DESENVOLVIMENTO DE UM MODELO LINEAR DE EFEITO MISTO NA ESTIMATIVA DO CRESCIMENTO E PRODUÇÃO DE POVOAMENTOS CLONAIS DE *Eucalyptus*.

Natalino Calegario[1], Richard F. Daniels[2], Romualdo Maestri[3], Rodolfo Neiva[4]

(Recebido: 28 de novembro de 2002; aceito: 19 de maio de 2004)

RESUMO: O enfoque principal do trabalho foi o desenvolvimento de um modelo linear de efeito misto para a estimativa do crescimento e da produção em área basal, para povoamentos clonais de *Eucalyptus grandis e Eucalyptus urophylla*. Utilizando uma base de dados de povoamentos clonais localizados na região costal brasileira, um modelo linear misto para área basal foi proposto. Após a modelagem da heterogeneidade da variância entre unidades amostrais e entre clones, verificou-se uma significativa melhoria dos parâmetros das informações estatísticas (CIA e CIB) e do logaritmo da máxima verossimilhança. Também, após a modelagem da autocorrelação, tais estatísticas tiveram melhoria significativa. Portanto, a modelagem, tanto da heteroscedasticidade quanto da autocorrelação, implicou em melhor performance do modelo linear.

Palavras-chave: crescimento de *Eucalyptus*, modelo linear misto, área basal, heterocedasticidade e autocorrelação.

THE DEVELOPMENT OF A LINEAR MIXED-EFFECT MODEL TO ESTIMATE GROWTH AND YIELD OF CLONAL *Eucalyptus* STANDS

***ABSTRACT**: The main purpose of this study was to develop a linear mixed-effects model to estimate the basal area growth and yield, for clonal Eucalytus stands. After modeling the variance among sample plots and clones, it was verified a significant improvement of the statistic information parameters (AIC and BIC) and the likelihood logarithm value. Also, after modeling both heteroscedasticity and autocorrelation, such statistic criteria had a significant improvement. Thus, the modeling process improved significantly the estimated parameters in the linear model.*

Key words: Eucalyptus growth, linear mixed-effect model, basal area, heteroscedasticity, autocorrelation

[1] Professor do Departamento de Ciências Florestais, Universidade Federal de Lavras, C.P. 3037, CEP 37200-000, Lavras-MG, calegari@ufla.br;
[2] Warnell School of Forest Resources, University of Georgia, Athens, 30606, Georgia, USA (ddaniels@smokey.forestry.uga.edu);
[3] Aracruz Celulose S/A, Rodovia Aracruz/Barra do Riacho, Km 25, CEP 29197-000, Aracruz-ES rmaestri@aracruz.com.br;
[4] Bahia Sul Celulose S/A, Rodovia BR 101, Km 880 – Jerusalém, CEP 45995-970, Teixeira de Freita-BA rodolfoneiva@bahiasul.com.br;

1 INTRODUÇÃO

Os modelos lineares de efeito misto têm sido utilizados recentemente na modelagem de processos longitudinais, espaciais e espaço-temporais, em vários campos científicos, tais como medicina (Verbeke & Lesaffre, 1997), biologia (Christman & Jernigan, 1997), engenharia (Pinheiro & Bates, 2000), agricultura (Littell et al., 1996) e outras.

Na ciencia florestal, Gregoire (1995) aplicou a teoria de modelos mistos para estimar o crescimento em área basal de *Pinus strobus* L. e *Pseudotsuga menziensii* (Mirb.) Franco, com sensíveis melhorias no ajuste quando comparada com modelos correntemente usados. Mais recentemente, Fang & Bailey (2001) aplicaram esta abordagem para a modelagem do crescimento em área basal de *Pinus elliottii* Engelm., após a aplicação de intensivos tratamentos silviculturais. Apesar da existência dos estudos em florestas, citados anteriormente, os autores não se preocuparam em desenvolver e mostrar a evolução da citada abordagem. Também, em povoamentos de *Eucalyptus*, não se têm informações do uso de tal metodologia na modelagem do crescimento e da produção.

Portanto, o principal objetivo deste estudo é o desenvolvimento de um modelo linear de efeito misto na estimativa de crescimento e da produção em área basal para povoamentos clonais de *Eucalyptus*.

2 MATÉRIAL E MÉTODOS

2.1 Modelo linear geral de efeito misto

A forma paramétrica geral apresentada aqui é baseada na de Laird & Ware (1982), citada por Davidian & Giltinan (1995), com algumas adaptações e mudanças para o estudo do crescimento e da produção florestal.

Supondo que m unidades amostrais foram alocadas de uma população florestal e que as mesmas foram medidas repetitivamente no tempo, por exemplo, t vezes. Se t é o mesmo para cada unidade, nós temos uma série balanceada de dados, podendo ser processada e analisada de uma forma relativamente simples, com t x m valores disponíveis. Porém, séries não balanceadas de dados são mais comuns em estudos de crescimento e de produção florestal. Então, o número de medidas repetidas no tempo irá variar e t_i representará o número de medidas tomadas para a i-ésima unidade amostral. Por exemplo, se a unidade i for medida anualmente durante j anos, o que é uma clássica situação em povoamentos de *eucalyptus*, o valor de $t_i=j$. No caso dos dados serem balanceados, $t_1=t_2=\ldots=t_m=j$. Considere que $\mathbf{y}_i$ representa o vetor de resposta para a i-ésima unidade amostral. Então, $\mathbf{y}_i$ tem dimenssão (j x 1) e esta sutuação pode ser modelada utilizando-se um modelo de efeito misto na sua forma linear (1).

$$y_i = X_i ß + Z_i b_i + e_i \qquad (1)$$

A variável resposta $\mathbf{y}_i$ será um vetor com dimensão (t_i x 1), a matriz $\mathbf{X}_i$ terá dimensão (t_i x p): t_i linhas com repetidas medidas e p colunas de covariantes ou variáveis independentes, incluindo a coluna de intercepto. O vetor **ß** terá dimensão (p x 1), representando p parâmetros de efeito fixo. Representando os efeitos aleatórios, Z_i será uma matriz de dimensão (t_i x k), conectando $\mathbf{y}_i$ com os efeitos aleatórios $\mathbf{b}_i$, o qual tem dimensão máxima (p x 1) e parâmetros estimados para a unidade amostral i. Na literatura baseada em modelos de efeito misto é comum representar os efeitos fixos com letras gregas e os efeitos aleatórios com letras latinas. Portanto, neste caso particular, a unidade amostral i terá p efeitos fixos (**ß**), incluindo o intercepto e até p efeitos aleatórios (**b**). Se tivermos um total de m unidades

amostrais, cada unidade terá os mesmos valores para os efeitos fixos e, possivelmente, diferentes valores para os efeitos aleatórios.

Em estudos de crescimento e produção florestal, uma clássica aplicação de modelo linear é relacionar o logaritmo da área basal (ln(G)), como variável resposta, com variáveis de povoamento, tais como o inverso da idade (1/idade), o logaritmo da altura dominante (ln(H)), o logaritmo do número de árvores (ln(N)) e as interações entre estes covariantes (4). Se uma unidade amostral *i*, por exemplo, é medida anualmente durante 5 anos, a variável dependente $\mathbf{y_i}$=ln(G) será um vetor de dimensão (5 x 1), a matriz $\mathbf{X_i}$ será uma matriz (5 x *p*) e o vetor **ß** terá dimensão (*p* x 1). O valor de *p* depende do número de covariantes que são significativos na equação mais o intercepto.

Nas presuposições padrões da análise de regressão, $e_i \sim N(0,\mathbf{S}_i)$, em que $\mathbf{S}_i$ representa a matriz de covariantes dentro da i-ésima unidade amostral. Se as observações são independentes, $\mathbf{S}_i = s^2 \mathbf{I}_{ti}$, em que $\mathbf{I}_{ti}$ representa uma matriz identidade de dimensão (t_i x t_i). Em nosso exemplo, se as 5 observações da mesma unidade amostral medidas em idades diferentes forem independentes e com mesma variância, $\mathbf{I}_{ti}$ será uma matriz (5 x 5) e $\mathbf{S}_i$ será uma matriz diagonal com s^2 . Na prática, $\mathbf{S}_i$ tem muitas variações. Condicional a $\mathbf{b}_i$, (1) resulta em

$$E(y_i|b_i) = \mathbf{X_i ß + Z_i b_i} \qquad (2a)$$

$$Cov(y_i|b_i) = \mathbf{S}_i \qquad (2b)$$

Se o vetor $\mathbf{b}_i$ de efeitos aleatórios tiver distribuição normal com média zero, matriz de dispersão **?** (*k* x *k*) e independentes entre eles e dos erros $\mathbf{e}_i$, a média marginal e a covariância de $\mathbf{y}_i$ será:

$$E(\mathbf{y}_i) = E\{E(\mathbf{y}_i|\mathbf{b}_i)\} = \mathbf{X_i ß} \qquad (3a)$$

$$Cov(\mathbf{y}_i) = E\{Cov(\mathbf{y}_i|\mathbf{b}_i)\} + Cov\{E(\mathbf{y}_i|\mathbf{b}_i)\} = \mathbf{S}_i + \mathbf{Z_i ?\ Z}^T_i = \mathbf{V}_i \qquad (3b)$$

Se o exemplo for expandido para dois grupos (clones=*c*), com 10 unidades amostrais (*i*) em cada grupo, 5 observações (*j*) em diferentes idades para cada unidade amostral e assumindo o seguinte modelo de área basal,

$$\ln(G)_{cij} = (b_{oc} + b_{oci}) + (b_{1c} + b_{1ci}) \times \frac{1}{Idade_{cij}} + (b_{2c} + b_{2ci}) \times \ln(H_{cij}) + (b_{3c} + b_{3ci}) \times \frac{1}{Idade_{cij}} \times \ln(H_{cij}) + e_{cij} \qquad (4)$$

com *c*=1…2; *i*=1…10; *j*=1…5, o modelo pode ser reestruturado da seguinte forma:

$$\mathbf{Y}^T = (\mathbf{Y}^T_{1,1}, \ldots, \mathbf{Y}^T_{1,10}, \mathbf{Y}^T_{2,1}, \ldots, \mathbf{Y}^T_{2,10})$$

$$\mathbf{Y}^T_{c,i} = (Y_{c,i,1}, Y_{c,i,2}, Y_{c,i,3}, Y_{c,i,4}, Y_{c,i,5})^T.$$

Então, $\mathbf{Y}^T$ terá dimensão de (1 x 100), em que **T** representa a transposição do vetor. E

$$\mathbf{X}^T = (\mathbf{X}^T_{1,1} | \ldots | \mathbf{X}^T_{1,10} | \mathbf{X}^T_{2,1} | \ldots | \mathbf{X}^T_{2,10}),$$

Em que,

$$\mathbf{X}_{1,i} = \begin{bmatrix} 1 & 0 & 1/Idade_{1,i,1} & 0 & \ln(H_{1,i,1}) & 0 & (1/Idade_{1,i,1})\ln(H_{1,i,1}) & 0 \\ 1 & 0 & 1/Idade_{1,i,2} & 0 & \ln(H_{1,i,2}) & 0 & (1/Idade_{1,i,2})\ln(H_{1,i,2}) & 0 \\ 1 & 0 & 1/Idade_{1,i,3} & 0 & \ln(H_{1,i,3}) & 0 & (1/Idade_{1,i,3})\ln(H_{1,i,3}) & 0 \\ 1 & 0 & 1/Idade_{1,i,4} & 0 & \ln(H_{1,i,4}) & 0 & (1/Idade_{1,i,4})\ln(H_{1,i,4}) & 0 \\ 1 & 0 & 1/Idade_{1,i,5} & 0 & \ln(H_{1,i,5}) & 0 & (1/Idade_{1,i,5})\ln(H_{1,i,5}) & 0 \end{bmatrix}$$

$$\mathbf{X}_{2,i} = \begin{bmatrix} 0 & 1 & 0 & 1/Idade_{2,i,1} & 0 & \ln(H_{2,i,1}) & 0 & (1/Idade_{2,i,1})\ln(H_{2,i,1}) \\ 0 & 1 & 0 & 1/Idade_{2,i,2} & 0 & \ln(H_{2,i,2}) & 0 & (1/Idade_{2,i,2})\ln(H_{2,i,2}) \\ 0 & 1 & 0 & 1/Idade_{2,i,3} & 0 & \ln(H_{2,i,3}) & 0 & (1/Idade_{2,i,3})\ln(H_{2,i,3}) \\ 0 & 1 & 0 & 1/Idade_{2,i,4} & 0 & \ln(H_{2,i,4}) & 0 & (1/Idade_{2,i,4})\ln(H_{2,i,4}) \\ 0 & 1 & 0 & 1/Idade_{2,i,5} & 0 & \ln(H_{2,i,5}) & 0 & (1/Idade_{2,i,5})\ln(H_{2,i,5}) \end{bmatrix}$$

$\boldsymbol{ß}^{\mathbf{T}} = (ß_{10}, ß_{20}, ß_{11}, ß_{21}, ß_{12}, ß_{22}, ß_{13}, ß_{23})^T$;

$\mathbf{Z} = I_{20} \otimes \mathbf{1}_{5,}$

em que I_n é a matriz identidade de ordem n, $\otimes$ representa o produto direto de Kronecker e $\mathbf{1}_n$ denota um vetor de (n x 1) com todas as entradas iguais a um.

Posteriormente,

$\mathbf{b} = (b_1, b_2, \ldots, b_{10}, b_{11}, \ldots, b_{20})^T$

e

$e^T = (e_{1,1,1}, e_{1,1,2}, e_{1,1,3}, \ldots, e_{2,10,5})^T$

Os termos dos erros são independentes com variância s^2 e matriz de variância-covariância **S.** Os efeitos aleatórios também são independentes com variância s^2_b, **?** representando a matriz de variância e covariância e **V** sendo a matriz de variância-covariância para a variável resposta **Y.** As seguintes representações são úteis para fins computacionais:

$\mathbf{?} = s^2_b I_{20}$,

$\mathbf{S} = s^2 I_{100}$,

$$\mathbf{V} = \mathbf{Z}\,\mathbf{?}\,\mathbf{Z}^T + \mathbf{S} = s^2_b(I_{20} \otimes J_5) + s^2(I_{20} \otimes I_5) = I_{20} \otimes (s^2_b J_5 + s^2 I_5)$$

2.2 Estimativa dos efeitos fixo e aleatório

Na seção anterior foi visto que os valores marginais para $\mathbf{y}_i$ tem distribuição normal com média $\mathbf{X_i ß}$ e matriz de variância-covariância $\mathbf{V_i} = \mathbf{S}_i + \mathbf{Z_i}\,\mathbf{?}\,\mathbf{Z}^T_i$. Seguindo Verbeck & Molenberghs (1997), suponha que **a** denote o

vetor de todos os components de variância e covariância em $\mathbf{V_i}$, isto é, **a** terá todos os diferentes elementos de **?** e todos os parâmetros de $\mathbf{S_i}$. Em nosso exemplo, para a i-ésima unidade amostral, **?** é representado por 10 diferentes parâmetros (4*(4+1)/2) e $\mathbf{S_i}$ por 5, considerando o mesmo como uma matriz diagonal. Sendo **? = (ß^T, a^T)^T** um vetor representando todos os parâmetros no modelo marginal, a abordagem clássica é minimizar a função marginal da verossimilhança com respeito a **?**.

$$\mathbf{L_{ML}}(q) = \prod_{i=1}^{N} \left\{ (2p)^{-ni/2} \mid \mathbf{V_i}(a) \mid^{-0.5} \times \exp\left(-\frac{1}{2} (\mathbf{Y_i} - \mathbf{X_i} b)^T \mathbf{V_i}^{-1}(a)(\mathbf{Y_i} - \mathbf{X_i} b) \right) \right\} \quad (5)$$

Se **a** é conhecido, o estimador de **ß** por máxima verossimilhança, obtido pela maximização de (5), é dado por

$$\hat{b} = \left(\sum_{i=1}^{N} \mathbf{X_i^T V_i^{-1}}(a) \mathbf{X_i} \right)^{-1} \sum_{i=1}^{N} \mathbf{X_i^T V_i^{-1}}(a) \mathbf{y_i} \quad (6)$$

e sua matriz de variância-covariância será

$$\text{var}(\hat{b}) = \left(\sum_{i=1}^{N} \mathbf{X_i^T V_i^{-1}}(a) \mathbf{X_i} \right)^{-1} \quad (7)$$

Quando **a** é desconhecido, mas temos uma estimativa disponível, $V_i^{-1}(\hat{a})$ pode substituir $V_i^{-1}(a)$. Para se estimar **a**, os métodos da máxima verossimilhança (MV) ou da máxima verossimilhança restrita (MVR) são usados. Detalhes sobre estes métodos podem ser encontrados em Searle et al. (1992), Davidian & Giltinan (1995) e Vonesh & Chinchilli (1997).

Usando análise de regressão, quando p=posto(**X**) = 4, o método da máxima verossimilhança gera menor erro médio quadrático para s^2 quando comparado com o método da máxima verossimilhança restrita. O contrário é verdadeiro se p>4 e (n-p) for relativamente grande (Verbeck & Molenberghs, 1997). Adicionalmente, a máxima verossimilhança restrita se ajusta melhor para a perda de graus de liberdade devido à estimativa dos efeitos fixos. A estimativa da MVR pode ser vista como uma estimativa dos componentes de variância baseados nos resíduos estimados somente após ajustados os efeitos fixos (Davidian & Giltinan, 1995).

Desde que os efeitos aleatórios são considerados variáveis aleatórias, é comum estimá-los com base em técnicas Baysianas. A distribuição marginal de $\mathbf{b_i}$ é normal multivariada com média zero e matriz de variância-covariância **? ,** e sua distribuição é referida como distribuição anterior de $\mathbf{b_i}$. Após os valores de $\mathbf{y_i}$ serem observados, a distribuição posterior de $\mathbf{b_i}$ pode ser estimada como:

$$\mathbf{f}(\mathbf{b_i} \mid \mathbf{y_i}) \equiv \mathbf{f}(\mathbf{b_i} \mid \mathbf{Y_i} = \mathbf{y_i}) = \frac{\mathbf{f}(\mathbf{y_i} \mid \mathbf{b_i}) \mathbf{f}(\mathbf{b_i})}{\int \mathbf{f}(\mathbf{y_i} \mid \mathbf{b_i}) \mathbf{f}(\mathbf{b_i}) \mathbf{db_i}} \quad (8)$$

A expressão (8) é a função de densidade de uma distribuição multivariada normal (Smith, 1973) e $\mathbf{b_i}$ é estimado pela média posterior de $\mathbf{b_i}$.

$$\hat{\mathbf{b}}_\mathbf{i}(q) = \mathbf{E}[\mathbf{b_i} \mid \mathbf{Y_i} = \mathbf{y_i}] = \int \mathbf{b_i} \mathbf{f}(\mathbf{b_i} \mathbf{y_i}) \mathbf{db_i} = \mathrm{Y} \mathbf{Z_i^{-1}} \mathbf{V_i^{-1}}(a)(\mathbf{y_i} - \mathbf{X_i} \hat{b}) \quad (9)$$

Esta estimativa é considerada como o melhor preditor linear não-tendencioso (BLUP) para $\mathbf{b_i}$ (Searle et al., 1992). A matriz de variância-covariância para $\mathbf{b_i}$ será:

$$\mathrm{cov}(\hat{\mathbf{b}}_i) = \mathrm{Y}\,\mathbf{Z}_i^{\mathrm{T}}\left\{\mathbf{W}_i - \mathbf{W}_i\mathbf{X}_i\left(\sum_{i=1}^{N}\mathbf{X}_i\mathbf{W}_i\mathbf{X}_i\right)^{-1}\mathbf{X}_i^{\mathrm{T}}\mathbf{W}_i\right\}\mathbf{Z}_i\,\mathrm{Y}, \quad (10)$$

e, para a verificação da variação da diferença entre os efeitos aleatórios estimados e observados (Laird & Ware, 1982), a seguinte expressão pode ser usada

$$\mathrm{cov}(\hat{\mathbf{b}}_i - \mathbf{b}_i) = \mathrm{Y} - \mathrm{var}(\hat{\mathbf{b}}_i)$$

2.3 Testes de hipóteses e intervalos de confiança

Neste caso, testes são úteis para a avaliação da precisão das estimativas e a significância dos termos no modelo. O primeiro teste discutido aqui é o teste da razão da máxima verossimilhança (TRMV). Embora chamado de teste da máxima verossimilhança, este teste pode também ser usado para comparar modelos aninhados ajustados por máxima verossimilhança restrita, mas os modelos têm que possuir os mesmos efeitos fixos (Pinheiro & Bates, 2000). Modelos aninhados ocorrem quando um modelo representa um caso especial de outro. Se L_2 representar o maior valor da máxima verossimilhança para um modelo mais geral e L_1 for o menor valor para um modelo restrito, o valor de TRMV será:

$$TRMV = 2\log(L_2 / L_1) = 2\,[\log(L_2) - \log(L_1)] \quad (11)$$

Desde que $L_2 > L_1$, o valor de TRMV será positivo e, se k_i é o número de parâmetros no i-ésimo modelo, a distribuição de TRMV será $?^2$ com (k_2-k_1) graus de liberdade. O valor de TRMV é comparado com o valor crítico de $?^2(k_2-k_1,a)$ e, se TRMV> $?^2(k_2-k_1,a)$, gerando valor-p (<0,05), o modelo mais geral é preferido quando comparado com o modelo restrito.

A precisão do modelo pode ser avaliada por técnicas de critérios de informações estatísticas. Estes 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). Estes critérios são avaliados como

$$CIA = -2\log(MV) + 2n_{par} \quad (12a)$$

$$CIB = -2\log(MV) + n_{par}\log(N) \quad (12b)$$

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

2.4 Funções de variâncias e estruturas de correlações

Com base nas pressuposições dos modelos de efeito misto, os erros dentro de um mesmo grupo são independentes e normalmente distribuídos, com média zero e variância s^2. Os efeitos aleatórios são também normalmente distribuídos com média zero e matriz de covariância **?**, e são independentes para diferentes grupos. Quando estas pressuposições são violadas, é necessário o uso de técnicas para se modelar a verdadeira estrutura dos dados.

A primeira técnica utilizada para a solução deste problema é modelar a estrutura da variância dos erros dentro dos grupos usando covariantes. Davidian & Giltinan (1995) apresentaram as seguintes expressões para a definição de uma expressão geral da

função de variância dos erros dentro de grupos:

$$\text{var}(e_{ij} \mid b_i) = s^2 g^2(m_{ij}, n_{ij}, d),\ \text{i=1,…,M, j=1,…,n}_i \quad (13)$$

em que M é o número de grupos, n_i é o número de observações para o ith grupo, $\mu_{ij} = E(y_{ij}|b_i)$, v_{ij} é o vetor de covariantes da variância, d é o vetor de parâmetros da variância e g(.) é a função de variância. Na ciência florestal, é comum a variabilidade dentro de grupos aumentar com o valor absoluto de um covariante. Por exemplo, a variabilidade do volume aumenta com o aumento da variável combinada diâmetro quadrado e altura. Esta técnica seria aplicada para corrigir esta e outras variações.

Na análise de correlação, dentre muitas famílias de estruturas de correlação, a estrutura auto-regressiva de média móvel (ARMA) (Box et al., 1994) é uma das mais utilizadas e conhecidas para este tipo de análise. A estrutura geral é dada por:

$$e_t = \sum_{i=1}^{p} f_i e_{t-i} + \sum_{j=1}^{q} q_j a_{t-j} + a_t \quad (14)$$

em que e_t refere-se a uma observação no tempo *t* e a_t é o termo de erro. A primeira parte da expressão refere-se ao modelo autoregressivo (AR(p)) e a segunda parte ao de movimento médio (MA(q)). Se p=0, nós teremos uma dituação MA(q) e, ao contrário, se q=0, a situação seria de AR(p). Na parte AR(p), f representa os parâmetros de correlação com ordem *p* e (*t-i*) representa a distância entre duas observações (lag). A tendência é que os valores de f 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. Na parte do movimento da média (MA(q)), o modelo assume que uma observação atual é uma função linear dos termos de erro (a_t), identicamente e independentemente distribuídos.

2.5 Dados

A base de dados utilizada no estudo é proveniente de povoamentos comerciais de clones do gênero *Eucalyptus grandis e E. urophylla*, da região costal brasileira, nos estados do Espírito Santo e Bahia, localizados entre as coordenadas17°48' S e 40 °17' W. Na Figura 1, cada clone é representado por um gráfico e cada unidade amostral é representada por uma linha nas coordenadas dos gráficos. Por exemplo, o clone número 6039 é representado por 4 unidades amostrais. Cada unidade foi permanentemente amostrada entre 3 e 10 vezes, com idades variando entre 2 e 10 anos, no período de 1992 a 2001, e a área das unidades amostrais variou de 131 a 200 m^2. Com base nestas variações, a base de dados pode ser classificada como longitudinal, irregularmente espaçada e desbalanceada.

Também a Figura 1 mostra a relação entre 1/idade e ln(G) para cada unidade amostral dentro de cada clone. Pode-se perceber um decréscimo linear consistente no logaritmo da área basal (ln(G)) com o aumento do inverso da idade, mas com variações no intercepto e, ou, na inclinação das curvas, para combinações de unidade amostral/clone. Com base nesta tendência, alguns clones poderiam ser agrupados. Por exemplo, os clones 6039, 6054, 3903, 2747 e 1030 poderiam ser agrupados, baseando-se no potencial de crescimento dos mesmos. Isto não foi feito porque cada clone poderia ter diferente regime de manejo e, ou, características tecnológicas para diferentes produtos finais.

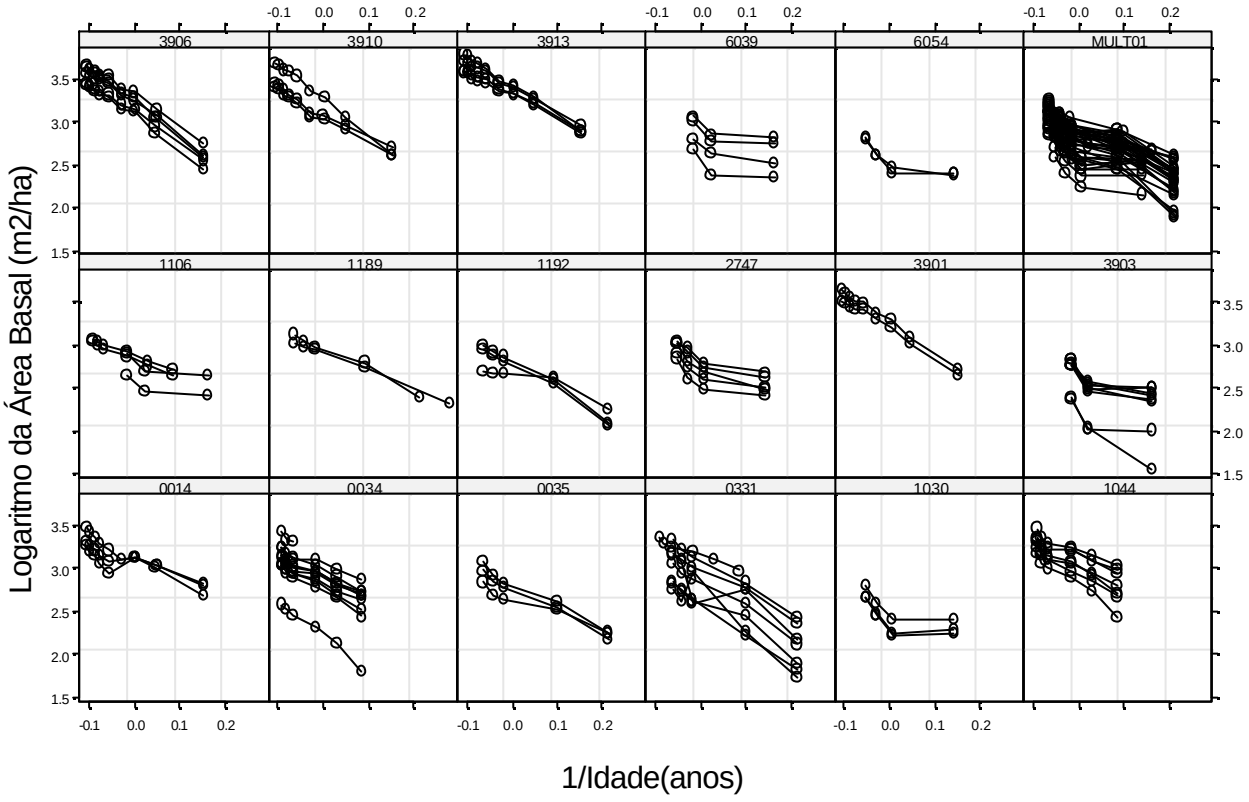

Figura 1. Relação linear entre o logaritmo da área basal (m^2/hectare) e o inverso da idade (1/idade), para 18 clones

Figure 1. *Linear relationship of the natural logarithm of Basal Area (m2/hectare) measured over time (1/age), for 18 different clones*

3 RESULTADOS E DISCUSSÃO

3.1 Ajuste de modelos lineares de efeito misto

A variável resposta foi o ln(G) e o efeito fixo será representado por 1/idade. Os efeitos aleatórios foram representados pelas unidades amostrais, ou parcelas, as quais foram selecionadas ao acaso da população alvo. Portanto, neste caso específico, foi ajustado um modelo de um único nível.

O primeiro passo foi ajustar equações baseadas no modelo linear simples relacionando ln(G) como resposta e 1/idade como covariante para todos as unidades amostrais, ignorando a estrutura de grupos. Após o ajuste, verificaram-se uma considerável variabilidade e alguns pontos de influência e, ou, *outliers*, como mostrado na Figura 2. Observações como 288, 26 e 123 podem ser consideradas de alta influência nos valores dos parâmetros estimados. Adicionalmente, a representação gráfica dos quantis da distribuição normal padrão indicou uma certa assimetria na distribuição dos

resíduos. Estes resultados indicam que o modelo de regressão linear simples não é adequado para a representação da estrutura dos dados.

Para se verificar diferenças entre clones, foram ajustadas equações com os parâmetros intercepto e inclinação para cada clone. Os resultados do efeito das interações são mostrados na Tabela 1. Todas as interações clone x (1/idade) tiveram valor de probabilidade significativo, sugerindo que os padrões de crescimento são diferentes para cada clone. Devido ao fato de os dados serem de medidas repetidas para cada unidade amostral, a pressuposição básica de independência nos modelos lineares pode ter sido violada. Para uma visualização prática da situação em estudo, três diferentes clones foram representados na Figura 3. Quando os resíduos de cada unidade amostral são representados graficamente, pode-se perceber, na representação, que o sinal tende a ser o mesmo para cada clone. Esta característica foi uma das motivações para se usar a modelagem de efeito misto no caso em questão.

O próximo passo foi a implementação de uma análise preliminar para decidir quais efeitos aleatórios devem ser incluídos no modelo e qual a estrutura de covariância mais apropriada. Para eliminar a possível correlação entre os parâmetros para cada unidade, os dados foram centrados em 4 anos (1/idade=0,25). Os valores estimados e os

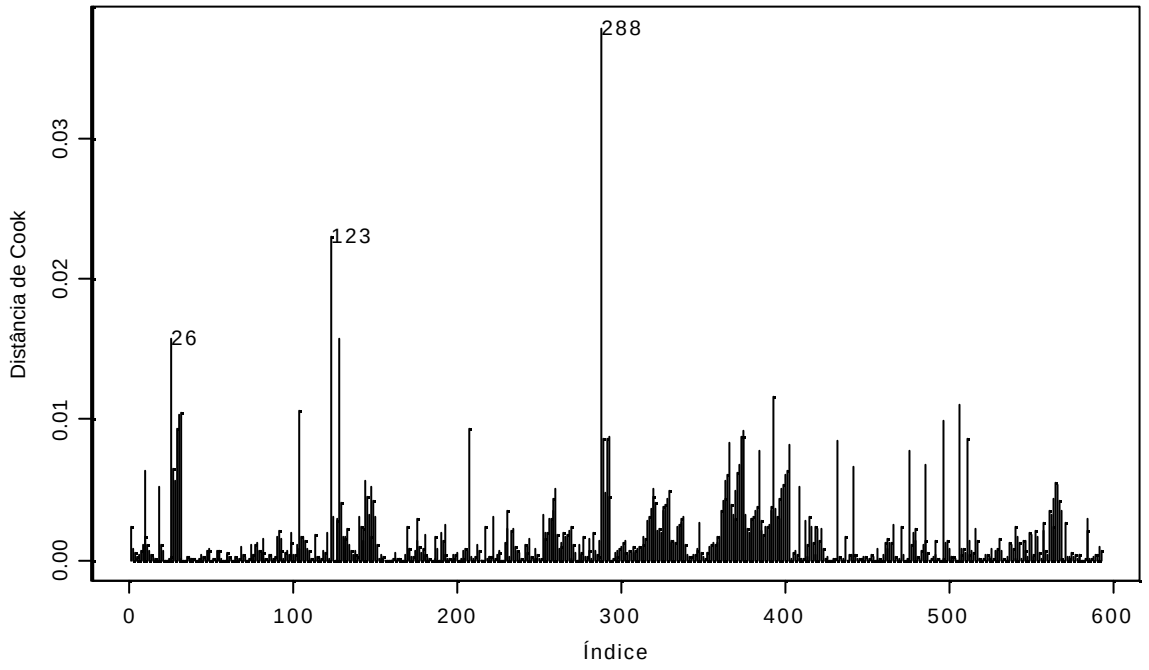

Figura 2. Diagnóstico gráfico de uma regressão linear simples ajustada para a variável ln(G) em função da variável 1/idade, para 595 unidades amostrais representando 18 clones

Figure 2. *Diagnostic plots for the simple linear regression model fitted on ln(BA) versus 1/age, for 595 plots representing 18 clones*

Tabela 1. Significância da variável clone e da interação entre clone e 1/idade, representando o banco de dados utilizado na análise

Table 1. *Significance of the Clones and interaction between Clone and 1/age representing the data set used in the analysis*

Coeficientes	Valores	Erro Padrão	Valor t	Pr(>|t|)
Intercepto	2,7628	0,0205	134,8305	< 0,0001
Clone0014	-0,1425	0,0504	-2,8278	0,0057
Clone0034	-0,0715	0,0365	-1,9599	0,0529
Clone0331	-0,0706	0,0210	-3,3541	0,0011
Clone1030	0,0361	0,0124	2,9166	0,0044
Clone2747	0,0312	0,0111	2,8066	0,0061
Clone3901	-0,0235	0,0055	-4,2767	< 0,0001
Clone3903	0,0212	0,0061	3,4765	0,0008
Clone3906	0,0162	0,0072	2,2644	0,0258
Clone3910	0,0306	0,0058	5,2548	< 0,0001
Clone6039	-0,0165	0,0072	-2,2912	0,0241
Clone6054	-0,0035	0,0019	-1,7990	0,0751
(1/idade)	-2,4075	0,0764	-31,5007	< 0,0001
Clone0014:(1/idade)	-1,8937	0,2103	-9,0046	< 0,0001
Clone0034:(1/idade)	-3,2333	0,1793	-18,0317	< 0,0001
Clone0035:(1/idade)	-3,2125	0,2125	-15,1206	< 0,0001
Clone0331:(1/idade)	-3,1377	0,1470	-21,3491	< 0,0001
Clone1030:(1/idade)	-4,3190	0,2659	-16,2444	< 0,0001
Clone1044:(1/idade)	-2,1929	0,1994	-10,9973	< 0,0001
Clone1106:(1/idade)	-3,0534	0,2345	-13,0193	< 0,0001
Clone1189:(1/idade)	-2,5108	0,2452	-10,2378	< 0,0001
Clone1192:(1/idade)	-3,3173	0,2140	-15,5048	< 0,0001
Clone2747:(1/idade)	-3,1551	0,2133	-14,7900	< 0,0001
Clone3901:(1/idade)	-1,2704	0,2689	-4,7238	< 0,0001
Clone3903:(1/idade)	-3,7837	0,1703	-22,2177	< 0,0001
Clone3906:(1/idade)	-1,5131	0,1877	-8,0601	< 0,0001
Clone3910:(1/idade)	-1,7063	0,2267	-7,5266	< 0,0001
Clone3913:(1/idade)	-0,6142	0,2027	-3,0296	0,0026
Clone6039:(1/idade)	-2,7458	0,2284	-12,0233	< 0,0001
Clone6054:(1/idade)	-3,8854	0,3196	-12,1560	< 0,0001
CloneMUL1:(1/idade)	-2,9523	0,1002	-29,4767	< 0,0001

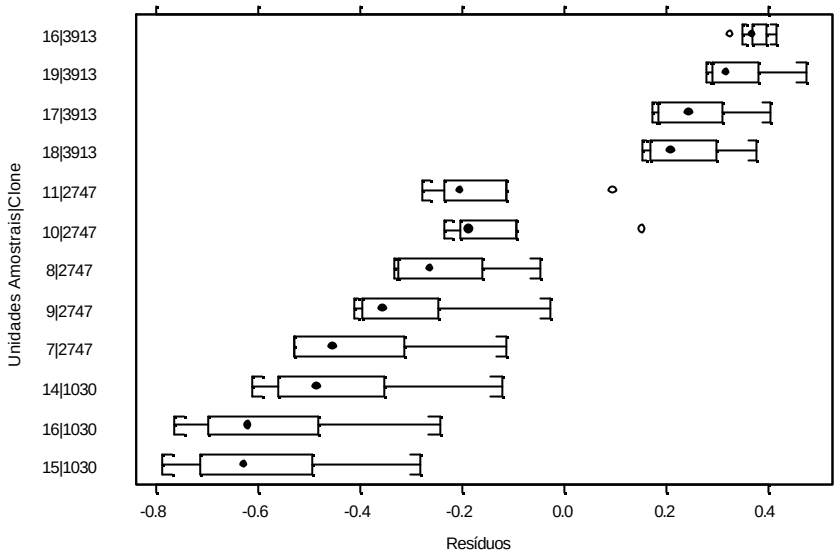

Figura 3. Distribuição dos resíduos do modelo linear por unidade amostral

Figure 3. *Residual distribution of the linear model by subject*

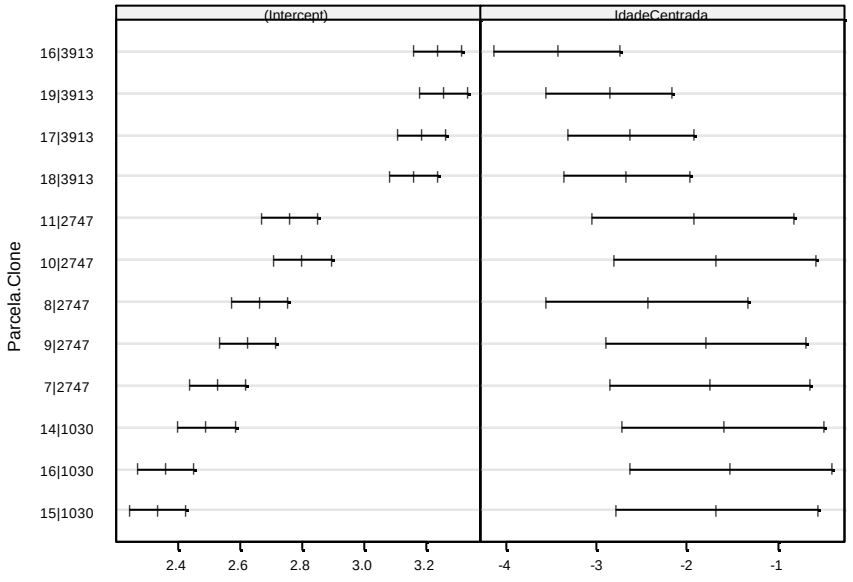

Figura 4. Intervalos de confiança do interpecto e da inclinação da reta para cada unidade amostral, com a=0,05

Figure 4. *Confidence intervals for intercept and slope for each plot, with a=0,05*

respectivos intervalos, para 12 unidades amostrais, foram representados graficamente para visualizar como os parâmetros estimados variam entre indivíduos (Figura 4). Os intervalos de confiança geraram uma clara indicação de que é necessário estimar interceptos separados e, em alguns casos, a inclinação para cada unidade, para contabilizar a variabilidade entre unidades amostrais. Pelo fato de terem sido usadas apenas 12 unidades amostrais nesta representação, os intervalos de confiança tiveram grande sobreposição. Se tivessem sido utilizadas todas as 115 unidades, as sobreposições seriam menos freqüentes, indicando que ambos os parâmetros podem ser considerados como de efeito aleatório.

Com base nas análises anteriores, têm-se suficientes argumentos para usar o modelo de efeitos misto, considerando as unidades amostrais como efeito aleatório.

O modelo de efeito misto foi primeiramente ajustado considerando apenas o inverso da idade como covariante. Os valores dos critérios de informações de CIA, CIB e MVR foram de -747,3824, -721,0611 e 379,6912, respectivamente. Visando uma confirmação de que o padrão de crescimento em área basal é diferente entre clones, uma nova equação foi ajustada incluindo o clone e a interação clone-inverso da idade como variáveis categóricas. Os novos respectivos valores para CIA, CIB e MVR foram -644,8446, -471,7272, e 362,4223. Com base nestes critérios, o primeiro ajuste teve melhor performance. Entretanto, quando os valores de probabilidade p foram analisados, de um total de 18 clones, 10 tiveram valores menores do que 0,05, indicando diferentes padrões de crescimento para cada clone. Portanto, decidiu-se manter as variáveis clone e clone x (1/idade) como covariantes. Dessa maneira, o propósito final foi desenvolver um modelo que considera a variabilidade entre clones.

3.2 Verificação das pressuposições de distribuição

As duas pressuposições básicas relacionadas à distribuição a serem verificadas referem-se às distribuições dos erros dentro dos grupos e dos efeitos aleatórios. Os erros dentro dos grupos são considerados independentes e identicamente/normalmente distribuídos, com média zero e variância s^2, e independentes dos efeitos aleatórios. Os efeitos aleatórios devem ser testados se os mesmos possuem distribuição normal com média zero e matriz de variância e covariância **?** , a qual não depende do grupo e são independentes entre grupos. Como observado por Pinheiro & Bates (2000), o método mais prático para a verificação da validade dessas pressuposições baseia-se na representação gráfica dos resíduos, dos valores estimados e nos efeitos aleatórios estimados. Também, testes de hipóteses podem ser utilizados para esta verificação, mas raramente contradizem as informações geradas com a representação gráfica.

Como pode-se verificar na Figura 5, em 12 unidades amostrais, representando 3 dos 18 clones na análise, os resíduos estão distribuídos em torno da linha representada pelo zero, confirmando a pressuposição de que E[**e**]=**0**. Na análise dos resíduos, fica aparente uma diferente variabilidade entre unidades amostrais. Portanto, a pressuposição de variância constante dentro dos grupos foi violada, sendo necessário modelar esta variabilidade para melhorar o desempenho do modelo. Caso sejam usados todos os 18 clones na análise, a distribuição dos resíduos se apresentou mais agrupada. Também a Figura 5 mostra alguns outliers nas combinações de parcela:clone 16|3913 e 2|2747 e resíduo relativamente elevado para o clone 1030. A representação gráfica dos resíduos padronizados versus os valores ajustados mostra que a variabilidade entre unidades para alguns clones é maior do que unidades para outros (Figura 6).

Com base nas observações das Figuras 5 e 6, a primeira iniciativa foi modelar a variância por clone para o erro entre grupos. A Tabela 2 representa os diferentes valores estimados para os modelos homocedástico e heterocedástico. Menores valores para CIA e CIB, maiores valores para o logaritmo da máxima verossimilhança e valores muito pequenos para valor-p do TRMV confirmam que o modelo heterocedástico melhorou sensivelmente o ajuste e explica melhor a variação dos dados quando comparado com o modelo homoscedástico.

O próximo passo foi acessar as pressuposições baseadas nos efeitos aleatórios. Por meio de análise gráfica (Figura 7), verificou-se que a pressuposição de normalidade parece razoável para os efeitos aleatórios. Uma segunda presuposição associada aos efeitos aleatórios também pode ser verificada na Figura 7. Pode ser visto que os pares inclinação-intercepto para todas as combinações unidade:clone têm média próxima de zero e variância aproximadamente constante.

Como os dados representam informações longitudinais, com medidas repetidas por unidade amostral e ou dados espaciais, com informações indexadas por localização espacial, o próximo passo foi verificar a estrutura de correlação e modelar tal estrutura. Devido ao fato de os dados não serem igualmente espaçados no tempo, a correlação espacial foi usada para se ajustar modelos temporais contínuos. A Figura 8 é uma representação gráfica do semivariograma amostrado. Os valores do semivariograma aumentam até 0,10 e depois decrescem. Essa tendência foi modelada utilizando-se cinco padrões de correlação espacial: exponencial, gaussiana, linear, racional quadrática e esférica. O modelo de correlação esférica teve melhor ajuste nesta situação específica. A representação gráfica resultante (Figura 9) mostra uma variabilidade aleatória em torno de y=0,7, sem um padrão definitivo. Esta variabilidade sugere que o modelo esférico é adequado. Os valores de CIA and CIB foram menores, confirmando que o modelo heterocedástico com autocorrelação melhorou a representação dos dados (Tabela 3). Também o valor maior do TRMV confirma a evidência de dependência dos dados, gerando um valor de probabilidade de 0,0004.

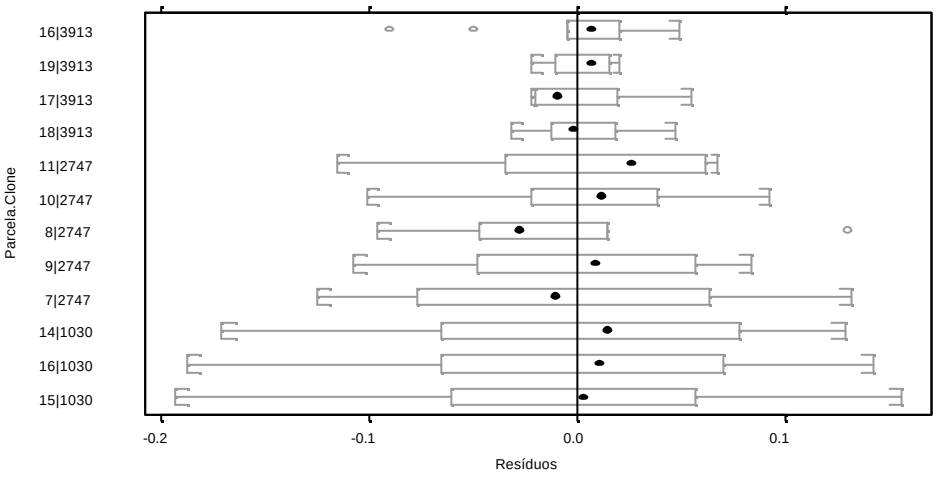

Figura 5. Boxplots dos resíduos para o modelo linear misto

Figure 5. *Boxplots of the residuals for linear mixed-effect model*

Tabela 2. Comparação entre o critério de informação de Akaike (CIA), critério de informação baysiana (CIB) e o logaritmo da máxima verossimilhança (LMV) para os modelos homocedástico e heterocedástico

Table 2. *Comparison among Akaike Information Criterion (AIC), Bayesian Information Criterion(BIC) and LogLikelihood (logLik) for homoscedastic and heteroscedastic models*

Modelo	GL	CIA	CIB	LMV	Teste da Razão		P-Value
1 – Homocedástico	6	-747,38	-721,06	379,69	-	-	
2 – Heterocedástico	70	-901,24	-798,28	520,62	1 vs 2	281,86	<0,0001

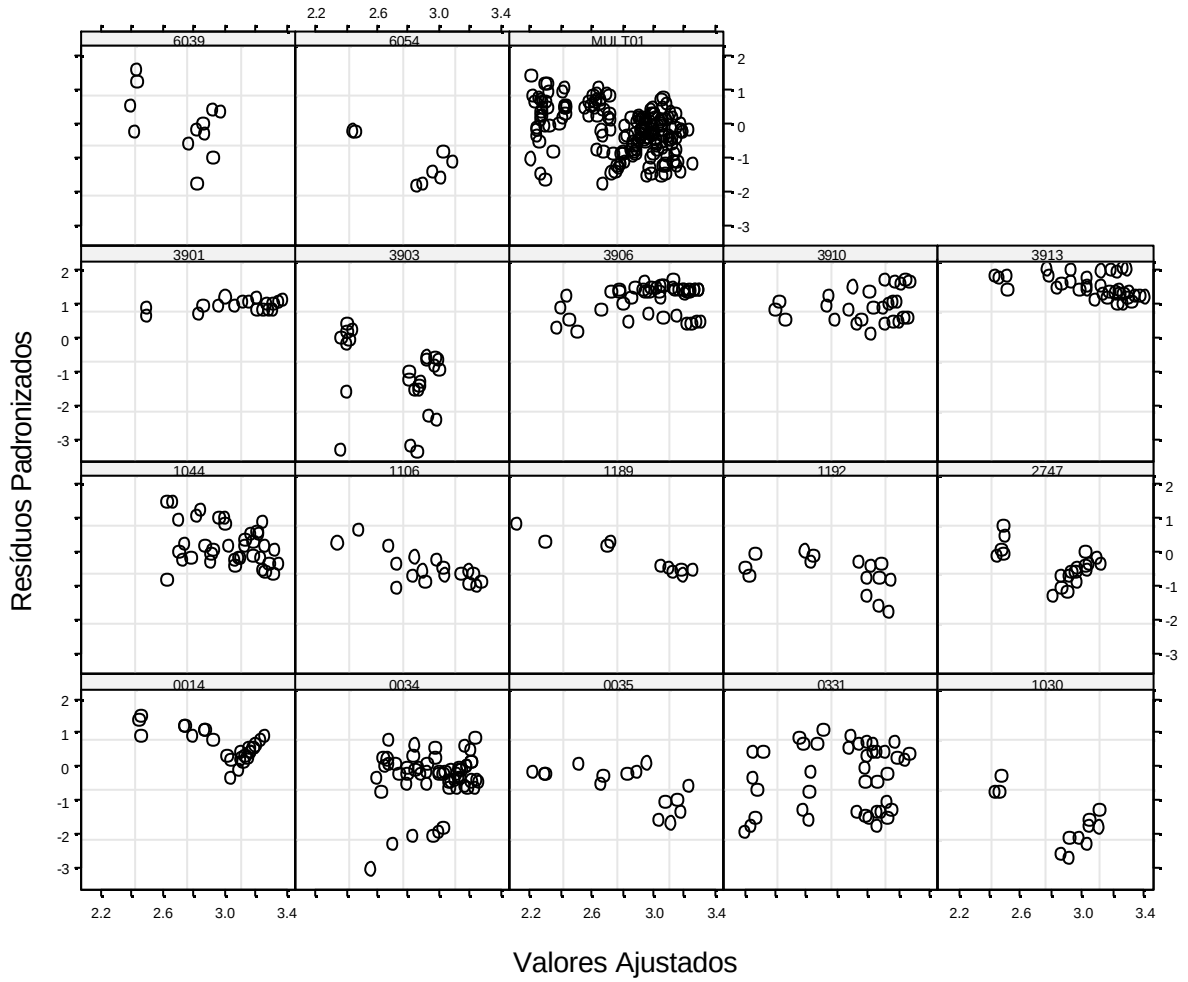

Figura 6. Valores residuais versus valores ajustados para o modelo linear misto por clone

Figure 6. *Residuals versus fitted values for linear mixed-effect model by clone*

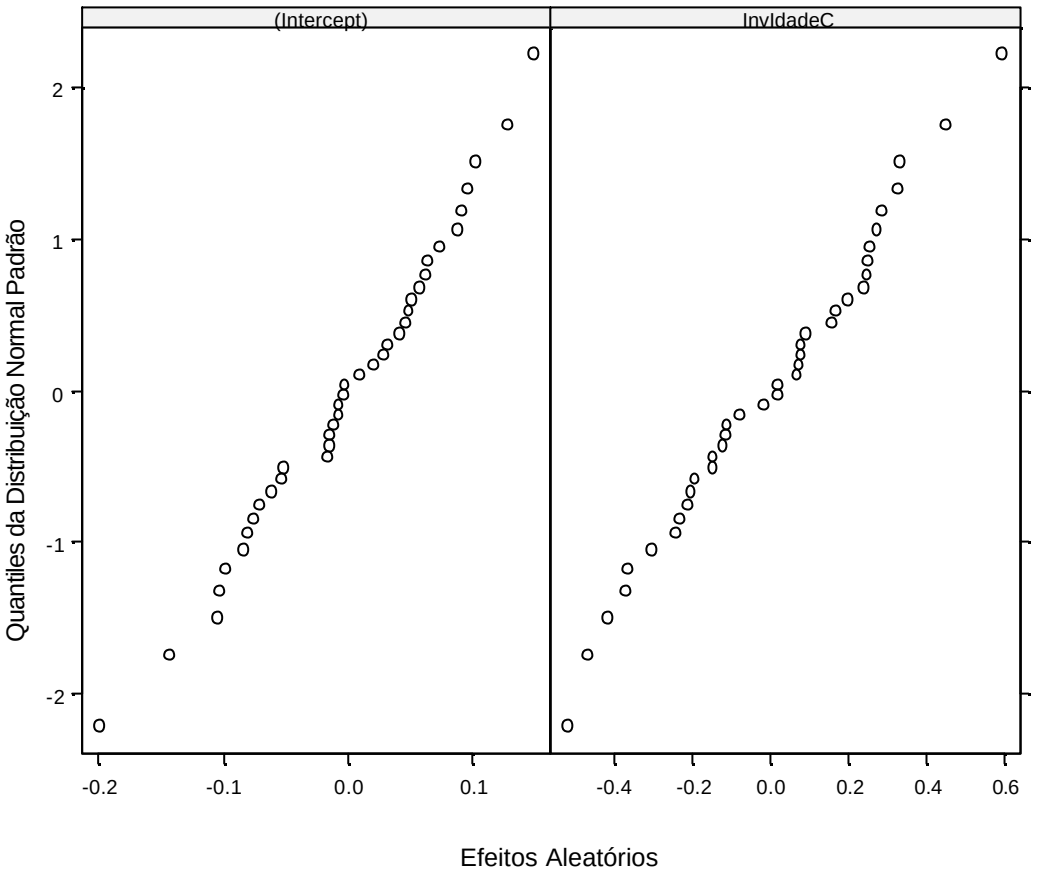

Figura 7. Distribuição normal dos efeitos aleatórios para o modelo heterocedástico ajustado

Figure 7. *Normal plot of estimated random effect for heterocedastic fitted model*

A pressupossição de normalidade para cada clone pode ser confirmada pela Figura 10. Com algumas excessões, os resíduos tiveram uma distribuição normal consistente, confirmando a melhoria obtida com a modelagem da heterocedastidade e da autocorrelação.

Os parâmetros estimados finais para os efeitos fixos e aleatórios podem ser vistos na Tabela 4. Considerando todas as 115 unidades amostrais, ambos os efeitos aleatórios para o intercepto e para a inclinação tiveram valores alternados de positivo e negativo. As estimativas para cada unidade amostral foram obtidas adicionando-se os efeitos fixos aos aleatórios. Portanto, considerando os valores dos parâmetros, as curvas para unidade amostral tiveram diferentes interceptos e inclinações

Tabela 3. Comparação entre o critério de informação de Akaike (CIA), critério de informação baysiano (CIB), logaritmo da máxima verossimilhança (LMV) e o teste da razão da máxima verossimilhança (TRMV) para os modelo homocedástico, heterocedástico e heterocedástico com autocorrelação

Table 3. *Comparing Akaike Information Criterion (AIC), Bayesian Information Criterion(BIC), LogLikelihood (logLik) and Likelihood Ratio Teste (LRT) for homocedastic, heterocedastic and heterocedastic-autocorrelation models*

Modelo	GL	CIA	CIB	LMV	TRMV		Valor-p
1- Homocedástico	6	-747,38	-721,06	379,69	-	-	-
2- Heterocedástico	70	-901,24	-798,28	520,62	1 vs 2	281,86	< 0,0001
3-Heterocedástico/ Autocorrelação	37	-1034,72	-872,41	554,36	2 vs 3	67,49	0,0004

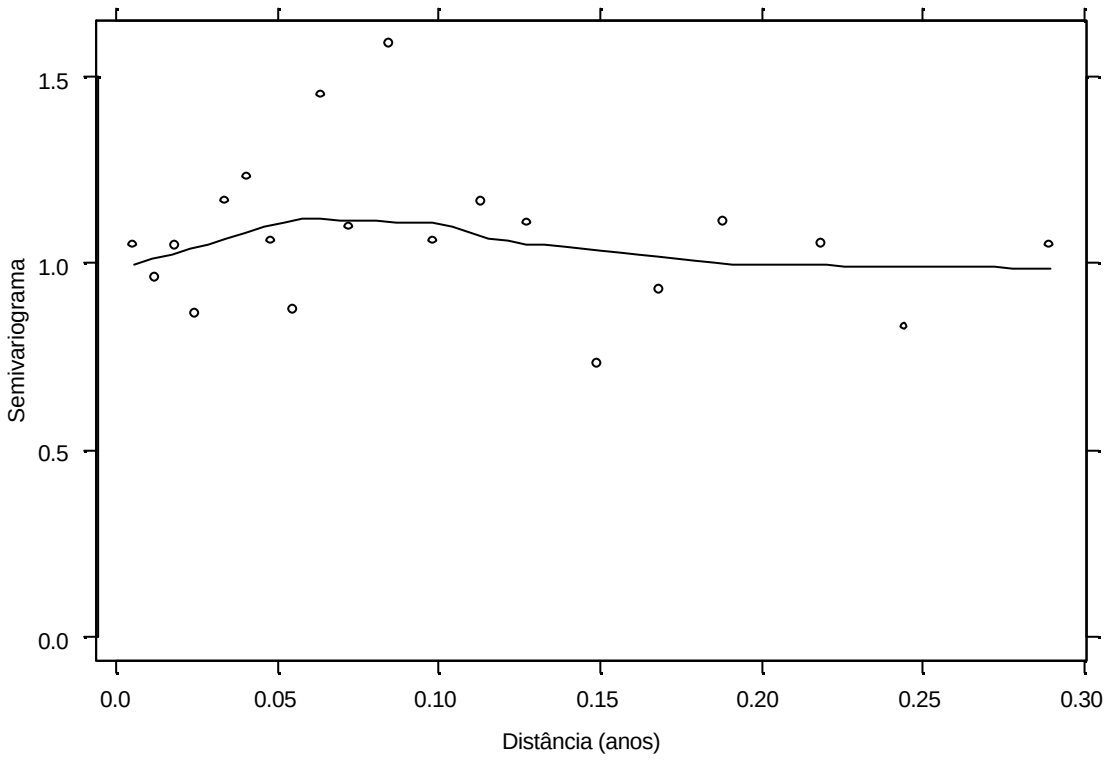

Figura 8. Estimativas do semivariograma amostrado para o modelo linear de efeito misto antes da modelagem da autocorrelação

Figure 8. *Sample semivariogram estimates for linear mixed-effect model before modeling the spatial autocorrelation*

Tabela 4. Estimativas dos parâmetros de efeitos fixo e misto para o modelo final, com modelagem da heterocedasticidade e da autocorrelaçao, representando 26 unidades amostrais em um total de 115

Table 4. *Fixed and mixed parameter estimates for the final model, with heteroscedastic and autocorrelation modeled, representing 26 plots in a total of the 115*

Parcela:Clone	Intercepto fixo $ß_o$	Intercepto aleatório b_{oi}	Inclinação fixa $ß_1$	Inclinação aleatória b_{1i}
1:0331	2,7505	0,173135	-2,494754	-0,602890
2:0331	2,7505	-0,326031	-2,494754	-0,820808
3:0331	2,7505	0,086599	-2,494754	-0,437708
4:0331	2,7505	-0,112138	-2,494754	-0,759980
5:0331	2,7505	0,003331	-2,494754	-1,161210
6:0331	2,7505	-0,468503	-2,494754	-0,934774
7:0331	2,7505	-0,339641	-2,494754	-1,133078
8:0331	2,7505	0,303319	-2,494754	0,701643
1:1030	2,7505	-0,153020	-2,494754	0,456104
1:1044	2,7505	-0,062380	-2,494754	-0,320059
2:1030	2,7505	-0,300282	-2,494754	0,375090
2:1044	2,7505	0,037841	-2,494754	-0,644155
1:1106	2,7505	-0,172280	-2,494754	0,907539
3:1030	2,7505	-0,279085	-2,494754	0,517948
3:1044	2,7505	0,113330	-2,494754	0,077612
2:1106	2,7505	0,075706	-2,494754	0,766970
1:1189	2,7505	0,031074	-2,494754	0,462230
1:1192	2,7505	-0,135062	-2,494754	-0,746007
4:1044	2,7505	0,273362	-2,494754	0,690275
3:1106	2,7505	0,001544	-2,494754	0,186253
2:1189	2,7505	0,054275	-2,494754	-0,143285
2:1192	2,7505	-0,166382	-2,494754	-0,713032
5:1044	2,7505	0,343911	-2,494754	0,488235
4:1106	2,7505	-0,043321	-2,494754	0,330997
3:1192	2,7505	-0,227671	-2,494754	0,774958
6:1044	2,7505	0,137643	-2,494754	0,491384

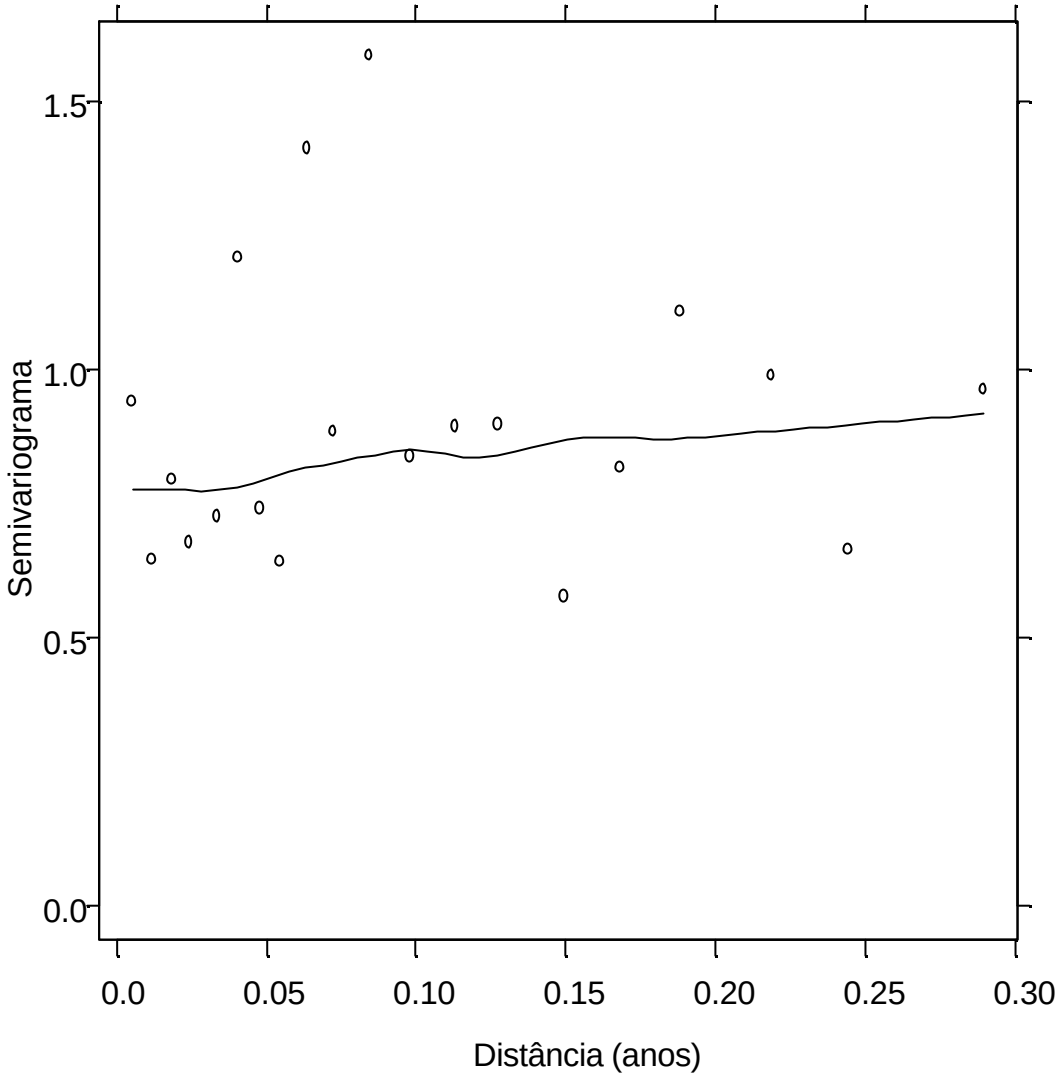

Figura 9. Estimativas do semivariograma amostrado para o modelo linear misto após a modelagem da autocorrelação.

Figure 9. *Sampled semivariogram estimates for linear mixed-effect model after modeling the spatial autocorrelation*

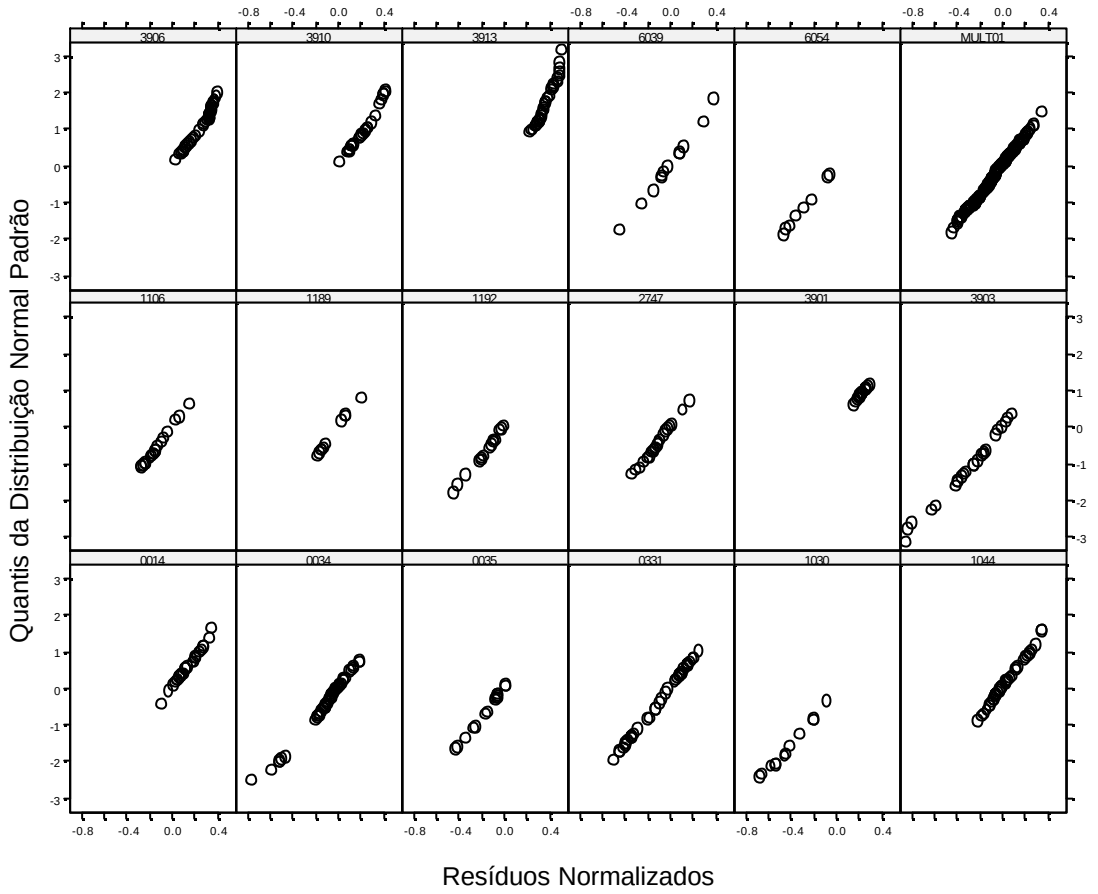

Figura 10. Normalização dos resíduos para cada clone

Figure 10*. Normal plots of residuals for each clone*

4 CONCLUSÕES

O modelo linear de efeito misto gerou estimativas precisas tanto para o efeito fixo (1/idade) quanto para o efeito aleatório (unidade amostral). Uma maior variabilidade foi verificada entre diferentes clones do que entre unidades amostrais para o mesmo clone, indicando a necessidade de se incluir o efeito do clone na análise. As estimativas individuais geradas com o modelo linear misto tendem a ser mais compactas porque as estimativas representam os efeitos individuais e as estimativas referentes aos efeitos fixos, associadas com as médias populacionais, proporcionam uma certa robustez ao processo.

Embora não tenham sido verificados problemas com a distribuição dos resíduos dentro das unidades amostrais, a heterocedasticidade entre unidades foi modelada e as estatísticas utilizando critérios de informações e máxima verossimilhança tiveram uma melhoria significativa. Também, devido ao fato das unidades amostrais representarem diferentes localizações, com

variações ambientais, o padrão de correlação espacial foi modelado. Mais uma vez, as estatísticas baseadas nos critérios de informações e na máxima verossimilhança tiveram melhoria significativa. A representação gráfica do semivariograma e dos quantis para a distribuição normal comprovou a superioridade do modelo que considerou heterocedasticidade e autocorrelação.

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. 380 p.

CHRISTMAN, M.C.; JERNIGAN, R. J. Spatial correlation models as applied to evolutionary biology. **Modeling longitudinal and spatially correlated data.** New York: Springer-Verlag, 1977. 410 p.

DAVIDIAN, M.; GILTINAN, D. M. **Nonlinear models for repeated measurement data.** London: Chapman and Hall, 1995. 359 p.

FANG, Z. X.; BAILEY, R. L. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. **Forest Science,** Bethesda, v. 47, n. 3, p. 287-300, Aug. 2001.

GREGOIRE, T. G.; SCHABENBERGER, O.; BARRETT, J. P. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. **Canadian Journal Forest Research,** Ottawa, v. 25, n. 1, p. 137-156, Jan. 1995.

LAIRD, N. M.; WARE, J. H. Random effects models for longitudinal Data. **Biometrics,** Washington, v. 38, n. 4, p. 963-974, 1982.

LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. **SAS system for mixed models.** Cary, NC: SAS Institute, 1996. 633 p.

PINHEIRO, J. C.; BATES, D. M. **Mixed-effects models in S and SPlus.** New York: Springer-Verlag, 2000. 528 p.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. **Akaike information criterion statistics.** Dordrecht: Kluwer Academic Publishers, 1986. 240 p.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics,** v. 6, P. 461-464, 1978.

SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance components.** New York: Wiley, 1992. 245 p.

STRAM, D. O.; LEE, J. W. Variance components testing in the longitudinal mixed-effects models. **Biometrics,** Washington, v. 50, p. 1171-1177, 1994.

VERBEKE, G.; LESAFFRE, E. A linear mixed-effects model with heterogeneity in the random-effects population. **Journal of the American Statistical Association,** Alexandria, v. 91, n. 433, p. 217-221, Mar. 1996.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed effects model in practice:** a SAS-oriented approach. New York: Springer-verlag, 306 p.

VONESH, E. F.; CHINCHILLI, V. M. **Linear and nonlinear models for the analysis of repeated measurements.** New York: Marcel Dekker, 19997. 324 p.