



DANIEL DANTAS

**ESTIMADOR GEOESTATÍSTICO E MODELOS DE EFEITO
MISTO PARA MODELAGEM HIPSOMÉTRICA E
VOLUMÉTRICA DE POVOAMENTOS FLORESTAIS**

LAVRAS - MG

2019

DANIEL DANTAS

**ESTIMADOR GEOESTATÍSTICO E MODELOS DE EFEITO
MISTO PARA MODELAGEM HIPSOMÉTRICA E
VOLUMÉTRICA DE POVOAMENTOS FLORESTAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Manejo Florestal, para a obtenção do título de Mestre.

Prof. Dr. Natalino Calegario

Orientador

LAVRAS - MG

2019

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Dantas, Daniel.

Estimador geoestatístico e modelos de efeito misto para
modelagem hipsométrica e volumétrica de povoamentos florestais /
Daniel Dantas. - 2019.

120 p. : il.

Orientador(a): Natalino Calegario.

Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2019.

Bibliografia.

1. Manejo florestal. 2. Biometria. 3. Inventário florestal. I.
Calegario, Natalino. II. Título.

DANIEL DANTAS

**ESTIMADOR GEOESTATÍSTICO E MODELOS DE EFEITO MISTO PARA
MODELAGEM HIPSOMÉTRICA E VOLUMÉTRICA DE POVOAMENTOS
FLORESTAIS**

**GEOSTATISTICAL ESTIMATOR AND MIXED MODELS FOR HYPOMETRIC
AND VOLUMETRIC MODELING OF FOREST STANDS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Manejo Florestal, para a obtenção do título de Mestre.

APROVADA em 19 de fevereiro de 2019.

Dr. Natalino Calegario – UFLA

Dr. Fausto Weimar Acerbi Júnior - UFLA

Dr. Samuel de Pádua Chaves Carvalho - UFMT

Prof. Dr. Natalino Calegario
Orientador

**LAVRAS - MG
2019**

À minha mãe, Normalice.

Ao meu pai, Alisson.

Aos meus irmãos, Isadora e Davi.

Aos meus avós, Adelize, Oscar, Eramita e Darci.

Dedico.

AGRADECIMENTOS

A Deus, por permitir que tudo isso se realizasse, me guiando, protegendo e sendo presença constante na minha vida. “Sem Ti, Senhor, o agora não seria”.

Aos meus pais, Alisson e Norma, e aos meus irmãos, Isadora e Davi, por todo amor, compreensão e apoio incondicional. Obrigado pela educação, por acreditarem e confiarem em mim em todos os momentos, isso foi fundamental para que conseguisse ir em busca dos meus objetivos. A vocês, meu amor e gratidão eternos!

Ao meu orientador, Natalino Calegario, pelos muitos ensinamentos transmitidos, pela boa convivência, paciência e orientação.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelo apoio financeiro.

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Engenharia Florestal, pela oportunidade de realizar esta importante etapa da minha vida; aos seus professores, pela minha formação e todo aprendizado adquirido ao longo deste período, em especial ao Willian, Bruno, Fausto e Lucas. Agradeço também à professora Dulcinéia, pela amizade.

Ao professor Marcio Leles, da UFVJM, pela amizade, apoio e ensinamentos.

À banca examinadora da defesa desta dissertação, professores Fausto e Samuel, pela participação e pelas contribuições.

Aos funcionários do Departamento de Ciências Florestais, em especial à Chica, por todo carinho, amizade e pelas boas comidas; à Adriana, pelas risadas; e ao Juliano, pela disponibilidade.

Aos colegas do Laboratório de planejamento e manejo intensivo de florestas, pela convivência e apoio, em especial ao Mario, Isaac e Elliezer.

A todos os meus colegas de curso, pelo companheirismo, amizade e pela interação produtiva. Em especial ao Talles, Luiz, Natielle e Lorena.

A todos os meus amigos: os velhos, os novos, os distantes e os mais próximos, que, sem citar nomes, sabem a importância que têm na minha vida.

Aos meus grandes amigos: Lucas Rocha, Douglas e Taiana, por tudo que vivemos e pela cumplicidade ao longo deste período; Anny, por todo apoio, incentivo e pelas contribuições; Lauriano e Lucas Baldo, pelo companheirismo.

A todos que contribuíram, direta ou indiretamente, para a conclusão deste trabalho:

O meu muito obrigado!

RESUMO

A informação de volume em um povoamento florestal é imprescindível para o manejo de florestas, levando em conta que grande parte dos processos produtivos em uma empresa é dependente dos dados de produtividade que são obtidos por meio dos inventários florestais. Nesse contexto, os procedimentos de mensuração e processamento dos dados são elementos fundamentais e merecem atenção, por serem a base para a determinação do volume em uma área. O setor florestal vive uma constante busca por técnicas que possibilitem atingir a excelência na estimativa da produção madeireira, otimizando os processos e maximizando os lucros. Neste trabalho foram propostos e avaliados o desempenho de um modelo extraído do treinamento de uma rede neural artificial e de novos modelos para estimar a altura total de árvores de eucalipto; de modelos mistos com adoção de estrutura na matriz de variância e covariância para estimativas hipsométricas e volumétricas; e de um estimador geoestatístico para predição do volume por hectare, considerando diferentes intensidades amostrais. As técnicas apresentaram desempenhos satisfatórios, resultando no ganho em precisão nas estimativas de altura total, volume individual e volume do povoamento. Os resultados do estudo permitem a redução da intensidade amostral em inventários florestais pré-corte, indicando o potencial e aplicabilidade destas técnicas na solução de problemas de mensuração e manejo florestal.

Palavras-chave: Redes neurais artificiais. Modelos lineares. Modelos não-lineares. Modelos de efeitos mistos. Geoestatística.

ABSTRACT

Volume information in a forest stand is essential for the forest management, taking into consideration that most of the productive processes in a company are dependent on the productivity data that are obtained through forest inventories. In this context, the measurement procedures and data processing are fundamental elements and deserve attention, since they are the basis for determining the volume in an area. The forest sector is constantly searching for techniques that enable it to achieve excellence in the estimation of timber production, optimizing processes and maximizing profits. In this work was proposed and evaluated the performance of a model extracted from the training of an artificial neural network and of new models to estimate the total height of eucalyptus trees; of mixed models with adoption of structure of the variance and covariance matrix for hypsometric and volumetric estimates; and a geostatistical estimator for volume prediction per hectare, considering different sample intensities. The techniques presented satisfactory performances, resulting in the gain in accuracy in the estimates of total height, individual volume and stand volume. The results of this study allow the reduction of the sample intensity in pre-cut forest inventories, indicating the potential and applicability of these techniques in the solution of problems of measurement and forest management.

Keywords: Artificial neural networks. Linear models. Nonlinear models. Mixed models. Geostatistics.

SUMÁRIO

	PRIMEIRA PARTE	11
1	INTRODUÇÃO	11
1.1	Objetivos gerais	12
1.1.1	Objetivos específicos	12
1.2	Hipóteses	12
2	REVISÃO BIBLIOGRÁFICA	14
2.1	Redes neurais artificiais	14
2.2	Modelos mistos	17
2.2.1	Modelos lineares mistos	17
2.2.2	Modelos não lineares mistos	17
2.2.3	Estruturas de variância e covariância	19
2.3	Modelo geoestatístico	21
2.3.1	Semivariograma	23
2.3.1.1	Modelos teóricos de semivariograma	26
2.3.1	Método de estimativa espacial: Krigagem	27
	REFERÊNCIAS	28
	SEGUNDA PARTE - ARTIGOS	33
	ARTIGO 1 - ACCURACY OF TREE HEIGHT ESTIMATION WITH ARTIFICIAL NEURAL NETWORK MODELS AND REGRESSION MODELS	33
1	INTRODUCTION	36
2	MATERIAL AND METHODS	37
2.1	Training and validation of ANNs	37
2.2	Adjustments of linear and non-linear models	40
2.3	Adjusting the Parabolic Model	41
2.4	Analysis of the estimates generated by the model extracted from ANN and hypsothetic models	42
3	RESULTS	44
4	DISCUSSION	50
5	CONCLUSION	51
	REFERENCES	52
	ARTIGO 2 - MODELOS MISTOS EM MULTINÍVEL PARA ESTIMATIVA DA ALTURA TOTAL DE ÁRVORES DE <i>Eucalyptus</i> spp.	57
1	INTRODUÇÃO	60
2	MATERIAL E MÉTODOS	62
2.1	Dados	62
2.2	Ajustes dos modelos linear e não-linear	62
2.3	Ajustes de modelos multinível de efeito misto	63
3	RESULTADOS E DISCUSSÃO	68
4	CONCLUSÃO	74
	REFERÊNCIAS	75
	ARTIGO 3 - MODELO NÃO LINEAR MISTO PARA ESTIMATIVA DO VOLUME DE ÁRVORES DE <i>Eucalyptus</i> spp.	78

1	INTRODUÇÃO	81
2	MATERIAIS E MÉTODOS	83
2.1	Dados	83
2.2	Modelo volumétrico	84
2.3	Modelo volumétrico multinível de efeito misto	84
3	RESULTADOS E DISCUSSÃO	88
4	CONCLUSÃO	93
	REFERÊNCIAS	94
	ARTIGO 4 - ESTIMADOR GEOESTATÍSTICO PARA PREDIÇÃO VOLUMÉTRICA POR MEIO DE DIFERENTES INTENSIDADES AMOSTRAIS	96
1	INTRODUÇÃO	99
2	MATERIAIS E MÉTODOS	101
2.1	Dados	101
2.1.1	Área de estudo	101
2.1.2	Base de dados	101
2.2	Análise exploratória	103
2.3	Intensidade amostral	103
2.4	Análise variográfica	104
2.5	Krigagem	105
3	RESULTADOS E DISCUSSÃO	106
4	CONCLUSÃO	117
	REFERÊNCIAS	118

PRIMEIRA PARTE

1 INTRODUÇÃO

A demanda por produtos de origem florestal é cada vez maior, especialmente devido ao crescente aumento demográfico mundial. Neste sentido, o planejamento e administração dos plantios florestais são importantes e sua avaliação precisa e eficiente é decisiva para ações adequadas de manejo florestal, atendendo aos objetivos técnicos e econômicos das empresas.

A informação de volume de madeira em um povoamento florestal é imprescindível para o manejo de florestas, levando em conta que grande parte dos processos produtivos em uma empresa é dependente dos dados de produção. A obtenção desta informação se dá por meio de inventário florestal, que consiste na determinação de um método de amostragem, alocação de parcelas na área e medição das variáveis de interesse.

Inventários florestais possuem custo elevado devido ao tempo necessário para os levantamentos, alocação das parcelas na área e medição de um grande número de árvores. Além disso, é susceptível a erros. Os erros de amostragem são controláveis e são influenciados, por exemplo, pelo método de amostragem e pelo tamanho da amostra. Por outro lado, os erros de não-amostragem são de difícil identificação, e várias podem ser as fontes deste segundo tipo de erro, dentre elas o uso de equações inadequadas para estimar a altura e o volume de árvores individuais.

O setor florestal vive uma constante busca por técnicas que possibilitem atingir a excelência na estimativa da produção madeireira, otimizando os processos e maximizando os lucros. Destaca-se também a existência da demanda por uma intensidade amostral capaz de reduzir o número de parcelas por hectare e ao mesmo tempo proporcionar uma boa representatividade da população, para que essa redução não interfira de forma negativa na qualidade das estimativas.

Nesse contexto, os procedimentos de mensuração e processamento dos dados são elementos fundamentais e merecem atenção, sendo necessários estudos sobre métodos que proporcionem estimativas precisas da variável altura, que são a base para a determinação do volume em uma área, e que forneçam subsídios para a tomada de decisões, por parte do gestor florestal, na condução do inventário florestal e processamento dos dados.

Esta dissertação está dividida em duas partes. A primeira busca situar o leitor diante dos objetivos deste trabalho e fazer uma abordagem teórica sobre os temas trabalhados, incluindo termos importantes que foram utilizados ao longo da segunda parte. A segunda parte foi

organizada em quatro artigos. O primeiro artigo comprova a eficiência de se extrair um modelo de uma rede neural artificial treinada e propõe a utilização de dois novos modelos hipsométricos, linear e não linear. O segundo artigo avalia o ajuste dos dois modelos hipsométricos propostos com a inclusão de efeitos aleatórios e adoção de estrutura na matriz de variância e covariância. No terceiro artigo, o modelo volumétrico não linear de Schumacher e Hall é ajustado por meio de regressões fixa e mista com adoção de estrutura na matriz de variância e covariância. O quarto artigo apresenta a eficiência do estimador geoestatístico para estimativas de volume por hectare, considerando diferentes intensidades amostrais.

1.1 Objetivo geral

- a) Propor e avaliar o desempenho de modelos para estimar a altura total e o volume de madeira de árvores de eucalipto; e avaliar a influência da intensidade amostral na precisão do inventário florestal com a utilização de geoestatística.

1.1.1 Objetivos específicos

- a) Propor e avaliar o desempenho de um modelo extraído de uma rede neural artificial e de novos modelos hipsométricos linear e não linear;
- b) avaliar o ajuste de novos modelos hipsométricos com regressão mista em multinível;
- c) avaliar o ajuste de um modelo volumétrico misto em multinível;
- d) avaliar o desempenho do estimador geoestatístico para estimar o volume de madeira em nível de povoamento, considerando diferentes intensidades amostrais.

1.2 Hipóteses

Diante dos objetivos propostos foram formuladas as seguintes hipóteses:

- a) é possível extrair um modelo hipsométrico eficiente de uma rede neural artificial;
- b) modelos mistos em multinível com estrutura na matriz de variância e covariância proporcionam aumento na precisão das estimativas de altura total das árvores;
- c) modelos mistos em multinível com estrutura na matriz de variância e covariância proporcionam aumento na precisão das estimativas volumétricas de árvores;
- d) o uso de geoestatística permite a redução da intensidade amostral de um inventário florestal pré-corte sem comprometer a precisão das estimativas de volume da área.

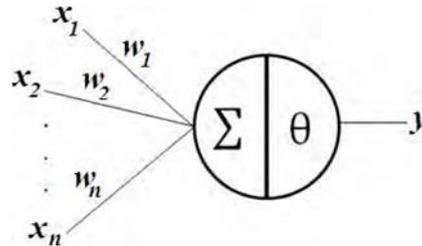
2 REVISÃO BIBLIOGRÁFICA

2.1 Redes neurais artificiais

As Redes Neurais Artificiais (RNA) têm sido utilizadas como alternativa aos modelos hipsométricos para a modelagem e projeção da produção de florestas. Segundo Braga et al. (2007), uma rede neural artificial é um processador computacional constituído por um grande

número de processadores simples (neurônios artificiais) (FIGURA 1), interconectados, baseados em neurônios encontrados nos seres humanos.

Figura 1 – Estrutura geral de um neurônio artificial.

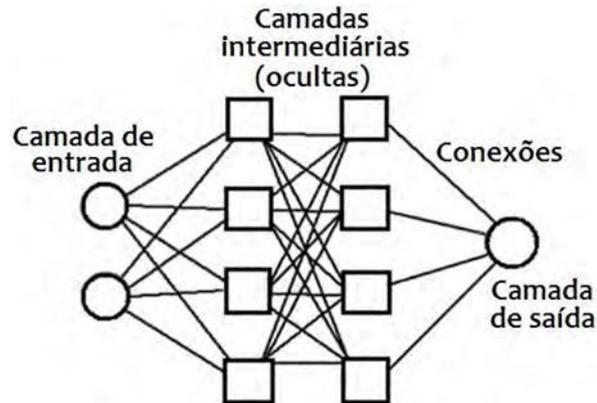


Fonte: Adaptado de Braga et al. (2000).

Um neurônio artificial é a unidade de processamento de informação de uma rede neural artificial. O primeiro modelo de neurônio artificial, proposto por McCulloch e Pitts em 1943, é uma simplificação do que se sabia a respeito do neurônio biológico na época. Constituído de n entradas x_1, x_2, \dots, x_n (dendritos) e uma saída y (axônio). As entradas são associadas a pesos w_1, w_2, \dots, w_n que representam as sinapses, que podem ser negativas ou positivas. É calculada a soma ponderada dos valores $x_i w_i$ recebidos e, se o valor obtido ultrapassa o limiar (*threshold*) do neurônio, este “dispara”, ou seja, ativa a função de ativação e gera a saída.

Os neurônios artificiais distribuem-se em camadas (FIGURA 2), o que justifica a denominação da rede neural artificial, cujas camadas conectam-se por pesos, que captam a influência de cada variável de entrada na variabilidade da variável resposta e atribuem um valor para cada variável, se o peso for alto, a entrada apresenta alta influência. Então, são armazenados os padrões de comportamento e as relações entre as variáveis independentes e a variável dependente (HAYKIN, 2009). Com isso, ao se ajustar os pesos de um neurônio artificial, o conhecimento adquirido torna-se disponível para uso e é possível obter a saída desejada para entradas específicas.

Figura 2 - Estrutura de uma rede neural artificial.



Fonte: Adaptado de Braga et al. (2000).

Nas camadas ocultas estão as funções de ativação, cujo objetivo é, além de introduzir não-linearidade à RNA e encontrar a função que melhor se adapta à distribuição dos dados, limitar os valores de cada neurônio para que a RNA não seja paralisada por neurônios divergentes (BRAGA, 2007). As funções de ativação mais utilizadas para aproximação de funções são a limiar, tangente hiperbólica, parabólica e sigmoideal (HAYKIN, 2009).

As características que mais chamam a atenção nas RNA são a sua aprendizagem e generalização de informações. Em outras palavras, as RNA, por meio de um exemplo aprendido, detectam as relações entre as variáveis e podem generalizar o conhecimento assimilado para um conjunto de dados desconhecidos. Outra característica interessante das RNA é a facilidade de extrair características não explícitas de um grupo de informações que lhes são fornecidas como exemplos (DIAMANTOPOULOU, 2005; GORGENS, 2014).

A forma como um determinado problema é representado internamente e o paralelismo de sua arquitetura pode tornar a solução de problemas com RNA melhor do que as soluções convencionais (LEK et al., 1996; BRAGA et al., 2000; HAYKIN, 2001; GORGENS, 2006). Outras características importantes das RNA são: mapeamento de entrada e saída; tolerância a falhas (neurônios e, ou, conexões danificados) e ruídos; adaptabilidade dos pesos das conexões às modificações do ambiente, ou seja, uma rede treinada pode ser re-treinada ou uma rede pode ser projetada para modificar seus pesos em tempo real (ambiente não estacionário); capacidade de informar a confiança de uma decisão tomada (probabilidade); aprendizagem a partir de exemplos e generalização para dados desconhecidos; e analogia neurobiológica (HAYKIN, 2001; BULLINARIA, 2009; GORGENS, 2015).

Os problemas tratáveis por meio de redes neurais artificiais enquadram-se nas categorias de aprendizagem, como, aproximação de funções, classificação de padrões, agrupamento de dados, predição (séries temporais), otimização, recuperação de conteúdo e controle (JAIN et

al., 1996; HAYKIN, 2001). A prognose da produção florestal se enquadra na tarefa de aproximação de função que consiste em projetar uma rede neural que aproxime a função desconhecida $f(x)$, que descreve o mapeamento dos pares de entrada-saída $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ de um conjunto de n padrões de treinamento.

Dentre as diversas aplicações de redes neurais artificiais, podem ser citados o controle de tráfego aéreo pode ser automatizado com a localização, altitude, direção e velocidade de cada radar utilizado como entrada para a rede (KUMAR et al., 2015); o comportamento animal e os ciclos populacionais podem ser adequados para análise por redes neurais; avaliação e valorização de propriedades, construções, automóveis, entre outros (OZÇELIK et al., 2013); o padrão de eco dos instrumentos sonar, radar e magnéticos pode ser usado para prever seus alvos; previsão do tempo (KUMAR et al., 2015); reconhecimento de voz pode ser obtido através da análise do padrão de áudio; fluxos de tráfego podem ser previstos de modo que o tempo do sinal possa ser otimizado (NAGHDI; GHAJAR, 2012); e diversas outras aplicações.

Nos últimos anos, diversos estudos no setor florestal têm testado e aplicado as redes neurais artificiais como uma ferramenta de modelagem e prognose do crescimento e produção florestal. Para Diamatopoulou (2005), a qualidade das estimativas obtidas por meio das RNA se deve à capacidade destas em modelar diversas variáveis e superar certos problemas encontrados em dados florestais, como relações não-lineares, distribuições não gaussianas, *outliers* e falhas (ou ausências) nos dados.

Binoti et al. (2013), em estudo sobre o efeito da redução das medições de altura na precisão obtida pelas RNA, avaliaram as estimativas obtidas reduzindo-se o número de parcelas com alturas medidas e concluíram que é possível reduzir o número de medições sem perda de precisão. Ainda segundo os autores, é possível reduzir o custo do inventário por meio da aplicação da RNA na estimativa das alturas das árvores. Lafeté et al. (2018) destacaram a eficiência das redes neurais artificiais na modelagem de biomassa de eucalipto sob diferentes idades e espaçamentos. Vieira et al. (2018), ao aplicarem redes neurais artificiais em dados contínuos de inventários florestais de eucalipto, observaram exatidão nas estimativas de crescimento em diâmetro e altura.

2.2 Modelos mistos

Em um processo de modelagem, procura-se explicar a variabilidade de uma variável dependente por meio dos efeitos que se atribuem a um outro conjunto de variáveis, denominadas independentes. Estes efeitos podem ser considerados como fixos, que são

constantes a serem estimadas, ou aleatórios, que são ocorrências de uma variável aleatória com distribuição de probabilidade conhecida.

Segundo Searle (1987), modelos lineares nos parâmetros possuem ao menos um efeito aleatório, conhecido por erro experimental. Se um modelo apresenta todos os demais componentes fixos é chamado de modelo fixo; se, no entanto, todos os demais fatores forem aleatórios o modelo é chamado de aleatório; quando o modelo apresenta tanto efeitos aleatórios quanto fixos, é denominado de modelo misto. Os efeitos aleatórios podem ser definidos como sendo aqueles que ao se retirar uma amostra aleatória de um conjunto infinito de níveis, as inferências estatísticas são válidas para toda a população e os efeitos fixos são aqueles atribuídos a um conjunto finito de níveis de tratamentos em teste no experimento (SEARLE; CASELLA; MCCULLOCH, 1992).

Existem situações em que os dados possuem comportamentos dependentes entre si, ou seja, uma observação depende de outra. De maneira geral, são chamados de dados agrupados, que incluem dados longitudinais, medidas repetidas e outros. Neste sentido, os modelos usuais, tanto lineares como não-lineares, tornam-se inadequados, uma vez que a pressuposição de independência das observações fica comprometida (BARBOSA, 2009).

As técnicas de ajuste têm permitido cada vez mais flexibilidade aos modelos com a presença de dados agrupados, levando em consideração a possível correlação existente entre as observações, sendo possível modelar as correlações entre e intragrupos. Tais modelos são denominados mistos, em que a parte aleatória é incorporada por meio de uma matriz de variância e covariância (LITTEL et al., 1996).

2.2.1 Modelos lineares mistos

Os modelos lineares mistos são fundamentados em três aspectos: ajustes e teste de hipóteses sobre os parâmetros de efeito fixo; predição dos parâmetros de efeito aleatório; e estimação dos componentes de variância. A vantagem desse método está fortemente associada ao exame dos efeitos aleatórios e à modelagem da estrutura de variância e covariância (LAIRD; WARE, 1982; CAMARINHA FILHO, 2003; CUNHA et al., 2013).

A abordagem de forma matricial de modelos lineares mistos pode ser representada, de acordo com Mrode (2015), por:

$$y = X\beta + Zb + \varepsilon \quad (1)$$

em que: y é um vetor $n \times 1$ de observações da variável resposta ou variável dependente, sendo n o número de observações; X é uma matriz $n \times p$ de incidência dos efeitos fixos, sendo p o número de parâmetros de efeitos fixos; β é um vetor $p \times 1$ de parâmetros de efeitos fixos; Z é uma matriz $n \times q$ de incidência dos efeitos aleatórios, sendo q o número de parâmetros de efeitos aleatórios; b é um vetor $q \times 1$ de parâmetros dos efeitos aleatórios; ε é um vetor $n \times 1$ de erros aleatórios, com:

$$b \sim N(0, G) \text{ e } \varepsilon \sim N(0, R) \quad (2)$$

em que G é a matriz de variâncias - covariâncias dos efeitos aleatórios e R é a matriz de variâncias - covariâncias residuais, as quais podem assumir diferentes estruturas.

Desse modo tem-se,

$$y \sim N(X\beta, ZGZ' + R) \quad (3)$$

2.2.2 Modelos não-lineares mistos

Os modelos não lineares de efeitos mistos têm recebido grande atenção devido à sua capacidade de trabalhar com dados que possuem medidas repetidas e não balanceadas. Nesta classe de modelos é permitido que alguns parâmetros sejam fixos enquanto outros apresentem variações entre os diferentes grupos, por meio dos efeitos aleatórios. Além disso, estes modelos conseguem modelar a correlação existente entre as observações, dentro de um mesmo grupo, que é frequentemente observada nesse tipo de dados. Os clássicos modelos de regressão assumem independência das observações, logo, não são adequados para dados agrupados, além de estimarem incorretamente as estimativas dos erros padrão em casos onde ocorre agrupamento (LITTELL; HENRY; AMMERMAN, 1998).

De acordo com Lindstrom e Bates (1990), a abordagem de forma matricial de modelos não-lineares mistos pode ser representada por:

$$y_{ij} = f(\phi_{ij}, V_{ij}) + \varepsilon_{ij}, i = 1, \dots, m; j = 1, \dots, n_i. \quad (4)$$

em que: f é uma função geral, real e diferenciável de um grupo específico de vetores do parâmetro ϕ_{ij} e um vetor de covariantes v_{ij} ; e ε_{ij} é o erro aleatório normalmente distribuído

dentro dos grupos; m é o número de grupos; n_i é o número de observações no i -ésimo grupo; n_j é o número de observações no j -ésimo grupo.

O vetor de parâmetros varia de indivíduo para indivíduo. Em um segundo estágio, o vetor ϕ_{ij} pode ser expresso pela equação 5.

$$\phi_{ij} = A_{ij}\beta_i + B_{ij}\beta_i + B_{ij}\beta_{ij}, \beta_i \approx N(0, \Psi_1) \text{ e } \beta_{ij} \approx N(0, \Psi_2) \quad (5)$$

em que: β é um vetor ($p \times 1$) de efeitos fixos; B_i é um vetor ($q_1 \times 1$) de efeitos aleatórios independentemente distribuídos com uma matriz de covariância-variância ψ_1 ; B_{ij} é um vetor ($q_2 \times 1$) de efeitos aleatórios independentemente distribuídos com uma matriz de covariância-variância ψ_1 , e pressuposto ser independente dos efeitos aleatórios do primeiro nível; A_{ij} e B_{ij} são matrizes de incidência; ε_{ij} , dentro dos grupos, são independentemente distribuídos e são independentes dos efeitos aleatórios.

2.2.3 Estruturas de variância e covariância

O processo de definição da estrutura de variância e covariância adequada é importante na modelagem dos modelos mistos, pois visa selecionar uma estrutura parcimoniosa, que explique bem a variabilidade dos dados e a correlação entre as medidas com um número pequeno de parâmetros. Essa seleção pode afetar diretamente as estimativas dos parâmetros, os erros padrão de efeitos fixos e aleatórios, diagnósticos e inferências (TORAL; ALENCAR; FREITAS, 2006) e é dependente da forma como os dados estão estruturados, de informações empíricas e da disponibilidade computacional. As modelagens dessas estruturas para o modelo linear misto e não linear misto são semelhantes e mais detalhes podem ser encontrados na literatura, como em Davidian e Giltinan (1995) e Pinheiro e Bates (2000).

A Tabela 1 apresenta algumas estruturas de covariâncias para G e R, comumente utilizadas.

Tabela 1 – Estruturas da matriz de variância e covariância. (Continua)

Estrutura	Exemplo
Componente de variância (CV)	$\Psi = \begin{vmatrix} \sigma_A^2 & 0 & 0 & 0 \\ 0 & \sigma_A^2 & 0 & 0 \\ 0 & 0 & \sigma_B^2 & 0 \\ 0 & 0 & 0 & \sigma_B^2 \end{vmatrix}$
Simétrica composta (SC)	$\Psi = \begin{vmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{vmatrix}$
Simétrica composta heterogênea (SCH)	$\Psi = \begin{vmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{vmatrix}$
Autorregressiva de primeira ordem AR(1)	$\Psi = \sigma^2 \begin{vmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{vmatrix}$
Autorregressiva de primeira ordem heterogênea ARH(1)	$\Psi = \begin{vmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{vmatrix}$
Não estruturada (NE)	$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$

Tabela 1 – Estruturas da matriz de variância e covariância. (Conclusão)

Estrutura	Exemplo
Huynh-Feldt (HF)	$\begin{pmatrix} \sigma_1^2 & \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_2^2 + \sigma_1^2}{2} - \lambda & \sigma_2^2 & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_3^2 + \sigma_1^2}{2} - \lambda & \frac{\sigma_3^2 + \sigma_2^2}{2} - \lambda & \sigma_3^2 & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_4^2 + \sigma_1^2}{2} - \lambda & \frac{\sigma_4^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_4^2 + \sigma_3^2}{2} - \lambda & \sigma_4^2 \end{pmatrix}$
Toeplitz (TOEP)	$\Psi = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$
Toeplitz heterogênea (TOEP)	$\Psi = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho_1 & \sigma_1 \sigma_3 \rho_2 & \sigma_1 \sigma_4 \rho_3 \\ \sigma_2 \sigma_1 \rho_1 & \sigma_2^2 & \sigma_2 \sigma_3 \rho_1 & \sigma_2 \sigma_4 \rho_2 \\ \sigma_3 \sigma_1 \rho_2 & \sigma_3 \sigma_2 \rho_1 & \sigma_3^2 & \sigma_3 \sigma_4 \rho_1 \\ \sigma_4 \sigma_1 \rho_2 & \sigma_4 \sigma_2 \rho_2 & \sigma_4 \sigma_3 \rho_1 & \sigma_4^2 \end{bmatrix}$

Fonte: Adaptado de West, Welch e Gatecki, (2015).

em que: σ é a variância; ρ é o parâmetro relacionado à covariância e representa a correlação entre as medições; λ é a diferença entre a média das variâncias e a média das covariâncias.

As aplicações dos modelos não lineares mistos podem ser diversas. Em povoamentos florestais, podem ser utilizados para a estimativa e predição do crescimento em altura (ADAME et al., 2008; CALEGARIO et al., 2005; CRECENTE-CAMPO et al., 2010; MENDONÇA; CARVALHO; CALEGARIO, 2015), estimativa e predição do crescimento em diâmetro (TIMILSINA; STAUDHAMMER, 2013), para avaliar o perfil longitudinal do fuste da árvore (CARVALHO et al., 2014; PIRES; CALEGARIO, 2007), entre outras.

2.3 Modelo Geoestatístico

A necessidade do desenvolvimento e utilização de ferramentas estatísticas, capazes de caracterizar a estrutura de variabilidade espacial de um determinado fenômeno, motivou o surgimento da teoria das variáveis regionalizadas ou, simplesmente, geoestatística (NOGUEIRA, 2013). De acordo com Couto et al. (1997), a geoestatística tem grande utilidade para caracterizar e mapear a variação espacial das propriedades dos sistemas.

Segundo Landim (2003), as raízes da geoestatística estão na indústria de minérios, em meados da década de 50, durante pesquisas desenvolvidas para cálculo de reservas minerais na

África do Sul por Daniel G. Krige e o estatístico H. S. Sichel, para dados de mineração do ouro. Concluiu-se que as variâncias dos dados possuíam uma estruturação que dependia da amostragem, e que as variâncias obtidas através da abordagem clássica de amostragem, não faziam sentido se não considerasse as distâncias entre as amostras. Baseado nessas observações, Matheron (1963) desenvolveu a Teoria das Variáveis Regionalizadas, definida como variáveis cujos valores estão relacionados de algum modo, com a posição espacial que ocupam, ou seja, que variam de um lugar a outro com certa continuidade.

Esta variável atua num espaço geométrico em que foi definida e onde será estudada sua variação, apresentando localização, continuidade e anisotropia, que estão ligadas ao fenômeno natural que elas representam (JOURNAL, 2003). Nogueira (2013) afirma que as variáveis regionalizadas apresentam dupla característica: aleatoriedade, uma vez que seus valores numéricos observados podem variar, consideravelmente, de um ponto a outro no espaço e, espacialidade, pois estes valores não são inteiramente independentes, apresentando uma aparente continuidade espacial, na qual podem ser captados e modelados matematicamente por técnicas geoestatísticas.

Segundo Mello (2004), a teoria das variáveis regionalizadas pressupõe que cada dado $Z(x_i)$ é modelado com uma variável aleatória que pode ser expressa pela soma de três componentes: um componente estrutural, associado a um valor médio constante ou a uma tendência constante; uma componente aleatória, espacialmente correlacionada; e um erro aleatório constante.

Se x representa uma posição em uma, duas ou três dimensões, então o valor da variável Z em x é dado por:

$$Z(x_i) = M(x) + \varepsilon'_x + \varepsilon_i \quad (6)$$

em que $M(x)$ é uma função determinística que descreve a componente estrutural Z em x ; ε'_x é um termo estocástico, que varia localmente e depende espacialmente de $M(x)$; e ε_i é um ruído aleatório não correlacionado, com distribuição normal com média igual a zero e variância σ^2 , atuando em um termo estocástico ε'_x (MELLO, 2004).

Para Vieira (2000), a geoestatística tem por objetivo identificar, na aparente desordem entre as amostras, uma medida da correção espacial entre os atributos, permitindo avaliar padrões de amostras adequadas e realizar estimativas de valores locais não amostrados com

base em alguns valores conhecidos na população, técnica denominada por estudiosos como Krigagem. Para Silva et al. (2010), esta técnica trata-se de um método de interpolação da geoestatística, o qual usa dependência espacial expressa no variograma entre amostras vizinhas para estimar valores em qualquer posição dentro do campo, sem tendência e com variância mínima.

De acordo com Isaaks e Srivastava (1989), a geoestatística fundamenta-se na esperança de que, na média, as amostras próximas, no tempo e no espaço, sejam mais similares entre si do que as que estiverem distantes. Trata-se de uma ferramenta de análise espacial que possui como finalidade estimar valores não amostrados, utilizando modelos probabilísticos associados à localização espacial dos pontos amostrados possibilitando, desta forma, uma análise da variabilidade entre os valores observados (MELLO, 2004).

A geoestatística vem se desenvolvendo nas últimas décadas e apresenta-se como uma ferramenta confiável para a estimativa de dados para a grande maioria dos fenômenos naturais, uma vez que se distingue por capturar o padrão de continuidade dos dados geoposicionados (SILVA et al., 2010).

A preocupação com técnicas e estimadores mais eficientes, que determinem as características de uma área, ou uma porção da superfície terrestre, são alvos de estudo de várias áreas do conhecimento (WOICIECHOWSKI et al., 2009).

A aglomeração de métodos estatísticos presentes na geoestatística ajudam a descrever e analisar dados que incluem medidas simultâneas de muitas variáveis correlacionadas, explicando a complexidade de um sistema biológico (NEVES, 2013), pois possuem grande utilidade na caracterização e mapeamento da variação espacial das propriedades dos sistemas (COUTO et al., 1997). As técnicas da geoestatística e da estatística clássica possibilitam o armazenamento, o tratamento, a sobreposição, a análise e a visualização, em conjunto, dessas informações espaciais, sob a forma de mapas digitais, permitindo tomadas de decisões rápidas, ágeis e com maior eficiência, minimizando custos, otimizando os recursos e as atividades produtivas.

2.3.1 Semivariograma

Segundo Correia (2013), o semivariograma é uma ferramenta geoestatística que permite verificar e modelar a dependência espacial de uma variável, pois permite representar, quantitativamente, a variação de um fenômeno regionalizado no espaço. Segundo Salviano (1996), o semivariograma analisa o grau de dependência espacial entre as amostras dentro de

um campo experimental, e define parâmetros necessários para a estimativa de valores para locais não amostrados, por meio da técnica de krigagem. Zimback (2011) afirma ser possível investigar a magnitude da correlação entre as amostras e sua similaridade ou não com a distância.

Para Amaral (2010), o semivariograma descreve matematicamente o relacionamento entre a variância de pares de observações (pontos) e a distância que separa estas observações (h), em que o semivariograma depende da distância h entre x e $x+h$ (VIEIRA, 2000).

De acordo com Maciel (2012), o estimador de semivariância $\hat{\gamma}(h)$ (7) é igual a metade da média aritmética das diferenças entre pares de valores experimentais ao quadrado, em todas os pontos separados pela distância h . Nesse estimador, $N(h)$ representa o número de pares separados por uma distância h , $Z(x)$ é o valor da variável observado no ponto x e $Z(x+h)$ é o valor da variável medida a uma distância h do ponto x (JOURNEL; HUIJBREGTS, 1978).

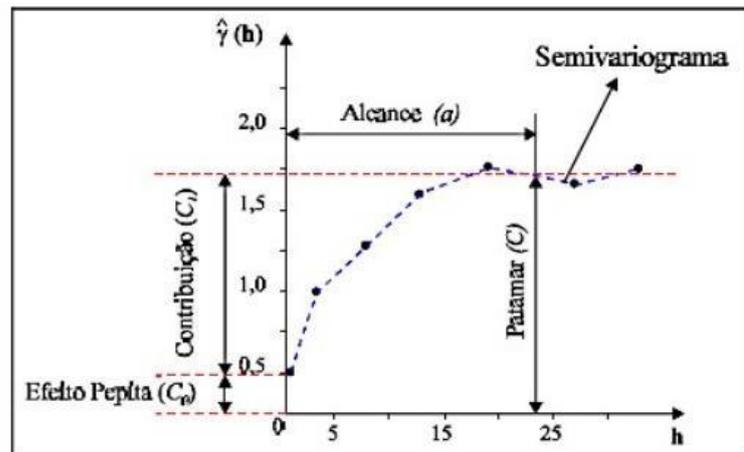
$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (7)$$

em que: $\gamma(h)$ é a semivariância estimada; $N(h)$ é o número de pares de valores medidos separados por um vetor h [$Z(x_i) - Z(x_i+h)$].

De acordo com Vieira (2000), as características de um semivariograma ideal ocorrem quando a diferença [$Z(x_i) - Z(x_i+h)$] diminui à medida que o valor de h decresce. Quando a distância é nula ($h = 0$), o valor da semivariância será zero ($\gamma = 0$). Todavia, isso não ocorre, visto que à medida em que a distância diminui o valor da semivariância tende a um valor positivo conhecido como efeito pepita (MACIEL, 2012).

De acordo com Cressie (1993), os parâmetros do semivariograma (FIGURA 4) que auxiliam na descrição da dependência espacial são o alcance, patamar, efeito pepita e contribuição.

Figura 4 - Exemplo de semivariograma.



Fonte: Camargo (2001).

em que:

Alcance (a): representa a distância em metros em que as amostras se apresentam correlacionadas espacialmente (WOJCIECHOWKI, 2006). Segundo Guerra (1988), corresponde ao conceito da zona de influência ou de dependência espacial de uma amostra, marcando a distância a partir da qual as amostras tornam-se independentes.

Patamar (c): é o valor do semivariograma correspondente ao seu alcance (AMARAL, 2010). É uma estimativa da variância da população, e desse ponto em diante considera-se que não existe mais dependência espacial entre as amostras, pois a variância da diferença entre pares de amostras torna-se invariante com a distância (AMARAL, 2010).

Efeito pepita (C_0): representa a descontinuidade do semivariograma para escalas menores do que a menor distância entre as amostras (XAVIER, 2010). Parte desta descontinuidade pode ser também devida a erros de medição, contudo, é impossível quantificar se a maior contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem (WOJCIECHOWKI, 2006).

Contribuição (C_1): é a diferença entre o Patamar (C) e o Efeito Pepita (C_0) (CAMARGO, 2001).

Os estimadores geoestatísticos podem ser classificados em três tipos: observado ou experimental (obtido a partir das amostras de campo), o verdadeiro (real, mas desconhecido) e o teórico (de referência, utilizado para o ajuste do modelo) (VIEIRA, 2000).

O semivariograma experimental é ajustado à curva que proporciona a máxima correlação possível com os pontos plotados, enquanto que o modelo ajustado é chamado de

teórico (WOJCIECHOWKI, 2006). De acordo com Vieira (2000), é importante que o modelo de semivariograma ajustado represente a tendência das semivariâncias calculadas, de tal forma que as estimativas obtidas a partir da Krigagem sejam mais exatas e, portanto, mais confiáveis.

Zimback (2001) afirma que a sensibilidade dos semivariogramas, com a finalidade de detectar a variabilidade espacial das amostras, está diretamente ligada ao melhor ajuste dos dados experimentais ao modelo teórico do variograma. Logo, dentre as ferramentas geoestatísticas o ponto central é o semivariograma, uma vez que este descreve variações quantitativas e interpola os pontos não amostrados. É importante escolher o modelo apropriado para a estimativa do semivariograma devido ao fato de cada modelo apresentar valores diferentes para alcance, efeito pepita e variância, sendo estes parâmetros críticos para a qualidade da krigagem (RIBEIRO JUNIOR, 1995).

2.3.1.1 Modelos teóricos de semivariograma

A sensibilidade dos semivariogramas para detectar a variabilidade espacial das amostras está diretamente ligada ao melhor ajuste de modelo teórico ao semivariograma empírico. Segundo Journel e Huijbregts (2003), os modelos teóricos de semivariogramas são superpostos à sequência de pontos obtidos no semivariograma empírico, de modo que a curva que melhor se ajusta aos pontos representa a magnitude. Wojciechowki (2006) afirma que dentre os modelos básicos, denominados isotrópicos por Isaaks e Srivastava (1989), são: modelos esférico (8); exponencial (9); e gaussiano (10).

$$\gamma_{\text{sph}}(h) = C_0 + C_1 \left[\frac{3}{2} \left(\frac{h}{\phi} \right) - \left(\frac{1}{2} \left(\frac{|h|}{\phi} \right)^3 \right) \right] \quad (8)$$

$$\gamma_{\text{exp}}(h) = C_0 + C_1 \left[1 - \exp \left(-3 \left(\frac{|h|}{\phi} \right) \right) \right] \quad (9)$$

$$\gamma_{\text{gauss}}(h) = C_0 + C_1 \left[1 - \exp \left(-3 \left(\frac{|h|}{\phi} \right)^2 \right) \right] \quad (10)$$

em que: γ = semivariância da variável de interesse; C_0 = efeito pepita; C = patamar; ϕ = alcance; h = distância; exp = exponencial.

2.3.2 Método de estimativa espacial: Krigagem

A krigagem é o método de interpolação da geoestatística, o qual usa a dependência espacial expressa no semivariograma entre amostras vizinhas para estimar valores em qualquer posição dentro do campo, sem tendência e com variância mínima (SILVA et al., 2010). Estas duas características fazem da krigagem um interpolador ótimo (ISAACKS; SRIVASTAVA, 1989). De acordo com Correia (2013), este método de inferência espacial estima dados em pontos não amostrados a partir de pontos amostrados, levando em consideração a estrutura de dependência espacial do fenômeno em estudo.

Segundo Soares (2009), existem vários interpoladores para estimar valores em pontos não amostrados, como krigagem simples, krigagem ordinária, krigagem universal, krigagem indicadora. Akhavan et al. (2010) afirmam que esses métodos constituem a âncora dos procedimentos geoestatísticos.

Maciel (2012) afirma que na krigagem simples as médias locais são constantes e com valor semelhante à média populacional, ou seja, a média da população é utilizada em conjunto com os pontos vizinhos para a estimativa do atributo. A Krigagem ordinária, de acordo com Motomiya et al. (2006), é o método de interpolação de dados, que usa a dependência espacial entre as amostras vizinhas, expressa no semivariograma, para estimar valores em qualquer posição dentro do espaço analisado, ao qual o modelo do semivariograma foi ajustado, sem tendência e com variância mínima. Em suma, esse estimador nada mais é do que uma média pondera dos valores observados. Enquanto que a krigagem universal é um método eficiente para os casos onde não ocorre a estacionariedade, ou seja, o processo apresenta uma tendência. Na krigagem ordinária desconsidera-se a média populacional e utiliza somente os pontos vizinhos para a estimativa, além de considerar a estacionariedade dos dados (MACIEL, 2012).

Desta forma a krigagem é uma técnica usada na geoestatística com o objetivo de estimar valores de variáveis para locais onde as mesmas não foram medidas a partir de valores adjacentes interdependentes, enfatizando-se que para que esta ferramenta seja usada é necessário que exista a dependência espacial definida pelo variograma (SALVIANO, 1996).

REFERÊNCIAS

- ADAME, P.; DEL-RÍO, M.; CAÑELLAS, I. A mixed nonlinear height-diameter model for pyrenean oak (*Quercus pyrenaica* Willd.). **Forest Ecology and Management**, Amsterdam, v. 256, n. 1/2, p. 88-98, 2008.
- AKHAVAN, R.; ZAHEDI AMIRI, G.; ZOBEIRI, M. Spatial variability of forest growing stock using geostatistics in the Caspian region of Iran. **Caspian Journal of Environmental Sciences**, v. 8, n. 1, p. 43-53, 2010.
- AMARAL, L. de P. **Geoestatística na caracterização do solo e da vegetação em Floresta Ombrófila Mista**. 2010. 154 p. Dissertação (Mestrado em Agronomia) – Universidade Estadual do Centro-Oeste, Guarapuava-PR, 2010.
- BARBOSA, M. **Uma abordagem para análise de dados com medidas repetidas utilizando modelos lineares mistos**. 2009. 119 p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) – Universidade de São Paulo, São Paulo, 2009.
- BHERING, L. L. et al. Application of neural networks to predict volume in eucalyptus. **Crop Breeding And Applied Biotechnology**, [S.L], v. 15, n. 3, p. 125-131, 2015.
- BINOTI, M. L. M. DA S. et al. Aplicação de Redes Neurais Artificiais para Estimativa da Altura de Povoamentos Equiâneos de Eucalipto. **Revista Árvore**, v. 37, n. 4, p. 639–645, 2013.
- BRAGA, A. de P.; CARVALHO, A. P. de L. F. de; LUDEMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro, RJ. Editora LTC, 262 p. 2000.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDEMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro: Editora LTC, 2007. 226p.
- BULLINARIA, J. A. **Introduction to Neural Computation**. Notas de aula. 2009. Disponível em: <<http://www.cs.bham.ac.uk/~jxb/inc.html>>. Acesso em: 20 de nov. 2017.
- CALEGARIO, N. et al. Estimativa do crescimento de povoamentos de *Eucalyptus* baseada na teoria dos modelos não lineares em multinível de efeito misto. **Ciência Florestal, Santa Maria**, v. 15, n. 3, p. 285-292, 2005.
- CAMARINHA FILHO, J. A. Utilização de Modelos Lineares Mistos em Agronomia. **Revista de Matemática e Estatística**. v. 21, p. 34-43, 2003.
- CAMARGO, E. C. G. Geoestatística: Fundamentos e aplicações. INPE: São José dos Campos, 2001. Disponível: <http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/5geoest.pdf>. Acesso em 25 de jan. de 2018.
- CARVALHO, S. P. C. et al. Modelagem não linear mista para descrever o afilamento de árvores clonais de *Eucalyptus* sp. **Scientia Forestalis**, Piracicaba, v. 42, n. 104, p. 605-614, dez., 2014.

CORREIA, M. D. **Modelagem geoestatística da distribuição de carbono do solo e biomassa de herbáceas em sistema silvipastoril**. 2013. 101 p. Dissertação (Mestrado em Biometria e estatística aplicada) – Universidade Rural de Pernambuco, Recife, 2013.

COUTO, E.G., STEIN, A.; KLAMT, E. Large area spatial variability of soil chemical properties in central Brazil. **Agriculture Ecosystems and Environment**, v. 66, p. 139- 152, 1997.

CRECENTE-CAMPO, F. et al. A generalized nonlinear mixed-effects height–diameter model for *Eucalyptus globulus* L. in northwestern Spain. **Forest Ecology and Management**, Amsterdam, v. 259, p. 943–952, 2010.

CRESSIE, N. **Statistics for spatial data**. New York: Wiley, 1993. 900p.

CUNHA, T. A; FINGER, C. A. G.; SCHNEIDER, P. R. Linear mixed model to describe the basal area increment for individual cedro (*Cedrela odorata* L.) trees in occidental Amazon, Brazil. **Ciência Florestal**, Santa Maria, v. 23, n. 3, p. 461-470, 2013

DAVIDIAN, M.; GILTINAN, D. M. **Nonlinear models for repeated measurement data**. New York: Chapman and Hall, 1995. 360 p.

DIAMANTOPOULOU, M. J. Artificial neural networks as an alternative tool in pine bark volume estimation. **Computers and electronics in agriculture**, v. 10, p. 235-244, 2005.

GORGENS, E. B. **Estimação do volume de árvores utilizando redes neurais artificiais**. 2006. 84 p. Dissertação (Mestrado em Ciência Florestal). Universidade Federal de Viçosa, Viçosa, MG, 2006.

GORGENS, E. B.; MONTAGHI, A.; RODRIGUEZ, L. C. E. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. **Computers and Electronics in Agriculture**, v. 116, p. 221-227, 2015.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. 900 p.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied geostatistics**. New York: Oxford University Press, 1989.

JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial Neural Network: A Tutorial. **Computer**, v. 29, n. 3, p. 31-44, 1996.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London: Academic, 1978. 600 p.

JOURNEL, A. G.; HUIJBREGTS, C.J. **Mining geostatistics**. London: Academic, 2003. 600 p.

KUMAR, S.; DESPANDE, S. M., DAS, S. S.; KONWAR, M.; CHAKRAVAERTY, K.; KALAPUREDDY, M. C. R. Temporal and structural evolution of a tropical monsoon cloud

system: A case study using X-band radar observations. **Journal of atmospheric and solar-terrestrial physics**. v. 133, p. 157-168, 2015.

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

LANDIM, P. M. B. **Análise estatística de dados geológicos**. 2. Ed. Ver e ampl. São Paulo: Editora UNESP, 2003.

LEK, S. et al. Application of neural networks to modelling nonlinear relationships in ecology. **Ecological Modelling**, v. 90, p. 39-52, 1996.

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

LITTEL, R. C. et al. **SAS system for mixed models**. Cary: SAS Institute, 1996. 633p.

LITTELL, R. C.; HENRY, P. R.; AMMERMAN, C.B. Statistical analysis of repeated measures using SAS procedures. **Journal of Animal Science**, v. 76, p. 1216-1231, 1998.

MACIEL, S. M. **Análise espacial do carbono em um fragmento florestal com predominância de *Anadenanthera* sp.** 2012. 135 p. Dissertação (Mestrado em Tecnologia da Madeira) – Universidade Federal de Lavras, Lavras. 2012.

MATHERON, G. Principles of geostatistics. **Economic Geology**. El paso, v. 58, p. 11246-66, 1963.

MCCULLOCH S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115-133, 1943.

MELLO, J. M. de. **Geoestatística aplicada ao inventário florestal**. 2004. 122 p. Tese (Doutorado em Recursos Florestais) – Universidade de São Paulo, Piracicaba. 2004.

MENDONÇA, A. R.; CARVALHO, S. P. C.; CALEGARIO, N. Modelos hipsométricos generalizados mistos na predição da altura de *Eucalyptus* sp. **Cerne**, Lavras, v. 21, n. 1, p. 107-115, 2015.

MÔRA, R. **Funções de afilamento de forma variável e modelagem de efeitos mistos em fustes de *Pinus taeda* e *Eucalyptus saligna***. 277 p. Tese (Doutorado em Engenharia Florestal) - Universidade Federal do Paraná, Curitiba, 2015.

MOTOMIYA, A. V. de A.; CORÁ, J. E.; PEREIRA, G. T. Uso da krigagem indicatriz na avaliação de indicadores de fertilidade do solo. **Revista Brasileira de Ciência do Solo**, n. 30, p. 485-496, 2006.

MRODE, R.A. **Linear models for the prediction of animal breeding values**. 3 ed. Wallingford: CAB International, 2015. 343 p.

NAGHDI, R.; GHAJAR, I. Application of Artificial Neural Network in the Modeling of Skidding Time Prediction. **Advanced Materials Research**. v. 403-408, n. 2012, p. 3538-3543, 2012. doi:10.4028/www.scientific.net/AMR.403-408.3538

NEVES, D. A. **Geoestatística aplicada ao estudo da variabilidade de espécies nativas em Fragmentos do Cerrado Brasileiro no Estado de São Paulo**. 2013. 172 p. Tese (Doutorado em Ciências da Terra e Meio Ambiente) – Universidade de A Coruña, Espanha. 2013.

NOGUEIRA, C. H. **Análise de variância com dependência espacial sob uma abordagem geoestatística**. 2013. 124 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2013.

ÖZÇELİK, R. et al. Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. **Forest Ecology and Management**, v. 306, p. 52–60, 2013.

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

PIRES, L. M.; CALEGARIO, N. Ajuste de modelos estocásticos lineares e não-lineares para a descrição do perfil longitudinal de árvores. **Revista Árvore**, Viçosa, v. 31, n. 5, p. 845-852, 2007.

RIBEIRO JÚNIOR, P. J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99 p. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiros”. 1995.

SALVIANO, A.A.C. **Variabilidade de atributos de solo e de *Crotalaria juncea* em solo degradado do município de Piracicaba-SP**. 1996. 91 p. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, Universidade de São Paulo. 1996.

SEARLE, S.R. **Linear models for unbalanced data**. New York: John Wiley, 1987. 536 p.

SEARLE, S.R.; CASELLA, G.; MCCULLOCH, C. **Variance components**. New York: John Wiley, 1992. 501 p.

SILVA, A. F. da.; ZIMBACK, C. R. L.; OLIVEIRA, R. B. de. Cokrigagem na estimativa da evapotranspiração em Campinas (SP). **Tékhn e Lógos**, Botucatu, São Paulo, v. 2, n. 1, 2010.

SOARES, P. **Levantamento fitossociológico de regeneração natural em reflorestamento misto no noroeste de Mato Grosso**. 2009. 50 p. Dissertação (Mestrado em Ciências Florestais e Ambientais) – Universidade Federal de Mato Grosso – Cuiabá. 2009.

TIMILSINA, N.; STAUDHAMMER, C. L. Individual tree-based diameter growth model of slash pine in Florida using nonlinear mixed modeling. **Forest Science**, Lawrence, v. 59, n. 1, p. 27-37, fev. 2013.

TORAL, F.L.B.; ALENCAR, M.M.; FREITAS, A.R. Arranjos para efeitos fixos e estruturas de (co)variâncias residuais para análises de medidas repetidas do peso de bovinos da raça Canchim. **Revista Brasileira de Zootecnia**, v. 35, n. 5, p. 1951-1958, 2006.

VIEIRA, S. R. **Geoestatística em estudos de variabilidade espacial**. 2000.

WEST, B. T.; WELCH, K. B.; GATECKI, A. T. **Linear mixed models: a practical guide using statistical software**. 2 ed. Ann Arbor: Chapman and Hall/CRC, 2015. 405 p.

WOJCIECHOWSKI, J. C. **Gesestatística aplicada ao estudo das características físico – químicas do solo em áreas de floresta estacional decidual**. 2006. 103 p. Dissertação (Mestrado em Geomática) – Universidade Federal de Santa Maria, Santa Maria, 2006.

WOJCIECHOWSKI, J. C. et al. Geoestatística aplicada ao estudo físico – químico do solo em áreas de floresta estacional decidual. **Ciência Florestal**, Santa Maria, v. 19, n. 4, p. 383-391, 2009.

XAVIER, A. C.; CECÍLIO, R. A.; LIMA, J. S. S. Módulos em *Matlab* para interpolação espacial pelo método da krigagem ordinária e do inverso da distância. **Revista Brasileira de Cartografia**, n 62/01, 2010.

ZIMBACK, C. R. L. **Análise especial de atributos químicos de solo para fins de mapeamento da fertilidade**. 2001. 114 p. Tese (Livre Docência em Levantamento do solo e fotopedologia), FCA, Botucatu, 2001.

SEGUNDA PARTE – ARTIGOS

**ARTIGO 1 - ACCURACY OF TREE HEIGHT ESTIMATION WITH ARTIFICIAL
NEURAL NETWORK MODELS AND REGRESSION MODELS**

**ARTIGO FORMATADO DE ACORDO COM AS NORMAS DA REVISTA
CANADIAN JOURNAL OF FOREST RESEARCH**

ABSTRACT

The height variable is commonly used as input attribute to estimate other variables. In order to guarantee less susceptibility to errors, it is necessary to obtain it correctly. The objective of this study was to propose and evaluate the performance of a model extracted from artificial neural network training and new regression models to estimate the total height of eucalyptus trees. The data used came from temporary forest inventories carried out in eucalyptus stands in Minas Gerais, Brazil. Artificial neural networks of the Multilayer Perceptron type were trained and, from the best performance network, a nonlinear equation was extracted to predict the total height of trees. Hypsometric models, both linear and nonlinear, were constructed and fitted, using regressors related to the individual tree (DAP) and stand (basal area of plot, age and site index). The new hypsometric models proposed in this study presented satisfactory performance and are efficient for estimating the total height of eucalyptus trees. As a highlight, the model extracted from the artificial neural network and the nonlinear regression model.

Keywords: Hypsometric relationship; Forest Inventory; eucalyptus.

RESUMO

A variável altura é comumente utilizada como atributo de entrada para estimativa de outras variáveis, a fim de propiciar menor susceptibilidade a erros faz-se necessária a sua correta obtenção. O objetivo deste estudo foi propor e avaliar o desempenho de um modelo extraído do treinamento da rede neural artificial e de novos modelos para estimar a altura total de árvores de eucalipto. Os dados utilizados neste estudo foram provenientes de inventários florestais temporários realizados em povoamentos de eucalipto, em Minas Gerais. Foram treinadas redes neurais artificiais do tipo *Multilayer Perceptron* e da rede de melhor desempenho extraiu-se uma equação não-linear para predição da altura total de árvores. Foram construídos e ajustados modelos hipsométricos, linear e não-linear, contemplando variáveis relativas à árvore individual (DAP) e ao povoamento (área basal da parcela, idade e índice de sítio). Os novos modelos hipsométricos propostos neste trabalho apresentaram desempenho satisfatório e mostram-se eficientes para estimativa da altura total de árvores de eucalipto. Com destaque para o modelo extraído da rede neural artificial e o modelo não linear.

Palavra-chave: Relação hipsométrica; Inventário florestal; Eucalipto.

INTRODUCTION

Productivity and estimated forest volume are intrinsically related to height, and are essential components for decision-making (RIBEIRO et al., 2010). However, height measurement is considered an onerous activity in forest inventories, requiring a longer time to be measured and being more prone to errors, due to lack of visibility of the base and top of the tree, winds, as well as the need to know the distance between operator and individual trees (BINOTI; BINOTI; LEITE, 2013; MENDONÇA et al., 2011; VENDRUSCOLO et al., 2015).

In order to optimize and reduce costs, through a more efficient operation (MENDONÇA; CHAVES; CARVALHO, 2015; CALEGARIO, 2015; RIBEIRO et al., 2010), the hypsometric relationship, considering the biological basis between height and diameter, has become significant in the forest inventory (HESS et al., 2014), allowing an estimation of total tree height with other variables (MENDONÇA et al., 2011).

Environmental factors and population characteristics such as age, genetic material, region, local productive potential, planting density, sociological position, management regime, as well as the silvicultural practices performed, can directly interfere with the height variable and, consequently, the hypsometric relation (ROTHMAN et al., 2007). This stand information influences the precise estimation and biological realism when considering the hypsometric relation, and can then use the equation for estimates at different locations (BINOTI; BINOTI; LEITE, 2013; MARTINS et al., 2016).

As an alternative to regression, and often presenting superior results, artificial neural networks (ANNs) have been widely applied in the forest area for variable estimation, parameters and modeling (BINOTI; BINOTI; LEITE, 2013; CASTAÑO-SANTAMARÍA et al., 1990). This technique has high learning capacity and generalization in this study. It is considered an efficient tool for solving biologically based problems, and considered a good estimator (ÖZÇELİK et al., 2013).

When learning, ANNs have the ability to identify the relationship between input and output attributes (MARTINS et al., 2016), and are advantaged by inserting categorical variables in the training (VENDRUSCOLO et al., 2015).

The variable height is commonly used as input attribute to estimate other variables, in order to provide less susceptibility to errors (MARTINS et al., 2016). The objective of this study is to propose and evaluate the performance of a model extracted from ANNs training and new linear and nonlinear models to estimate the total height of eucalyptus trees.

MATERIAL AND METHODS

The data based used in this study came from temporary forest inventories carried out in eucalyptus stands in Minas Gerais, Brazil. It included a total of 4,229 individuals distributed in 347 plots. Numerical variables were considered: age (months), height (meters), diameter at breast height (cm), basal area of plot (m² / ha) and site index (m) (Table 1).

Table 1. Used descriptive statistics of forest inventory data of *Eucalyptus* spp. Where: H is height in meters; DBH is diameter in centimeters, at 1.30 meters of level height; I is age in months; G is the basal area in m² / ha; and S, the site index in meters

Variable	Minimum	Average	Maximum	Standard deviation	CV%
H	5,00	21,26	31,50	3,94	18,54
DBH	5,00	14,51	26,20	2,97	20,44
I	36	59	78	9	15,93
G	6,05	17,28	34,21	5,11	29,59
S	12,50	25,84	32,50	3,27	12,67

Training and validation of ANNs

ANNs were trained considering four input variables: DBH, age, site index, and baseline area; and one exit variable: height. The training process consisted in adjusting its weights through a learning algorithm that extracts characteristics from the data provided and aims to generate an ANN that performs the task of interest (BINOTI et al., 2014). The training was performed in software R, version 3.4.1, through the neuralnet package (FRITSCH; GÜNTHER, 2008).

The trained ANNs were of the Multilayer Perceptron type (MLP), composed of an input layer, an intermediate layer and an output layer. The algorithm used was the resilient backpropagation, in which the learning rate was defined automatically by the neuralnet package, with values varying between 0.01 and 1.12.

The choice of the number of neurons in the intermediate layer was made through the use of k-fold. This methodology randomly subdivides the database into k subgroups (ALI; PAZZANI, 1996; CIGIZOGLU; KISI, 2006). The k value was 10 subgroups, with the proportion of 90% for training and 10% for testing (DIAMANTOPOLOU, 2010), applying cross-validation. Different numbers of neurons, within the range of 1 to 20, were tested.

The activation function used was logistic, or sigmoidal, within a range of 0 to 1, which implies limiting the amplitude of outputs and inputs. For this reason, data was normalized, i.e., values were converted to ciphers from 0 to 1. Linear normalization was obtained through equation 1 (SOARES et al, 2011), which considers the minimum and maximum value of each variable for their conversion, maintaining the original data distribution (VALENCIA, 2010).

$$x' = \frac{(x - x_{min}) * (b - a)}{(x_{max} - x_{min})} + a \quad (1)$$

Where: x' : normalized value; x : original value; x_{min} : minimum value of the variable; x_{max} : maximum value of the variable; a : lower limit of the normalization interval; b : upper limit of the normalization interval.

Data was divided in two groups, 70% for training and 30% for generalization. Among data destined to ANNs, 70% was used in the training phase and 30% in the testing phase.

The criterion for stopping the ANNs' training process was either the maximum number of 100,000 cycles, or the mean squared error of less than 1%. Once one of those was met, training ceased and the best ANN was selected, based on the smallest mean square error.

Figure 1 illustrates the ANN architecture that presented the lowest error among the other evaluated, composed of five neurons in the hidden layer.

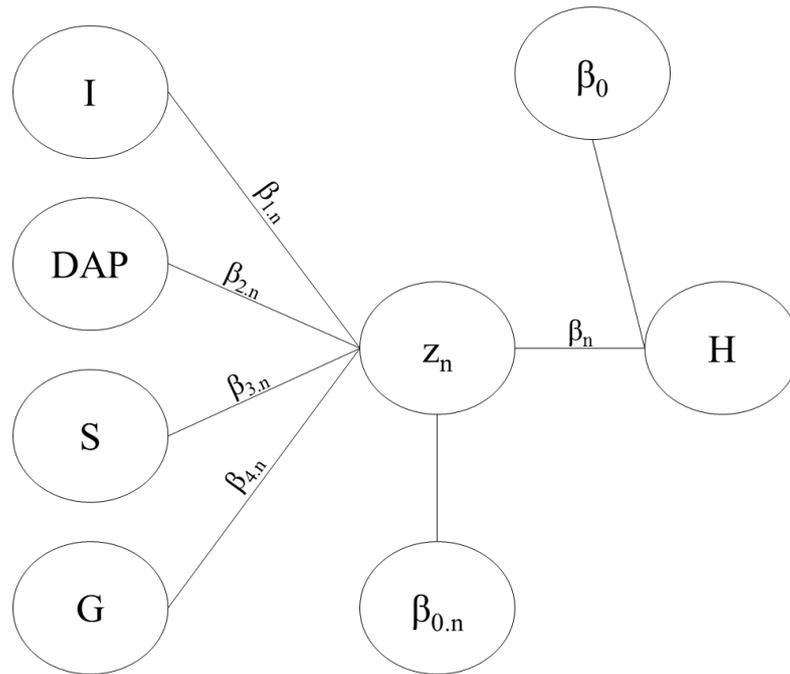


Figure 1. ANNs architecture. Where: H is height in meters; DAP is diameter in centimeters at chest height; I is age in months; G is the plot's basal area in m^2 ; and S, the site index in meters; n is neuron amount.

A nonlinear equation was extracted from the ANN to predict the total height of trees. A system of equations with coefficients resulting from the weights generated by neural network neurons was generated. This system was used to predict tree height that made up the generalized database.

Model (2) expresses the relation between the hidden layer and the response variable, where β_0 is the bias, and other coefficients are the weights related to each neuron. Model (3) represents the activation function used in each neuron of the hidden layer, derived from the logistic model. Finally, model (4) is the result of the relation between input variables and the respective hidden layer neurons, generating a model for each neuron.

$$H' = \beta_0 + \beta_1 * z_1 + \beta_2 * z_2 + \beta_3 * z_3 + \beta_4 * z_4 + \beta_5 * z_5 + \varepsilon \quad (2)$$

$$z_n = \left[\frac{1}{1 + e^{-w_i}} \right] \quad (3)$$

$$w_i = \beta_{0.n} + \beta_{1.n} * I_i' + \beta_{2.n} * DAP_i' + \beta_{3.n} * S_{ij}' + \beta_{4.n} * G_i' \quad (4)$$

Where: H' : standard height; β_0 : bias; β_n : coefficient associated with the neuron n ; $\beta_{k.n}$: coefficient of the model between the input variable k and the neuron n ; z_n : response of n -th hidden layer neuron; w_i : sum of the products between the weights and the inputs; I_i' : normalized age of the i -th tree (months); DAP_i' : diameter at the height of the normalized breast of the i -th tree (cm); S_{ij}' : site index of the j -th sampling unit containing the i -th tree (m), standardized; G_i' : normalized basal area of the i -th plot i -th (m^2/ha); ε : error.

Adjustments of linear and non-linear models

Data processing was performed in statistical software R, version 3.4.1, in order to compare the performance of ANN and hypsometric models. Data used in model adjustment were the same ones used in the ANN training. 70% of the data were allocated to the adjustment and 30% to the generalization, that is, the total height prediction in a set of data not used in the adjustments.

In order to estimate the total height of eucalyptus trees, linear and non-linear models were constructed and adjusted, considering variables related to individual tree (DAP) and stand (basal area, age and site index), based on the following functional relation:

$$f(H_i) = f(I_i, DBH_i, G_i, S_{ij})$$

Where: H_i : i -th tree height (m); DBH_i : i -th tree diameter at breast height (cm); I_i : i -th tree age (months); G_i : i -th plot basal area (m^2/ha); S_{ij} : site index of the j -th sample unit containing the i -th tree (m).

For model construction and adjustment, interactions among the independent variables were tested and transformed, so that there was no conflict with the positive or negative impact of each variable over H . The models were accepted when the interaction between all independent variables were significant according to the t-Student test, at a significance level of

0.05. The adjustments of the hypsometric models were also evaluated with ANOVA, regression, parametric, and residual analysis. In the case of the dependent variable undergoing transformation by the Napierian logarithm, the equations were compared by the Furnival index, in which the lower values indicate better performances (FURNIVAL, 1961).

$$IF = \exp \left[\frac{\sum_{i=1}^n \ln(Hreal_i)}{n} \right] * S_{YX} \quad (5)$$

Where: IF is the Furnival index; $Hreal_i$ is the actual height of the i -th individual in meters; S_{YX} is the estimate's standard error; and n is the number of trees sampled.

The constructed hypsometric models, linear (6) and nonlinear (7), are as follows:

$$\begin{aligned} \ln H_i = & \beta_0 + \beta_1 \left(\frac{1}{DBH_i} \right) + \beta_2 \left(\frac{1}{I_i} \right) + \beta_3 \left(\frac{1}{G_i} \right) + \beta_4 \left(\frac{1}{S_{ij}} \right) + \beta_5 \left(\frac{1}{DBH_i} * \frac{1}{I_i} \right) + \\ & \beta_6 \left(\frac{1}{DBH_i} * \frac{1}{G_i} \right) + \beta_7 \left(\frac{1}{I_i} * \frac{1}{G_i} \right) + \beta_8 \left(\frac{1}{DBH_i} * \frac{1}{I_i} * \frac{1}{G_i} \right) + \varepsilon_i \end{aligned} \quad (6)$$

$$H_i = \left[\left(\beta_0 + \beta_1 * S_{ij} + \beta_2 * \left(\frac{1}{I_i} \right) \right) * e^{\beta_3 + \beta_4 * \left(\frac{1}{G_i} \right) + \beta_5 * \left(\frac{1}{I_i} * \frac{1}{DBH_i} \right)} \right] + \varepsilon_i \quad (7)$$

Where: H_i : i -th tree height (m); β_k : regression coefficients; DBH_i : i -th tree diameter at breast height (cm); I_i : i -th tree age (months); G_i : i -th plot basal area (m²/ha); S_{ij} : site index of the j -th sample unit containing the i -th tree (m); ε : random error (m).

Adjusting the Parabolic Model

The hypsometric Parabolic model, in its original (8) and modified version (9), was used as reference in this study. The purpose of the modification was to standardize the variables used with the other techniques analyzed.

$$H_i = \beta_0 + \beta_1 DBH_i + \beta_2 DBH_i^2 + \varepsilon_i \quad (8)$$

$$H_i = \beta_0 + \beta_1 DBH_i + \beta_2 I_i + \beta_3 G_i + \beta_4 S_{ij} + \beta_5 DBH_i^2 + \beta_6 DBH_i * I_i + \beta_7 DBH_i * G_i + \beta_8 I_i * G_i + \beta_9 DBH_i * I_i * G_i + \varepsilon_i \quad (9)$$

Where: H_i : i-th tree height (m); β_k : regression coefficients; DBH_i : i-th tree diameter at breast height (cm); I_i : i-th tree age (months); G_i : i-th plot basal area (m²/ha); S_{ij} : site index of the j-th sample unit containing the i-th tree (m); ε : random error (m).

Analysis of the estimates generated by the model extracted from ANN and hypsometric models

The quality of the estimates by the models was evaluated based on ANOVA's, graphical residual analysis, Pearson correlation, Root Mean Square Error (RMSE) (SIIPILEHTO, 2000; LEITE; ANDRADE, 2002) (10), in the information criterion Akaike (SAKAMOTO et al., 1986) (11), and Graybill F Test (GRAYBILL, 1976) (12).

$$RMSE (\%) = \left[\sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} / \bar{Y} \right] * 100 \quad (10)$$

$$AIC = -2 * \ln(mv) + 2p \quad (11)$$

Where: *RMSE*: root of the mean square error; Y_i : observed value; \hat{Y}_i estimated value; n : number of cases; \bar{Y} : mean of observed values; *AIC*: Akaike information criterion; \ln : neperian logarithm; mv = maximum likelihood value; p : model parameter number.

For comparison between model predictions and actual values, Graybill's F test uses the coefficients generated by equation (13).

$$F(H_0) = \frac{\left(\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)' \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)}{2 * QM_{Res}} \quad (12)$$

$$Y = \beta_0 + \beta_1 x + \varepsilon \quad (13)$$

Where: $\hat{\beta}_0$ e $\hat{\beta}_1$: estimated parameters of the model (11); Y : estimated variable; β_0 e β_1 : model parameters; X : observed variable; ε : random error.

RESULTS

System coefficients of equations extracted from ANN are shown in Table 2.

Table 2. ANN parameters

Model	β_0	β_1	β_2	β_3	β_4	β_5
ANN	-0,0316	1,4592	-0,9360	0,2718	-0,0872	-0,5098
Neuron 1	-0,0536	-0,1007	2,1378	0,5594	-0,0774	-
Neuron 2	0,1142	-0,6396	2,2807	-0,4052	-1,1075	-
Neuron 3	1,0134	4,6225	-0,3106	0,7100	-0,7713	-
Neuron 4	4,3076	-31,5286	-12,5045	-11,3089	47,1370	-
Neuron 5	-4,2068	20,0465	-6,0806	-0,4135	-0,3370	-

The adjusted parameters for the linear model (6) encompassed the interactions between DAP inversions, age and basal area in their different combinations (Table 3). The coefficients related to the independent variables were significant, according to t-student test at a significance of 0.05.

Table 3. Significance of parameters adjusted for linear model (6)

F.V.	Parameters	Standard error	t value	Pr (>t)	
Intercept	4,8490	0,0891	54,4310	<0,0000	*
1/DBH	-10,9600	1,1190	-9,7920	<0,0000	*
1/I	-47,3200	5,3400	-8,8610	<0,0000	*
1/G	-5,8410	1,3230	-4,4140	<0,0000	*
1/S	-12,4900	0,4235	-29,4850	<0,0000	*
(1/DBH)(1/I)	288,2000	65,7300	4,3840	<0,0000	*
(1/DBH)(1/G)	68,7100	14,4900	4,7420	<0,0000	*
(1/I)(1/G)	137,800	69,4800	1,9830	0,0474	*
(1/DBH)(1/I)(1/G)	-2139,0000	729,6000	-2,9320	0,0034	*

Where: FV: source of variation; DBH: diameter at the breast height (cm); I: age (m); G: basal area (m²/ha); S: site index; *: significant at 0.05.

Table 4 shows the fitted parameters and the significance of non-linear model parameters (7). All coefficients were significant at 0.05 according to the t-Student test.

Table 4. Significance of the adjusted parameters for non-linear model (7)

Where: FV: source of variation; β_{\square} : regression coefficients; *: significant at 0.05.

F.V.	Parameters	Standard error	t value	Pr (>t)	
β_0	36,1200	1,1400	31,6860	<0,0000	*
β_1	0,6065	0,0177	34,2970	<0,0000	*
β_2	-1314,0000	58,7800	-22,3500	<0,0000	*
β_3	-8,4390	0,5648	-14,9430	<0,0000	*
β_4	-10,3400	1,7530	-5,8980	<0,0000	*
β_5	272,9000	34,0700	8,0090	<0,0000	*

The coefficients of the parabolic model were significant, except for the coefficient associated to DAP2 (Table 5).

Table 5. Significance of the adjusted parameters for the Parabolic model (8)

F.V.	Parameters	Standard error	t value	Pr (>t)	
Intercepto	6,5105	0,7621	8,5430	<0,0000	*
DBH	1,0025	0,1048	9,5630	<0,0000	*
DBH ²	0,0011	0,0035	0,3100	0,7570	ns

Where: FV: source of variation; *: significant at 0.05; ns: not significant.

The coefficients of the modified parabolic model are shown in Table 6. Among model parameters, six were not significant at a significance level of 0.05.

Table 6. Significance of the adjusted parameters for the modified parabolic model (9)

F.V.	Parameters	Standard error	t value	Pr (>t)	
Intercepto	-32,2400	19,3700	-1,6640	0,0962	ns
S	0,4245	0,0130	32,6610	<0,0000	*
G	2,3930	1,2780	1,8730	0,0612	ns
I	0,6025	0,3345	1,8010	0,0718	ns
DBH	8,0950	4,6870	1,7270	0,0842	ns
DBH ²	-0,7828	0,3741	-2,0920	0,0365	*
G*I	-0,0446	0,0183	-2,4320	0,0151	ns
DBH*G	-0,5141	0,2890	-1,7790	0,0754	*
I*DBH	-0,1300	0,0788	-1,6510	0,0989	ns
G*DBH ²	0,0431	0,0215	2,0070	0,0448	*
I*DBH ²	0,0132	0,0061	2,1660	0,0304	*
DBH*DBH ²	0,0228	0,0098	2,3160	0,0206	*
G*DBH*I	0,0092	0,0041	2,2560	0,0242	*
G*I*DBH ²	-0,0007	0,0003	-2,4410	0,0147	*
G*DBH*DBH ²	-0,0011	0,0005	-2,1700	0,0301	*
I*DBH*DBH ²	-0,0004	0,0002	-2,4830	0,0131	*
G*DBH*I*DBH ²	0,0000	0,0000	2,6010	0,0094	*

Where: FV: source of variation; *: significant at 0.05; ns: not significant.

Table 7 shows the models' variance analysis, where all were significant at a 0.05 level. The parabolic model had the largest mean square of the residue while the model extracted from ANN showed the lowest. The same behavior was evidenced by the determination coefficient (R^2). According to the Furnival index (IF) the equations had low values, except for the parabolic equation.

Table 7. Analysis of variance of the evaluated models

ANOVA ANN (R² 85,22; Syx:1,69m; IF:1,69)				
Source	DF	SS	MS	F
Regression	25	38.847,32	1.553,89	546,17
Residue	2.339	6.654,59	2,85	-
TOTAL	2.364	45.501,91		
ANOVA Linear regression (R² 84,27; Syx: 1,74m; IF:1,78)				
Source	DF	SS	MS	F
Regression	8	38.368,61	4.796,08	1.584,06
Residue	2.356	7.133,30	3,03	-
TOTAL	2.364	45.501,91		
ANOVA Non-linear regression (R² 83,88; Syx:1,76m; IF:1,57)				
Source	DF	SS	MS	F
Regression	6	38.187,37	6.364,56	2.051,75
Residue	2.358	7.314,54	3,10	-
TOTAL	2.364	45.501,91		
ANOVA Modified parabolic (R² 83,87; Syx: 1,76m; IF:1,57)				
Source	DF	SS	MS	F
Regression	15	38.207,90	2.547,19	820,31
Residue	2.349	7.294,01	3,11	-
TOTAL	2.364	45.501,91		
ANOVA Parabolic (R² 60,22; Syx: 2,77m; IF:2,47)				
Source	DF	SS	MS	F
Regression	2	27.417,77	13.708,88	1.790,54
Residue	2.362	18.084,14	7,66	-
TOTAL	2.364	45.501,91		

Where: DF: degree of freedom; SS: sum of squares; MS: mean squares; F: significance test.

Figure 2 shows the waste dispersion plots for the different techniques used. It is verified that there is a greater tendency of error in larger heights. Best results were derived from the

equation extracted from ANN, with greater residue homogeneity, and by the nonlinear equation, with lower dispersion of residues. The parabolic model resulted in greater dispersion.

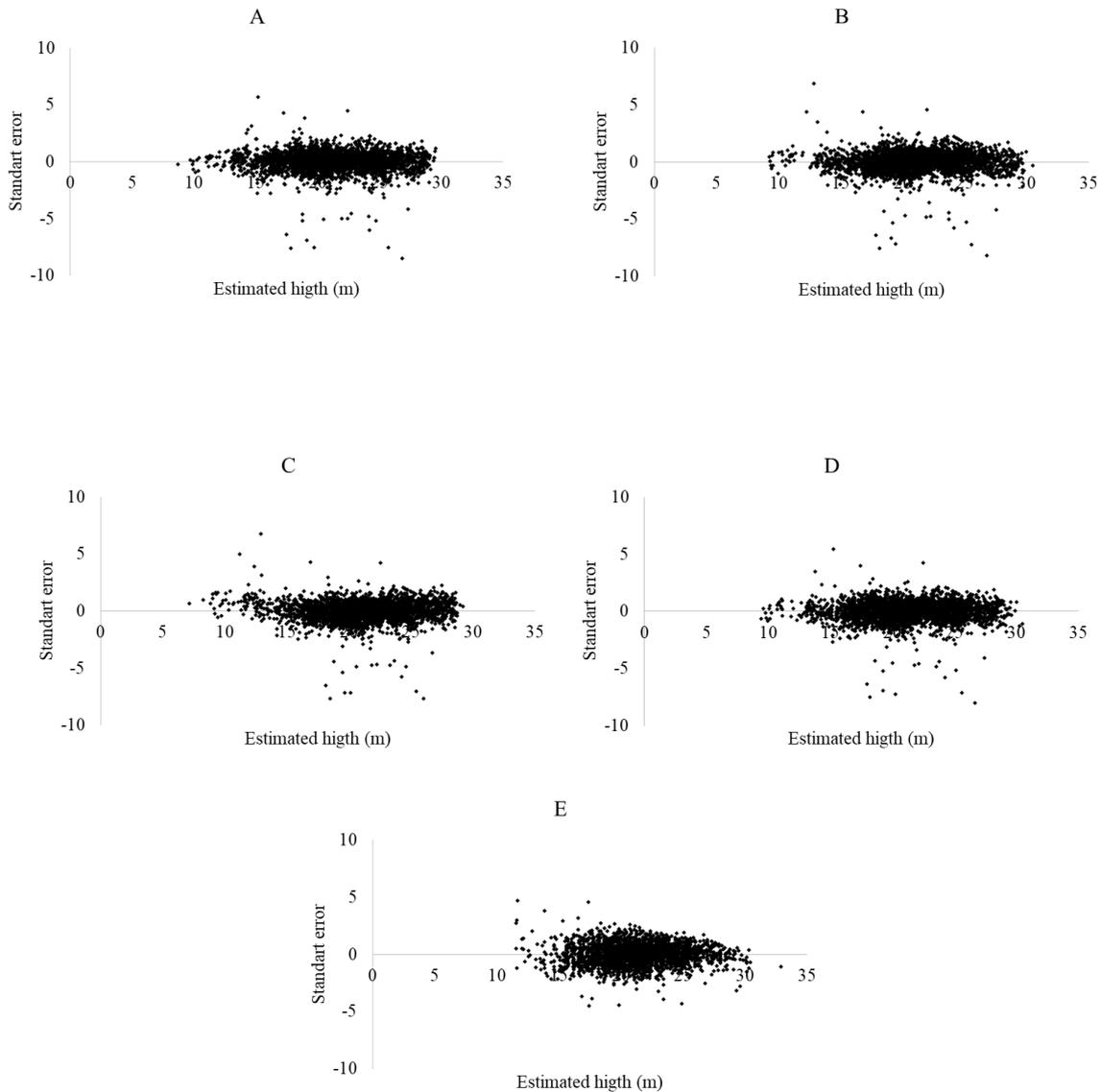


Figure 2. Residue scatter plot of adjusted equations (A): Artificial Neural Network, (B): Linear, (C): Nonlinear, (D): Parabolic modified, (E): Parabolic.

Table 8 shows the evaluation statistics of the generalization quality of the equations. The parabolic equation, used as reference in this study, presented the worst statistics when compared with the others.

Table 8. Criteria for evaluating equations

	Correlation	RMSE (%)	AIC	Graybill's F test
ANN	0,93	6,95	2,19	6,45E-04 ^{ns}
Linear regression	0,93	7,05	2,26	1,84E-06 ^{ns}
Non-linear regression	0,92	7,16	2,32	2,04E-04 ^{ns}
Modified parabolic	0,92	7,21	2,35	3,98E-05 ^{ns}
Parabolic	0,79	11,66	6,15	3,03E-03 ^{ns}

^{ns}: not significant at 0,05.

The estimates of the equations were strongly correlated with observed values, with an error magnitude lower than 10%, except for the parabolic equation. The lower the RMSE, the greater the accuracy of the estimates. The optimal situation is when it equals zero (MEHTÄTALO et al., 2006). RMSE for proposed equations (ANN, linear and non-linear regression) and the modified parabolic were close to each other. The Akaike information criterion for the equations proposed in this study did not present significant differences among them except for the parabolic equation, according to criteria presented by Burnham e Anderson (2002), in which differences smaller or equal to 2 do not indicate statistical difference. Graybill's F test indicated no statistical difference between values observed and estimated by the equations evaluated.

DISCUSSION

ANNs are widely distributed in the forest sector and have a high capacity to describe the behavior of the relationship between the independent and dependent variables (BINOTI et al, 2017; REIS et al., 2017; NANDY et al., 2017; VIEIRA et al., 2018). However, the discussions are not based on mathematical relationships between neurons and studied variables, which can be considered to obtain a model and, consequently, their regression statistics, allowing their comparison with others.

The coefficients of the equation extracted from the ANN in this study were efficient for estimating the total height. This can be attributed to the non-linear behavior generated by the hyperbolic tangent activation function, which can describe the biological relationship between height and independent variables properly.

Generally, classical hypsometric models encompass only the use of DAP variables, as in the parabolic model. Retslaff et al. (2015) report that the inclusion of other variables, such as age and site index, provides greater precision for height and is necessary to know the behavior of the influence of the independent variables on the response variable.

In order to obtain the linear model here proposed, inverse transformations of the independent variables and the dependent neperian logarithm were used, which contributed to the improvement of the correlation between the variables and generated a linear behavior. Therefore, variables were efficient for height prediction. As the main interaction (1/DAP, 1/I and 1/G) presented significant character, it is necessary to decompose it, that is, the presence of the secondary interactions between these variables (6), which allows a better understanding of the effects on the dependent variable (PREACHER; CURRAN; BAUER, 2006). The interaction of the site index with other continuous variables was not significant due to its categorical character (AERTSEN et al., 2010); hence, it was not considered. However, the site index is efficient in the model in an isolated way, as explained in equation (6), helping in the best line positioning. The influence of this variable on the linear coefficient was also observed in the model used by Nogueira et al. (2014).

The linear models are diffused for modeling independent variables, since they are easy to adjust and interpret. However, they do not allow biological interpretation, since relations between the dependent and independent variables are nonlinear (GEVREY et al., 2003).

Regarding the proposed nonlinear model, the influence of site index and age in the intercept of the equation was verified. These variables classify the population according to their hypsometric growth, influencing the location of the adjusted curve when categorizing the data.

The site index is strictly related to stand height, contributing to the homogeneity of the residues (RETSLAFF et al., 2015) and to the height prediction (SENA et al., 2015; TÉO et al., 2017). According to Silva et al. (2007), in addition to the linear coefficient of the equation, there is also an influence of the age variable on the angular coefficient, since there is a greater contribution of this variable in the height variation in higher DAP values. Regarding the basal area of the parcel, this variable represents the occupation degree of the area, indicating the competitiveness among the individuals. The influence of competitiveness behaves differently according to the sociological position of the trees, with a greater interference in the dominated trees and a lower one in the dominant ones (BARTOSZECK et al., 2004). In this sense, there is influence on the slope of the adjusted curve.

The statistics derived from the parabolic equation indicate the importance of the inclusion of variables that allow describing height variability, as presented in the models proposed in this study. The model extracted from ANN and the nonlinear model presented better performance by describing the biological behavior among the variables.

CONCLUSION

New hypsometric models here proposed had satisfactory performance and are efficient for estimating the total height of eucalyptus trees, with emphasis on the model extracted from the ANN and the nonlinear model.

REFERENCES

- Aertsen, W., Kint, V., Orshoven, J.V., Özkan, K., and Muys, B. 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecol. model.* **221**(8): 1119-1130. <https://doi.org/10.1016/j.ecolmodel.2010.01.007>.
- Ali, K.M., and Pazzani, M.J. 1996. Error reduction through learning multiple descriptions. *Machine learning*, **24**(3): 173-202. Available from <http://www.nrcresearchpress.com/page/cjfr/authors#9d> [accessed 24 September 2018].
- Bartoszeck, A.C.P.S., Machado, S.A., Figueiredo Filho, A., and Oliveira, E.B. 2004. Dinâmica da relação hipsométrica em função da idade, do sítio e da densidade inicial de povoamentos de bracinga da Região Metropolitana de Curitiba, PR. *Revista Árvore.* **28**: 517-533. <http://dx.doi.org/10.1590/S0100-67622004000400006>.
- Binoti, D.H.B., Binoti, M.L.M.S., and Leite, H.G. 2014. Configuração de redes neurais artificiais para estimação do volume de árvores. *Braz. J. Wood Sci.* **05**: 58-67. <http://dx.doi.org/10.15210/cmadv5i1.4067>.
- Binoti, D.H.B., Duarte, P.J., Silva, M.L.M., Silva, G.F., Leite, H.G., Mendonça, A.R., Andrade, V.C.L., and Vega, A.E.D. 2018. Estimation of height of eucalyptus trees with neuroevolution of augmenting topologies (NEAT). *Revista Árvore.* **41**: 1-8. <http://dx.doi.org/10.1590/1806-90882017000300014>.
- Binoti, M.L.M.S., Binoti, D.H.B., Leite, H.G. 2013. Aplicação de Redes Neurais Artificiais para Estimação da Altura de Povoamentos Equiâneos de Eucalipto. *Revista Árvore.* **37**: 639–645. <http://dx.doi.org/10.1590/S0100-67622013000400007>.
- Burnham, K.P., and Anderson, D.R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. Springer, New York, US.
- Castaño-Santamaría, J., Crecente-Campo, F., Fernández-Martínez, J.L., Barrio-Anta, M., and Obeso, J.R. 2013. Tree height prediction approaches for uneven-aged beech forests in

northwestern Spain. For. Ecol. Manage. **307**: 63–73.

<https://doi.org/10.1016/j.foreco.2013.07.014>.

Cigizoglu, H.K., and Kisi, Ö. 2006. Methods to improve the neural network performance in suspended sediment estimation. Journal of Hydrology. **317**: 221-238.

<https://doi.org/10.1016/j.jhydrol.2005.05.019>.

Diamantopoulou, M.J. 2010. Filling gaps in diameter measurements on standing tree boles in the urban forest of Thessaloniki, Greece. Environ. Model. Softw. **25**: 1857-1865.

<https://doi.org/10.1016/j.envsoft.2010.04.020>.

Furnival, G.M. 1961. An index for comparing equations used in constructing volume tables.

Forest Science. **7**: 337-341. <https://doi.org/10.1093/forestscience/7.4.337>.

Gevrey, M., Dimopoulos, I., and Lek, S. 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. model. **160**: 249-264.

[https://doi.org/10.1016/S0304-3800\(02\)00257-0](https://doi.org/10.1016/S0304-3800(02)00257-0).

Graybill, F.A. 1976. Theory and application of the linear model, Duxbury Press, Belmont, US.

Günther, F., and Fritsch, S. 2010. Neuralnet: Training of neural networks. The R journal. **2**(1): 30-38. Available from <https://journal.r-project.org/archive/2010/RJ-2010-006/RJ-2010-006.pdf> [accessed 24 September 2018].

Hess, A.F., Braz, E.M, Thaines, F., and Mattos, P.P. 2014. Ajuste de relação hipsométrica para espécies da Floresta Amazônica. Ambiência. **10**: 21–29.

<https://doi.org/10.5935/ambiencia.2014.01.02>.

Leite, H.G., and Andrade, V.C.L. 2002. Um método para condução de inventários florestais sem o uso de equações volumétricas. Revista Árvore. **26**: 321-328.

<http://dx.doi.org/10.1590/S0100-67622002000300007>.

Martins, E.D.R., Binoto, M.L.M.S., Leite, H.G., Binoti, D.H.B., and Dutram G.C. 2016. Configuração de redes neurais artificiais para estimação da altura total de árvores de eucalipto.

Revista Brasileira de Ciências Agrárias. **11**: 117–123.

<https://doi.org/10.5039/agraria.v11i1a5354>.

Mehtätalo L., Maltamo M., and Kangas, A. 2006. The use of quantile trees in the prediction of the diameter distribution of a stand. *Silva Fennica*. **40**: 501–516.

<https://doi.org/10.14214/sf.333>.

Mendonça, A.R., Calegario, N., Silva, G.F., Borges, L.A.C., and Carvalho, S.P.C. 2011. Modelos hipsométricos e de crescimento em altura das árvores dominantes e codominantes para *Pinus caribaea* var. *hondurensis*. *Sci. For.* **39**: 151–160. Available from

<http://www.ipef.br/publicacoes/scientia/nr90/cap03.pdf> [accessed 24 September 2018].

Mendonça, A.R., Carvalho, S.P.C., and Calegario, N. 2015. Modelos Hipsométricos Generalizados Mistos na Predição da Altura de *Eucalyptus* sp. *Cerne*. **21**: 107–115.

<http://dx.doi.org/10.1590/01047760201521011191>.

Nandy, S., Singh, R., Ghosh, S., Watham, T., Kushwaha, S.P.S., Kumar, A.S., and Dadhwal, V.K. 2017. Neural network-based modelling for forest biomass assessment. *Carbon Manage.*

8: 305–317. <https://doi.org/10.1080/17583004.2017.1357402>.

Nogueira, G.S., Marshall, P.L., Leite, H.G., and Campos, J.C.C. 2015. Thinning intensity and pruning impacts on *Eucalyptus* plantations in Brazil. *Int. J. For. Res.* **2015**: 1–10.

<http://dx.doi.org/10.1155/2015/168390>.

Özçelik, R., Diamantopoulou, M.J., Crecente-Campo, F., and Eler, U. 2013. Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models.

For. Ecol. Manage. **306**: 52–60. <https://doi.org/10.1016/j.foreco.2013.06.009>.

Preacher, K.J., Curran, P.J., and Bauer, D.J. 2006. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *J. Educ. Behav. Stat.* **31**: 437–448.

<https://doi.org/10.3102/10769986031004437>.

- R Development Core Team, 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Reis, L.P., Souza, A.L., Reis, P.C.M., Mazzei, L., Soares, C.P.B., Torres, C.M.M.E., Silva, L. F., Ruschel, A.R., Rego, L.J.S., and Leite, H.G. 2018. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. *Ecol. Eng.* **112**: 140-147. <https://doi.org/10.1016/j.ecoleng.2017.12.014>.
- Retslaff, F.A.S., Figueiredo Filho, A., Dias, A.N., Bernett, L.G., and Figura, M.A. 2015. Curvas de sítio e relações hipsométricas para *Eucalyptus grandis* na região dos Campos Gerais, Paraná. *Cerne*. **21**: 219-225. <http://dx.doi.org/10.1590/01047760201521021349>.
- Ribeiro, A., Ferraz Filho, A.C., Mello, J.M., Ferreira, M.Z., Lisboa, P.M.M., and Scolforo, J.R.S. 2010. Estratégias e metodologias de ajuste de modelos hipsométricos em plantios de *Eucalyptus* sp. *Cerne*. **16**: 22–31. <http://dx.doi.org/10.1590/S0104-77602010000100003>.
- Sakamoto, Y., Ishiguro, M., and Kitagawa, G. 1986. Akaike information criterion statistics, Dordrecht, The Netherlands.
- Sena, A.L.M., Silva Neto, A.J., Oliveira, G.M.V., and Calegario, N. 2015. Linear and nonlinear models with use of covariates for hypsometric relationship of two tropical pine species. *Ciência Florestal*. **25**: 969-980. <http://dx.doi.org/10.5902/1980509820651>.
- Siipilehto, J. 2000. A comparison of two parameter prediction methods for stand structure in Finland. *Silva Fennica*. **34**: 331-349. Available from <http://www.metla.fi/silvafennica/abs/sa34/sa344331.htm> [accessed 24 September 2018].
- Silva, G.F., Xavier, A.C., Rodrigues, F.L., and Peternelli, L.A. 2007. Análise da influência de diferentes tamanhos e composições de amostras no ajuste de uma relação hipsométrica para *Eucalyptus grandis*. *Revista Árvore*. **31**: 685-694. <http://dx.doi.org/10.1590/S0100-67622007000400013>.

- Soares, F.A.A.M.N., Flôres, E.L., Cabacinha, C.D., Carrijo, G.A., and Veiga, A.C.P. 2011. Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks. *Comput. Eléctron. agric.* **78**: 19-27. <https://doi.org/10.1016/j.compag.2011.05.008>.
- Sousa, G.T.O., Azevedo, G.B., Barreto, P.A.B., and Conceição Júnior, V. 2013. Relações hipsométricas para *Eucalyptus urophylla* conduzidos sob regime de alto fuste e talhadia no sudoeste da Bahia. *Scientia Plena.* **9**: 1–7. Available from <https://www.scientiaplenu.org.br/sp/article/view/771/671> [accessed 24 September 2018].
- Téo, S.J., Machado, S.A., Figueiredo Filho, A., Tomé, M., 2017. General height-diameter equation with biological attributes for *Pinus taeda* L. stands. *Cerne.* **23**: 403-411. <http://dx.doi.org/10.1590/01047760201723042414>.
- Valença, M. 2010. Fundamentos das redes neurais: exemplos em Java. 2nd ed. Livro Rápido, Olinda, BR.
- Vendruscolo, D.G.S., Drescher, R., Souza, H.S., Moura, J.P.V.M., Mamoré, F.M.D., and Siqueira, T.A.S. 2015. Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais artificiais. *Rev. Bras. Biom.* **33**: 556–569. Available from http://jaguar.fcav.unesp.br/RME/fasciculos/v33/v33_n4/A9_DiogoRonaldo.pdf [accessed 24 September 2018].
- Vieira, G.C.V., Mendonça, A.R., Silva, G.F., Zanetti, S.S., Silva, M.M., and Santos, A.R. 2018. Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. *Sci. Total Environ.* **619-620**: 1473-1481. <https://doi.org/10.1016/j.scitotenv.2017.11.138>.

**ARTIGO 2 – MODELOS MISTOS EM MULTINÍVEL PARA ESTIMATIVA DA
ALTURA TOTAL DE ÁRVORES DE *Eucalyptus* spp.**

ARTIGO FORMATADO DE ACORDO COM A NBR 6022 (ABNT, 2003).

RESUMO

O setor florestal vive uma constante busca por técnicas que possibilitem atingir a excelência na estimativa da produção madeireira, otimizando os processos e maximizando os lucros. Nesse contexto, os procedimentos de mensuração são elementos fundamentais e merecem atenção, especialmente métodos que proporcionem estimativas precisas da altura total das árvores, por serem a base para a determinação do volume de madeira em uma área. O objetivo deste trabalho foi avaliar o desempenho de modelos linear e não-linear de efeitos mistos, em multinível, adotando estrutura na matriz de variância e covariância para estimar a altura total de árvores de *Eucalyptus* spp. A base de dados foi proveniente de inventários florestais temporários realizados em povoamentos localizados entre os municípios de Bom Jardim de Minas, Lima Duarte e Arantina, em Minas Gerais, Brasil. Os modelos linear e não linear foram ajustados com regressão fixa e com a inclusão de efeitos aleatórios, em nível de indivíduo e unidade de manejo. Os ajustes dos modelos mistos foram realizados adotando-se estruturas distintas para a matriz de variância e covariância. A seleção das melhores equações baseou-se no Critério de Informação de Akaike (AIC), Teste da Razão de Máxima Verossimilhança (TRMV) e Teste de Vuong. De acordo com os critérios adotados, os modelos mistos com estrutura Huynh-Feldt (H-F) na matriz de variância e covariância foi superior aos modelos de efeitos fixos. Verificou-se uma redução na raiz do erro quadrático médio de 7,05 % para 0,64%, no modelo linear, e de 7,16 % para 0,54 %, no modelo não linear. Sendo assim, os modelos mistos com estrutura na matriz de variância e covariância mostram-se bastante eficientes para modelagem da altura total de *Eucalyptus* spp., sendo recomendada a sua utilização, devido ao ganho em precisão.

Palavra-chave: Relação hipsométrica. Modelo linear. Modelo não linear. Inventário florestal.

ABSTRACT

The forest sector is constantly searching for techniques that enable it to achieve excellence in the estimation of timber production, optimizing processes and maximizing profits. In this context, measurement procedures are fundamental elements and deserve attention, especially methods that provide accurate estimates of the total height of the trees, since they are the basis for determining the volume of wood in an area. The objective of this work was to evaluate the performance of models, linear and nonlinear, fitted with mixed effects, in multilevel, adopting structure in the matrix of variance and covariance to estimate the total height of Eucalyptus trees. The database was derived from temporary forest inventories carried out in settlements located between the cities of Bom Jardim de Minas, Lima Duarte and Arantina, in Minas Gerais, Brazil. The linear and nonlinear models were adjusted with fixed regression and with the inclusion of random effects in multilevel. The adjustments of the mixed models were performed adopting distinct structures of variance and covariance matrix. The selection of the best equations was based on the Akaike Information Criterion (AIC), Maximum Likelihood Ratio Test (TRMV) and Vuong Test. According to the adopted criteria, the mixed models with Huynh-Feldt structure (H-F) in the variance and covariance matrix were superior to the fixed effects models. There was a reduction in root mean square error from 7.05% to 0.64% in the linear model and from 7.16% to 0.54% in the nonlinear model. Thus, the mixed models with structure in the variance and covariance matrix are very efficient for modeling the total height of Eucalyptus, and its use is recommended due to gain in precision.

Keywords: Hypsometric relationship. Linear model. Nonlinear model. Forest Inventory.

1 INTRODUÇÃO

O conhecimento da produtividade local e a estimativa do volume de madeira em uma floresta está intrinsecamente relacionado à variável altura, sendo estes componentes essenciais para as tomadas de decisão relacionadas ao manejo florestal (RIBEIRO et al., 2010). No entanto, a mensuração da altura é tida como uma atividade onerosa na realização do inventário florestal, com maior demanda de tempo para ser aferida e propícia à erros, devido à falta de visibilidade da base e topo da árvore, ocorrência de ventos e necessidade do conhecimento da distância entre o operador e indivíduo (BINOTI; BINOTI; LEITE, 2013; MENDONÇA et al., 2011; VENDRUSCOLO et al., 2015).

Visando a otimização e redução dos custos, devido à uma operacionalização mais eficiente, a relação hipsométrica, que considera o fundamento biológico existente entre a altura e o diâmetro, permite estimar a altura total da árvore em função de outras variáveis (MENDONÇA et al., 2011; HESS et al., 2014; MENDONÇA; CHAVES; CARVALHO, 2015; CALEGARIO, 2015; RIBEIRO et al., 2010).

Fatores ambientais e características do povoamento florestal como idade, material genético, região, potencial produtivo do local, densidade do plantio, posição sociológica, regime de manejo, assim como as práticas silviculturais realizadas, podem interferir diretamente na variável altura e, conseqüentemente, na relação hipsométrica (MENDONÇA et al., 2011; RIBEIRO et al., 2010; SOUSA; AZEVEDO; BARRETO, 2013). Estas informações provenientes do povoamento florestal influenciam na estimativa precisa e com maior realismo biológico ao se considerar a relação hipsométrica, podendo então utilizar a equação para estimativas em diferentes locais (BINOTI; BINOTI; LEITE, 2013; MARTINS et al., 2016).

Uma alternativa para considerar dados correlacionados, no espaço e/ou no tempo, e modelar explicitamente a estrutura de covariância é a utilização de modelos mistos. Com os modelos mistos é possível generalizar estruturas de correlações e variância não constante. Os modelos mistos são uma das técnicas mais sofisticadas de regressão, tendo o trabalho de Lappi (1991) como pioneiro na área florestal.

O modelo misto possibilita incluir, na estrutura de seu ajuste, um conjunto de variáveis não observáveis, denominado de efeitos aleatórios, por exemplo, incorporar a variabilidade da árvore, de maneira a permitir, junto às variáveis observáveis, denominadas de efeitos fixos, o ajuste de um modelo mais consistente (PINHEIRO; BATES, 2000). Nas ciências florestais, os modelos mistos são aplicados em problemas não lineares, como crescimento em altura de

eucaliptos e avaliações genéticas (CALEGARIO et al., 2005; COSTA et al., 2005; MORA, 2006).

A variável altura é comumente utilizada como atributo de entrada para estimativa de outras variáveis, a fim de propiciar menor susceptibilidade a erros (MARTINS et al., 2016). Portanto, objetivo deste trabalho é comparar o desempenho de modelos lineares e não-lineares ajustados por regressão fixa e regressão mista em nível de indivíduo e povoamento para estimar a altura total de árvores de eucalipto.

2 MATERIAL E MÉTODOS

2.1 Dados

A área de estudo abrange 21 unidades de manejo implantadas com clones de híbridos de *Eucalyptus* spp., localizados nos municípios de Bom Jardim de Minas, Lima Duarte e Arantina, em Minas Gerais, totalizando 1090 ha de área inventariada.

O clima da região é caracterizado como tropical de altitude, tipo Cwb, segundo a classificação de Köppen, com temperatura média anual de 20,1 °C, com invernos secos e frios, com ocorrência de geadas em algumas áreas, e verões chuvosos com temperaturas moderadamente altas. A precipitação total média anual é de 1456 mm (ALVARES et al., 2013).

Os dados utilizados neste estudo foram provenientes de inventários florestais temporários e pré-corte. Considerou-se um total de 27299 indivíduos, distribuídos em 565 parcelas, tendo como variáveis numéricas: idade (meses), altura (metros), diâmetro à altura do peito (centímetros), área basal da parcela (m²/ha) e índice de sítio (m) (TABELA 1).

Tabela 1 - Estatísticas descritivas dos dados amostrais utilizados no estudo.

Variável	Mínimo	Média	Máximo	Desvio Padrão	CV%
H	5,00	21,26	31,50	3,94	18,54
DAP	5,00	14,51	26,20	2,97	20,44
I	36	59	78	9	15,93
G	6,05	17,28	34,21	5,11	29,59
S	12,50	25,84	32,50	3,27	12,67

Fonte: Do autor.

em que H é a altura em metros; DAP é o diâmetro, em centímetros, a 1,30 m de altura do solo; I é a idade, em meses; G é a área basal, em m²/ha; e S, o índice de sítio, em metros.

2.2 Ajustes dos modelos linear e não-linear

O processamento dos dados foi realizado no *software* R, versão 3.4.1. Foram ajustados os modelos linear (1) e não-linear (2) para a estimativa da altura total de árvores de eucalipto, contemplando variáveis relativas à árvore individual (DAP) e ao povoamento (área basal, idade e índice de sítio).

$$\ln H_i = \beta_0 + \beta_1 \left(\frac{1}{DAP_i} \right) + \beta_2 \left(\frac{1}{I_i} \right) + \beta_3 \left(\frac{1}{G_i} \right) + \beta_4 \left(\frac{1}{S_{ij}} \right) + \beta_5 \left(\frac{1}{DAP_i} * \frac{1}{I_i} \right) + \beta_6 \left(\frac{1}{DAP_i} * \frac{1}{G_i} \right) + \beta_7 \left(\frac{1}{I_i} * \frac{1}{G_i} \right) + \beta_8 \left(\frac{1}{DAP_i} * \frac{1}{I_i} * \frac{1}{G_i} \right) + \varepsilon_i \quad (1)$$

$$H_i = \left[\left(\beta_0 + \beta_1 * S_{ij} + \beta_2 * \left(\frac{1}{I_i} \right) \right) * e^{\beta_3 + \beta_4 * \left(\frac{1}{G_i} \right) + \beta_5 * \left(\frac{1}{I_i} * \frac{1}{DAP_i} \right)} \right] + \varepsilon_i \quad (2)$$

em que: H_i : altura da i-th árvore (m); β_k : coeficientes da regressão; DAP_i : diâmetro à altura do peito da i-th árvore (cm); I_i : idade da i-th árvore (meses); G_i : área basal da i-th parcela (m²/ha); S_{ij} : índice de sítio da j-th unidade amostral contendo a i-th árvore (m); ε : erro aleatório (m).

2.3 Ajustes de modelos multinível de efeito misto

A base de dados utilizada neste trabalho é classificada como não-balanceada, ou seja, o número de árvores entre as unidades amostrais é diferente (CALEGARIO, 2005). O procedimento de modelagem de efeitos mistos multinível foi empregado incorporando nos modelos linear e não linear a variabilidade existente entre os indivíduos e entre as parcelas. Para isso, os modelos foram reajustados, por meio de regressão mista, pelo método da máxima verossimilhança. O método da máxima verossimilhança foi idealizado por Fisher, segundo Searle (1987), e consiste na obtenção de estimadores que maximizem a função densidade de probabilidade das observações, em relação aos efeitos fixos e aos componentes de variância.

A abordagem de forma matricial de modelos lineares mistos pode ser representada, de acordo com Mrode (2015), por:

$$y = X\beta + Zb + \varepsilon, \quad (3)$$

em que: y é um vetor $n \times 1$ de observações da variável resposta ou variável dependente, sendo n o número de observações; X é uma matriz $n \times p$ de incidência dos efeitos fixos, sendo p o número de parâmetros de efeitos fixos; β é um vetor $p \times 1$ de parâmetros de efeitos fixos; Z é uma matriz $n \times q$ de incidência dos efeitos aleatórios, sendo q o número de parâmetros de efeitos aleatórios; b é um vetor $q \times 1$ de parâmetros dos efeitos aleatórios; ε é um vetor $n \times 1$ de erros aleatórios, com

$$b \sim N(0, G) \text{ e } \varepsilon \sim N(0, R) \quad (4)$$

sendo que G é a matriz de variâncias - covariâncias dos efeitos aleatórios e R é a matriz de variâncias - covariâncias residuais.

Desse modo tem-se,

$$y \sim N(X\beta, ZGZ' + R). \quad (5)$$

Já os modelos não lineares mistos (6), podem ser representados com a variável resposta y_{ij} representando os grupos aleatórios i e j , em que i representa a i -ésima unidade amostral e j a j -ésima árvore. Então, $i = 1, \dots, m$; e $j = 1, \dots, n_i$; em que m é o número total de unidades amostrais e n_i é o número de árvores dentro da i -ésima unidade amostral; f é uma função geral, real e diferenciável de um grupo específico de vetores do parâmetro ϕ_{ij} e um vetor de covariantes v_{ij} ; e ε_{ij} é o erro aleatório normalmente distribuído dentro dos grupos (PINHEIRO; BATES, 2000).

$$y_{ij} = f(\phi_{ij}, V_{ij}) + \varepsilon_{ij}, i = 1, \dots, m; j = 1, \dots, n_i. \quad (6)$$

O vetor de parâmetros varia de indivíduo para indivíduo. Em um segundo estágio, o vetor ϕ_{ij} pode ser expresso pela equação 7.

$$\phi_{ij} = A_{ij}\beta_i + B_{ij}\beta_i + B_{ij}\beta_{ij}, \beta_i \approx N(0, \Psi_1) \text{ e } \beta_{ij} \approx N(0, \Psi_2) \quad (7)$$

em que: β é um vetor ($p \times 1$) de efeitos fixos; B_i é um vetor ($q_1 \times 1$) de efeitos aleatórios independentemente distribuídos com uma matriz de covariância-variância ψ_1 ; B_{ij} é um vetor ($q_2 \times 1$) de efeitos aleatórios independentemente distribuídos com uma matriz de covariância-variância ψ_1 , e pressuposto ser independente dos efeitos aleatórios do primeiro nível; A_{ij} e B_{ij} são matrizes de incidência; ε_{ij} , dentro dos grupos, são independentemente distribuídos e são independentes dos efeitos aleatórios.

Na modelagem dos modelos mistos, uma etapa importante é a definição da estrutura de variância e covariância, pois ela visa obter uma estrutura parcimoniosa, que explique bem a variabilidade dos dados e a correlação entre as medidas com um número reduzido de parâmetros

(TORAL; ALENCAR; FREITAS, 2006). Essa escolha pode afetar diretamente as estimativas dos parâmetros, os erros padrão de efeitos fixos e aleatórios, diagnósticos e inferências. Sua seleção depende das estruturas dos dados, de informações empíricas e da disponibilidade computacional.

As modelagens destas estruturas para o modelo linear misto e não linear misto são semelhantes. Este processamento foi feito por meio do pacote *nlme*, do *software* R, e sua função *correlation* (PINHEIRO; BATES, 2000). Foram utilizadas 14 estruturas, entre as quais Componente de Variância (CV), Não Estruturada (NE), Simétrica Composta (SC), Autorregressiva de Primeira Ordem (AR(1)), Autorregressiva de Primeira Ordem Heterogênea (ARH(1)), Simétrica Composta Heterogênea (CSH), Toeplitz (TOEP), Autorregressiva de Primeira Ordem Médias Móveis (ARMA(1,1)), Toeplitz Heterogênea (TOEPH), Antependência de Ordem 1 (ANTE(1)), Correlação sem Estrutura (CSE), Espacial (SP(POW)(c-list)), Diagonal Principal (UN(1)) e Huynh-Feldt (H-F). Mais detalhes sobre essas estruturas podem ser encontrados em (PINHEIRO; BATES, 2000; DAVIDIAN; GILTINAN, 1995).

Para a seleção da estrutura da matriz de variância e covariância, foi utilizado o critério de informação de Akaike (AIC) (8) (SAKAMOTO et al., 1986), o teste da razão da máxima verossimilhança e o teste de Vuong (1989).

$$AIC = -2 * \ln(mv) + 2p \quad (8)$$

em que: *AIC* é critério de informação Akaike; \ln é o logaritmo neperiano; *mv* é o valor de máxima verossimilhança; e *p* é o número de parâmetros do modelo.

O teste de razão da máxima verossimilhança (TRMV) (9) consiste em comparar modelos dois a dois, nos quais o seu valor é obtido pela diferença dos valores das suas funções de verossimilhança (PINHEIRO; BATES, 2000).

$$TRMV = 2 * [\ln(mv_2) - \log(mv_1)] \quad (9)$$

em que: \ln é o logaritmo neperiano; *mv*₂ é o valor de função de verossimilhança do modelo 2; *mv*₁ é o valor de função de verossimilhança do modelo 1.

Para comparação entre os modelos, por meio do teste de razão de verossimilhança - T_{RLNN} de Vuong (1989) (10), utilizou-se a estatística descrita a seguir:

$$T_{RLNN} = \frac{1}{\sqrt{n}} \frac{\log(\widehat{m}\widehat{v}, \widehat{\beta})}{\widehat{\omega}^2} \quad (10)$$

em que:

$$\widehat{\omega}^2 = \frac{1}{n} \sum_{i=1}^n \left(\log \frac{f(y_i|x_i, \widehat{m}\widehat{v})}{g(y_i|x_i, \widehat{\beta})} \right)^2 - \left[\frac{1}{n} \sum_{i=1}^n \left(\log \frac{f(y_i|x_i, \widehat{m}\widehat{v})}{g(y_i|x_i, \widehat{\beta})} \right) \right]^2 \quad (11)$$

é um estimador para variância de $\frac{1}{\sqrt{n}} \log(\widehat{m}\widehat{v}, \widehat{\beta})$ e $\log(\widehat{m}\widehat{v}, \widehat{\beta})$ é o teste da razão de verossimilhança. A estatística tem, assintoticamente,

$$TLRNN \xrightarrow{d} N(0,1) \quad (12)$$

sob

$$H_0 : \log \frac{f(y_i|x_i, \widehat{m}\widehat{v})}{g(y_i|x_i, \widehat{\beta})} = 0 \quad (13)$$

ou seja, os modelos são equivalentes ao nível α de significância, sendo $Z_{\alpha/2}$ o valor crítico da distribuição normal padrão, rejeitando-se a hipótese nula se $|T_{RLNN}| > Z_{1-\alpha/2}$.

Selecionada a melhor estrutura da matriz de variância e covariância para cada modelo misto, linear e não linear, os ajustes dos modelos foram avaliados com base na Raiz do Erro Quadrático Médio (REQM) (SIPILEHTO, 2000; LEITE; ANDRADE, 2002) (14), correlação entre valores estimados e observados e Teste F de Graybill (GRAYBILL, 1976) (15). Além disso, foi realizada a análise gráfica dos resíduos padronizados, comparando os modelos com regressão fixa e mista em multinível com adoção de estrutura na matriz de variância e covariância.

$$REQM (\%) = \left[\sqrt{\frac{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n-p}}{\bar{Y}}} \right] * 100 \quad (14)$$

em que: *REQM*: raiz do erro quadrático médio; Y_i : valor observado; \hat{Y}_i : valor estimado; n : número de casos; \bar{Y} : média dos valores observados.

O teste F de Graybill (15), para comparação entre as predições dos modelos e os valores reais, utiliza os coeficientes gerados pela equação (16).

$$F (H_0) = \frac{\left(\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)' \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)}{2 * QM_{Res}} \quad (15)$$

$$Y = \beta_0 + \beta_1 x + \varepsilon \quad (16)$$

em que: $\hat{\beta}_0$ e $\hat{\beta}_1$ são os parâmetros estimados do modelo (16); Y é a variável estimada; β_0 e β_1 são os parâmetros do modelo; X é a variável observada; ε é o erro aleatório.

3 RESULTADOS E DISCUSSÃO

Os parâmetros ajustados para o modelo linear englobaram as interações existentes entre os inversos do DAP, da idade e da área basal e suas diferentes combinações (TABELA 2). Os coeficientes relacionados às variáveis independentes foram significativos, pelo teste *t-student* a um nível de significância de 0,05.

Tabela 2 - Significância dos parâmetros ajustados para modelo linear (1).

F.V.	Parâmetros	Erro padrão	t value	Pr (>t)	
Intercepto	4,8490	0,0891	54,4310	<0,0000	*
1/DAP	-10,9600	1,1190	-9,7920	<0,0000	*
1/I	-47,3200	5,3400	-8,8610	<0,0000	*
1/G	-5,8410	1,3230	-4,4140	<0,0000	*
1/S	-12,4900	0,4235	-29,4850	<0,0000	*
(1/DAP)(1/I)	288,2000	65,7300	4,3840	<0,0000	*
(1/DAP)(1/G)	68,7100	14,4900	4,7420	<0,0000	*
(1/I)(1/G)	137,800	69,4800	1,9830	0,0474	*
(1/DAP)(1/I)(1/G)	-2139,0000	729,6000	-2,9320	0,0034	*

Fonte: Do autor.

em que: F.V.: fonte de variação; DAP: diâmetro a 1,30 m do solo (cm); I: idade (m); G: área basal (m²/ha); S: índice de sítio; *: significativo a 0,05;

A Tabela 3 indica os parâmetros ajustados e a significância dos parâmetros do modelo não linear (2), sendo que todos foram significativos a um nível de 0,05 para o teste *t-Student*.

Tabela 3 - Significância dos parâmetros ajustados para modelo não linear (2).

F.V.	Parâmetros	Erro padrão	t value	Pr (>t)	
β_0	36,1200	1,1400	31,6860	<0,0000	*
β_1	0,6065	0,0177	34,2970	<0,0000	*
β_2	-1314,0000	58,7800	-22,3500	<0,0000	*
β_3	-8,4390	0,5648	-14,9430	<0,0000	*
β_4	-10,3400	1,7530	-5,8980	<0,0000	*
β_5	272,9000	34,0700	8,0090	<0,0000	*

Fonte: Do autor.

em que: F.V.: fonte de variação; β_i : coeficientes da regressão; *: significativo a 0,05.

As Tabelas 4 e 5 apresentam os critérios de seleção para as estruturas das matrizes de variância e covariância avaliadas, para os modelos mistos linear e não linear. Dentre elas, a que melhor se ajustou à estimação da altura das árvores de eucalipto foi a Huynh-Feldt (H-F), que apresentou menor valor de AIC e maior valor do logaritmo da máxima verossimilhança. Além disso, realizando o teste da razão de verossimilhanças e tomando tal estrutura como hipótese alternativa, comparou-se com as outras estruturas.

Tabela 4 - Seleção das estruturas da matriz de variância e covariância para o modelo linear misto.

Modelo	AIC	LogLik	Teste	TRMV	p-valor
1. Toeplitz (TOEP)	10345,09	5811,04	1 vs 6	555,58	<0,0000
2. Toeplitz H. (TOEPH)	10391,45	5813,12	2 vs 6	502,92	<0,0000
3. Não estruturada	10377,32	5811,37	3 vs 6	518,33	<0,0000
4. Correlação sem estrutura	10392,75	5815,25	4 vs 6	773,58	<0,0000
5. Simétrica C.H. (CSH)	10012,96	6347,95	5 vs 6	1,18	<0,8801
6. Huynh-Feldt (H-F)	10014,31	6348,86	-	-	-

Fonte: Do autor.

Tabela 5 - Seleção das estruturas da matriz de variância e covariância para o modelo não linear misto.

Modelo	AIC	LogLik	Teste	TRMV	p-valor
1. Toeplitz (TOEP)	8746,91	5995,03	1 vs 6	492,7	<0,0000
2. Toeplitz H. (TOEPH)	8747,95	5997,11	2 vs 6	440,04	<0,0000
3. Não estruturada	8757,40	5995,36	3 vs 6	455,45	<0,0000
4. Correlação sem estrutura	8662,33	5999,24	4 vs 6	510,7	<0,0000
5. Simétrica C.H. (CSH)	8490,56	6531,94	5 vs 6	0,88	0,7902
6. Huynh-Feldt (H-F)	8473,34	6532,85	-	-	-

Fonte: Do autor.

É possível observar que as hipóteses nulas foram praticamente todas rejeitadas (p -valor < 0,05), com exceção da estrutura Simétrica Composta Heterogênea (CSH), que não apresentou diferença significativa em relação à Huynh-Feldt (H-F). Isso foi verificado em ambos os modelos, linear e não linear, em que os valores de p -valor foram 0,8801 e 0,7902, respectivamente. Indicando que estas duas estruturas para a matriz de variância e covariância são as mais adequadas e poderiam ser utilizadas. Estas estruturas consideram que árvores

pertencentes a uma mesma parcela têm maior correlação entre si do que entre árvores pertencentes a outras parcelas, havendo covariâncias diferentes. De acordo com West, Welch e Gatecki (2015), a estrutura Huynh-Feldt (HF) é caracterizada pelas variâncias desiguais entre parcelas e covariâncias obtidas pela média aritmética entre as variâncias e subtraindo λ , sendo λ a diferença entre a média das variâncias e a média das covariâncias. Na estrutura Simétrica Composta Heterogênea (CSH) aplicam-se diferentes variâncias e algumas covariâncias desiguais, ajustadas por um coeficiente de correlação ρ entre os indivíduos.

Optou-se pela estrutura Huynh-Feldt (H-F) por apresentar menor valor de AIC e LogLik.

Nas Tabelas 6 e 7 estão apresentados os resultados dos ajustes (parâmetros fixos) dos modelos mistos linear e não linear, respectivamente, por meio do método da máxima verossimilhança adotando estrutura Huynh-Feldt (H-F) na matriz de variância e covariância, em que todos os parâmetros apresentaram significância a um nível de 0,05.

Tabela 6 - Significância dos parâmetros fixos ajustados para o modelo linear misto.

F.V.	Parâmetros	Erro padrão	t value	Pr (>t)	
Intercepto	4,9892	0,0668	48,1873	<0,0000	*
1/DAP	-13,1745	0,8558	-11,0189	<0,0000	*
1/I	-58,4056	3,9870	-9,7067	<0,0000	*
1/G	-7,8977	0,9907	-5,1407	<0,0000	*
1/S	-11,6701	0,3138	-17,1883	<0,0000	*
(1/DAP)(1/I)	438,5311	49,7992	6,3029	<0,0000	*
(1/DAP)(1/G)	92,9600	10,9458	5,6855	<0,0000	*
(1/I)(1/G)	264,3745	51,7783	3,3448	0,0048	*
(1/DAP)(1/I)(1/G)	-3736,6734	548,3373	-4,5672	0,0009	*

Fonte: Do autor.

em que: F.V.: fonte de variação; DAP: diâmetro a 1,30 m do solo (cm); I: idade (m); G: área basal (m²/ha); S: índice de sítio; *: significativo a 0,05;

Tabela 7 - Significância dos parâmetros fixos ajustados para modelo não linear misto.

F.V.	Parâmetros	Erro padrão	t value	Pr (>t)	
β_0	37,9895	0,2069	31,4761	<0,0000	*
β_1	0,5669	0,0276	20,5269	<0,0000	*
β_2	-1430,1869	33,7983	-25,1801	<0,0000	*
β_3	-9,3829	0,5402	-17,3691	<0,0000	*
β_4	-10,8414	0,6561	-4,0816	<0,0000	*
β_5	364,2581	19,0100	10,7081	<0,0000	*

Fonte: Do autor.

em que: F.V.: fonte de variação; β_i : coeficientes da regressão; *: significativo a 0,05.

Na Tabela 8 são apresentados os critérios de avaliação da qualidade dos modelos.

Tabela 8 - Critérios de avaliação das equações.

Modelo	Coefficiente de correlação	REQM (%)	Teste F de Graybill
Linear fixo	0,9323	7,05	1,84E-06 ^{ns}
Não linear fixo	0,9255	7,16	2,04E-04 ^{ns}
Linear misto	0,9990	0,64	6,14E-06 ^{ns}
Não linear misto	0,9993	0,54	5,01E-07 ^{ns}

Fonte: Do autor.

em que: ^{ns}: não significativo a 0,05.

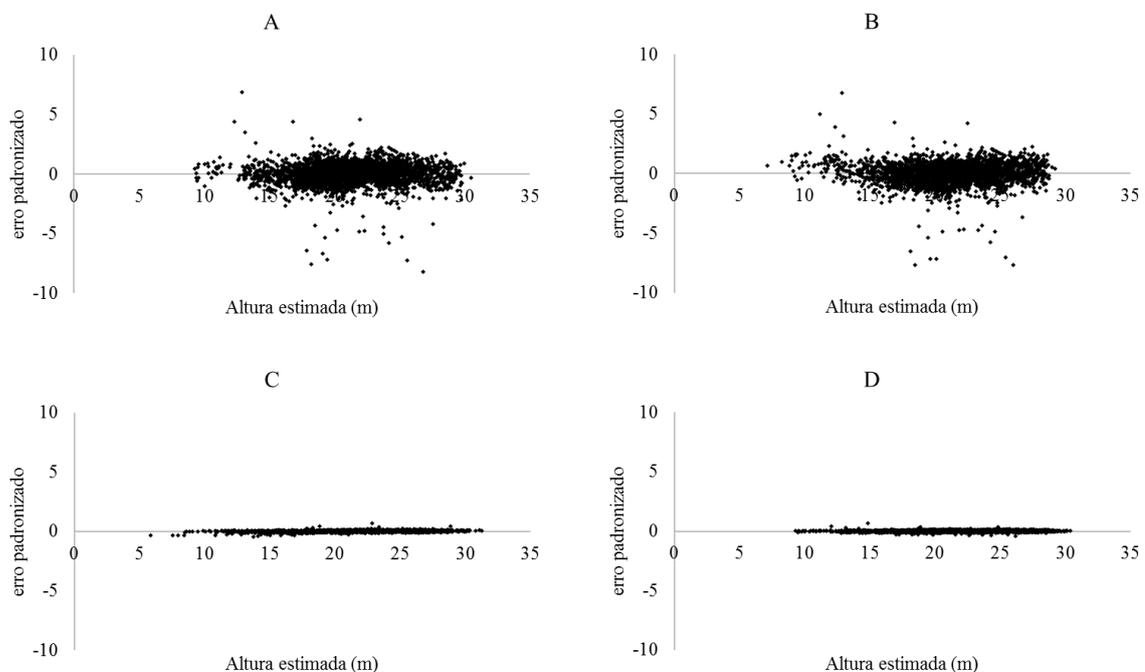
As estimativas das equações apresentaram-se fortemente correlacionadas com os valores observados, com valores de 0,92 e 0,93 para os modelos de regressão fixa e 0,99 para os modelos de regressão mista. Para todos os modelos, verificou-se magnitudes de erros inferiores a 10%, sendo que quanto menor o REQM, maior a precisão das estimativas, e a situação ótima quando este é igual a zero (MEHTÄTALO et al., 2006). Os modelos fixos apresentaram valores de REQM 7,05 % (linear) e 7,16 % (não linear), enquanto os modelos mistos linear e não linear apresentaram expressiva redução, com valores inferiores a 0,7 % e próximos entre si. O teste de F de Graybill indicou a não existência de diferença estatística entre os valores observados e estimados em todas as equações avaliadas.

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. Em Calegario et al. (2005), que estudaram a área basal de clones de *Eucalyptus* spp., a redução foi de aproximadamente 15

vezes. Neste trabalho, também se observou redução do erro padrão residual, que foi de 1,7400 m no modelo linear de regressão fixa para 0,1582 m no modelo linear de regressão mista, uma redução de 11 vezes. Já para os modelos não lineares a redução foi de 1,7600 m para o modelo fixo para 0,1332 m para o modelo misto, uma redução de 13,2 vezes. Com isso, destaca-se o ganho em precisão com a utilização dos modelos ajustados por regressão mista, uma vez que estes modelos permitem estimar os efeitos fixos, prever os efeitos aleatórios e estimar os componentes de variância, considerando a variabilidade existente em cada árvore e entre as diferentes parcelas.

Como próximo passo de avaliação dos modelos, foi analisada a distribuição dos resíduos. A Figura 1 apresenta os gráficos de dispersão de resíduos padronizados para as diferentes abordagens trazidas no estudo. Em geral, verifica-se uma boa distribuição dos resíduos, com dispersão em torno de zero e homogeneidade da variância. No entanto, os modelos mistos proporcionaram menores dispersões de resíduos, indicando considerável superioridade na qualidade de suas estimativas.

Figura 1 - Gráfico de dispersão de resíduos das equações ajustadas, (A): Linear, (B): Não Linear, (C): Linear misto, (E): Não linear misto.



Fonte: Do autor.

Os gráficos confirmam a eficiência da inclusão de efeitos aleatórios, em multinível, com adoção de estrutura Huynh-Feldt (H-F) na matriz de variância e covariância, nos modelos linear e não linear propostos neste trabalho, para estimação da altura total de árvores de eucalipto, devido ao aumento expressivo da precisão das estimativas, quando comparados com os modelos com efeitos fixos. Sendo, portanto, recomendada a utilização destes modelos.

De acordo com West, Welch e Gatecki (2015), as estruturas de variância e covariância permitem uma maior flexibilidade nas correlações. Em qualquer análise de dados, deve-se buscar escolher uma estrutura correta para as matrizes G e R que seja mais adequada e parcimoniosa, com conhecimento dos dados observados e sobre as relações entre as observações de cada parcela, uma vez que diferentes números de variâncias e covariâncias devem ser estimados (LITTELL et al., 2006; PERRI; IEMMA, 1999).

Os resultados desta pesquisa corroboram com os estudos de Mendonça, Carvalho e Calegario (2015). Os autores comprovaram a eficiência dos modelos mistos ao compararem o método tradicional de regressão não linear dos modelos assintótico e sigmoidal, com a inclusão de efeitos mistos, em que observaram aumento na acurácia dos ajustes do crescimento em altura. Adame et al. (2008), Calegario et al. (2005a; 2005b) e Crecente-Campo et al. (2010) observaram redução do erro padrão residual e, conseqüentemente, aumento na precisão da estimativa do crescimento em altura de árvores de eucalipto com inclusão de efeitos aleatórios. Sharma e Breidenbach (2015) ajustaram funções não lineares da relação diâmetro e altura com a inclusão de efeitos aleatórios para três espécies: *Picea abies* (L.) Karst., *Pinus sylvestris* L., e *Betula pubescens* (Ehrh.), na Noruega, e perceberam um ganho significativo na qualidade da predição da altura por meio de modelos mistos.

4 CONCLUSÃO

A inclusão de efeitos aleatórios, em multinível, nos modelos linear e não linear propostos neste trabalho proporcionou significativa redução do erro padrão residual, quando comparados com os modelos ajustados com regressão fixa.

A escolha de uma estrutura da matriz de variância e covariância é um passo importante no ajuste de modelos de efeito misto. Neste estudo, a estrutura Huynh-Feldt foi a que apresentou as melhores estatísticas, melhorando, portanto, a predição da altura de árvores de clones de eucalipto.

REFERÊNCIAS

- ADAME, P.; DEL-RÍO, M.; CAÑELLAS, I. A mixed nonlinear height-diameter model for pyrenean oak (*Quercus pyrenaica* Willd.). **Forest Ecology and Management**, Amsterdam, v. 256, n. 1/2, p. 88-98, 2008.
- ALVARES, C.A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, p. 711-728, 2013.
- BINOTI, M. L. M. DA S.; BINOTI, D. H. B.; LEITE, H. G. Aplicação de Redes Neurais Artificiais para Estimção da Altura de Povoamentos Equiâneos de Eucalipto. **Revista Árvore**, v. 37, n. 4, p. 639-645, 2013.
- CALEGARIO, N. et al. Estimativa de crescimento de povoamentos de Eucalyptus baseada na teoria dos modelos não-lineares em multinível de efeito misto. **Ciência Florestal**, Santa Maria, v. 15, n. 3, p. 285-292, 2005a.
- CALEGARIO, N. et al. Melhoria da qualidade de ajuste de modelos biométricos florestais pelo emprego da teoria dos modelos não lineares generalizados. **Scientia Forestalis**, Piracicaba, n. 69, p. 38-50, 2005b.
- CALEGARIO, N. et al. Modeling dominant height growth based on nonlinear mixed-effects model: a clonal Eucalyptus plantation case study. **Forest Ecology and Management**, Amsterdam, v. 204, p. 11-20, 2005c.
- CARVALHO, S. P. C. et al. Modelagem não linear mista para descrever o afilamento de árvores clonais de *Eucalyptus* sp. **Scientia Forestalis**, Piracicaba, v. 42, n. 104, p. 605-614, 2014.
- COSTA, B. R. et al. Avaliação genética de indivíduos de erva-mate (*Ilex Paraguariensis* St.-Hil.) na região de Caarapó, MS, pelo procedimento REML/BLUP. **Revista Ciência Florestal**, Santa Maria, v. 16, n. 2, p. 145 - 151, 2005.
- CRECENTE-CAMPO, F. et al. A generalized nonlinear mixed-effects height– diameter model for *Eucalyptus globulus* L. in northwestern Spain. **Forest Ecology and Management**, Amsterdam, v. 259, p. 943-952, 2010.
- DAVIDIAN, M.; GILTINAN, D. M. **Nonlinear models for repeated measurement data**. New York: Chapman and Hall, v. 1., 1995. 360 p.
- GRAYBILL, F.A. **Theory and application of the linear model**. Belmont: Duxbury Press, 1976. 704p.
- HESS, A. F. et al. Ajuste de relação hipsométrica para espécies da Floresta Amazônica. **Ambiência**, v. 10, n. 1, p. 21-29, 2014.
- LAPPI, J. Calibration of height and volume equations with random parameters. **Forest Science**, Lawrence, v. 37, n. 3, p. 781 - 801, 1991.

LEITE, H. G.; ANDRADE, V. C. L. Um método para condução de inventários florestais sem o uso de equações volumétricas. **Revista Árvore**, Viçosa, v. 26, n. 3, 2002.

LITTELL, R. C. et al. SCHABENBERGER, O. **SAS® for mixed models**. 2 ed. Cary: SAS Institute Inc., 2006. 795 p.

MARTINS, E. D. R. et al. Configuração de redes neurais artificiais para estimação da altura total de árvores de eucalipto. **Revista Brasileira de Ciências Agrárias**, v. 11, n. 2, p. 117–123, 2016.

MEHTÄTALO L., MALTAMO M., KANGAS A. The use of quantile trees in the prediction of the diameter distribution of a stand. **Silva Fennica**, v. 40, n. 3, article id 333, 2006. <https://doi.org/10.14214/sf.333>

MENDONÇA, A. R. DE et al. Modelos hipsométricos e de crescimento em altura das árvores dominantes e codominantes para *Pinus caribaea* var. *hondurensis*. **Scientia Forestalis**, v. 39, n. 90, p. 151–160, 2011.

MENDONÇA, A. R. DE; CHAVES E CARVALHO, S. D. P.; CALEGARIO, N. Modelos Hipsométricos Generalizados Mistos na Predição da Altura de *Eucalyptus* sp. **Cerne**, v. 21, n. 1, p. 107–115, 2015.

MORA, F. Herebabilidade y valor genético (REML/BLUP) en genotipos de un eucalipto tolerante a la sequía en el Norte de Chile. **Revista Ciência Florestal**, Santa Maria, v. 16, n. 2, p. 145 - 151, 2006.

MRODE, R.A. **Linear models for the prediction of animal breeding values**. 3 ed. Wallingford: CAB International, 2015. 343 p.

PINHEIRO, J. C.; BATES, D. M. **Mixed-effects models in S and S-PLUS**. New York: Springer-Verlag, 2000. v. 1. 528 p.

PERRI, S. H. V; IEMMA, A. F. Procedimento MIXED do SAS® para análise de modelos mistos. **Scientia Agricola**, v. 56, n. 4, p. 959-967, 1999.

RIBEIRO, A. et al. Estratégias e metodologias de ajuste de modelos hipsométricos em plantios de *Eucalyptus* sp. **Cerne**, v. 16, n. 1, p. 22–31, 2010.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. **Akaike information criterion statistics**. Dordrecht, The Netherlands: D. Reidel, 1986. 81 p.

SEARLE, S.R. **Linear models for unbalanced data**. New York: John Wiley, 1987. 536 p.

SHARMA, M.; BREIDENBACH, J. Modeling height-diameter relationships for Norway spruce, Scots pine, and downy birch using Norwegian national forest inventory data Ram P.. **Forest Science and Technology**, Seul, v. 11, n. 1, p. 44, 2015.

SOUSA, G. T. O.; AZEVEDO, G. B.; BARRETO, P. A. B. Relações hipsométricas para *Eucalyptus urophylla* conduzidos sob regime de alto fuste e talhadia no sudoeste da Bahia. **Scientia Plena**, v. 9, n. 4, p. 1–7, 2013.

SIIPILEHTO, J. A comparison of two parameter prediction methods for stand structure in Finland. **Silva Fennica**, v. 34, n. 4, 2000.

TORAL, F.L.B.; ALENCAR, M.M.; FREITAS, A.R. Arranjos para efeitos fixos e estruturas de (co)variâncias residuais para análises de medidas repetidas do peso de bovinos da raça Canchim. **Revista Brasileira de Zootecnia**, v. 35, n. 5, p. 1951-1958, 2006.

VENDRUSCOLO, D. G. S. et al. Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais artificiais. **Revista Brasileira de Biometria**, v. 33, n. 4, p. 556–569, 2015.

VUONG, Q. Likelihood ratio tests for model selection and nontested hypotheses. **Econometrica**, v. 57, p. 307 - 333, 1989.

WEST, B. T.; WELCH, K. B.; GATECKI, A. T. **Linear mixed models: a practical guide using statistical software**. 2 ed. Ann Arbor: Chapman and Hall/CRC, 2015. 405 p.

**ARTIGO 3 – MODELO NÃO LINEAR MISTO PARA ESTIMATIVA DO VOLUME
DE ÁRVORES DE *Eucalyptus* spp.**

ARTIGO FORMATADO DE ACORDO COM A NBR 6022 (ABNT, 2003).

RESUMO

A estimativa volumétrica dos clones de eucalipto, em geral, é baseada em equações que têm como variáveis independentes o diâmetro mensurado à altura do peito e a altura total da árvore. Esta relação é modelada por regressão clássica não permitindo por exemplo, generalizações de estruturas de variância e covariância. Uma alternativa para considerar dados possivelmente correlacionados e modelar explicitamente a sua estrutura de covariância é o uso de modelos mistos. O presente estudo teve como objetivo avaliar a qualidade das estimativas volumétricas de árvores de *Eucalyptus* spp por meio de um modelo com efeitos mistos. A base de dados foi proveniente de um povoamento localizado nos municípios de Bom Jardim de Minas, Lima Duarte e Arantina, em Minas Gerais. Foram cubadas rigorosamente, pelo método de Smalian, 818 árvores. O modelo de Schumacher e Hall foi ajustado com regressão fixa e com a inclusão de efeitos aleatórios. O ajuste do modelo misto foi realizado adotando-se estruturas distintas para a matriz de variância e covariância. A seleção da melhor equação baseou-se no Critério de Informação de Akaike (AIC), Teste da Razão de Máxima Verossimilhança (TRMV) e Teste de Vuong. De acordo com os critérios adotados, o modelo misto com estrutura Huynh-Feldt (HF) na matriz de variância e covariância foi superior ao modelo clássico com regressão fixa, o qual apresentou uma redução no erro relativo médio de 6,61 % para 0,43%. O modelo misto não linear com estrutura na matriz de variância e covariância mostra-se eficiente para modelagem do volume de clones de *Eucalyptus* spp, sendo recomendada a sua utilização, devido ao ganho em precisão.

Palavras-chave: Modelo de Schumacher e Hall. Regressão clássica. Regressão com efeitos mistos. Volume de árvores.

ABSTRACT

The volumetric estimation of eucalyptus clones is usually based on equations that have as their independent variables the diameter at the height of the chest and the total height of the tree. This relationship is modeled by classical regression, not allowing, for example, generalizations of variance and covariance structures. An alternative to consider possibly correlated data and to explicitly model its covariance structure is the use of mixed models. The present study had as objective to evaluate the quality of the volumetric estimates of *Eucalyptus* spp. trees by means of a model with mixed effects. The database came from a settlement located in the municipalities of Bom Jardim de Minas, Lima Duarte and Arantina, in Minas Gerais, Brazil. 818 trees were scaled by the Smalian method. The model of Schumacher and Hall was adjusted with fixed regression and with the inclusion of random effects. The adjustment of the mixed model was done by adopting distinct structures in the variance and covariance matrix. The selection of the best equation was based on the Akaike Information Criterion (AIC), Maximum Likelihood Ratio Test (TRMV) and Vuong Test. According to the adopted criteria, the mixed model with Huynh-Feldt structure (HF) in the variance and covariance matrix was superior to the classic model with fixed regression, which presented a reduction in the mean relative error from 6.61% to 0.43%. The non-linear mixed model with structure in the variance and covariance matrix is efficient for *Eucalyptus* spp. volume modeling, and its use is recommended due to gain in accuracy.

Keywords: Schumacher and Hall model. Classic regression. Regression with mixed effects. Volume of trees.

1 INTRODUÇÃO

Uma das informações de maior importância para o conhecimento do potencial produtivo de uma floresta, em uma dada região, é a variável volume de madeira, sendo imprescindível para o planejamento do manejo sustentável de florestas. O volume individual de madeira é um ponto de partida para a avaliação do conteúdo de madeira em um povoamento florestal (MACHADO et al., 2002) e serve como base para as decisões relacionadas a tratamentos silviculturais, colheita e transporte da produção volumétrica da madeira. Portanto, é fundamental que a obtenção do volume de madeira seja realizada corretamente, representando bem a população amostrada (ANDRADE, 2001).

A estimativa volumétrica dos clones de eucaliptos é baseada, normalmente, em equações que têm como variáveis independentes o diâmetro à altura do peito (DAP) e a altura total da árvore (H). De acordo com Scolforo et al. (2004) e Thomas et al. (2006), o modelo de Schumacher e Hall destaca-se na estimativa do volume de árvores. No entanto, os modelos de regressão clássicos possuem como premissa, a independência entre as observações e a homogeneidade da variância, que podem, em alguns casos, não serem atendidas.

Uma alternativa para considerar dados correlacionados, no espaço e/ou no tempo, e modelar explicitamente a sua estrutura de covariância, se dá por meio dos modelos mistos. Algumas das abordagens possíveis com os modelos mistos é a generalização das estruturas de correlações e da variância. Os modelos mistos são técnicas sofisticadas de regressão, tendo o trabalho de Lappi (1991) como pioneiro na área florestal.

O modelo misto possibilita incluir em sua estrutura um conjunto de variáveis não observáveis, denominadas de efeitos aleatórios. Com isso, tem-se um ajuste mais consistente do modelo ao se incorporar a variabilidade existente em cada árvore junto às variáveis observáveis, denominadas de efeitos fixos (PINHEIRO; BATES, 2000). Segundo Littell et al. (1996) e Wolfinger (1993), os modelos mistos podem descrever blocos incompletos, parcelas subdivididas, dados espaço-temporais, coeficientes aleatórios, curvas polinomiais e análise de curvas de crescimento.

Geralmente, nas ciências florestais, os modelos mistos são aplicados em problemas não lineares, como crescimento em altura de eucaliptos e avaliações genéticas (CALEGARIO et al., 2005; COSTA et al., 2005; MORA, 2006).

Verificada a relevância e sofisticação desta classe de modelos, o objetivo deste trabalho foi avaliar o desempenho do modelo volumétrico não linear de Schumacher e Hall ajustado por

meio de regressão fixa e regressão mista com modelagem da matriz de variância e covariância dos dados.

2 MATERIAIS E MÉTODOS

2.1 Dados

A área de estudo abrange 21 unidades de manejo implantadas com clones de híbridos de *Eucalyptus* spp., localizados nos municípios de Bom Jardim de Minas, Lima Duarte e Arantina, em Minas Gerais, Brasil, totalizando 1090 ha de área inventariada.

O clima da região é caracterizado como tropical de altitude, tipo Cwb, segundo a classificação de Köppen, com temperatura média anual de 20,1 °C, com invernos secos e frios, com ocorrência de geadas em algumas áreas, e verões chuvosos com temperaturas moderadamente altas. A precipitação total média anual é de 1456 mm (ALVARES et al., 2013).

Os dados utilizados neste estudo foram provenientes da cubagem rigorosa de 818 árvores, realizada entre os anos de 2010 e 2016. As informações obtidas foram a altura total (*Ht*), em metros, o diâmetro à altura do peito (*DAP*), em centímetros, e os diâmetros na base das árvores (a 0,1 m de altura) e nas alturas de 0,5 m; 1 m; 1,5 m e 2 m; e, a partir desta secção, a cada 2 m. Os volumes individuais foram obtidos pela fórmula de Smalian. As estatísticas descritivas dos dados são apresentadas na Tabela 1.

Tabela 1 - Estatísticas descritivas dos dados de volume do fuste de árvores de *Eucalyptus* spp.

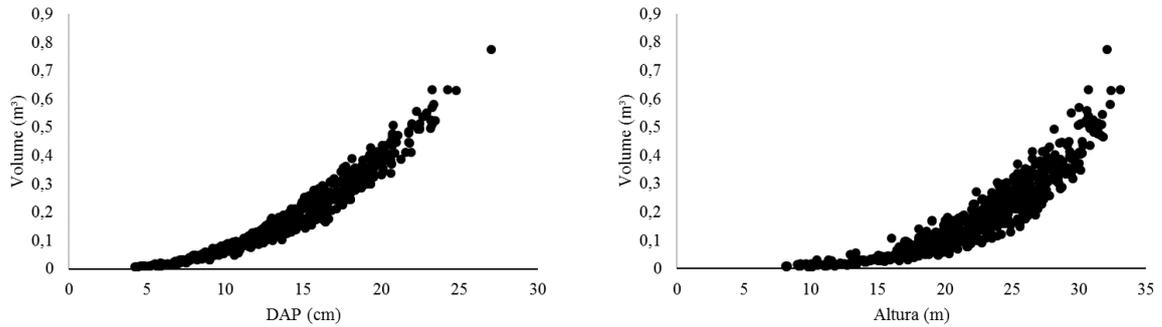
Variável	Mínimo	Médio	Máximo	Desvio padrão	CV%
Volume (m ³)	0,0066	0,1785	0,7752	0,1335	74,79
DAP (cm)	4,19	13,60	27,00	4,45	32,72
Altura (m)	8,10	21,82	33,08	5,10	23,37

Fonte: Do autor

em que: DAP é o diâmetro a 1,30 de altura do solo.

Os gráficos de dispersão dos volumes de madeira em função do *DAP* e *Ht* são apresentados na Figura 1, indicando um comportamento não linear entre estas variáveis.

Figura 1 - Gráficos de dispersão dos volumes individuais de *Eucalyptus* spp. em função do *DAP*, *Ht*.



Fonte: Do autor.

2.2 Modelo volumétrico

O modelo volumétrico não linear de Schumacher e Hall (1933) (1) foi ajustado aos dados de volume. O processamento foi realizado no *software* estatístico R, versão 3.4.1.

$$V_i = \beta_0 * DAP_i^{\beta_1} * Ht_i^{\beta_2} + \varepsilon_i \quad (1)$$

em que: V é o volume em m^3 ; DAP é o diâmetro, em cm, a 1,30 m de altura; Ht é a altura total em m; β_0 , β_1 e β_2 são os parâmetros do modelo; ε é o erro aleatório.

2.3 Modelo volumétrico multinível de efeito misto

Posteriormente, com o intuito de avaliar o emprego de modelos de efeito misto para estimar o volume individual de madeira de *Eucalyptus*, o modelo de Schumacher e Hall foi reajustado, com a incorporação da variabilidade existente em cada árvore e em cada unidade de manejo. Com isso, tem-se o modelo não linear de efeitos mistos com parâmetros fixos e aleatórios. O ajuste foi realizado pelo método da máxima verossimilhança, idealizado por Fisher, segundo Searle (1987), que consiste na obtenção de estimadores que maximizem a função densidade de probabilidade das observações, em relação aos efeitos fixos e aos componentes de variância.

Os modelos não lineares mistos (2) podem ser representados com a variável resposta y_{ij} representando os grupos aleatórios i e j , em que i representa a i -ésima unidade de manejo e j a j -ésima árvore. Então, $i = 1, \dots, m$; e $j = 1, \dots, n_i$; em que m é o número total de unidades de

manejo e n_i é o número de árvores dentro da i -ésima unidade de manejo; f é uma função geral, real e diferenciável de um grupo específico de vetores do parâmetro ϕ_{ij} e um vetor de covariantes v_{ij} ; e ε_{ij} é o erro aleatório normalmente distribuído dentro dos grupos (PINHEIRO; BATES, 2000).

$$y_{ij} = f(\phi_{ij}, V_{ij}) + \varepsilon_{ij}, i = 1, \dots, m; j = 1, \dots, n_i. \quad (2)$$

O vetor de parâmetros varia de indivíduo para indivíduo. Em um segundo estágio, o vetor ϕ_{ij} pode ser expresso pela equação 3.

$$\phi_{ij} = A_{ij}\beta_i + B_{ij}\beta_i + B_{ij}\beta_{ij}, \beta_i \approx N(0, \Psi_1) \text{ e } \beta_{ij} \approx N(0, \Psi_2) \quad (3)$$

em que: β é um vetor ($p \times 1$) de efeitos fixos; B_i é um vetor ($q_1 \times 1$) de efeitos aleatórios independentemente distribuídos com uma matriz de covariância-variância ψ_1 ; B_{ij} é um vetor ($q_2 \times 1$) de efeitos aleatórios independentemente distribuídos com uma matriz de covariância-variância ψ_1 , e pressuposto ser independente dos efeitos aleatórios do primeiro nível; A_{ij} e B_{ij} são matrizes de incidência; ε_{ij} , dentro dos grupos, são independentemente distribuídos e são independentes dos efeitos aleatórios.

Na modelagem dos modelos mistos, uma etapa importante é a definição da estrutura de variância e covariância, pois ela visa obter uma estrutura parcimoniosa, que explique bem a variabilidade dos dados e a correlação entre as medidas com um número reduzido de parâmetros (TORAL; ALENCAR; FREITAS, 2006). Essa escolha pode afetar diretamente as estimativas dos parâmetros, os erros padrão de efeitos fixos e aleatórios, diagnósticos e inferências. Sua seleção depende das estruturas dos dados, de informações empíricas e da disponibilidade computacional.

Foram inseridas estruturas de variância e covariância no modelo não linear misto de Schumacher e Hall, visto que é coerente pensar em uma maior correlação entre as árvores pertencentes a uma mesma unidade de manejo. Este processamento foi feito por meio do pacote *nlme*, do *software* R, e sua função *correlation* (PINHEIRO; BATES, 2000). Foram utilizadas 14 estruturas, entre as quais Componente de Variância (CV), Não Estruturada (NE), Simétrica Composta (SC), Autorregressiva de Primeira Ordem (AR(1)), Autorregressiva de Primeira Ordem Heterogênea (ARH(1)), Simétrica Composta Heterogênea (CSH), Toeplitz (TOEP), Autorregressiva de Primeira Ordem Médias Móveis (ARMA(1,1)), Toeplitz Heterogênea

(TOEPH), Ante-Dependência de Ordem 1 (ANTE(1)), Correlação sem Estrutura (CSE), Espacial (SP(POW)(c-list)), Diagonal Principal (UN(1)) e Huynh-Feldt (H-F). Mais detalhes sobre essas estruturas podem ser encontrados em (PINHEIRO; BATES, 2000; DAVIDIAN; GILTINAN, 1995).

Para a escolha da estrutura da matriz de variância e covariância, foi utilizado o Critério de Informação de Akaike (AIC) (SAKAMOTO et al., 1986) (4), em que o melhor modelo é o que apresentar o menor valor de AIC; o Teste de Razão da Máxima Verossimilhança (TRMV) (PINHEIRO; BATES, 2000); e o Teste de Vuong (1989).

$$AIC = -2 * \ln(mv) + 2p \quad (4)$$

em que: AIC é o Critério de Informação de Akaike; \ln é o logaritmo neperiano; mv = valor de máxima verossimilhança; p é o número de parâmetros do modelo.

O teste de razão da máxima verossimilhança (TRMV) (5) consiste em comparar modelos dois a dois, nos quais o seu valor é obtido pela diferença dos valores das suas funções de verossimilhança (PINHEIRO; BATES, 2000).

$$TRMV = 2 * [\ln(mv_2) - \log(mv_1)] \quad (5)$$

em que: \ln é o logaritmo neperiano; mv_2 é o valor de função de máxima verossimilhança do modelo 2; mv_1 é o valor de função de máxima verossimilhança do modelo 1.

Para comparação entre os modelos, por meio do teste de razão de verossimilhança - T_{RLNN} de Vuong (1989) (6), utilizou-se a estatística descrita a seguir:

$$T_{RLNN} = \frac{1}{\sqrt{n}} \frac{\log(\widehat{mv}, \widehat{\beta})}{\widehat{\omega}^2} \quad (6)$$

em que:

$$\widehat{\omega}^2 = \frac{1}{n} \sum_{i=1}^n \left(\log \frac{f(y_i|x_i, \widehat{mv})}{g(y_i|x_i, \widehat{\beta})} \right)^2 - \left[\frac{1}{n} \sum_{i=1}^n \left(\log \frac{f(y_i|x_i, \widehat{mv})}{g(y_i|x_i, \widehat{\beta})} \right) \right]^2 \quad (7)$$

é um estimador para variância de $\frac{1}{\sqrt{n}}\log(\widehat{mv}, \widehat{\beta})$ e $\log(\widehat{mv}, \widehat{\beta})$ é o teste da razão de verossimilhança. A estatística tem, assintoticamente,

$$TLRNN \xrightarrow{d} N(0,1) \quad (8)$$

sob

$$H_0 : \log \frac{f(y_i|x_i, \widehat{mv})}{g(y_i|x_i, \widehat{\beta})} = 0 \quad (9)$$

ou seja, os modelos são equivalentes ao nível α de significância, sendo $Z_{\alpha/2}$ o valor crítico da distribuição normal padrão, rejeitando-se a hipótese nula se $|TLRNN| > Z_{1-\alpha/2}$.

A avaliação dos modelos volumétricos não lineares de Schumacher e Hall, em suas formas fixa e mista, adotando estrutura na matriz de variância e covariância, baseou-se no erro padrão residual (S_{yx}) (10), na correlação entre valores estimados e observados, gráficos de dispersão dos resíduos e gráficos de dispersão dos volumes estimados e observados.

$$S_{yx}(\%) = \frac{\sqrt{\frac{SQR}{n-p-1}}}{\bar{Y}} * 100 \quad (10)$$

em que: n é o número de casos; \bar{Y} é a média dos valores observados; SQR é a soma de quadrados do resíduos.

3 RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta os parâmetros ajustados do modelo não linear de Schumacher e Hall, sendo que todos os coeficientes foram significativos a um nível de 0,05, pelo teste *t-Student*.

Tabela 2 - Significância dos parâmetros ajustados para o modelo não linear de Schumacher e Hall.

F.V.	Parâmetros	Erro padrão	t value	Pr (>t)	
β_0	4,15e ⁻⁵	2,81e ⁻⁶	14,77	<0,0000	*
β_1	1,7111	0,0246	69,59	<0,0000	*
β_2	1,2040	0,0358	33,64	<0,0000	*

Fonte: Do autor

em que: F.V.: fonte de variação; β_i : coeficientes da regressão; *: significativo a 0,05.

Após o ajuste inicial do modelo não linear de Schumacher e Hall, foram ajustados os modelos mistos, com a inclusão de efeitos aleatórios, considerando as 14 estruturas para a matriz de variância e covariância.

A Tabela 3 apresenta os critérios de seleção para as estruturas de matriz de variância e covariância avaliadas. Dentre elas, a que melhor se ajustou à estimação volumétrica dos eucaliptos foi a Huynh-Feldt (H-F), que apresentou menor valor de AIC e maior valor de logaritmo da verossimilhança (LogLik). Além disso, realizando o teste da razão de verossimilhança e tomando tal estrutura como hipótese alternativa, comparou-se com as demais estruturas.

Tabela 3 – Critérios de seleção das estruturas da matriz de variância e covariância no ajuste do modelo volumétrico de efeitos mistos.

Modelo	AIC	LogLik	Teste	TRMV	p-valor
1. Toeplitz (TOEP)	4201,92	2802,12	1 vs 7	555,58	<0,000
2. Diagonal	4149,67	2804,17	2 vs 7	502,92	<0,000
3. Simétrica C. (CS)	4155,61	2802,47	3 vs 7	518,33	<0,000
4. Correlação sem estrutura	4139,32	2804,45	4 vs 7	773,58	<0,000
5. Não estruturada	3951,12	3330,80	5 vs 7	2,13	<0,632
6. Simétrica C.H. (CSH)	3949,75	3330,45	6 vs 7	2,00	<0,563
7. Huynh-Feldt (H-F)	3940,31	3350,68	-	-	-

Fonte: Do autor.

É possível observar que as hipóteses nulas foram praticamente todas rejeitadas (p -valor $< 0,05$), com exceção para estrutura Simétrica Composta Heterogênea (CSH) (p -valor = 0,563) e Não Estruturada (NE) (p -valor = 0,632), que não apresentaram diferenças significativas em relação à Huynh-Feldt (H-F). Os resultados indicam que a utilização de qualquer uma destas três estruturas para a matriz de variância e covariância seria adequada ao conjunto de dados. De acordo com West, Welch e Gatecki (2015), a estrutura Huynh-Feldt (HF) é caracterizada pelas variâncias desiguais entre unidades de manejo e covariâncias obtidas pela média aritmética entre as variâncias e subtraindo λ , sendo λ a diferença entre a média das variâncias e a média das covariâncias. Na estrutura Simétrica Composta Heterogênea (CSH) aplicam-se diferentes variâncias e algumas covariâncias desiguais, ajustadas por um coeficiente de correlação ρ entre indivíduos. Na Não Estruturada também são atribuídas variâncias e covariâncias distintas para cada ocasião.

Decidiu-se utilizar a estrutura Huynh-Feldt (HF) devido ao critério do AIC e LogLike.

De acordo com West, Welch e Gatecki (2015), as estruturas de variância e covariância permitem uma maior flexibilidade nas correlações. Em qualquer análise de dados, deve-se buscar escolher uma estrutura correta para estas matrizes que seja mais adequada e parcimoniosa, com conhecimento dos dados observados e sobre as relações entre as observações de cada unidade amostral (LITTELL et al., 2006), neste caso a unidade de manejo, uma vez que diferentes números de variâncias e covariâncias devem ser estimados (PERRI; IEMMA, 1999).

Na Tabela 4, observam-se os resultados do ajuste (parâmetros fixos) do modelo misto não linear de Schumacher e Hall, por meio do método da máxima verossimilhança, adotando estrutura Huynh-Feldt (HF) na matriz de variância e covariância, em que todos os parâmetros apresentaram significância a um nível de 0,05, pelo teste *t-student*.

Tabela 4 - Significância dos parâmetros fixos ajustados para modelo não linear misto de Schumacher e Hall.

F.V.	Parâmetros	Erro padrão	t value	Pr (>t)	
β_0	$3,4645e^{-5}$	$1,59e^{-6}$	21,81	<0,0000	*
β_1	1,8184	0,0200	90,85	<0,0000	*
β_2	1,1655	0,0282	41,38	<0,0000	*

Fonte: Do autor.

em que: F.V.: fonte de variação; β_i : coeficientes da regressão; *: significativo a 0,05.

Os resultados obtidos de erro padrão residual para os modelos mistos foram inferiores aos valores obtidos a partir dos modelos fixos, como são apresentados na Tabela 5.

Tabela 5 - Critérios de avaliação da qualidade dos modelos não lineares como efeitos fixos e efeitos mistos.

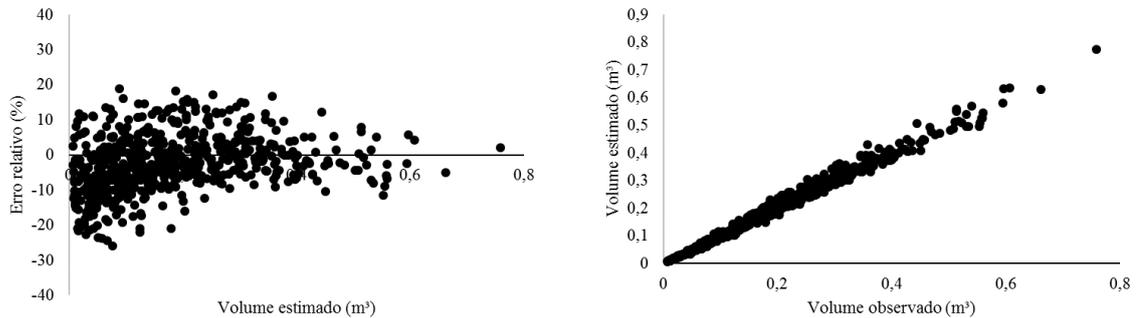
	Correlação	Syx (m³)	Syx (%)
Modelo não linear	0,9943	0,0142	7,9830
Modelo não linear misto	0,9999	0,0003	0,1516

Fonte: Do autor.

Entre as vantagens dos modelos que incluem os efeitos aleatórios, com relação aos modelos de efeitos fixos, é a redução do erro padrão residual. Calegario et al. (2005) estudaram a área basal de clones de *Eucalyptus* e observaram redução de, aproximadamente, 15 vezes. Neste trabalho, também foi observada uma redução expressiva do erro padrão residual, que foi de 0,0142 no modelo não linear de regressão fixa para 0,0003 no modelo não linear de regressão mista, uma diminuição de 53 vezes.

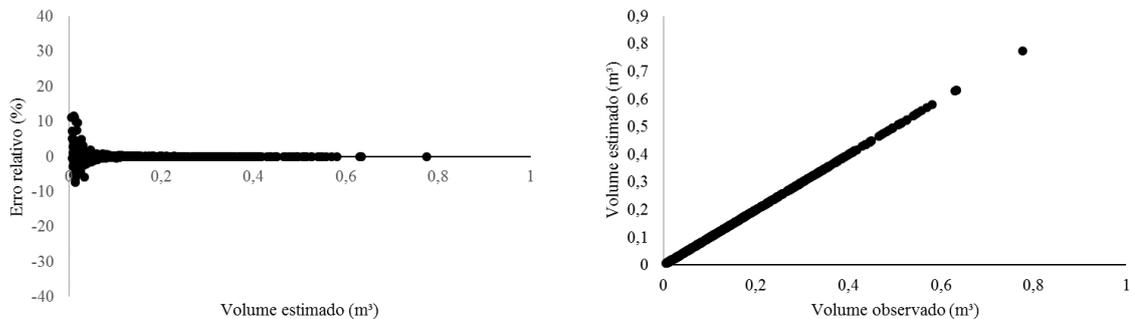
Dando sequência à avaliação dos modelos, o próximo passo foi a análise de resíduos. As Figuras 2 e 3 apresentam os gráficos de dispersão de resíduos e dos volumes estimados e observados.

Figura 2 - Gráficos de dispersão de resíduos e de volumes estimados *versus* observados para o modelo de Schumacher e Hall não linear fixo.



Fonte: Do autor.

Figura 3 - Gráficos de dispersão de resíduos e de volumes estimados *versus* observados para o modelo de Schumacher e Hall não linear misto com adoção de estrutura de covariância.



Fonte: Do autor.

Verifica-se que o modelo misto não linear com estrutura Huynh-Feldt (HF) na matriz de variância e covariância apresentou alta precisão quando comparado com o modelo não linear de regressão fixa, com resíduos bem distribuídos e média em torno de zero. Houve maior tendência de erro nos menores valores de volume, no entanto estes erros foram inferiores a 10 %. Já o modelo não linear fixo apresentou uma tendência de superestimar os menores valores de volume e subestimar os valores intermediários. O modelo de regressão fixa apresentou um erro relativo médio de 6,61 %, enquanto para o modelo de regressão mista este valor foi de 0,43 %. Carvalho et al. (2011), ao aplicarem modelagem mista na predição da área basal e do volume, verificaram uma redução do erro de 15% para 12% na predição da área basal e de 26% para 4% na predição do volume.

O modelo não linear misto mostrou-se bastante eficaz na modelagem do volume de clones de *Eucalyptus* spp, uma vez que este tipo de modelo permite estimar os efeitos fixos, prever os efeitos aleatórios e estimar os componentes de variância, considerando a

variabilidade existente em cada árvore e entre os diferentes unidades de manejo avaliadas, sendo recomendada a sua aplicação, devido ao ganho em precisão.

3 CONCLUSÃO

O modelo não linear misto de Schumacher e Hall com estrutura Huynh-Feldt (HF) na matriz de variância e covariância apresentou alta precisão nas estimativas do volume individual de árvores de *Eucalyptus* spp. quando comparado ao ajuste com apenas efeitos fixos.

REFERÊNCIAS

- ALVARES, C.A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, p. 711-728, 2013.
- ANDRADE, V. C. L. **Um método para descrever o perfil do tronco em árvores de eucalipto utilizando geometria analítica**. Dissertação (Mestrado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa. 2001. 70 p.
- CALEGARIO, N. et al. Estimativa do crescimento de povoamentos de *Eucalyptus* baseada na teoria dos modelos lineares em multiníveis de efeitos mistos. **Revista Árvore**, Santa Maria, v. 29, n. 2, p. 251 - 261, 2005.
- COSTA, B. R. et al. Avaliação genética de indivíduos de erva-mate (*Ilex paraguariensis* St.-Hil.) na região de Caarapó, MS, pelo procedimento REML/BLUP. **Revista Ciência Florestal**, Santa Maria, v. 16, n. 2, p. 145 - 151, 2005.
- DAVIDIAN, M.; GILTINAN, D. M. **Nonlinear models for repeated measurement data**. New York: Chapman and Hall, v. 1, 1995. 360 p.
- LAPPI, J. Calibration of height and volume equations with random parameters. **Forest Science**, Lawrence, v. 37, n. 3, p. 781 - 801, 1991.
- LITTEL, R. C. et al. **SAS system for mixed models**. Cary: SAS Institute, 1996. 633p.
- MACHADO, S. A; CONCEIÇÃO, M. B.; FIGUEIREDO, D. J. Modelagem do volume individual para diferentes idades e regimes de desbaste em plantações de *Pinus oocarpa*. **Revista Ciências Exatas e Naturais**, Curitiba, vol. 4, n. 2, 2002.
- MORA, F. Herebabilidad y valor genético (REML/BLUP) en genotipos de un eucalipto tolerante a la sequía en el Norte de Chile. **Revista Ciência Florestal**, Santa Maria, v. 16, n. 2, p. 145 - 151, 2006.
- PERRI, S. H. V; IEMMA, A. F. Procedimento MIXED do SAS® para análise de modelos mistos. **Scientia Agricola**, v. 56, n. 4, p. 959-967, 1999.
- PINHEIRO, J. C.; BATES, D. M. **Mixed-effects models in S and S-PLUS**. New York: Springer-Verlag, 2000. v. 1. 528 p.
- SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel, p. 81, 1986.
- SCHUMACHER, F. X.; HALL, F. S. Logarithmic expression of timber-tree volume. **Journal of Agricultural Research**, Lahore, v. 47, n. 9, p. 719 - 734, 1933.
- SCOLFORO, J. R. S. et al. Estimativa de volume, peso seco, peso de óleo e quantidade de moirões para a candeia (*Eremanthus erythropappus* (DC.). Macleish). **Cerne**, Lavras, v. 10 n. 1, p. 87 - 102, 2004.

SEARLE, S.R. **Linear models for unbalanced data**. New York: John Wiley, 1987. 536p.

THOMAS, C. et al. Comparação de equações volumétricas ajustadas com dados de cubagem e análise do tronco. **Ciência Florestal**, Santa Maria, v. 16, n. 3, p. 319 - 327, 2006.

TORAL, F.L.B.; ALENCAR, M.M.; FREITAS, A.R. Arranjos para efeitos fixos e estruturas de (co)variâncias residuais para análises de medidas repetidas do peso de bovinos da raça Canchim. **Revista Brasileira de Zootecnia**, v. 35, n. 5, p. 1951-1958, 2006.

VUONG, Q. Likelihood ratio tests for model selection and nontested hypotheses. **Econometrica**, v. 57, p. 307 - 333, 1989.

WOLFINGER, R. D. Covariance structure selection in general mixed models. **Communications in Statistics**, v. 22, p. 1079-1106, 1993.

WEST, B. T.; WELCH, K. B.; GATECKI, A. T. **Linear mixed models: a practical guide using statistical software**. 2 ed. Ann Arbor: Chapman and Hall/CRC, 2015. 405 p.

**ARTIGO 4 - ESTIMADOR GEOESTATÍSTICO PARA PREDIÇÃO VOLUMÉTRICA
POR MEIO DE DIFERENTES INTENSIDADES AMOSTRAIS**

ARTIGO FORMATADO DE ACORDO COM A NBR 6022 (ABNT, 2003).

RESUMO

Inventários florestais possuem custos elevados devido ao tempo necessário para os levantamentos, alocação de parcelas na área, e medição das variáveis de interesse. Nesse sentido, destaca-se a existência da demanda por uma intensidade amostral capaz de reduzir o número de parcelas por hectare e ao mesmo tempo proporcionar boa representatividade da população, para que essa redução não interfira na acurácia de estimação dos parâmetros populacionais. O objetivo deste trabalho foi avaliar o desempenho de um estimador geoestatístico para estimar a volumetria de um povoamento de eucalipto considerando diferentes intensidades amostrais. A base de dados foi proveniente de inventários florestais que antecederam em 3 meses as operações de corte na floresta, e foram realizados aos 8 anos de idade. A intensidade amostral foi de uma parcela a cada 5 ha aproximadamente, totalizando 220 parcelas inventariadas em campo. Os dados foram divididos em dois grupos: 80% para ajuste dos semivariogramas e 20% para validação. Dentre os dados destinados aos ajustes, foram selecionadas parcelas que representassem quatro diferentes intensidades amostrais (1:5, 1:10, 1:15 e 1:20), e posteriormente, ajustados os modelos de semivariogramas. Os modelos foram então utilizados para estimar a volumetria das parcelas destinadas à validação. Os resultados permitiram inferir que os modelos esféricos ajustados para as diferentes intensidades amostrais apresentaram desempenhos satisfatórios e próximos entre si, com erros inferiores a 10 %. O menor valor foi apresentado na intensidade amostral de 1:5, 7,33 %, e o maior na intensidade de 1:20, 8,90 %. Assim sendo, foi possível concluir que o estimador geoestatístico permitiu a redução da intensidade amostral em inventários que antecederam o corte de povoamentos clonais de *Eucalyptus*, sem grandes perdas na precisão.

Palavras-chave: Semivariograma. Amostragem Florestal. Modelo geoestatístico.

ABSTRACT

Forest inventories have high costs due to the time required for the planning, allocating of plots in the area and measuring of the variables of interest. In this sense, there is a demand for a sample intensity that reduces the number of parcels per hectare and at the same time provides a good representation of the population, so that this reduction does not interfere in a negative way in the quality of the estimates. The objective of this work was to evaluate the performance of a geostatistical estimator to estimate the wood volume per hectare of a settlement of eucalyptus, considering different sample intensities. The database was derived from pre-cut forest inventories, carried out at the age of eight years, with a sampling intensity of one plot per 4.95 ha (approximately, 1:5), totaling 220 plots. Data were divided into two groups: 80% for semivariograms adjustments and 20% for validation. Sampling units representing four different sample intensities (1:5, 1:10, 1:15 and 1:20) were selected and the semivariograms models were adjusted. The models were then used to estimate the volumes per hectare of the parcels intended for validation. It was verified that the spherical models adjusted for the different sample intensities presented satisfactory and similar performances with each other, with mean relative errors lower than 10%. The lowest value was presented in the sampling intensity of 1:5, 7.33%, and the highest in the intensity of 1:20, 8.90%. An average relative error difference of only 1.57%. Therefore, it is possible to reduce the sample intensity in pre-cut inventories for Eucalyptus stands, without great losses in accuracy, with the use of point kriging to obtain the volume of wood per hectare in non-sampled points.

Keywords: Semivariogram. Forest inventory. Allocation of plots. Spherical model.

1 INTRODUÇÃO

Uma das informações de maior importância para o conhecimento do potencial de uma floresta, em uma dada região, é a variável volume de madeira, sendo imprescindível para o planejamento do manejo sustentável de florestas. O volume individual é um ponto de partida para a avaliação do conteúdo de madeira em um povoamento florestal (MACHADO et al., 2002) e serve como base para as decisões relacionadas a tratos silviculturais, colheita e transporte da produção volumétrica da madeira. Assim, é fundamental que a obtenção do volume de madeira seja realizada corretamente, representando bem a população amostrada (ANDRADE, 2001).

A necessidade do desenvolvimento e da utilização de ferramentas estatísticas, capazes de caracterizar a estrutura de variabilidade espacial de um determinado fenômeno, motivou o surgimento da teoria das variáveis regionalizadas ou, simplesmente, geoestatística (NOGUEIRA, 2013). De acordo com Couto et al. (1997), a geoestatística tem grande utilidade para caracterizar e mapear a variação espacial das propriedades dos sistemas.

Segundo Landim (2003), as raízes da geoestatística estão na indústria de minérios, em meados da década de 50, durante pesquisas desenvolvidas para cálculo de reservas minerais na África do Sul por Daniel G. Krige e o estatístico H. S. Sichel, para dados de mineração do ouro. Concluiu-se que as variâncias dos dados possuíam uma estruturação que dependia da amostragem, e que as variâncias obtidas por meio da abordagem clássica de amostragem não faziam sentido se não considerasse as distâncias entre as amostras. Baseado nessas observações, Matheron (1963) desenvolveu a Teoria das Variáveis Regionalizadas, definida como variáveis cujos valores estão relacionados de algum modo, com a posição espacial que ocupam, ou seja, que variam de um lugar a outro com certa continuidade.

Esta variável atua num espaço geométrico em que foi definida e onde será estudada sua variação, apresentando localização, continuidade e anisotropia, que estão ligadas ao fenômeno natural que elas representam (JOURNAL, 2003). Nogueira (2013) afirma que as variáveis regionalizadas apresentam dupla característica: aleatoriedade, uma vez que seus valores numéricos observados podem variar, consideravelmente, de um ponto a outro no espaço e, espacialidade, pois estes valores não são inteiramente independentes, apresentando uma aparente continuidade espacial, na qual podem ser captados e modelados matematicamente por técnicas geoestatísticas.

Segundo Mello (2004), a teoria das variáveis regionalizadas pressupõe que cada dado é modelado com uma variável aleatória que pode ser expressa pela soma de três componentes:

um componente estrutural, associado a um valor médio constante ou a uma tendência constante; uma componente aleatória, espacialmente correlacionada; e um erro aleatório constante.

Um aspecto que deve ser considerado, com a adoção destas técnicas como ferramenta de modelagem no manejo florestal, é a possibilidade de redução do número de medições necessárias, sem perda na qualidade das estimativas. Isso acarretaria em uma diminuição no tempo e custo dos inventários florestais. Para isso, são necessários estudos que forneçam subsídios para o manejador no processamento de dados de inventário florestal.

O objetivo deste trabalho foi avaliar o desempenho do estimador geoestatístico para estimar a volumetria de um povoamento de eucalipto sob diferentes intensidades amostrais.

2 MATERIAIS E MÉTODOS

2.1 Dados

2.1.1 Área de estudo

A área de estudo contempla 21 unidades de manejo de clones de *Eucalyptus* spp, e está localizada nos municípios de Bom Jardim de Minas, Lima Duarte e Arantina, em Minas Gerais, Brasil, totalizando 1090 ha. O clima da região é do tipo tropical de altitude, do tipo Cwb, segundo a classificação de Köppen, com temperatura média anual de 20,1 °C, com invernos secos e frios, com ocorrência de geadas em algumas áreas, e verões chuvosos com temperaturas moderadamente altas. A precipitação total média anual é de 1456 mm (ALVARES, 2013).

2.1.2 Base de dados

Os dados utilizados neste estudo foram provenientes de inventários florestais temporários, em que foram mensurados 27299 indivíduos, distribuídos em 565 parcelas. As parcelas foram demarcadas aleatoriamente, de modo a representarem toda a área, tendo como variáveis numéricas: idade (meses), altura (metros), diâmetro à altura do peito (centímetros), área basal da parcela (m²/ha) e índice de sítio (m) (TABELA 1).

Tabela 1 - Estatísticas descritivas dos dados de inventário florestal de *Eucalyptus* spp. utilizados.

Variável	Mínimo	Média	Máximo	Desvio Padrão	CV%
H	5,00	21,26	31,50	3,94	18,54
DAP	5,00	14,51	26,20	2,97	20,44
I	36	59	78	9	15,93
G	6,05	17,28	34,21	5,11	29,59
S	12,50	25,84	32,50	3,27	12,67

Fonte: Do autor.

em que: H é a altura em metros; DAP é o diâmetro, em centímetros, a 1,30 metros de altura do solo; I é a idade, em meses; G é a área basal, em m²/ha; e S é o índice de sítio, em metros.

Para as árvores que não tiveram altura medida em campo, as alturas foram estimadas por meio de um modelo hipsométrico não linear misto (1).

$$H_{ij} = \left[\left(\beta_0 + \beta_1 * S_{ij} + \beta_2 * \left(\frac{1}{I_i} \right) \right) * e^{\beta_3 + \beta_4 * \left(\frac{1}{G_i} \right) + \beta_5 * \left(\frac{1}{I_i} * \frac{1}{DAP_{ij}} \right)} \right] + \varepsilon_{ij} \quad (1)$$

em que: H_{ij} é a altura da i -ésima árvore (m) da j -ésima unidade amostral; β_k são os coeficientes da regressão; DAP_{ij} é o diâmetro à altura do peito da i -th árvore (cm) da j -ésima unidade amostral; I_i é a idade da i -th árvore (meses); G_i é a área basal da i -th parcela (m^2/ha); S_{ij} é o índice de sítio da j -th unidade amostral contendo a i -th árvore (m); ε é o erro aleatório (m).

Os volumes de madeira das árvores foram estimados por meio do modelo não linear de Schumacher e Hall (1933) (2), ajustado com regressão mista, em que os efeitos aleatórios foram as variabilidades existentes em cada indivíduo e entre as diferentes unidades de manejo. Para o ajuste foram utilizados dados originados de cubagem rigorosa, realizada entre os anos de 2010 e 2016, composta por 818 árvores, em que foram medidos a altura total (Ht), em metros, o diâmetro à altura do peito (DAP), em centímetros, e os diâmetros na base das árvores (a 0,1 m de altura) e nas alturas de 0,5 m; 1 m; 1,5 m e 2 m; e, a partir desta secção, a cada 2 m. Os volumes individuais foram obtidos pela fórmula de Smalian.

$$V_i = \beta_0 * DAP_i^{\beta_1} * Ht_i^{\beta_2} + \varepsilon_i \quad (2)$$

em que: V_i é o volume em m^3 da i -ésima árvore; DAP_i é o diâmetro, em cm, a 1,30 m de altura da i -ésima árvore; Ht é a altura total, em m, da i -ésima árvore; β_0 , β_1 e β_2 são os parâmetros do modelo; ε_i é o erro aleatório (m^3).

As estatísticas descritivas dos dados são apresentadas na Tabela 2.

Tabela 2 - Estatísticas descritivas dos dados de cubagem rigorosa de *Eucalyptus* spp.

Variável	Mínimo	Médio	Máximo	Desvio padrão	CV%
Volume (m^3)	0,0066	0,1785	0,7752	0,1335	74,79
DAP (cm)	4,19	13,60	27,00	4,45	32,72
Altura (m)	8,10	21,82	33,08	5,10	23,37

Fonte: Do autor.

em que: DAP é o diâmetro a 1,30 de altura do solo.

A equação ajustada foi aplicada aos dados de inventário florestal, em que foram estimados os volumes de madeira de cada fuste e, posteriormente, estimado o volume de madeira por hectare de cada parcela. Foram selecionadas, dentre os dados utilizados neste estudo, parcelas georreferenciadas, oriundas de inventários florestais pré-corte realizados aos 8 anos de idade dos povoamentos, resultando em um total de 220 parcelas, que representam uma intensidade amostral de uma parcela a cada 4,95 ha (aproximadamente, 1:5).

2.2 Análise exploratória

Inicialmente, foi realizada uma análise exploratória, com a finalidade de observar a tendência espacial dos dados e detectar a presença de *outliers*. Foram calculadas as estatísticas descritivas de tendência central e de dispersão (média, mediana, moda, desvio padrão e coeficiente de variação), e analisados os comportamentos da variável regionalizada em função da longitude e da latitude. Para verificar a hipótese de normalidade dos dados utilizou-se o teste de Shapiro-Wilk, que testa a hipótese de que a amostra é proveniente de uma população com distribuição normal, ou seja, a característica apresenta distribuição unimodal, em geral simétrica, que se aproxima da distribuição normal (SHAPIRO; WILK, 1965).

2.3 Intensidade amostral

Para caracterizar o padrão espacial da variável volume, e avaliar a influência da intensidade amostral de parcelas na precisão das estimativas volumétricas a nível de povoamento, procedeu-se às análises geoestatísticas.

Os dados foram divididos em dois grupos: 80% para ajuste dos semivariogramas e 20% para validação. Dentre os dados destinados aos ajustes dos semivariogramas, foram selecionadas unidades amostrais, distribuídas de forma a representar a área, que simulassem quatro diferentes intensidades amostrais: 1 parcela a cada 5 ha; 1 parcela a cada 10 ha; 1 parcela a cada 15 ha e 1 parcela a cada 20 ha (MELLO et al., 2009). Para cada intensidade amostral foram calculados o erro padrão da média, erro do inventário e, conseqüentemente, o intervalo de confiança, de acordo com os estimadores clássicos da Amostragem Casual Simples (ACS) (COCHRAN, 1977).

2.4 Análise variográfica

Para cada intensidade amostral, foram ajustados três modelos de semivariograma: esférico (3), exponencial (4) e gaussiano (5), pelo método da máxima verossimilhança (WONNACOTT; WONNACOTT, 1972). O processamento foi feito no *software* R, versão 3.4.1, por meio do pacote geoR (RIBEIRO JÚNIOR; DIGLLE, 2001).

$$\gamma_{\text{sph}}(h) = C_0 + C_1 \left[\frac{3}{2} \left(\frac{h}{\phi} \right) - \left(\frac{1}{2} \left(\frac{|h|}{\phi} \right)^3 \right) \right] \quad (3)$$

$$\gamma_{\text{exp}}(h) = C_0 + C_1 \left[1 - \exp \left(-3 \left(\frac{|h|}{\phi} \right) \right) \right] \quad (4)$$

$$\gamma_{\text{gauss}}(h) = C_0 + C_1 \left[1 - \exp \left(-3 \left(\frac{|h|}{\phi} \right)^2 \right) \right] \quad (5)$$

em que: γ é a semivariância da variável de interesse; C_0 é o efeito pepita; C é o patamar; ϕ é o alcance; h é a distância; *exp* é o exponencial.

Para a escolha do modelo mais adequado a cada conjunto de dados, o critério utilizado foi o menor valor do Critério de Informação de Akaike (AIC) (6) (SAKAMOTO et al., 1986; MC BRATNEY; WEBSTER, 1986). Segundo Biase et al., (2012) a utilização deste critério na seleção do modelo é mais indicada, devido à sua maior sensibilidade. Se a diferença entre os modelos for superior a dois, indica que estes são consideravelmente diferentes entre si (BURNHAN; ANDERSON, 2002).

$$AIC = -2 * \ln(mv) + 2p \quad (6)$$

em que: *AIC* é critério de informação Akaike; \ln é o logaritmo neperiano; *mv* é o valor de máxima verossimilhança; e *p* é o número de parâmetros do modelo.

2.5 Krigagem

Após o ajuste e seleção dos modelos de semivariogramas, para cada intensidade amostral, foram realizadas interpolações dos dados pelo método da krigagem ordinária pontual, que considera a dependência espacial e estima valores sem tendência e com variância mínima (VIEIRA, 2000).

Os modelos de semivariogramas foram então utilizados para prever a volumetria das parcelas que constituem o conjunto de dados destinados à validação. Esse processamento foi realizado no software ArcGIS 10.1 (ESRI, 2011), por meio da ferramenta *kriging* (*Geostatistical Analyst*). Para avaliação da precisão dos modelos foram calculados os Erros Relativos Médios (ERM%) (7) entre os volumes de madeira obtidos pelas diferentes intensidades amostrais (V_{est}) e o volume de madeira observado (V_{obs}); gráficos de distribuição dos volumes de madeira estimados e observados; gráficos de erros percentuais em função dos valores observados; e os coeficientes de correlação entre os volumes de madeira estimados e observados.

$$ERM \% = \frac{(V_{est} - V_{obs})}{V_{obs}} \times 100 \quad (7)$$

em que: V_{est} é o volume de madeira estimado e V_{obs} é o volume de madeira observado.

3 RESULTADOS E DISCUSSÃO

A Tabela 3 apresenta os resultados da análise exploratória dos dados de volumetria do povoamento considerando os dados de ajuste e validação. Verificou-se pelo teste de Shapiro-Wilk que os dados se aproximam da distribuição normal ($w < 0,95$).

Tabela 3 - Estatísticas descritivas dos dados volumétricos do povoamento de eucalipto.

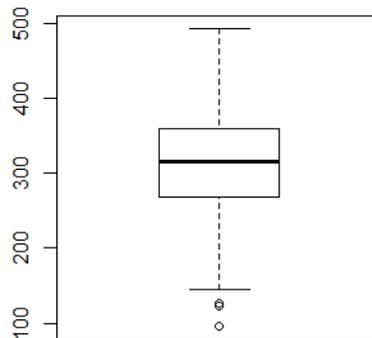
Estatística	Volume (m³.ha⁻¹)
Média (m ³ .ha ⁻¹)	315,45
Mediana (m ³ .ha ⁻¹)	316,25
Desvio padrão (m ³ .ha ⁻¹)	77,83
Coefficiente de variação (%)	24,36
Teste de Shapiro-Wilk	0,4268

Fonte: Do autor.

Segundo Cressie (1993) e Diggle e Ribeiro Júnior (2007), a aplicação da geoestatística não depende de uma distribuição normal dos dados, no entanto, a presença da normalidade permite que sejam feitas inferências com propriedades diferentes, como por exemplo, o uso da Máxima Verossimilhança, como foi feito neste trabalho.

Foram observados dados que podem ser considerados *outliers*, ou seja, que apresentam comportamento diferente em relação aos demais (FIGURA 1). Segundo Diggle e Ribeiro Júnior (2000), estes dados provocam forte impacto no ajuste do semivariograma e, portanto, foram removidos da base de dados para que não comprometessem o estudo variográfico.

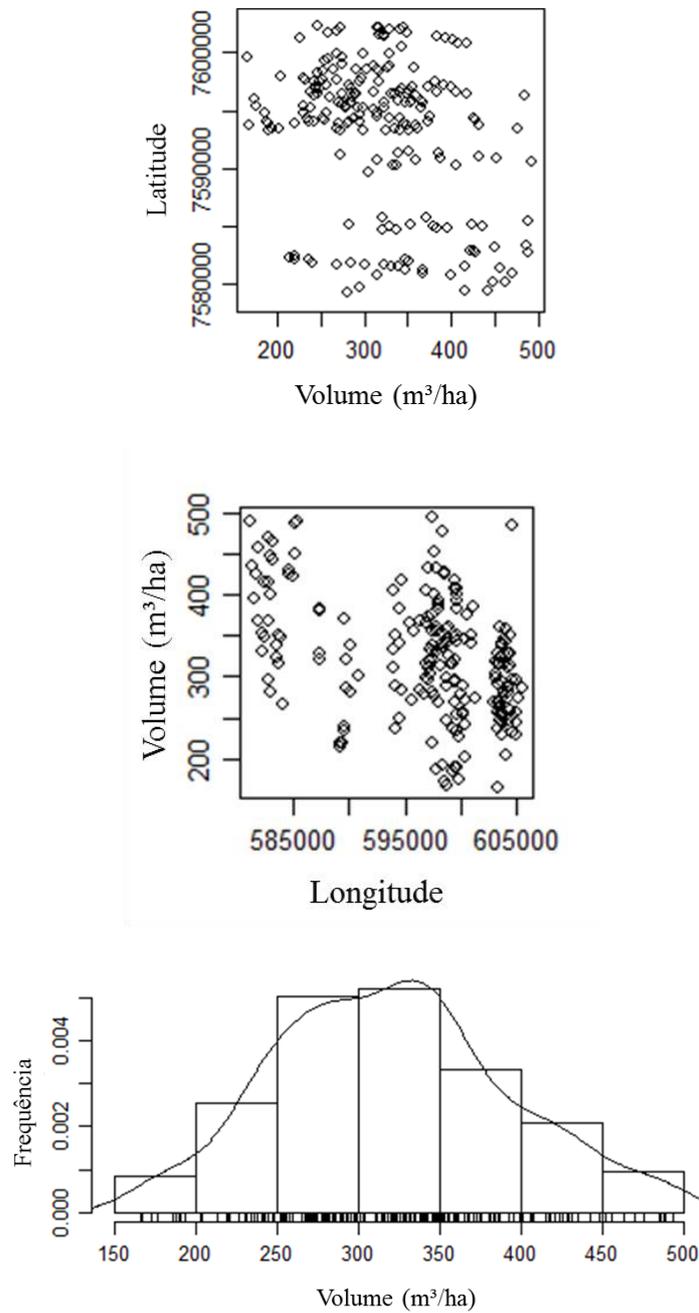
Figura 1 - Box plot de dados volume por hectare.



Fonte: Do autor.

Os principais gráficos a serem observados na análise exploratória de dados são apresentados na Figura 2. Os gráficos A e B mostram o comportamento da variável regionalizada (volume de madeira por hectare) em função da latitude e da longitude. Verificou-se inexistência de tendência para a variável volume nos dois sentidos. Esta é uma característica importante em estudos geoestatísticos, pois indica a possibilidade de se aplicar a Hipótese Intrínseca descrita por: Isaaks e Srivastava (1989); Journel e Huijbrets (1978); Ribeiro Júnior (1995); e Vieira (2000). O gráfico C mostra que os dados utilizados têm uma distribuição aproximadamente normal, corroborando os resultados da Tabela 3.

Figura 2 - Gráficos exploratórios da variável regionalizada volume por hectare; A: volume em função da latitude; B: volume em função da longitude; C: histograma de frequência.



Fonte: Do autor.

As estatísticas geradas pelo processamento dos estimadores da amostragem casual simples (TABELA 4) indicaram que, em todas as quatro intensidades amostrais, o erro de amostragem se manteve abaixo de 10 %. O menor valor de erro de amostragem foi apresentado pela intensidade 1:5, 3,06 %, e o maior valor pela intensidade 1:20, 9,50 %. Pode-se considerar que a intensidade de 1:20 proporcionou uma suficiência amostral, para um erro admissível de

10 %, que normalmente é o aceitável pelas empresas de base florestal e para os órgãos públicos florestais. No entanto, isso gerou uma considerável perda na precisão, visto que quanto maior o erro de amostragem, maior a amplitude do intervalo de confiança. A amplitude do intervalo de confiança com o processamento dos dados com intensidade 1:5 foi de 19,45 m³.ha⁻¹, enquanto para a intensidade de 1:20, este valor aumentou para 57 m³.ha⁻¹.

Tabela 4 - Estatísticas do processamento pelos estimadores da Amostragem Casual Simples para cada uma das quatro intensidades amostrais.

Resultados da ACS	Intensidades			
	1:5	1:10	1:15	1:20
Número de parcelas (n)	180	90	45	23
Média	14,50	14,47	15,37	13,67
Variância	12,82	15,12	13,00	15,69
Desvio Padrão	3,58	3,89	3,61	3,96
Coefficiente de variação	24,69	26,87	23,46	28,97
Erro Padrão da Média	0,27	0,41	0,54	0,76
Erro de amostragem	0,44	0,68	0,90	1,30
Erro de amostragem (%)	3,06	4,70	5,87	9,50
Número Ótimo de Parcelas	17	20	16	21
Intervalos de Confiança:				
Por parcela (m³)				
Volume médio	14,50	14,47	15,37	13,67
Limite superior	14,94	15,15	16,27	14,97
Limite inferior	14,06	13,79	14,46	12,37
Por hectare (m³/ha)				
Volume médio	317,95	317,43	336,98	299,87
Limite superior	327,68	332,34	356,76	328,37
Limite inferior	308,23	302,51	317,19	271,37
Para a população (m³)				
Média	346.569	345.994	367.303	326.858
Limite superior	357.166	362.251	388.867	357.919
Limite inferior	335.973	329.738	345.740	295.797

Fonte: Do autor.

Considerando os resultados obtidos pelo processamento dos estimadores da ACS, seria plausível considerar que as intensidades 1:5 e 1:10 apresentaram desempenhos satisfatórios, por apresentar intervalos de confiança próximos entre si. Em contrapartida, as intensidades de 1:15 e 1:20 proporcionaram perdas expressivas na precisão das estimativas. No entanto, vale ressaltar que os estimadores da ACS não consideram a autocorrelação entre as parcelas e que a utilização de um interpolador geoestatístico pode aumentar a eficiência da estimativa, mesmo com uma redução na intensidade amostral.

As análises variográficas indicaram que há uma estrutura espacial da variável volume de madeira por hectare, com semivariância de comportamento modelável, para todas as intensidades amostrais. Isso está de acordo, em partes, com os resultados observados por Mello et al. (2009), que detectaram continuidade espacial do volume de madeira por hectare em povoamentos de *Eucalyptus* spp., para as intensidades amostrais de 1:4 e 1:7. Em seu estudo, os autores, no entanto, não observaram estrutura espacial para a intensidade amostral de uma parcela a cada 10 ha e recomendaram tratar as informações de forma independente, com a aplicação da estatística clássica.

Na Figura 3 são apresentados os semivariogramas gerados para as diferentes intensidades amostrais. Verifica-se que as linhas dos modelos esférico (azul), exponencial (preta) e gaussiano (vermelha) praticamente se confundem após a estabilização da curva. Esta diferenciação é mais perceptível nas estimativas dos efeitos pepita.

Figura 3 - Modelos esférico (linha azul), exponencial (linha preta) e gaussiano (linha vermelha) ajustados aos semivariogramas para as intensidades amostrais 1:5, 1:10, 1:15 e 1:20. (Continua)

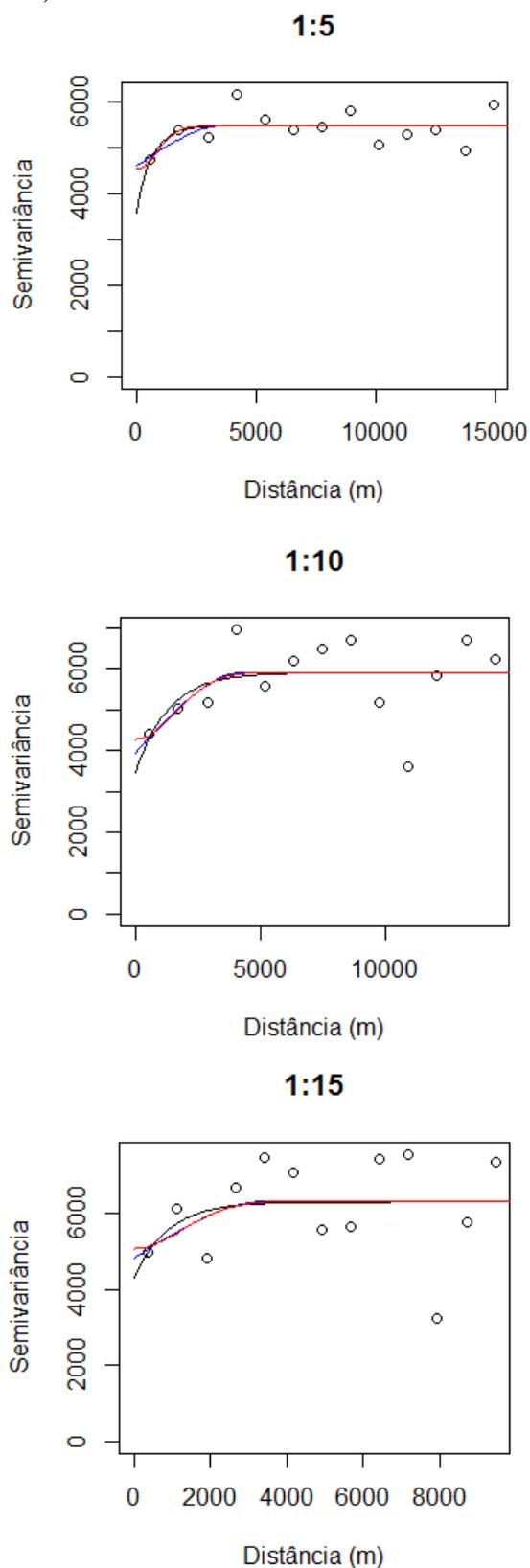
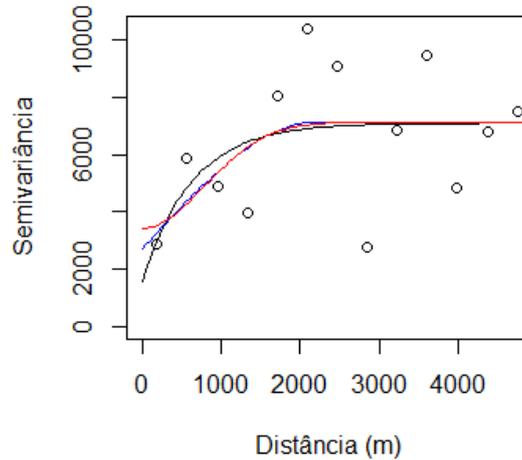


Figura 3 - Modelos esférico (linha azul), exponencial (linha preta) e gaussiano (linha vermelha) ajustados aos semivariogramas para as intensidades amostrais 1:5, 1:10, 1:15 e 1:20. (Conclusão)

1:20



Fonte: Do autor.

A seleção de um modelo com um bom desempenho, que represente de forma adequada as semivariâncias, especialmente nas pequenas distâncias, é importante para um melhor resultado no processo de krigagem (NIELSEN; WENDROTH, 2003). Por isso, além de uma análise visual do ajuste dos modelos aos diferentes semivariogramas, baseou-se também no Critério de informação de Akaike, o qual indicou que, para todas as intensidades amostrais avaliadas, o modelo esférico foi o que apresentou o melhor ajuste às semivariâncias, com menores valores de AIC. Os parâmetros obtidos com o ajuste dos modelos esférico, exponencial e gaussiano, para cada intensidade amostral, estão apresentados na Tabela 5.

Tabela 5 - Estimativa dos parâmetros efeito pepita (τ^2), patamar (σ^2) e alcance (θ) dos modelos esférico, exponencial e gaussiano ajustados aos semivariogramas da variável volume de madeira por hectare, para cada intensidade amostral.

Intensidade	Modelo	τ^2	σ^2	θ	AIC
1:05	esférico	4606,01	5496,43	3793,31	-1026,3
	exponencial	3595,66	5483,43	638,42	-993,96
	gaussiano	4540,53	5481,11	1212,21	-1015
1:10	esférico	3947,7	5907,81	4403,46	-993,09
	exponencial	3442,29	5901,98	1305,57	-960,75
	gaussiano	4269,43	5911	2230,2	-981,75
1:15	esférico	4813,07	6326,62	3568,14	-927,09
	exponencial	4283,33	6297,88	883,16	-896,75
	gaussiano	5071,04	6323,79	1784,12	-918,75
1:20	esférico	2702,74	7142,48	2216,61	-977,19
	exponencial	1552,08	7095,56	630,17	-904,85
	gaussiano	3430,01	7138,42	111,02	-965,85

Fonte: Do autor.

Com o intuito de avaliar o desempenho dos modelos de semivariogramas, ajustados com diferentes intensidades amostrais, para estimar o volume por hectare, procedeu-se a uma validação preditiva. Esta validação consiste em fazer a estimativa de valores já conhecidos, mas que não foram utilizados no ajuste dos modelos.

Após a estimação dos volumes por hectare para as 40 parcelas destinadas à validação, por meio da krigagem pontual, foram calculados os erros relativos e seus desvios padrão, que são apresentados na Tabela 6.

Tabela 6 - Erros relativos da validação preditiva, gerados pelos modelos de semivariogramas ajustados para cada intensidade amostral.

Intensidade	Erro Relativo (%)			
	Mínimo	Médio	Máximo	Desvio padrão
1:5	0,12	7,33	18,69	5,83
1:10	0,07	7,97	24,57	6,38
1:15	0,16	7,71	21,30	5,44
1:20	0,83	8,90	25,34	6,46

Fonte: Do autor.

Verifica-se que os modelos esféricos ajustados para as quatro diferentes intensidades amostrais apresentaram desempenhos satisfatórios e próximos entre si, com erros relativos médios inferiores a 10 %. O menor valor foi obtido na intensidade amostral de 1:5 (7,33%), e o maior valor na intensidade de 1:20 (8,90%). Uma diferença de erro relativo médio de 1,57 %. Essa diferença pode ser considerada pequena, ao se comparar a diferença no esforço amostral em se alocar uma parcela a cada 5 ha e uma parcela a cada 20 ha.

A Figura 4, por meio de gráficos de dispersão dos resíduos e de volumes de madeira estimados versus observados, reforça os resultados observados na Tabela 6, indicando uma similaridade nos desempenhos dos modelos de semivariogramas com a utilização de diferentes intensidades amostrais. Os resíduos estão distribuídos em torno de zero, com uma dispersão relativamente pequena.

Figura 4 - Gráficos de dispersão de resíduos e de volumes de madeira por hectare estimados versus observados, para as diferentes intensidades amostrais. (Continua)

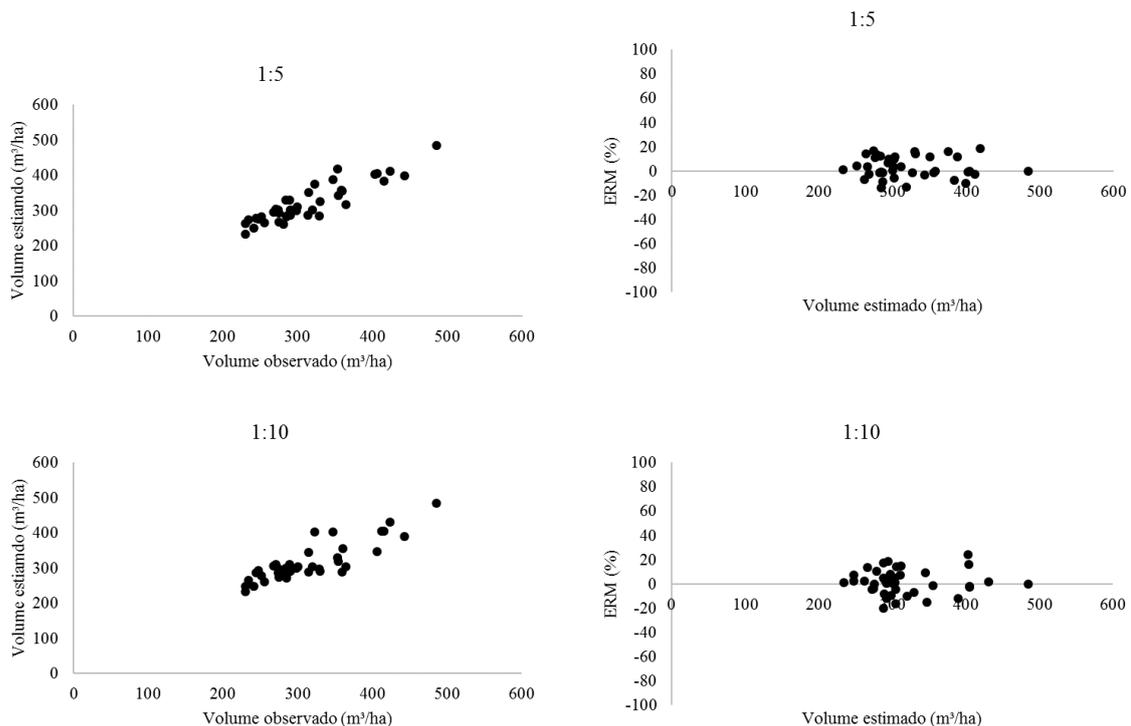
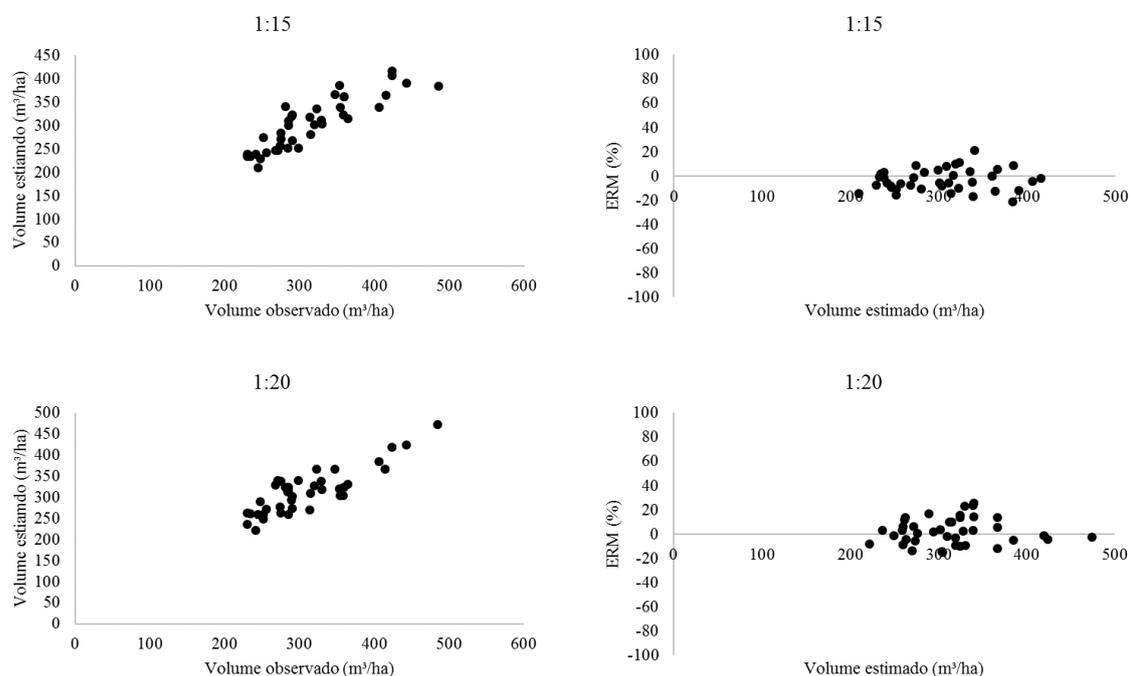


Figura 4 - Gráficos de dispersão de resíduos e de volumes de madeira por hectare estimados versus observados, para as diferentes intensidades amostrais. (Conclusão)



Fonte: Do autor.

O estimador geoestatístico, ao considerar a correlação entre as unidades amostrais, proporcionou estimativas precisas da variável volume de madeira por hectare. O desempenho dos modelos foi semelhante nas quatro intensidades amostrais avaliadas, não havendo diferença expressiva na qualidade das estimativas com a diminuição da intensidade amostral, diferentemente do que foi verificado com a utilização do estimador clássico da amostragem casual simples. Estes resultados estão de acordo com Mello et al. (2006), que modelaram a estrutura de correlação espacial existente entre as observações da variável volume de madeira por talhão, em plantios de *Eucalyptus* spp. no estado de São Paulo, e observaram que o erro no método geoestatístico variou de 8 a 12% enquanto, no estimador clássico de amostragem, variou de 9 a 32%.

Conforme Cochran (1977) e Thompson (1992), desconsiderando a componente espacial, ocorre superestimação da precisão, acarretando em erros de planejamento, com possíveis aumentos nos custos do inventário florestal.

Usando os benefícios da correlação espacial entre as parcelas, pode-se aumentar a precisão das estimativas dos inventários florestais e a intensidade de uma parcela a cada 20 ha pode atender às necessidades do inventário pré-corte e, possivelmente, também do contínuo. Com essa intensidade amostral seria possível satisfazer duas demandas importantes do planejamento florestal da empresa, que são a precisão das informações com relação ao

crescimento da floresta (monitoramento) e o fornecimento de informações imediatamente antes do corte.

A diminuição da intensidade amostral reduz o custo associado a um inventário florestal, com menos gastos na alocação de parcelas, medições e processamento dos dados (GUEDES et al., 2012). No entanto, os bons resultados gerados no manejo florestal não estão apenas associados aos processos de amostragem dos plantios florestais, mas também aos métodos de modelagem.

Diversos estudos utilizaram a geoestatística aplicando com sucesso a técnica de krigagem para buscar informações adicionais que melhor caracterizem os povoamentos florestais e comprovem sua eficiência (SINGH; DAS, 2014). Palmer et al. (2009) aplicaram a krigagem ordinária para prever o volume médio de madeira, o incremento anual e a altura máxima média para *Pinus radiata*. Viana et al. (2012) estimaram a biomassa da copa de povoamentos de *Pinus pinaster*. Hernández e Emery (2009) abordaram o uso da geoestatística para garantir a representatividade da amostragem no inventário florestal, em que o atributo medido foi a taxa de sobrevivência das árvores.

4 CONCLUSÃO

A geoestatística é uma técnica eficiente para estimativa da volumetria, nível povoamento, para plantios clonais de *Eucalyptus* spp.

O modelo esférico é preferível aos modelos exponencial e gaussiano para predição do volume para plantios de *Eucalyptus* spp aos 8 anos de idade com erros médios inferiores a 10 %.

É possível a redução da intensidade amostral em inventários pré-corte para povoamentos de *Eucalyptus* spp., com garantia de precisão nas predições por meio da krigagem pontual.

REFERÊNCIAS

- ALVARES, C.A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, p. 711-728, 2013.
- ANDRADE, V. C. L. **Um método para descrever o perfil do tronco em árvores de eucalipto utilizando geometria analítica**. 2001. 70 p. Dissertação (Mestrado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa, 2001.
- BIASE, A. G.; SANTANA, T. V. F. de. **Geoestatística análise de dados em outro software Gamma Desing GS+**. Escola superior de Agricultura Luiz de Queiroz – Pós-Graduação em Estatística e Experimentação Agrônômica, Piracicaba, São Paulo, 2012. 120 p.
- BURNHAN, K.P.; ANDERSON, D.R. **Model selection and multimodel inference**. 2.ed. New York: Springer, 2002. 350p.
- CRESSIE, N. **Statistics for spatial data**. New York: Wiley, 1993. 900p.
- COCHRAN, W. G. **Sampling Techniques**. 3 ed. New York, 1977.
- COUTO, E.G., STEIN, A.; KLAMT, E. Large area spatial variability of soil chemical properties in central Brazil. **Agriculture Ecosystems and Environment**, v. 66, p. 139- 152. 1997.
- DIGGLE, P. J.; RIBEIRO JÚNIOR, P. J. Model-based geostatistics. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14., Caxambú, 2000. Resumos, Caxambú: Associação Brasileira de Estatística, 2000, p.192.
- DIGGLE, P. J.; RIBEIRO JÚNIOR, P. J. **Model-based geostatistics**. Londres: Springer, 2007. 230 p.
- ESRI. **ArcGIS Desktop**: Release 10. Redlands, CA: Environmental Systems Research Institute. 2011.
- GUEDES I. C. L. et al. Técnicas geoestatísticas e interpoladores espaciais na estratificação de povoamentos de *Eucalyptus* sp. [Geostatistical techniques and spatial interpolators in the stratification of *Eucalyptus* sp. stands]. **Ciência Florestal**. v. 22. n. 3, p. 541–550. 2012.
- Hernández J. Emery X. A geostatistical approach to optimize sampling designs for local forest inventories. **Canadian Journal of Forest Research**, v. 39, p. 1465–1474. 2009.
- ISAAKS, E. H.; SRIVASTAVA, R. M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989. 560p.
- JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London: Academic Press, 1978. 600p.
- JOURNEL, A. G.; HUIJBREGTS, C.J. **Mining geostatistics**. London: Academic, 600 p. 2003.

LANDIM, P. M. B. **Análise estatística de dados geológicos**. 2. Ed. Ver e ampl. São Paulo: Editora UNESP, 2003.

MACHADO, S. A.; CONCEIÇÃO, M. B.; FIGUEIREDO, D. J. Modelagem do volume individual para diferentes idades e regimes de desbaste em plantações de *Pinus oocarpa*. **Revista Ciências Exatas e Naturais**, Curitiba, vol. 4, n. 2, 2002.

MATHERON, G. Principles of geostatistics. **Economic Geology**. El paso, v. 58, p. 1124-66, 1963.

Mc BRATNEY, A.G.; WEBSTER, A.G. Choosing functions for semi-variograms and fitting them to sampling estimates. **Journal of Soil Science**, v. 37, p. 617-639, 1986.

MELLO, J. M. de. **Geoestatística aplicada ao inventário florestal**. 2004. 122 p. Tese (Doutorado em Recursos Florestais) – Universidade de São Paulo, Piracicaba. 2004.

MELLO, J. M. de.; OLIVEIRA, M. S. de; BATISTA, J. L. F.; JÚNIOR, P. J. R.; JÚNIOR, H. K. Uso do estimador geoestatístico para predição volumétrica por talhão. **Floresta**, Curitiba, PR. v. 36, n. 2, p. 251-260, 2006.

MELLO, J. M. et al. Métodos de amostragem geoestatística para estimativa do número de fustes e volume em plantios de *Eucalyptus grandis*. **Floresta**, Curitiba, v. 39, n. 1, p. 157-166, jan./mar. 2009.

NIELSEN, D. R.; WENDROTH, O. **Spatial and temporal statistics: sampling field soils and their vegetation**. Reiskirchen: Catena Verlag, 2003. 398p.

NOGUEIRA, C. H. **Análise de variância com dependência espacial sob uma abordagem geoestatística**. 2013. 124 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2013.

PALMER, D. J. et al. Comparison of spatial prediction techniques for developing *Pinus radiata* productivity surfaces across New Zealand. **Forest Ecology and Management**, v. 258, p. 2046–2055. 2009.

RIBEIRO JÚNIOR, P.J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. Piracicaba, 1995. 99 p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo. 1995.

RIBEIRO JÚNIOR, P. J.; DIGGLE, P. J. geoR: A package for geostatistical analysis. **R-NEWS**, v. 1, n. 2, p. 15-18, 2001.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel, p. 81, 1986.

SCHUMACHER, F. X.; HALL, F. S. Logarithmic expression of timber-tree volume. **Journal of Agricultural Research**, Lahore, v. 47, n. 9, p. 719 - 734, 1933.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3/4, p. 591-611, 1965.

SINGH T. P.; DAS S. Predictive analysis for vegetation biomass assessment in Western Ghat region (WG) using geospatial techniques. **Journal of the Indian Society of Remote Sensing**, v. 42, n. 3, p. 549–557. 2014.

THOMPSON, S.K. **Sampling**. New York: Wiley, 1992. 343p.

VIANA, H. et al. Estimation of crown biomass of *Pinus pinaster* stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models. **Ecology Modeling**, v. 226, p. 22–35. 2012.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F.; ALVAREZ V., V. H.; SCHAEFER, C. E. G. R., (Eds.) Tópicos em ciências do solo. Viçosa, MG, **Sociedade Brasileira de Ciência do Solo**, 2000. v. 1, p. 2-54.

WONNACOTT, T. H; WONNACOTT, R. J. **Introductory Statistics**. 2 ed. New York: John Wiley, 1972. 510 p.