



Ciência Rural

ISSN: 0103-8478

cienciarural@mail.ufsm.br

Universidade Federal de Santa Maria

Brasil

Piovesan, Pamela; Borges de Araújo, Lucio; Santos Dias, Carlos Tadeu dos
Validação cruzada com correção de autovalores e regressão isotônica nos modelos de efeitos
principais aditivos e interação multiplicativa
Ciência Rural, vol. 39, núm. 4, julio, 2009, pp. 1018-1023
Universidade Federal de Santa Maria
Santa Maria, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=33115802050>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe , Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

Validação cruzada com correção de autovalores e regressão isotônica nos modelos de efeitos principais aditivos e interação multiplicativa

Cross-validation with eigenvalue correction and isotonic regression in the additive main effect and multiplicative interaction model

Pamela Piovesan^I Lucio Borges de Araújo^{II} Carlos Tadeu dos Santos Dias^{III}

RESUMO

Neste trabalho é apresentada a aplicação dos modelos AMMI com o propósito de analisar a interação entre genótipo e ambiente em experimentos agronômicos multiambientais com dados balanceados. São apresentados dois métodos de validação cruzada e o aperfeiçoamento desses métodos por meio da correção de autovalores, sendo estes ordenados por meio da regressão isotônica. É realizado um estudo comparativo entre esses métodos por meio de dados reais. Os resultados mostram para esse conjunto de dados que o método de EASTMENT & KRZANOWSKI (1982) seleciona um modelo mais parcimonioso. Além disso, quando esse método é aperfeiçoado com a correção dos autovalores, o número de componentes não se altera. O método de GABRIEL (2002) seleciona um maior número de termos no modelo, e, quando se aplica a correção de autovalores, o número de termos diminui. O aperfeiçoamento desses métodos por meio da correção de autovalores traz um grande benefício para o pesquisador do ponto de vista prático, uma vez que a seleção do número de termos multiplicativos representa um ganho do número de blocos (ou repetições), quando se utiliza o modelo AMMI em vez do modelo completo, sendo, portanto, melhor utilizar o modelo AMMI com correção dos autovalores e selecionar o número de componentes por meio do método de Eastment e Krzanowski.

Palavras-chave: interação genótipo × ambiente, experimentos agronômicos multiambientais, análise multivariada.

ABSTRACT

This paper presents an application of AMMI models - Additive Main effects and Multiplicative Interaction model -

for a thorough study about the effect of the interaction between genotype and environment in multi-environments experiments with balanced data. Two methods of crossed validation are presented and the improvement of these methods through the correction of eigenvalues, being these rearranged by the isotonic regression. A comparative study between these methods is made, with real data. The results show that the EASTMENT & KRZANOWSKI (1982) method selects a more parsimonious model and when this method is improved with the correction of the eigenvalues, the number of components are not modified. GABRIEL (2002) method selects a huge number of terms to hold back in the model, and when this method is improved by the correction of eigenvalue, the number of terms diminishes. Therefore, the improvement of these methods through the correction of eigenvalues brings a great benefit from the practical point of view for the analyst of data proceeding from multi-ambient, since the selection of numbers of multiplicative terms represents a profit of the number of blocks (or repetitions), when the model AMMI is used, instead of the complete model.

Key words: genotype × environment interaction, multi-environments experiments, multivariate analysis.

INTRODUÇÃO

Em experimentos agrícolas, os ensaios são realizados em vários ambientes para, por meio de uma análise estatística adequada, isolar o componente de variabilidade devido à interação entre genótipos e ambientes ($G \times E$). Os programas de melhoramento têm como principal objetivo buscar técnicas capazes de

^IDepartamento de Ciências Exatas (LCE), Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), Piracicaba, SP, Brasil.

^{II}Departamento de Bioestatística, Universidade Estadual Paulista "Júlio e Mesquita Filho" (UNESP), Botucatu, SP, Brasil. LCE, ESALQ, USP, Piracicaba, SP, Brasil.

^{III}LCE, ESALQ, USP. Av. Pádua Dias, 11, 13418-900, Piracicaba, SP, Brasil. E-mail: ctsdias@esalq.usp.br. Autor para correspondência.

detectar de maneira aprofundada essas interações para a seleção de genótipos consistentes e de elevada produtividade.

O método AMMI (Additive Main effects and Multiplicative Interaction model) surge, então, com a finalidade de estudar detalhadamente as interações ($G \times E$) por meio da decomposição ortogonal da soma de quadrados dessas interações, fato que o torna vantajoso se comparado aos métodos tradicionais. Além disso, esse método apresenta uma boa capacidade preditiva. O AMMI é um procedimento unimultivariado que vem ganhando destaque, pois combina num único modelo componentes aditivos para os efeitos principais e componentes multiplicativos para os efeitos da interação genótipo-ambiente, combinando seqüencialmente a análise de variância (técnica univariada) e a análise de componentes principais (técnica multivariada), respectivamente. Esse procedimento permite que a análise de componentes principais seja diretamente tratada por meio da decomposição por valores singulares (DVS). A interação ($G \times E$) é representada pelos termos multiplicativos do modelo AMMI, e a seleção do número ótimo desses termos pode ser feita de duas maneiras: por meio de testes de hipóteses ou de formas de validação cruzada.

A validação cruzada é um procedimento estatístico livre de distribuição teórica que vem ganhando aplicação recente. GAUCH & ZOBEL (1988) introduziu a chamada validação preditiva por meio da validação cruzada a fim de selecionar o melhor modelo AMMI.

Assim, o presente trabalho tem como objetivos: 1) aplicar o modelo AMMI com o propósito de analisar a estrutura de interação entre genótipo e ambiente em experimentos agronômicos multiambientais com dados balanceados; 2) apresentar e aplicar técnicas de validação cruzada para a escolha do número de termos multiplicativos do modelo AMMI; 3) aperfeiçoar as técnicas de validação cruzada a partir de três métodos de correção de autovalores, ordenados por meio da regressão isotônica, e avaliar os métodos de validação cruzada aqui modificados por meio do estudo de dados reais, tendo em vista que há viés na estimativa dos autovalores (ARAÚJO & DIAS, 2006).

MATERIAL E MÉTODOS

Os dados utilizados são referentes a ensaios internacionais multiambientais de cultivares obtidos pelo CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo). A variável resposta avaliada foi produção em kg. Os experimentos foram conduzidos

em um delineamento aleatorizado em blocos e são referentes a 20 genótipos de trigo avaliados em 34 ambientes com quatro blocos.

O modelo AMMI postula componentes aditivos para os efeitos principais de genótipos e ambientes e componentes multiplicativos para a interação entre genótipo e ambiente. Esse modelo utiliza a matriz de médias de dados Y e é dado por:

$$\bar{Y}_{ij} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ij},$$

em que \bar{Y}_{ij} é a resposta média do genótipo i no ambiente j , μ é a performance média geral, g_i é o efeito fixo do genótipo i ($i = 1, \dots, g$), e_j é o efeito fixo da interação específica do genótipo i e o ambiente j , na qual $(ge)_{ij}$ é representada por: $\sum_{k=1}^m \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij}$, sob as restrições, $\sum_i g_i = \sum_j e_j = \sum_i (ge)_{ij} = \sum_j (ge)_{ij} = \sum_i \sum_j (ge)_{ij} = 0$ pois a estimativa da soma de quadrado de resíduos é feita por meio do método de mínimos quadrados.

O ajuste do modelo AMMI para dados balanceados é obtido em duas etapas (ZOBEL et al., 1988): a primeira etapa consiste na obtenção das estimativas dos efeitos principais (g_i e e_j) e da média geral μ , estimativas que formam a parte aditiva do modelo, sendo ajustadas por uma análise de variância comum (ANOVA) aplicada à matriz de médias Y . Com a aplicação da ANOVA, obtém-se um resíduo de não-aditividade ε_{ij} e o efeito de interação recai nesse resíduo, formando a matriz de interações GE. Cada elemento da matriz de interações GE é obtido por:

$$\varepsilon_{ij} = (ge)_{ij} = Y_{ij} - \bar{Y}_{i*} - \bar{Y}_{*j} + \bar{Y}_{**},$$

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. Essa matriz de interação GE é então analisada pela decomposição por valores singulares (DVS). A partir daí, a análise AMMI centra-se na decomposição da soma de quadrados dos elementos da matriz GE, obtida de forma direta por DVS. Assim, λ_k é o k -ésimo valor singular de GE, α_{ik} é o i -ésimo elemento do vetor singular coluna $\alpha_{k(g \times 1)}$ associado a λ_k e γ_{jk} é o j -ésimo elemento do vetor singular linha $\gamma_{k(1 \times e)}$ associado a λ_k (MANDEL, 1971; PIEPHO, 1995).

Os dois métodos de validação cruzada utilizam a decomposição por valores singulares e consistem em eliminar cada elemento ge_{ij} da matriz GE com dimensão $g \times e$, predizer o valor \widehat{ge}_{ij} de ge_{ij} para cada possível escolha de m (o número de componentes) e medir a discrepância entre o valor atual e o valor predito por $PRESS(m) = \frac{1}{ge} \sum_{i=1}^g \sum_{j=1}^e (\widehat{ge}_{ij} - ge_{ij})^2$. Como o objetivo é estimar cada elemento da matriz para evitar

viés, os dados de ge_{ij} não devem ser utilizados nos cálculos de ge_{ij} para cada i e j .

O procedimento de validação cruzada de Eastment e Krzanowski é descrito como: Sejam $\mathbf{GE}^{(i)}$ a matriz resultante após deletar a i -ésima linha de \mathbf{GE} e centralizá-la em torno das médias das linhas e $\mathbf{GE}_{(-i)}$ a matriz resultante após deletar a j -ésima coluna de \mathbf{GE} e centralizá-la em torno das médias das colunas. Por meio da decomposição por valores singulares, podemos escrever: $\mathbf{GE}^{(i)} = \bar{\mathbf{U}} \bar{\mathbf{D}} \bar{\mathbf{V}}$; com $\bar{\mathbf{U}} = (\bar{u}_{st})$, $\bar{\mathbf{V}} = (\bar{v}_{st})$ e $\bar{\mathbf{D}} = diag(\bar{d}_1, \dots, \bar{d}_p)$, $\mathbf{GE}_{(-j)} = \tilde{\mathbf{U}} \tilde{\mathbf{D}} \tilde{\mathbf{V}}$; com $\tilde{\mathbf{U}} = (\tilde{u}_{st})$, $\tilde{\mathbf{V}} = (\tilde{v}_{st})$ e $\tilde{\mathbf{D}} = diag(\tilde{d}_1, \dots, \tilde{d}_m)$.

O objetivo é estimar os elementos ge_{ij} da matriz de interações dados por $\widehat{ge}_{ij} = \sum_{t=1}^m u_{it} d_t v_{jt}$, em

que u_{it} e v_{jt} são elementos das matrizes \mathbf{U} e \mathbf{V} , respectivamente, tal que $\mathbf{GE} = \mathbf{UDV}'$.

Para a estimativa de ge_{ij} , ao invés de utilizar a informação da i -ésima linha \mathbf{U} e j -ésima coluna \mathbf{V} , são utilizados $\tilde{\mathbf{U}}$ e $\tilde{\mathbf{V}}$, respectivamente. Para a parte central \mathbf{D} , é utilizada uma composição das duas matrizes $\tilde{\mathbf{D}}$ e \mathbf{D} , sendo o valor predito dado por:

$$\widehat{ge}_{ij} = \sum_{t=1}^m (\tilde{u}_{it} \sqrt{\tilde{d}_t})(\tilde{v}_{jt} \sqrt{d_t}).$$

$m = 0, 1, \dots, p$. O procedimento de GABRIEL (2002) toma uma mistura de regressão e aproximação de uma matriz de posto inferior como base para sua predição. O algoritmo é descrito a seguir: Seja a matriz \mathbf{GE} de interação e considere a partição:

$$\mathbf{GE} = \begin{bmatrix} ge_{11} & \mathbf{ge}_{1*} \\ \mathbf{ge}_{*1} & \mathbf{GE}_{11} \end{bmatrix}.$$

A submatriz \mathbf{GE}_{11} de posto m , após aplicar a DVS, pode ser escrita como $\mathbf{GE}_{11} = \sum_{k=1}^m u_{ik} d_k v_{ik} = \mathbf{UDV}'$, em que $\mathbf{U} = [u_1, \dots, u_m]$, $\mathbf{V} = [v_1, \dots, v_m]$ e $\mathbf{D} = diag(d_1, \dots, d_m)$. A estimativa do elemento ge_{11} utilizando o modelo de regressão é dada por: $ge_{11} = \mathbf{ge}_{1*} \hat{\beta}$, em que o vetor $\hat{\beta}$ é referente ao modelo de regressão relativo à submatriz \mathbf{GE}_{11} e utilizando \mathbf{ge}_{*1} na estimativa desse parâmetro tem-se $\hat{\beta} = (\mathbf{GE}'_{11} \mathbf{GE}_{11})^{-1} \mathbf{GE}'_{11} \mathbf{ge}_{*1}$ e como $\mathbf{GE}'_{11} = \mathbf{UDV}'$ segue que $\hat{\beta} = \mathbf{V} \mathbf{D}^{-1} \mathbf{U}' \mathbf{g}_{*1}$. Então, prediz-se ge_{11} por: $ge_{11} = \mathbf{ge}_{1*} \mathbf{VD}^{-1} \mathbf{U}' \mathbf{g}_{*1}$. Analogamente, o valor ajustado da validação cruzada é obtido para todos os outros elementos ge_{ij} , $i = 1, \dots, g$; $j = 1, \dots, m$; ($i,j \neq (1,1)$).

A qualidade do ajuste é dada pelas estatísticas: $PRESS(m) = \frac{1}{ge} \sum_{i=1}^g \sum_{j=1}^e (\widehat{ge}_{ij}^m - ge_{ij})^2$ e $PRECORR(m) = Corr(ge_{ij}, ge_{ij} | \forall i, j)$, em que $PRESS(m)$ (*Predictive Sum of Squares*) é a soma de quadrados da predição de um modelo ajustado com m componentes, e o $PRECORR(m)$ (*Predictive Correlation*) é a correlação entre os valores observados e os valores preditos pelo modelo com m

componentes e cada valor ajustado requer uma diferente partição de \mathbf{GE} . Tanto para o procedimento de validação cruzada de EASTMENT & KRZANOWSKI (1982), quanto para o procedimento de GABRIEL (2002), a escolha do número de termos multiplicativos m pode ser baseada em alguma função apropriada de $PRESS(m)$. Porém, o comportamento das estatísticas difere para os métodos. O método de GABRIEL (2002) produz valores que primeiro decrescem e depois (normalmente) crescem com m , sendo, então, sugerida a escolha do valor m que minimiza a função. O método de EASTMENT & KRZANOWSKI (1982) produz, geralmente, um conjunto de valores que é monotonicamente não-crescente. Devido a isso é sugerido o uso da estatística dada por:

$$W_m = \frac{PRESS(m-1) - PRESS(m)}{D_m} \div \frac{PRESS(m)}{D_r}$$

em que D_m é o número de graus de liberdade requerido para ajustar o m -ésimo componente e é dado por: $D_m = g + e - 2m$. D_r é o número de graus de liberdade remanescentes após ajustar o m -ésimo componente e é dado por: $D_r = D_{r-1} - [g + e - (m-1)2]$, $r = 2, 3, \dots, (g-1)$ e $D_1 = (g-1)e$. A estatística W_m representa o aumento na informação preditiva média suprida pelo m -ésimo componente, dividido pela informação preditiva média em cada um dos componentes. Então, o número de componentes importantes m são aqueles em que m é o maior entre todos os m tal que W_m é maior que a unidade (EASTMENT & KRZANOWSKI, 1982).

Os três métodos apresentados por MUIRHEAD (1987) para a correção dos autovalores, utilizados nesse caso para a correção dos autovalores provenientes da matriz $(\mathbf{GE})(\mathbf{GE})'$ ou $(\mathbf{GE})'(\mathbf{GE})$ em $(\mathbf{GE})(\mathbf{GE})' \sim W(e, \Sigma)$, são dados a seguir:

i) O primeiro método de correção é dado por:

$$\phi_k^{(1)} = \frac{e \hat{\lambda}_k^2}{\alpha_k^{(1)}} = \frac{e \hat{\lambda}_k^2}{e - g + 1 + 2 \hat{\lambda}_k^2 \sum_{k \neq k^*}^p \frac{1}{\hat{\lambda}_k^2 - e \hat{\lambda}_{k^*}^2}} \quad (1)$$

ii) O segundo método de correção é dado por:

$$\phi_k^{(2)} = \frac{e \hat{\lambda}_k^2}{\alpha_k^{(2)}} = \frac{e \hat{\lambda}_k^2}{e - g - 1 + 2 \hat{\lambda}_k^2 \sum_{k \neq k^*}^p \frac{1}{\hat{\lambda}_k^2 - e \hat{\lambda}_{k^*}^2}} \quad (2)$$

iii) O terceiro método de correção é dado por:

$$\begin{aligned} \phi_k^{(3)} &= \frac{e \hat{\lambda}_k^2}{e + g + 1 - 2k} - \frac{e \hat{\lambda}_k^2 \ln(e \hat{\lambda}_k^2)}{b + \sum_{k=1}^p [\ln(e \hat{\lambda}_k^2)]^2} = \\ &\frac{e \hat{\lambda}_k^2 (b + \sum_{k=1}^p [\ln(e \hat{\lambda}_k^2)]^2) - (e + g + 1 - 2k) \ln(e \hat{\lambda}_k^2)}{(e + g + 1 - 2k)(b + \sum_{k=1}^p [\ln(e \hat{\lambda}_k^2)]^2)} = \frac{\alpha_k^{(3)}}{d_k} \end{aligned} \quad (3)$$

$$\text{em que: } b = \frac{5,8(g-2)^2}{(e+g-1)^2}, c = \frac{6(g-2)}{5(e+g-1)^2},$$

$$\alpha_k^{(3)} = e\hat{\lambda}_k^2(b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2 - (e+g+1-2k)c\ln(e\hat{\lambda}_k^2))$$

$$\text{e } d_k = (e+g+1-2k)(b + \sum_{k=1}^p [\ln(e\hat{\lambda}_k^2)]^2).$$

Nos três métodos de correção, temos que: g é o número de genótipos do experimento; e é o número de ambientes do experimento e $\hat{\lambda}_k^2$ é o k -ésimo autovalor da matriz $(GE)(GE)'$ ou $(GE)'(GE)$ com $k = 1, 2, \dots, p$, sendo $p = \min\{g-1, e-1\}$. Porém, nem sempre os autovalores obtidos com os métodos de correção apresentados estão em ordem decrescente ($\phi_1 \geq \phi_2 \geq \dots \geq \phi_p$) e para reordená-los será utilizada a regressão isotônica.

O algoritmo para ordenar os autovalores, obtidos pelos três métodos de correção, é apresentado por LIN & PERLMAN (1985) e é descrito como segue: os produtos $e\hat{\lambda}_k^2$ (numerador dos métodos de correção (1) e (2) ou d_k (denominador do método de correção (3)) são listados em uma coluna e em outra coluna é listado o denominador do método de correção (1), $\alpha_k^{(1)}$, ou o denominador do método de correção (2), $\alpha_k^{(2)}$, ou ainda o numerador do método de correção (3), $\alpha_k^{(3)}$, como segue:

$$\begin{array}{ll} e\hat{\lambda}_1^2 \text{ ou } d_1 & \alpha_1^{(1)} \text{ ou } \alpha_1^{(2)} \text{ ou } \alpha_1^{(3)} \\ e\hat{\lambda}_2^2 \text{ ou } d_2 & \alpha_2^{(1)} \text{ ou } \alpha_2^{(2)} \text{ ou } \alpha_2^{(3)} \\ \vdots & \vdots \\ e\hat{\lambda}_p^2 \text{ ou } d_p & \alpha_p^{(1)} \text{ ou } \alpha_p^{(2)} \text{ ou } \alpha_p^{(3)} \end{array}$$

Passo 1) Fazendo todos α_k 's positivos: a) Inicia-se pelo final da lista e procura-se para cima até que se encontre o primeiro par ($e\hat{\lambda}_k^2, \alpha_k$) com α_k negativo; b) Soma-se esse par com o par imediatamente acima dele, substituindo-os pelo par ($e\hat{\lambda}_k^2 + e\hat{\lambda}_{k-1}^2, \alpha_k + \alpha_{k-1}$) para que na lista um par seja menor do que o próximo; c) Repetem-se (a) e (b) para a nova lista até que todos α_k sejam positivos.

Passo 2) Re-ordenando as razões $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ de

forma que estejam em ordem decrescente: Listam-se as razões $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ à direita de cada par ($e\hat{\lambda}_k^2, \alpha_k$) obtido no passo 1. Um par ($e\hat{\lambda}_k^2, \alpha_k$) (exceto o par no final da lista) é chamado de par violado se a razão $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ não é maior do que a razão $\frac{e\hat{\lambda}_{k+1}^2}{\alpha_{k+1}}$. a) Inicia-se pelo final da

lista encontrada no passo 1 e procede-se para cima até o primeiro par violado ser encontrado; b) Soma-se esse par violado com o par imediatamente acima dele,

substituem-se esses dois pares e suas razões pelo par ($e\hat{\lambda}_k^2 + e\hat{\lambda}_{k-1}^2, \alpha_k + \alpha_{k-1}$) e sua razão $\frac{e\hat{\lambda}_k^2 + e\hat{\lambda}_{k-1}^2}{\alpha_k + \alpha_{k-1}}$, formando uma nova lista; c) Reinicia-se no par imediatamente após o par trocado em (b) e procede-se para cima até o próximo par violado ser encontrado e repete-se (b); d) Repete-se (c) até todas as razões $\frac{e\hat{\lambda}_k^2}{\alpha_k}$ estarem em ordem decrescente.

Passo 3) Cada razão no final da lista foi obtida por bloco acumulado de um ou mais pares consecutivos ($e\hat{\lambda}_k^2, \alpha_k$) na lista original.

Na maioria dos estudos, existe um grande interesse na comparação do Erro Médio Quadrático do modelo (QM_{EM} (modelo)) selecionado com o Erro Médio Quadrático (QM_{EM}) do experimento. Para NACHIT et al., (1992), uma aproximação para o número de blocos (ou repetições) que falta para o modelo AMMI completo apresentar uma performance igual ao modelo AMMI selecionado é uma forma de medir a eficiência da correção e validação cruzada por meio do número de blocos que se ganha ao analisar os dados com o modelo selecionado. Essa medida é obtida por:

$$R_{AMMI} = \frac{QM_{EM}}{QM_{EM}(\text{modelo})}.$$

Para fazer uma estimativa do QM_{EM} (modelo), PIEPHO (1994) sugere a seguinte expressão: $QM_{EM}(\text{modelo}) = (RMSPD)^2 - (QM_{EM})$. A medida RMSPD (Root Mean Square Prediction Difference) apresentada por CORNÉLIUS et al. (1993) refere-se ao poder de predição do modelo selecionado, podendo, assim, fazer uma comparação aproximada com outros RMSPDs provenientes de modelos ajustados por validação cruzada. O ajuste é feito por:

$$RMSPD_{(PRESS)} = \sqrt{PRESS + \frac{(r-1)}{r} QM_{EM}}$$

sendo r o número de blocos em cada ambiente do experimento.

RESULTADOS E DISCUSSÃO

A tabela 1 mostra a análise de variância conjunta do experimento. Como pode ser observado, há efeito significativo entre blocos, genótipos, ambientes e também interação entre genótipos e ambientes. A análise AMMI focaliza-se no estudo dessas interações, que neste experimento corresponde a 9,97% da soma de quadrados total. Obtendo a matriz de interações, sabemos que a soma de quadrados dos autovalores dessa matriz, obtida por meio da DVS, fornece exatamente a soma de quadrados da interação. Porém, busca-se nessa análise descartar um ruído, deixando no modelo somente os componentes necessários para a explicação dessa interação.

Validação cruzada com correção de autovalores e regressão isotônica...

Tabela 1 - Análise de variância conjunta do experimento para a variável produção de trigo em kg.

Fonte de Variação	GL	SQ	QM	F	P
Blocos d. ambiente (B d. E.)	102	257862519	2528064	7,14	<0.0001
Genótipo (G)	19	89066441	4687707	13,23	<0.0001
Ambiente (E)	33	4333925428	131331074	370,67	<0.0001
Interação (G×E)	627	594108485	947541	2,67	<0.0001
Resíduo	1938	68664646195	354307		

Na tabela 2, são apresentados os resultados para o experimento quando são aplicados os procedimentos de EASTMENT & KRZANOWSKI (1982) e GABRIEL (2002). Observa-se que a estatística PRESS difere para os dois métodos. Para o método de Krzanowski, na medida em que o número de componentes aumenta, a estatística PRESS normalmente diminui. Para o método de Gabriel, os valores da estatística oscilam conforme aumenta o número de componentes. Devido a essa diferença, para o método de Krzanowski utiliza-se a estatística W_m , cujo número de componentes m a serem selecionados é aquele maior valor de m em que W_m é maior que um. Por meio dessa medida, observamos que o número de componentes a ser retido para esse experimento é três. Para o método de Gabriel, seleciona-se o número de termos em que a estatística PRESS minimiza a função, portanto, para esse experimento são selecionados cinco

componentes. A estatística PRECORR para os dois métodos apresenta valores próximos conforme aumenta o número de componentes.

Os métodos de correção de autovalores 1 e 2 não se mostraram eficientes nos dois procedimentos de validação cruzada, fornecendo autovalores negativos em que o processo de regressão isotônica, por meio do algoritmo de isotonização de Stein, não conseguiu torná-lo positivo para posteriormente reordená-lo na ordem decrescente. Portanto, não foi possível utilizar esses dois métodos. O método 3, apresentado por MUIRHEAD (1987), foi o único que corrigiu com sucesso todos os autovalores utilizados para estimar cada valor da matriz de interações, tanto no método de validação cruzada dado por EASTMENT & KRZANOWSKI (1982), quanto naquele dado por GABRIEL (2002). Os resultados dos dois procedimentos de validação cruzada com o método 3 de correção de autovalores são apresentados na tabela 2.

Tabela 2 - Validação cruzada sem correção e com correção de autovalores pelo método 3.

M	Sem correção						Com correção					
	Eastment e Krzanowski			Gabriel			Eastment e Krzanowski			Gabriel		
M	PRESS	W	PRECORR	PRESS	PRECORR	PRESS	PRESS	W	PRECORR	PRESS	PRECORR	PRESS
1	193891,3	1,44	0,946	195490,0	0,946	189228,61	1,76	0,947	207446,6	0,942		
2	192323,6	0,09	0,947	197841,8	0,945	181414,08	0,47	0,950	216701,4	0,940		
3	170883,2	1,28	0,953	171088,7	0,953	160751,13	1,31	0,956	192694,5	0,947		
4	158021,9	0,78	0,957	157381,9	0,956	147271,67	0,88	0,959	183652,4	0,950		
5	158224,1	-0,01	0,957	156528,2	0,957	145470,82	0,11	0,960	185235,5	0,950		
6	158383,9	-0,01	0,957	160492,9	0,956	143470,08	0,12	0,960	191413,8	0,948		
7	157092,2	0,06	0,958	176520,3	0,951	142202,69	0,07	0,961	213263,5	0,942		
8	143497,2	0,67	0,961	161816,6	0,955	129800,16	0,68	0,964	198737,7	0,946		
9	143404,5	0,00	0,961	167634,1	0,954	129116,69	0,03	0,964	205425,8	0,944		
10	140609,6	0,11	0,962	183000,6	0,949	126191,09	0,13	0,965	223115,7	0,940		
11	135448,8	0,19	0,964	193150,9	0,947	122076,47	0,17	0,966	236129,7	0,937		
12	128569,7	0,23	0,966	195596,9	0,946	115438,74	0,25	0,968	240506,1	0,936		
13	123189,6	0,16	0,967	192330,1	0,947	110533,06	0,16	0,970	235487,4	0,937		
14	121481,3	0,04	0,968	196925,9	0,946	109219,71	0,03	0,970	238557,4	0,937		
15	120075,0	0,02	0,968	231282,3	0,937	108174,97	0,02	0,970	264484,3	0,930		
16	116127,2	0,03	0,969	209015,1	0,943	103750,95	0,04	0,972	243167,4	0,936		
17	115898,1	0,00	0,969	222919,8	0,940	104939,12	0,00	0,971	250590,2	0,934		
18	113713,0	-0,02	0,970	313613,4	0,917	102336,08	-0,03	0,972	299998,6	0,922		

Os autovalores corrigidos não alteraram o número de componentes a serem retidos no modelo quando se aplica o método de Eastment e Krzanowski. Já para o método de Gabriel, o número de componentes passou a ser 4, diminuindo um componente se comparado ao mesmo método sem a correção de autovalores.

A medida RAMMI foi calculada tanto para a validação cruzada de EASTMENT & KRZANOWSKI (1982), quanto para a validação cruzada de GABRIEL (2002), quando os autovalores são corrigidos pelo método 3 de correção. Essa medida representa o número de blocos que se ganha ao analisar os dados com o modelo AMMI selecionado em vez do modelo AMMI completo. De acordo com essa medida, ao selecionar o modelo AMMI pelo método de EASTMENT & KRZANOWSKI (1982), tem-se um ganho de aproximadamente cinco blocos ($R_{AMMI} = 4,91$). Para o método de GABRIEL (2002), o ganho é de aproximadamente quatro blocos ($R_{AMMI} = 3,73$). Logo esses valores podem ser utilizados como um indicativo de que o método proposto por Eastment e Krzanowski, combinado com a correção de autovalores, é melhor que o método proposto Gabriel juntamente com os autovalores corrigidos.

CONCLUSÕES

Quando são aplicados os métodos de validação cruzada e quando estes são aperfeiçoados, os resultados dos experimentos permitem afirmar que o método de EASTMENT & KRZANOWSKI (1982) seleciona modelos mais parcimoniosos. Quando esse método é aperfeiçoado por meio da correção de autovalores pelo método 3 e a regressão isotônica é utilizada para reordenar os autovalores corrigidos, observa-se que o número de componentes a ser retido no modelo não é alterado. Já o método de GABRIEL (2002) seleciona um maior número de termos para reter no modelo. Porém, quando são utilizadas a correção de autovalores pelo método 3 e a regressão isotônica para reordená-los, observa-se que o número de termos diminui.

A seleção de termos multiplicativos é de grande importância, pois, na prática, significa um ganho do número de blocos (ou repetições) quando o modelo AMMI é selecionado em vez do modelo completo. Portanto, o aperfeiçoamento desses métodos por meio da correção de autovalores traz um grande benefício do ponto de vista prático para o analista de dados provenientes de ensaios multi-ambientais.

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro a todos os autores

REFERÊNCIAS

- ARAÚJO, L.B.; DIAS, C.T.S. Métodos de correção de autovalores e regressão isotônica nos modelos AMMI. *Revista de Matemática e Estatística*, Jaboticabal, v.24, n.2, p.71-89, 2006.
- CORNELIUS, P.L. et. al. Test and estimators of multiplicative models for variety trials. In: ANNUAL KANSAS STATE UNIVERSITY CONFERENCE ON APPLIED STATISTICS IN AGRICULTURE, 5., 1993, Manhattan. *Proceedings...* Manhattan : KS, 1993. p.156-166.
- EASTMENT, H.T.; KRZANOWSKI, W.J. Cross-validatory choice of the number of components from a principal components analysis. *Technometrics*, Alexandria, v.24, p.73-77, 1982.
- GABRIEL, K.R. Le biplot- outil d'exploration de données multidimensionnelles. *Journal de la Societe Francaise de Statistique*, Paris, v.143, p.5-55, 2002.
- GAUCH, H.G.; ZOBEL, R.W. Predictive and postdictive sucess of statistical analysis of yield trials. *Theoretical and Applied Genetics*, New York, v.76, p.1-10, 1988.
- LIN, S.P.; PERLMAN, M.D. A Monte Carlo comparison of four estimators of a covariance matrix. In: KRISHNAIAH, P.R. *Multivariate analysis-VI*. Amsterdam: Elsevier, 1985. Cap.26, p.411-429.
- MANDEL, J. A new analysis of variance model for non-additive data. *Technometrics*, v.13, n.1, p.1-18, 1971.
- MUIRHEAD, R.J. Developments in eigenvalue estimation. In: GUPTA, A.K. *Advances in Multivariate Statistical*. Boston: Reidel, 1987. p.277-288.
- NACHIT, M.M. et al. Use of AMMI and linear regression models to analyze genotype-environment interaction in durum wheat. *Theoretical and Applied genetics*, New York, v.83, p.597-601, 1992.
- PIEPHO, H.P. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. *Theoretical and Applied Genetics*, New York, v.89, p.647-654, 1994.
- PIEPHO, H.P. Robustness of statistical test for multiplicative terms in additive main effects and multiplicative interaction model for cultivar trial. *Theoretical and Applied Genetics*, New York, v.90, p.438-443, 1995.
- ZOBEL, R.W. et. al. Statistical analysis of a yield trial. *Agronomy Journal*, Madson, v.80, p.388-393, 1988.