostos por TUCKER(1966), que fornece uma decomposição trilinear dos dados organizados no arranjo.

Na maioria dos estudos, devido a falta de uma ferramenta adequada para estudar a interação entre genótipos, locais e anos, os pesquisadores combinam os fatores locais e anos (criando outro fator, denominado ambiente) e fazem o estudo da interação genótipos × ambientes utilizando uma decomposição em valores singulares, chamada de os modelos AMMI (Additive main effects and multiplicative interaction) por GAUCH & ZOBEL(1988).

Assim, o trabalho tem os objetivos de propor uma sistemática para o estudo e a interpretação da estabilidade e adaptabilidade fenotípica, por meio do modelo de Tucker3, e avaliar o uso do gráfico *joint plot*.

2 MATERIAL E MÉTODOS

^I Faculdade de Matemática (FAMAT), Universidade Federal de Uberlândia (UFU), Uberlândia, MG, Brasil. * autor para correspondência E-mail: lucio@famat.ufu.br

^{II} Departamento de Ciências Exatas (LCE), Universidade de São Paulo (USP), Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Piracicaba, SP, Brasil,

Os dados referem-se a 13 genótipos de feijão que foram conduzidos em 9 experimentais distintos constituídos pelos anos agrícolas de 2000/2001, 2001/2002 e 2005/2006, pelos municípios de Dourados e Aquidauana, sendo que os experimentos foram instalados na época das águas (Dourados) e também na época da seca (Dourados e Aquidauana). Cada local é constituído de município e uma época de instalação. Têm-se ainda que em cada experimento foi utilizado um delineamento em blocos ao acaso, com 3 blocos em cada experimento. Neste trabalho foi considerado a variável produtividade de grãos em ton/h.

O modelo AMMI pressupõe componentes aditivos para os efeitos principais de genótipos e ambientes e componentes multiplicativos para o efeito de interação. Então, a resposta média sobre b blocos do i -ésimo genótipo no j -ésimo ambiente é representada por:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^q \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij} + \varepsilon_{ij}, \text{ sendo que: } Y_{ij}: \text{ é a resposta média do } i\text{-ésimo}$$

genótipo no j -ésimo ambiente, com $i=1,2,\dots,g$ e $j=1,2,\dots,e$; μ : é uma constante, geralmente a média; g_i : é o efeito do i -ésimo genótipo; e_j : é o efeito do j -ésimo ambiente; o

efeito da interação do i -ésimo genótipo com j -ésimo ambiente é $(ge)_{ij} = \sum_{k=1}^q \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij}$, em

que: λ_k : é a raiz quadrada do k -ésimo autovalor da matriz $(GE)(GE)'$ (ou $(GE)'$ (GE)), com $k=1,2,\dots,q$ e onde $q < p$ determina uma aproximação de mínimos quadrados para a matriz GE pelos q primeiros termos da DVS e $p = \min\{g-1, e-1\}$; α_{ik} : é o i -ésimo elemento do vetor coluna \mathbf{a}_k associado a λ_k ; γ_{jk} : é o j -ésimo elemento do vetor linha $\mathbf{\gamma}_k$ associado a λ_k ; ρ_{ij} : é o resíduo adicional; ε_{ij} : é erro experimental associado ao i -ésimo genótipo no j -ésimo ambiente, assumido ser independente e $\varepsilon_{ij} \sim N(0, \frac{\sigma^2}{b})$;

A matriz GE é a interação entre genótipos \times ambiente (matriz de resíduos), em que cada elemento $(ge)_{ij}$ é encontrado pela seguinte relação: $(ge)_{ij} = Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_{..}$, 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 ; $\bar{Y}_{..}$: é a média geral do experimento.

Uma possível generalização do modelo de componentes principais para dados de duas entradas é para um arranjo de três entradas \mathbf{X} , com elementos

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk}, \text{ sendo que } e_{ijk} \text{ é um elemento do arranjo de resíduos } \mathbf{E}$$

$(I \times J \times K)$; a_{ip} , b_{jq} e c_{kr} são elementos das matrizes de cargas $A(I \times P)$, $B(J \times Q)$ e $C(K \times R)$; e g_{pqr} é um elemento do arranjo núcleo $\mathbf{G}(P \times Q \times R)$. Este é o modelo Tucker3 de $\mathbf{X}(P, Q, R)$, em que P , Q , R indicam o número de componentes em cada entradas.

Uma vez fixado o número de componentes nas matrizes A , B e C , a estimação dos parâmetros a_{ip} , b_{jq} e c_{kr} do modelo de Tucker3 é feito através do método iterativo de mínimos quadrados alternados (SMILDE et al, 2004), em que cada conjunto de parâmetros é estimado condicionalmente aos demais parâmetros. Para determinar a melhor solução do modelo de Tucker3, utilizou-se o método proposto por TIMMERMAN & KIERS (2000).

O modelo AMMI para a interação tripla pressupõe componentes aditivos para os efeitos principais de genótipos, locais e anos e componentes multiplicativos para os efeitos de interações duplas e triplas. Então, a resposta média sobre b repetições ou blocos do i -ésimo genótipo no j -ésimo local e no k -ésimo ano é representada por:

$$Y_{ijk} = \mu + g_i + l_j + a_k + (gl)_{ij} + (ga)_{ik} + (la)_{jk} + \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + \rho_{ijk} + \varepsilon_{ijk}, \text{ sendo que:}$$

Y_{ijk} : é a resposta média do i -ésimo genótipo no j -ésimo local no k -ésimo ano, com $i = 1, 2, \dots, g$, $j = 1, 2, \dots, l$ e $k = 1, 2, \dots, a$; μ : é uma constante; g_i : é o efeito do i -ésimo genótipo; l_j : é o efeito do j -ésimo local; a_k : é o efeito do k -ésimo ano; $(gl)_{ij}$: é o efeito da interação do i -ésimo genótipo com o j -ésimo local; $(ga)_{ik}$: é o efeito da interação do i -ésimo genótipo com o k -ésimo ano; $(la)_{jk}$: é o efeito da interação do j -ésimo local com o k -ésimo ano; a_{ip} : é um elemento da matriz de componentes A ; b_{jq} : é um elemento da matriz de componentes B ; c_{kr} : é um elemento da matriz de componentes C ; ρ_{ijk} : é o resíduo adicional da interação tripla que não é explicado pelo Tucker3 (P, Q, R); o efeito da interação do i -ésimo genótipo no j -ésimo local no k -ésimo ano é $(gla)_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + \rho_{ijk}$; ε_{ijk} : é erro experimental associado ao i -ésimo genótipo no j -ésimo local no k -ésimo ano, assumido ser $\varepsilon_{ijk} \sim N(0, \frac{\sigma^2}{b})$ e todos os ε_{ijk} independentes;

Seja \underline{Z} é um arranjo com as interações entre genótipos \times locais \times anos (arranjo de resíduos), de modo que cada elemento $(gla)_{ijk}$ do arranjo de três-entradas \underline{Z} é estimado como:

$(\widehat{gla})_{ijk} = Y_{ijk} - \bar{Y}_{ij.} - \bar{Y}_{i.k} - \bar{Y}_{.ij} + \bar{Y}_{i..} + \bar{Y}_{.j.} + \bar{Y}_{..k} - \bar{Y}_{...}$, em que: $(\widehat{gla})_{ijk}$: é o efeito da interação tripla estimada para o genótipo i no local j e no ano k ; Y_{ijk} : é a média das b repetições do genótipo i no local j e no ano k ; $\bar{Y}_{ij.}$: é a média dos elementos da i -ésima linha com a j -ésima coluna do arranjo de interação; $\bar{Y}_{i.k}$: é a média dos elementos da i -ésima linha com o k -ésimo tubo do arranjo de interação; $\bar{Y}_{.jk}$: é a média dos elementos da j -ésima coluna com o k -ésimo tubo do arranjo de interação; $\bar{Y}_{i..}$: é a média dos elementos da i -ésima fatia horizontal do arranjo de interação; $\bar{Y}_{.j.}$: é a média dos elementos da j -ésima fatia vertical do arranjo de interação; $\bar{Y}_{..k}$: é a média dos elementos da k -ésima fatia frontal do arranjo de interação; $\bar{Y}_{...}$: é a média geral do experimento.

Biplot é uma representação em que as linhas e as colunas são apresentadas em um gráfico com duas ou três dimensões, sendo que a construção do *biplot* é baseada na decomposição em valores singulares (DVS) da matriz de dados (GABRIEL, 1971). Assim é possível representar, os efeitos de interação de cada genótipo com cada ambiente.

O *joint plot* (KROONENBERG, 1983) é utilizado para representar graficamente os modelos de Tucker3, sendo semelhante a um *biplot*. A diferença entre eles é que na construção do *joint plot* é construído um *biplot* para dois fatores dada a matriz de componente do modelo Tucker3 referente ao terceiro fator, ou seja, cada *joint plot* é construído usando diferentes fatias do arranjo núcleo. Para construir um *joint plot*, é necessário obter uma matriz $\Delta_r = A G_r B_r' = A_r^* B_r^{*}$ de dimensão $I \times J$, com $r = 1, 2, \dots, R$ e então, através da DVS, representa-se Δ_r por meio de um *biplot*, obtendo-se o *joint plot*.

Quanto a interpretação de um *joint plot* (VARELA, et al., 2006), suponha um gráfico que é projetado sobre r -ésimo componente principal da terceira entrada tal que, no *joint plot* aparecem todos os níveis das duas primeiras entradas. Em seguida, selecione a partir da matriz C , os níveis deste fator com maior peso no r -ésimo componente (positivos ou negativos). Suponha que a matriz C tem um valor positivo e elevado associado ao k -ésimo nível da terceira entrada, então proximidades entre os níveis da primeira e da segunda entrada, indicam que a interação tripla entre i -ésimo nível da primeira entrada, j -ésimo nível da segunda entrada e k -ésimo nível da terceira entrada é positiva. Em contrapartida, se o i -ésimo nível do primeiro fator está muito longe do j -ésimo nível do segundo fator, indica que a interação tripla associada a

estes três fatores é negativa. Caso a matriz C tenha um valor alto negativo associado ao k -ésimo nível do terceiro fator, a interação tripla serão o oposto de quando o valor é positivo. Em geral, os níveis de uma entrada localizado no centro do *joint plot* são considerados um conjunto que tem um desempenho médio em todos os outros modos.

4 RESULTADOS E DISCUSSÃO

A análise de variância conjunta considerando 13 genótipos, 9 ambientes (combinação de 3 locais e 3 anos) mostra que a interação genótipos \times ambiente é significativa ($SQ_{G \times E} = 58,64$ com 96 graus de liberdade) e representa 28% da variabilidade total dos dados. A soma de quadrados do resíduo é 8,51 com 216 graus de liberdade.

A próxima etapa da análise corresponde ao ajuste da interação pela decomposição em valor singular aplicada a matriz GE , que terá posto $p = \min(12,8) = 8$. Assim um modelo AMMI com dois componentes explica 64,5% da $SQ_{G \times E}$, como resposta padrão

A última etapa da análise AMMI consiste na representação gráfica dos genótipos e ambientes em um gráfico denominado de *biplot*. A Figura 1a ilustra o *biplot* resultante e a partir dele são feitas as devidas interpretações procurando identificar genótipos e ambientes, que menos contribuem para a interação entre genótipos \times ambientes. Logo, por este gráfico, nota-se que os genótipos que menos contribuíram para a interação (pontos próximos da origem, que indicam quais são genótipos estáveis) foram os genótipos G_2, G_3, G_4, G_6, G_7 e G_{11} , mas, entretanto para fins de recomendação de cultivares deseja-se uma alta performance na produtividade, que pode ser avaliada pelas médias (DUARTE; VENKOVSKY, 1999). Assim, entre estes genótipos destaca-se com os genótipos G_2 e G_3 , que tiveram médias altas, e os genótipos G_4, G_6, G_7 e G_{11} apresentaram as piores médias. Os demais genótipos tiveram adaptações específicas: G_1 adaptou-se especificamente ao ambiente E_2 , G_5 ao ambiente E_3 , G_8 aos ambientes E_1 e E_4 , G_9 aos ambientes E_4 e E_5 , G_{10} ao ambiente E_9 , G_{12} aos ambientes E_8 e E_9 e G_{13} ao ambiente E_9 .

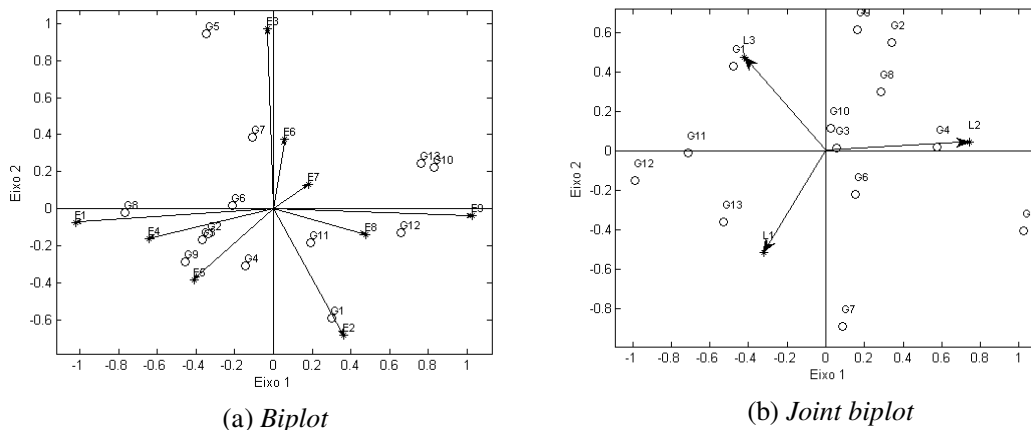


Figura 1 – *Biplot* e *joint plot* projetados dentro do primeiro componente do terceiro modo.

Por outro lado o ranqueamento dos genótipos num ambiente estável será de maior confiança para o melhorista. Assim entre os ambientes destaca-se E_6 (combinação do local Aquidauana na época das “secas” com o ano de 2001/2002), E_7 (combinação do local Dourados na época das “águas” com o ano de 2005/2006) e E_8 (combinação do local Dourados na época das “secas” com o ano de 2005/2006) como ambientes estáveis.

relação ao ano 3 (c_{31} é positivo), tem-se que: o genótipo 1 teve uma interação positiva com L_3 , negativa com L_2 e não interagem com o local L_1 ; para os genótipos 11, 12 e 13, observa-se uma interação negativa com L_2 e positiva com L_1 e L_3 ; os genótipos 4 e 5 teve uma interação negativa com L_1 e L_3 e positiva com L_2 ; para os genótipos 2, 8 e 9, nota-se que a interação é positiva com L_2 e L_3 e negativa com L_1 ; o genótipo 7 teve uma interação positiva com L_1 , negativa com L_3 e não interagem com o local L_2 ; e com relação aos genótipos 3, 6 e 10, que estão no centro deste gráfico, pode-se dizer que todos tem uma baixa interação com todos os locais no ano 1 e no ano 3 e conseqüentemente são genótipos estáveis.

4 CONCLUSÕES

A sistemática de análise proposta com o uso dos modelos de Tucker3 mostrou-se eficiente e adequada para separar o padrão de resposta técnica e o ruído contidos no arranjo de interação tripla, bem como a interpretação da estabilidade e adaptabilidade fenotípica. O gráfico *joint plot* facilita o entendimento da interação tripla, além de trazer informações adicionais sobre estas, comparativamente aos gráficos *biplot* obtidos pela modelagem AMMI de duas entradas, permitindo identificar quais combinações de genótipos, locais e anos contribuem ou não para essa interação.

AGRADECIMENTOS

Os autores agradecem a FAPEMIG pelo apoio financeiro

REFERÊNCIAS

- DUARTE, J.B.; VENCOVSKY, R. **Interação genótipo x ambiente: uma introdução à análise "AMMI"**. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60p. (Série Monografias, 9).
- GABRIEL, K.R. The biplot graphic display of matrices with applications to principal components analysis. **Biometrika**, Cambridge, v.58, p.453-467, 1971
- GAUCH, H.G.; ZOBEL, R.W. Predictive and postdictive success of statistical analysis of yield trials. **Theoretical and Applied Genetics**, New York, v. 76, p. 1-10, 1988.
- KROONENBERG, P.M. **Three-Mode Principal Component Analysis: Theory and applications**. Leiden: DSWO Press, 1983, 398p.
- SMILDE, A.; BRO, R.; GELADI, P. **Multi-way Analysis With Applications in the Chemical Sciences**. Chichester: John Wiley & Sons, Ltd, 2004. 369p.
- TIMMERMAN, M. E.; KIERS, H.A.L. Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. **British Journal of Mathematical and Statistical Psychology**, London, v.53, n.1, p.1-16, 2000
- TUCKER, L. Some mathematical notes on three-mode factor analysis. **Psychometrika**. New York, v.31, p.279-311, 1966.
- VARELA, M. et. al. Analysis of a three-way interaction including multi-attributes. **Australian Journal of Agricultural Research**. Sydney, v.57, p.1185-1193, 2006.