

**MODELO LINEAR MISTO PARA AVALIAÇÃO GENÉTICA DE CLONES  
PELA METODOLOGIA DFREML E AM-BLUP  
MIXED LINEAR MODEL FOR THE GENETIC EVALUATION OF CLONES  
RESING THE METHODOLOGY DFREML AND AM-BLUP**

**Resende, M.D.V. de**

Research Scientist-EMBRAPA/Florestas

**RESUMO**

No presente trabalho foram apresentados estimadores e preditores para estimação de componentes de variância e predição de valores genotípicos a partir de testes clonais. Foram considerados dois modelos, associados à avaliação de clones aparentados e não aparentados. A esta classe de modelos denominou-se modelos de clones repetidos (RCM).

**Palavras-chave:** BLUP; REML; dominância, valores genotípicos, seleção.

**ABSTRACT**

In this paper, estimators and predictors for variance components and genotypic values, respectively, were presented aiming to ranking clones. Models associated to evaluation of related and unrelated clones were considered. This class of models was named repeated clone model (RCM).

**Key words:** BLUP; REML; dominance, genotypic values, selection.

**INTRODUÇÃO**

O procedimento "melhor predição linear não viciada" (BLUP) sob um modelo de seleção requer o conhecimento dos componentes de variância antes da seleção, ou seja, na população base não selecionada. Estimativas de componentes de variância obtidos a partir de análise de variância (quadrados mínimos) são viciados pela seleção. Entretanto, os procedimentos iterativos máxima verossimilhança (ML) e máxima verossimilhança restrita (REML), os quais iteragem nas equações de modelo misto (EMM) do procedimento BLUP são livres do vício devido à seleção, embora o ML seja viciado devido à estimação dos efeitos fixos (GRASER *et al.*, 1987).

Dessa forma, atualmente, o procedimento padrão para estimação de componentes de variância e predição de valores genéticos é o

REML-BLUP. A realização simultânea e de forma dependente dessas duas etapas, originou o termo denominado "avaliação genética". Avaliação genética envolve todos os conceitos necessários à tomada de decisão na seleção, por parte dos melhoristas, ou seja, covariância genética entre parentes, estimação de componentes de covariância e parâmetros genéticos (herdabilidades, repetibilidades e correlações genéticas), estimação de médias (efeitos fixos), predição de valores genéticos (efeitos aleatórios), estimação da acurácia seletiva, do progresso genético e da tendência genética.

Três classes de modelos mistos visando à avaliação genética de candidatos à seleção foram propostos por melhoristas de animais. O "modelo de reprodutor" (sire model-SM) foi proposto por HENDERSON (1973) e baseia-se na avaliação do mérito genético aditivo de parentais, transmissíveis às suas filhas (o mérito é função de 1/4 da variação genética aditiva). QUAAS & POLLAK (1980) propôs o "modelo animal" (animal model-AM), o qual baseia-se na predição de valores genéticos aditivos de todos os indivíduos (parentais ou não parentais) envolvidos na avaliação genética. Também QUAAS & POLLAK (1980) desenvolveram o "modelo animal reduzido" (reduced animal model-RAM) em que reduz-se o número de equações à quantidade igual ao número de parentais e posteriormente prediz-se também o valor genético aditivo dos indivíduos não parentais.

Na avaliação genética de clones não se tem interesse apenas no mérito genético aditivo (valores genéticos) dos indivíduos mas sim, no mérito genético total (valores genotípicos). Também não se tem interesse na avaliação genética de todas as árvores (clones e rametes) do experimento (teste clonal), mas apenas no mérito genético de cada clone com base na avaliação de todos os seus rametes. Assim, o modelo SM não se adequa à situação devido

ao primeiro aspecto ressaltado e os modelos **AM** e **RAM** não se adequam devido ao segundo aspecto.

Por outro lado, os modelos **AM** e **RAM** permitem a incorporação do efeito aleatório de dominância (HENDERSON, 1985; KENNEDY, 1990; KENNEDY & SCHAEFFER, 1990), ou de um efeito aleatório adicional representando os efeitos não aditivos mais o efeito de ambiente permanente (HENDERSON, 1982). Neste último caso, o modelo é adequado à situação em que se tem medidas repetidas em um mesmo indivíduo e, portanto, são considerados conjuntamente os parâmetros herdabilidade e repetibilidade. Entretanto, no caso de testes clonais, trata-se de medidas repetidas em um mesmo genótipo e não no mesmo indivíduo (fenótipo) e portanto não existe o efeito de ambiente permanente afetando as medidas repetidas, de forma que o modelo de repetibilidade também não é adequado.

Com base no exposto, o presente trabalho tem por objetivo desenvolver um procedimento de avaliação genética de clones através do desenvolvimento de um novo modelo.

## 2. MATERIAL E MÉTODOS

### 2.1. Modelo misto

O novo modelo será denominado modelo de clones repetidos (**RCM**). O modelo linear misto univariado do **RCM** é:

$$Y = Xb + Z_1 A + Z_1 D + Z_2 e + d$$

onde:

- Y** - vetor dos dados (**ncb x 1**)
- b** - vetor de efeitos fixos (média geral e de blocos [(b+1) x 1])
- A** - vetor aleatório de valores genéticos aditivos (**c x 1**)
- D** - vetor aleatório de efeitos genéticos de dominância (**c x 1**)
- e** - vetor aleatório dos efeitos ambientais de parcela (**cb x 1**)
- d** - residuo aleatório
- Z<sub>1</sub>** - matriz de incidência, associando os vetores **a** e **d**, ao vetor dos dados (**Y**) (**c x ncb**)
- Z<sub>2</sub>** - matriz de incidência, associando o vetor **e** ao vetor dos dados (**Y**) (**cb x ncb**)
- n**, **b** e **c** - referem-se ao número de plantas por parcela, de blocos e de clones, respectivamente.

O modelo apresenta os seguintes momentos:

$$E \begin{bmatrix} Y \\ A \\ D \\ e \\ d \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; V \begin{bmatrix} Y \\ A \\ D \\ e \\ d \end{bmatrix} = \begin{bmatrix} V & Z_1 A_1 (1/\lambda_1) & Z_1 D_1 (1/\lambda_2) & Z_2 I (1/\lambda_3) & I (1/\lambda_4) \\ A_1 (1/\lambda_1) Z_1' & A_1 (1/\lambda_1) & 0 & 0 & 0 \\ D_1 (1/\lambda_2) Z_1' & 0 & D_1 (1/\lambda_2) & 0 & 0 \\ I (1/\lambda_3) Z_2' & 0 & 0 & I (1/\lambda_3) & 0 \\ I (1/\lambda_4) & 0 & 0 & 0 & I (1/\lambda_4) \end{bmatrix} (\sigma_d^2 + \sigma_e^2)$$

onde:

$$V = [Z_1' A_1 (1/\lambda_1) Z_1' + Z_1 D_1 (1/\lambda_2) Z_1' + Z_2 I (1/\lambda_3) Z_2' + I (1/\lambda_4)] (\sigma_d^2 + \sigma_e^2)$$

$$V = Z_1 A_1 Z_1' \sigma_A^2 + Z_1 D_1 Z_1' \sigma_D^2 + Z_2 I Z_2' \sigma_e^2 + I \sigma_d^2$$

onde:

- I** - matriz identidade de ordem apropriada
- A<sub>1</sub>** - matriz de parentesco genético aditivo entre os clones
- D<sub>1</sub>** - matriz de parentesco genético de dominância entre os clones.

$$\lambda_1 = \frac{\sigma_d^2 + \sigma_e^2}{\sigma_A^2}; \lambda_2 = \frac{\sigma_d^2 + \sigma_e^2}{\sigma_D^2}; \lambda_3 = \frac{\sigma_d^2 + \sigma_e^2}{\sigma_e^2}; \lambda_4 = \frac{\sigma_d^2 + \sigma_e^2}{\sigma_d^2}$$

$\sigma_A^2, \sigma_D^2, \sigma_o^2, \sigma_d^2$  - variâncias genética aditiva, genética de dominância, ambiental a nível de parcelas e ambiental a nível de indivíduo, respectivamente.

As EMM para estimação de  $\hat{b}, \hat{A}, \hat{D}, \hat{e}$  equivalem:

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + A^{-1}\lambda_1 & Z_1'Z_1 & Z_1'Z_2 \\ Z_1'X & Z_1'Z_1 & Z_1'Z_1 + D_1^{-1}\lambda_2 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_1 & Z_2'Z_2 + 1\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{A} \\ \hat{D} \\ \hat{e} \end{bmatrix} = \begin{bmatrix} X'Y' \\ Z_1'Y \\ Z_1'Y \\ Z_2'Y \end{bmatrix}$$

Verifica-se que o modelo RCM é similar ao modelo SM no que diz respeito à utilização de indivíduos (subclasse) dentro de acesso (classe). Isto pode ser verificado através da matriz  $Z_1$  que apresenta dimensão equivalente a um modelo SM e não é similar a  $I$ , como no modelo AM. Entretanto, o modelo RCM apresenta efeitos adicionais. Pelo modelo RCM, o mérito genético total ( $\hat{g}$ ) é dado por  $\hat{g} = \hat{A} + \hat{D}$ .

## 2.2. Matrizes de parentesco genético aditivo e de dominância

As matrizes  $A_1$  e  $D_1$  são matrizes quadradas que contemplan os parentescos aditivos e de dominância entre os clones avaliados. Se dois clones são irmãos germanos o parentesco entre eles na matriz  $A_1$  é 1/2 e na matriz  $D_1$  é 0. Uma vez que são necessárias as matrizes  $A_1^{-1}$  e  $D_1^{-1}$ , recomenda-se a inclusão dos parentais dos clones na avaliação genética, pois isto permite a computação direta de  $A_1^{-1}$  sem necessidade de se escrever  $A_1$  (ver RESENDE & PRATES, 1995). Para evitar a inversão de  $D_1$ , as EMM podem ser simplificadas para:

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + A_1^{-1}\lambda_1 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + 1\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{A} \\ \hat{e} \end{bmatrix} = \begin{bmatrix} X'Y' \\ Z_1'Y \\ Z_2'Y \end{bmatrix}$$

onde:

$$K = (I + D_1 A_1^{-1} \sigma_D^2 / \sigma_A^2)$$

Neste caso a solução para  $\hat{D}$  é obtida por  $\hat{D} = D_1 A_1^{-1} \hat{A} \sigma_D^2 / \sigma_A^2$  (HENDERSON, 1984; KENNEDY & SCHAEFFER, 1990). Este procedimento requer a computação de  $D_1$  e  $A_1^{-1}$ , mas evita a computação de  $D_1^{-1}$ .

O procedimento descrito é adequado à populações não endógamas.

Para o caso em que os clones não são aparentados, o modelo misto RCM assume a forma:

$$Y = Xb + Z_1(A + D) + Z_2e + d \text{ e as EMM equivalem a:}$$

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + I\lambda & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{A} + \hat{D} \\ \hat{e} \end{bmatrix} = \begin{bmatrix} X'Y \\ Z_1'Y \\ Z_2'Y \end{bmatrix}$$

onde:

$$\lambda = \frac{\sigma_d^2 + \sigma_e^2}{\sigma_A^2 + \sigma_D^2}$$

Neste caso, não são necessárias as matrizes de parentesco e o modelo torna-se ainda mais similar ao modelo SM, mas diferenciado devido à consideração do mérito genético total.

### 2.3. Estimação de componentes de variância pelo método da máxima verossimilhança restrita

Os algoritmos mais utilizados na estimação por REML são o EM (maximização de esperanças) e o DF (livre de derivados ou não derivativo), propostos por DEMPSTER et al. (1977) e GRASER et al. (1987), respectivamente.

O algoritmo EM-REML para o modelo RCM, apresenta os seguintes estimadores:

$$\hat{\sigma}_d^2 = (Y'Y - \hat{b}'X'Y - \hat{A}'Z_1'Y - \hat{D}'Z_2'Y - \hat{e}'Z_2'Y) / [ncb - r(x)]$$

$$\hat{\sigma}_e^2 = (\hat{e}'e + \sigma_d^2 \text{tr } C_{ee}) / cb$$

$$\hat{\sigma}_A^2 = (\hat{A}'\hat{A} + \sigma_d^2 \text{tr } (A_1^{-1}C_{AA})) / c$$

$$\hat{\sigma}_D^2 = (\hat{D}'\hat{D} + \sigma_d^2 \text{tr } (D_1^{-1}C_{DD})) / c$$

onde:

$C_{ee}$ ,  $C_{AA}$ ,  $C_{DD}$ , representam as partições da inversa da matriz dos coeficientes, referentes aos efeitos de parcela, do valor genético aditivo e de dominância, respectivamente.  $r(x)$  é oposto da matriz  $x$ .

Para o modelo simplificado, adequado à avaliação de clones são apresentados os estimadores EM-REML:

$$\hat{\sigma}_d^2 = (Y'Y - \hat{b}'X'Y - (\hat{A} + \hat{D})'Z_1'Y - \hat{e}'Z_2'Y) / [ncb - r(x)]$$

$$\hat{\sigma}_e^2 = (\hat{e}'e + \sigma_d^2 \text{tr } C_{ee}) / cb$$

$$(\hat{\sigma}_A^2 + \hat{\sigma}_D^2) = [\hat{A} + \hat{D}]'(\hat{A} + \hat{D}) + \sigma_d^2 \text{tr } C_{(A+D)(A+D)} / c$$

onde:

$C_{(A+D)(A+D)}$  - refere-se à partição da inversa da matriz dos coeficientes, referente ao mérito genético total.

O algoritmo DFREML para o modelo RCM (mais de um fator aleatório além do resíduo),

baseia-se na avaliação da razão  $\theta$  entre o componente de variância e a variância fenotípica (MEYER, 1989). No presente caso necessitam ser avaliados os parâmetros  $\theta_A$ ,  $\theta_D$  e  $\theta_B$ , relativos a  $\sigma_A^2$ ,  $\sigma_D^2$  e  $\sigma_e^2$ , respectivamente. Assim, para a determinação do máximo da função de verossimilhança é necessário avaliar apenas a função:

$$-2 \log_e = \text{const} + \mathbf{y}'\mathbf{P}\mathbf{y}|\sigma_d^2 + \log_e |\mathbf{C}^*| + [(\text{ncb} - r(\mathbf{x}) - 2\mathbf{c} - \mathbf{cb})] \\ \log_e \sigma_d^2 + \mathbf{c}(\log_e \sigma_A^2 + \log_e \sigma_D^2) + \mathbf{cb}(\log_e \sigma_e^2 + \log_e |\mathbf{A}_1| + \log_e |\mathbf{D}_1|)$$

Para o modelo simplificado necessitam ser avaliados os parâmetros  $\theta_{A+D}$  e  $\theta_e$ , associados a  $(\sigma_A^2 + \sigma_D^2)$  e  $\sigma_e^2$ , respectivamente. Neste caso, para a determinação do máximo da função de verossimilhança, basta avaliar a função:

$$-2 \log_e L = \text{const} + \mathbf{y}'\mathbf{P}\mathbf{y}|\sigma_d^2 + \log_e |\mathbf{C}^*| + [(\text{ncb} - r(\mathbf{x}) - \mathbf{c} - \mathbf{cb})] \\ \log_e \sigma_d^2 + \mathbf{c} \log_e (\sigma_A^2 + \sigma_D^2) + \mathbf{cb} \log_e \sigma_e^2$$

onde:

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} - \text{matriz de projeção ortogonal.}$$

$|\mathbf{C}^*|$  - determinante de uma submatriz não singular de  $\mathbf{C}$  com posto máximo, onde  $\mathbf{C}$  é a matriz dos coeficientes das EMM.

As constantes  $\log_e |\mathbf{A}_1| + \log_e |\mathbf{D}_1|$  não dependem dos parâmetros a serem estimados e portanto não necessitam ser calculados visando maximizar o  $\log_e L$ . Procedimentos para avaliação  $\theta_A$ ,  $\theta_D$ ,  $\theta_e$  e  $\theta_{A+D}$  foram descritos por MEYER (1989).

### 3. CONSIDERAÇÃO FINAL

O modelo apresentado é extremamente útil do ponto de vista prático, especialmente para avaliação de clones aparentados, fato comum em gerações avançadas de melhoramento.

### 4. REFERÊNCIAS BIBLIOGRÁFICAS

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

GRASER, H. V.; SMITH, S. P.; TIER, B. A derivative free approach for estimating variance components in animal models by restricted maximum likelihood. Journal of Animal Science, Champaign, v.64, n.5, p.1362-1370, 1987.

HENDERSON, C. R. Sire evaluation and genetic trends. In: Animal Breeding and Genetics Symposium in Honor of J. Lush. American Society of Animal Science, Champaign, III. 1973. p.10-41.

HENDERSON, C. R. Avaliação de vacas e touros. In: SIMPÓSIO BRASILEIRO DE MELHORAMENTO GENÉTICO DE BOVINOS, 1. Coronel Pacheco, 1982. Anais. Coronel Pacheco, EMBRAPA/CNPGC, 1982. p.137-168.

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

HENDERSON, C. R. Best linear unbiased prediction of nonadditive genetic merit in noninbred populations. Journal of Animal Science, Champaign, v.60, n.1, p.111-117, 1985.

KENNEDY, B.W. Use of mixed model methodology in analysis of designed experiments. In: GIANOLA, D.; HAMMOND, K. Advances in statistical methods for genetic improvement of livestock. Springer-Verlag, Berlin, 1990, p.77-97.

KENNEDY, B.W.; SCHAEFFER, L.R. Reproductive technology and genetic evaluation. In: GIANOLA, D.; HAMMOND, K. Advances in statistical methods for genetic improvement of livestock. Springer-Verlag, Berlin, 1990, p.507-532.

MEYER, K. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genetique, Selection, Evolution, v.21, p.317-340, 1989.

QUAAS, R.L.; POLLAK, E.J. Mixed model methodology for farm and ranch beef cattle testing programs. Journal of Animal Science, v.51, p.1277-1287, 1980.

RESENDE, M.D.V.; PRATES, D.F. Avaliação genética de Pinus através das metodologias REML livre de derivados (DFREML) e BLUP sob modelo animal (AM-BLUP). Floresta, 1995 (no prelo).