

# Métodos de análise de dados longitudinais para o melhoramento genético da pinha

Keny Henrique Mariguele<sup>(1)</sup>, Marcos Deon Vilela de Resende<sup>(2)</sup>, José Marcelo Soriano Viana<sup>(3)</sup>, Fabyano Fonseca e Silva<sup>(4)</sup>, Paulo Sérgio Lima de Silva<sup>(5)</sup> e Filipe de Castro Knop<sup>(6)</sup>

<sup>(1)</sup>RiceTec Sementes Ltda., Rua 18 de Novembro, nº 341, Bairro Navegantes, CEP 90240-040 Porto Alegre, RS. E-mail: kmariguele@hotmail.com <sup>(2)</sup>Embrapa Florestas, Estrada da Ribeira, Km 111, CEP 83411-000 Colombo, PR. E-mail: deon@cnpf.embrapa.br <sup>(3)</sup>Universidade Federal de Viçosa (UFV), Departamento de Biologia Geral, Avenida P.H. Rolfs, s/nº, CEP 36570-000 Viçosa, MG. E-mail: jmsviana@ufv.br <sup>(4)</sup>UFV, Departamento de Estatística. E-mail: fabyanofonseca@ufv.br <sup>(5)</sup>Universidade Federal Rural do Semi-Árido, Departamento de Ciências Vegetais, BR 110, Km 47, Bairro Presidente Costa e Silva, CEP 59625-900 Mossoró, RN. E-mail: paulosergio@ufersa.edu.br <sup>(6)</sup>UFV, Departamento de Fitotecnia. E-mail: fcknop@yahoo.com.br

Resumo – O objetivo deste trabalho foi comparar formas de análise de medidas repetidas para o melhoramento da produção de frutos de pinha (*Annona squamosa*). Vinte progênies de meias-irmãs foram avaliadas por três anos (2003, 2004 e 2005) em delineamento de blocos ao acaso, com cinco repetições, com cada parcela constituída de quatro plantas. A característica avaliada foi o número de frutos por indivíduo. Os modelos de simetria composta, de simetria composta com variâncias heterogêneas, autorregressivo com variâncias heterogêneas, e antedependência estruturada, foram analisados com o programa ASReml. A estimação dos componentes de variância e a predição dos valores genéticos foram feitas com o procedimento REML/BLUP. A comparação dos modelos foi realizada pelo teste de razão de verossimilhança e pelo critério de Akaike. O modelo antedependência estruturada, para os fatores progênie e parcela, e o modelo multivariado, para o fator resíduo, são as melhores abordagens para a análise dos dados, pois propiciam eficiência e parcimônia em relação ao modelo multivariado completo. Com o modelo antedependência estruturada, é possível a identificação de famílias superiores, em cada colheita, e também de famílias com maior número total de frutos.

Termos para indexação: *Annona squamosa*, Akaike, matriz de variância e covariância, medidas repetidas, REML/BLUP, valores genéticos.

## Methods of longitudinal data analysis for the genetic improvement of sugar apple

Abstract – The objective of this work was to compare the ways of analyzing repeated measures to improve the production of sugar apple (*Annona squamosa*). Twenty half-sib progenies were evaluated, over three years (2003, 2004 and 2005), in a randomized block design with five replicates, and each plot was constituted of four plants. The evaluated trait was the number of fruit per individual. The models of compound symmetry, autoregressive with heterogeneous variance, the structured ante-dependence, and compound symmetry with heterogeneous variance were analyzed using the ASReml software. The estimation of variance components and the prediction of breeding values were made by the REML/BLUP. The comparison of the models was done by the likelihood ratio test and Akaike's information criterion. The structured ante-dependence model, for the factors progeny and parcel, and the multivariate model, for the residual factor, are the best approaches for data analysis, providing efficiency and parsimony over the full multivariate model. With the structured ante-dependence model, it is possible to identify superior families in each harvest, and also the families with larger total number of fruit.

Index terms: *Annona squamosa*, Akaike, variance and covariance matrix, repeated measure, REML/BLUP, genetic value.

## Introdução

A análise de experimentos de medidas repetidas nas parcelas e indivíduos apresenta peculiaridades, pelo fato de as várias colheitas serem correlacionadas entre si e pela possibilidade de haver heterogeneidade de variâncias e de covariâncias entre as várias colheitas

ou safras. Um modelo completo e adequado para analisar um conjunto de dados dessa natureza é o multivariado, também denominado modelo com matriz de covariância não estruturada entre colheitas (UN), o qual trata cada colheita como se fosse uma variável diferente. Essa estrutura de covariância é aplicada a todos os fatores aleatórios do modelo, tais como os

efeitos genotípicos de tratamento, efeitos de parcelas e efeitos residuais (Gilmour et al., 2004). Porém, quando se considera um número relativamente grande (três ou mais) de colheitas, tal modelo é difícil de ser ajustado por uma questão de convergência. Além disso, pode ser superparametrizado, ou seja, depender da estimativa de um grande número de parâmetros (Resende, 2007).

Há várias alternativas (Resende & Sturion, 2001; Resende, 2002, 2007; Resende & Thompson, 2003; Gilmour et al., 2004) para a predição de efeitos e modelagem da estrutura de correlação entre as medidas repetidas em cada indivíduo (ou tratamento), ao longo do tempo. Geralmente, as diferentes abordagens são implementadas via metodologia de modelos mistos.

O procedimento BLUP (best linear unbiased prediction) assume que os componentes de variância são conhecidos, e as propriedades dos preditores somente são asseguradas nessas condições (Henderson, 1984). Assim, na prática, os componentes de variância devem ser estimados da maneira mais fidedigna possível, para que as estimativas possam substituir adequadamente os parâmetros (Resende et al., 2000). Nesse caso, o procedimento padrão para estimação dos componentes de variância, no contexto dos modelos lineares mistos, é o da máxima verossimilhança restrita (REML), descrito em detalhes por Searle et al. (1992) e Rao (1997).

O uso do modelo de repetibilidade (o mais simples e parcimonioso), em comparação ao modelo multivariado (o mais completo e complexo), possibilita grande eficiência, pois as perdas são de apenas 0 a 5%, quando a correlação genética entre medições sucessivas é alta (acima de 0,80). Nesses casos, a correlação entre ordenamentos pelo modelo de repetibilidade e pelo modelo multivariado é muito próxima de 1 (Mrode, 2005). O grande benefício do modelo de repetibilidade, comparado ao multivariado, refere-se à necessidade de estimativas de poucos parâmetros genéticos. Um modelo misto de repetibilidade, que contempla simultaneamente a herdabilidade e a repetibilidade do caráter, foi proposto por Resende et al. (2000) para o caso de delineamento experimental de blocos ao acaso com várias plantas por parcela. No entanto, estruturas intermediárias entre os modelos de repetibilidade e multivariado podem ser mais eficientes.

Em plantas perenes, o número de medições realizadas varia tipicamente de 3 a 6 anos de colheitas, pois um número maior de colheitas compromete a

eficiência dos programas de melhoramento por unidade de tempo (Resende, 2002). Assim, as opções mais interessantes aos melhoristas de plantas perenes são os modelos de: repetibilidade, quando as suposições são atendidas; repetibilidade + interação genótipos x medições; simetria composta (CS); autorregressivo com variâncias heterogêneas (ARH); antedependência estruturada (SAD); e de simetria composta, com variâncias heterogêneas (CSH).

A estrutura CS assume tanto a homogeneidade de variâncias como a de covariâncias entre colheitas (Littell et al., 1998). Essa abordagem é desejável porque depende do menor número possível de parâmetros. No entanto, pode ser ineficiente no caso de heterogeneidade de variância e de covariância entre colheitas, algumas vezes decorrentes de efeitos de escala de medida de uma colheita a outra. Esse procedimento considera simultaneamente a heterogeneidade de variância genética e a fenotípica entre colheitas, sendo penalizado pela alta variância fenotípica e capitalizado pela alta variância genética, o que equivale aproximadamente a padronizar pelo desvio-padrão ambiental de cada colheita e multiplicar pela média desses desvios-padrão. Dessa forma, a aplicação de estruturas CS fornece resultados semelhantes aos que se obtêm quando se aplica a estrutura CSH.

As alternativas de modelo ARH, SAD e multivariado podem ser aplicadas aos vários fatores aleatórios do modelo estatístico base. No contexto de estatística experimental com efeitos de tratamentos considerados fixos, estas modelagens, em geral, se aplicam somente aos resíduos. Mas no caso do melhoramento, em que os tratamentos genéticos são considerados de efeitos aleatórios, essas modelagens podem ser aplicadas aos efeitos residuais e também genéticos (Gilmour et al., 2004). Inclusive, diferentes modelagens podem ser aplicadas aos efeitos genéticos e residuais. Para algumas espécies, a modelagem dos efeitos genéticos como ARH ou SAD e a dos erros por um modelo multivariado mostram-se eficientes (Resende et al., 2006).

Na literatura, há escassez de informação sobre a estimação de parâmetros genéticos e dados de medidas repetidas em fruteiras, embora a importância dos modelos de repetibilidade para a avaliação genética em algumas espécies perenes tenha sido enfatizada (Resende et al., 2000). Assim, os resultados obtidos no presente trabalho, com a pinha, contribuem para

preencher uma lacuna e propiciam informações sobre modelagens adicionais que podem ser aplicadas às fruteiras.

O objetivo deste trabalho foi comparar formas de análise de medidas repetidas para o melhoramento genético de pinha, por meio de diferentes estruturas de matriz de covariância para os fatores progênie, parcela e resíduo.

## Material e Métodos

Foram avaliadas 20 progênies de meias-irmãs de pinheiras, em três anos de colheitas (2003, 2004 e 2005), cultivadas em delineamento de blocos ao acaso, com cinco repetições, em que cada parcela era constituída de quatro plantas. A característica avaliada foi número de frutos por indivíduo, em colheitas anuais (produção total de cada ano).

O modelo de repetibilidade (Resende, 2002) foi  $y = Xb + Zg + Wc + Tp + \varepsilon$ , em que:  $y$  é o vetor de dados;  $b$  é o vetor dos efeitos das combinações ano-repetição, assumidos como fixos, conforme recomendação de Littell et al. (1998), e somados à média geral;  $g$  é o vetor dos efeitos genotípicos de progênies, assumidos como aleatórios;  $c$  é o vetor dos efeitos aleatórios de ambiente permanente (parcelas no caso);  $p$  é vetor dos efeitos aleatórios permanentes de indivíduo; e  $\varepsilon$  é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

As distribuições e estruturas de médias e variâncias foram:

$$y|b, V \sim N(Xb, V); g|\sigma_g^2 \sim N(0, \sigma_g^2); c|\sigma_c^2 \sim N(0, I\sigma_c^2)$$

$$p|\sigma_p^2 \sim N(0, I\sigma_p^2); e|\sigma_e^2 \sim N(0, I\sigma_e^2), e,$$

$$\text{Cov}(g, c') = 0; \text{Cov}(g, p') = 0; \text{Cov}(g, e') = 0;$$

$$\text{Cov}(p, c') = 0; \text{Cov}(p, e') = 0; \text{Cov}(c, e') = 0,$$

ou seja:

$$E \begin{bmatrix} y \\ g \\ c \\ p \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \text{Var} \begin{bmatrix} y \\ g \\ c \\ p \\ e \end{bmatrix} = \begin{bmatrix} V & ZG & WC & TP & R \\ GZ & G & 0 & 0 & 0 \\ CW' & 0 & C & 0 & 0 \\ PT' & 0 & 0 & P & 0 \\ R & 0 & 0 & 0 & R \end{bmatrix},$$

em que:

$$G = I\sigma_g^2; P = I\sigma_p^2; C = I\sigma_c^2; R = I\sigma_e^2, e,$$

$$V = Z\sigma_g^2 Z' + WI\sigma_c^2 W' + TI\sigma_p^2 T' + I\sigma_e^2.$$

A estrutura da matriz de covariância entre medidas repetidas, para o fator progênie, é:

$$\Sigma = \begin{bmatrix} \sigma_g^2 & \sigma_g^2 & \sigma_g^2 \\ & \sigma_g^2 & \sigma_g^2 \\ & & \sigma_g^2 \end{bmatrix}$$

e a estrutura da matriz de correlação, para o fator progênie, é:

$$H = \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}.$$

O modelo de simetria composta (Resende, 2007) foi  $y = Xb + Zg + Wp + Tp + Qi + \varepsilon$ , em que:  $y$  é o vetor de dados;  $b$  é o vetor dos efeitos das combinações ano-repetição, assumidos como fixos, conforme recomendação de Littell et al. (1998), e somados à média geral;  $g$  é o vetor dos efeitos genotípicos de progênies, assumidos como aleatórios;  $c$  é vetor dos efeitos aleatórios de ambiente permanente (parcelas no caso);  $p$  é vetor dos efeitos aleatórios permanentes de indivíduo;  $i$  é o vetor dos efeitos da interação genótipos x anos; e  $\varepsilon$  é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

A estrutura de covariância é similar à anterior, porém, com o acréscimo da matriz de covariância dos efeitos da interação  $GA = I\sigma_{ga}^2$ . A estrutura da matriz de covariância, para o fator progênie, é

$$\Sigma = \begin{bmatrix} \sigma_g^2 + \sigma_{gxa}^2 & \sigma_g^2 & \sigma_g^2 \\ & \sigma_g^2 + \sigma_{gxa}^2 & \sigma_g^2 \\ & & \sigma_g^2 + \sigma_{gxa}^2 \end{bmatrix},$$

e a estrutura da matriz de correlação entre medidas repetidas, para o fator progênie, é:

$$H = \begin{bmatrix} 1 & \sigma_g^2/(\sigma_g^2 + \sigma_{gxa}^2) & \sigma_g^2/(\sigma_g^2 + \sigma_{gxa}^2) \\ & 1 & \sigma_g^2/(\sigma_g^2 + \sigma_{gxa}^2) \\ & & 1 \end{bmatrix}$$

As estruturas ARH, SAD e multivariada (UN) são descritas a seguir, com base em Resende (2007). Em todas essas abordagens, foram contemplados os efeitos de progênie, de parcela e de resíduos.

Para o modelo autorregressivo com variâncias heterogêneas (ARH), conforme Pletcher & Geyer (1999) e Jaffrézic & Pletcher (2000), a estrutura da matriz de covariância é:

$$\Sigma = \begin{bmatrix} \sigma_{g1}^2 & \sigma_1 \sigma_2 \rho^{(1-2)} & \sigma_1 \sigma_3 \rho^{(1-3)} \\ & \sigma_{g2}^2 & \sigma_2 \sigma_3 \rho^{(2-3)} \\ & & \sigma_{g3}^2 \end{bmatrix}$$

Este modelo assume correlações diferentes entre idades e reconhece a existência de correlações seriais entre medidas repetidas. A estrutura geral de correlação de três diferentes safras é:

$$H = \begin{bmatrix} 1 & \rho^{[t_2-t_1]} & \rho^{[t_3-t_1]} \\ & 1 & \rho^{[t_3-t_2]} \\ & & 1 \end{bmatrix}$$

Verifica-se que o modelo AR estima uma só correlação e projeta-a para os demais lags. Neste caso, a matriz de covariância genética é dada por  $G = SRS$ , em que  $S$  é uma matriz diagonal, com elementos equivalentes à raiz quadrada da variância genética em cada colheita.

Para o modelo antedependência estruturada (SAD), conforme Núñez-Antón & Zimmermann (2000), a matriz de covariância é:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho_1 & \sigma_1 \sigma_3 \rho_1 \rho_2 \\ & \sigma_2^2 & \sigma_2 \sigma_3 \rho_2 \\ & & \sigma_3^2 \end{bmatrix}$$

e a matriz de correlação é:

$$H = \begin{bmatrix} 1 & \rho_1 & \rho_1 \rho_2 \\ & 1 & \rho_2 \\ & & 1 \end{bmatrix}$$

Verifica-se que a correlação (ou distância) entre as medidas 1 e 3 equivale ao produto entre a correlação da medida 1 com a 2 ( $\rho_1$ ) e a correlação da medida 2 com a 3 ( $\rho_2$ ).

Para o modelo multivariado (Resende, 2002), a estrutura geral da matriz de covariância é:

$$\Sigma = \begin{bmatrix} \sigma_{g1}^2 & \sigma_{12} & \sigma_{13} \\ & \sigma_{g2}^2 & \sigma_{23} \\ & & \sigma_{g3}^2 \end{bmatrix}$$

e a estrutura geral de correlação, entre as três diferentes safras, é dada por:

$$H = \begin{bmatrix} 1 & a & b \\ & 1 & c \\ & & 1 \end{bmatrix}$$

em que  $a$ ,  $b$  e  $c$  representam diferentes valores de correlação.

Todos os modelos foram analisados com o programa ASReml (Gilmour et al., 1997, 2004), e a estimação dos componentes de variância foi feita por meio do REML/BLUP.

Na análise de modelos mistos com dados desbalanceados, os efeitos do modelo não são testados via teste F, tal como se faz no modelo da análise de variância. Nesse caso, para efeitos aleatórios, o recomendado é o da razão de verossimilhança (LRT) (Resende, 2007).

Para o teste de razão de verossimilhança (LRT), foi usada a inferência probabilística ou critério de deviance (DIC). Além de sua utilidade na estimação, o princípio de verossimilhança também permite comparar a adequabilidade de vários modelos, desde que tenham uma estrutura hierárquica ou aninhada (Resende, 2007).

Os seguintes passos foram adotados: obtenção do ponto de máximo do logaritmo da função de verossimilhança residual (Log L), para os modelos a serem comparados; obtenção da deviance  $D = -2 \text{ Log L}$ , para os dois modelos; obtenção da diferença entre as deviances, para se calcular o qui-quadrado e realizar o teste da razão de verossimilhança (LRT); teste da significância dessa diferença, por meio do teste de qui-quadrado, com o número de graus de liberdade

igual à diferença entre o número de parâmetros entre os dois modelos testados.

O critério de informação de Akaike (AIC) penaliza a verossimilhança pelo número de parâmetros independentes ajustados. Por esse critério, qualquer parâmetro extra deve aumentar a verossimilhança, ao menos em uma unidade, para que entre no modelo. O AIC é dado por  $AIC = -2\log L + 2p$ , em que  $p$  é o número de parâmetros estimados (Akaike, 1974), conforme Littell et al. (1998).

A comparação entre os modelos de análise visou encontrar um modelo que melhor represente os dados e seja parcimonioso. Isto foi feito pelo teste da razão de verossimilhança e pelo critério de Akaike (AIC).

## Resultados e Discussão

A comparação entre os modelos de repetibilidade e CS mostrou que este último apresentou a menor deviance (Tabela 1). A escolha do modelo CS foi confirmada pelo AIC, que apresentou os valores 6.403,86 e 6.356,42, respectivamente. Como esperado, a estimativa da  $\sigma^2_{g_s}$  no modelo mais simples, foi superior à obtida no modelo CS, pois, no modelo simples de repetibilidade, ocorre a superestimação pelo fato de a interação estar embutida. O teste de  $\chi^2 = 49,44^{**}$  foi significativo a 1% de probabilidade, o que mostra que a diferença entre os dois modelos é importante e, portanto, é compensador acrescentar-se um parâmetro a mais ( $\sigma^2_{g_{xa}}$ ). Resultados similares foram obtidos por Resende & Thompson (2003).

Na comparação com o modelo CS, o ARH forneceu deviance menor (6.306,44) (Tabela 2). O valor de AIC para o ARH foi 6334,44, e é inferior ao obtido no modelo CS (6.346,42). O teste de  $\chi^2 = 39,98^{**}$  mostrou diferença significativa entre os modelos a 1% de probabilidade, e que as estimativas obtidas por este modelo são mais confiáveis, apesar do aumento do número de parâmetros.

Os anos três, um e dois, nesta ordem, apresentaram variabilidade genética decrescente. No caso do ARH (Tabela 2), as baixas estimativas de variâncias das parcelas conduziram a correlações entre os anos iguais a 1 e, assim, não são confiáveis. As maiores estimativas de correlação genética foram obtidas entre as colheitas 1 e 2 (0,57, 0,78 e 0,90), pelos modelos ARH, SAD e multivariado, respectivamente. Esses valores são

mais coerentes com a natureza quantitativa do caráter avaliado e com o controle genético diferenciado do mesmo caráter, ao longo das idades, conforme observado para outras espécies fruteiras como o cupuaçu (Alves et al., 2010) e o açaí (Farias Neto et al., 2011).

A deviance obtida pelo SAD foi de 6.137,36 (Tabela 2), inferior à obtida pelo ARH. Mais uma vez o AIC de 6.169,36 foi inferior ao obtido pelo SAD (6334,44). O teste  $\chi^2 = 169,08^{**}$  revelou diferença significativa entre os modelos e, apesar do aumento de dois parâmetros, as estimativas obtidas são mais confiáveis. Assim, as inferências práticas e decisões de seleção devem ser feitas a partir dos resultados e predições realizadas com base nesse modelo.

Os modelos SAD e multivariado apresentaram quase a mesma deviance (6.137,36 e 6.136,76) (Tabela 2). O modelo multivariado poderia ser escolhido, por apresentar valor ligeiramente menor da deviance. Entretanto, os valores de AIC foram 6.169,36 e 6.172,72 e, portanto, menor para o SAD. No entanto, pelo critério da parcimônia, confirmado pelo teste  $\chi^2 = 0,6^{ns}$ , não há diferença significativa entre os modelos. Assim, prefere-se o modelo menos parametrizado que, no caso, é o SAD, que apresenta dois parâmetros a menos. A superioridade do modelo SAD foi relatada também por Núñez Antón & Zimmermann (2000).

Além disso, o modelo multivariado completo e não estruturado, tende a apresentar problemas de convergência, em virtude das altas correlações, geralmente verificadas entre medidas repetidas, e do elevado número de parâmetros a ser estimado. A abordagem multivariada torna-se proibitiva quando o número de colheitas ou safras é elevado. Algumas vezes, mesmo com pequeno número de safras,

**Tabela 1.** Estimativas das variâncias associadas ao modelo de repetibilidade e simetria composta (CS), aplicado aos dados originais de número de frutos, em três colheitas de *Annona squamosa* L. Número total de parâmetros de cada modelo igual a 4 e 5, respectivamente.

Variância	Modelo	
	Repetibilidade	CS
$\sigma_g^2$	5,77	2,98
$\sigma_p^2$	2,97	2,96
$\sigma_e^2$	7,52	10,21
$\sigma_{g_{xa}}^2$	-	8,38
$\sigma^2$	67,45	59,38
Deviance	6.395,86	6.346,42

as matrizes de covariância não estruturadas não são positivas definidas e conduzem a correlações maiores que 1. Medidas repetidas, altamente correlacionadas, aumentam o risco de obtenção de matrizes de covariância não positivas definidas e de não convergência, no processo de estimação de componentes de variância. Também, quanto maior a ordem da matriz de covariância, maior é a chance de que ela não seja positiva definida (Resende, 2007).

Com correlações genéticas discrepantes entre pares de colheitas e muitas colheitas, outras estruturas intermediárias entre UN e CS podem ser usadas. Para caracteres de plantas perenes, em que as correlações seguem alguns padrões em função das distâncias entre as medidas consideradas, em geral havendo decréscimo das correlações em função do aumento dessas distâncias, os modelos ARH e SAD tendem a ser os mais adequados (Resende & Thompson, 2003; Resende et al., 2006). O modelo ARH pode ser também ser adequado, nas situações em que as correlações apresentam a mesma magnitude em intervalos de mesma dimensão, por exemplo, correlações iguais entre as colheitas 1 e 2 e entre as colheitas 2 e 3. Em geral, com a estabilização do caráter, a correlação entre as colheitas 2 e 3 tende a ser maior que a correlação entre as colheitas 1 e 2 e, portanto, o modelo SAD parece ser mais coerente pois contempla isso.

Cecon et al. (2008) trabalharam com a produção de cinco anos de 50 clones de café 'Conilon' e verificaram que o modelo que proporcionou o melhor ajuste foi o CSH. Malheiros (1999), em estudo teórico, verificou que o modelo de matriz de covariância mais apropriado é o multivariado não estruturado. Resende et al. (2006) trabalharam com 1.800 indivíduos de erva-mate, em três colheitas, e concluíram que os modelos ARH, SAD e multivariado apresentaram basicamente a mesma deviance, porém, o ARH foi o escolhido por ser mais parcimonioso pelo AIC. Pletcher & Geyer (1999) e Jaffrézic & Pletcher (2000) também relataram a superioridade do método ARH. Entretanto, Gilmour et al. (2004) destacaram ARH e SAD como modelagens mais favoráveis de maneira geral.

O teste de famílias é uma das estratégias mais exploradas para a seleção em espécies perenes (Oliveira et al., 2006). No presente trabalho, pelo modelo escolhido (SAD), com intensidade de seleção de 25% das famílias, observa-se que, na colheita 1, as melhores foram as famílias 3, 4, 15, 11 e 10, na colheita 2, as famílias 3, 4, 11, 10 e 2 e, na colheita 3, as famílias 11, 6, 20, 8 e 4. Pelas médias das colheitas, as melhores famílias foram a 11, 4, 6, 3 e 15. As famílias 4 e 11 estiveram entre as que apresentaram os maiores valores genéticos nas três colheitas individuais (Tabela 3).

**Tabela 2.** Estimativas dos parâmetros de variância, covariância e correlações, associados aos modelos - autorregressivo de primeira ordem com variâncias heterogêneas (ARH), antedependência estruturada (SAD) e multivariado completo - aplicados aos dados originais de número de frutos, em três colheitas de *Annona squamosa*. Número total de parâmetros dos modelos ARH, SAD e multivariado completo igual a 14, 16 e 18, respectivamente.

Tratamento (progênie)			Parcela			Resíduo		
Covariância	Variância	Correlação	Covariância	Variância	Correlação	Covariância	Variância	Correlação
Modelo ARH								
2,58	8,85	0,57	0,32E-13	0,99E-13	-	-4,14	67,18	-0,06
4,46	2,31	0,32	0,48E-6	0,10E-13	-	-0,66	74,26	-0,01
4,06	21,94	0,57	0,15E-6	22,56	-	5,86	51,59	0,09
Deviance				6.306,44				
Modelo SAD								
4,04	8,51	0,78	-1,71	5,98	-0,41	10,28	47,64	0,26
3,18	3,18	0,24	-1,76	3,00	-0,20	12,09	33,56	0,16
2,50	20,81	0,31	3,09	12,98	0,49	15,09	116,83	0,24
Deviance				6.137,36				
Modelo multivariado								
3,35	6,31	0,90	-0,52	1,01	-0,76	10,42	47,62	0,26
3,32	2,19	0,35	-0,38	0,46	-0,27	11,34	33,72	0,15
1,25	14,30	0,22	0,72	1,99	0,75	14,74	116,30	0,24
Deviance				6.136,76				

**Tabela 3.** Valores genéticos obtidos a partir das estimativas do modelo SAD, quanto ao número de frutos das 20 progênieis de meias-irmãs de pinheiras (*Annona squamosa*).

Família	Colheita 1	Colheita 2	Colheita 3	Geral
1	10,9583	12,0233	7,5383	10,1733
2	14,8385	15,4516	14,3663	14,8855
3	20,5083	17,9033	11,6853	16,6990
4	19,8643	16,8733	17,3113	18,0163
5	15,1084	14,5757	15,0277	14,9039
6	14,8927	15,4056	20,1953	16,8312
7	12,1393	13,0583	10,9493	12,0490
8	15,1880	13,4803	17,8943	15,5209
9	11,4383	12,6753	15,8423	13,3186
10	16,0593	15,8003	13,3953	15,0850
11	16,2263	16,5863	23,9203	18,9110
12	15,3689	14,9183	14,3662	14,8845
13	14,1159	14,9596	13,2643	14,1133
14	13,9403	14,7173	14,2993	14,3190
15	16,4583	14,9395	15,6724	15,6901
16	15,4703	15,2915	14,2862	15,0160
17	11,3553	12,9923	15,4082	13,2519
18	12,2683	14,2471	7,6353	11,3836
19	14,1521	14,3123	12,5523	13,6722
20	13,8149	13,9540	18,5553	15,4414

### Conclusões

1. O modelo SAD, para os fatores progênie e parcela, e o modelo multivariado, para o fator resíduo, são as melhores abordagens para a análise dos dados, pois propiciam eficiência e parcimônia em relação ao modelo multivariado completo.

2. A escolha do melhor modelo (SAD), com base nas diferentes estruturas de covariância, é necessária na análise de medidas repetidas para maximizar a eficiência do melhoramento genético da pinha.

3. Com o modelo SAD, é possível uma identificação precisa de famílias superiores em cada colheita e, também, com maior produtividade total de frutos.

### Referências

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions Automatic Control**, v.19, p.716-723, 1974.

ALVES, R.M.; RESENDE, M.D.V. de; BANDEIRA, B. dos S.; PINHEIRO, T.M.; FARIAS, D.C.R. Avaliação e seleção de progênieis de cupuaçuzeiro (*Theobroma grandiflorum*), em Belém, Pará. **Revista Brasileira de Fruticultura**, v.32, p.204-212, 2010.

CECON, P.R.; SILVA, F.F. e; FERREIRA, A.; FERRÃO, R.G.; CARNEIRO, A.P.S.; DETMANN, E.; FARIA, P.N.; MORAIS, T.S. da S. Análise de medidas repetidas na avaliação de clones de

café 'Conilon'. **Pesquisa Agropecuária Brasileira**, v.43, p.1171-1176, 2008.

FARIAS NETO, J.T. de; RESENDE, M.D.V. de; OLIVEIRA, M. do S.P. de. Seleção simultânea em progênieis de açaizeiro irrigado para produção e peso do fruto. **Revista Brasileira de Fruticultura**, v.33, p.532-539, 2011.

GILMOUR, A.R.; CULLIS, B.R.; VERBYLA, A.P. Accounting for natural and extraneous variation in the analysis of field experiments. **Journal of Agricultural, Biological and Environmental Statistics**, v.2, p.269-293, 1997.

GILMOUR, A.R.; CULLIS, B.R.; WELHAM, S.J.; GOGEL, B.J.; THOMPSON, R. An efficient computing strategy for prediction in mixed linear models. **Computational Statistics and Data Analysis**, v.44, p.571-586, 2004.

HENDERSON, C.R. **Applications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462p.

JAFFRÉZIC, F.; PLETCHER, S.D. Statistical models for estimating the genetic basis of repeated measures and other function valued traits. **Genetics**, v.156, p.913-922, 2000.

LITTELL, R.C.; HENRY, P.R.; AMMERMAN, C.B. Statistical analysis of repeated measures using SAS procedures. **Journal of Animal Science**, v.76, p.1216-1231, 1998.

MALHEIROS, E.B. Precisão da análise de dados longitudinais, com diferentes estruturas para a matriz de variâncias e covariâncias, quando se utiliza o esquema em parcelas subdivididas. **Revista de Matemática e Estatística**, v.17, p.263-273, 1999.

MRODE, R.A. **Linear models for the prediction of animal breeding values**. 2<sup>nd</sup> ed. Wallingford: CAB International, 2005. 386p.

NÚÑEZ ANTÓN, V.; ZIMMERMANN, D.L. Modeling nonstationary longitudinal data. **Biometrics**, v.56, p.699-705, 2000.

OLIVEIRA, A.N.; SILVA, A.C. da; ROSADO, S.C. da S.; RODRIGUES, E.A.C. Variações genéticas para características do sistema radicular de mudas de baru (*Dipteryx alata Vog.*). **Revista Árvore**, v.30, p.905-909, 2006.

PLETCHER, S.D.; GEYER, C.J. The genetic analysis of age dependent traits: modeling the character process. **Genetic**, v.153, p.825-833, 1999.

RAO, P.S.R.S. **Variance components: mixed models, methodologies and applications**. Boca Raton: CRC, 1997. 350p.

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

RESENDE, M.D.V. de. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007. 561p.

RESENDE, M.D.V. de; SIMEÃO, R.M.; FERNANDES, J.S.C. Blup individual com medidas repetidas aplicado ao melhoramento de espécies perenes. **Ciência e Agrotecnologia**, v.24, p.986-997, 2000.

RESENDE, M.D.V. de; STURION, J.A. **Análise genética de dados com dependência espacial e temporal no melhoramento**

**de plantas perenes via modelos geoestatísticos e de séries temporais empregando REML/BLUP ao nível individual.** Colombo: Embrapa Florestas, 2001. 80p.

RESENDE, M.D.V. de; THOMPSON, R. **Multivariate spatial statistical analysis of multiple experiments and longitudinal data.** Colombo: Embrapa Florestas, 2003. 126p. (Embrapa Florestas. Documentos, 90).

RESENDE, M.D.V. de; THOMPSON, R.; WELHAM, S. Multivariate spatial statistical analysis of longitudinal data in perennial crops. **Revista de Matemática e Estatística**, v.24, p.147- 169, 2006.

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

---

Recebido em 16 de julho de 2011 e aprovado em 21 de novembro de 2011