



Revista Árvore

ISSN: 0100-6762

r.arvore@ufv.br

Universidade Federal de Viçosa
Brasil

Calegario, Natalino; Daniels, Richard F.; Lopes Souza, Agostinho; Maestri, Romualdo
Estimativa do crescimento de povoamentos de Eucalyptus baseada em modelos lineares em
multiníveis de efeito misto
Revista Árvore, vol. 29, núm. 2, março-abril, 2005, pp. 251-261
Universidade Federal de Viçosa
Viçosa, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=48829208>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

ESTIMATIVA DO CRESCIMENTO DE POVOAMENTOS DE *Eucalyptus* BASEADA EM MODELOS LINEARES EM MULTINÍVEIS DE EFEITO MISTO¹

Natalino Calegario², Richard F. Daniels³, Agostinho Lopes Souza⁴ e Romualdo Maestri⁵

RESUMO – A proposta do presente estudo foi a modelagem do crescimento de povoamentos clonais de *Eucalyptus*, com base em modelos lineares mistos em multiníveis. A base de dados utilizada foi proveniente de povoamentos homogêneos, localizados na região costal brasileira, nos Estados do Espírito Santo e da Bahia. Foram utilizados dois níveis aleatórios de modelagem: unidades amostrais e árvores individuais dentro das unidades amostrais. Como exemplo de aplicação do método foi utilizado o logaritmo da área basal, como resposta, em função do inverso da idade, da altura total e da interação entre elas. Com a utilização de técnicas de modelagem baseada nos efeitos fixos e mistos, as estimativas dos parâmetros foram melhoradas significativamente. Também, com a modelagem da autocorrelação e da heterogeneidade da variância, partindo-se de um modelo homoscedástico para um modelo heteroscedástico auto-regressivo, com estrutura de variância positiva definida no nível 1 e com estrutura de correlação auto-regressiva de primeira ordem (AR1), os valores do logaritmo da máxima verossimilhança foram significativamente elevados.

Palavras-chave: Modelos mistos em multiníveis, *Eucalyptus*, autocorrelação e heteroscedasticidade.

Eucalyptus STAND GROWTH ESTIMATE BASED ON MULTILEVEL LINEAR MIXED-EFFECTS MODELS

Abstract – The main purpose of this study was to apply the mixed-effects theory in modeling *Eucalyptus* stand growth. The database is from homogeneous stands located on Brazilian coastal region, Espírito Santo and Bahia States. We used two levels of sampling: sample units and individual trees inside of sample units. As an example, we considered the logarithm of the basal area as response variable and inverse of age, logarithm of the total height and their interaction as covariates. The parameter estimate quality of the linear model was significantly improved. Also, by modeling the heteroscedasticity and the autocorrelation (AR(1)), the log-likelihood values were increased significantly.

Key words: Multilevel mixed-effects model; *Eucalyptus*; autocorrelation; heteroscedasticity.

¹ Recebido em 01.09.2003 e aceito para publicação em 25.11.2004.

² Departamento de Ciências Florestais – UFLA, 37200-000 Lavras, Minas Gerais. (calegari@ufla.br).

³ School of Forest Resources, University of Georgia, Athens, 30606, Georgia, USA (ddaniels@smokey.forestry.uga.edu).

⁴ Departamento de Engenharia Florestal da Universidade Federal de Viçosa. 36570-000 Viçosa, Minas Gerais, Brasil. (alsouza@ufv.br).

⁵ Aracruz Celulose, Aracruz, Espírito Santo, Brasil (maestri@aracruz.br).

1. INTRODUÇÃO

Os modelos lineares em multiníveis são baseados em fatores de classificação aninhados. No caso de crescimento de povoamentos homogêneos, normalmente se têm dois níveis aleatórios: as unidades amostrais (nível 1) e as árvores dentro das unidades amostrais (nível 2). Baseando-se na definição de Daniels e Burkhardt (1988), os modelos mistos poderiam ser considerados um nível de resolução intermediário entre modelos de árvore individual dependente da distância e não-dependente da distância. Embora a base de dados tenha um nível de resolução independente da distância, a abordagem da modelagem inclui efeitos aleatórios e processos espaço-temporais, podendo estes ser considerados um nível superior de resolução comparados com os modelos de efeito fixo, cotidianamente utilizados no processo de estimativa.

A aplicação dos modelos em multiníveis tem crescido nos últimos anos. Autores como Goldstein (1987, 1995), Lindstrom e Bates (1988), Karlsson e Sheiner (1994) e Longford (1993) vêm publicando vários estudos baseados em mais de um nível aleatório na modelagem.

Na área florestal, modelos em multiníveis de efeito misto ainda são pouco usados devido, principalmente, ao fato de a sua complexidade ser uma metodologia relativamente recente e a limitações em software. Uma exceção é o trabalho de Fang (1999), que modelou o crescimento de *Pinus* sp., baseando-se nesses modelos. Em outro trabalho mais recente, Jayaraman e Lappi (2001) utilizaram os modelos em multiníveis para a estimativa da relação hipsométrica em povoamentos de Teca (*Tectona grandis* L.f.).

Em povoamentos homogêneos de *Eucalyptus* não se têm informações de sua aplicação. Portanto, os objetivos deste artigo foram aplicar a teoria dos modelos lineares mistos em multiníveis para a modelagem do crescimento em área basal, testar e corrigir heteroscedasticidade e autocorrelação de clones de eucaliptos, visando melhorar a qualidade do ajuste do modelo.

2. MATERIAL E MÉTODOS

2.1. Base de dados

A base de dados utilizada é proveniente de povoamentos comerciais de clones do gênero *Eucalyptus*, da região costal brasileira, nos Estados do Espírito Santo e da Bahia, localizados entre as

coordenadas 17°48' S e 40° 17' W. Cada combinação parcela/clone é por um gráfico, e cada árvore é representada por uma linha nas coordenadas dos gráficos, relacionando a área basal, ou seccional, por árvore, com a idade, em anos (Figura 1). Por exemplo, o clone de número AR4 está presente em duas unidades amostrais, as de números 18771 e 18764. Cada unidade foi permanente e aleatoriamente amostrada com idade variando entre 2 e 7,5 anos, no período de 1996 a 2002, e a área das unidades amostrais variou de 131 a 200 m². Considerando tais variações, a base de dados pode ser classificada como longitudinal, irregularmente espaçada e desbalanceada. Longitudinal devido ao fato de as medidas serem dos mesmos indivíduos repetidas no tempo e desbalanceada em razão de a distância, em tempo, entre as diferentes medidas não ser a mesma.

Para o estudo, devido ao grande número de observações, a base de dados foi reduzida para sete clones, oito unidades amostrais e, em média, 30 árvores dentro de cada unidade amostral.

2.2. Modelo linear em multiníveis de efeito misto generalizado

A forma generalizada desses modelos é dada pela expressão (1). Nesse caso, a variável resposta y_{ijk} é medida para o primeiro nível i , para o segundo nível j e na ocasião k :

$$y_{ijk} = X_{ijk}\beta + Z_{i,jk}b_i + Z_{ijk}b_{ij} + \varepsilon_{ijk} \quad (1)$$

em que:

X_{ijk} são matrizes de efeito fixo com dimensão $(n_{ijk} \times p)$; β representa o vetor de parâmetros fixos com dimensão $(p \times 1)$; $Z_{i,jk}$ são matrizes com dimensão $(n_i \times q_1)$, associadas com o primeiro nível de efeito aleatório b_i ; Z_{ijk} são matrizes com dimensões $(n_i \times q_2)$ associadas com o segundo nível de efeito aleatório b_{ij} ; e ε_{ijk} são erros estocásticos. Mais detalhes sobre modelos em multiníveis podem ser encontrados em Davidian e Giltinan (1995), Vonesh e Chinchilli (1997), Littell et al. (1996) e Calegario (2002).

Baseado na série de dados apresentada na Figura 1, o modelo em dois níveis foi ajustado, tendo como variável resposta o logaritmo da área basal ($\ln(g)$), por árvore individual, e como principal regressor o inverso da idade ($1/t$). Também foram incluídos como regressores

o logaritmo da altura e a interação entre o inverso da idade e o logaritmo da altura (2).

$$\ln(g)_{ijk} = (\beta_0 + b_{0i} + b_{0j}) + (\beta_1 + b_{1i} + b_{1j}) \frac{1}{t_{ijk}} + (\beta_2 + b_{2i} + b_{2j}) \ln(h)_{ijk} + (\beta_3 + b_{3i} + b_{3j}) \frac{1}{t_{ijk}} \ln(h)_{ijk} + \varepsilon_{ijk}$$

$$b_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \\ b_{3i} \end{bmatrix} \sim N(0, \Psi_1), b_j = \begin{bmatrix} b_{0j} \\ b_{1j} \\ b_{2j} \\ b_{3j} \end{bmatrix} \sim N(0, \Psi_2) \text{ e } \varepsilon_{ijk} \sim N(0, \sigma^2). \quad (2)$$

Os parâmetros β_0 , β_1 e β_2 são de efeito fixo e b_i representa o vetor dos efeitos aleatórios em nível de unidade amostral, b_j representa os efeitos aleatórios em nível de árvore e ε_{ijk} são os erros dentro dos grupos. Assume-se que os b_i são independentes para diferentes unidades amostrais, os b_j são independentes para diferentes árvores, diferentes unidades amostrais e independentes de b_i . Os erros ε_{ijk} são independentes para diferentes unidades amostrais, árvores e observações, e são independentes dos efeitos aleatórios.

2.3. Modelagem da heterogeneidade de variância e da autocorrelação

Na Figura 1, representa um aumento da variabilidade em torno da curva média de crescimento à medida que a idade dos indivíduos aumenta, podendo classificar

essa base de dados como heteroscedástica. Tal característica viola as pressuposições básicas de homogeneidade da variância, com a ressalva de que esta deve ser corrigida.

Além da violação da homogeneidade de variância, a base de dados da Figura 1 também viola a pressuposição de independência das observações, por se tratar de verificações proveniente de remedições de árvores individuais no tempo, ou seja, dados longitudinais.

As técnicas apresentadas aqui na modelagem de tais violações serão baseadas nas propostas de Davidian e Giltinan (1995), as quais apresentaram as seguintes expressões:

$$\text{var}(y_i) = \sigma^2 g^2(\mu_i, z_i, \Phi), \quad \mu_i = f(x_i, \beta) \quad (3)$$

A variância da resposta em (1) é função de g que, por sua vez, o é da média da resposta, de fatores fixos z , que podem ser representados por parte ou por todos os componentes de x e do vetor de parâmetros Φ da função de variância. Como a resposta média é função dos parâmetros da regressão β , a variância também é função desses parâmetros. A função $g(\cdot)$ pode possuir várias formas. As mais comuns são a forma exponencial, o parâmetro como uma potência da média e a forma com dois componentes.

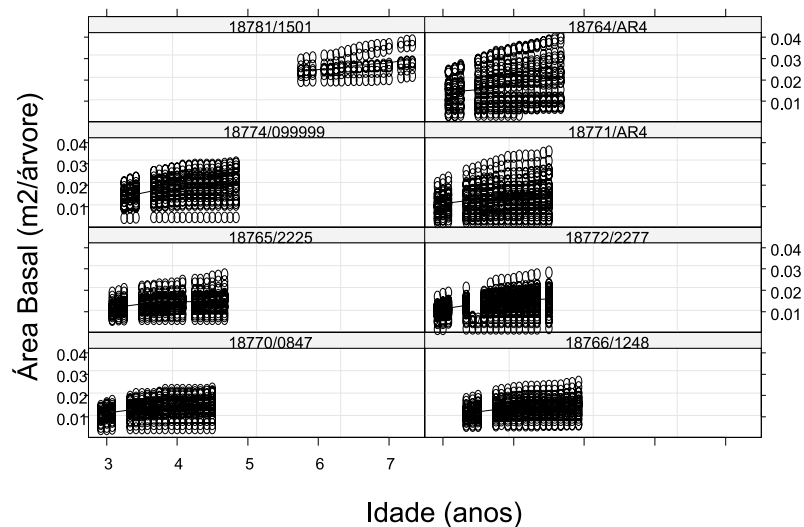


Figura 1 – Relação entre a área basal e a idade da árvore.
Figure 1 – Relationship between the basal area and age.

O processo de estimativa da função de variância é baseado nos mínimos quadrados generalizados. Após a estimativa do parâmetro Φ e da escolha dos valores iniciais para β , um processo iterativo gera valores definitivos para os parâmetros pela minimização da função de pseudoverossimilhança, conforme a seguinte expressão:

$$PV(\beta^{(0)}, \phi, \Phi) = \sum_{i=1}^n \left(-\frac{\{y_i - f(x_i, \beta^{(0)})\}^2}{\hat{\sigma}^2 g^2\{f(x_i, \beta^{(0)}), z_i, \Phi\}} + \log[\hat{\sigma}^2 g^2\{f(x_i, \beta^{(0)}), z_i, \Phi\}] \right) \quad (4)$$

Tecnicamente, a minimização de (4) significa a maximização da verossimilhança em $\beta^{(0)}$. Para a minimização, por iteração, é necessário o conhecimento de Φ . Alguns métodos de estimativa do citado parâmetro pode ser encontrado em Carroll e Ruppert (1988). Independentemente da variância e da sua função $g(\cdot)$, a minimização em (4) implica minimização das somas de quadrados dos erros ($\{y_i - f(x_i, \beta^{(0)})\}^2$). Porém, quanto mais apropriados os valores estimados da variância e da sua função $g(\cdot)$, menor a soma do quadrado dos erros.

Para a modelagem da autocorrelação, a seguinte expressão pode ser utilizada:

$$Corr(e) = \Gamma(\alpha) \quad (5)$$

Ou em forma matricial,

$$\Gamma(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \dots & \alpha^{n-1} \\ & 1 & \alpha & \alpha^2 & \dots & \alpha^{n-2} \\ & & 1 & \alpha & \dots & \alpha^{n-3} \\ & & & 1 & \dots & \alpha^{n-4} \\ & & & & \ddots & \alpha^{n-5} \\ & & & & & \ddots & \alpha^{n-6} \\ & & & & & & \ddots & \alpha^{n-7} \\ & & & & & & & \ddots & \alpha^{n-8} \\ & & & & & & & & \ddots & \alpha^{n-9} \\ & & & & & & & & & \ddots & \alpha^{n-10} \\ & & & & & & & & & & \ddots & \alpha^{n-11} \\ & & & & & & & & & & & \ddots & \alpha^{n-12} \\ & & & & & & & & & & & & \ddots & \alpha^{n-13} \\ & & & & & & & & & & & & & \ddots & \alpha^{n-14} \\ & & & & & & & & & & & & & & \ddots & \alpha^{n-15} \\ & & & & & & & & & & & & & & & \ddots & \alpha^{n-16} \\ & & & & & & & & & & & & & & & & \ddots & \alpha^{n-17} \\ & & & & & & & & & & & & & & & & & \ddots & \alpha^{n-18} \\ & & & & & & & & & & & & & & & & & & \ddots & \alpha^{n-19} \\ & & & & & & & & & & & & & & & & & & & \ddots & \alpha^{n-20} \\ & \ddots & \alpha^{n-21} \\ & \ddots & \alpha^{n-22} \\ & \ddots & \alpha^{n-23} \\ & \ddots & \alpha^{n-24} \\ & \ddots & \alpha^{n-25} \\ & \ddots & \alpha^{n-26} \\ & \ddots & \alpha^{n-27} \\ & \ddots & \alpha^{n-28} \\ & \ddots & \alpha^{n-29} \\ & \ddots & \alpha^{n-30} \\ & \ddots & \alpha^{n-31} \\ & \ddots & \alpha^{n-32} \\ & \ddots & \alpha^{n-33} \\ & \ddots & \alpha^{n-34} \\ & \ddots & \alpha^{n-35} \\ & \ddots & \alpha^{n-36} \\ & \ddots & \alpha^{n-37} \\ & \ddots & \alpha^{n-38} \\ & \ddots & \alpha^{n-39} \\ & \ddots & \alpha^{n-40} \\ & \ddots & \alpha^{n-41} \\ & \ddots & \alpha^{n-42} \\ & \ddots & \alpha^{n-43} \\ & \ddots & \alpha^{n-44} \\ & \ddots & \alpha^{n-45} \\ & \ddots & \alpha^{n-46} \\ & \ddots & \alpha^{n-47} \\ & \ddots & \alpha^{n-48} \\ & \ddots & \alpha^{n-49} \\ & \ddots & \alpha^{n-50} \end{bmatrix} \quad (6)$$

A expressão (6), na realidade, representa a situação de correlação auto-regressiva de ordem 1 (AR1). Caso se tenham, por exemplo, duas medidas de área basal do i-ésimo indivíduo nos tempos $t_{i,1}$ e $t_{i,2}$, os componentes da expressão (6) poderiam ser substituídos por $\alpha^{|t_{i,1}-t_{i,2}|}$. Existem vários outros padrões de autocorrelação disponíveis na literatura.

Outro padrão de correlação seria a combinação da auto-regressiva com a média móvel – ARMA (BOX et al., 1994). Esse padrão pode ser generalizado pela seguinte estrutura:

$$\varepsilon_t = \sum_{i=1}^p \phi_i \varepsilon_{t-i} + \sum_{j=1}^q \eta_j a_{t-j} + a_t \quad (7)$$

em que ε_t se refere a uma observação no tempo t , e a_t é o erro-termo. A primeira parte da expressão (7) refere-se ao modelo auto-regressivo (AR p) e a segunda parte, ao de média móvel (MA q). Se $p=0$, tem-se uma MA q situação e, ao contrário, se $q=0$, a situação seria de AR p . Na AR p parte, ϕ representa os parâmetros de correlação com ordem p e $(t-i)$, a distância entre duas observações (lag). A tendência é de que os valores de ϕ decresçam com o tempo, indicando que observações próximas no tempo são mais correlacionadas do que observações distantes, o que é comum em estudos de dados longitudinais dessa natureza. Na parte do movimento da média (MA q), o modelo assume que uma observação atual é uma função linear dos erros termos (a_t), idêntica e independentemente distribuídos.

2.4 Critério de seleção do melhor modelo

A seleção dos modelos gerados foi baseada no Teste da Razão da Máxima Verossimilhança (TRMV), o que é uma técnica de seleção para modelos aninhados. Modelos aninhados ocorrem quando um modelo representa um caso especial de outro. Se MV_2 corresponder ao maior valor da máxima verossimilhança para um modelo mais geral e MV_1 for o menor valor para um modelo restrito, o valor de TRMV será:

$$TRMV = 2 \log(MV_2 / MV_1) = 2 [\log(MV_2) - \log(MV_1)] \quad (8)$$

em que MV_2 e MV_1 são os valores da máxima verossimilhança para os modelos mais geral e mais restrito, respectivamente. Desde que i -ésimo $MV_2 > MV_1$, o valor de TRMV será positivo e, se k_1 for o número de parâmetros no i -ésimo modelo, a distribuição de TRMV será χ^2 com $(k_2 - k_1)$ graus de liberdade. O valor de TRMV é comparado com o valor crítico de $\chi^2(k_2 - k_1, \alpha)$ e, se $TRMV > \chi^2(k_2 - k_1, \alpha)$, gerando valor- p ($< 0,05$), o modelo mais geral será preferido quando comparado com o restrito.

A precisão do modelo pode ser avaliada por técnicas de Critérios de Informações Estatísticas. Tais critérios são basicamente representados por dois métodos: Critério de Informação de Akaike (CIA) (SAKAMOTO et al., 1986) e Critério de Informação Bayesiana (CIB) (SCHWARZ, 1978). Tais critérios são avaliados como:

$$CIA = -2 \log(MV) + 2 n_{\text{par}} \quad (9)$$

$$CIB = -2 \log(MV) + n_{\text{par}} \log(N) \quad (10)$$

para cada modelo, em que n_{par} é o número de parâmetros no modelo e N o número total de observações. Menores valores de ambos os critérios implicam melhor ajuste. Como esses critérios são conservativos (STRAM e LEE, 1994), gerando maiores valores-p do que deveriam, é aconselhável usar um valor de α de 10% para selecionar o melhor modelo.

3. RESULTADOS E DISCUSSÕES

3.1. Ajuste do modelo linear de efeito misto em dois níveis

Devido ao elevado número de observações, número de árvores dentro das unidades amostrais e 20 componentes de variância-co-variância, o primeiro modelo de efeito misto foi ajustado assumindo-se que Ψ_1 e Ψ_2 são matrizes diagonais, as quais tornaram o perfil da máxima verossimilhança mais estável. Os parâmetros estimados e os respectivos desvios-padrão são apresentados na Tabela 1. Os valores estimados para os efeitos fixos estão de acordo com a tendência de mudança do logaritmo da área basal com o aumento do inverso da idade. O valor negativo do intercepto ($\beta_0 = -4,1151$) se deu pelo fato de os valores da área basal por árvore individual serem menores do que 1 e, como consequência, o seu logaritmo é negativo. Os parâmetros associados ao inverso da idade ($\beta_1 = 19,3416$) e ao logaritmo da altura total ($\beta_2 = -0,0442$) também apresentaram valores negativos, pelo fato de o logaritmo da área basal reduzir seu valor com o aumento de ambas as quantidades. O valor significativo da interação ($\beta_3 = 5,5481$) indica que o crescimento em área basal é diferente nos dois regressores. Os efeitos fixos apresentaram significância estatística no nível 0,05 de probabilidade (p-valor < 0,05), com exceção de β_2 ,

que teve um p-valor igual a 0,0603. A variável $\ln(h)$ foi mantida no modelo devido a sua significativa interação com o inverso da idade. Os componentes de co-variância entre os efeitos aleatórios são considerados zero na estrutura diagonal utilizada. Os valores não muito reduzidos das estimativas dos desvios-padrão em ambos os níveis indicam que os efeitos aleatórios devem ser mantidos no modelo.

Uma das grandes vantagens dos modelos que incluem os efeitos aleatórios, com relação aos modelos de efeitos fixos, é a redução do erro-padrão residual. Com o ajuste do modelo incluindo apenas os efeitos fixos, ou seja, o inverso da idade, o logaritmo da altura e a interação entre eles, o erro-padrão gerado foi de 0,1964. Com a inclusão dos efeitos aleatórios, esse erro foi reduzido para 0,013350 (Tabela 1), sendo esta uma diminuição de aproximadamente 15 vezes na citada quantidade.

3.2. Verificação das pressuposições distribucionais

As pressuposições distribucionais são baseadas nos erros intragrupos (ϵ_{ijk}) e nos efeitos aleatórios (b_i e b_{ij}). Os erros intragrupos são considerados independentes e normalmente distribuídos, com média zero e variância σ^2 , e os efeitos aleatórios também possuem distribuição normal, com média zero, matriz de variância Ψ , além de serem independentes em diferentes grupos.

Tais pressuposições podem ser verificadas graficamente, com análises de resíduos. Na Figura 2, mostra-se a distribuição dos resíduos intragrupos, a qual indica que os resíduos estão aproximadamente distribuídos em torno de zero, mas pode não ter variância constante. Também se podem verificar alguns *outliers* e, ou, pontos de influência. O mesmo padrão pode ser

Tabela 1 – Estimativa dos parâmetros dos efeitos fixos, efeitos aleatórios e desvios-padrão pelo modelo linear misto em multiníveis

Table 1 – Parameter estimates of the fixed effects, random effects and standard errors for the linear multilevel mixed-effects model

Parâmetros Estimados	Efeitos Fixos					Desvio-Padrão dos Efeitos Aleatórios	
	Valor	Erro-padrão	G.L.	Valor-t	p-valor	Unidades Amostrais	Árvores dentro das Unidades Amostrais
β_0	-4,1151	0,084581	4032	-48,65	<0,0001	0,16484	0,45943
β_1	-19,3416	3,713849	4032	-5,20	<0,0001	10,40285	1,12316
β_2	-0,0442	0,023544	4032	-1,87	0,0603	0,03953	0,129426
β_3	5,5481	1,260866	4032	4,40	<0,0001	3,53865	0,318676
Residual	-	0,196400	-	-	-	-	0,013350

verificado na distribuição dos resíduos padronizados (Figura 3). Embora os resíduos totais estejam simetricamente distribuídos, os resíduos por unidade amostral têm diferentes padrões, os quais violam as pressuposições de homoscedasticidade, tendo que modelar esse padrão. Tal padrão foi confirmado pelo teste de Multiplicativo de Lagrange, gerando um valor significativo de probabilidade (p-valor<0,0001).

A distribuição residual do modelo heteroscedástico

está representada na Figura 4. Após a modelagem do padrão, a distribuição entre unidades amostrais se mostrou mais similar. O valor do logaritmo da máxima verossimilhança aumentou de 10,261 para 10,362, gerando um valor da razão dos logaritmos das máximas verossimilhanças de 203,15, com um p-valor < 0,0001, confirmando que o modelo heteroscedástico explica os dados significativamente melhor do que o modelo homoscedástico.

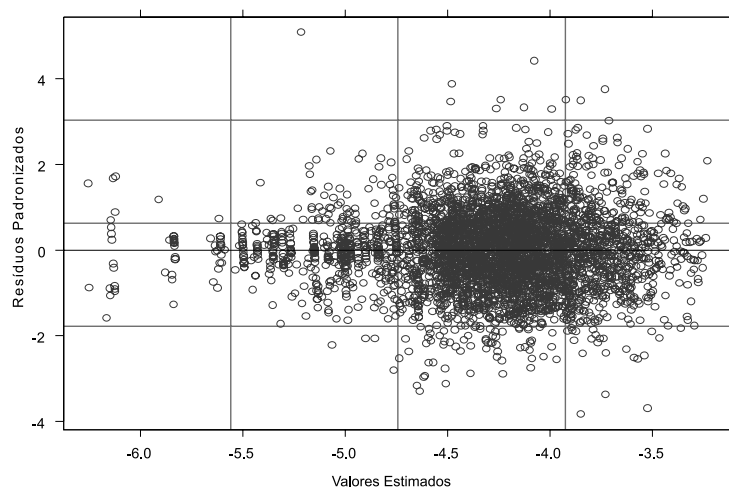


Figura 2 – Resíduos padronizados versus valores ajustados para o modelo linear em multiníveis.
Figure 2 – Standard residuals versus fitted values for the linear multilevel mixed model.

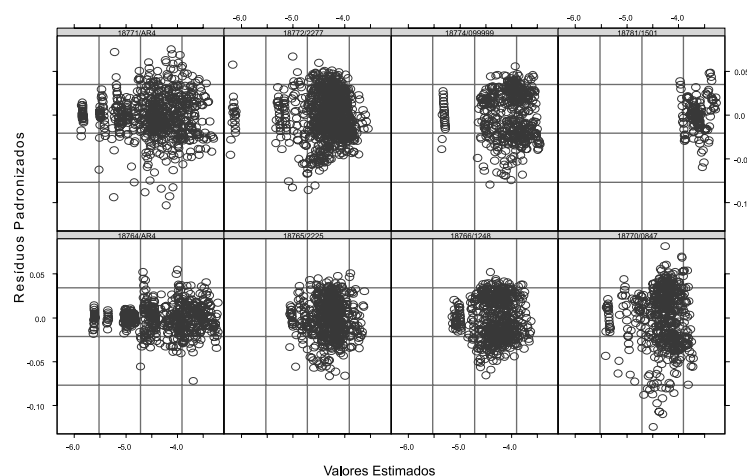


Figure 3 – Resíduos por unidade amostral do modelo linear em multiníveis.
Figure 3 – Residuals by sampling unit for the linear multilevel model.

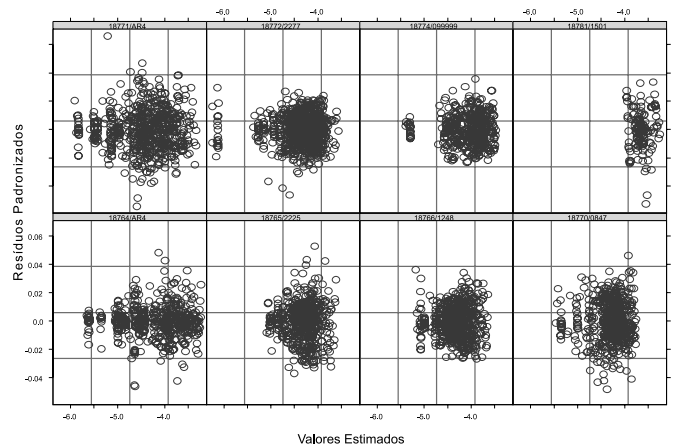


Figura 4 – Resíduos padronizados por unidade amostral do modelo linear em multiníveis após a modelagem da variância entre unidades amostrais (modelo heteroscedástico).

Figure 4 – Standard residuals per sampling unit for the linear multilevel model after modeling the heteroscedasticity between sampling units.

Para a verificação das pressuposições dos efeitos aleatórios, assumindo que estes são independentes, o modelo (2) foi inicialmente ajustado usando-se uma estrutura diagonal para a matriz de variância. Pode ser verificado na Figura 5 que os efeitos aleatórios não são independentes para o nível 1 (unidade amostral), com forte correlação entre b_{1i} e b_{3i} , o mesmo ocorrendo no nível 2 (Figura 6).

O modelo com estrutura positiva definida para o nível 1 foi ajustado para a checagem das pressuposições de independência. Os resultados estão na Tabela 2. Os valores da máxima verossimilhança estão dispostos em ordem crescente. Baseando nas informações das correlações dos erros dentro dos grupos e nas co-variâncias dos efeitos aleatórios, tentou-se modelar essas características levando em consideração os níveis 1 (unidade amostral) e 2 (árvores dentro das unidades amostrais). O melhor modelo foi gerado pela modelagem do padrão da heteroscedasticidade entre unidades amostrais e também contabilizando a correlação entre os efeitos aleatórios no nível 1, o qual teve um valor da máxima verossimilhança de 10379 e um valor p-valor significativo ($<0,0001$), quando comparado com o modelo heteroscedástico, o qual tem estrutura diagonal em ambos os níveis.

A estrutura de autocorrelação empírica foi modelada com base na visualização mostrada na Figura 7. Os

valores são positivos nos dois primeiros intervalos (lags), indicando que um modelo auto-regressivo de ordem 1 (AR(1)) pode ser adequado para a modelagem da correlação dentro dos grupos. Um valor inicial adequado para a correlação está em torno de 0,26, o qual é aproximadamente o valor da autocorrelação empírica no intervalo 1. Tal correlação foi confirmada pelo teste de Durbin-Watson, gerando um valor $d=0,157$, sendo esse valor menor do que o valor tabelado $d=1,66$, com $\alpha=0,01$, gerando um valor significativo de probabilidade (p-valor $<0,0001$).

O valor da máxima verossimilhança aumentou de 10359 para 10952, em comparação com o modelo heteroscedástico com estrutura positiva definida, com e sem a modelagem da estrutura de correlação, respectivamente. O teste da razão da máxima verossimilhança foi 1146,70, com um significativo valor de probabilidade ($<0,0001$), indicando que a estrutura auto-regressiva de ordem 1 representa a dependência dentro dos grupos.

As estimativas dos efeitos fixos dos diferentes modelos estão representadas na Tabela 3. Como pode ser verificado nessa tabela, os valores estimados dos efeitos fixos mudam com a modelagem da estrutura de co-variância, gerando valores estimados diferentes. Outra mudança ocorre no erro-padrão residual, e o modelo de melhor “performance” apresentou menor erro.

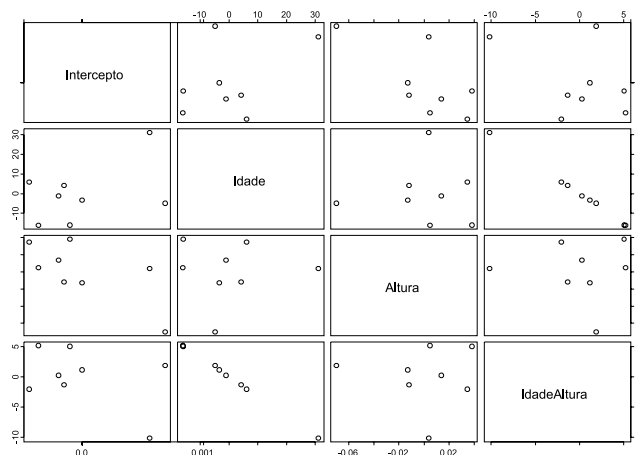


Figura 5 – Representação das correlações entre os efeitos aleatórios do primeiro nível estimado no modelo em multiníveis.
Figure 5 – Correlations between the random effects for the first level estimated in the multilevel model.

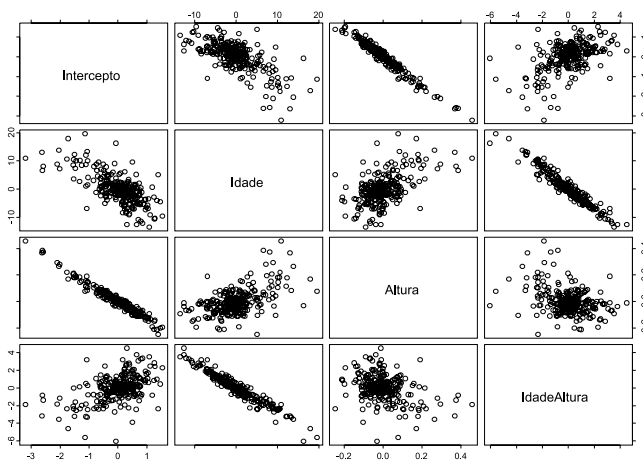


Figura 6 – Representação das correlações entre os efeitos aleatórios do segundo nível estimados no modelo em multiníveis.
Figure 6 – Correlations between the random effects of the second level estimated in the multilevel model.

As curvas estimadas para os dois níveis estão representadas nas Figuras 8 e 9. Para o nível 1, mesmo embora a curva passe pelo centro dos dados observados, a variabilidade em torno das linhas é relativamente grande e diferente entre unidades amostrais.

No nível 2, representado na Figura 9 e nas primeiras 25 árvores, é fácil verificar que o modelo em dois níveis gerou previsões que seguiram de perto os dados

observados, indicando que o modelo explica o crescimento da área basal com bastante acuracidade. Também, quando as inclinações das retas geradas nos níveis 1 e 2 são comparadas, a variação no padrão é muito intensa, tendo linhas paralelas e não-paralelas, com diferentes distâncias entre elas. Quanto maior a distância entre essas linhas de ajuste, pior vai ser a predição.

Tabela 2 – Informações estatísticas e logaritmo da máxima verossimilhança dos seis diferentes modelos em multiníveis
Table 2 – Statistical information and the log-likelihood for six different multilevel models

Modelo	gl	CIA	CIB	logMax	Teste	Razão	p-valor
1-Heteroscedástico/Positiva Definida							
Nível 1/Bloco Níveis 1 e 2	28	-20125	-19947	10090	-	-	-
2-Homoscedástico/Diagonal							
Níveis 1 e 2	13	-20496	-20413	10261	1vs2	340,56	<0,0001
3-Heteroscedástico/Positiva Definida Níveis 1 e 2	32	-20596	-20392	10330	2vs3	137,93	<0,0001
4-Heteroscedástico/Diagonal Níveis 1,2	20	-20685	-20557	10362	3vs4	64,99	<0,0001
5-Heteroscedástico/Positiva Definida Nível 1	26	-20707	-20541	10379	4vs5	30,08	<0,0001
6-Heteroscedástico/Positiva Definida Nível 1/AR(1)	27	-21851	-21680	10952	5vs6	1146,70	<0,0001

Em que: gl=graus de liberdade; CIA=Critério de Informação de Akaike; CIB=Critério de Informação Baysiano; logMax=logaritmo da Máxima Verossimilhança; Teste=Teste da Razão da Máxima Verossimilhança entre os modelos a *versus* b; Razão=Valor da Razão no teste; e p-valor=valor da probabilidade para o teste qui-quadrado.

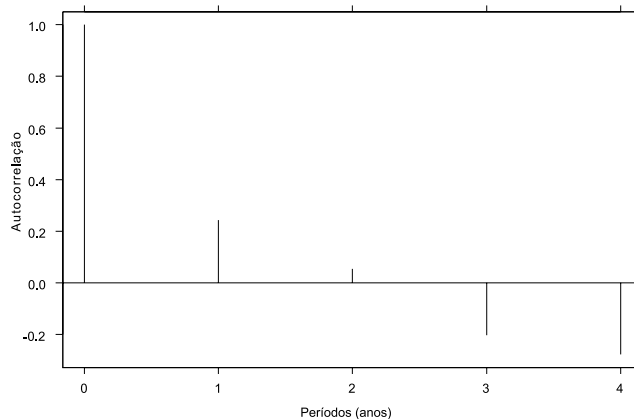


Figura 7 – Autocorrelação empírica correspondendo aos resíduos normalizados com quatro períodos (*lags*).
Figure 7 – Empiric autocorrelation representing the normalized residuals with four lags.

Tabela 3 – Estimativas dos parâmetros e o erro-padrão residual (EPR) do efeito fixo dos diferentes modelos
Table 3 – Parameter estimates and the Residual Standard Error (EPR) for the fixed effect of the different models

Modelo	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	EPR
1-Heteroscedástico/Positiva Definida	-4,154230	-20,675400	-2,249000	5,387987	0,017860
Nível 1/Bloco Níveis 1 e 2					
2-Homoscedástico/Diagonal	-4,115100	-19,341600	-0,044200	5,548100	0,013350
Níveis 1 e 2					
3-Heteroscedástico/Positiva Definida Níveis 1 e 2	-3,114829	-18,806650	-3,382550	5,440016	0,022675
4-Heteroscedástico/Diagonal Níveis 1,2	-3,176540	-19,549020	-3,028420	5,108540	0,022377
5-Heteroscedástico/Positiva Definida Nível 1	-3,907652	-17,328712	-2,873160	5,569021	0,020176
6-Heteroscedástico/Positiva Definida Nível 1/AR(1)	-3,365167	-15,982240	-0,259027	4,535554	0,016745

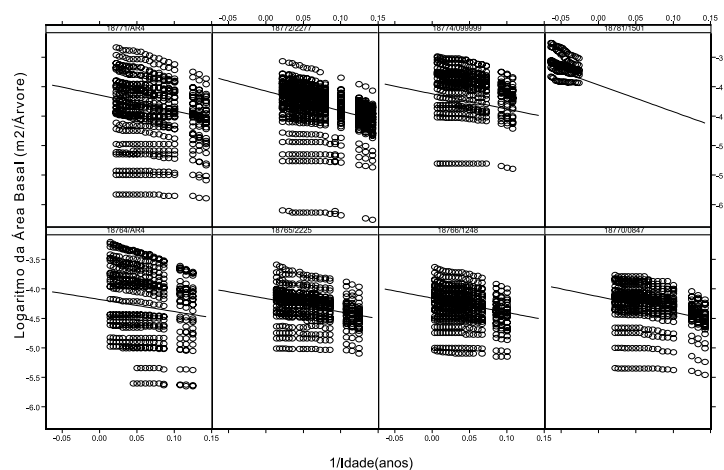


Figura 8 – Valores de $\ln(g)$ estimados para o primeiro nível (unidade amostral).
Figure 8 – $\ln(BA)$ values estimated for the first level (sample unity).

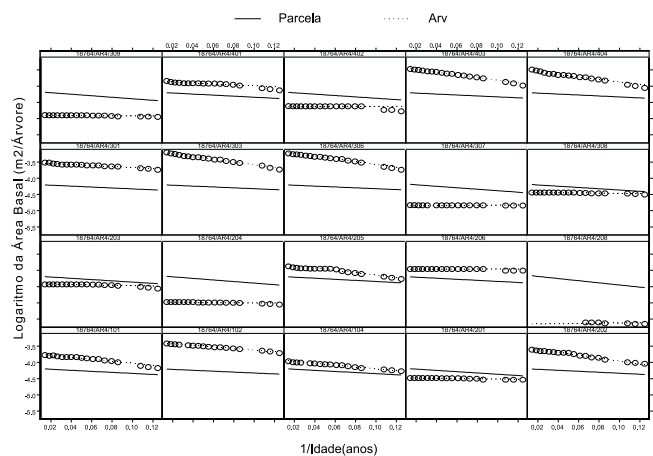


Figura 9 – Valores estimados de $\ln(g)$ para o primeiro e segundo níveis (20 árvores).
Figure 9 – $\ln(g)$ estimated values for first and second levels (20 trees).

4. CONCLUSÃO

Com base nos resultados, pode-se concluir que:

Na modelagem da área basal, o modelo em multinível, incluindo efeitos aleatórios, apresentou erro-padrão residual 15 vezes menor quando comparado com o modelo de efeito fixo.

A modelagem da heterogeneidade da variância nos níveis 1 e 2 melhorou os critérios estatísticos, elevando significativamente o valor do logaritmo da máxima verossimilhança.

Os critérios estatísticos também foram melhorados na modelagem da autocorrelação, resultando significativo o teste da razão da máxima verossimilhança.

As estimativas da área basal no segundo nível (árvore) foram extremamente precisas, com os valores estimados praticamente iguais aos valores observados.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. **Time series analysis: forecasting and control**, 3. ed. San Francisco: Holden-Day, 1994. 500p.
- CALEGARIO, N. **Modeling Eucalyptus stand growth based on linear and nonlinear mixed-effects models**. 2002. 109 p. Dissertation (PhD in Forest Science). University of Georgia, Athens, 2002.
- CARROLL, R.J.; RUPPERT, D. **Transformation and Weighting in Regression**. London: Chapman and Hall, 1988. 310p.
- DANIELS, R. F.; BURKHART, H. E. An integrated system of forest stand models. **Foresty Ecology Management**, v.23, p.159-177, 1988.
- DAVIDIAN, M.; GILTINAN, D. M. **Nonlinear models for repeated measurement data**. London: Chapman and Hall, 1995. 359p.
- FANG, Z. **A simultaneous system of linear and nonlinear mixed effects models for forest growth and yield prediction**. 1999. 171f. Dissertation. (PhD in Forest Science). University of Georgia. Athens, 1999.
- GOLDSTEIN, H. **Multilevel models in education and social research**. Oxford: Oxford University Press, 1987. 265 p.
- GOLDSTEIN, H. **Multilevel statistical models**. New York: Halstead Press, 1995. 325 p.
- JAYARAMAN, K.; LAPPI, J. Estimation of height-diameter curves through multilevel models with special reference to even-aged teak stands. **Forest Ecology and Management**, v.142, n.1-3, p.155-162, 2001.
- KARLSSON, M.O.; SHEINER, L.B. The importance of modeling inter-occasion variation in population pharmacokinetics. **Journal of Pharmacokinetics and Biopharmaceutics**, v.22, p.735-750, 1994.
- LINDSTROM, M.J.; BATES, D.M. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. **Journal of the American Statistical Association**, n.83, p.1014-1022, 1988.
- LITTELL, R.C. et al. **SAS System for mixed models**. Cary: SAS Institute, 1996. 633 p.
- LONGFORD, N.T. **Random coefficient models**. Oxford: Oxford University Press, 1993. 317p.
- SAKAMOTO, Y. et al. **Akaike Information Criterion Statistics**. Dordrecht: Reidel, 1986. 250 p.
- SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v.6 p.461-464. 1978.
- STRAM, D.O.; LEE, J.W. Variance components testing in the longitudinal mixed-effects models. **Biometrics**, v.50, p.1171-1177. 1994.
- VONESH, E.F.; CHINCHILLI, V.M. **Linear and nonlinear models for the analysis of repeated measurements**. New York: Marcel Dekker, 1997. 412 p.