

Documentos

EFEITOS FIXOS OU ALEATÓRIOS DE REPETIÇÕES NO CONTEXTO DOS MODELOS MISTOS NO MELHORAMENTO DE PLANTAS PERENES

Marcos Deon Vilela de Resende

Colombo, PR
2002

Exemplares desta publicação podem ser adquiridos na:

Embrapa Florestas

Estrada da Ribeira, km 111 - CP 319

83411-000 - Colombo, PR - Brasil

Fone: (41) 666-1313

Fax: (41) 666-1276

Home page: www.cnptf.embrapa.br

E-mail: sac@cnptf.embrapa.br

Para reclamações e sugestões *Fale com o ouvidor*:

www.embrapa.br/ouvidoria

Comitê de Publicações da Unidade

Presidente: Moacir José Sales Medrado

Secretário-Executivo: Guiomar Moreira Bragunia

Membros: Antônio Carlos de S. Medeiros, Edilson B. de Oliveira,

Erich G. Schaitza, Honorino R. Rodighieri, Jarbas Y. Shimizu,

José Alfredo Sturion, Patricia P. de Mattos, Sérgio Ahrens,

Susete do Rocio C. Penteado

Supervisor editorial: Moacir José Sales Medrado

Revisor de texto: Glaici Kokuka

Normalização bibliográfica: Elizabeth Câmara Trevisan

Lidia Woronkoff

Foto(s) da capa: arquivos da biblioteca

Editoração eletrônica: Marta de Fátima Vencato

1ª edição

1ª impressão (2002): 500 exemplares

Todos os direitos reservados.

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

CIP-Brasil. Catalogação-na-publicação.

Embrapa Florestas

Resende, Marcos Deon Vilela de

Efeitos fixos ou aleatórios de repetições no contexto dos modelos mistos no melhoramento de plantas perenes / Marcos Deon Vilela de Resende. Colombo : Embrapa Florestas, 2002. 23 p. (Embrapa Florestas. Documentos, 68).

Inclui bibliografia

ISSN 1517-536X

1. Planta perene – melhoramento – metodologia. I. Título.
II. Série.

CDD 634.956 (21. ed.)

© Embrapa, 2002

Autor

Marcos Deon Vilela de Resende

Engenheiro-agrônomo, Pesquisador da *Embrapa
Florestas*.

deon@cnpf.embrapa.br

Apresentação

A metodologia de modelos mistos (REML/BLUP) ao nível individual tem sido uma ferramenta bastante utilizada, atualmente, no melhoramento de plantas perenes. Em função de sua importância várias questões têm sido levantadas sendo uma das mais importantes a definição os efeitos fixos ou aleatórios de repetições sobre o modelo.

Este trabalho discute esta importante questão à luz de princípios básicos da predição de valores genéticos, da influência da heterogeneidade ambiental entre repetições e de vivos na comparação genética através de repetições.

Vitor Afonso Hoeflich
Chefe Geral
Embrapa Florestas

Sumário

1. Introdução	9
2. Princípio básico da predição de valores genéticos	9
3. Influência da heterogeneidade ambiental entre repetições e tamanho da repetição na eficiência dos procedimentos alternativos de predição	11
4. Vício na comparação genética através das repetições	18
5. Considerações Finais	19
6. Referências bibliográficas	22

Efeitos fixos ou aleatórios de repetições no contexto dos modelos mistos no melhoramento de plantas perenes

Marcos Deon Vilela de Resende

1 Introdução

A estimação de componentes de variância e a predição de valores genéticos no contexto do melhoramento de plantas perenes, demanda o uso da metodologia de modelos mistos (REML/BLUP) ao nível individual, a qual vem sendo aplicada com eficiência ao melhoramento de várias espécies perenes. Uma questão fundamental na aplicação dessa metodologia refere-se à definição dos efeitos do modelo estatístico como fixos ou aleatórios. No caso do efeito de repetições experimentais esta definição nem sempre é trivial. Assim, esta questão é abordada no presente artigo.

2 Princípio básico da predição de valores genéticos

Um modelo estatístico geral para uma observação fenotípica é dado por:

$$y = \mu + r + g + e, \text{ em que:}$$

y : valor fenotípico;

μ : média geral;

r : efeitos ambientais identificáveis ou atribuíveis a determinadas causas como repetições, locais, anos, etc.

g : efeito genético;

e : efeitos ambientais não identificáveis ou completamente aleatórios.

Na predição dos efeitos genéticos (g), o primeiro passo a ser adotado é a correção dos dados observados (y), para os efeitos ambientais identificáveis (r) (Van Vleck et al., 1987).

Ignorando a média geral, esta correção é dada por: $(y - r) = (g + e)$.

A partir desta correção, a obtenção de g , livre de e , é dada pela multiplicação de $(g + e)$ pela herdabilidade (h^2), ou seja, obtendo-se o preditor

$$\hat{g} = h^2 (y - r) = h^2 (g + e) = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} (y - r).$$

A acurácia ($r_{\hat{g}g}$) da predição do valor genético, nesta situação simples, equivale

à raiz quadrada da herdabilidade. Assim, quanto maior a herdabilidade, maior é a acurácia seletiva e a precisão na seleção, visto que a variância do erro de

predição do valor genético é dada por $PEV = 1 - \frac{r_{\hat{g}g}^2}{\sigma_g^2}$, ou seja, quanto maior a

acurácia, menor é a variância do erro de predição. Em situações mais complexas, a acurácia continua sendo função direta da herdabilidade, ou seja, quanto maior a h^2 , maior também é $r_{\hat{g}g}$ (Van Vleck et al., 1987).

Matricialmente, o modelo geral é dado pelo seguinte modelo linear misto:

$$y = Xr + Zg + e$$

y : vetor de observações fenotípicas;

r : vetor de efeitos fixos ou efeitos ambientais identificáveis;

g : vetor de efeitos genéticos, aleatório;

e : vetor de efeitos ambientais não identificáveis, aleatório;

X e Z : matrizes de incidência para r e g , respectivamente.

No contexto desse modelo linear misto, a correção dos dados é dada por $y - E(y) = (y - Xr)$. $E(y)$ ou Xr equivalerá à média da repetição (sob um modelo de médias e não de efeitos) quando o efeito da repetição for considerado fixo e equivalerá à média geral quando o efeito de repetição for considerado aleatório.

Entretanto, nas situações práticas, os efeitos ambientais identificáveis não são sempre puramente ambientais, mas, por vezes, incluem também frações de efeitos genéticos, confundidos com os efeitos ambientais. Assim, devem ser classificados como efeitos ambientais identificáveis àqueles predominantemente ambientais, ou seja, aqueles em que as frações de efeitos genéticos confundidos são desprezíveis. E esta classificação nem sempre é trivial.

3 Influência da heterogeneidade ambiental entre repetições e tamanho da repetição na eficiência dos procedimentos alternativos de predição

Pela metodologia de modelos mistos, a predição BLUP (melhor predição linear não viciada) dos valores genéticos pode ser obtida, alternativamente, considerando o efeito de repetição (r) como fixo ou aleatório. Considerando tal efeito como fixo ou predominantemente ambiental, as equações de modelo misto são dadas por:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1} \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{r} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}, \text{ em que:} \quad (1)$$

$$\frac{\sigma_e^2}{\sigma_g^2} = \frac{1 - h^2}{h^2}$$

Verifica-se que a predição BLUP, no caso, depende do conhecimento dos

componentes de variância σ_g^2 e σ_e^2 ou da herdabilidade $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$, que é

uma herdabilidade individual dentro da repetição.

Considerando o efeito de repetição como aleatório, as equações de modelo misto são dadas por:

$$\begin{bmatrix} X'X + I \frac{\sigma_e^2}{\sigma_r^2} & X'Z \\ Z'X & Z'Z + A + \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{r} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}, \text{ em que:} \quad (2)$$

$$\frac{\sigma_c^2}{\sigma_g^2} = \frac{1 - h^{2*} - r^2}{h^{2*}}, \text{ sendo que } h^{2*} \text{ é dada por } h^{2*} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2 + \sigma_r^2}, \text{ que é uma}$$

herdabilidade individual no experimento como um todo. Note-se que $h^2 \geq h^{2*}$, ou seja, h^2 só será igual a h^{2*} quando $\sigma_r^2 = 0$. O componente r^2 é dado por

$$r^2 = \frac{\sigma_r^2}{\sigma_g^2 + \sigma_e^2 + \sigma_r^2}.$$

Verifica-se que a equação (1) é um caso particular de (2) quando σ_r^2 é muito grande em relação a σ_e^2 , ou seja, quando $\sigma_r^2 \rightarrow \infty$. Assim, com grande variação ambiental entre repetições, ou seja, com grande heterogeneidade ambiental entre repetições, tem-se:

- (i) o modelo (2) passa a equivaler ao (1) e, portanto, a h^{2*} estimada por REML (máxima verossimilhança restrita) simultaneamente à predição BLUP é equivalente a h^2 ;
- (ii) os dois modelos conduzem a idênticas predições e acurácias de predição;
- (iii) o modelo efetivamente usado é o (1), ou seja, com efeito fixo de repetição.

Quando σ_r^2 é muito pequeno em relação a σ_e^2 , ou seja, $\sigma_r^2 \rightarrow 0$, não existe heterogeneidade ambiental entre repetições. Neste caso, não existe efeito de repetição e pelo modelo (2) pode-se concluir pela não significância de tal efeito,

pelo teste da razão de verossimilhança (LRT). Nesta situação, pode-se concluir:

- (i) não há necessidade de consideração e/ou ajuste para o efeito de repetição;
- (ii) corrigir cada observação pela média do bloco ou pela média geral conduzirá a resultados idênticos;
- (iii) $\hat{h}^{2*} = \hat{h}^2$;
- (iv) os dois modelos conduzirão a resultados e precisões idênticas.

Entretanto, na prática, σ_r^2 apresenta magnitude intermediária, de forma que os modelos são efetivamente diferentes e, para comparar os modelos (1) e (2) é necessário inferir sobre a fração de σ_g^2 que fica confundida no efeito de bloco.

A julgar apenas pelos denominadores de h_g^{2*} e h^2 concluir-se-ia pela superioridade do modelo (1), aquele com efeito fixo de repetição. No entanto, é preciso verificar qual modelo faz pleno uso de σ_g^2 , o numerador da herdabilidade.

Resende & Higa (1994) decompuseram a σ_g^2 total em frações retidas nos vários efeitos do modelo estatístico. As várias frações obtidas foram:

- (i) Efeito do indivíduo dentro de parcela ($Y_{ijk} - \bar{Y}_{ij}$):

$$\frac{n-1}{n} (1-\rho) \sigma_k^2$$

- (ii) Efeito de progênie ($Y_i - \bar{Y}$):

$$\frac{p-1}{p} \frac{1+(nb-1)\rho}{nb} \sigma_g^2$$

- (iii) Efeito de parcela ($\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}$):

$$\frac{p-1}{p} \frac{b-1}{b} \frac{1-\rho}{n} \sigma_g^2$$

(iv) Efeito de bloco ($\bar{Y}_j - \bar{Y}_\cdot$):

$$\frac{b-1}{b} \frac{1-p}{np} \sigma_g^2$$

Considerando σ_g^2 a variância genética aditiva, ρ é o coeficiente de correlação genética aditiva entre indivíduos da progênie, ou seja, $\rho = 0,25$ para meios-irmãos e $\rho = 0,5$ para irmãos germanos. As quantidades n , p e b referem-se ao número de plantas por parcela, número de progênies e número de repetições, respectivamente.

Verifica-se que a fração da variação genética retida no efeito de repetição depende inversamente do tamanho da repetição ($N = np$). Com $p = 100$ progênies e $n = 5$ plantas por parcela, as frações da variação genética retida serão $0,75/500 = 0,0015$ para progênies de meios irmãos e $0,50/500 = 0,0010$ para progênies de irmãos germanos. Estas frações são desprezíveis e, portanto, neste caso, o modelo (1) (com efeito fixo de repetição) é superior ao modelo (2), visto que as observações serão ajustadas para a média da repetição e h^2 é maior que h^{2*} . Em outras palavras, nesta situação, não existem efeitos genéticos a serem recuperados nos efeitos de repetições.

Com uma planta por parcela, o efeito de repetições retém $\frac{1-p}{p} \sigma_g^2$. Neste caso, com as mesmas 100 progênies as frações de variância genética retida serão $0,75/100 = 0,0075$ para progênies de meios irmãos e $0,50/100 = 0,0050$ para progênies de irmãos-germanos. Estas frações representam menos de 1% da variação genética total e são igualmente desprezíveis. Com teste clonal em vez de teste de progênies, tem-se $p = 1$ e nenhuma variação genética será retida entre repetições.

Mesmo que a sobrevivência não seja 100% estas inferências são válidas, visto que as repetições são altamente conectadas, com conexão próxima a 100%, conforme conceito de conectabilidade de Weeks & Williams (1964). Com sobrevivência muito baixa e especialmente nos delineamentos com uma planta por parcela, as repetições tenderão a ser incompletas e caracterizarão blocos incompletos. Nesta situação, os blocos incompletos são melhor modelados como efeitos aleatórios (visando a recuperação da informação genética inter-blocos),

sendo a mesma afirmativa válida para os delineamentos em látice (Resende & Fernandes, 2000). É importante relatar que a definição de desbalanceamento apresentada por Visscher & Goddard (1993), refere-se ao caso em que alguma progênie não esteja representada em todas as repetições. Em outras palavras, tal conceito de desbalanceamento refere-se a repetições incompletas. A alocação não aleatória dos materiais genéticos nas repetições também caracteriza um delineamento desbalanceado.

Resende & Fernandes (1999) demonstraram a equivalência entre índices multi-efeitos ótimos derivados para o caso balanceado por Resende & Higa (1994) e suas correspondentes predições BLUP via metodologia de modelos mistos. Assim, índices multi-efeitos incluindo todos os efeitos aleatórios do modelo estatístico podem ser utilizados para derivar os seus correspondentes BLUP, para vários modelos alternativos. Para o modelo (1), porém para o caso de delineamentos com várias plantas por parcela (fato que requer a expansão do modelo (1) para incluir o efeito de parcela), o correspondente índice multi-efeitos é dado por:

$$\hat{g}_1 = b_1 (Y_{ijk} - \bar{Y}_{ij.}) + b_2 (\bar{Y}_{i.} - \bar{Y}_{..}) + b_3 (Y_{ij.} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})$$

Para o modelo (2) expandido, o correspondente índice multi-efeitos é dado por:

$$\hat{g}_2 = b_1 (Y_{ijk} - \bar{Y}_{ij.}) + b_2 (\bar{Y}_{i.} - \bar{Y}_{..}) + b_3 (Y_{ij.} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}) + b_4 (\bar{Y}_{.j} - \bar{Y}_{..})$$

Os ponderadores dos vários efeitos aleatórios dos índices multi-efeitos são:

$$b_1 = h_d^2 = \frac{(1-\rho) \sigma_g^2}{\sigma_d^2} : \text{herdabilidade do efeito de indivíduo dentro de parcela;}$$

$$b_2 = h_p^2 = \frac{1 + (nb - 1) \rho}{nb} \frac{\sigma_g^2}{\sigma_p^2 + \sigma_c^2 / b + \sigma_d^2 / nb} : \text{herdabilidade do efeito de progênie;}$$

$$b_3 = h_c^2 = \frac{1-\rho}{n} \frac{\sigma_g^2}{\sigma_d^2 / n + \sigma_c^2} : \text{herdabilidade do efeito de parcela;}$$

$$b_4 = h_r^2 = \frac{(1-\rho) \sigma_g^2}{\sigma_e^2 + \sigma_c^2/p + \sigma_d^2/np} : \text{herdabilidade do efeito de bloco, em que:}$$

σ_p^2 , σ_c^2 , σ_g^2 e σ_e^2 : variância entre progênies, entre parcelas, dentro de parcelas e entre repetições, respectivamente.

No caso de uma planta por parcela os índices multi-efeitos são:

$$\hat{g}_1 = b_5 (Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}) + b_6 (\bar{Y}_i - \bar{Y}), \text{ que equivale}$$

exatamente ao BLUP pelo modelo (1) para o caso balanceado.

$$\hat{g}_2 = b_5 (Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}) + b_6 (\bar{Y}_i - \bar{Y}) + b_7 (\bar{Y}_j - \bar{Y}), \text{ que}$$

equivale exatamente ao BLUP pelo modelo (2) para o caso balanceado.

Os ponderadores dos vários efeitos são:

$$b_5 = \frac{(1-\rho) \sigma_g^2}{\sigma_e^2} : \text{herdabilidade dentro de progênie;}$$

$$b_6 = \frac{1 + (b-1) \rho}{b} \frac{\sigma_g^2}{\sigma_p^2 + \sigma_e^2/b} : \text{herdabilidade entre progênies;}$$

$$b_7 = \frac{1-\rho}{p} \frac{\sigma_g^2}{\sigma_e^2 + \sigma_e^2/p} : \text{herdabilidade do efeito de bloco, em que:}$$

σ_e^2 : variância residual no delineamento com uma planta por parcela.

Conforme esperado, verifica-se que, quando σ_e^2 é muito grande ($\sigma_e^2 \rightarrow \infty$), $\hat{g}_2$ converge para $\hat{g}_1$ (pois $b_6 = b_7 = 0$), assim como o modelo (2) converge para o modelo (1) nas equações de modelo misto para a predição BLUP.

Adicionalmente, vale comentar novamente que esta convergência também ocorre quando o tamanho da repetição é grande (np alto), independentemente da magnitude de σ_e^2 , visto que o numerador de b_6 e b_7 tenderiam a zero (e não

haveria informação genética a ser recuperada no efeito de repetição). Este resultado foi confirmado por um estudo de simulação realizado por Ugarte et al.

(1992), que concluíram que um modelo do tipo BLUP(2) é mais vantajoso que um modelo do tipo BLUP(1) somente quando cada nível do efeito em consideração (repetição) for pequeno e, simultaneamente, σ_e^2 / σ_r^2 , for de alta magnitude, ou seja, quando existe pequena heterogeneidade ambiental entre repetições.

Em uma situação real de campo, na experimentação com espécies perenes, os fatores repetição grande (np alto) e grande heterogeneidade ambiental entre repetições (σ_r^2 de grande magnitude) contribuem, simultaneamente, para que b_j e b_j tendam a zero e, portanto, para que a predição seja mais acurada pelo modelo (1) (ou pelo índice $\hat{g}_j$) do que pelo modelo (2), visto que as observações estarão sendo ajustadas à média da repetição e, que a variância do desvio ($Y_{ijk} - Y_j$) é menor que a variância do desvio ($Y_{ijk} - Y$). Tais variâncias são os denominadores de h^2 e $h^{2'}$, respectivamente. Dessa forma, parece prudente tratar os efeitos de repetições completas (e logicamente bem conectadas) como fixos, conforme tem sido realizado (Van Vleck et al., 1987; Henderson, 1984; White & Hodge, 1989; Gianola & Hammond, 1990). A concepção original do BLUP trata-r como efeito fixo (Henderson, 1973).

É importante relatar também que, considerando o efeito de repetição como aleatório, o modelo não será mais misto e, sim aleatório, segundo a classificação tradicional dos modelos em fixos, aleatórios e mistos (Henderson, 1953; Searle, 1971). Sob um modelo aleatório (tendo somente a média geral μ como efeito fixo), o fato de considerar μ como conhecido ou estimá-lo a partir dos dados, conduz aos mesmos resultados práticos (Visscher & Goddard, 1993). Neste caso, o BLUP e o BLP (melhor preditor linear) são equivalentes para propósitos de seleção. Também neste caso, em termos de estimação de componentes de variância, os métodos de máxima verossimilhança (ML) e REML se equivalem, pois a divisão por η ou por $(\eta - 1)$ nos estimadores da variância residual por ML e REML, respectivamente, não conduzem a diferenças práticas nas estimativas (Blasco, 2001). Em resumo, no modelo aleatório, o procedimento ML/BLP é adequado para a estimação/predição.

Outra forma de enxergar a correção dos dados para os efeitos ambientais identificáveis é através da expressão alternativa para o BLUP de $\hat{g}$, dada por: $\hat{g} = C'V^{-1}(y - X\hat{\beta})$, em que C é uma matriz de covariância entre g e y e V é a matriz de variância de y . Percebe-se claramente, que os dados são previamente corrigidos para os efeitos ambientais antes de serem ponderados por uma função

da herdabilidade. E com alta heterogeneidade entre repetições, esta correção dos dados deve ser feita em relação às médias de repetições e não pela média geral.

Adicionalmente, considerar os efeitos de repetições como fixos apresenta menos restrições do ponto de vista estatístico, pois evita suposições tais quais a normalidade dos efeitos de repetições ($r \sim (0, I\sigma_r^2)$). normalidade esta assumida no procedimento REML/BLUP para a estimação de σ_r^2 e predição de r . Além disso, é importante relatar que a predição é realmente BLUP somente quando os componentes de variância são conhecidos. Entretanto, na prática, estes são estimados dos dados amostrais. Assim, a rigor, nas aplicações práticas não se tem BLUP, mas, sim EBLUP, significando BLUP empírico, devido à substituição dos valores paramétricos (segundos momentos) por suas estimativas. E considerando o efeito de repetição como aleatório, tem-se a necessidade de estimar um componente de variância (σ_r^2) a mais. Este fato introduz mais erro na predição em relação ao modelo com efeito fixo de repetição.

4 Vício na comparação genética através das repetições

A consideração dos efeitos de repetição como fixos propicia predições invariantes aos efeitos fixos, removendo assim vícios nas comparações genéticas através das repetições. Tal consideração é sobretudo importante quando existe uma alocação não aleatória dos materiais genéticos nas repetições e, neste caso, o BLUP minimiza o vício causado por associações entre efeitos de repetições e o nível genético dos materiais que nelas se desenvolvem. Por outro lado, quando não existem associações entre efeitos de repetições e o nível genético dos materiais, não existem vícios (PEV aproximadamente igual ao erro quadrático médio) mesmo quando se consideram como aleatórios os efeitos de repetições.

O tipo de alocação dos materiais genéticos nas repetições, torna-se então relevante, embora muitas vezes não esteja sob controle do melhorista. Considerando que o delineamento em blocos completos balanceado propicia uma boa conectabilidade (capacidade para estimar diferenças ambientais entre repetições, conforme Weeks & Williams, 1964), a adoção do modelo com efeito fixo de repetição permite, com segurança, contornar este possível fator

indesejável associado à correlação entre efeitos de repetições e de materiais genéticos. É importante relatar que, na prática do melhoramento florestal, em geral, as repetições são estabelecidas com mudas de padrão similar de crescimento, ou seja, as melhores mudas são estabelecidas em uma mesma repetição o mesmo acontecendo com as piores. Isto caracteriza uma alocação não aleatória de materiais genéticos nas repetições.

Visscher & Goddard (1993) relatam que, quando os melhores materiais genéticos são alocados nas piores (em termos de ambiente) repetições, a correlação entre os valores genéticos preditos e os verdadeiros pode se tornar negativa. Assim, em situações de alocação não aleatória dos materiais genéticos nas repetições, recomendam tratar os efeitos de repetição como fixos para propósitos práticos de avaliação genética.

A grande questão envolvida na comparação não viciada de materiais genéticos através das repetições é o confundimento entre efeitos genéticos e ambientais na média da repetição. Com alocação não aleatória, torna-se mais difícil o ajuste para os efeitos ambientais e a separação entre os efeitos genéticos e ambientais confundidos. Nesta situação, os efeitos de repetição devem ser tratados como fixos para que se obtenha o BLUP. Segundo Foulley et al. (1990), uma simples razão para explicar isto, refere-se ao fato do BLUP ser invariante à translação nos efeitos considerados fixos. Isto pode ser explicado também invocando um argumento Bayesiano, de forma que o BLUP com r fixo equivale a um estimador Bayesiano usando uma distribuição a priori não informativa para r .

5 Considerações Finais

Em resumo, a comparação entre os modelos BLUP(1) e BLUP(2) depende de três fatores: (i) do tamanho da repetição; (ii) do tipo de alocação dos materiais genéticos nas repetições (aleatória ou não); (iii) da relação σ_a^2 / σ_e^2 .

Considerar efeitos de repetições completas (balanceadas ou não em termos do número de plantas por progênie) como fixos ou aleatórios é um compromisso entre garantir não vício (o U do BLUP) ou recuperar informação genética retida no referido efeito, respectivamente. Considerar como fixo garante não vício mas pode perder informação genética quando o tamanho da repetição é muito pequeno. Considerar aleatório garante o uso de toda informação genética mas não garante efetivamente o não vício.

Neste sentido, o tamanho da repetição é de grande importância na decisão. O trabalho de Visscher & Goddard (1993) considera bem esta questão e mostra que a potencial vantagem do uso do efeito como aleatório é: (i) praticamente nula quando N (o tamanho da repetição) é maior que 24; (ii) pouco expressiva para N entre 4 e 24; (iii) pequena ($0,035 = 0,847 - 0,812$, ou seja, acurácia passando de 0,812 para 0,847, ou seja 4%) para $N = 4$.

Na situação de pequeno tamanho de repetição, a diferença nos valores de acurácia associados ao BLUP(1) e BLUP(2) são devidas a dois fatores: (i) aumento na acurácia do BLUP(2) devido ao uso adicional do efeito de repetição e (ii) redução na acurácia do BLUP(1) em função da estimação do efeito de repetição, a partir de um menor número de observações do que aquele usado na estimação do efeito da média geral pelo BLUP(2).

A estimação dos efeitos fixos simultaneamente à predição dos valores genéticos conduz a um acréscimo (em relação ao BLP que supõe os efeitos fixos como conhecidos) na PEV, porque um erro adicional é introduzido em função do uso de uma estimativa destes efeitos e não de seu valor verdadeiro. A PEV pode ser expressa por $PEV = G - C'V'C + C'V'X(X'V'X)^{-1}X'V'C$ (em que $G = Var(a)$ e $C = COV(a, y)$) (Henderson, 1984) em que a terceira parte desta expressão refere-se ao acréscimo devido à estimação dos efeitos fixos. Verifica-se, assim, que a contribuição para a PEV, associada à estimação dos efeitos fixos, é maior para o BLUP(1) do que para o BLUP(2). Este resultado é devido ao menor número de observações usados na estimação das médias de repetição pelo BLUP(1). De maneira geral, à medida que o número de observações em cada nível do efeito fixo aumenta, diminui o acréscimo na PEV (Kennedy & Trus, 1993). Com o aumento do número de tratamentos tanto a unidade repetição quanto experimento aumentam e, conseqüentemente, isto conduz a uma redução na contribuição da estimação dos efeitos fixos para a PEV tanto para o BLUP(1) quanto para o BLUP(2).

O vetor de efeitos fixos deve contemplar os efeitos ambientais identificáveis, sendo que o principal efeito ambiental identificável nos experimentos com plantas perenes é o efeito de repetição, o qual praticamente não contribui (Resende e Higa, 1994) com informação genética a respeito dos indivíduos (visto que as repetições, normalmente, apresentam grande número de indivíduos). Por outro lado, o bloco é o principal estrato homogêneo para comparação de indivíduos, de forma que dentro de blocos os indivíduos são

prontamente comparáveis de uma maneira não viciada. Para comparação de indivíduos através dos blocos torna-se necessário ajustar os valores individuais para os efeitos de blocos, de maneira que as comparações sejam não viciadas. A forma de se conseguir, isto é, tratar os efeitos de repetições como fixos, fato que torna as avaliações genéticas invariantes aos efeitos de repetições e não viciadas devido à associação (correlação) entre efeitos de blocos e nível genético dos indivíduos que nele se desenvolvem.

A obtenção de avaliações genéticas invariantes aos efeitos de repetições, ou seja, comparáveis de maneira não viciada nas diferentes repetições, depende da consideração do efeito de repetições como fixo e também de uma boa conectabilidade entre as repetições, o que é conseguido procurando-se fazer a maioria das progênies estar representada em cada nível do efeito fixo.

A definição dos efeitos de repetições como fixos parece contraditória em relação ao nome do delineamento: blocos ao acaso. Entretanto, a rigor, o nome correto é "delineamento de tratamentos aleatorizados em blocos", uma vez que os blocos podem ser assumidos como fixos ou aleatórios, dependendo da situação experimental. Assim, os tratamentos é que são casualizados e não necessariamente os blocos.

Os seguintes fatores apontam para uma preferencial definição dos efeitos de repetições completas como fixos, no contexto do melhoramento de plantas perenes:

- (i) alta heterogeneidade ambiental entre repetições;
- (ii) ausência de informação genética a ser recuperada nos efeitos de repetições, dado o tamanho das mesmas;
- (iii) alocação não aleatória dos materiais genéticos nas repetições;
- (iv) necessidade de suposição de normalidade dos efeitos de repetições, quando os mesmos forem ajustados como aleatórios;
- (v) necessidade de estimação de mais um componente de variância, quando os efeitos de repetições forem ajustados como aleatórios, fato que introduz mais erro na predição;

Entretanto, em situações especiais, os dois modelos alternativos podem ser ajustados para posterior seleção daquele que propiciar maior acurácia seletiva.

É importante relatar que tratar os efeitos de repetição como fixos não limita as inferências dos resultados para o conjunto de repetições analisado. Os efeitos genéticos são elementos aleatórios preditos ($\hat{g}$), os quais podem ser extrapolados para quaisquer ambientes, bastando realizar a soma $\hat{g} + E(y)$, que fornece o valor genético predito para um ambiente com média $E(y)$.

6 Referências Bibliográficas

BLASCO, A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, v. 79, p. 2023-2046, 2001.

FOULLEY, J. L.; BOUIX, J.; GOFFINET, B.; ELSEN, J. M. Connectedness in genetic evaluation. In: GIANOLA, D.; HAMMOND, K. (Ed.), **Advances in statistical methods for genetic improvement of livestock**. Berlin: Springer-Verlag, 1990, p. 277-308.

GIANOLA, D.; HAMMOND, K. **Advances in statistical methods for genetic improvement of livestock**. Berlin: Springer-Verlag, 1990. 534 p.

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

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. Champaign, 1973. [Symposium...], Champaign: American Society of Animal Science, 1973, p. 10-41.

KENNEDY, B. W.; TRUS, D. Considerations on genetic connectedness between management units under an animal model. **Journal of Animal Science**, v. 71, p. 2341-2352, 1993.

RESENDE, M. D. V. de; FERNANDES, J. S. C. Análises alternativas envolvendo o procedimento BLUP e o delineamento experimental de blocos incompletos ou látice. **Revista de Matemática e Estatística**, São Paulo, v. 18, p. 103-124, 2000.

RESENDE, M. D. V. de; FERNANDES, J. S. C. Procedimento BLUP individual para delineamentos experimentais aplicados ao melhoramento florestal. **Revista de Matemática e Estatística**, São Paulo, v. 17, p. 89-107, 1999.

RESENDE, M. D. V. de; HIGA, A. R. Maximização da eficiência da seleção em testes de progêies de *Eucalyptus* através da utilização de todos os efeitos do modelo matemático. **Boletim de Pesquisa Florestal**, Colombo, v. 28/29, p. 37-55, 1994.

SEARLE, S. R. **Linear models**. New York: J. Wiley & Sons, 1971. 532 p.

UGARTE, E.; ALENDIA, R.; CARABAÑO, M. J. Fixed or random contemporary groups in genetic evaluations. **Journal of Dairy Science**, v. 75, n. 1, p. 269-278, 1992.

VISSCHER, P. M.; GODDARD, M. E. Fixed and random contemporary groups. **Journal of Dairy Science**, v. 76, n. 5, p. 1444-1454, 1993.

VLECK, L. D. van; POLLAK, E. J.; OLTENACU, E.A. B. **Genetics for the animal sciences**. New York: W. H. Freeman, 1987. 391 p.

WEEKS, D. L.; WILLIAMS, D. R. A note on the determination of connectedness in an N-way cross classification. **Technometrics**, v. 6, p. 319-324, 1964.

WHITE, T. L.; HODGE, G. **Predicting breeding values with application in forest tree improvement**. London: Kluwer, 1989. 367 p.