

ESTIMAÇÃO DE COMPONENTES DE VARIÂNCIA E PREDIÇÃO DE VALORES GENÉTICOS PELO MÉTODO DA MÁXIMA VEROSSIMILHANÇA RESTRITA (REML) E MELHOR PREDIÇÃO LINEAR NÃO VICIADA (BLUP) EM *Pinus*

Marcos Deon Vilela de Resende^{*}
Douglas Fabrício Prates^{**}
Cibele Keiko Yamada^{***}
Andréia de Jesus^{***}

RESUMO

Foram comparados três procedimentos de estimação de componentes de variância visando a predição de valores genéticos: quadrados mínimos (**LS**), máxima verossimilhança (**ML**) e máxima verossimilhança restrita (**REML**). A estimação pelo método **ML** apresentou convergência mais rápida do que pelo método **REML**. As herdabilidades obtidas pelos métodos **ML** e **REML** foram bastante similares e de magnitudes superiores à obtida pelo método **LS**. O procedimento **REML**, embora computacionalmente mais complicado, foi o mais preciso. A seleção e estimação de ganhos genéticos devem ser realizadas a partir do procedimento **REML** iteragindo nas equações de modelo misto (**BLUP**).

PALAVRAS-CHAVE: máxima verossimilhança, inferência estatística, modelos mistos, parâmetros genéticos.

ABSTRACT

VARIANCE COMPONENT ESTIMATION AND PREDICTION OF BREEDING VALUES BY THE RESTRICTED MAXIMUM LIKELIHOOD (REML) AND BEST LINEAR UNBIASED PREDICTION (BLUP) METHODS IN *Pinus*

Three procedures of variance component estimation for the prediction of breeding values were compared: least square (**LS**), maximum likelihood (**ML**), and restricted maximum likelihood (**REML**). The estimation by **ML** showed faster convergence than **REML**. Heritabilities estimated through **ML** and **REML** were similar and superior than those estimated **LS**. **REML** was the most precise but also the most computationally difficult. Selection and estimation of genetic gains must be based on **REML** iterating on mixed model equations (**BLUP**).

* Eng. Agrônomo, Mestre, CREA n^o 50602/D, Pesquisador da Embrapa - Centro Nacional de Pesquisa de Florestas.
** Acadêmico de Estatística, Setor de Ciências Exatas, Universidade Federal do Paraná-UFPR.
*** Acadêmica de Ciência da Computação, Setor de Ciências Exatas, PUC-PR.

KEYWORDS: maximum likelihood, statistical inference, mixed models, genetic parameters.

1. INTRODUÇÃO

A predição de valores genéticos e os métodos de seleção dependem, essencialmente, de estimativas de componentes de variância. Os métodos de predição de valores genéticos denominados índice de seleção, melhor predição linear (**BLP**) e melhor predição linear não viciada (**BLUP**) são fundamentados no conhecimento ou estimação precisa dos componentes de variância genética e fenotípica (HENDERSON, 1977; HAYES & HILL, 1980). Esses três métodos são comumente considerados no melhoramento de espécies perenes (RESENDE et al., 1995) e, via de regra, os componentes de variância são estimados pelo método dos quadrados mínimos, tais como os procedimentos de análise de variância (métodos I, II e III) descritos por HENDERSON (1953). Estes procedimentos (de solução explícita) permitem trabalhar com um certo nível de desbalanceamento e se reduzem à usual análise de variância quando os dados são balanceados.

Por outro lado, existe uma quarta situação de predição de valores genéticos, em que as variâncias são desconhecidas e os desbalanceamentos são tais que a estimação pelo método dos quadrados mínimos apresenta eficiência desconhecida. Nestas situações, outros procedimentos de estimação de componentes de variância são demandados e a predição de valores genéticos deve ser realizada simultaneamente à estimação dos componentes de variância.

Em inferência estatística, os procedimentos adequados de estimação devem apresentar as seguintes propriedades desejáveis: não vício, consistência, "completude", suficiência e eficiência. O método de estimação por quadrados mínimos (**LS**), na situação de dados desbalanceados, apresenta a propriedade de não vício, mas suas demais propriedades estatísticas são pouco conhecidas (SEARLE, 1971). Outra grande desvantagem do método **LS** é a possibilidade de obtenção de estimativas negativas de componentes de variância. Assim, recentemente (após 1967) outros processos têm sido considerados para a estimação de componentes de variância, notadamente aqueles ligados à máxima verossimilhança (**ML**). HARTLEY & RAO (1967) derivaram um procedimento de máxima verossimilhança para estimação de variâncias no contexto dos modelos lineares mistos. Este foi, posteriormente, modificado e aplicado por melhoristas animais (HENDERSON, 1973; HILL & NICHOLAS, 1974; SCHAEFFER, 1976; THOMPSON, 1973, 1976, 1979; GIANOLA et al., 1989).

Os estimadores de máxima verossimilhança apresentam as seguintes propriedades desejáveis (KENNEDY, 1981): translação invariante (não afetados por mudanças nos efeitos fixos); suficiência, consistência e eficiência. Outra vantagem da **ML** é a geração de estimativas não negativas de componentes de variância. Por outro lado, os estimadores **ML** são viciados em decorrência de dois aspectos (SHAW, 1987): imposição de restrição de não negatividade e perda de graus de liberdade devido à estimação dos efeitos fixos.

Visando eliminar esta segunda fonte de vício, PATTERSON & THOMPSON (1971) apresentaram uma modificação ao método **ML**, o qual passou a ser denominado de máxima verossimilhança restrita ou "residual" (**REML**) por HARVILLE

(1977). Além do aperfeiçoamento de natureza estatística, o método **REML** mantém as demais propriedades do **ML**. **REML** é equivalente a **ML** para um conjunto de dados que tenham sido padronizados para média zero. A modificação realizada conduz a estimativas idênticas aquelas obtidas por análise de variância, se o delineamento for balanceado e as restrições de não negatividade forem ignoradas. Outros procedimentos (não iterativo, portanto de solução explícita) de estimação de componentes de variância referem-se à estimação quadrática não viciada de norma mínima (MINQUE) (RAO, 1970) e estimação quadrática não viciada de variância mínima (MIVQUE). Conforme relatado para a **REML**, para dados balanceados, o método MINQUE, também conduz a resultados idênticos aos da análise de variância. Entretanto, é computacionalmente mais complicado e teoricamente pouco entendido (THOMPSON, 1979). ANDERSON (1984) relata que os procedimentos MINQUE & MIVQUE são de pouca utilidade prática, pois dependem do conhecimento prévio dos valores dos componentes de variância a serem estimados. Em resumo, os métodos de máxima verossimilhança são mais flexíveis, não exigindo delineamentos balanceados e geralmente conduzindo a maiores eficiências que o método de quadrados mínimos. Além disso, sugere que os estimadores **ML** e **REML** podem ser robustos à desvios da normalidade (HARVILLE, 1977). Assim, o método **REML** tornou-se, recentemente, o principal método de estimação de componentes de variância e tem sido amplamente empregado em melhoramento animal (SCHAEFFER et al., 1978; THOMPSON, 1979; KENNEDY, 1981; MEYER, 1989a; MEYER & HILL, 1991; MEYER, 1991a; THOMPSON & HILL, 1990; SEARLE et al., 1992).

O método **REML** é iterativo e, portanto, demanda grandes recursos computacionais. Por outro lado, programas que economizam computação através da consideração de estruturas particulares de dados tem sido desenvolvidos (MEYER, 1988; MEYER, 1991b; BOLDMAN et al., 1993; GROENEVELD et al., 1990; MARTINS et al., 1994; LOPES, 1994) e a aplicação da **REML** tem sido grande. No Brasil, a aplicação do método iniciou-se com os trabalhos de FREITAS (1991), FREITAS & VENCOSKY (1993), LOPES et al. (1993a; 1993b) e MARTINS et al. (1993).

O presente trabalho tem como objetivos a estimação de componentes de variância e predição de valores genéticos, simultaneamente, através dos métodos **REML** e **BLUP** empregando-se as equações de modelos mistos, para a seleção de progênieis ou parentais superiores de *Pinus maximinoi*.

2. MATERIAL E MÉTODOS

2.1. Modelo misto e estimação pelo método da máxima verossimilhança restrita (REML).

Exceção feita a delineamentos completamente balanceados, o método **REML** demanda procedimentos iterativos, realizados através de equações de modelo misto, pelo procedimento **BLUP** (HENDERSON, 1973, 1975). Assim, o procedimento **REML** sob modelo misto realiza, simultaneamente, as operações de estimação (**BLUE**) dos efeitos fixos por quadrados mínimos generalizados, de estimação de componentes

de variância por **REML** e de predição (**BLUP**) de valores genéticos. Assim é uma ferramenta poderosa e elegante de seleção.

Vários algoritmos computacionais para obtenção de componentes de variância por **REML** como **MS** (Method of scoring, de PATTERSON & THOMPSON, 1971; THOMPSON, 1973; SCHAEFFER et al., 1978); **EM** (Expectation -Maximization, de DEMPSTER et al., 1977); **DFREML** (Derivate-free Restricted Maximum Likelihood, de SMITH & GRASER, 1986 e GRASER et al., 1987), tem sido desenvolvidos visando maior facilidade de processamento (MEYER, 1989a; HARVILLE & CALLANAN, 1990). Dentre estes, os mais usados são o **EM** e **DFREML**. O algoritmo **EM** é muito estável, numericamente, apresentando convergência mesmo que os valores iniciais não tenham sido adequados. Entretanto, uma inconveniência do algoritmo **EM**, é a lentidão para estimativas próximas ao limite do espaço de definição paramétrica (por exemplo, quando uma variância tende a zero). Os algoritmos livres de derivadas (derivative-free) apresentam vantagens computacionais, pois evitam a inversão de matrizes de elevada ordem. Os estimadores obtidos pelo **EM** são idênticos aos de HARVILLE (1977) e MEYER (1989a). Se valores iniciais positivos forem utilizados, a convergência para valores não negativos é garantida (HARVILLE, 1977).

O emprego do algoritmo **EM** para estimação de componentes de variância descrito neste trabalho é similar ao procedimento relatado por KENNEDY (1981).

Considerando o modelo linear mais simples (RESENDE et al., 1995), tem-se:

$$Y = X\beta + Z\mu + \Sigma, \quad (1)$$

onde:

Y é o vetor dos dados, β é o vetor de efeitos fixos, μ é o vetor de valores genéticos, Σ é o vetor de erros e X e Z são matrizes de delineamento ou incidência. Para o caso univariado e sendo $R = \text{Var}(e) = I\sigma_e^2$ e $V = \text{Var}(\mu) = A\sigma_A^2$, tem-se que $V(y) = ZAZ' \sigma_A^2 + I\sigma_e^2$; β e μ podem ser estimados a partir das seguintes equações de modelo misto:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}(\sigma_e^2 / \sigma_A^2) \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix}, \quad (2)$$

onde: I é a matriz identidade, A é o numerador da matriz de parentesco entre os indivíduos (progênies) avaliados, σ_e^2 é a variância ambiental e σ_A^2 é função da variância genética aditiva. Assim, desde que σ_A^2 e σ_e^2 ou a herdabilidade ($h^2 = \sigma_A^2 / (\sigma_A^2 + \sigma_e^2)$) sejam conhecidos, β e μ podem ser estimados a partir de (2).

Sendo A a matriz identidade, o método iterativo em **BLUP** (através das equações de modelo misto) segue os seguintes passos:

a) Partindo-se de um valor inicial (estimado previamente ou assumido) de h^2 (ou σ_e^2 / σ_A^2), obtém-se μ e β a partir de (2).

(b) Obtém-se σ_A^2 e σ_e^2 a partir das seguintes expressões (HARVILLE, 1977;

MEYER, 1989a):

$$\hat{\sigma}_e^2 = [Y'Y - \hat{\beta}'X'Y - \hat{\mu}'Z'Y] / [N - r(X)]$$

$$\hat{\sigma}_A^2 = [\hat{\mu}'\hat{\mu} + \sigma_e^2 \text{ tr } C_{22}] / q,$$

onde:

N = número total de observações.

$r(X)$ = posto de X ou número de colunas linearmente independentes de X .

C_{22} é da forma

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{11} & C_{22} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}(\sigma_e^2 / \sigma_A^2) \end{bmatrix}^{-1}$$

q = número de elementos aleatórios (nº. de indivíduos ou progênies ou valores genéticos a serem preditos).

tr = operação traço matricial ou soma dos elementos da diagonal da matriz.

(c) Substitui-se em (2), $\hat{\sigma}_e^2$ e $\hat{\sigma}_A^2$ obtidos em (b) e obtém-se novos valores para

$\hat{\beta}$ e $\hat{\mu}$.

(d) Entra-se novamente em (b), usando $\hat{\beta}$, $\hat{\mu}$ e $\hat{\sigma}_e^2 / \hat{\sigma}_A^2$ e se repete o processo até a convergência.

Na convergência, examinam-se os parâmetros $\hat{\beta}$, $\hat{\mu}$, $\hat{\sigma}_A^2$ e $\hat{\sigma}_e^2$. A velocidade de convergência é tanto maior quanto mais acertada (próxima da verdadeira) for a informação inicial utilizada.

Na maioria dos casos, $\mathbf{A}\sigma_A^2$ e $I\sigma_e^2$ em (2) devem ser substituídos por matrizes de variância-covariância \mathbf{G} e \mathbf{R} , pois não existe homogeneidade de variâncias genéticas e ambientais associadas aos diferentes dados.

Na predição de valores genéticos de vários caracteres, \mathbf{G} e \mathbf{R} (matrizes não singulares) encerram também covariâncias entre caracteres e os requerimentos computacionais são enormes. Assim, procedimentos que implicam em redução de computações tem sido desenvolvidos, tais como a transformação do **REML** multivariado para univariado, através do método denominado transformação canônica (THOMPSON & HILL, 1990; FREITAS et al., 1994).

2.2. Aplicação a dados experimentais

Consideram-se neste estudo, dados referentes a um teste de procedência e progênes de *Pinus maximoi* (material genético pertencente à rede experimental da CAMCORE - Central America and Mexico Coniferous Resources), obtidos a partir de SAMPAIO (1995). O experimento foi conduzido no delineamento de blocos casualizados com nove repetições e seis plantas por parcela. Foram avaliadas oito populações e 102 progênes de polinização aberta (meios-irmãos) em dois experimentos distintos em uma mesma fazenda, conforme Tabela 1.

TABELA 1. Estrutura experimental associada aos dados empregados no estudo.

Experimento 1		Experimento 2	
Procedências	Progênes	Procedências	Progênes
1	1-16	2	64-71 (+ 6 comuns)
2	17-31	3	72-82 (+ 2 comuns)
3	32-45	6	83-89
4	46-57	7	90-96
5	58-63	8	97-102

Visando a seleção de progênes ou parentais, trabalhou-se a nível de médias de progênes em cada experimento. Conforme modelo linear apresentado por RESENDE et al. (1995) define-se o vetor aleatório μ a ser predito, o vetor fixo β e componentes de variância a serem estimados:

$$\beta' = [\mu, \ell_1, \ell_2, P_1, \dots, P_8]; \mu' = [VA_{11}, \dots, VA_{102}]; \text{Var} \begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} I\sigma_A^2 & 0 \\ 0 & I\sigma_e^2 \end{pmatrix};$$

σ_A^2 = variância dos valores genéticos aditivos de progênes. Equivale a ¼ da variância genética aditiva;

σ_e^2 = variância residual ao nível de médias de parcela e equivalente a $\frac{\sigma_d^2}{nb} + \frac{\sigma^2}{b}$, onde σ_d^2 é a variância dentro de parcela, σ^2 é a variância entre parcelas e **n** e **b**, são os números de plantas por parcelas e de blocos, respectivamente.

Define-se ainda:

- μ = efeito fixo da média geral.
- ℓ_1 = efeito fixo do experimento 1.
- ℓ_2 = efeito fixo do experimento 2.
- P_i = efeito da procedência i.
- V_{Ai} = valor genético associado à progênie i, integrante do vetor μ de valores genéticos

Nesta situação tem-se a seguinte estrutura de variâncias dos dados: $\text{VAR}(Y) = V = ZUZ' + R$. Uma vez aceita a condição de homogeneidade de variâncias genéticas e ambientais, tem-se $V = ZAZ' \sigma_A^2 + I \sigma_c^2$. Mas como as progênes não são aparentadas entre si, $V = ZIZ' \sigma_A^2 + I \sigma_c^2 = ZZ' \sigma_A^2 + I \sigma_c^2$. Assim, o sistema de equações lineares mais adequado a esta situação é definido por:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I(\sigma_c^2 / \sigma_A^2) \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix} \quad (3)$$

sendo

$\hat{\beta}$ e $\hat{\mu}$ obtidos a partir da solução deste sistema. Solução equivalente é estimar $\hat{\beta} = (X'X)^{-1} X'Y$ e $\hat{\mu} = [Z'Z + I(\sigma_c^2 / \sigma_A^2)]^{-1} Z'(Y - X\hat{\beta})$, onde, com homogeneidade de variâncias, a estimação GLS de β equivale à OLS. A relação $\lambda = \sigma_c^2 / \sigma_A^2 = \frac{1-h^2}{h^2}$ pode ser determinada a partir da herdabilidade (h^2). Neste caso, h^2 a nível de médias de família foi o parâmetro inicial usado no processo iterativo. $\lambda = 0,6294$ foi a estimativa prévia utilizada, tendo sido obtida pelo método de quadrados mínimos ordinários (Modelo III de HENDERSON, 1953).

Após a convergência, os valores finais de $\hat{\beta}$ e $\hat{\mu}$ foram utilizados na predição dos valores genéticos das progênie através da expressão $\hat{g}_{iK} = \hat{\beta}_i + \hat{\mu}_K$, onde $\hat{\beta}_i$ é o efeito da procedência i e $\hat{\mu}_K$ é o efeito da progênie K da procedência i , conforme RESENDE et al. (1995). Os resultados foram apresentados em termos de $\hat{t} = \hat{\mu} + g_{iK}$.

2.3. Comparação de metodologias de estimação de componentes de variância

Para efeito de comparação com o procedimento REML, foram preditos valores genéticos considerando variâncias conhecidas (na realidade usando $\lambda=0,6294$, conforme estimativa obtida pelo método III de Henderson) e, também, estimadas pelo método da máxima verossimilhança (ML) irrestrita. Segundo o algoritmo EM, os estimadores de σ_A^2 e σ_c^2 por ML foram (SCHAEFFER, 1976):

$$\left. \begin{aligned} \sigma_c^2 &= [Y'Y - \hat{\beta}'X'Y - \hat{\mu}'Z'Y] / N \\ \sigma_A^2 &= \left[\mu'\hat{\mu} + \sigma_c^2 \operatorname{tr} [Z'Z + I(\sigma_c^2 / \sigma_A^2)]^{-1} \right] / q \end{aligned} \right\}$$

As computações foram realizadas empregando-se o software MATLAB (Matrix

Laboratory) desenvolvido pela Universidade de Stanford-USA (MATLAB,1992).

O critério de convergência utilizado foi de 4 casas decimais para os parâmetros λ e h^2 e 7 casas decimais, para os componentes σ_A^2 e σ_e^2 . O número de decimais foi observado através da diferença entre o resultado obtido em uma iteração e na anterior.

3. RESULTADOS E DISCUSSÃO

3.1. Estimação de componentes de variância e convergência

As estimativas dos componentes de variância e herdabilidade ao nível de média de famílias obtidas pelos métodos iterativos **ML** e **REML** estão apresentadas nas Tabelas 1 e 2, respectivamente. O critério de convergência adotado produziu resultados concordantes com os obtidos por LOPES et al. (1994), os quais concluíram que o uso de 4 decimais é adequado. Por outro lado, SCHENKEL (1991) usou, como critério de convergência, a diferença entre os valores genéticos preditos em uma determinada iteração e na anterior, implicando em um maior número de iterações.

TABELA 1. Estimativas de componentes de variância e herdabilidade para o caráter volume de madeira em *Pinus maximinoi*, obtidas através do método da máxima verossimilhança restrita (REML)*.

Nº da Iteração	$\sigma_A^2 (x10^{-4})$	$\sigma_e^2 (x10^{-4})$	λ	h^2
1	4,1960	2,6410	0,6294	0,6137
2	4,8373	3,0200	0,6243	0,6156
3	4,8531	3,0059	0,6194	0,6175
4	4,8685	2,9922	0,6146	0,6193
5	4,8836	2,9788	0,6100	0,6211
6	4,8982	2,9657	0,6055	0,6229
...
36	5,1307	2,7609	0,5381	0,6502
37	5,1384	2,7542	0,5360	0,6510
...
57	5,2175	2,6854	0,5147	0,6602
58	5,2222	2,6813	0,5135	0,6607
...
95	5,3328	2,5859	0,4849	0,6734
96	5,3332	2,5856	0,4848	0,6735
97	5,3336	2,5855	0,4848	0,6735

* Simbologia conforme definida no material e métodos.

TABELA 2. Estimativas de componentes de variância e herdabilidade para o caráter volume de madeira em *Pinus maximinoi*, obtidas através do método da máxima verossimilhança (ML)*.

Nº da Iteração	$\sigma_A^2 (\times 10^{-4})$	$\sigma_c^2 (\times 10^{-4})$	λ	h^2
1	4,1960	2,6410	0,6294	0,6137
2	4,4423	2,7729	0,6242	0,6157
3	4,4571	2,7597	0,6192	0,6176
4	4,4714	2,7469	0,6143	0,6195
5	4,4855	2,7343	0,6096	0,6213
6	4,4991	2,7221	0,6050	0,6231
⋮	⋮	⋮	⋮	⋮
80	4,8812	2,3860	0,4868	0,6717
81	4,8816	2,3857	0,4887	0,6717
82	4,8819	2,3854	0,4886	0,6718
⋮	⋮	⋮	⋮	⋮
87	4,8837	2,3832	0,4881	0,6720
88	4,8840	2,3836	0,4880	0,6720
89	4,8844	2,3833	0,4880	0,6720

* Simbologia conforme definida no material e métodos.

TABELA 3. Valores genéticos preditos através de três procedimentos:

\hat{t}_1 - considerando variâncias conhecidas; \hat{t}_2 - usando variâncias estimadas pelo método ML; \hat{t}_3 - usando variâncias estimadas pelo método REML.

Progênes	\hat{t}_1	\hat{t}_2	\hat{t}_3
1	0,2265	0,2267	0,2268
2	0,2247	0,2248	0,2248
3	0,2165	0,2159	0,2159
4	0,2099	0,2086	0,2086
5	0,1788	0,1746	0,1745
6	0,2668	0,2709	0,2710
7	0,2161	0,2154	0,2154
8	0,2190	0,2186	0,2186
9	0,2185	0,2180	0,2180
10	0,2450	0,2471	0,2471
11	0,2358	0,2369	0,2370
12	0,2077	0,2062	0,2062
13	0,2051	0,2033	0,2033
14	0,2304	0,2310	0,2311
15	0,2518	0,2545	0,2545
16	0,2219	0,2218	0,2218
17	0,2251	0,2250	0,2250
18	0,2218	0,2470	0,2214
19	0,2183	0,2214	0,2208
20	0,2818	0,2208	0,2625
21	0,2144	0,2624	0,2133
22	0,2355	0,2233	0,2365
23	0,1961	0,2365	0,2062
24	0,2398	0,2063	0,2412
25	0,2091	0,2411	0,2075
26	0,2252	0,2075	0,2253
27	0,2182	0,2253	0,2175
28	0,2359	0,2323	0,2323
29	0,2238	0,1920	0,2245
30	0,2459	0,2630	0,2478
31	0,1949	0,2483	0,1919
32	0,2608	0,2397	0,2631
33	0,2538	0,2524	0,2483
34	0,2395	0,2411	0,2397
35	0,2511	0,2626	0,2524
36	0,2408	0,2612	0,2412
37	0,2592	0,2390	0,2513
38	0,2397	0,2193	0,2390
39	0,2209	0,2349	0,2192
40	0,2351	0,2354	0,2349
41	0,2356	0,2409	0,2354

Progênie	\hat{t}_1	\hat{t}_2	\hat{t}_3
42	0,2406	0,2553	0,2409
43	0,2538	0,2227	0,2553
44	0,2240	0,2274	0,2226
45	0,2283	0,2137	0,2274
46	0,2135	0,2362	0,2137
47	0,2340	0,2237	0,2362
48	0,2227	0,2240	0,2238
49	0,2229	0,1872	0,2241
50	0,1894	0,2522	0,1872
51	0,1723	0,1686	0,1685
52	0,2069	0,2064	0,2064
53	0,2373	0,2397	0,2398
54	0,1973	0,1959	0,1959
55	0,1882	0,1871	0,1870
56	0,2326	0,2346	0,2347
57	0,2210	0,2219	0,2219
58	0,2320	0,2357	0,2358
59	0,1730	0,1711	0,1711
60	0,2017	0,2025	0,2025
61	0,1862	0,1855	0,1855
62	0,1930	0,1930	0,1930
63	0,1727	0,1707	0,1707
64	0,2059	0,2040	0,2039
65	0,2293	0,2296	0,2296
66	0,2310	0,2316	0,2316
67	0,2222	0,2219	0,2219
68	0,2176	0,2168	0,2168
69	0,2212	0,2208	0,2208
70	0,2494	0,2516	0,2517
71	0,2266	0,2267	0,2267
72	0,2596	0,2617	0,2618
73	0,1937	0,1896	0,1895
74	0,2468	0,2477	0,2477
75	0,2247	0,2234	0,2234
76	0,2093	0,2066	0,2065
77	0,2143	0,2121	0,2120
78	0,2500	0,2512	0,2512
79	0,2357	0,2356	0,2355
80	0,2238	0,2225	0,2224
81	0,2541	0,2556	0,2556
82	0,2472	0,2481	0,2482
83	0,2217	0,2199	0,2199
84	0,2434	0,2437	0,2437
85	0,2383	0,2382	0,2382
86	0,2454	0,2459	0,2460
87	0,2604	0,2623	0,2624
88	0,2394	0,2393	0,2393
89	0,2308	0,2299	0,2299

Progênes	\hat{t}_1	\hat{t}_2	\hat{t}_3
90	0,2286	0,2277	0,2277
91	0,2580	0,2599	0,2599
92	0,2385	0,2385	0,2385
93	0,2223	0,2208	0,2208
94	0,2276	0,2266	0,2266
95	0,2449	0,2456	0,2456
96	0,2447	0,2454	0,2454
97	0,2011	0,1991	0,1991
98	0,2280	0,2285	0,2286
99	0,2178	0,2174	0,2174
100	0,2177	0,2173	0,2173
101	0,2532	0,2561	0,2562
102	0,2148	0,2142	0,2141

Entretanto, foram necessárias 89 e 97 iterações para se obter uma precisão de 10^{-4} nas estimativas de herdabilidade, pelos métodos **ML** e **REML**, respectivamente. Utilizando o algoritmo **EM**, FREITAS & VENCOSKY (1993) necessitaram de 40 iterações para a convergência pelos métodos **ML** e **REML** e JENSEN & MAO (1988) necessitaram de mais de 100 iterações para a convergência pelo método **REML**.

O método **ML** foi mais vantajoso pois demandou menor número de iterações, além de ser computacionalmente mais simples que o **REML**. Ambos convergiram para valores similares de herdabilidade (0,6720 para **ML** e 0,6735 para **REML**). Entretanto, o método **ML** tem a desvantagem de não considerar a perda de graus de liberdade na estimação dos efeitos fixos do modelo linear misto. Isto subestima o componente σ_e^2 , conduzindo também, à subestimação do componente σ_A^2 . Estas subestimações são tanto maiores quanto maior o número (posto de **X**) de equações independentes para os efeitos fixos em relação ao número (**N**) total de observações. Como neste caso o posto de **X** é pequeno em relação a **N**, ambos os métodos conduziram a resultados similares, embora o método **REML** seja mais preciso. FREITAS & VENCOSKY (1993) também obtiveram resultados similares por esses métodos.

Em comparação com o método tradicional de estimação por quadrados mínimos ($h^2 = 0,6137$), verifica-se que a herdabilidade associada ao conjunto de dados foi maior. Entretanto, devido ao pequeno desbalanceamento, a diferença entre as duas estimativas (por quadrados mínimos e por **REML**) foi de pequena magnitude. Sob completo balanceamento, os dois métodos conduzem a resultados idênticos, desde que não sejam impostas restrições de não negatividade ao método **REML** (ANDERSON, 1979).

3.2. Predição de valores genéticos

Nas comparações entre os procedimentos de predição de valores genéticos: (i) com variâncias conhecidas (estimadas pelo método de quadrados mínimos (\hat{t}_1)); (ii)

com variâncias estimadas pelo método **ML** (\hat{t}_2); (iii) com variâncias estimadas pelo método **REML** (\hat{t}_3); verificou-se que para as melhores progênes, os métodos **ML** e **REML** sob **BLUP** conduziram a maiores valores genéticos preditos, em relação ao método de quadrados mínimos. Portanto, aceitar as variâncias como conhecidas, quando na realidade elas não eram, conduziram a subestimativas de ganhos genéticos.

Dentre os procedimentos de predição estudados, o mais acurado foi o **REML** sob **BLUP**. Portanto, o mesmo deve ser utilizado para seleção e estimação dos ganhos genéticos em *Pinus maximinoi*.

4. CONCLUSÕES

- A estimação de componentes de variância pelo método **ML** apresenta convergência mais rápida do que pelo método **REML**.
- As estimativas de herdabilidade obtidas pelos métodos iterativos **ML** e **REML** são similares, porém, de magnitudes superiores a obtida pelo método explícito de quadrados mínimos.
- O procedimento **ML**, embora viciado, é satisfatório, na presente situação.
- O procedimento **REML**, embora computacionalmente mais complicado, é o mais preciso.
- A seleção e estimação de ganhos genéticos devem ser realizadas a partir do procedimento **REML** iteragindo nas equações de modelo misto (**BLUP**).

5. REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, R.D. Estimating variance components from balanced data: Optimum properties of REML solutions and MIVQUE estimators. In: VLECK, L.D. van; SEARLE, S.R., eds. **Variance components and animal breeding**. Ithaca: Cornell University, 1979. p.205-215.
- ANDERSON, R.D. Variance components, In: QUAAS, R.L.; ANDERSON, R.D., GILMOUR, A.R. **Use of mixed models for prediction and for estimation of (co)variance components**. Armidale: University of New England-AGBU, 1984. p.77-145.
- BOLDMAN, K.G.; KRIESE, L.A.; VLECK, L.D. van.; KACHMAN, S.D. **A manual for use of MTDFREML**; a set of programs to obtain estimates of variances and covariances. Whashington: ARS-USDA, 1993. 120p.
- 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.
- FREITAS, A.R. **Comparação de métodos de estimação de componentes de variâncias e parâmetros afins de múltiplos caracteres em bovinos**. Piracicaba: ESALQ/USP, 1991. 170p. Tese Doutorado.

- FREITAS, A.R.; VENCOVSKY, R. Comparação de métodos de estimação de componentes de variância e parâmetros afins de múltiplos caracteres em bovinos. **Pesquisa Agropecuária Brasileira**, Brasília, v.28, n.4, p.453-463, 1993.
- FREITAS, A.R.; FAVORETTI, A.C.; ALENCAR, M.M.; PEGORIN, M.J. Uso da máxima verossimilhança restrita e transformação canônica para estimação de parâmetros genéticos de características de crescimento em bovinos. **Revista da Sociedade Brasileira de Zootecnia**, Viçosa, v.23, n.3, p.394-401, 1994.
- GIANOLA, D.; FERNANDO, R.L.; IM, S.; FOULLEY, J.L. Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems. **Genome**, Ottawa, v.31, p.768-777, 1989.
- 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.
- GROENEVELD, E.; KOVAC, M.; WANG, T. PEST, a general purpose BLUP package for multivariate prediction and estimation. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 4., 1990, Edinburgh. Edinburgh: The University of Edinburgh, 1990, v.13, p.488-491.
- HARTLEY, H.O.; RAO, J.N.K. Maximum likelihood estimation for the mixed analysis of variance model. **Biometrika**, London, v.54, n.1/2, p.93-108, 1967.
- HARVILLE, D.A. Maximum likelihood approaches to variance component estimation and to related problems. **Journal of the American Statistics Association**, Washington, v.72, p.320-328, 1977.
- HARVILLE, D.A.; CALLANAN, T.P. Computational aspects of likelihood-based inference for variance components. In: GIANOLA, D.; HAMMOND, K. **Advances in statistical methods for genetic improvement of livestock**. Heidelberg: Springer-Verlag, 1990. p.136-176.
- HAYES, J.F.; HILL, W.G. A reparameterization of a genetic selection index to locate its sampling properties. **Biometrics**, Washington, v.36, p.237-248, 1980.
- HENDERSON, C.R. Estimation of variance and covariance components. **Biometrics**, v.9, p.226-252, 1953.
- 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, Ill. 1973. p.10-41.
- HENDERSON, C.R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, Washington, v.31, p.423-447, 1975.
- HENDERSON, C.R. Prediction of future records. In: POLLACK, E.; KEMPTHORNE, O.; BAILEY, I. In: INTERNATIONAL CONFERENCE ON QUANTITATIVE GENETICS, 1977, Ames. **Proceedings**. Ames: IOWA Stte University, 1977. p.615-638. IOWA STATE UNIVERSITY PRESS, Ames, IOWA. 1977, p.615-638.

- HILL, W.G.; NICHOLAS, F.W. Estimation of heritability by both regression of offspring on parent and intra-class correlation of sibs in one experiment. **Biometrics**, Washington, v.30, p.447-468, 1974.
- JENSEN, J.; MAO, I.L. Transformation algorithms in analysis of single trait and of multitrait models with equal designs matrices and one random factor per trait: a review. **Journal of Animal Science**, Champaign, v.66, p.2750-2761, 1988.
- KENNEDY, B.W. Variance component estimation and prediction of breeding values. **Canadian Journal of Genetics and Cytology**, Ottawa, v.23, n.4, p.565-578, 1981.
- LOPES, P.S.; MARTINS, E.N.; SILVA, M.A.; REGAZZI, A.J. **Estimação de componentes de variância**. Viçosa: Imprensa Universitária, 1993a. 61p.
- LOPES, P.S.; MARTINS, E.N.; SILVA, M.A.; RAGGI, L.A. **Métodos de resolução de sistemas de equações lineares**. Viçosa: Imprensa Universitária, 1993b. 55p.
- LOPES, P.S. **Avaliação genética de suínos utilizando a metodologia de modelos mistos**. Viçosa: UFV, 1994. 98p. Tese Doutorado.
- LOPES, P.S.; SILVA, M.A.; REGAZZI, A.J.; RAGGI, L.A.; MARTINS, E.N. Critérios de convergência para a resolução de equações de modelo misto. **Revista da Sociedade Brasileira de Zootecnia**, Viçosa, v.23, n.6, p.1009-1021, 1994.
- MARTINS, E.N.; LOPES, P.S.; SILVA, M.A.; REGAZZI, A.J. **Modelo linear misto**. Viçosa: Imprensa Universitária, 1993. 46p.
- MARTINS, E.N.; SILVA, M.A.; LOPES, P.S.; REGAZZI, A.J.; RAGGI, L.A.; CRUZ, C.D. Desenvolvimento de um programa computacional para estimação de componentes de variância em coelhos usando a máxima verossimilhança restrita. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 31., 1994, Maringá, **Resumos**. Maringá: UEM, 1994. p.223.
- MATLAB. **High-performance numeric computation and visualization software** - reference guide. Natick: The Math Works, 1992.
- MEYER, K. DFREML - a set of programs to estimate variance components under an individual animal model. **Journal of Dairy Science**, Champaign, Suppl.2, p.33-34, 1988.
- MEYER, K. Estimation of genetic parameters. In: HILL, W.G.; MACKAY, T.F.C. **Evolution and animal breeding**. Wallingford: CAB International, 1989a, p.161-167.
- MEYER, K. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm **Genetique, Selection, Evolution**, Paris, v.21, p.317-340, 1989b.
- MEYER, K. DFREML - Programs to estimate variance components by REML using a derivative-free algorithm, user notes. Armidale: University of New England-AGBU, 1991a. 97p.

- MEYER, K. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. **Genetique, Selection, Evolution**, Paris, v.23, p.67-83, 1991b.
- MEYER, K.; HILL, W.G. Mixed model analysis of a selection experiment for food intake in mice. **Genetical Research**, Cambridge, v.57, p.71-81, 1991.
- PATTERSON, H.D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, London, v.58, p.545-554, 1971.
- RAO, C.R. Estimation of heterocedastic variances in linear models. **Journal of the American Statistical Association**, Washington, v.65, p.161-172, 1970.
- RESENDE, M.D.V; PRATES, D.F.; JESUS, A.; YAMADA, C.K. Melhor predição linear não viciada (BLUP) de valores genéticos no melhoramento de *Pinus*. 1995 (em preparação).
- SAMPAIO, P.T.B. Variação genética e métodos de seleção envolvendo procedências e progênies de 3 espécies de *Pinus* tropicais. Curitiba: UFPR, 1996. 250p. Tese Doutorado (em preparação).
- SCHAEFFER, L.R. Maximum likelihood estimation of variance components in dairy cattle breeding research. **Journal of Dairy Science**, Champaign, v.59, p.2146-2151, 1976.
- SCHAEFFER, L.R.; WILTON, J.W.; THOMPSON, R. Simultaneous estimation of variance and covariance components from multitrait mixed model equations. **Biometrics**, Washington, v.34, p.199-208, 1978.
- SCHENKEL, F.S. **Aplicação da metodologia de modelos mistos na avaliação genética de suínos**. Porto Alegre: UFRGS, 1991. 218p. Dissertação Mestrado.
- SEARLE, S.R. Topics in variance component estimation. **Biometrics**, Washington, v.27, p.1-76, 1971.
- SEARLE, S.R.; CASELLA, G.; McCULLOCH, C.E. **Variance components**. New York: J. Wiley, 1992. 528p.
- SHAW, R.G. Maximum-likelihood approaches to quantitative genetics of natural populations. **Evolution**, Lawrence, v.41, p.812-826, 1987.
- SMITH, S.P.; GRASER, H.V. Estimating variance components in a class of mixed models by restricted maximum likelihood. **Journal of Dairy Science**, Champaign, v.69, p.1156-1165, 1986.
- THOMPSON, R. The estimation of variance and covariance components with an application when records are subject to culling. **Biometrics**, Washington, v.29, p.527-550, 1973.
- THOMPSON, R. Design of experiments to estimate heritability when observations are available on parents and offspring. **Biometrics**, Washington, v.32, p.283-304, 1976.
- THOMPSON, R. Sire evaluation. **Biometrics**, Washington, v.35, p.339-353, 1979.

THOMPSON, R.; HILL, W.G. Univariate REML analysis for multivariate data with the animal model. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 4., 1990, Edinburgh. Edinburg: The University of Edinburgh, 1990, v.13, p.484-487.