

**ESTIMATIVA DO CRESCIMENTO DE POVOAMENTOS DE *Eucalyptus* BASEADA NA TEORIA DOS MODELOS NÃO LINEARES EM MULTINÍVEL DE EFEITO MISTO****GROWTH ESTIMATE OF *Eucalyptus* STANDS BASED ON NONLINEAR MULTILEVEL MIXED-EFFECTS MODEL THEORY**Natalino Calegario<sup>1</sup> Romualdo Maestri<sup>2</sup> Cristina L. Leal<sup>3</sup> Richard F. Daniels<sup>4</sup>**RESUMO**

O presente estudo foi baseado na utilização da teoria dos modelos não-lineares de efeito misto em multiníveis para a modelagem do crescimento em altura de povoamentos clonais de *Eucalyptus*. A base de dados utilizada representa medições de árvores individuais, tomadas em diferentes locais e no tempo, sendo a mesma classificada de longitudinal, irregularmente espaçada, não-balanceada, com autocorrelação e com a presença de heterogeneidade de variância. O modelo logístico de três parâmetros foi utilizado para a estimativa do crescimento em altura. Os parâmetros do modelo foram estimados como de efeitos fixos e aleatórios em dois níveis: unidade amostral (nível 1) e árvore individual dentro das unidades amostrais (nível 2). A inclusão de tais níveis reduziu sensivelmente o erro padrão da estimativa para o modelo. As estimativas também foram melhoradas com a modelagem da heterogeneidade da variância, utilizando estruturas diagonais, e da autocorrelação, com a estrutura ARMA(2,1).

**Palavras-chave:** modelos mistos em multiníveis; crescimento de *Eucalyptus*; heterogeneidade de variância; autocorrelação.

**ABSTRACT**

This study was based on the application of the nonlinear multilevel mixed-effects theory in modeling the height growth of *Eucalyptus* plantation. The database was from individual tree measurements, taken from different sites and over time. This type of database is considered as longitudinal, irregularly spaced, unbalanced, with autocorrelation and heteroscedasticity. The tree-parameter logistic model was used to estimate the height growth with fixed and random effects in two levels: sample units (level 1) and trees inside sample units (level 2). By including both levels, the standard error of estimate was reduced significantly. Also, the estimates were improved by modeling the variance heterogeneity and the autocorrelation, using the ARMA(2,1) structure.

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

**INTRODUÇÃO**

Os modelos não-lineares em multiníveis de efeito misto vêm sendo utilizados com bastante frequência a partir da última década, para a modelagem de diversas situações, em diferentes áreas do conhecimento. Tal metodologia, que considera a presença de efeitos fixos e aleatórios na base de dados, se mostra bastante flexível e de maior precisão, quando comparada com métodos que utilizam apenas os efeitos fixos nos modelos.

Na ciência florestal, tais modelos ainda são de uso restrito. Algumas exceções são os trabalhos de Fang(1999) que modelaram o crescimento de *Pinus* sp. após a aplicação de diferentes tratamentos silviculturais, Gaber e Maguire (2003) que modelaram o perfil longitudinal de três espécies florestais e Gregoire e Schabenberger (1994), que utilizaram tal método na estimativa volumétrica para árvores individuais. Nenhum dos trabalhos citados utilizou mais de um nível aleatório e, no crescimento de povoamentos de *Eucalyptus*, não existem informações ainda do uso da citada metodologia.

1. Engenheiro Florestal, PhD., Professor Adjunto, Departamento de Ciências Florestais, Universidade Federal de Lavras, CEP: 37200-000, Lavras (MG). calegari@ufla.br.
2. Engenheiro Florestal, Dr., Técnico da Empresa Aracruz Celulose S.A., Av. Brigadeiro Faria Lima, 2277, 3º e 4º andares, Ed. Plaza Iguatemi, Bairro Jardim Paulistano, CEP: 01452-200, São Paulo (SP). aestri@aracruz.br.
3. Economista, PhD., Departamento de Administração e Economia, Universidade Federal de Lavras, CEP: 37200-000, Lavras (MG). cleal@ufla.br.
4. Bachelor of Science in Forest Resources, PhD., Professor of University of Georgia, Athens, 30606, Georgia, USA. ddaniels@smokey.forestry.uga.edu.

Recebido para publicação em 3/09/2003 e aceito em 16/08/2005.

O principal objetivo do presente trabalho é a modelagem do crescimento em altura de povoamentos de *Eucalyptus* baseada na metodologia em multinível de efeito misto, com modelagem da heterogeneidade da variância e da autocorrelação serial. Espera-se, com tal modelagem, uma melhoria significativa na qualidade do ajuste.

## MATERIAL E MÉTODOS

### Dados

A base de dados é proveniente de povoamentos clonais de *Eucalyptus*, localizados na costa leste brasileira, nos estados do Espírito Santo e Bahia, de propriedade da Aracruz Celulose S.A. Foram registradas informações longitudinais por árvore individual, conforme Figura 1, com um total de 4289 observações, representando 254 árvores e 7 unidades amostrais. As medições foram realizadas mensalmente, com a idade dos povoamentos variando de 1,8 até 5 anos. Portanto, trata-se de uma base de dados longitudinais, irregularmente espaçada, não-balanceada, com presença de autocorrelação e, possivelmente, heterogeneidade de variância.

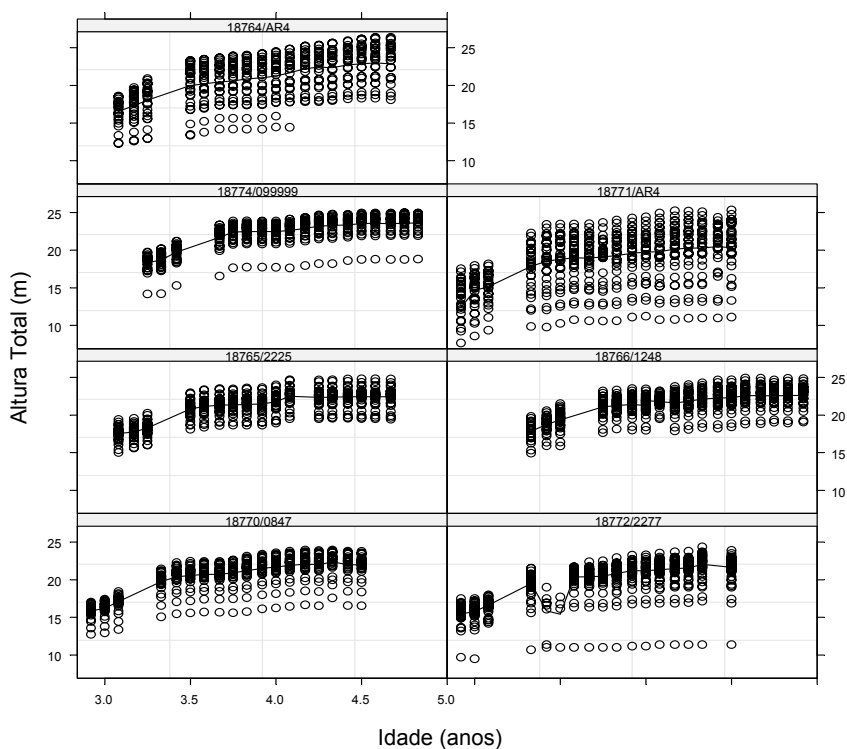


FIGURA 1: Crescimento em altura total para árvore individual representando diferentes combinações de unidades amostrais e clones.

FIGURE 1: Total height growth for individual tree representing different combination of unit samples and clones.

### Modelos generalizados não-lineares em multiníveis de efeito misto

Para a representação da forma generalizada dos modelos em multiníveis, foram seguidas as expressões apresentadas por Davidian e Giltinan (1995) e Lindstrom e Bates (1990) como estruturas básicas para esse tipo de modelagem, com algumas adaptações para o caso específico do crescimento florestal. Será considerada a variável resposta  $y_{ijk}$  representando os grupos aleatórios  $i$  e  $j$  ( $i$ -ésima unidade amostral e  $j$ -ésima árvores em nosso caso) medidas no tempo  $k$ . Então,  $i = 1, \dots, m$ ,  $j = 1, \dots, n_i$  e  $k = 1, \dots, p_{ij}$ , em que  $m$  é o número total de unidades amostrais,  $n_i$  é o número de árvores dentro da  $i$ -ésima unidade amostral e  $p_{ij}$  é o número de remedições da  $j$ -ésima árvore dentro da  $i$ -ésima unidade amostral. Normalmente, na ciência

florestal, esse tipo de base de dados é classificado como não-balanceado, ou seja, o número de árvores entre as unidades amostrais são diferentes e, com menor frequência, o número de medidas repetidas em cada árvore também é diferente. O modelo  $y_{ijk} = f(\phi_{ij}, \nu_{ijk}) + \varepsilon_{ijk}$  é uma generalização para representar a relação não-linear entre a variável resposta e os covariantes dentro dos diferentes grupos. A função  $f$  é geral de valores reais, diferenciável para um grupo específico do vetor de parâmetros  $\phi_{ij}$ , um vetor de covariantes  $\nu_{ijk}$  e  $\varepsilon_{ijk}$  são os erros dentro dos grupos, com distribuição normal. A função  $f$  deve ser não-linear em pelo menos um componente do vetor de parâmetro  $\phi_{ij}$  para um grupo específico, os quais tem a seguinte forma:

$$\phi_{ij} = A_{ijk} \beta + B_{i,jk} b_i + B_{ij,k} b_{ij}, \quad b_i \approx N(0, \Psi_1) \text{ e } b_{ij} \approx N(0, \Psi_2), \tag{1}$$

em que  $\beta$  é um vetor ( $p \times 1$ ) de efeitos fixos e  $b_i$  é um vetor ( $q_1 \times 1$ ) de efeitos aleatórios associados com o  $i$ -th grupo;  $b_{ij}$  é um vetor ( $q_2 \times 1$ ) associado com o segundo nível de efeitos aleatórios e pressuposto ser independente dos efeitos aleatórios do primeiro nível e  $A_{ijk}$  e  $B_{ijk}$  são matrizes de incidência. Como pressuposições básicas, os erros dentro dos grupos são independentemente distribuídos com média zero e variância  $\sigma^2$  e são independentes dos efeitos aleatórios.

A função logística com três parâmetros fixos será utilizada como exemplo de notação, tendo  $y$  como variável resposta (ou dependente) e  $x$  como regressor (ou independente). Supondo que o modelo final tenha três efeitos aleatórios e três efeitos fixos, o modelo e as matrizes da expressão  $\phi_{ij} = A_{ijk} \beta + B_{i,jk} b_i + B_{ijk} b_{ij}$  terão o seguinte formato:

$$\left. \begin{aligned} y_{ijk} &= \frac{\phi_{1ij}}{1 + \exp[-(x_{ijk} - \phi_{2ij}) / \phi_{3ij}]} + \varepsilon_{ijk}, \\ \begin{bmatrix} \phi_{1ij} \\ \phi_{2ij} \\ \phi_{3ij} \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_{1ij} \\ b_{2ij} \\ b_{3ij} \end{bmatrix}, \\ b_i &\approx N \left( 0, \begin{bmatrix} \psi_{111} & \psi_{112} & \psi_{113} \\ \psi_{112} & \psi_{122} & \psi_{123} \\ \psi_{131} & \psi_{132} & \psi_{133} \end{bmatrix} \right), \\ b_{ij} &\approx N \left( 0, \begin{bmatrix} \psi_{211} & \psi_{212} & \psi_{213} \\ \psi_{212} & \psi_{222} & \psi_{223} \\ \psi_{231} & \psi_{232} & \psi_{233} \end{bmatrix} \right), \\ \varepsilon_{ijk} &\approx N(0, \sigma^2). \end{aligned} \right\} \tag{2}$$

O vetor  $\beta$  representa os efeitos fixos,  $b_{pi}$  os efeitos aleatórios para o nível 1 e  $b_{pij}$  os efeitos aleatórios para o nível 2. No modelo logístico, as interpretações dos parâmetros são as seguintes:  $\Phi_{1ij}$  é a assíntota horizontal superior quando a idade tende para o infinito;  $\Phi_{2ij}$  é a idade na qual a resposta é  $\Phi_{1ij} / 2$ , representando o ponto de inflexão da curva sigmóide;  $\Phi_{3ij}$  representa o parâmetro escala, o qual é o valor da distância no eixo  $x$  entre o ponto de inflexão e o ponto onde a resposta é, aproximadamente,  $0,73 \Phi_{1ij}$ . Como pressuposições básicas, os efeitos aleatórios têm distribuição normal e são independentes para diferentes grupos e os erros são normais, independentes e identicamente distribuídos e são independentes dos efeitos aleatórios. Será verificado se as matrizes  $\Psi_1$  e  $\Psi_2$  possuem distribuição normal e são diagonais, isto é, se os efeitos aleatórios são independentes. Extensões para mais de dois níveis seguem o mesmo procedimento.

A precisão do modelo pode ser avaliada por critérios de informações estatísticas. Neste trabalho, foram utilizados dois critérios: Critério de Informação de Akaike (AIC) (Sakamoto *et al.*, 1986) e Critério de Informação Bayesian (BIC) (Schwarz, 1978). Esses critérios são estimados com as seguintes expressões:

$$AIC = -2\log(MV) + 2n_{\text{par}} \quad (3)$$

$$BIC = -2\log(MV) + n_{\text{par}} \log(N) \quad (4)$$

Em que: MV = valor da máxima verossimilhança;  $n_{\text{par}}$  = número de parâmetros no modelo. Menores valores para ambos os critérios implicam em melhor ajuste. Desde que esses critérios são conservativos (Stram and Lee, 1994), gerando maiores valores-p do que deveriam, é aconselhável usar-se um valor de  $\alpha$  de 10% para se selecionar o melhor modelo.

## RESULTADOS E DISCUSSÃO

### Ajuste do modelo não linear de efeito misto em dois níveis

Na Figura 1, pode-se notar que existe uma tendência não-linear evidente no crescimento em altura e possíveis diferentes parâmetros entre unidades amostrais e entre árvores dentro das unidades amostrais. A função logística, apresentada em (2), foi utilizada na modelagem do citado crescimento por ter interpretação biológica dos parâmetros e gerar excelentes ajustes em tal situação. A variável dependente  $y_{ijk}$  representa a altura total, em metros, para a  $i$ -ésima unidade amostral,  $j$ -ésima árvore e no tempo  $k$ . A variável independente  $x_{ijk}$  representa a idade, em anos, para a  $i$ -ésima unidade amostral,  $j$ -ésima árvore e tempo  $k$ .

O modelo não-linear em dois níveis ajustado está representado na Tabela 1. Considerou-se uma estrutura diagonal para a matriz de variância-covariância, pressupondo que os efeitos aleatórios nos dois níveis são independentes. Para os efeitos fixos, pode-se observar que os valores das probabilidades (valor-p) foram significativos para os três parâmetros, confirmando a importância deles no modelo.

TABELA 1: Parâmetros estimados para os efeitos fixos e erros padrões para os efeitos aleatórios.  
TABLE 1: Parameter estimated for the fixed effects and standard errors for the random effects.

| Parâmetro Estimado | Efeitos Fixos |             |      |         |         | Erro Padrão dos Efeitos Aleatórios |                       |
|--------------------|---------------|-------------|------|---------|---------|------------------------------------|-----------------------|
|                    | Valor         | Erro Padrão | G.L. | Valor-t | Valor-p | UA                                 | AUA                   |
| $\Phi_1$           | 21,69225      | 0,509483    | 684  | 42,57   | <0,0001 | 0,22063                            | 3,28324               |
| $\Phi_2$           | 2,67183       | 0,033037    | 684  | 80,87   | <0,0001 | 0,08285                            | 0,04256               |
| $\Phi_3$           | 0,37902       | 0,012367    | 684  | 30,64   | <0,0001 | 0,02788                            | $6,82 \times 10^{-7}$ |
| Resíduos           | -             | -           | -    | -       | -       | -                                  | 0,29723               |

Em que: UA = Unidade amostral; AUA = Árvores dentro das unidades amostrais.

O valor relativamente pequeno do erro padrão do efeito aleatório  $\Phi_3$  para o segundo nível pode ser uma evidência de que tal parâmetro pode ser retirado do modelo como efeito aleatório. Isso será verificado posteriormente.

Inicialmente, foi checada a correlação entre os efeitos aleatórios. A Figura 2 mostra uma correlação positiva entre o parâmetro de resposta média com o parâmetro escala, para o nível 2. Para contabilizar tal correlação, a matriz de variância-covariância para o segundo nível foi modelada como matriz bloco diagonal, tendo o parâmetro assíntota como primeiro bloco e os parâmetros resposta média e escala como o segundo bloco. Os resultados são apresentados na Tabela 2.

O ganho na modelagem da variância-covariância como bloco diagonal para o segundo nível não foi significativa (p-valor = 0.1067), indicando que a estrutura diagonal é a melhor opção.

Na seqüência, a heterogeneidade da variância foi avaliada. A primeira importante observação é a distribuição dos resíduos na Figura 3. Para cada unidade amostral, os resíduos estão aproximadamente distribuídos em torno de zero, não indicando violações de homocedasticidade para este nível aleatório. Entretanto, comparando as distribuições entre unidades amostrais, o padrão de distribuição se apresenta consideravelmente variável, com unidades amostrais apresentando maior variabilidade quando comparadas com outras. Baseado nesse padrão, a estratégia foi modelar a variância dentro das unidades amostrais (modelo heteroscedástico).

TABELA 2: Informações estatísticas e valores da máxima verossimilhança para seis diferentes modelos.

TABLE 2: Statistic information and likelihood values for six different models.

| Modelo                             | GL | AIC | BIC | logLik | Teste | Razãolog | Valor-p |
|------------------------------------|----|-----|-----|--------|-------|----------|---------|
| Homoscedástico/Diagonal Níveis 1,2 | 10 | 746 | 791 | -363   | -     | -        | -       |
| Homosc./Diag.Niv.1 e BlocoNiv.2    | 11 | 745 | 795 | -361   | 1vs2  | 2,602    | <0,1067 |
| Heteroscedástico/Diag. Nível 1 e 2 | 16 | 721 | 794 | -344   | 2vs3  | 34,086   | <0,0001 |
| Heter./Diag. Nível 1,2/AR(1)       | 17 | 562 | 640 | -264   | 3vs4  | 161,178  | <0,0001 |
| Heter./Diag. Nível 1,2/ ARMA(1,1)  | 18 | 550 | 632 | -257   | 4vs5  | 13,967   | <0,0002 |
| Heter./Diag. Nível 1,2 / ARMA(2,1) | 19 | 547 | 634 | -254   | 5vs6  | 4,819    | <0,0281 |

Em que: GL = graus de liberdade; AIC = critério de informação de Akaike; BIC = critério de informação Bayesiano; logLik = logaritmo da máxima verossimilhança; Razãolog = razão entre as máximas verossimilhanças dos modelos testados; Valor-p = valor da probabilidade acima do valor encontrado no teste da razão para a distribuição qui-quadrado; AR(1) = estrutura autoregressiva de ordem 1; ARMA(1,1) = estrutura autoregressiva e de média móvel, ambas com ordem 1.

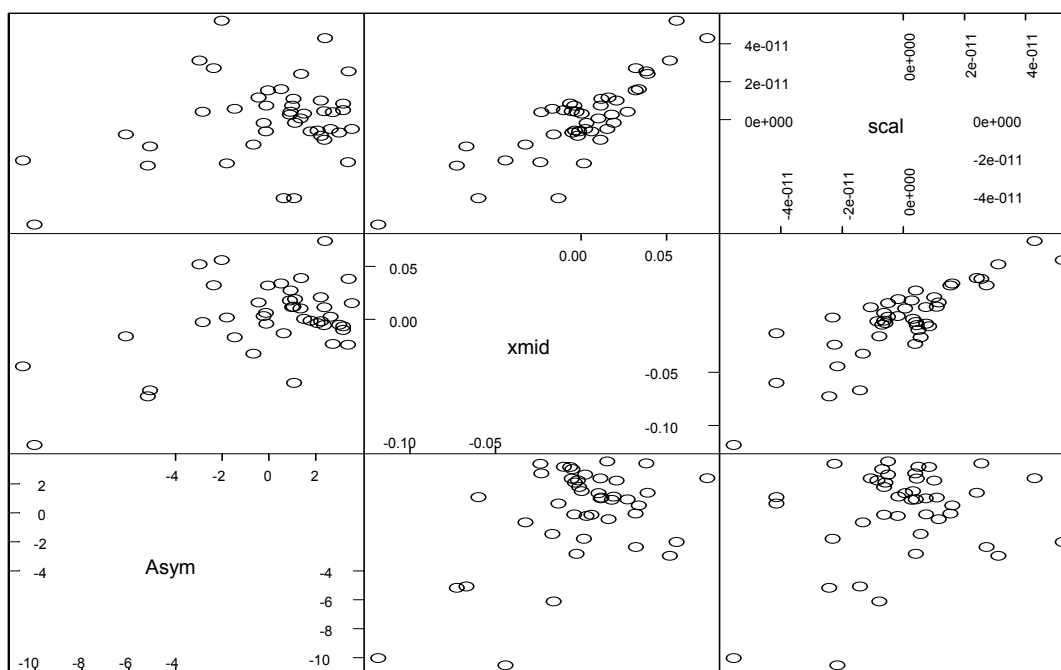


FIGURA 2: Correlações entre os efeitos aleatórios, em que Asym se refere ao parâmetro assintota, xmid é o parâmetro do ponto de inflexão e scal é o parâmetro escala.

FIGURE 2: Pairs plot showing the correlation between the random effects, where Asym is the asymptote parameter, xmid is the inflection point parameter, and scal is the scale parameter.

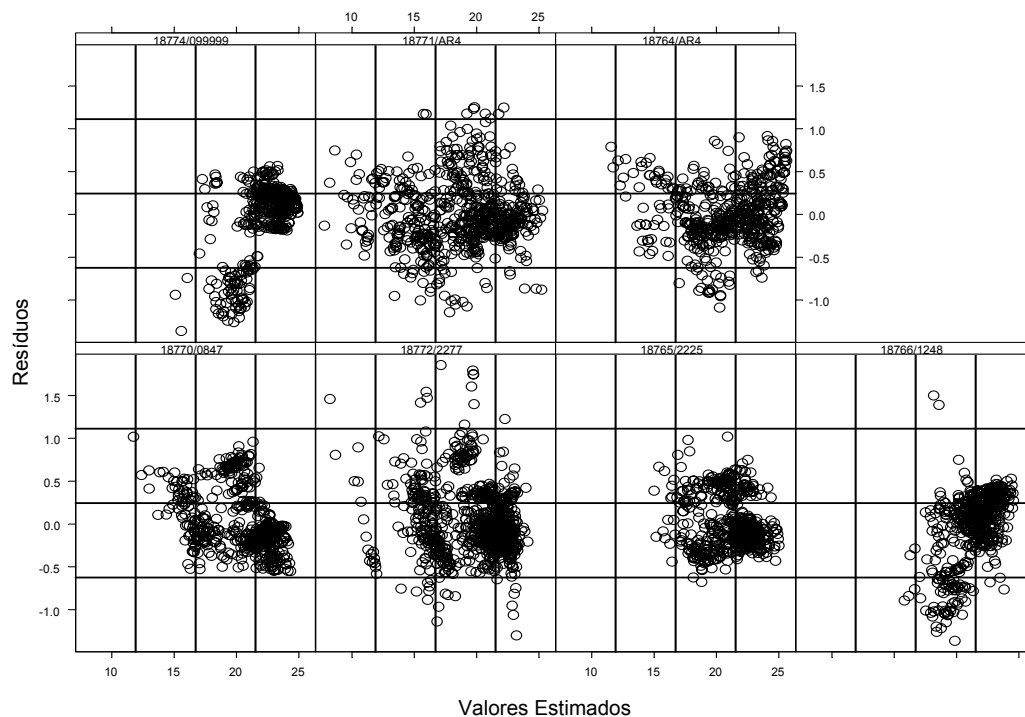


FIGURA 3: Distribuição dos resíduos padronizados por unidade amostral.

FIGURE 3: Standard residual distribution by sample plot.

A Tabela 2 mostra a diferença entre os dois padrões de modelagem. Os menores valores dos Critérios de Informação de Akaike (AIC) e Bayesiano (BIC), e o maior valor do logaritmo da máxima verossimilhança, indicam que o modelo heteroscedástico explica melhor os dados do que o modelo homoscedástico. Isso pode ser confirmado com o valor reduzido do valor-p ( $<0,0001$ ).

Após a definição de uma estrutura diagonal para a matriz de variância-covariância e assumindo o modelo como heteroscedástico, o próximo passo foi a modelagem da autocorrelação. Considerando se tratar de dados longitudinais de medidas repetidas, espera-se uma melhoria dos parâmetros pela modelagem da correlação serial. Por conseguinte, vários modelos foram testados, com variações de autoregressivos (AR(p)), média móvel (MA(q)) e a combinação entre eles (ARMA(p,q)). Os resultados estão representados na Tabela 2. Maior valor do logaritmo da máxima verossimilhança para o modelo ARMA(2,1), gerando um valor do teste da razão das máximas verossimilhanças de 4,81, quando comparada com o modelo ARMA(1,1), e um significativo valor-p de 0,0281, indicam que o modelo é preferido. Para posteriores análises, o modelo ARMA(2,1) será utilizado.

O acesso geral da capacidade do modelo (2) pode ser verificado pela representação gráfica dos valores estimados para ambos níveis versus os valores observados (Figura 4).

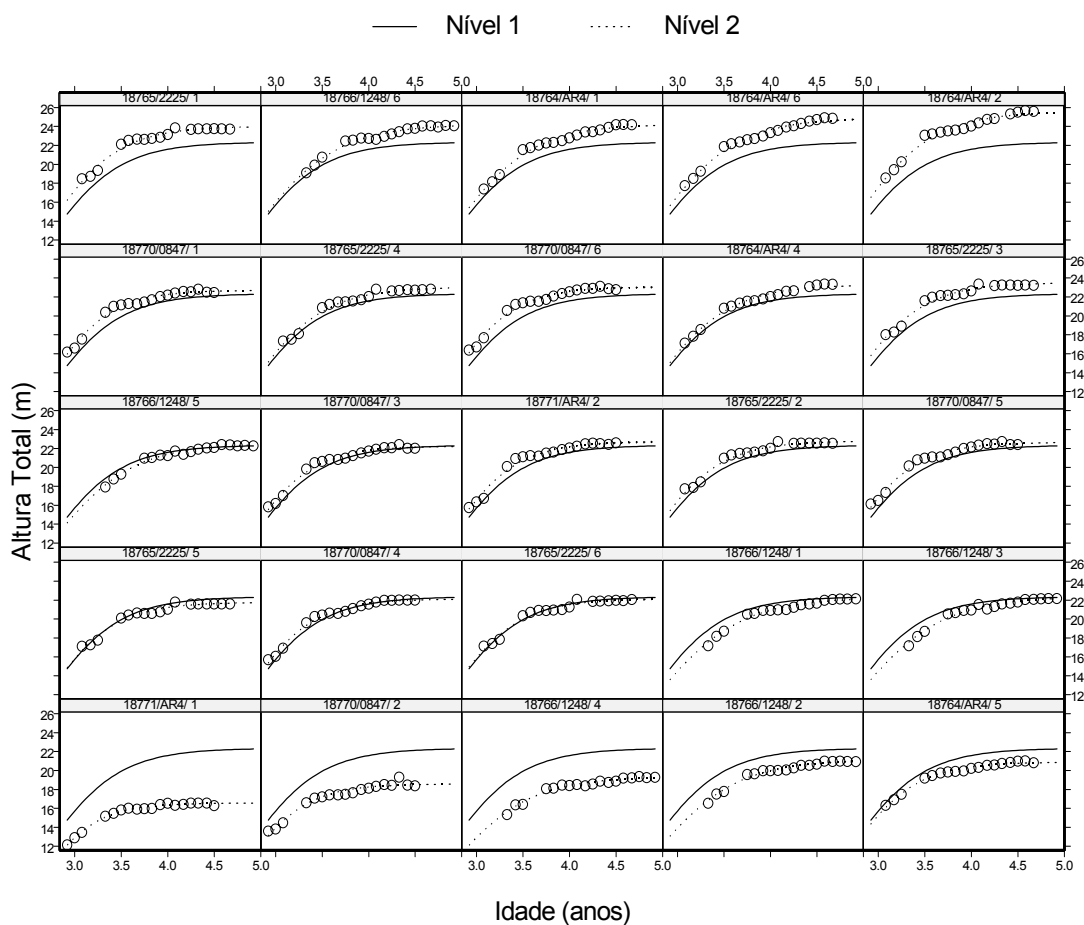


FIGURA 4: Valores estimados para altura total (h), em metros, por unidade amostral (nível 1) e por árvore dentro das unidades amostrais (nível 2), para 25 árvores com exemplo.

FIGURE 4: Estimates for total height (h), in meters, by sample unit (level 1) and by tree inside the sample unit (level 2), representing an example of 25 trees.

Para um melhor entendimento, o quadro do lado inferior esquerdo da Figura 4, cujo código é 18771/AR4/1, representa a parcela de número 18771, clone AR4 e árvore número 1. A figura mostra que o modelo heteroscedástico, com correlação diagonal entre os efeitos aleatórios, para ambos os níveis e com estrutura de correlação autoregressiva e de média móvel, descreve os dados de forma precisa. Adicionalmente, pode-se observar que o modelo em nível de árvore individual tem a habilidade de representar as variações entre árvores com incrível flexibilidade, mesmo em situações mais extremas quando uma específica árvore se afasta muito da média do padrão de crescimento.

## CONCLUSÕES

Com a modelagem do crescimento em altura, utilizando a função logística de três parâmetros, verificou-se a necessidade de se incluir os efeitos fixos e os aleatórios nas estimativas. Com isso, houve redução significativa do erro padrão residual. Além da utilização da teoria dos modelos mistos não-lineares em multinível, o sucesso da modelagem do crescimento se deu também por causa do uso da função logística, a qual possui parâmetros com interpretações biológicas.

Associado à inclusão dos efeitos fixos e aleatórios na modelagem, o processo de estimativa foi melhorado significativamente com a modelagem da heterogeneidade da variância e da autocorrelação. Como são dados biológicos longitudinais, tais características eram de ser esperadas. Isso foi confirmado graficamente. Com tal modelagem, os valores da máxima verossimilhança foram elevados significativamente, gerando testes significativos para a razão do citado valor para diferentes modelos. O

modelo final selecionado possui três fatores fixos, três aleatórios, com modelagem da heterogeneidade da variância e com uma estrutura de autocorrelação ARMA(2,1).

Tal metodologia proposta pode ser utilizada em um grande número de aplicações na ciência florestal, em que a modelagem estiver presente. Na biometria e no manejo florestal a aplicação do citado método tem um potencial bastante promissor, sobretudo nos sistemas de predição e projeção para fins de planejamento florestal.

#### REFERÊNCIA BIBLIOGRÁFICA

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*. Athens. UGA. 1999. 177p. Dissertation (PhD) – University of Georgia, 1997.

GARBER, S. M.; MAGUIRE, D. A. Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. **Forest Ecology and Management**, v. 179, p. 507-522, 2003.

GREGOIRE, T. G.; SCHABENBERGER, O. Fitting bole-volume equations to spatially correlated within-tree data. In: ANNUAL CONFERENCE ON APPLIED STATISTICS IN AGRICULTURE, 6., 1994, Manhattan, **Proceedings** ...Kansas, 1994. p. 120-133.

LINDSTROM, J. M.; BATES, D. M. Nonlinear mixed effects models for repeated measures data. **Biometrics**, v. 46, p. 673-687, 1990.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. **Akaike Information Criterion Statistics**. Dordrecht, Holland :Kluwer Academic Publishers, 1986, 256p.

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.