

Acurácia preditiva de testes clonais de *Eucalyptus* spp.
utilizando efeitos aditivos do parentesco e validação cruzadaPredictive accuracy of *Eucalyptus* spp. clonal trials
using additive kinship effects and cross validationRafael Tassinari Resende¹, Marcos Deon Vilela de Resende²,
Fabyano Fonseca e Silva³ e Elizabete Keiko Takahashi⁴

Resumo

O uso de extensos plantios clonais é uma prática consensual nas empresas florestais. No entanto, os modelos estatísticos e o tamanho das parcelas utilizadas na seleção dos clones superiores ainda é uma questão entre os melhoristas florestais. Com base em 11 experimentos contendo parcelas semelhantes aos plantios comerciais e 63 clones, este trabalho objetivou avaliar o uso de informação de pedigree, heterogeneidade genética entre ambientes e a possibilidade de redução de parcelas, a fim de otimizar a predição do incremento médio anual (IMA) realizando validação cruzada entre ambientes. Os ambientes proporcionaram altas herdabilidades no sentido-restrito (entre 0,65 a 0,95) e forte alteração dos rankings entre experimentos. A inclusão de parentesco e variâncias genéticas particulares para cada ambiente ao modelo é um procedimento eficiente na melhoria das acurácias realizadas. Além disso, 4 a 16 plantas avaliadas por parcela é uma quantidade confiável para prever a produtividade dos talhões florestais.

Palavras-chave: Modelos mistos; Pedigree; Parcelas Quadradas; Estruturas de Covariância; Validação Cruzada; Redução de Parcela

Abstract

Large clonal planting is a practice already established by the timber companies. However, statistical models, the plot size used in selection of superior clones is still a topic of discussion between forest breeders. Based on 11 trials containing plots similar to commercial plantations and 63 clones, this study aimed to evaluate the use of pedigree information, genetic heterogeneity between environments and the possibility of plot reduction in order to optimize the predictive accuracies of the mean annual increment (MAI) performing cross-validation between environments. All trials showed high narrow-sense heritabilities (from 0.65 to 0.95) and a strong change of clonal rankings was verified. It was observed that the inclusion of pedigree information and particular genetic variances for each trial on model is an effective procedure to increase the predictive accuracies. Moreover, between 4-16 trees evaluated per plot is a reliable size for the prediction of forest stands.

Keywords: MET mixed-models; Pedigree; Big plots; Covariance Structures; Cross-validation; Plot Reduction.

INTRODUÇÃO

A inclusão de informação detalhada de parentesco é uma tendência no melhoramento genético moderno. Nesse contexto, dados de pedigree conduzem a informações precisas dos cruzamentos, indicando o relacionamento genético dos indivíduos independentemente de delineamento genético (EL-KASSABY et al., 2011). Em florestas artificiais de *Eucalyptus* spp. com fins de produção é vital a utilização de clones, por questões de homogeneidade de plantio e elevada produtividade. No melhoramento florestal, a seleção clonal é feita assumindo-se não-parentesco dos clones, valendo-se de delineamentos genéticos apenas em etapas precedentes aos experimentos clonais (RESENDE; BARBOSA, 200).

¹Pesquisador Doutor em Genética e Melhoramento. UFV - Universidade Federal de Viçosa. Av. Peter Henry Rolfs, s/n - Campus Universitário - 36570-900 - Viçosa, MG, Brasil. E-mail: rafael.tassinari@gmail.com

²Pesquisador Doutor. EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa de Florestas. Estrada da Ribeira, Km 111 - 83411000 - Colombo, PR, Brasil. marcos.deon@gmail.com

³Professor Associado do Departamento de Zootecnia. UFV - Universidade Federal de Viçosa. Av. Peter Henry Rolfs, s/n - Campus Universitário - 36570-900 - Viçosa, MG, Brasil. E-mail: fabyanobr@gmail.com

⁴Pesquisadora Doutora. Celulose Nipo Brasileira S.A. Rodovia de Ligação BR 381 a Belo Oriente Km 3 - Caixa-postal: 100 - 35195-000 - Belo Oriente, MG, Caixa-postal: 100. E-mail: elizabete.takahashi@cenibra.com.br

Os clones em avaliação no programa de melhoramento muitas vezes não estão dispostos em todos os experimentos e/ou ambientes da empresa, isso porque ocupam uma considerável área de plantio demandando altos custos de implantação e medições. *Multi-environment trials* (METs) são comumente usados em programas de melhoramento florestal (OGUT et al., 2014) na seleção de genótipos superiores e na avaliação das interações genótipo por ambiente (GA), sendo desejável o conhecimento do potencial genético de um clone em todos os locais disponíveis (PIEPHO; MÖHRING 2007). Os dados de MET são geralmente desbalanceados, onde somente parte dos genótipos estão presentes em todos os testes (OGUT et al., 2014), atendendo assim a realidade dos plantios clonais florestais. Nesse sentido, o uso de matriz de parentesco em modelos estatísticos possibilita predições de genótipos não-observados baseando-se no potencial parental de outros indivíduos (HENDERSON, 1984).

Outra questão discutível em testes clonais é o tamanho das parcelas, posto que para atingir confiabilidade nas predições de produtividade dos talhões, é de praxe implantar experimentos com muitas plantas dentro de parcelas e também muitas repetições (ZHANG et al., 2015). Esses procedimentos adotados visam criar experimentos mais fidedignos aos plantios comerciais e contornar efeitos de competição entre as árvores (SILVA; KERR, 2013). Considerando modelos animais, que realizam predições acuradas a partir de indivíduos não reprodutíveis (MRODE, 2014), é possível supor eficiência da matriz de parentesco na redução de parcelas experimentais.

O presente trabalho teve dois objetivos principais: i) verificar a eficiência da inclusão da matriz "A" de numeradores dos coeficientes genético-aditivos de parentesco (HAYES et al., 2009) nas acurácias preditivas entre ambientes e na eficiência da predição de indivíduos não instalados nos ambientes utilizados no modelo preditivo; e ii) verificar a possibilidade de redução das parcelas com auxílio da matriz A. Para alcançar tais objetivos, foram avaliados modelos que contabilizam variâncias genéticas homogêneas e heterogêneas entre ambientes.

MATERIAL E MÉTODOS

Descrição dos experimentos

Onze testes clonais ampliados (TC) contendo 36, 70 e 100 plantas por parcela (Tabela 1) foram selecionados no conjunto de dados da empresa CENIBRA S.A. Todos os experimentos foram delineados em blocos casualizados com três repetições. Entre os 11 testes, seis deles foram implementados com 10m²/planta e os demais com 6m²/planta. Além disso, seis testes foram implantados em região de alta altitude e cinco em baixa altitude (variando entre 300 a 870 m acima do nível do mar). Sessenta e três clones de híbridos de eucalipto foram distribuídos entre os testes, sendo que deste total, 34 estavam presentes em mais de um teste. Os experimentos foram medidos em idades próximas aos 6 anos, e suas medições foram realizadas entre setembro de 2008 e março de 2010.

Tabela 1. Descrição dos experimentos utilizados no ajuste dos modelos estatísticos.

Table 1. Description of trials used on statistical models.

Teste	Clones	Altitude	NAP	EP	IMA
TC01	28	Baixa	100	10	38,90
TC02	12	Alta	36	10	44,41
TC03	12	Alta	100	10	52,33
TC04	20	Baixa	100	6	32,73
TC05	10	Baixa	100	6	44,56
TC06	10	Alta	100	6	54,65
TC07	8	Alta	100	6	47,12
TC08	10	Alta	36	10	36,48
TC09	8	Baixa	100	10	39,25
TC10	6	Alta	100	6	61,96
TC11	7	Baixa	70	10	34,97

TC é a abreviação de teste clonal ampliado; Clones representa o número de clones por experimento; Altitude Alta e Baixa variam aproximadamente entre 300 a 870 m acima do nível do mar; NAP é o número de árvores por parcela; EP é o espaçamento de plantio (em m²/planta); IMA é o incremento médio anual (medido em m³/ha/ano). TC is an abbreviation of big clonal test; Clones is the number of clones in experiment; Altitude is the type of region elevation (varying between 300-870m above sea level); NAP is number of trees within plot; EP is the planting spacing (m²/plant); IMA is the mean annual increment (MAI in m³/ha/year).

A variável base utilizada no trabalho foi o incremento médio anual (IMA) da parcela, que é o volume de madeira extrapolado para um hectare dividido pelo tempo de crescimento. Os valores de volume (*vol*) em m³ foram calculados pela equação simplificada para árvores individuais [1] (SCHUMACHER; HALL, 1933), em que, *DAP* é o diâmetro na altura do peito (cm), *H* é a altura da árvore (m), *f* é o fator de forma (adotado 0,45), e ϵ é a proporção entre o perímetro e o diâmetro de uma circunferência (adotado 3,14159).

$$vol = DAP^2 \frac{\pi}{40000} Hf \quad [1]$$

Análises estatísticas

Quatro modelos foram comparados nesse estudo, tais modelos formam a combinação entre: consideração (*Pedigree*) e não consideração (*Unrelated*) de parentesco entre os clones *versus* variância genética homogênea (ID) e heterogênea (DIAG) entre ambientes (Tabela 1). Todos os modelos foram ajustados utilizando o pacote *asreml* do R (BUTLER et al. 2009). Em termos de análise multi-experimentos (MET) e considerando *g* genótipos em 11 TC's, o modelo linear misto utilizado pode ser representado pela forma reduzida matricial abaixo:

$$y = Xb + Zg + e, \quad [2]$$

Em que *y* é o vetor de observações em todos experimentos; *b* é o vetor de efeitos fixos (intercepto e os experimentos); *g* é o vetor aleatório de genótipos dentro de experimentos; e *e* representa o vetor de resíduos dentro de experimentos. *X* e *Z* são as matrizes de incidência relacionando as observações com os efeitos fixos e aleatórios, respectivamente. Vetores contendo os efeitos aleatórios assumem distribuição normal multivariada, centradas em zero com as seguintes estruturas de covariância: $Cov(g, g') = G_0 \otimes K$, $Cov(\epsilon, \epsilon') = I_g \otimes R_0$ e $Cov(g, \epsilon') = 0$. Sendo I_g uma matriz identidade de tamanho *g*; G_0 uma matriz de covariância do efeito genético dos clones em cada experimento; e \otimes denota o produto de *Kronecker* entre matrizes. *K* é assumido como I_g (matriz identidade de ordem $g \times g$) para os modelos 1 e 3 e *A* (matriz formada pelo denominador dos coeficientes aditivos de parentesco de ordem $g \times g$, obtidos por meio da informação de *pedigree* (HAYES et al., 2009) para os modelos 1 e 2. A $Cov(g, g') = G_0 \otimes K$ é representada pela Eq. [3] para os modelos 1 e 3, e pela Eq. [4] para os modelos 3 e 4, sendo σ_g^2 a variância genética total para os modelos 1 e 3, e a variância genética aditiva para os modelos 2 e 4; a notação $\sigma_{g_j}^2$ é utilizada para distinguir variâncias genéticas (ou genéticas aditivas) entre experimentos nos modelos do tipo DIAG (modelos 2 e 4), sendo $j = \{1, 2, \dots, 11\}$.

$$G_0 \otimes K = I_g \sigma_g^2 \otimes K \quad [3]$$

$$G_0 \otimes K = \begin{bmatrix} \sigma_{g_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{g_2}^2 & & \\ \vdots & & \ddots & \\ 0 & & & \sigma_{g_{11}}^2 \end{bmatrix} \otimes K \quad [4]$$

Os resíduos são independentes e normalmente distribuídos (IND) dentro dos experimentos com matriz bloco diagonal entre os diferentes ambientes fazendo-se $R = \sigma_{\epsilon_1}^2 I_{n_1} \otimes \sigma_{\epsilon_2}^2 I_{n_2} \otimes \dots \otimes \sigma_{\epsilon_{11}}^2 I_{n_{11}}$ (sendo $\sigma_{\epsilon_1}^2 \neq \sigma_{\epsilon_2}^2 \neq \dots \neq \sigma_{\epsilon_{11}}^2$). Os ajustes dos modelos foram comparados com base no *Akaike Information Criteria* (AIC) (AKAIKE, 1998). As herdabilidades no sentido restrito, em cada experimento, foram calculadas com base na Eq. [5] a partir dos componentes de variância obtidos via procedimento REML- *Average Information* (GILMOUR et al., 1974), em que $\sigma_{g_j}^2$ e $\sigma_{\epsilon_j}^2$ são as variâncias genéticas aditivas e as variâncias residuais no *j*-ésimo experimento, respectivamente.

$$h_j^2 = \frac{\sigma_{g_j}^2}{\sigma_{g_j}^2 + \sigma_{\epsilon_j}^2} \quad [5]$$

Para avaliar as acurácias preditivas (realizadas) dos quatro modelos, foi adotada a técnica de validação cruzada (KOHAVI, 199) tomando 10 testes como população de treinamento e o teste

remanescente como população de validação. As acurácias preditivas foram calculadas a partir da correlação de *Pearson* entre os *Estimated Breeding Values* (EBV) dos modelos MET e os EBV's dos clones da população de validação. Com base na matriz de parentesco, foram calculadas também a habilidade dos modelos 2 e 4 em prever indivíduos não-observados na população de treinamento.

Na verificação da possibilidade de redução das parcelas foi realizado o procedimento *bootstrap* (HESTERBERG et al., 2000) para amostrar parcelas menores nos testes. Os tamanhos de parcelas (tratamentos) utilizados foram: 1 (parcela de árvore única ou *single tree plot* – STP); 4 (2×2); 9 (3×3); 16 (4×4); 25 (5×5); e 36 (6×6) árvores dentro da parcela. Foram realizadas um máximo de 100 iterações por tratamento (i.e., quanto maior a parcela menor o número de possibilidades de combinações entre as árvores). Para validar os resultados da redução da parcela, foram calculadas as acurácias preditivas por meio de validação cruzada descrita no parágrafo anterior.

RESULTADOS E DISCUSSÃO

Acurácia preditiva entre ambientes contrastantes.

Os onze experimentos apresentaram diferenças marcantes com relação à produtividade, variando entre 32,7 a 62,0 m³/ha/ano. Na empresa, tipicamente experimentos implantados em altitude elevada apresentam maior produtividade em IMA (JESUS et al., 2015). No presente trabalho, em ambientes de maior altitude a produtividade foi em média 51,0 m³/ha/ano contra 37,7 m³/ha/ano dos experimentos em altitude baixa (Tabela 1). Além disso, os experimentos mais adensados (6 m²/planta) produziram ligeiramente mais, em média 44,1 contra 41,6 m³/ha/ano dos experimentos com 10 m²/planta.

Tabela 2. Comparação do ajuste dos 4 modelos utilizados.
Table 2. Adjusted comparison between the four models used.

Modelo	Estrutura de Covariâncias	nPar	AIC	loglik
1	Unrelated, ID	12	1627,54	-801,77
2	Pedigree, ID	12	1609,84	-792,92
3	Unrelated, DIAG	22	1626,67	-791,34
4	Pedigree, DIAG	22	1605,78	-780,89

Unrelated: modelos com clones assumidos como não-aparentados; Pedigree: modelos com inclusão de informação de parentesco entre clones; ID: variância genética homogênea entre experimentos; DIAG: variância genética heterogênea entre experimentos; nPar: número de componentes de variância estimados por modelo; AIC critério de comparação de modelos Akaike Information Criteria; loglik é o log na base 10 do verossimilhança do procedimento REML. Unrelated: models with clones defined as unrelated; Pedigree: models with the inclusion of pedigree information of clones; ID: homogeneous genetic variance across environments; DIAG: heterogeneous genetic variance across environments; nPar is the number of variance components estimated in the model; AIC is the Akaike Information Criteria; loglik is the likelihood log of REML procedure.

Os parâmetros de ajuste dos quatro modelos MET utilizados são mostrados na Tabela 2. Os modelos 2 e 4 apresentaram melhor ajuste (AIC's menores), sendo que o de melhor ajuste foi o modelo 4 (informação de parentesco e variâncias genéticas heterogêneas entre ambientes).

Bons resultados utilizando a matriz de parentesco no cálculo de parâmetros genéticos e ganhos de seleção em crescimento e forma da árvore em *Eucalyptus regnans* foram obtidos por Suontama et al. (2015). Estes autores encontraram herdabilidades (h^2) em torno de 0,30 tanto para DAP como para altura (H). Bush et al. (2015) utilizaram informação de parentesco baseada em marcadores moleculares para cálculo de parâmetros genéticos em progênies de meios-irmãos em *E. cladocalyx* e encontraram h^2 's iguais a 0,18 e 0,36 para DAP e H, respectivamente (utilizando parentesco via marcadores), e valores iguais a 0,36 e 0,53 utilizando informação de famílias. Valores de h^2 's mais baixos como aos desses dois trabalhos são tipicamente encontrados em populações iniciais ou intermediárias de melhoramento (RESENDE, 2002), contrastando por este motivo, com os valores de h^2 's dos testes clonais apresentados neste trabalho (resultados descritos a seguir).

A partir do modelo 4, foram calculadas as variâncias genéticas aditivas e residuais, em cada ambiente, bem como as respectivas herdabilidades (h^2) no sentido restrito (Tabela 3). O ambiente que proporcionou maior h^2 foi o TC09 e o de menor o TC02. No entanto, de acordo com Resende (2002), todas as h^2 's no sentido restrito encontradas neste trabalho podem ser consideradas como de alta magnitude.

Tabela 3. Componentes de variância (genética aditiva e residual) e herdabilidades no sentido-restrito.
Table 3. Variance components (additive genetic and residual) and narrow-sense heritabilities.

Teste	σ_g^2	σ_e^2	h^2
TC01	55,024	8,849	0,861 ± 0,028
TC02	17,317	9,408	0,648 ± 0,090
TC03	61,390	8,454	0,879 ± 0,034
TC04	40,224	6,480	0,861 ± 0,032
TC05	34,091	7,969	0,811 ± 0,056
TC06	48,568	18,653	0,723 ± 0,082
TC07	114,783	35,263	0,765 ± 0,075
TC08	40,572	17,634	0,697 ± 0,083
TC09	188,199	9,125	0,954 ± 0,016
TC10	258,416	24,730	0,913 ± 0,033
TC11	24,474	2,362	0,912 ± 0,031

Exp: experimento avaliado; σ_g^2 : variância genética aditiva; σ_e^2 : variância residual; h^2 : herdabilidade no sentido restrito.

Exp: evaluated experiment; σ_g^2 : genetic additive variance; σ_e^2 : residual variance; h^2 : narrow sense heritability.

A interação genótipo por ambiente (GA) é observada na Figura 1 por meio dos valores genéticos aditivos (EBV's) calculados pelo modelo 4. Observa-se uma interação complexa com forte alteração da classificação (rankings) dos clones entre ambientes. Os ambientes são mostrados de forma ordenada do de menor para o de maior média de EBV. Outros autores também avaliaram a interação genótipo por ambiente, além de calcular correlações genéticas entre ambientes para variáveis de crescimento de *Eucalyptus* spp. Suontama et al. (2015) encontraram correlações genéticas entre ambientes iguais a 0,72 e 0,83 para altura e diâmetro (variáveis que compõe o IMA) em famílias de meios-irmãos de *E. regnans* na Nova Zelândia. Lambeth, et al. (1994) avaliaram IMA em clones de *E. grandis* e encontraram correlação genética entre ambientes igual a 0,96.

Além de avaliações entre ambientes, correlações genéticas para fins de seleção precoce (em idade de 3 para 6 anos) foram estimadas para IMA em clones de *E. grandis* (OSORIO et al., 2003) e diâmetro e altura em clones de *E. urophylla* (PINTO et al., 2014), esses autores encontraram valores entre 0,84 e 0,97. As acurácias realizadas calculadas por meio de validação cruzada também demonstraram alto índice de acerto entre ambientes para os quatro modelos utilizados (Tabela 4), resultado compatível com as altas h^2 's do caráter IMA (LAMBETH et al. 1994). Neste aspecto, os ambientes com menores estimativas de h^2 's (TC02 e TC08) também foram os que proporcionaram menores acurácias. Embora as acurácias em todos ambientes tenham sido elevadas, em especial, o modelo 4 apresentou superioridade numérica em comparação com demais modelos, sendo indicado na seleção clonal em etapas finais do programa de melhoramento do eucalipto. O componente de variância genética de plantas de reprodução vegetativa se divide em variância genética aditiva, variância devido a desvios de dominância e variância epistática (RESENDE 2007), no entanto, este trabalho se dispôs a estudar os ganhos em acurácia realizada (via validação cruzada) focando apenas no componente aditivo. Demais modelos teóricos que contabilizam os três componentes podem ser vistos em Resende (2002; 2007)

A predição de indivíduos não-observados com base em parentesco também é uma habilidade dos modelos de seleção genômica (MEUWISSEN et al., 2011), no entanto, a genotipagem de todos os indivíduos com marcadores moleculares ainda é um processo caro (GROVER; SHARMA, 2016). Nesse sentido, a utilização da matriz A (com base em pedigree) ainda é uma alternativa interessante e mais viável em detrimento da matriz G (com base em marcadores moleculares). A predição de progênies com base na matriz aditiva A não é algo novo no setor florestal e diversos autores já realizaram trabalhos relevantes com esse objetivo principalmente para os gêneros *Pinus* e *Eucalyptus* (ZHANG et al., 2015; SUONTAMA et al., 2015; OGUT et al., 2014). Os modelos 1 e 3 (Tabela 2) não são capazes de realizar a predição para outros locais de indivíduos não-observados nestes modelos. Isso ocorre simplesmente devido à falta de informação destes níveis (i.e., clones) no fator genotípico dos modelos (MRODE, 2014). Por outro lado, graças às informações de parentesco, os modelos 2 e 4 foram capazes de prever EBV para indivíduos não-observados nos experimentos utilizados no modelo de treinamento (acurácias apresentadas na Tabela 4). As acurácias de predição dos indivíduos dos testes de validação (observados e não observados no modelo de treinamento) foram muito parecidas com as acurácias de predição dos indivíduos observados. Esses resultados indicam que os modelos 2 e 4 foram eficientes na predição, mantendo as acurácias elevadas (Tabela 4).

Tabela 4. Acurácia predita (realizada) após a validação-cruzada. É apresentado os valores para os modelos com (Pedigree) e sem (Unrelated) consideração de parentesco entre clones e variância genética homogênea (ID) e heterogênea (DIAG) entre ambientes.

Table 4. Prediction accuracy after cross-validation considering and not considering the kinship between the clones and with (DIAG) and without (ID) heterogeneous genetic variance between trials.

σ_g^2 entre ambientes →		ID		DIAG	
Parentesco →		Unrelated	Pedigree	Unrelated	Pedigree
Experimento	P / V	modelo 1	modelo 2	modelo 3	modelo 4
TC01	24 / 28	0,762	0,854 / 0,834	0,766	0,857 / 0,839
TC02	7 / 12	0,651	0,700 / 0,611	0,647	0,698 / 0,610
TC03	8 / 12	0,876	0,881 / 0,882	0,879	0,882 / 0,884
TC04	12 / 20	0,672	0,876 / 0,864	0,670	0,872 / 0,862
TC05	7 / 10	0,818	0,631 / 0,747	0,823	0,634 / 0,747
TC06	9 / 9	0,820	0,841 / 0,841	0,826	0,846 / 0,846
TC07	8 / 8	0,823	0,903 / 0,903	0,824	0,906 / 0,906
TC08	6 / 10	0,497	0,593 / 0,693	0,495	0,590 / 0,691
TC09	7 / 8	0,856	0,788 / 0,791	0,873	0,826 / 0,827
TC10	6 / 6	0,879	0,922 / 0,922	0,888	0,928 / 0,928
TC11	7 / 7	0,804	0,863 / 0,863	0,800	0,854 / 0,854
Todos	34 / 63	0,769	0,805 / 0,814	0,772	0,808 / 0,817

P é a quantidade de clones em sobreposição entre o modelo de predição e o modelo de treinamento; V é a quantidade de clones no modelo de validação (total de clones do experimento). Entre os valores de acurácia, os ao lado esquerdo da barra são as acurácias preditas para P e ao lado direito as acurácias preditas para V. Note que para os modelos do tipo Unrelated, não há valores de acurácia, isso ocorre por limitação do método. P is the number of overlapping clones between the prediction model; validation test; V is the total number of clones in the validation test. Among the accuracy values; on the left side of the bar are the values for P; on the right side, the values for V. Note that for Unrelated models, there are no accuracy values due to method limitation.

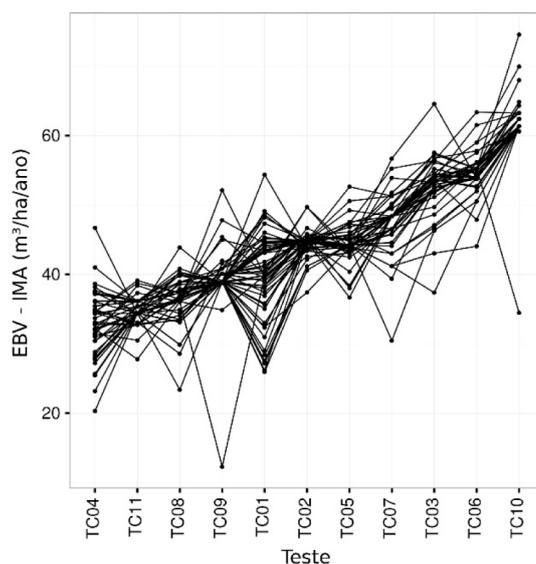


Figura 1. Interação Genótipo por Ambiente (GA). EBV são os valores genéticos aditivos do caráter MAI (incremento médio anual) para cada clone nos 11 testes avaliados, obtidos por meio do modelo 4. Os experimentos foram ordenados pela média dos EBV's de cada experimento, do menor para o maior, para melhor visualização.

Figure 1. Genotype by environment interaction (GE). EBV are the estimated breeding values of trait MAI (mean annual increment) for each clone in 11 evaluated tests, obtained by through the model 4. Environmental trials were ordered by clone's EBV means, from smallest to largest.

Redução das parcelas experimentais

Os talhões florestais são áreas extensas preenchidas com a replicação de um único clone. A necessidade de reproduzir experimentos com características semelhantes aos plantios comerciais é um tema questionável no meio científico (ZHANG et al., 2015). Nossos resultados demonstram que foi possível alcançar acurácias de predição elevadas avaliando menos plantas dentro de parcelas. Scarpinati et al. (2009) relatam que a utilização de diferentes formas de parcela (STP, linear e quadrada) não alteram o ordenamento dos clones e recomendam a utilização de STP na seleção clonal. Uma clara estabilização das acurácias preditivas é observada a partir de parcelas com apenas 4 árvores

avaliadas (Figura 2). Tomando o número intermediário de 16 plantas dentro de parcela, as acurácias foram em média (linha tracejada na Figura 2) iguais a 0,74, 0,79, 0,75 e 0,80 para os modelos 1 a 4, respectivamente. As acurácias dos modelos com todas as árvores foram de 0,77, 0,80, 0,77 e 0,81 para os modelos 1 a 4, respectivamente. Com base nesses valores, o modelo 4 foi novamente eficiente, sendo também útil na avaliação de experimentos com menos plantas por parcela.

O número ótimo de parcelas está atrelado ao número de repetições experimentais (PARANAIBA et al., 2009). Nesse sentido, embora a inclusão da matriz A tenha se mostrado eficiente em modelos com e sem contabilização da variância genética para cada ambiente e em diferentes números de plantas avaliadas por parcela, o número de repetições não foi avaliado neste trabalho. Todavia, Gezan et al. (2006) e Stanger et al. (2011) demonstram que a utilização de menos plantas por parcelas com aumento no número de repetições é uma maneira eficiente de manter a acurácia preditiva reduzindo-se os gastos. Com base nos resultados obtidos por esses autores, é possível supor que a aplicação da matriz A em experimentos com maior número de repetições e menos plantas por parcela pode ser uma maneira complementar de alcançar acurácias ainda maiores.

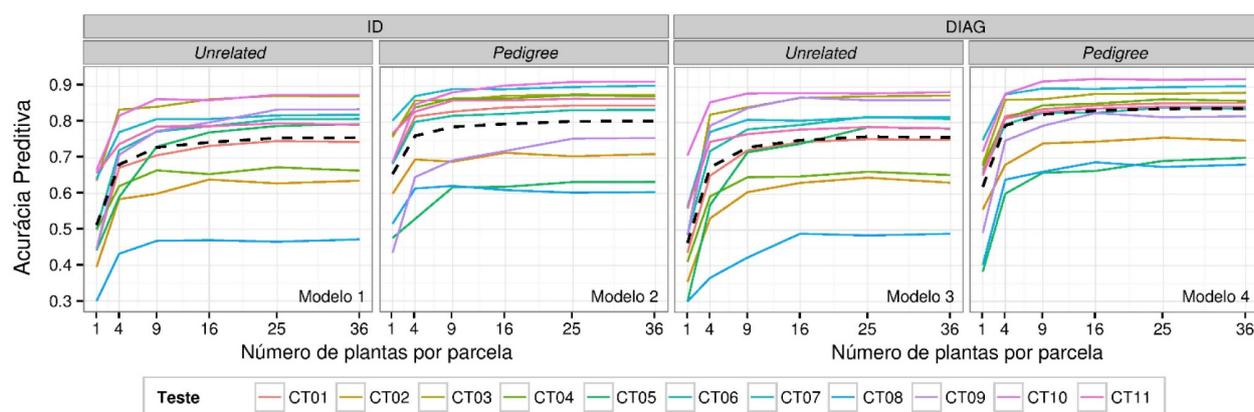


Figura 2. Acurácias preditivas após a redução das avaliações em cada parcela. ID representa os modelos de variância genética homogênea entre experimentos. DIAG representa os modelos com variância genética heterogênea entre experimentos. Para os dois casos foi assumido ausência de parentesco entre os clones (Unrelated) e parentesco entre clones (Pedigree). A linha tracejada é a média entre todos os experimentos.

Figure 2. Prediction accuracies after plot reductions. ID depicts the model with no genetic variance structure. DIAG depicts the model that uses genetic heterogeneity variance among experiments. For both cases, the relationship among clones as unrelated was adopted; the kinship matrix based on pedigree information. The dashed line is the mean of all lines.

No TC08 houve menores acurácias preditivas para todas as quantidades de árvores por parcela (Figura 2), haja vista que este experimento é justamente o de menor número de plantas por parcela (36). Mesmo assim, os modelos 2 e 4 demonstraram acurácias preditivas superiores quando comparadas com os demais modelos, além disso, no modelo 4 é possível observar uma aproximação das acurácias deste teste às dos demais testes avaliados.

CONCLUSÕES

É possível alcançar elevadas acurácias preditivas entre ambientes utilizando os modelos do tipo MET (*multi-environment trials*), com acerto dos valores genéticos igual a 77%. A incorporação de informação de parentesco entre os clones e assumindo-se variâncias genéticas heterogêneas entre ambientes a taxa de acerto foi de 81%. Informações de parentesco são úteis na predição de clones não observados nos modelos de preditivos, mantendo elevadas as acurácias preditivas dos materiais genéticos.

Nos testes clonais ampliados, a avaliação de 4 a 16 plantas por parcela, com 3 repetições, é suficiente para garantir acurácias preditivas semelhantes à avaliação de 100 plantas por parcela.

AGRADECIMENTOS

Os autores agradecem a empresa CENIBRA S.A. pela disponibilidade dos dados fenotípicos e as instituições CAPES, CNPq e CENIBRA S.A. pelo suporte financeiro.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, New York, v. 19, n. 6, p. 716-723, 1974.
- BUSH, D.; KAIN, D.; KANOWSKI, P. MATHESONM C. Genetic parameter estimates informed by a marker-based pedigree: a case study with *Eucalyptus cladocalyx* in southern Australia. **Tree genetics & genomes**, v. 11, n. 1, p. 1-16, 2015.
- BUTLER, D. G.; CULLIS, B. R.; GILMOUR, A. R.; GOGEL, B. J. **ASReml-R reference manual**. Queensland, AU: Queensland Department of Primary Industries, 2009.
- EL-KASSABY, Y. A; CAPP, E. P.; LIEWLAKSANEYANAWIN, C. K LAPSTE, J.; LSTIBUREK, M. Breeding without breeding: is a complete pedigree necessary for efficient breeding?. **PLoS One**, v. 6, n. 10, 2011.
- GEZAN, S. A.; WHITE, T. L.; HUBER, D. A. Comparison of experimental designs for clonal forestry using simulated data. **Forest Science**, Bethesda, v. 52, n. 1, p. 108-116, 2006.
- GILMOUR, A. R.; THOMPSON, R.; CULLIS, B. R. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Raleigh, v. 51, n. 4, p. 1440-1450, 1995.
- GROVER, A.; SHARMA, P. C. Development and use of molecular markers: past and present. **Critical reviews in biotechnology**, v. 36, n. 2, p. 290-302, 2016.
- HAYES, B. J.; VISSCHER, P. M.; GODDARD, M. E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research**, Cambridge, v. 91, n. 1, p. 47-60, 2009.
- HENDERSON, H. C. R. **Applications of Linear Models in Animal Breeding**. Ontario CA: University of Guelph, Guelph, 1984.
- HESTERBERG, T.; MONAGHAN, S.; MOORE, D. S.; CLIPSON, A.; EPSTEIN, R. **Bootstrap methods and permutation tests**. New York: W. H. Freeman and Company, 2005. 70 p. (Introduction to the Practice of Statistics, v. 5).
- JESUS, G. L.; SILVA, I. R.; ALMEIDA, L. F. J.; SANTOS, M. A.; LEITE, F. P.; NEVES, J. C. L. Eucalyptus productivity, soil physical properties and organic matter fractions influenced by traffic intensity and harvest residues. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 39, n. 4, p. 1190-1203, 2015.
- KOHAVI, R. **A study of cross-validation and bootstrap for accuracy estimation and model selection**. 1995. Disponível em: < <http://ai.stanford.edu/~ronnyk/accEst.pdf> >. Acesso em 10 out. 2016.
- LAMBETH, C.; ENDO, M.; WRIGHT, J. Genetic analysis of 16 clonal trials of *Eucalyptus grandis* and comparisons with seedling checks. **Forest Science**, Bethesda, v. 40, n. 3, p. 397-411, 1994.
- MEUWISSEN T.H.E.; HAYES, B. J.; GODDARD M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Pittsburgh, v. 157, n. 4, p. 1819-1829, 2001.
- MRODE, R. A. **Linear models for the prediction of animal breeding values**. 2 ed. Oxfordshire: Cabi, 2014.344 p.
- OGUT, F; MALTECCA, C.; ROSS, W.; MCKEAND, S.; ISIK, F. Genetic analysis of Diallel progeny test data using factor analytic linear mixed models. **Forest Science**, Bethesda, v. 60, n. 1, p. 119-127, 2014.

OSORIO, L.; WHITE, T.; HUBER, D. Age–age and trait–trait correlations for *Eucalyptus grandis* Hill ex Maiden and their implications for optimal selection age and design of clonal trials. **Theoretical and Applied Genetics**, Berlin, v. 106, n. 4, p. 735-743, 2003.

PARANAIBA, P. E.; FERREIRA, D. F.; DE MORAIS, A. R. Tamanho ótimo de parcelas experimentais: proposição de métodos de estimação. **Revista Brasileira de Biometria**, Lorena, v. 27, n. 2, p. 255-268, 2009.

PIEPHO, H. P.; MÖHRING, J. Computing heritability and selection response from unbalanced plant breeding trials. **Genetics**, Pittsburgh, v. 177, n. 3, p. 1881-1888, 2007.

PINTO, D. S. RESENDE, T. T.; MESQUITA, G. G.; ROSADO, A. M.; CRUZ, C. D. Early selection in tests for growth traits of *Eucalyptus urophylla* clones test. **Scientia Forestalis**, Piracicaba, v. 42, n. 102, p. 251-257, 2014.

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

RESENDE, M. D. V.; BARBOSA, M. H. P. **Melhoramento genético de plantas de propagação assexuada**. Colombo: Embrapa Florestas, 2005. 130 p.

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

SCARPINATI, E. A.; PERECIN, D.; PAULA, R. C.; BONINE, C. A. V.; PAVAN, B. E.; CANDIDO, L. S. Influences of experimental plot forms and statistical model analysis in *Eucalyptus* spp Clone selections. **Revista Árvore**, Viçosa, v. 33, n. 4, p. 769-776, 2009.

SCHUMACHER, F.; HALL, F. Logarithmic expression of timber-tree volume. **Journal of agricultural research**, Washington, v.47, n. 719-734, 1933.

SILVA, J. C.; KERR, R. J. Accounting for competition in genetic analysis, with particular emphasis on forest genetic trials. **Tree genetics & genomes**, v. 9, n. 1, p. 1-17, 2013.

STANGER, T. K.; GALLOWAY, G. M.; RETIEF, E. C. L. Final results from a trial to test the effect of plot size on *Eucalyptus* hybrid clonal ranking in coastal Zululand, South Africa. **Southern Forests: a Journal of Forest Science**, v. 73, n. 3-4, p. 131-135, 2011.

SUONTAMA, M.; LOW, C. B.; DUNGEY, H. Genetic parameters and genetic gains across three breeding cycles for growth and form traits of *Eucalyptus regnans* in New Zealand. **Tree Genetics & Genomes**, v. 11, n. 6, p. 1-14, 2015.

ZHANG, J.; PETER, G. F.; POWELL, G.; WHITE, T.; GEZAN, S. A. Comparison of breeding values estimated between single-tree and multiple-tree plots for a slash pine population. **Tree Genetics & Genomes**, v. 11, n. 3, p. 1-10, 2015.

Recebido em 27/02/2017

Aceito para publicação em 01/07/2016