

DANIELLE SILVA PINTO

**MODELOS MISTOS NO MELHORAMENTO VEGETAL COM A
UTILIZAÇÃO DO SOFTWARE R**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2013

Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV

T

P659m Pinto, Danielle Silva, 1983-
2013 Modelos mistos no melhoramento vegetal com a utilização
do software R / Danielle Silva Pinto. – Viçosa, MG, 2013.
vii, 45 f. : il. ; 29 cm.

Inclui anexo.

Orientador: Cosme Damião Cruz.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Melhoramento vegetal. 2. Plantas - Melhoramento genético. 3. BLUP. 4. Métodos de simulação. I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 631.52

DANIELLE SILVA PINTO

**MODELOS MISTOS NO MELHORAMENTO VEGETAL COM A
UTILIZAÇÃO DO SOFTWARE R**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 12 de agosto de 2013.

Leonardo Lopes Bhering

Lívia Gracielle Oliveira Tomé

Carlos Eduardo Magalhães dos Santos

Moysés Nascimento
(Coorientador)

Cosme Damião Cruz
(Orientador)

À Deus.

Ao meu pai, Aguielo Barbosa Pinto.

À minha mãe, Rosilene de Sousa e Silva.

Aos meus irmãos Rose Kelly, Rudi, Anderson e Rafael.

Aos meus sobrinhos Caio Gabriel e Cauê.

Dedico.

AGRADECIMENTOS

A Deus pela infinita graça e misericórdia, por sempre me preencher de alegria, amor e compreensão, permitindo conquistar os meus sonhos.

A minha família pelo apoio incondicional ao longo de todo esses anos.

Ao professor Cosme Damião Cruz, pela orientação séria e comprometida, pela amizade e confiança depositada em mim.

À Universidade Federal de Viçosa, particularmente ao programa de pós-graduação Genética e Melhoramento, pela formação acadêmica e oportunidade de crescimento intelectual e profissional.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro durante o período do doutorado.

À Klabin S.A em nome da melhorista da empresa Regiane Abjaud Estopa, pelo fornecimento de conjunto de dados para execução das análises.

Ao pesquisador e amigo Marcos Deon Vilela de Resende por sua colaboração e orientação durante o curso e pelo apoio na realização deste trabalho.

Ao professor e amigo Moysés Nascimento pela orientação e colaboração na realização deste trabalho.

Aos membros da banca Leonardo Lopes Bhering, Carlos Eduardo Magalhaes dos Santos e Lívia Gracielle Oliveira Tomé, por terem aceitado o convite e por suas contribuições oportunas, que certamente enriqueceram o trabalho.

Ao amigo Rafael Simões Tomaz e Janeo Eustáquio de Almeida Filho pelo esforço incondicional e dedicação nas programações e convivência no laboratório.

Ao secretário da Pós-Graduação Marco Túlio pela dedicação em seus serviços e ajuda constante.

Aos professores de graduação e de pós-graduação, pela atenção e ensinamentos.

As amigas Ana Carolina Campana Nascimento, Jacqueline Glasenapp, Penha Fonseca, Aline Galdino, Lorêta Buuda, Águida Morales que foram grandes amigas em todos os momentos, desde as discussões científicas as descontrações casuais.

Aos meus colegas e amigos da UFV, principalmente ao pessoal do laboratório de Bioinformática Caio Salgado, Eliel Ferreira, Felipe Barrera, Digner Cedillo, David Baffa, João Filipi, Rafael Tassinari, Angelica Plata, Marciane Oliveira, Vinicius Quintão, Luiza da Matta, Lívia Torres, Amálio Gonzalez, Gabi Silva, Isabela Sant'Anna, Gislayne Faria e Ísis Almeida pela amizade e convívio agradável.

Ao meu querido grupo de convivência, que por todo esse tempo sempre me incentivou, acompanhou todas as dificuldades e alegrias e me fez sentir um no meio de vocês! Valeu!!!

Às queridas amigas de república, desde o período de 2007 a 2013, Lidiane, Daniela, Sara, Lú, Paula, Andiará, Joice, Bruna, Ana Cláudia e Milena. Obrigada pelo carinho e pela convivência agradável por todos esses seis anos! Fui muito feliz com vocês!

E a todos que direta ou indiretamente colaboraram para o sucesso deste trabalho.

OBRIGADA!

SUMÁRIO

RESUMO	vi
ABSTRACT	vii
INTRODUÇÃO GERAL	1
REFERÊNCIAS.....	4
ARTIGO 1: Uso do pedigree para análises de modelos mistos no melhoramento vegetal	
Resumo	06
1. Introdução.....	06
2. Base teórica da estimação e predição	07
3. Metodologia.....	09
3.1 Material.....	10
3.2 Descrição do script.....	11
4. Conclusão.....	21
5. Apêndice.....	22
6. Referências.....	25
ARTIGO 2: Análises comparativas entre os métodos de seleção combinada e o procedimento REML/BLUP em ensaios simulados de progênies de meio irmãos	
Resumo.....	27
1. Introdução.....	27
2. Material e Métodos.....	29
3. Resultados e Discussão.....	33
4. Conclusão.....	43
5. Referências.....	43
CONCLUSÃO GERAL.....	45

RESUMO

PINTO, Danielle Silva, D.Sc., Universidade Federal de Viçosa, agosto de 2013. **Modelos mistos no melhoramento vegetal com a utilização do software R.** Orientador: Cosme Damião Cruz. Coorientadores: Marcos Deon Vilela de Resende e Moysés Nascimento.

Esta pesquisa apresentou como objetivos, a elaboração de um script no software R para avaliação de dados no melhoramento vegetal utilizando o pacote *pedigreemm*, com a construção de um arquivo padrão para análise e aplicação de correção do ajuste, em delineamentos genéticos de progênies de famílias de meio irmãos e de irmãos germanos. Este foi fundamentado em análises para o melhoramento do *Eucalyptus urophylla* S.T. Blake. Para os quatro tipos de delineamentos apresentados no trabalho existe uma correção específica, que ajusta saída do pacote gerando os valores genéticos preditos corretos do BLUP. O segundo objetivo foi comparar os métodos de seleção combinada e o procedimento REML/BLUP em delineamento de progênies de meio irmãos simuladas. Baseado na ferramenta de simulação e da teoria de genética quantitativa, a partir de uma população base, progênies de meio irmãos foram geradas com o diferentes níveis de desbalanceamento variando de 10 a 50%. Três características quantitativas foram simuladas, com diferentes herdabilidades de médias de progênies. Os resultados obtidos possibilitaram perceber que os métodos de seleção combinada e via REML/BLUP apresentaram respostas muito semelhantes em todas as avaliações realizadas e a utilização do método de índice combinado mostra-se adequado para a seleção de progênies de meio irmãos, com confiabilidades similares ao procedimento REML/BLUP.

ABSTRACT

PINTO, Danielle Silva, D.Sc., Universidade Federal de Viçosa, August, 2013. **Mixed models in plant breeding using the R software.** Adviser: Cosme Damião Cruz. Co-advisers: Marcos Deon Vilela de Resende and Moysés Nascimento.

This research aimed to development script trough R software for data evaluation in plant breeding using the pedigreemm package, by the construction of a standard file for analysis and correction adjustment application in replicated genetic progenies of full subling and half-sibs families. This was based on an analysis for *Eucalyptus urophylla* ST Blake breeding. For the four types of designs presented in the work there is a specific constant, which adjusts the output generates by the package correcting the the genetic values predicted by BLUP. The second objective was to compare combined selection methods and REML/BLUP delineation of different simulated half-sibs. Based on the simulation tool and quantitative genetics theory, through a basis population, different half-sibs were generated with different levels of imbalance ranging from 10 to 50%. Three quantitative traits were simulated with different heritability average by progenies. The results enabled us to realize that the combined selection methods and REML/BLUP showed very similar responses in all assessments and the use of the combined index method is adequate for selecting different half-sibs, with reliabilities similar to the procedure by REML/BLUP.

INTRODUÇÃO GERAL

O melhoramento genético é arte e a ciência que visam a modificação gênica das plantas para torná-las mais úteis ao homem (Borém e Miranda, 2007). É a base para que a agricultura possa disponibilizar de forma sustentável alimentos, fibras e bioenergia para a sociedade (Borém e Fritsche- Neto, 2013).

O melhoramento genético florestal é uma ciência que teve seu maior desenvolvimento, no cenário mundial, a partir da década de 1950 e, no Brasil, a partir de 1967, com a lei de incentivos fiscais ao reflorestamento (Resende, 1999).

De acordo com Resende (2002), o sucesso prático de um programa de melhoramento genético de espécies perenes é fundamentado no conhecimento do produto final de interesse, das relações industriais e mercadológicas, exigências qualitativas e formas de uso pelo consumidor. Para a obtenção do produto final é necessário o conhecimento do germoplasma disponível, da sua variação biológica entre espécies, entre populações dentro de espécies e dentro de populações. É necessário ainda conhecer os fatores ambientais que interferem na expressão fenotípica; e conhecimentos de métodos de seleção e de melhoramento.

Ao longo dos anos, várias metodologias de seleção foram desenvolvidas, para garantir o sucesso dos programas de melhoramento. No contexto do melhoramento florestal com a seleção recorrente, Resende (2007) relata cronologicamente a evolução dos métodos de seleção, a seleção baseada em famílias: i) entre médias de famílias, ii) melhor predição linear-BLP (Best Linear Prediction) de famílias, iii) BLP de famílias em vários locais e iv) máxima verossimilhança restrita/melhor predição linear não-viesada (REML/BLUP) de famílias. A seleção baseada no indivíduo: i) entre famílias e dentro de parcelas, ii) entre famílias corrigidas e dentro de família no experimento, iii) individual pela seleção combinada, iv) individual pelo índice multi-efeitos (IME) ou

BLUP individual, v) individual pela seleção combinada modificada e vi) REML/BLUP individual.

Com a exigência de métodos mais acurados de seleção, atualmente a metodologia de modelos mistos REML/BLUP desenvolvida por Henderson (1975) é o procedimento ótimo de avaliação genética no melhoramento florestal, devido suas importantes características como correção simultânea dos dados para os efeitos ambientais, comparação de indivíduos no tempo e espaço, produção de resultados não viciados e maximização da acurácia de seleção. Esta metodologia é indicada para dados desbalanceados, pois maximiza o ganho genético e a eficiência dos programas de melhoramento além de considerar o parentesco genético entre os indivíduos (Resende, 2007; Pires et al., 2011).

O REML/BLUP tem sido aplicado a diversas culturas perenes como o eucalipto (Rosado et al., 2009), pinus (Missio et al., 2005), pupunheira (Farias Neto e Resende, 2001), seringueira (Arantes et al., 2010), teca (Costa et al., 2007), dendê (Purba et al., 2001) e também a culturas anuais como o feijão (Resende et al., 2001) e o milho (Viana et al., 2010) demonstrando a importância desse procedimento na avaliação genética.

Para auxiliar o melhorista na avaliação de dados, é necessária a utilização de softwares voltados para a análise genética. Os softwares como o ASREML (Butler et al., 2009) e SAS (Statistical Analysis System) permitem o ajuste de um número infinito de modelos e modelos mais complexos. Entretanto, o SAS apresenta, atualmente, limitação em lidar com matriz de parentesco e prever valores genéticos em nível individual (Resende, 2002). Softwares como Selegen REML/BLUP (Resende, 2002) e GENES (Cruz, 2013) são intensamente utilizados em pesquisas em Universidades, empresas particulares e institutos públicos de pesquisa em nosso país. Possuem a característica de serem gratuitos para instituições públicas e são de fácil manuseio.

Outro software que possui característica de ser livre para *download* e não, necessariamente, voltado para análises genéticas, mas auxilia por sua ampla gama de pacotes estatísticos é o R (R development Core Team, 2013), que possui uma linguagem dirigida à análise gráfica e estatística, semelhante a linguagem S e tem sido amplamente utilizado na área de genética e melhoramento (Paternelli et al., 2009; Vazquez et al., 2009). Diversos pacotes voltados a análise genética são criados no R e disponibilizados gratuitamente no site www.r-project.org. Neste trabalho, maior atenção é dada aos pacotes que lidam com ajustes de modelos lineares mistos, como o *lme4* (Bates et al., 2012) e *pedigreemm* (Vazquez et al., 2009), que utiliza as mesmas funções do *lme4*, porém inclui a matriz de parentesco no ajuste do modelo.

A elaboração do pacote *pedigreemm* foi realizada para análise de dados de melhoramento animal e sua utilização para o melhoramento vegetal é possível, porém depende de pequenas modificações no arquivo de dados e correção da saída dos resultados em relação ao valores genéticos preditos.

Neste contexto, o presente trabalho teve como objetivo geral o estudo dos modelos lineares mistos no melhoramento vegetal. Os objetivos específicos foram:

1. Objetivou-se a elaboração de um *script* no *software* R utilizando o pacote *pedigreemm* para avaliação de dados de melhoramento vegetal, aplicando-se as modificações no arquivo e correções dos valores genéticos preditos pelo *pedigreemm*, para cada tipo de família considerada e o fornecimento de um exemplo prático de análise com progênies de meio irmãos de *Eucalyptus urophylla* S.T. Blake.;
2. Investigar a melhor estimativa da acurácia de predição entre as metodologias de seleção combinada, modelos mistos progênies de meio irmãos simuladas.

REFERÊNCIAS

- Arantes, F.C., Gonçalves, P.S., Scaloppi Junior, E.J., Moraes, M.L.T., Resende, M.D.V. (2010). Ganho genético com base no tamanho efetivo populacional de progênies de seringueira. **Pesq. agropec. bras.**, Brasília, v.45, n.12, 1419-1424p.
- Bates, D., Maechler, M., Bolker, B. (2012). lme4: Linear Mixed-Effects Models Using Eigen and S4. **R package**, version 0.999999-0, URL <http://CRAN.R-project.org/>. Acesso: 23 abril 2013.
- Borém, A., Fristche-Neto, R. (2013). **Biotecnologia aplicada ao melhoramento de plantas**. 1. ed. Visconde do Rio Branco, MG: Suprema Gráfica e Editora, v. 1. 336p.
- Borém, A., Miranda, G.V. (2005). **Melhoramento de plantas**. 4 ed. Viçosa: UFV, 525p.
- Butler, D.G, Cullis, B.R., Gilmour, A.R., Gogel, B.J. (2009). **Mixed Models for S language environments ASReml-R reference manual**. Version 3. URL <http://www.vsnr.co.uk/software/asreml/user-area/tutorials/>. Acesso: 23 abril 2013.
- Costa, R.B., Resende, M.D.V., Silva, V.S.M. (2007). Experimentação e seleção no melhoramento genético de TECA (*Tectona grandis* L.f.). **Floresta e ambiente**, v.14, n.1, 76-92p.
- Cruz, C.D. (2013). GENES, a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, Maringá, v.35, 271-276p. Doi 10.4025/actasciagron. V35i3.21251
- Farias Neto, J.T., Resende, M.D.V. (2001). Aplicação da metodologia de modelos mistos (REML/BLUP) na estimação de componentes de variância e predição de valores genéticos em pupunheira (*Bactris gasipaes*). **Revista Brasileira de Fruticultura**, Jaboticabal - SP, v. 23, n. 2, 320-324p.
- Henderson, C.R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, Vol. 31, n. 2, 423-447p. URL <http://www.jstor.org/stable/2529430>. Acesso: 09 setembro 2009.
- Missio, R.F., Silva, A.M., Dias, L.A.S., Moraes, M.L.T., Resende, M.D.V. (2005). Estimates of genetic parameters and prediction of additive genetic values in *Pinus kesiya* progenies. **Crop Breeding and Applied Biotechnology**, 394-401p.
- Peternelli, L.A., Ferreira, F.M., Rocha, R.B., Barros, W.S., Barbosa, M.H.P. (2009). Melhoramento genético vegetal: análise dos coeficientes de endogamia e de parentesco para qualquer nível de ploidia usando o pacote estatístico R. **Bragantia**, Campinas, v.68, n.4, 849-855p.

Pires, I.E., Resende, M.D.V., Silva R.L., Resende Júnior, M.F.R. (2011). **Genética florestal**. Viçosa-MG: Arka. 318p.

Purba, A.R., Flori, A., Baudouin, L., Hamon, S. (2001). Prediction of oil palm (*Elaeis guineensis*, Jacq.) agronomic performances using the best linear unbiased predictor (BLUP). **Theoretical and Applied Genetics**, 787–792p.

R Development Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.r-project.org>.

Resende, M.D.V., Silva, F.F., Lopes, P.S., Azevedo C. F.(2012). **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: Universidade Federal de Viçosa/ Departamento de Estatística. 291p. URL http://www.det.ufv.br/ppestbio/corpo_docente.php

Resende, M.D.V. (2007). **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 362p.

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

Resende, M.D.V. (2002). **Software Selegen –REML/BLUP**. Documentos Embrapa Florestas, Colombo, n.77, 67p.

Resende, M.D.V., Furlani-Júnior, E., Moraes, M.L.T., Fazuoli, L.C. (2001). Estimativas de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/BLUP. **Bragantia**, Campinas, 185-193p.

Resende, M.D.V de. (1999). **Predição de valores genéticos, componentes de variância, delineamentos de cruzamento e estrutura de populações no melhoramento florestal**. Tese de Doutorado. Universidade Federal do Paraná, Curitiba-PR, 434p.

Rosado, A.M., Rosado, T.B., Resende Júnior, M.F., Bhering, L.L., Cruz, C.D. (2009). Ganhos genéticos preditos por diferentes métodos de seleção em progênies de *Eucalyptus urophylla*. **Pesquisa agropecuária brasileira**, Brasília, v.44, n.12, 1653-1659p.

Viana, J.M.S., Almeida, I.F., Resende, M.D.V., Faria, V.R., Silva, F.F. (2010). BLUP for genetic evaluation of plants in non-inbred families of annual crops. **Euphytica**, 174:31–39p.

Vazquez, A.I., Bates, D.M., Rosa, G.J.M, Gianola, D., Weigel, K.A.(2009). Technical note: An R package for fitting generalized linear mixed models in animal breeding. **Journal of Animal Science**.88:497 -504. Doi:10.2527/jas.2009-1952.

Artigo 1

Uso do pacote *pedigreemm* para análises de modelos mistos no melhoramento vegetal

Resumo - A metodologia de modelos mistos com as equações de Henderson é bastante difundida e utilizada por diversos segmentos da pesquisa, tanto animal como vegetal. Softwares como o R, tem sido amplamente utilizado na modelagem genética-estatística. Com o objetivo de otimizar e difundir a utilização do pacote *pedigreemm* no *software* R, objetivou-se a elaboração de um *script* para avaliação de dados de melhoramento vegetal, aplicando-se as modificações no arquivo e correções dos valores genéticos preditos pelo *pedigreemm*, para cada tipo de família considerada e o fornecimento de um exemplo prático de análise com progênies de meio irmãos de *Eucalyptus urophylla* S.T. Blake. A construção do *script* foi demonstrado passo a passo para o exemplo prático considerado, com as explicações necessárias para o entendimento do usuário. Com isso, apresentou-se a utilização do *pedigreemm* para estimação de componentes de variância e predição de valores genéticos, para dados provenientes de melhoramento vegetal via REML/BLUP.

Palavras-chave: valor genético aditivo, REML/BLUP, progênies, lme4.

1. Introdução

A avaliação genética de caracteres quantitativos no melhoramento vegetal requer grande precisão experimental e utilização de metodologias estatísticas adequadas para a seleção dos indivíduos. E nesse contexto, o parâmetro mais importante a ser utilizado para seleção genética é a acurácia (Resende, 2007).

No melhoramento de espécies perenes, em que a população de melhoramento possui longos ciclos de avaliação e a mensuração sempre ocorre mais de uma vez no mesmo indivíduo, além de ocorrer perdas nas parcelas por diversos motivos, a utilização dos modelos mistos via REML/BLUP (máxima verossimilhança restrita e melhor predição linear não viesada) tem sido o procedimento mais adequado, para a seleção de indivíduos superiores por considerar estas peculiaridades (Resende, 2007).

Na modelagem genética-estatística softwares como o R (R Development Core Team, 2013) são amplamente utilizados para o desenvolvimento de pacotes relacionados a esta. Dentre os pacotes criados, para a utilização de modelos lineares mistos voltados à estimação de componentes de variância e predição de valores genéticos, pode se citar o *pedigreemm* (Vazquez et al., 2009). Este tem a capacidade do *lme4*, para ajustar modelos lineares mistos, com a diferença de que permite a correlação genética entre os níveis dos efeitos aleatórios do modelo, com a incorporação da matriz de parentesco.

Entretanto, a utilização desse pacote em dados de melhoramento vegetal, com delineamentos genéticos de famílias, não é viável, devido o *pedigreemm* não considerar a informação dos genitores na matriz de incidência dos efeitos aditivos, gerando predições incorretas para os indivíduos. Resende et al. (2012) criaram equações que corrige a saída do *pedigreemm* para famílias de meio irmãos e irmãos germanos, considerando também algumas modificações, como a inclusão das informações dos genitores e a duplicação da última linha do arquivo de dados, para que o pacote pudesse ler e ajustar as referidas informações de famílias.

Para otimizar e difundir a utilização do pacote *pedigreemm* no *software* R, objetivou-se a elaboração de um *script* para avaliação de dados de melhoramento vegetal, aplicando-se modificações no arquivo de dados e correções dos valores genéticos preditos pelo *pedigreemm*, para cada tipo de família considerada e o fornecimento de um exemplo prático de análise com progênies de meio irmãos de *Eucalyptus urophylla* S.T. Blake.

2. Base teórica da estimação e predição

A acurada estimação de parâmetros e predição do valor genético aditivo de indivíduos é desejável em um programa de melhoramento. A estimação de parâmetros genéticos permite identificar a natureza da ação gênica, responsáveis pelo controle de caracteres quantitativos, sendo de grande utilidade na avaliação das diferentes estratégias de melhoramento (Cruz e Carneiro, 2006).

A metodologia de modelos mistos, com a Melhor Estimação/Predição Linear Não-Viesada REML/BLUP de Henderson (1975), tem sido amplamente utilizada no melhoramento de espécies perenes. Isso é devido às vantagens práticas que esse procedimento oferece como a comparação de indivíduos ou variedades através do tempo e espaço, permitir a simultânea correção para os efeitos ambientais, estimação de componentes de variância e predição dos valores genéticos, permite também lidar com

estruturas complexas de dados (medidas repetidas, diferentes anos, locais e delineamentos), entre outros (Resende, 2007).

A avaliação genética deve ser realizada sobre valores genéticos e não em médias fenotípicas, pois essas não se repetirão em plantios futuros (Resende e Duarte, 2007). Por isso é enfatizado que o melhor procedimento de seleção de indivíduos superiores é o REML/BLUP, que maximiza a acurácia de predição e tem a propriedade *shrinkage* da média, que é uma propriedade de estimadores relatada na teoria de modelos lineares mistos, que compreende ao “encurtamento” da distribuição das médias ajustadas de tratamentos, em torno da média geral, quando passa de uma análise assumindo-os como de efeitos fixos para outra em que tais efeitos são tidos como aleatórios (Duarte e Vencovsky, 2001).

Considerando a seguinte equação de um modelo linear misto:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

em que \mathbf{y} é o vetor de dados; $\boldsymbol{\beta}$ é o vetor dos efeitos de repetição (fixos) somados a média geral, \mathbf{a} é vetor dos efeitos genéticos aditivos individuais (aleatórios), \mathbf{e} é o vetor de erros ou resíduos (aleatórios). As letras \mathbf{X} e \mathbf{Z} representam as matrizes de incidência para os referidos efeitos, respectivamente. Assumindo-se como conhecidos a matriz de variância e covariâncias dos efeitos aleatórios \mathbf{G} e a matriz de variância e covariâncias residual \mathbf{R} , a simultânea estimação dos efeitos fixos e predição dos efeitos aleatórios podem ser obtidas pelas equações de modelos mistos dados por (Henderson, 1984):

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Assume-se que $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$; $E(\mathbf{a}) = E(\mathbf{e}) = 0$ e os efeitos residuais, que incluem efeitos ambientais e genéticos não aditivos são independentemente distribuídos, com variância σ_e^2 , entretanto $\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2 = \mathbf{R}$; $\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$; $\text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2 = \mathbf{G}$ e $\text{cov}(\mathbf{a}, \mathbf{e}) = \text{cov}(\mathbf{e}, \mathbf{a}) = 0$, em que \mathbf{A} é o numerador do coeficiente de parentesco de Wright.

Esse modelo apresentado por Henderson é, atualmente, a forma padrão do modelo misto referenciado na maioria das pesquisas estatísticas e implementado em programas computacionais (Littell, 2002).

Os efeitos fixos podem ser estimados pelo Melhor Estimador Linear Não - Viesado (BLUE), enquanto que os efeitos aleatórios podem ser preditos pelo Melhor Preditor Linear Não - Viesado (BLUP). Na prática BLUE e BLUP são substituídos por

suas estimativas, BLUE e BLUP empíricos, respectivamente, sendo obtidos preferencialmente pelo método da Máxima Verossimilhança Restrita (REML).

3. Metodologia

Neste trabalho é proposta a estimação de componentes de variância e predição de valores genéticos aplicáveis a modelos de progênies de meio irmãos e de irmãos germanos, avaliadas em delineamento de blocos casualizados (DBC), aos quais foram utilizados quatro tipos de rotinas de programação no R descritas a seguir:

Caso 1: Modelos de progênies de meio irmãos com uma planta na parcela

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \quad (1)$$

Em que \mathbf{y} é o vetor de observações; \mathbf{b} é o vetor dos efeitos de repetição (fixos) somados a média geral, \mathbf{a} é o vetor dos efeitos genéticos aditivos individuais (aleatórios), \mathbf{e} é o vetor de erros ou resíduos (aleatórios). As letras \mathbf{X} e \mathbf{Z} representam as matrizes de incidência para os referidos efeitos, respectivamente.

Caso 2: Modelos de progênies de meio irmãos com várias plantas por parcela

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wp} + \mathbf{e} \quad (2)$$

Em que \mathbf{y} é o vetor de observações; \mathbf{b} é o vetor dos efeitos de repetição (fixos) somados a média geral, \mathbf{a} é o vetor dos efeitos genéticos aditivos individuais (aleatórios), \mathbf{p} é o vetor dos efeitos de parcela (aleatórios) \mathbf{e} é o vetor de erros ou resíduos (aleatórios). As letras \mathbf{X} , \mathbf{Z} e \mathbf{W} representam as matrizes de incidência para os referidos efeitos, respectivamente.

As equações de modelos mistos para os dois casos são:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda_1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}, \text{ uma planta} \quad (3)$$

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\lambda_2 & Z'W \\ W'X & W'Z & W'W + I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}, \text{ várias plantas} \quad (4)$$

$$\lambda_1 = \frac{1-h_a^2}{h_a^2}, \lambda_2 = \frac{1-h_a^2-c^2}{h_a^2}, \lambda_3 = \frac{1-h_a^2-c^2}{c^2};$$

Em que σ_a^2 é a variância aditiva, σ_e^2 é a variância residual e σ_p^2 é a variância de parcela; λ_1 , λ_2 e λ_3 são fatores associados à propriedade *shrinkage* dos efeitos aleatórios

de indivíduo e o último, aos efeitos aleatórios de parcelas e c^2 é o coeficiente de determinação dos efeitos da parcela, dado por $c^2 = \frac{\sigma_p^2}{\sigma_{fenotípica}^2}$.

No caso 1, a expressão da herdabilidade individual no sentido restrito é dada por $h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$; e no caso 2, $h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_p^2 + \sigma_e^2}$

Caso 3: Modelos de progênies de irmãos germanos com uma planta na parcela;

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e} \quad (5)$$

Em que \mathbf{y} é o vetor de observações; \mathbf{b} é o vetor dos efeitos de repetição (fixos) somados a média geral, \mathbf{g} é o vetor dos efeitos genotípicos individuais (aleatórios), \mathbf{e} é o vetor de erros ou resíduos (aleatórios). As letras \mathbf{X} e \mathbf{Z} representam as matrizes de incidência para os referidos efeitos, respectivamente.

Caso 4: Modelos de progênies de irmãos germanos com várias plantas na parcela.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{Wp} + \mathbf{e} \quad (6)$$

Em que \mathbf{y} é o vetor de observações; \mathbf{b} é o vetor dos efeitos de repetição (fixos) somados a média geral, \mathbf{g} é o vetor dos efeitos genotípicos individuais (aleatórios), \mathbf{p} é o vetor dos efeitos de parcela e \mathbf{e} é o vetor de erros ou resíduos (aleatórios). As letras \mathbf{X} , \mathbf{Z} e \mathbf{W} representam as matrizes de incidência para os referidos efeitos, respectivamente.

As equações de modelos mistos para as famílias de irmãos germanos são idênticas às apresentadas anteriormente nas equações (3) e (4), com exceção, que aqui será considerada a presença do vetor dos efeitos genotípicos individuais.

As distribuições e estruturas de médias e variâncias são:

$$y|b, V \sim N(Xb, V)$$

$$g|\sigma_g^2 \sim N(0, \sigma_g^2)$$

$$a|A, \sigma_a^2 \sim N(0, A\sigma_a^2)$$

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

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

3.1 Material

Como aplicação de um dos *scripts* desenvolvido, foi utilizado como exemplo o conjunto de dados de teste de progênies de meio irmãos de *Eucalyptus urophylla* S.T.

Blake do programa de melhoramento da empresa Klabin S.A., plantado no ano de 2006, com quatro anos de idade, instalado na fazenda da empresa em Telêmanco Borba-PR.

O experimento consistiu no delineamento de blocos ao acaso com uma única planta por parcela (Single Tree Plot), com 30 repetições, 122 progênies testadas provenientes de uma área de produção de sementes, no espaçamento 3,0 x 2,5 m. As variáveis respostas utilizadas foram altura (H), medida em metros (m) e diâmetro à altura do peito (DAP), medido em centímetros (cm).

Tabela 1: Exemplo de planilha de dados fornecido pelo usuário para a construção do arquivo padrão a ser utilizado pelo pacote *pedigreemm*

Id	Prog	Bloc	DAP	H
1	1	1	0	0
2	1	2	9.87	16.3
3	1	3	19.74	23.6
4	1	4	6.05	8.0
.
.
.
3660	122	30	18.14	22.0

Id- indivíduo, Prog- progênie, Bloc- bloco, DAP- diâmetro a altura do peito, H- altura.

3.2 Descrição do Script

Nesta seção será ilustrada a construção do arquivo padrão, a partir dos dados utilizados como exemplo de aplicação acima citados, para a correção da matriz de incidência dos efeitos aleatórios dos indivíduos para as análises das progênies de meio irmãos.

O arquivo padrão é criado a partir do próprio conjunto de dados fornecido, sendo necessário que a ordenação da planilha seja igual ao da Tabela 1. Para o caso de mais de uma informação na parcela, a coluna de parcela deve ser colocada após a coluna de blocos. O arquivo padrão será lido e processado da seguinte forma no R:

Passo 1: Estruturação do arquivo de dados

Devem ser seguidos os seguintes procedimentos:

- 1) Carregar o pacote e proceder a leitura do conjunto de dados originais

```
library(pedigree)
dados<-read.table("MODELO2_FMI.txt",h=T)
```

- 2) Montagem do arquivo, segundo o procedimento do pacote *pedigreemm*

Devem-se criar linhas iniciais no arquivo de dados (122 linhas), correspondentes às informações de genitores para a geração da matriz de incidência (**Z**) dos efeitos

aleatórios do modelo individual, e cada genitor receberá um único bloco diferente dos demais. Lembrando-se que essa matriz terá dimensão 3660 (indivíduos mensurados) x 3782 (genitores – referentes as 122 linhas iniciais e indivíduos mensurados).

A inclusão das linhas dos genitores no conjunto é necessária para que os componentes de variância e os valores genéticos sejam corrigidos.

3) Criar linhas adicionais referentes a informações de genitores

```
Id : Identificação do indivíduo - numérico, de 1 até o número de
indivíduos.
# Prog :

num.individuos <- nrow(dados)# número de linhas do arquivo dados
num.familias <- length(unique(dados$Prog))
num.blocos <- length(unique(dados$Bloc))

num.variaveis <- ncol(dados)-3
col.parc <- 0

colnames(dados)[1:2]<-c("Id","G1") # nomeando as colunas

G2<- rep(NA,num.individuos)
dados<-cbind(dados,G2)
if (sum(colnames(dados)=="Parc")==TRUE) {
  dados<-(dados[,c(1:2,ncol(dados), (ncol(dados)-(2+num.variaveis)):
(ncol(dados)-1) )]) # caso exista parcela

  }else {

  dados<-(dados[,c(1:2,ncol(dados), (ncol(dados)-
(1+num.variaveis)):ncol(dados)-1) )]) }
```

4) Criar e organizar colunas

```
Id<-seq(1:num.familias);
G1<-rep(NA,num.familias); #criação da coluna de genitor masculino (sire)
G2<-rep(NA,num.familias); # e feminino (dam)
Bloc<-seq(1:num.familias);

variaveis<- matrix(data= (0), nrow = num.familias, ncol = num.variaveis)
colnames(variaveis) <- colnames(dados[(5+col.parc):ncol(dados)]) #
identifica os nomes das colunas de variáveis
```

5) Criar cabeçalho das novas linhas criadas, incluindo a coluna de G2

```
Cabecalho<- (cbind(Id,G1,G2,Bloc,variaveis))

dados$Id<- dados$Id + num.familias
dados$Bloc<- dados$Bloc + num.familias

dados<-rbind(Cabecalho, dados)
```

6) Duplicação da linha do último indivíduo do conjunto de dados

Essa duplicação é necessária, sendo utilizada como um artifício computacional, uma vez que o pacote *pedigreemm* foi elaborado para análise de dados de melhoramento animal, que possui uma estrutura de medidas repetidas, em que cada indivíduo tem várias medições no curto espaço de tempo. Sendo que essa estrutura de mensuração nem sempre é aplicada para o melhoramento vegetal. Essa última linha, se ajustada como

efeito fixo atribuindo um bloco específico, terá suas informações desconsideradas, não prejudicando na estimação e predição do REML/BLUP.

É importante frisar que essa linha duplicada receba, quando existir parcela, qualquer número diferente das parcelas existentes no arquivo, mesmo para as linhas referentes aos genitores.

```
dados<- rbind(dados,dados[nrow(dados),])
dados[nrow(dados),4]<-99999
```

Dessa forma, o arquivo padrão criado será conforme a Tabela 2 abaixo:

Tabela 2: Exemplo de arquivo padrão para a utilização do pacote *pedigreemm* para o ajuste de dados de delineamentos de blocos ao acaso, em testes de progênies de famílias de meio irmãos, com uma planta na parcela

	Id	G1	G2	Bloc	DAP	H
Linhas iniciais de genitores	1	NA	NA	1	0	0
	2	NA	NA	2	0	0
	3	NA	NA	3	0	0
	4	NA	NA	4	0	0

	122	NA	NA	122	0	0
Progênies	123	1	NA	123	7.05	12.7
	124	1	NA	124	9.87	16.3
	125	5	NA	125	19.74	23.6

	3782	122	NA	152	18.14	22
Linha duplicada	3782	NA	NA	99999	18.14	22

(.): simbologia usada para representar as informações dos indivíduos não apresentados na tabela;

Id: indivíduo; G1 e G2: genitores masculino e feminino, respectivamente; Bloc: bloco; DAP e H: variáveis diâmetro a altura do peito e altura, respectivamente.

Os indivíduos com numeração de 1 a 122 correspondem aos genitores, cujas informações de genealogia não são conhecidas (2° e 3° colunas), por isso a presença de “NA”. A coluna **Bloc** refere-se ao bloco, sendo que cada genitor deve estar em blocos diferentes ao do delineamento experimental relacionado ao conjunto de dados.

Na coluna **Id** (indivíduos), os genitores devem estar sempre nas primeiras linhas que antecedem as progênies e receber a numeração de 1 a r genitores, fato que foi considerado no script e pode ser observada uma nova numeração dessas progênies. Os genitores **G1** (genitores masculinos - sire) e **G2** (genitores femininos - dam) devem ter dados próprios, mesmo que não existam informações sobre eles, necessitando de criação

de dados fictícios e, posteriormente, o resultado gerado pelo *pedigreemm* será corrigido, de acordo com as correções desenvolvidas por Resende et al., (2012).

Para o caso de mais de uma informação na parcela (**Parc**), os genitores receberão valores fictícios (por ex. 99999).

Passo 2: Estabelecimento da genealogia dos indivíduos

O *pedigreemm* gera as informações da genealogia dos indivíduos e no caso do arquivo padrão, a última linha deve ser retirada.

```
dados$Bloc<- as.factor(dados$Bloc)
genealogia<- dados[1:(nrow(dados)-1),1:3]      #Retira-se o indivíduo
repetido do fim do arquivo padrão
G1 <-genealogia[,2];G2 <- genealogia[,3];Ind<-genealogia[,1]
Genealogia<-pedigree(sire<-as.integer(G1),dam<-as.integer(G2),label<-
as.character(Id))
```

Passo 3: Predição dos valores genéticos

Após a geração do arquivo padrão e das informações da genealogia dos indivíduos, é realizado o ajuste do modelo (caso 1) para predição de efeitos genéticos de cada indivíduo e genitor.

1) Ajustar o modelo individual

```
for (n in 1:num.variaveis){      # o ajuste será realizado para n
variáveis presentes no arquivo
ajuste<-pedigreemm(H ~ -1 + Bloc + (1|Id) ,
data = dados,
pedigree = list(Id = genealogia))
```

Na equação o valor -1 se refere ao ajuste do modelo, desconsiderando-se a média geral.

2) Obter os componentes de variância e efeitos fixos

```
summary (ajuste)
```

3) Obter os BLUP's de cada indivíduo

```
ranef (ajuste)
```

4) Corrigir dos valores genéticos preditos

A correção dos valores genéticos preditos para os genitores é dada por:

$$\hat{a}_{gen-correto} = (y - \hat{B}loc) \quad (7)$$

Resende et al. (2012) desenvolveram as correções para os valores genéticos preditos dos indivíduos gerados pelo *pedigreemm*, para progênes de meio irmãos e de irmãos germanos, que podem ser aplicados para cada caso apresentado neste trabalho, conforme a seguir:

Caso 1: Para progênes de meio irmãos com uma planta na parcela

$$\hat{a}_i = \hat{a}_{iR} + \hat{a}_{gen-incorreto} \frac{(3/4)h_a^2}{(4-h_a^2)} + \hat{a}_{gen-correto} \frac{(1/2)(1-h_a^2)}{(1/4)(4-h_a^2)} \quad (8)$$

Caso 2: Para progênies de meio irmãos com n plantas na parcela

$$\hat{a}_i = \hat{a}_{iR} + \hat{a}_{gen-incorreto} \left(\frac{3h_a^2}{16-4h_a^2-16c^2} \right) + \hat{a}_{gen-correto} \left(\frac{2-2c^2-2h_a^2}{4-h_a^2-4c^2} \right) \quad (9)$$

Caso 3: Para progênies de irmãos germanos com uma planta na parcela

$$\hat{a}_i = \hat{a}_{iR} + (a_{G1-não-corrigido} + a_{G2-não-corrigido}) \frac{(1/4)h_a^2}{(2-h_a^2)} + (a_{G1-corrigido} + a_{G2-corrigido}) \frac{(1-h_a^2)}{(2-h_a^2)} \quad (10)$$

Caso 4: Para progênies de irmãos germanos com n planta na parcela

$$\hat{a}_i = \hat{a}_{iR} + (a_{G1-não-corrigido} + a_{G2-não-corrigido}) \frac{h_a^2}{8-4h_a^2-8c^2} + (a_{G1-corrigido} + a_{G2-corrigido}) \frac{(1-c^2-h_a^2)}{2-h_a^2-2c^2} \quad (11)$$

Em que:

\hat{a}_i : vetor de valores genéticos corretos preditos dos indivíduos;

\hat{a}_{iR} : vetor de valores genéticos incorretos dos indivíduos preditos pelo *pedigreemm*;

$\hat{a}_{gen-correto}$: vetor de valores genéticos corretos dos genitores.

$\hat{a}_{gen-incorreto}$: vetor de valores genéticos incorretos dos genitores preditos pelo *pedigreemm*; para irmãos germanos, tem-se pai (G1) e mãe (G2), $a_{G1-corrigido}$ e $a_{G2-corrigido}$, respectivamente.

h_a^2 : herdabilidade individual no sentido restrito;

c^2 : coeficiente de determinação dos efeitos da parcela.

5) Estimar os parâmetros e pesos para correção dos valores genéticos

```
VCi<- VarCorr(ajuste)      # componentes de variância
VCi[[1]][1] ->var.ind
attr(VCi,"sc")^2 -> var.res
h2.ind <- var.ind/(var.ind + var.res)
```

6) Calcular os pesos (refere – se às relações das herdabilidade) relacionados à equação de correção

```
peso1<- (3/4*h2.ind)/(4-h2.ind)
peso2<- (1/2*(1-h2.ind))/(1/4*(4-h2.ind))
blup<- ranef(ajuste)[[1]][1] # vetor ( $\hat{a}_{iR}$ )
ef.bloc<- fixef(ajuste)
```

7) Aplicar a equação de correção (8) para o caso de famílias de meio irmãos com uma planta na parcela

```
valor.genetico<- c(NULL)
for(j in 1:nrow(resultado)){
  valor.genetico[j] <- resultado[j,3] +
  peso1*resultado[j,4]+peso2*resultado[j,5]
}
```



```
resultado<- cbind(resultado, valor.genetico)
```

8) Calcular os parâmetros genéticos e ambientais

```
var.fen <- (var.ind + var.res)
media<- mean(dados[(num.familias+1):(nrow(dados)-1),4+n])
CV.gen.ind<- (sqrt(var.ind)/media)*100
CV.res <- (sqrt(0.75*var.ind+var.res)/media)*100
CV.gen.prog<- (sqrt(0.25*var.ind)/media)*100
h2.mp <- (0.25*var.ind/ (0.25*var.ind + (0.75* var.ind +
var.res)/(num.blocos*1) ) )

parametros<rbind(var.ind,var.res,var.fen,h2.ind,h2.mp,CV.gen.ind,CV.gen.
prog,CV.res,media)
colnames(parametros) <- "PARAMETROS"
```

9) Estimar o ganho de seleção e nova média

```
ganho<- NULL
for (i in 1:nrow(resultado)){
  ganho[i] <- mean(resultado[1:i,4])
}
nova.media <- ganho + media
resultado<- cbind(resultado, ganho, nova.media)
```

10) Calcular o tamanho efetivo (Ne)

A expressão para o cálculo do tamanho efetivo para progênes de meio irmãos é dada por:

$$Ne = \frac{4 N_f \bar{k}_f}{\bar{k}_f + 3 + (\sigma_{k_f}^2 / \bar{k}_f)}$$

em que: N_f – número famílias selecionados; \bar{k}_f – número médio de indivíduos selecionados por família; $\sigma_{k_f}^2$ – variância do número de indivíduos selecionados por família.

```
NE <- NULL
for(i in 1:nrow(resultado)) {
  fam.selec<- unique(resultado$Prog[1:i])
  num.fam <- length ( fam.selec )
  new.resultado<- resultado[1:i,1:2]
  num.ind.p.fam<- NULL
  for ( j in 1:length(fam.selec) ) {num.ind.p.fam[j] <-
nrow( new.resultado[new.resultado[,2]==fam.selec[j],] )
}

  if(num.fam == 1){
    kf<- mean(num.ind.p.fam) # número médio de ind. selecionado por
família
    NE[i] <- (4*(num.fam*kf))/(kf+3) } else {
    kf<- mean(num.ind.p.fam) # número médio de ind.
selecionado por família
    var.kf<- var(num.ind.p.fam) # variância do num. de ind.
selecionado por família
    NE[i] <- (4*(num.fam*kf))/(kf+3+(var.kf/kf))
}
}
resultado<- cbind(resultado, NE)
```

11) Realizar o teste da Razão de Verossimilhança

Os efeitos aleatórios foram testados pelo teste da razão de verossimilhança (LRT) para comparar a adequabilidade dos modelos com o efeito aleatório a ser testado e o modelo sem o referido efeito através da diferença entre as deviances dos referidos modelos. A deviance é definida como $D = -2 \log_e L$. O número de graus de liberdade é dado pela diferença no número de componentes de variância entre os modelos.

$$LRT = 2[\log_e L_{p+i} - \log_e L_p]$$

O LRT foi construído da seguinte forma:

```
#1- Obter as deviances de cada modelo;
Modelo completo (ajuste)
Modelo reduzido (ajuste0)
#2- Fazer a diferença do modelo reduzido com o modelo completo
```

Para obter o modelo reduzido, nesse caso de parcela com uma planta, utilizou-se o pacote *glm* do *lme4* do *software* R, para o ajuste do modelo sem o efeito aleatório de indivíduo e a sintaxe padrão está definida em `ajuste 0`. Para esse pacote é utilizado o arquivo padrão, porém sem as linhas adicionais de genitores e sem a duplicação da última linha do arquivo. O exemplo demonstrado abaixo se refere a variável altura (H).

```
ajuste <- pedigreemm(H ~ -1 + Bloc + (1|Id),
                    data = dados,
                    pedigree = list(Id = genealogia))

ajuste0 <-glm(H ~ -1 + Bloc,data=dados[(num.familias+1):
                                       (nrow(dados)-1),])

devianceML <-ajuste@deviance[1][1] # valor da deviance por ML
LRT <-deviance (ajuste0)-devianceML # (353415 - 28826.64)
LRT
324588.4
```

A significância desse valor de LRT é então testada, usando o teste de qui-quadrado com 1 grau de liberdade.

```
qchisq(0.95, 1) # 0.95 é o quantil referente a significância(5%) e 1
é o grau de liberdade.
```

```
##  $\chi^2$  tabelado 3,84 e 6,63 de significativo a 5% e 1%, respectivamente.
```

Passo 4: Saída dos Resultados

Os resultados REML/BLUP encontrados foram comparados aos Selegen REML-BLUP (Resende, 2002) e ASREML-R (Butler et al., 2009) para as devidas verificações do processamento.

1) Estimação dos componentes de variância (REML)

Será apresentada a análise da variável altura como exemplo de aplicação do script criado. A estimação dos componentes de variância via REML pode ser visualizado a seguir:

```

Linear mixed model fit by REML
Formula: H ~ -1 + Bloc + (1|Id)
Data: dados
AIC      BIC      logLik  deviance  REMLdev
28315    29288    -14002  28827     28005
Random  effects:
Groups  Name      Variance  Std.Dev.
Id      (Intercept) 8.8087   2.9624
Residual                84.7192  9.2043
Numberofobs: 3783          groups: Id, 3782

```

Através do ajuste do modelo, informações adicionais são fornecidas, como os critérios para escolha de modelos, Akaike's Information Criterion (AIC), Schwartz's Bayesian Information Criterion (BIC), Log-Verossimilhança (LogLik), deviance do modelo pela máxima verossimilhança (deviance ML) e deviance pela máxima verossimilhança restrita (REMLdev) (Bates, 2008). Para pequenos níveis do fator bloco pode ser utilizado a deviance ML para o LRT.

- 2) As estimativas dos parâmetros genéticos e ambientais podem ser visualizadas a seguir (Tabela 3):

Tabela 3: Estimativas dos parâmetros genéticos e ambientais da análise de teste de progênes de meio irmãos de *Eucalyptus urophylla*.

Estimativas dos parâmetros variável Altura	
var.ind	8.8087
var.res	84.7187
var.fen	93.4947
h2.ind	0.0942
h2.mp	0.4197
CV.gen.ind	20.6501
CV.gen.prog	10.3250
CV.res	66.4912
média	14.3724

var.ind: variância individual; var.res: variância residual; var.fen: variância fenotípica; h2.ind: herdabilidade individual; h2.mp: herdabilidade média de progênes; CV.gen.ind: coeficiente de variação genético individual; CV.gen.prog: coef. de variação genético de progênie; CV.res: coef. de variação residual e média geral do experimento para altura.

- 3) Estimação BLUE para os efeitos fixos (Blocos)

O vetor dos efeitos fixos é constituído por valores genéticos de cada genitor ($\hat{\alpha}_{gen-correto}$), seguidos pelas médias dos blocos do delineamento experimental. Isso é decorrente da criação das linhas fictícias para os genitores. Os blocos individuais

criados para cada genitor (Tabela 2) tem a função de fornecer corretamente os efeitos de cada família, para serem utilizados nas equações de correção desenvolvidas por Resende et al. (2012) para a predição dos valores genéticos corretos de cada indivíduo (\hat{a}_i)

Com a função `fixed` (ajuste) são obtidas as estimativas dos efeitos fixos.

	Fixed effects:		
	Estimate	Std. Error	t value
Bloc1	-0.91355	9.47946	-0.096
Bloc2	-2.23718	9.47946	-0.236
Bloc3	-1.54598	9.47946	-0.163
Bloc4	2.64037	9.47946	0.279
Bloc5	0.55838	9.47946	0.059
Bloc6	-2.05809	9.47946	-0.217
Bloc7	-1.10105	9.47946	-0.116
Bloc8	-1.55438	9.47946	-0.164
Bloc9	0.66472	9.47946	0.07
Bloc10	-2.7241	9.47946	-0.287
Bloc11	-3.17183	9.47946	-0.335
...
...
Bloc121	-3.79027	9.47946	-0.4
Bloc122	0.3597	9.47946	0.038
Bloc123	12.56387	0.85873	14.631
Bloc124	13.48907	0.85873	15.708
Bloc125	13.77883	0.85873	16.046
Bloc126	13.02923	0.85873	15.173
Bloc127	15.93789	0.85873	18.56
...
...
Bloc152	13.00954	0.85873	15.15
Bloc99999	21.72004	9.58958	2.265

Nota-se que do bloco 1 até o 122 são valores genéticos de cada genitor e do bloco 123 ao 152 (que equivalem aos blocos 1 a 30, dos dados originais) são as médias de cada bloco do experimento. Esses efeitos dos genitores são corretamente estimados, porém com o sinal invertido, o que deve ser corrigido (Tabela 4).

Tabela 4: Predição do valor genético aditivo dos 20 primeiros genitores, após correção do resultado do pacote *pedigreemm*.

Genitores	$\hat{a}_{gen-incorreto}$	$\hat{a}_{gen-correto}$
1	1.702941797	0.91222932
4	4.167844789	2.23262487
8	2.880675574	1.54311599
12	-4.91529668	-2.63301878
13	-1.038155399	-0.55611753
16	3.83432726	2.0539667
17	2.052092961	1.09926209
18	2.896309208	1.55149059
19	-1.236181432	-0.66219582
20	5.074595573	2.71835177
21	5.908389397	3.1649972
25	3.05264555	1.63523661
29	-0.381542762	-0.20438426

31	-0.949564805	-0.50866146
32	-3.904321668	-2.0914612
16	-2.278423712	-1.22050261
17	-2.486872168	-1.33216397
18	1.093230064	0.58561985
19	2.958843745	1.58498899
20	-2.580673973	-1.38241158

$\hat{a}_{gen-incorreto}$: vetor de valores genéticos incorretos dos genitores preditos pelo *pedigreemm*; $\hat{a}_{gen-correto}$: vetor de valores genéticos corretos dos genitores.

4) Predição BLUP (valores genéticos dos indivíduos)

As estimativas dos efeitos genéticos aditivos do pacote *pedigreemm* não possui informação de genitor, por isso a necessidade de criação das linhas próprias de genitores, apresentados anteriormente, para a construção da matriz de incidência (\mathbf{Z}) dos efeitos aditivos verdadeira para as famílias em questão. Com a função `ranef (ajuste)` são obtidas as predições dos valores genéticos de cada indivíduo do arquivo padrão.

```
ranef (ajuste)
$Id
      (Intercept)
1      1.704706E+00
2      4.174610E+00
3      2.884829E+00
4     -4.926958E+00
5     -1.041952E+00
6      3.840416E+00
7      2.054566E+00
...
...
120     1.469726E+00
121     7.072700E+00
122     -6.712051E-01
123     -9.419184E-01
124     1.893537E-01
125     7.424389E-01
...
...
3780     4.80323240E-01
3781     -1.191439109
3782     4.69820615 E-01
```

Esse vetor de predição dos efeitos aleatórios para indivíduos (\hat{a}_{iR}) gera as informações incorretas tanto dos genitores (Id 1 a 122), como as informações das progênes de meio irmãos. Com a expressão de correção (8), foram geradas as predições corrigidas para progênes de meio irmãos (Id 123 a 3782).

A seguir é apresentada a tabela de resultados gerada pelo script com as informações das 20 melhores progênes em ordem de maior valor genético aditivo (Tabela 5).

Tabela 5: Arquivo final gerado para seleção das famílias de meio- irmãos de *Eucalyptus urophylla*, considerando o modelo de blocos ao acaso com uma planta na parcela.

Id	Prog	Bloco	Valor genético	Ganho	Nova. média	Ne
821	70	19	2.867489	2.867489	17.20407	1
3611	291	19	2.743383	2.805436	17.14202	2
3521	288	19	2.733947	2.781606	17.11819	3
3614	291	21	2.733849	2.769667	17.10625	3.490909
3616	291	23	2.643525	2.744439	17.08102	3.658537
316	21	23	2.639832	2.727004	17.06359	4.645161
3526	288	23	2.554816	2.702406	17.03899	5.309255
3620	291	27	2.510109	2.678369	17.01495	5.333333
3609	291	17	2.48695	2.6571	16.99368	5.261164
3615	291	22	2.480576	2.639448	16.97603	5.150215
3617	291	24	2.463543	2.623456	16.96004	5.028571
3606	291	14	2.460904	2.60991	16.94649	4.909091
3619	291	26	2.458061	2.59823	16.93481	4.797161
3612	291	2	2.445412	2.587314	16.9239	4.694611
832	70	29	2.429595	2.576799	16.91338	5.245265
314	21	21	2.427476	2.567467	16.90405	5.818182
3605	291	13	2.423452	2.558995	16.89558	5.659731
2104	169	12	2.407203	2.550562	16.88715	6.312713
281	20	19	2.403813	2.542839	16.87942	6.976969

Id- indivíduo; Prog- progênie; Ne-tamanho efetivo.

5) Geração do arquivo de resultados

Para obter os arquivos de saída do processamento foram criados dois arquivos denominados de “parâmetros” e “resultado” da variável de interesse do estudo, para cada caso. O arquivo parâmetros conterá valores referentes às estimativas de componentes de variância estimados pelo REML e o arquivo resultado conterá as predições do valor genético individual, ganho com a seleção, nova média e tamanho efetivo.

```
write.table(resultado,paste("_Resultado_Variavel_",colnames(dados)[4+n],
".txt",sep = ""))
write.table(parametros, paste("_Parametros_Variavel_",
colnames(dados)[4+n], ".txt",sep = "")) }
```

4. Conclusão

A utilização do pacote *pedigreemm* fornecido no CRAN do software R é possível para análises genéticas de dados de melhoramento vegetal com as devidas correções aplicadas a cada caso desenvolvido, devendo-se respeitar a forma que foi elaborada os exemplos de cada script.

É importante frisar que a linha duplicada do conjunto de dados foi utilizada como recurso para viabilizar o uso computacional da forma original do pacote *pedigreemm* para os dados de melhoramento vegetal apresentados.

Os scripts criados para os delineamentos de blocos ao acaso, com uma ou várias informações na parcela para as famílias de meio irmãos e irmãos germanos, serão

disponibilizados no programa computacional GENES, que estará disponível para download no link <http://www.ufv.br/dbg/genes/genes.htm>, juntamente com os exemplos de cada análise.

Agradecimentos

Os autores agradecem a empresa Klabin S.A. pela concessão dos dados para a elaboração desse artigo e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro.

5. Apêndice

Script completo para progênies FMI uma plantas na parcela

```
rm(list=ls(all=TRUE))

# carregamento do pacote pedigreemm
library(pedigreeem)
# leitura dos dados BLUP-INDIVIDUO
#setwd("C:\\Dani_Pinto\\FMI - 22.02.13\\FMI_sem_parcela") #
definição da trilha de dados

dados=read.table("MODELO2_FMI.txt",h=T) # 1 planta por parcela

# Informações Básicas Sobre O Arquivo De Dados
# o arquivo deve ser construído da seguinte maneira:
# Id : Identificação do indivíduo - numérico, de 1 até o número de indivíduos.
# Prog :
num.individuos <- nrow(dados)
num.familias <- length(unique(dados$Prog))
num.blocos <- length(unique(dados$Bloc))
num.variaveis <- ncol(dados)-3

# Montagem do arquivo de dados segundo o modelo do pedigreeem
## Organizando a tabela de dados
colnames(dados)[1:2]<-c("Id","G1") # nomeando as colunas
G2<- rep(NA,num.individuos)
dados<-cbind(dados,G2)
if (sum(colnames(dados)=="Parc")==TRUE) {
  dados<- (dados[,c(1:2,ncol(dados), (ncol(dados)-(2+num.variaveis)) : (ncol(dados)-1) )])
} else {
  dados<- (dados[,c(1:2,ncol(dados), (ncol(dados)-(1+num.variaveis)) : (ncol(dados)-1) )])
}

## Criação das linhas iniciais do arquivo de dados modificado
Id<-seq(1:num.familias); G1<-rep(NA,num.familias); G2<-rep(NA,num.familias)
Bloc<-seq(1:num.familias);
variaveis <- matrix(data= (0), nrow = num.familias, ncol = num.variaveis)
colnames(variaveis) <- colnames(dados[(5+col.parc):ncol(dados)]) # pega os nomes das colunas de
variáveis
Cabecalho<- (cbind(Id,G1,G2,Bloc,variaveis))
dados$Id <- dados$Id + num.familias
dados$Bloc <- dados$Bloc + num.familias
dados<-rbind(Cabecalho, dados)

# duplicação da última informação
dados<- rbind(dados,dados[nrow(dados),])
```

```

# Construção do arquivo de pedigree a partir do arquivo de dados
dados$Bloc <- as.factor(dados$Bloc)
genealogia <- dados[1:(nrow(dados)-1),1:3] # Retira-se o indivíduo repetido do fim do arq. de dados
modificado
G1 = genealogia[,2];    G2 = genealogia[,3];    Ind = genealogia[,1]
genealogia = pedigree(sire= as.integer(G1), dam=as.integer(G2), label=as.character(Ind))
# Modelo para predição de efeitos genéticos de indivíduos
for (n in 1:num.variaveis){
# Modelo para predição de efeitos genéticos de indivíduos
ajuste= pedigreemm(dados[,4+n] ~ -1 + Bloc + (1|Id) ,
                    data = dados,
                    pedigree = list(Id = genealogia))
# Estimação de parâmetros e de pesos para correção dos valores genéticos
# Script para cálculo de efeitos, pesos e herdabilidades para delineamento DBC-FMI com 1 planta por
parcela
    VCi <- VarCorr(ajuste) # componentes de variância
    VCi[[1]][1] -> var.ind
    attr(VCi,"sc")^2 -> var.res
    h2.ind <- var.ind/(var.ind + var.res)
    peso1 <- (3/4*h2.ind)/(4-h2.ind)
    peso2 <- (1/2*(1-h2.ind))/(1/4*(4-h2.ind))
    blup <- ranef(ajuste)[[1]][1]
    ef.bloc <- fixef(ajuste)

#Confecção do arquivo de resultados
resultado <- NULL
# Script para cálculo de efeitos, pesos e herdabilidades para delineamento dbc-fmi com 1 planta por
parcela
    resultado <- cbind(dados[ (num.familias + 1):( num.individuos +
num.familias),1:2],blup[(num.familias + 1):( num.individuos + num.familias),])
    colnames(resultado) <- c("Id", "Prog", "VG.id inc.")
    v.gen.fam <- c(NULL)
        for(j in 1:nrow(resultado)){
            for(k in 1:num.familias){
                if (resultado[j,2] == k) { v.gen.fam[j] <- blup[[1]][k]
            }
        }
    resultado <- cbind(resultado, v.gen.fam )
    ef.bloc.fam <- -ef.bloc[1:num.familias]
    ef.bloc <- c(NULL)
        for(j in 1:nrow(resultado)){
            for(k in 1:num.familias){
                if (resultado[j,2] == k) { ef.bloc[j] <- ef.bloc.fam[k]
            }
        }
    resultado <- cbind(resultado, ef.bloc )
    valor.genetico <- c(NULL)
    for(j in 1:nrow(resultado)){
        valor.genetico[j] <- resultado[j,3] + peso1*resultado[j,4] +
peso2*resultado[j,5]
    }
    resultado <- cbind(resultado, valor.genetico)

## Formatação do arquivo de resultados
    resultado <- resultado[,-c(3:5)]
    Blocos <- as.numeric(dados$Bloc) ; Blocos <- Blocos[-c(1:num.familias)] ;
Blocos <- Blocos[-length(Blocos)] ; Blocos <- Blocos - num.familias
    resultado <- cbind(resultado[,1:2], Blocos,resultado[,3])
    resultado[,1] <- resultado[,1]-num.familias

```



```

colnames(resultado) <- c("Id", "Prog", "Bloc", "Valor_genetico")
#Impressão dos arquivos de resultados
resultado <- resultado[with(resultado, order(Valor_genetico, decreasing = TRUE)),]
# Resultado com valores genéticos ordenados
var.fen <- (var.ind + var.res)
media <- mean(dados[(num.familias+1):(nrow(dados)-1),4+n])
CV.gen.ind <- (sqrt(var.ind)/media)*100
CV.res <- (sqrt(0.75*var.ind+var.res)/media)*100
CV.gen.prog <- (sqrt(0.25*var.ind)/media)*100
h2.mp <- (0.25*var.ind / (0.25*var.ind + (0.75* var.ind + var.res)/(num.blocos*1) ) )
acuracia <- sqrt(h2.mp)
h2.ad <- (.75*var.ind) / (.75*var.ind + var.res)
PEV <- ((1- acuracia**2)*var.ind/4) # variância do erro de predição
SEP <- sqrt(PEV) # Desvio padrão do erro de predição
parametros <- rbind(var.ind,
var.res,
var.fen,
h2.ind,
h2.mp,
acuracia,
h2.ad,
PEV,
SEP,
CV.gen.ind,
CV.gen.prog,
CV.res,
media )
colnames(parametros) <- "PARAMETROS"
parametros
ganho <- NULL
for (i in 1:nrow(resultado)){
ganho[i] <- mean(resultado[1:i,4])
}
nova.media <- ganho + media
resultado <- cbind(resultado, ganho, nova.media)

## Cálculo do Tamanho EFETIVO
NE <- NULL
for(i in 1:nrow(resultado)) {
fam.selec <- unique(resultado$Prog[1:i])
num.fam <- length ( fam.selec )
new.resultado <- resultado[1:i,1:2]
num.ind.p.fam <- NULL
for ( j in 1:length(fam.selec) ) {
num.ind.p.fam[j]<-nrow
(new.resultado[new.resultado[,2]==fam.selec[j],] ) }
if(num.fam == 1){
kf <- mean(num.ind.p.fam) # número médio de ind. selec por familia
NE[i] <- (4*(num.fam*kf))/(kf+3)
} else {
kf <- mean(num.ind.p.fam) # número médio de ind. selec por
familia
var.kf <- var(num.ind.p.fam) # variância do num. de
ind. selec. por
NE[i] <- (4*(num.fam*kf))/(kf+3+(var.kf/kf))
}
}
resultado<- cbind(resultado, NE)

# Geração do arquivo de resultados
# Modelo para delineamento dbc-fmi com 1 planta por parcela
write.table(resultado, paste("_Resultado_Variavel_", colnames(dados)[4+n], ".txt", sep = ""))
write.table(parametros, paste("_Parametros_Variavel_", colnames(dados)[4+n], ".txt", sep = ""))

```

```
# Término do for para análise do número de variáveis  
} # for para o número de variáveis
```

6. Referências

Bates, D., Maechler, M. 2008. **The Comprehensive R Archive Network**. URL <http://cran.r-project.org/>.

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J. 2009. **Mixed models for S language environments: ASREML-R reference manual**. Version 3. URL: <http://www.vsnr.co.uk/downloads/asreml/release3/UserGuide.pdf>.

Cruz, C.D., Carneiro, P.C.S. 2006. **Modelos biométricos aplicados ao melhoramento genético**. Editora UFV, Viçosa, v.2, 585p.

Duarte, J.B., Vencovsky, R. 2001. Estimacão e predicão por modelo linear misto com ênfase na ordenacão de médias de tratamentos genéticos. **Scientia Agricola**, v.58, n.1, p.109-117.

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. **Biometrics** 31:423-447.

Henderson, C.R. 1984. **Applications of Linear Models in Animal Breeding**. 462 p. Guelph: University of Guelph.

Littell, R.C. 2002. Analysis of unbalanced mixed model data: a case study comparison of ANOVA versus REML/GLS. **Journal of Agricultural, Biological, and Environmental Statistics**, Vol. 7, Num 4, Pg 472–490. *Doi: 10.1198/108571102816*

Resende, M.D.V. de. 2002. Software SELEGEN-REML/BLUP. Colombo: Embrapa Florestas, 67 p. (Embrapa Florestas, Documentos 77).

Resende, M.D.V, Duarte, J.B. 2007. Precisão e controle de qualidade de experimentos de avaliacaão de cultivares. **Pesquisa agropecuária tropical**, 37 (3), pg182-194. URL <http://www.revistas.ufg.br/index.php/pat/article/download/1867/1773>

Resende, M.D.V. 2007. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 362p.

Resende, M.D.V., Silva, F.F., Lopes, P.S., Azevedo, C. F. 2012. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa: Universidade Federal de Viçosa/ Departamento de Estatística. 291p. URL http://www.det.ufv.br/ppestbio/corpo_docente.php

R Development Core Team (2013) **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.r-project.org>.

Vazquez, A.I., Bates,D.M., Rosa,G.J.M., Gianola D., Weigel, K.A. 2009. Technical note: An R package for fitting generalized linear mixed models in animal breeding. **Journal of Animal Science**.88:497 -504. *Doi:10.2527/jas.2009-1952*.

Análises comparativas entre os métodos de seleção combinada e o procedimento REML/BLUP em ensaios simulados de progênes de meio irmãos

RESUMO – A confiabilidade, que é o quadrado da acurácia pode ser utilizada para auxiliar o melhorista na escolha de método de seleção. O objetivo dessa pesquisa foi comparar os métodos de seleção combinada e o procedimento REML/BLUP em progênes de meio irmãos simuladas. Para isso, foi simulada uma população base de 10.000 indivíduos, genotipadas em relação a 100 locos manifestando dois alelos por loco. Em seguida foram tomadas ao acaso, 100 matrizes fornecedoras de gametas que foram fertilizados por um “pool” gamético produzido por toda população gerando 100 famílias de meio-irmãos com 100 indivíduos, com genótipos conhecidos. Também foram simuladas três características quantitativas com diferentes herdabilidades (0,20 0,40 e 0,60). Foram aplicadas diferentes níveis de desbalanceamentos variando de 10 a 50% nos dados originais, para as análises comparativas entre os métodos. As metodologias do índice combinado e REML/BLUP apresentaram confiabilidade semelhantes para seleção de progênes e de indivíduos e foram superiores a média fenotípica para seleção individual. Também foi verificado que a seleção combinada através do índice combinado e o procedimento via REML/BLUP apresentaram respostas muito semelhantes em todas as avaliações. Com o aumento da herdabilidade foi verificado que as metodologias apresentaram maiores percentuais de coincidência e menor mudança na classificação das melhores progênes selecionadas, quando comparadas com os valores genéticos verdadeiros. Os diferentes níveis de desbalanceamento tiveram maior influencia na seleção de indivíduos do que na seleção de progênes.

Palavras-chave: seleção, melhoramento vegetal, modelos mistos, avaliação genética.

1. Introdução

A prática da seleção é um processo importante nos programa de melhoramento, que deve ser praticado com acurácia elevada. Vários métodos de seleção para espécies

perenes foram desenvolvidos no Brasil, e evoluíram da seleção massal, seleção entre e dentro de progênes, seleção combinada para o uso da metodologia de modelos mistos, por meio da predição de valores genéticos em nível individual via procedimento BLUP (melhor predição linear não viesado) (Fonseca et al., 2010).

A seleção combinada é aquela em que são identificados genótipos superiores a partir da informação do indivíduo e de seus aparentados (Cruz e Carneiro, 2006). Várias pesquisas apontam a eficiência desse método sobre a seleção entre e dentro de famílias e seleção massal (Arnhold et al., 2009; Costa et al., 2000; Martins et al., 2005; Rocha et al., 2009). Porém Resende (2002) ressalta que o BLUP propicia o maior ganho genético possível em um ciclo de seleção.

O BLUP desenvolvido por Henderson (1963) consiste basicamente na predição dos valores genéticos, ajustando-se os dados para os efeitos fixos e número desigual de indivíduos, por meio da metodologia de modelos mistos (Lopes, 2005). O REML/BLUP (máxima verossimilhança restrita/ melhor preditor linear não viesado) é o procedimento ótimo para avaliação genotípica de plantas perenes, pois lidam naturalmente com o desbalanceamento e considera também as informações de parentesco entre os indivíduos (Resende, 2007).

Estudos de simulação no melhoramento genético têm sido muito utilizados na comparação de metodologias de estimação de componentes de variância e predição de valores genéticos. A grande vantagem da simulação em nível de genes é que podem ser obtidos os verdadeiros valores dos parâmetros, permitindo análise comparativa mais adequadas entre os métodos escolhidos (Carneiro Júnior et al., 2004). Assim, simular ensaios em que os valores genéticos são previamente estabelecidos tornaria mais elucidativo comparar a eficácia de diferentes metodologias em prever valores genéticos e identificar indivíduos e progênes superiores.

Na literatura científica vários métodos são comparados em relação a sua eficiência em selecionar os melhores indivíduos utilizando-se de informações fenotípicas. Entretanto, a comparação via simulação da seleção fenotípica e da seleção baseada nos valores preditos, permite uma maior compreensão da confiabilidade entre os métodos.

Dessa forma, o objetivo dessa pesquisa foi comparar os métodos de seleção do índice combinado e o procedimento REML/BLUP em progênes de meio irmãos simuladas.

2. Material e Métodos

Foi simulada uma população base de 10.000 indivíduos, genotipadas em relação a 100 locos manifestando dois alelos por loco. Em seguida foram tomadas, ao acaso, 100 matrizes fornecedoras de gametas que foram fertilizadas por um “pool” gamético produzido por toda população gerando 100 progênes de meio irmãos com 100 indivíduos, com genótipos conhecidos.

Foi considerado que as progênes simuladas foram avaliadas em experimentos em blocos ao acaso, com quatro repetições e parcelas constituídas por 25 indivíduos, semelhante ao que se faz com a cultura de milho. O efeito genotípico de cada loco e o efeito agregado foi estabelecido e, juntamente com a influência ambiental, determinaram o valor fenotípico final. Assim, o valor mensurado era dado por:

$$Y_{ijk} = \mu + g_i + b_j + \varepsilon_{ij} + \delta_{ijk}$$

Sendo:

Y_{ijk} : observação no k-ésimo indivíduo, avaliado no i-ésimo genótipo da j-ésima repetição;

μ : média geral do ensaio;

g_i : efeito da família i (i=1,2...100);

b_j : efeito do bloco j (j=1,2...4);

ε_{ij} : efeito da parcela ij;

δ_{ijk} : efeito do k-ésimo indivíduo, da i-ésima família no j-ésimo bloco (k=1,2...25).

É mostrado, na Tabela 1, o esquema do resultado da análise de variância esperado para este tipo de delineamento.

Tabela 1. Esquema da análise de variância para experimentos em blocos ao acaso, com a avaliação entre e dentro de progênes de meio irmãos

FV	GL	SQ	QM	F
Blocos	r-1	SQB	QMB	
Famílias	g-1	SQT	QMT	QMT/QME
Entre parcelas	(r-1)(g-1)	SQE	QME	
Dentro de parcelas	(n-1)gr	SQD	QMD	
Total	grn-1	SQTo		

$$CV\% = (100\sqrt{QMR}) / m ; m = \text{Média geral}$$

Para simulação dos valores fenotípicos foram adotadas pressuposições a respeito dos efeitos genéticos e ambientais. No aspecto genético, considerou-se ação aditiva dos

locos gênicos, ou seja, sem desvios de dominância entre os alelos e nenhuma interação epistática. Em relação aos efeitos ambientais, foram pressupostos ação independente e aleatória dos erros entre e dentro de parcelas, com distribuição normal. As magnitudes dos componentes de variância foram estabelecidas a partir de informações prévias sobre a média da característica iguais a 100, do coeficiente de variação experimental e da herdabilidade, em nível de médias.

Os componentes de variâncias dos efeitos aleatórios manifestavam sobre a variação fenotípica conforme as esperanças matemáticas descritas na Tabela 2.

Tabela 2. Esperanças matemáticas dos quadrados médios associados às fontes de variações de experimento em blocos ao acaso com informação entre e dentro de parcelas

FV	E(QM)		F
	Balancedos	Desbalanceados	
Blocos	$\sigma_d^2 + n\sigma_e^2 + ng\sigma_b^2$	$\sigma_d^2 + \bar{n}\sigma_e^2 + \bar{n}g\sigma_b^2$	-
Família	$\sigma_d^2 + n\sigma_e^2 + nr\sigma_g^2$	$\sigma_d^2 + \bar{n}\sigma_e^2 + \bar{n}r\sigma_g^2$	QMT/QME
Entre parcelas	$\sigma_d^2 + n\sigma_e^2$	$\sigma_d^2 + \bar{n}\sigma_e^2$	QME/QMD
Dentro de parcelas	σ_d^2	σ_d^2	

\bar{n} = média harmônica

Sendo $\frac{1}{\bar{n}} = \frac{1}{gr} \sum_{i=1}^g \sum_{j=1}^r \left(\frac{1}{n_{ij}} \right)$ e n_{ij} o número desbalanceado de indivíduos para a i-ésima progênie no j-ésimo bloco.

Para essas progênies foram simuladas três características fenotípicas (quantitativas), com herdabilidades de médias de 0,20, 0,40 e 0,60, com simulação de seus valores genéticos verdadeiros, seguindo distribuição normal $[N(0, \sigma_a^2)]$. Neste estudo foi preservado o conceito de que a herdabilidade expressa o quadrado da correlação entre os valores das médias de progênies e o valor genético verdadeiro.

Para o estudo comparativo entre os métodos, foram introduzidos diferentes níveis de desbalanceamentos aleatórios no arquivo das progênies simuladas, com a criação de um script no software R (R Core Team, 2013). Foram estabelecidos desbalanceamentos de 10, 20, 30, 40 e 50% para cada característica simulada e para efeito de comparação, também utilizou-se o arquivo balanceado. Deve ser ressaltado que a utilização do BLUP na predição de valores genéticos tem sido largamente enfatizada em ensaios desbalanceados, que normalmente ocorre na experimentação com

culturas perenes por diversas razões, sendo uma delas mortes de plantas durante seu ciclo de cultivo.

Para a predição dos valores genéticos utilizando a abordagem da seleção combinada, as informações do indivíduo e de sua progênie são combinadas em um índice pela atribuição de pesos ótimos a estas duas fontes de informação. Neste caso, foi utilizado o índice de Pires e Cruz (1996) (Cruz e Carneiro, 2006):

$$IC = bi(Y_{ijk} - \bar{Y}_{.j}) + bf(\bar{Y}_{i..} - \bar{Y}_{...})$$

Em que:

IC: índice combinado;

bi: peso do valor fenotípico individual;

bf: peso do valor fenotípico da média da progênie;

$(Y_{ijk} - \bar{Y}_{.j})$: desvio do valor fenotípico individual (Y_{ijk}) em relação à média do bloco ($\bar{Y}_{.j}$);

$(\bar{Y}_{i..} - \bar{Y}_{...})$: desvio do valor fenotípico da progênie ($\bar{Y}_{i..}$) em relação à média geral ($\bar{Y}_{...}$).

Para a predição de valores genéticos por meio da abordagem via BLUP, o modelo matricial adotado para as análises via modelo misto foi da seguinte forma:

$$y = Xb + Za + Wp + e$$

A forma matricial da equação de modelos mistos via análise univariada, está apresentada abaixo (Resende, 2002):

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W \\ W'X & W'Z & W'W + I\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}$$

$$\lambda_1 = \frac{1-h_a^2-c^2}{h_a^2}; \lambda_2 = \frac{1-h_a^2-c^2}{c^2};$$

Em que:

y: é o vetor de observações;

b: é o vetor dos efeitos de repetição (fixos) somados a média geral;

a: é o vetor dos efeitos genéticos aditivos individuais (aleatórios),

p: é o vetor dos efeitos de parcela (aleatórios);

e: é o vetor de erros ou resíduos (aleatórios).

X, *Z*, *W*: representam as matrizes de incidência para os referidos efeitos, respectivamente. *A* é o numerador do coeficiente de parentesco de Wright.

λ_1 e λ_2 : são fatores associados à propriedade *shrinkage* dos efeitos aleatórios de indivíduo;

h_a^2 : herdabilidade no sentido restrito;

c^2 : é o coeficiente de determinação dos efeitos da parcela, dado por $c^2 = \frac{\text{Variância da parcela}}{\text{Variância fenotípica}}$.

As distribuições e estruturas de médias e variâncias univariadas são:

$$y|b, V \sim N(Xb, V)$$

$$a|A, \sigma_a^2 \sim N(0, A\sigma_a^2)$$

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

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

Foram calculados o quadrado da correlação entre os valores genéticos reais (a) e os valores genéticos preditos (\hat{a}) das 20 melhores progênies e 100 melhores indivíduos, para medir a confiabilidade entre os métodos, conforme expressão abaixo:

$$r_{\hat{a}a} = \sqrt{\frac{\text{Cov}(\hat{a}, a)}{\sigma_{\hat{a}}^2 \sigma_a^2}}$$

Sendo:

$r_{\hat{a}a}$: correlação entre os valores genéticos reais e os preditos;

$\text{Cov}(\hat{a}, a)$: covariância entre os valores genéticos reais e os preditos;

σ_a^2 : variância genética aditiva paramétrica;

$\sigma_{\hat{a}}^2$: variância genética aditiva estimada.

Também foram obtidas as coincidências (%) dos selecionados, tanto das 20 progênies como dos 100 indivíduos, entre os valores genéticos verdadeiros e os obtidos pelo índice combinado, entre os valores genéticos verdadeiros e os preditos pelo REML/BLUP e entre os valores genéticos obtidos pelo índice combinado e o REML/BLUP. Para avaliar a classificação dos selecionados, para ambas metodologias, foi utilizada a correlação de Spearman.

As simulações da população genitora, progênies de meio irmãos e das características quantitativas e as predições dos valores genéticos via seleção combinada foram realizadas no aplicativo computacional GENES (Cruz, 2013), e a predição via modelos mistos foi realizada pelo software Selegen REML/BLUP (Resende, 2002).

3. Resultados e Discussão

Ao se promover seleção, além de ser importante ter-se uma idéia do que se espera com o processo, ou seja, a predição do resultado da seleção, é importante ter-se uma estimativa da segurança que se terá ao se adotar determinado procedimento. Esta segurança é definida pela acurácia da seleção. A acurácia refere-se a correlação entre o valor genotípico verdadeiro do material genético e aquele estimado ou predito a partir de experimentos.

Dessa forma, foram obtidas as acurácias dos valores genéticos verdadeiros para comparação das acurácias obtidas dos valores genéticos preditos pelo índice combinado e via REML/BLUP, dos dados balanceados representado pelo valor zero (0) e desbalanceados, para os diferentes valores de herdabilidades (Figura 1).

Como em seleção genética, o que se faz geralmente, é procurar identificar indivíduos geneticamente superiores via medidas fenotípicas, há necessidade de uma boa relação entre o fenótipo e o genótipo. Uma medida de associação entre duas variáveis é dada pelo coeficiente de correlação, que nesse caso, é a própria herdabilidade. Quanto mais relacionadas forem as duas variáveis, ou seja o fenótipo e o genótipo, mais acuradamente uma variável é predita pela outra. Logo, a acurácia de uma predição é diretamente relacionada com a magnitude do coeficiente de herdabilidade.

Dessa forma, verifica-se na Figura 1, que para as características com valor baixo de herdabilidade (0,20), as acurácias obtidas por ambos os métodos tiveram maior variação e menores magnitudes, e conforme aumentou-se a herdabilidade, essas variações foram diminuindo e percebeu-se o aumento na magnitude das acurácias.

Também foi observado a influencia dos desbalanceamentos na predição da acurácia pelos métodos e verificou-se que, em 20% de desbalanceamento para as diferentes herdabilidades, na combinação de 0,20 de herdabilidade e 40% desbalanceado e 0,60 de herdabilidade com 50% de desbalanceamento, o índice combinado apresentou maiores acurácias do que o REML/BLUP. Porém, essas diferenças na estimação da acurácia de predição, no geral, não foram tão marcantes para designar o melhor método neste estudo.

A característica com maior herdabilidade tem maior facilidade para a prática da seleção e percebeu-se na figura 1, que há pouca diferença entre as acurácias obtidas por ambos os métodos relativos a característica com herdabilidade 0,60 nos diferentes níveis de desbalanceamento. Isso indica que a utilização do índice combinado ou do REML/BLUP para a seleção e predição de ganhos será muito semelhante (Figura 1).

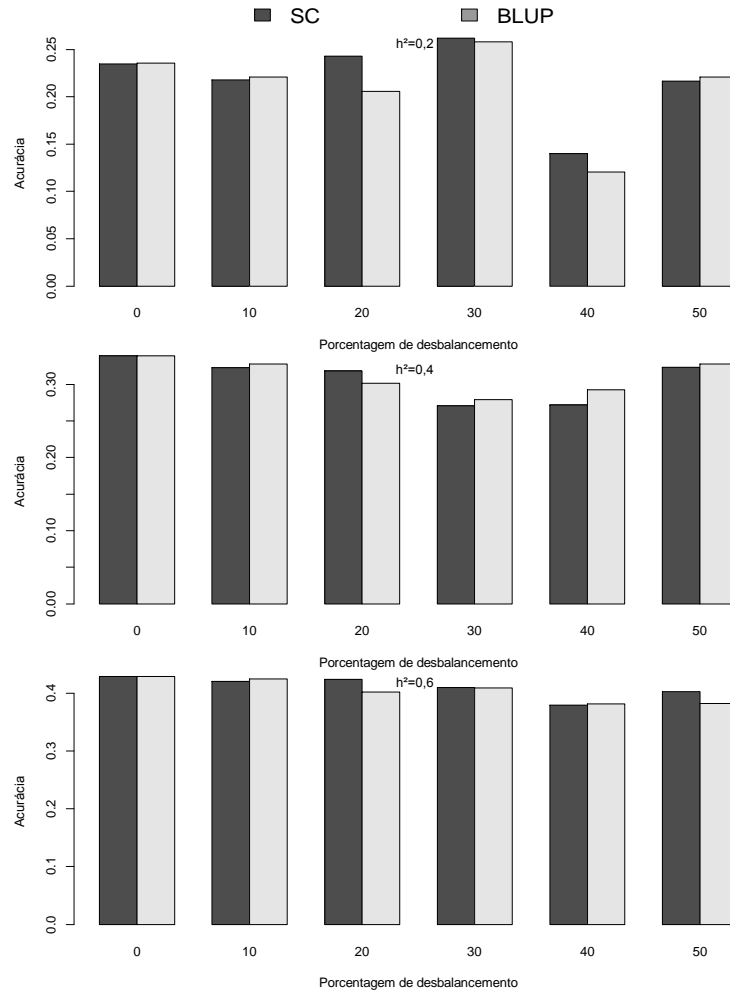


Figura 1: Acurácias de predição das análises pelo índice combinado (SC), modelos misto via REML/BLUP para as características com herdabilidades médias (h^2) de 0,20, 0,40 e 0,60 de progênes de meio irmãos combinadas com os níveis de desbalanceamento de 10 a 50% e com os dados balanceados, respectivamente.

Em relação a seleção de progênes foi analisado o quadrado da correlação do valor genético verdadeiro e das médias preditas para as progênes, para os valores fenotípicos médios (MF), índice combinado (IC) e REML/BLUP e os resultados estão apresentados na Tabela 3. Este valor expressa a herdabilidade da característica ou a confiabilidade do processo seletivo.

Tabela 3. Quadrado da correlação entre o valor genotípico verdadeiro (paramétrico) e a média predita por meio dos valores fenotípicos médios (MF), índice combinado (IC) e REML/BLUP para as características com herdabilidade média de progênes (h^2) de 0,20, 0,40 e 0,60 nos diferentes níveis de desbalanceamento

Desbalanceamento (%)	$h^2 = 0,20$			$h^2 = 0,40$			$h^2 = 0,60$		
	MF	IC	REML /BLUP	MF	IC	REML /BLUP	MF	IC	REML /BLUP
0	0,20	0,27	0,20	0,40	0,40	0,40	0,60	0,61	0,60
10	0,18	0,25	0,18	0,38	0,37	0,38	0,60	0,61	0,61
20	0,15	0,27	0,15	0,31	0,25	0,31	0,51	0,53	0,51
30	0,23	0,19	0,23	0,24	0,32	0,24	0,52	0,51	0,52
40	0,04	0,07	0,04	0,31	0,31	0,30	0,47	0,49	0,47
50	0,18	0,22	0,17	0,37	0,37	0,37	0,42	0,42	0,43

* o valor (0%), representa o nível balanceado

Observou-se que foi preservado o conceito de que a herdabilidade expressa o quadrado da correlação entre os valores das médias de progênes e o valor genético verdadeiro em ambos os métodos, pois ao nível balanceado (0%), verificou-se que as herdabilidades foram iguais as simuladas. Ainda foi observado uma tendência geral de redução dos valores de herdabilidades com o aumento do nível do desbalanceamento. Fato que era esperado, pois a perda de indivíduos afeta diretamente a estimativa desse parâmetro, tendo em vista que ele depende do tamanho amostral.

A correlação entre o valor genético e o valor predito é a própria acurácia e o quadrado dessa correlação expressa a confiabilidade da seleção, que também é apropriada para comparação de métodos de seleção. A confiabilidade dos métodos do índice combinado e REML/BLUP foram similares em relação a média predita para progênes, que também foram similares aos valores fenotípicos médios. Para um total máximo de 100 indivíduos por progênie, tanto o índice quanto o REML/BLUP indicaram que os valores fenotípicos médios foram bons preditores dos valores genotípicos. Freitas et al., (2009) obtiveram respostas semelhantes em relação a predição dos valores genotípicos através dos valores fenotípicos das médias de progênes de meio irmãos de eucalipto.

Em relação a seleção individual foi analisado o quadrado da correlação do valor genético verdadeiro e das médias preditas para os indivíduos, para os valores fenotípicos médios (MF), índice combinado (IC) e BLUP (Tabela 4).

Tabela 4. Quadrado da correlação entre o valor genotípico verdadeiro (paramétrico) e o valor predito do indivíduo por meio dos valores fenotípicos médios (MF), índice combinado (IC) e BLUP para as características com herdabilidade média (h^2) de progênes de 0,20, 0,40 e 0,60 nos diferentes níveis de desbalanceamento

Desbalanceamento (%)	$h^2= 0,20$			$h^2= 0,40$			$h^2= 0,60$		
	MF	IC	REML/BLUP	MF	IC	REML/BLUP	MF	IC	REML/BLUP
0	0,01	0,07	0,06	0,03	0,13	0,11	0,06	0,20	0,18
10	0,01	0,06	0,05	0,03	0,12	0,11	0,06	0,20	0,18
20	0,01	0,08	0,04	0,03	0,08	0,09	0,06	0,18	0,16
30	0,01	0,05	0,07	0,02	0,10	0,08	0,06	0,17	0,17
40	0,01	0,02	0,01	0,03	0,09	0,09	0,06	0,16	0,15
50	0,01	0,06	0,05	0,02	0,11	0,11	0,06	0,15	0,15

* o valor (0%), representa o nível balanceado

Em relação a avaliação do valor predito do indivíduo os métodos de seleção, que utilizam as informações das progênes, apresentaram melhores resultados que os valores fenotípicos médios, que tiveram valores da confiabilidade muito baixas e próximos de zero. Fato que já era esperado, pois a seleção fenotípica individual (ou massal) só é vantajosa, quando a característica em estudo apresentar herdabilidades altas, ademais como o fenótipo é uma função linear entre o genótipo, ambiente e da interação do genótipos x ambientes, o melhorista deve considerar todos esses fatores que podem interferir na expressão da característica (Pires et al., 2011).

O índice combinado e o REML/BLUP apresentaram estimativas muito próximas de confiabilidade. Isso foi observado para características de alta e baixa herdabilidade em todos os níveis de desbalanceamento.

No presente trabalho foi analisada a média genotípica de 20 progênes e de 100 indivíduos utilizando os valores genéticos verdadeiros de forma direta e indireta a partir dos valores dos indivíduos cuja superioridade foi estabelecida utilizando como critério de seleção os valores fenotípicos médios (MF), os escores do índice combinado (IC) e do REML/BLUP para todas as características com os seis níveis de desbalanceamento, (Tabela 5 e 6). Foi considerado o mesmo valor genético simulado (VG) para as características com as diferentes herdabilidades, visto que foram simuladas de forma independentes e sem correlação entre elas.

A média genotípica verdadeira, foi o máximo que se pôde obter da seleção, foi de pequena variação para as 20 progênes superiores identificadas nos diferentes níveis de desbalanceamento (Tabela 5). Foi observado que houve variação na seleção dos 100 melhores indivíduos, com menor média genotípica para o maior nível de desbalanceamento (50%) (Tabela 6).

Na média genotípica das 20 melhores famílias, como comentado anteriormente, todos os métodos de avaliação tiveram médias muito similares, com alguns empates entre os níveis de desbalanceamento e entre os métodos MF, IC e REML/BLUP. A menor variação entre as médias genotípicas calculadas com as médias genotípicas verdadeiras entre os métodos, foi observada para a característica de maior herdabilidade (0,60) (Tabela 5).

Com relação a análise das médias genotípicas dos 100 melhores indivíduos identificados, o IC e REML/BLUP apresentaram superioridade em relação ao valor fenotípico médio para a característica de 0,60 de herdabilidade, fato que corrobora com o que foi comentado na Tabela 4, em relação a confiabilidade dos métodos. Também foi observado, que quanto maior a herdabilidade menor é a variação da média genotípica verdadeira com a calculada para o IC e REML/BLUP.

Tabela 5. Médias genotípicas das 20 melhores progênes identificadas considerando como critério de seleção o valor genotípico paramétrico (VG), a média fenotípica (MF), o índice combinado (IC) e BLUP para as características com herdabilidade média (h^2) de progênes de 0,20, 0,40 e 0,60 nos diferentes níveis de desbalanceamento

Desbalanceamento (%)	$h^2 = 0,20$				$h^2 = 0,40$				$h^2 = 0,60$			
	VG	MF	IC	REML/BLUP	VG	MF	IC	REML/BLUP	VG	MF	IC	REML/BLUP
0	128,44	122,09	123,03	122,09	128,44	126,07	125,70	126,07	128,44	126,11	125,81	126,11
10	128,50	121,89	123,05	121,15	128,50	124,78	124,56	124,78	128,50	126,57	125,58	126,24
20	128,39	119,05	119,68	119,49	128,39	123,49	124,65	124,09	128,39	125,33	125,05	125,34
30	128,77	123,33	123,80	122,70	128,77	122,84	122,62	122,84	128,77	125,89	125,90	125,89
40	128,33	119,99	120,62	119,99	128,33	124,28	124,51	123,59	128,33	125,77	125,55	124,15
50	128,34	119,93	120,26	119,93	128,34	124,59	124,29	124,59	128,34	124,19	123,80	124,89

* a coluna VG é igual para todas as características.

Tabela 6. Médias genotípicas dos 100 melhores indivíduos identificados considerando como critério de seleção o valor genotípico paramétrico (VG), a média fenotípica (MF), o índice combinado (IC) e BLUP para as características com herdabilidade média (h^2) de progênes de 0,20, 0,40 e 0,60 nos diferentes níveis de desbalanceamento

Desbalanceamento (%)	$h^2 = 0,20$				$h^2 = 0,40$				$h^2 = 0,60$			
	VG	MF	IC	REML/BLUP	VG	MF	IC	REML/BLUP	VG	MF	IC	REML/BLUP
0	159,22	121,63	126,88	126,79	159,22	129,54	133,63	133,42	159,22	129,12	137,94	137,27
10	158,65	121,31	124,47	122,36	158,65	129,47	133,89	134,81	158,65	129,97	138,07	136,59
20	158,21	121,43	119,73	120,53	158,21	128,84	128,98	130,11	158,21	130,73	138,75	137,90
30	158,18	122,18	133,06	132,72	158,18	127,42	130,83	129,88	158,18	130,66	137,31	137,56
40	157,03	119,53	123,65	122,26	157,03	128,42	127,51	128,32	157,03	129,33	134,33	133,95
50	156,25	121,74	120,73	117,51	156,25	127,48	130,20	129,67	156,25	130,01	134,04	133,23

* a coluna VG é igual para todas as características.

A seleção das 20 melhores progênes foi realizada por ambos os métodos e comparadas com a seleção realizadas pelo valor genotípico médio (VGM) (Tabela 7). Dessa seleção foram avaliadas a percentagem de coincidência, para saber quantas progênes selecionadas foram as mesmas entre os métodos, e a correlação de Spearman (Tabela 8), que identifica a classificação dos selecionados entre os métodos. Essa mesma análise foi realizada para os indivíduos, sendo considerados apenas os 100 melhores.

As 20 melhores progênes simuladas classificados pelo valor genético paramétrico médio estão apresentadas na Tabela 7.

Tabela 7: Classificação das 20 melhores progênes simuladas, selecionadas pelo valor genético paramétrico médio (VGM)

Ordem de classificação	Progênes	VGM
1	16	133,08
2	60	132,65
3	23	131,75
4	64	130,64
5	17	130,50
6	62	130,36
7	41	130,25
8	82	130,06
9	13	129,45
10	84	128,23
11	22	128,03
12	49	128,02
13	34	127,81
14	36	127,17
15	38	126,44
16	83	125,17
17	24	125,10
18	43	124,95
19	71	124,64
20	75	124,40

Na análise da coincidência da seleção de progênes do valor genético verdadeiro com o índice combinado foi observado variação de 25 a 50% de coincidência entre as progênes selecionadas pelo índice para as características relativas a herdabilidade 0,20. Para a herdabilidade 0,40 a variação do percentual de coincidência foi de 50 a 70% e para herdabilidade 0,60 a variação foi de 55 a 75%.

Com o aumento da herdabilidade observou-se aumento do percentual de coincidência, o que também demonstra a importancia de um bom planejamento experimental que explore a herdabilidade máxima da população em relação a uma determinada característica de interesse.

Esses resultados foram similares no percentual de coincidência do REML/BLUP com o valor genético real.

Na comparação entre as metodologias de seleção do índice combinado com REML/BLUP, para todas as diferentes herdabilidades o percentual de coincidência foi elevado, atingindo o máximo de 100%, mesmo nos casos de desbalanceamento (Tabela 8). Dessa forma, a seleção das melhores progênies realizadas tanto pelo índice como pela metodologia REML/BLUP foram similares, mesmo nos casos em que ocorreram desbalanceamento nos dados, para as diferentes herdabilidades.

Verificou-se correlações de Spearman de magnitudes muito baixas, embora significativas, na sua maioria, para a herdabilidade 0,20 nos diferentes níveis de desbalanceamento, comparando-se o valor genético verdadeiro com o índice combinado e com o REML/BLUP. Os baixos valores estimados indicaram que os métodos classificam de forma diferente as progênies selecionadas da classificação pelo valor genético verdadeiro.

Para as características de herdabilidade 0,20 com níveis de 10, 40 e 50% de desbalanceamento não foram obtidas concordância na classificação das melhores progênies, pois a correlação de Spearman foi zero ou próximo de zero para o índice combinado. Respostas similares foram obtidas para a metodologia REML/BLUP nos casos de 30 e 50% de desbalanceamento, além disso, também as correlações foram de baixas magnitudes.

As correlações de Spearman tiveram a mesma tendência de apresentarem maiores magnitudes com o aumento da herdabilidade. As magnitudes das correlações de Spearman foram sempre superiores a 0,7 entre as metodologias do índice combinado e REML/BLUP. Isso demonstra a correspondência na classificação das progênies superiores selecionadas entre as metodologias.

A percentagem de coincidência e a correlação de Spearman na seleção de progênies para os dados balanceados foram iguais entre as metodologias, o que era previsto, visto que, em experimentos balanceados, os valores genéticos preditos com base na abordagem fundamentada em REML/BLUP e índice ou seleção combinada são equivalentes (Resende, 2002).

Em relação a seleção de indivíduos foi considerado a seleção dos 100 melhores em ambas metodologias, e calculados os percentuais de coincidência e a correlação de Spearman entre o valor genético real e o índice combinado e entre o REML/BLUP e entre o índice e o REML/BLUP (Tabela 9).

Os percentuais de coincidências para a seleção de indivíduos foram muito baixas em todas as combinações de herdabilidade com os desbalanceamentos entre as comparações dos valores genéticos com as metodologias do índice combinado e a REML/BLUP. Aliado a esses resultados, as correlações de Spearman também foram de baixas magnitudes, porém significativas em sua maioria, com a presença de correlações negativas para as herdabilidades de 0,20 e 0,40 e nos diferentes níveis de desbalanceamentos. Apesar desses resultados, a característica de maior herdabilidade, como no caso comentado anteriormente, apresentou melhores percentuais de coincidências e correlação de Spearman.

Em relação a percentagem de coincidência e a correlação de Spearman entre as metodologias do índice e REML/BLUP mantiverem sempre magnitudes altas, o que também para a seleção de indivíduos, como no caso da seleção de progênies, ambos os métodos selecionaram semelhantemente as progênies e os indivíduos.

Alguns autores relatam sobre comparações entre metodologias de seleção combinada e modelos mistos, como Rosado et al., (2009) que avaliaram progênies de meio irmãos de *Eucalyptus urophylla* e concluíram que a predição de ganhos por ambos os métodos foram similares e superiores a outras metodologias de seleção. Rocha et al., (2006) também avaliaram progênies de meio irmãos *E. urophylla* e nos dois procedimentos da máxima verossimilhança restrita e esperança de quadrados mínimos, os valores dos parâmetros genéticos foram semelhantes, exceto entre progênies, em que o procedimento REML proporcionou valores elevados com acurácia superior a 70% em todas as populações.

Tabela 8: Percentual de coincidência (C) e correlação de Spearman para as 20 progênes selecionadas pelas comparações do valor genotípico paramétrico (VG) com o índice combinado (IC), VG com REML/BLUP, e IC com REML/BLUP para as características com herdabilidade média (h^2) de progênes de 0,20, 0,40 e 0,60 nos diferentes níveis de desbalanceamento

desb (%)	$h^2=0,2$						$h^2=0,4$						$h^2=0,6$					
	VG x IC		VG x		IC x		VG x IC		VG x		IC x		VG x IC		VG x		IC x	
	C	r	C	r	C	r	C	r	C	r	C	r	C	r	C	r	C	r
0	40	0,06**	40	0,06**	100	1	70	0,65**	70	0,65**	100	1	75	0,68**	75	0,68**	100	1
10	45	0,07**	40	0,25**	95	0,98**	55	0,76**	55	0,76**	100	0,99**	75	0,79**	75	0,66**	90	0,73**
20	30	0,17**	30	0,07**	90	0,99**	50	0,34**	55	0,31**	95	0,98**	65	0,57**	65	0,59**	95	0,99**
30	50	0,38**	45	0,49**	95	0,98**	50	0,63**	50	0,63**	100	0,99**	60	0,67**	60	0,69**	100	0,97**
40	35	0,02ns	30	0,15**	95	0,91**	60	0,28**	55	0,41**	95	0,97**	60	0,27**	50	0,62**	90	0,93**
50	25	0	20	0,09**	95	0,94**	55	0,57**	55	0,53**	100	0,99**	55	0,57**	55	0,57**	90	0,97**

** , * : Significativo a 1 e 5% de probabilidade pelo teste t; o valor (0%), representa o nível balanceado

Tabela 9: Percentual de coincidência (C) e correlação de Spearman para os 100 melhores indivíduos selecionados pelas comparações do valor genotípico paramétrico (VG) com o índice combinado (IC), VG com REML/BLUP, e IC com REML/BLUP para as características com herdabilidade média (h^2) de progênes de 0,20, 0,40 e 0,60 nos diferentes níveis de desbalanceamento

desb (%)	$h^2=0,2$						$h^2=0,4$						$h^2=0,6$					
	VG x IC		VG x		IC x		VG x IC		VG x		IC x		VG x IC		VG x		IC x	
	C	r	C	r	C	r	C	r	C	r	C	r	C	r	C	r	C	r
0	2	0,19**	1	0,19**	95	0,94**	8	0	8	0,07**	96	0,98**	11	0,23**	11	0,22**	98	0,98**
10	3	0,09**	1	-0,06**	81	0,88**	8	0	9	-0,09**	95	0,98**	12	0,42**	10	0,26**	70	0,74**
20	1	-0,27**	1	-0,21**	91	0,91**	4	0,2**	5	0,05**	87	0,86**	12	0,2**	12	0,26**	96	0,96**
30	9	0,01 ^{ns}	8	0,28**	79	0,47**	7	0,07**	6	0,13**	83	0,94**	16	0,15**	16	0,12**	93	0,93**
40	3	-0,08**	5	-0,09**	89	0,91**	3	0	5	0	95	0,94**	10	0,08**	10	0,15**	92	0,95**
50	4	0,19**	3	-0,03*	83	0,62**	9	0,11**	8	0,12**	88	0,85**	15	0,16**	14	0,18**	90	0,90**

** , * : Significativo a 1 e 5% de probabilidade pelo teste t, o valor (0%), representa o nível balanceado

4. CONCLUSÃO

Os métodos de seleção combinada e modelos mistos via REML/BLUP apresentaram respostas muito semelhantes em todas as avaliações realizadas. A utilização do método de índice combinado mostra-se adequado para a seleção de progênies de meio irmãos, com confiabilidades similares ao procedimento REML/BLUP.

Com o aumento da herdabilidade foi verificado que as metodologias apresentaram maiores percentuais de coincidência e menor mudança na classificação das melhores progênies selecionadas, quando comparadas com os valores genéticos verdadeiros.

Os diferentes níveis de desbalanceamento tiveram maior influência na seleção de indivíduos do que na seleção de progênies.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro,

5. Referências

ARNHOLD, E., VIANA, J.M.S., SILVA R.G., MORA F. 2009. Eficiências relativas de métodos de seleção de famílias endogâmicas em milho-pipoca, **Acta Scientiarum, Agronomy**, Maringá, v. 31, n. 2, p. 203-207. DOI: 10.4025/actasciagron,v31i2,578

CARNEIRO JÚNIOR, J.M., EUCLYDES, R.F., LOPES, O.S., TORRES, R.A. 2004. Avaliação de Métodos de Estimação de Componentes de Variância Utilizando Dados Simulados, **Revista Brasileira de Zootecnia**, v,33, n,2, p,328-336.

COSTA, R.B., RESENDE, M.D.V., ARAÚJO, A.J., GONÇALVES, O.S., BORTOLETTO, N. 2000. Seleção combinada univariada e multivariada aplicada ao melhoramento genético da seringueira, **Pesquisa Agropecuária Brasileira**. Brasília, v.35, n.2, p. 381-388.

CRUZ, C.D., CARNEIRO, P.C.S. 2006. **Modelos biométricos aplicados ao melhoramento genético** v.2, Editora UFV, Viçosa, 585p.

CRUZ, C.D. 2012. GENES, a software package for analysis in experimental statistics and quantitative genetics, **Acta Scientiarum, Agronomy**, Maringá, v.35, p. 271-276. Doi 10.4025/actasciagron. V35i3,21251.

FREITAS, R.G., VASCONCELOS, E.S., CRUZ, C.D., ROSADO, A.M., ROCHA, R.B., TAKAMI, L.K. 2009. Predição de ganhos genéticos em progênies de polinização

aberta de *Eucalyptus urograndis* cultivadas em diferentes ambientes e submetidas a diferentes procedimentos de seleção. **Revista Árvore**, Viçosa, v.33, n.2, p.255-263.

FONSECA, S.M. da, RESENDE M.D.V., ALFENAS, A.C., GUIMARÃES, L.M.S., ASSIS, T.F., GRATTAPAGLIA, D. 2010. **Manual prático de melhoramento genético de eucalipto**, Viçosa: Ed, UFV, 200p.

HENDERSON, C. R. 1963. Selection index and expected genetic advance. NAS-NRC:982.

HENDERSON, C.R., QUAAS, R.L. 1976. Multiple trait evaluation using relative records, **Journal of Dairy Science**, v.43, 1188 p.

LOPES, P.S. 2005. **Teoria do melhoramento animal**, Belo Horizonte: Editora FEPMVZ - 118p.

MARTINS, I.S., CRUZ, C.D., ROCHA, M.G.B., REGAZZI, A.J., PIRES, I.E. 2005. Comparação entre os processos de seleção entre e dentro e o de seleção combinada, em progênies de *Eucalyptus grandis*, **Cerne**, Lavras, v, 11, n, 1, p, 16-24.

PIRES, I.E., RESENDE, M.D.V., SILVA, R.L., RESENDE JÚNIOR, M.F. 2011. **Genética florestal**. Viçosa: Arka, 318p.

R DEVELOPMENT CORE TEAM (2013) R: *A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL [http://www,r-project.org](http://www.r-project.org)

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

RESENDE, M.D.V. 2002. **Software SELEGEN-REML/BLUP**, Colombo: Embrapa Florestas, 67 p, (Embrapa Florestas, Documentos 77), 2002.

RESENDE, M.D.V. 2007. **Matemática e estatística na análise de experimentos e no melhoramento genético**, Colombo: Embrapa Florestas, 362p, 2007.

ROCHA, R.B., ROCHA, M.G.B., SANTANA, R.C., VIEIRA, A. H. 2009. Estimação de parâmetros e seleção de procedências e famílias de *Dipteryx alata* VOGEL (baru) utilizando metodologia de REML/BLUP e E(QM), **Cerne**, Lavras, v, 15, n, 3, p, 331-338.

ROCHA MGB; PIRES IE; XAVIER A; CRUZ CD; ROCHA, RB, Avaliação genética de progênies de meio-irmãos de *Eucalyptus urophylla* utilizando os procedimentos REML/BLUP e E(QM), **Ciência Florestal**, Santa Maria, v, 16, n, 4, p, 369-379, 2006,

ROSADO, A.M., ROSADO, T.B., RESENDE JÚNIOR, M.F., BHERING, L.L., CRUZ CD. 2009. Ganhos genéticos preditos por diferentes métodos de seleção em progênies de *Eucalyptus urophylla*, **Pesquisa agropecuária brasileira**, Brasília, v.44, n.12, p,1653-1659.

CONCLUSÃO GERAL

A utilização do pacote *pedigreemm* pode ser aplicado ao melhoramento vegetal observando-se as devidas correções para cada situação experimental considerada. O script elaborado apresenta os parâmetros genéticos de importância para a seleção genética de indivíduos, no caso de famílias de meio irmãos, que também pode ser aplicado as famílias de irmãos germanos.

A utilização da metodologia de seleção combinada apresentou eficiência semelhante a metodologia de modelos mistos para a estimação da confiabilidade de predição, mesmo em níveis extremos de desbalanceamento. Com o aumento da herdabilidade foi verificado que as metodologias apresentaram maiores percentuais de coincidência e menor mudança na classificação das melhores progênes selecionadas, quando comparadas com os valores genéticos verdadeiros.