

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS E TECNOLOGIAS PARA A SUSTENTABILIDADE
PROGRAMA DE PÓS-GRADUAÇÃO EM PLANEJAMENTO E USO DE RECURSOS
RENOVÁVEIS

LUCAS KRÖHLING BERNARDI

**INFERÊNCIA MULTIMODELOS NA PREDIÇÃO DE MULTIPRODUTOS EM
POVOAMENTOS DE *Eucalyptus* sp.**

Sorocaba

2020

LUCAS KRÖHLING BERNARDI

**INFERÊNCIA MULTIMODELOS NA PREDIÇÃO DE MULTIPRODUTOS EM
POVOAMENTOS DE *Eucalyptus* sp.**

Dissertação apresentada ao Programa de Pós-Graduação em Planejamento e Uso de Recursos Renováveis, para obtenção do título de Mestre em Planejamento e Uso de Recursos Renováveis.

Orientadora: Profa. Dra. Monica Fabiana Bento
Moreira Thiersch

Co-Orientador: Prof. Dr. Claudio Roberto
Thiersch

Financiamento: CAPES

Sorocaba

2020

Kröhling Bernardi, Lucas

Inferência Multimodelos na predição de multiprodutos em povoamentos de Eucalyptus sp. / Lucas Kröhling Bernardi. -- 2020.
65 f. : 30 cm.

Dissertação (mestrado)-Universidade Federal de São Carlos, campus Sorocaba, Sorocaba

Orientador: Monica Fabiana Bento Moreira Thiersch
Banca examinadora: Claudio Roberto Thiersch, Thelma Shirlen Soares, Pedro José Ferreira Filho
Bibliografia

1. Afilamento do fuste. 2. Regressão por mixagem de modelos . 3. Algoritmo. I. Orientador. II. Universidade Federal de São Carlos. III. Título.

Ficha catalográfica elaborada pelo Programa de Geração Automática da Secretaria Geral de Informática (SIn).

DADOS FORNECIDOS PELO(A) AUTOR(A)

Bibliotecário(a) Responsável: Maria Aparecida de Lourdes Mariano – CRB/8 6979

LUCAS KRÖHLING BERNARDI

**INFERÊNCIA MULTIMODELOS NA PREDIÇÃO DE
MULTIPRODUTOS EM POVOAMENTOS DE *Eucalyptus sp.***

Dissertação apresentada ao Programa de Pós-Graduação para obtenção do título de
mestre em Planejamento e Uso de Recursos Renováveis.
Universidade Federal de São Carlos.
Sorocaba, 28 de fevereiro de 2020.

Coorientador:



Prof. Dr. Claudio Roberto Thiersch
Universidade Federal de São Carlos – UFSCar *Campus Sorocaba*

Examinadores:



Prof. Dr. Pedro José Ferreira Filho
Universidade Federal de São Carlos – UFSCar *Campus Sorocaba*

Certifico que a sessão de defesa foi realizada com a participação à distância da Dra. Thelma Shirlen Soares (Universidade Federal de Goiás – Campus Jataí) e, depois das arguições e deliberações realizadas, a participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa de Dissertação de Lucas Kröhling Bernardi.



Prof. Dr. Claudio Roberto Thiersch
Presidente da Comissão Examinadora

*Dedico aos meus pais, Cris e Beto,
aos meus irmãos, Ana Júlia e Thiago
e à minha namorada, Bibi.*

AGRADECIMENTOS

Agradeço à Universidade Federal de São Carlos, campus Sorocaba, pela oportunidade.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de mestrado concedida.

Agradeço à minha orientadora, Profa. Dra. Monica Fabiana Bento Moreira Thiersch, e ao meu coorientador, Prof. Dr. Cláudio Roberto Thiersch, por terem aceitado me orientar e compartilhar seus conhecimentos, pela paciência, dedicação e compreensão ao me auxiliar durante esses dois anos de pesquisa e desenvolvimento deste trabalho, me servindo de exemplo e inspiração.

Aos meus pais, Cristiane e Vilberto, por serem os pilares nos quais meu caráter está sustentado, por me darem a oportunidade da vida e me criarem com o máximo amor e carinho, a vocês o meu eterno agradecimento. Aos meus irmãos Ana Júlia e Thiago, que são motivo de muito orgulho para mim e para nossos pais, e fazerem minha vida muito mais feliz.

Aos meus amigos Pedrão, Renan, Jose, Fabinho, Marcelinho, Gi e Mari pela amizade verdadeira, por me ajudarem nos momentos difíceis e por estarem comigo em todos os momentos alegres. Guardaremos para sempre essas lembranças e, juntos, construiremos muitas outras ainda.

Agradeço imensamente à minha namorada Bibi, que durante toda a minha pesquisa e estudo esteve sempre do meu lado, me dando apoio em cada momento difícil, me tranquilizando em momentos de apreensão, dizendo sempre o que eu precisava ouvir nos momentos certos e que compartilhou de todas as minhas alegrias a cada nova conquista a cada nova descoberta e a cada novo resultado positivo. Você foi a força que me manteve de pé, obrigado pela amizade, pela união, pelo companheirismo e pelo seu amor.

Por fim, agradeço ao Dr. Celso Charuri, por ser meu amigo e meu Mestre, meu maior exemplo e por me mostrar o caminho para um Mundo Bem Melhor.

RESUMO

BERNARDI, L. K. **Inferência Multimodelos na predição de multiprodutos em povoamentos de *Eucalyptus* sp.** 2020. 65 f. Dissertação (Mestrado) – Universidade Federal de São Carlos, Sorocaba, 2020.

Ao selecionar um modelo para realizar a estimativa de uma variável, condiciona-se essa estimativa a toda e qualquer característica deste modelo, podendo estar sujeito a subestimativas ou superestimativas, dependendo do modelo em questão e de suas possíveis tendenciosidades. Estudos recentes vêm sugerindo a realização da chamada Inferência Multimodelos, que consiste na obtenção do resultado através da contribuição de dois ou mais modelos matemáticos. Este estudo teve como objetivo propor uma nova metodologia de Inferência Multimodelos, baseada na modificação do algoritmo ARM (Adaptive Regression by Mixing), aplicada ao inventário florestal para estimativa do volume de madeira por meio de equações volumétricas e de afilamento do fuste. Para a estimativa do volume das árvores utilizou-se o algoritmo ARM modificado e para a estimativa do afilamento do fuste, utilizou-se o algoritmo proposto, *Mixed Taper Equations*, ou Algoritmo MTE. Para as duas bases de dados clonais de *Eucalyptus* sp. utilizadas no trabalho, o algoritmo ARM modificado foi superior à seleção de modelos e estimou o volume das árvores de forma mais acurada, apresentando RMSE de 3,53% e 3,86% enquanto o modelo selecionado obteve 3,99% e 3,97%. Para a estimativa de afilamento do fuste, com uma base de dados composta por três variedades clonais de *Eucalyptus* sp. o algoritmo MTE foi superior à seleção de modelos com RMSE de 4,11%, 8,96% e 4,91%, enquanto que os modelos selecionados para cada caso apresentaram RMSE de 4,49%, 8,88% e 5,02%, respectivamente, para os três materiais genéticos testados. Além disso, ambos, ARM modificado e MTE apresentaram melhor dispersão dos resíduos em comparação com os modelos selecionados e, no caso do algoritmo MTE, melhores resultados por árvore. A implementação do algoritmo MTE no *software* R demonstrou que a técnica é viável e de fato conseguiu calcular pesos distintos para os modelos testados ao longo das alturas de medição. Além disso, o método proposto para a estimativa dos volumes se utilizando dos pesos e das próprias equações de afilamento foi capaz de realizar a reconstituição do volume das árvores com resultados árvore a árvore mais acurados do que os modelos individualmente.

Palavras-chave: biometria, mixagem de modelos de regressão, cubagem, diâmetro.

ABSTRACT

BERNARDI, L. K. **Multimodel Inference to predict multiple wood products in *Eucalyptus* sp. stands.** 2020. 65 p. Thesis (Masters) – Federal University of São Carlos, Sorocaba, 2020.

When selecting a single model to estimate a variable, this estimate is conditioned to any and all characteristics of this model, which may be subject to underestimation or overestimation, depending on the model in question and its possible biases. Recent studies have suggested the so-called Multimodel Inference, which consists in obtaining the result through the contribution of two or more mathematical models. The aim of this study was to propose a new methodology for Multimodel Inference, based on the modification of the ARM (Adaptive Regression by Mixing) algorithm, applied to forest inventory to estimating wood volume using volumetric and taper equations. To estimate wood volume with volumetric equations, the modified ARM algorithm was used and to estimate tree taper was used the proposed algorithm, Mixed Taper Equations, or MTE Algorithm. For the two clonal *Eucalyptus* sp. databases used in this study, the modified ARM algorithm was superior to model selection estimating tree volume more accurately, presenting RMSE of 3.53% and 3.86% while the selected model obtained 3.99% and 3.97 %, respectively. To estimate trees taper, with a database composed of three clonal varieties of *Eucalyptus* sp. the MTE algorithm was superior to model selection with RMSE of 4.11%, 8.96% and 4.91%, while selected models for each case presented RMSE of 4.49%, 8.88% and 5, 02%, respectively, for the three genetic materials tested. In addition, both modified ARM and MTE showed better dispersion of residues compared to the selected models and, in the case of the MTE algorithm, better results per tree. The use of MTE algorithm within the software R demonstrated that the technique is viable and managed to calculate different weights for the models tested over the measurement heights. In addition, the proposed method for estimating volumes using weights and taper equations was able to perform the tree volume reconstitution with more accurate results per tree than the tested models, individually.

Keywords: biometrics, mixing regression models, upper-stem diameters.

SUMÁRIO

1.	INTRODUÇÃO	15
2.	REFERENCIAL TEÓRICO	17
2.1.	CRITÉRIOS DE SELEÇÃO DE MODELOS	17
2.1.1.	Teste de Razão de Verossimilhança	17
2.1.2.	Crítério de Informação de Akaike	18
2.1.3.	Crítério de Informação de Akaike de Segunda Ordem	19
2.1.4.	Crítério de Informação Bayesiano	20
2.1.5.	Outros critérios de seleção de modelos	21
2.1.	INFERÊNCIA MULTIMODELOS	22
2.2.	EQUAÇÕES DE VOLUME	23
2.3.	AFILAMENTO DO TRONCO	24
2.3.1.	Funções de afilamento não-segmentadas	25
2.3.2.	Funções de afilamento segmentadas	27
2.3.3.	Funções de afilamento de forma variável	28
3.	MATERIAL E MÉTODOS	31
3.1.	OBTENÇÃO DAS BASES DE DADOS	31
3.1.1.	Equações de volume	31
3.1.2.	Equações de afilamento	31
3.1.2.1.	Obtenção dos dados de volume real	31
3.2.	PROCESSAMENTO DOS DADOS	32
3.2.1.	Definições	32
3.2.2.	Inferência Multimodelos – Algoritmo	32
3.2.2.1.	Algoritmo ARM para predição volumétrica	32
3.2.2.2.	Algoritmo MTE (<i>Mixed Taper Equations</i>)	34
3.2.3.	Crítérios estatísticos utilizados	36
3.2.4.	Modelos volumétricos testados	37
3.2.5.	Modelos de afilamento testados	37
4.	RESULTADOS E DISCUSSÃO	39
4.1.	EQUAÇÕES VOLUMÉTRICAS	39
4.2.	EQUAÇÕES DE AFILAMENTO	41
5.	CONCLUSÕES	57

1. INTRODUÇÃO

No Brasil, as florestas plantadas ocupam 7,83 milhões de hectares, retêm 91% da madeira produzida para fins industriais, totalizando 1,3% de toda a riqueza gerada no País e 6,9% do PIB industrial. Dentre as áreas de florestas plantadas no Brasil, os plantios de eucalipto representam 5,7 milhões de hectares, situados principalmente em Minas Gerais, São Paulo e Mato Grosso do Sul, onde 36% desses plantios são destinados ao segmento de celulose e papel, 12% à siderurgia e carvão vegetal, 6% a painéis de madeira e pisos laminados, 4% a produtos sólidos de madeira, 10% a investidores financeiros e 29% a produtores independentes (IBÁ, 2019).

Empreendimentos florestais focados na maximização do lucro e agregação de renda, ao destinarem recursos a um único produto, ficam suscetíveis à variação de preço de mercado do mesmo. Dessa forma, o conceito de múltiplos usos da madeira é alternativo à essa oscilação de preço e instabilidade financeira e, para que isso seja possível, o uso de técnicas que produzam resultados confiáveis é de fundamental importância.

Entre as variáveis quantitativas de uma floresta, destaca-se o volume. Esta variável pode ser obtida com base em diversos métodos, onde a maioria concentra-se em estimativas a partir do uso de expressões matemáticas.

As equações volumétricas são amplamente utilizadas no âmbito do inventário florestal quando o objetivo é o cálculo do volume comercial de um povoamento.

Outra forma de estimar o volume de um povoamento é com equações de afilamento do tronco, comumente utilizadas para as estimativas dos múltiplos produtos da madeira por serem capazes de estimar o volume entre quaisquer alturas do tronco.

A escolha do modelo no meio florestal é baseada na técnica de seleção de modelos de regressão onde o modelo que apresenta os melhores resultados para o critério de seleção escolhido e em inspeções gráficas, ou seja, que apresentou resultados mais próximos do volume real da amostra realizada, é escolhido para estimar o volume de toda a população.

Entretanto, um modelo que seja selecionado para determinada base de dados específica nem sempre se comportará bem em qualquer outra base ou até mesmo em uma parcela dos dados já testados (HU, 2015; YANG, 2001), o modelo selecionado é sempre o melhor dentre o conjunto de modelos considerado, porém isso apenas se aplica para o resultado médio, visto que em alguma parcela dos dados outros modelos podem apresentar resultados melhores do que o selecionado.

Dada essa variação do melhor modelo dentro de parcelas específicas da população, Yang (2001) propôs o algoritmo (*ARM – Adaptive Regression by Mixing*) para o cálculo de

pesos para cada modelo ajustado e realiza a estimativa dos parâmetros populacionais utilizando-se de todos os modelos pré-selecionados.

Portanto, este estudo teve como objetivo a proposta de uma nova metodologia para a estimativa do volume comercial por meio do algoritmo ARM para a realização de uma inferência multimodelos, além da proposta de um novo algoritmo, baseando-se no ARM para a modelagem do perfil do fuste das árvores com equações de afilamento.

2. REFERENCIAL TEÓRICO

2.1. CRITÉRIOS DE SELEÇÃO DE MODELOS

Em análises estatísticas, múltiplos modelos são comumente considerados e testados, sendo apenas um deles selecionado baseando-se em testes de hipóteses, ou no uso de critérios estatísticos juntamente com inspeções gráficas. Sendo assim, a estimativa, interpretação e predição final são então baseadas no modelo selecionado (YANG, 2003).

A seleção de modelos busca encontrar aquele que envolva o mínimo possível de parâmetros a serem estimados e ainda que mais se aproxime do modelo verdadeiro, respeitando o princípio científico fundamentado na parcimônia (EMILIANO, 2013).

A escolha do modelo candidato mais próximo do modelo verdadeiro requer medida definida de distância ou informação (DAL BELLO, 2010) e, por isso, diversos critérios de seleção de modelos são apresentados na literatura (BOZDOGAN, 1987; BURNHAM e ANDERSON, 2002; LITTEL *et al.*, 2002; WOLFINGER, 1993), como os baseados no máximo da função de verossimilhança (MFV) os frequentemente mais utilizados, principalmente o Critério de Informação de Akaike (AIC) e o Critério Bayesiano de Schwarz (BIC).

Além disso, principalmente na área florestal, é muito comum utilizar valores que medem a qualidade do ajuste para selecionar o melhor modelo, como por exemplo o uso dos critérios estatísticos como o coeficiente de determinação (R^2), erro padrão residual na unidade da variável de interesse (S_{yx}) ou em porcentagem, o *Mean absolute bias* (MAB) para medir o viés segundo Kozak (2004) e a raiz do erro quadrático médio (RMSE).

2.1.1. Teste de Razão de Verossimilhança

O Teste de Razão de Verossimilhança (TRV) serve para comparar dois modelos aninhados, ou seja, casos onde o primeiro modelo (M1) seja um caso particular do segundo modelo (M2), ou se uma variável independente é significativa para o cálculo da variável dependente. Por exemplo, temos que M1 é um polinômio de terceiro grau e M2 um de quarto grau, neste caso a hipótese nula é de que M1 é preferível a M2. Como o nome diz, o TRV é a divisão entre a verossimilhança de M1 e a verossimilhança de M2. Rejeita-se a hipótese nula caso:

$$\lambda = -2 (\log f(M_1|\theta) - \log f(M_2|\theta)) > \lambda_c$$

Onde: $f(M_i|\theta)$ = máxima verossimilhança do modelo i .

Em que p-valor $P(Q > \lambda_c) = \alpha, Q \sim \chi^2_\gamma$.

Onde γ = número de parâmetros de M2 – número de parâmetros de M1.

2.1.2. Critério de Informação de Akaike

Akaike (1973) desenvolveu um critério de seleção de modelo, conhecido como Critério de Informação de Akaike (AIC) que se origina da minimização da informação (ou distância) de Kullback-Leibler (K-L) como base para a seleção de modelos, utilizando-se a função de log-verossimilhança em seu ponto máximo, acrescida da penalidade associada ao número de parâmetros do modelo, sendo o modelo selecionado aquele que apresentar o menor valor de *AIC*, definido por:

$$AIC = -2 \log f(x|\theta) + 2p$$

Onde: $f(x|\theta)$ = máxima verossimilhança da função; p = é o número de parâmetros do modelo.

Dessa forma, o AIC tem por princípio selecionar o modelo que apresentar as estimativas mais próximas do modelo verdadeiro, porém respeitando o princípio da parcimônia, penalizando os modelos mais complexos pela sua dificuldade de ajuste.

O Critério de Informação de Akaike é um dos critérios mais aceitos e utilizados para realizar a seleção de modelos, sendo utilizado por diversos autores em diversas áreas do conhecimento.

Yamaoka, Nakagawa e Uno (1978), estudando a farmacocinética, ou seja, o tempo de concentração de uma droga no plasma, com equações lineares, empregam o AIC como o principal critério de seleção do melhor modelo.

Silva *et al.* (2011) aplicando modelos não lineares para estimar o crescimento de bovinos da raça Nelore, utilizaram, dentre critérios como: coeficiente de determinação (R^2), quadrado médio do erro (QME), erro quadrático médio de predição (MEP), coeficiente de determinação de predição (R^2_p) e Critério de Informação Bayesiano (BIC); o AIC como um dos critérios para ranquear os modelos e selecionar o melhor.

Na área florestal, em trabalhos relacionados a afilamento florestal, o AIC é aplicado como critério de seleção de modelos.

Môra (2015) utiliza o AIC como um dos critérios de seleção do melhor modelo no afilamento do fuste de *Pinus taeda* e *Eucalyptus saligna*, testando desde funções de afilamento mais simples até as mais recentes e complexas.

Carvalho *et al.* (2014) também utilizaram o AIC como um dos critérios de seleção de modelo com a técnica de modelagem não linear mista para descrever o afilamento do fuste de árvores clonais de *Eucalyptus* sp. em diferentes classes diamétricas e idades.

Özçelik e Crecente-Campo (2015), estudando equações de afilamento para estimar, diâmetro, altura e volume comercial de árvores de cedro-do-Líbano (*Cedrus libani*), na Turquia,

utilizaram, além do R^2 e do RMSE, o AIC como um dos critérios de seleção das equações testadas. Os autores ranquearam os modelos de acordo com a metodologia apresentada por Poudel e Cao (2013), que leva em consideração o valor de cada critério em relação aos valores máximos e mínimos gerados para cada critério dentre todos os modelos, gerando um valor de ranking para cada critério de seleção, sendo o ranking final do modelo resultado da média entre os rankings para cada critério.

Entretanto, apesar do extenso uso, em aplicações práticas o AIC frequentemente indica a seleção dos modelos mais complexos, o que pode ser explicado por uma possível insuficiência na penalização da complexidade desses modelos, representadas pelo número de parâmetros destes (DAVISON, 2001), isso ocorre geralmente em casos onde a base de dados não é suficientemente grande, aproximadamente 40 vezes maior do que o número de parâmetros dos modelos (BURNHAM e ANDERSON, 2002).

2.1.3. Critério de Informação de Akaike de Segunda Ordem

De forma a aumentar a penalização relativa à complexidade do modelo, levando em conta o tamanho da amostra, Hurvich e Tsai (1989) desenvolveram o critério de informação de Akaike de segunda ordem, chamado de AICc (AIC corrigido), que é mais indicado para pequenos conjuntos de dados. No AICc a penalização do modelo pela sua complexidade é multiplicada pelo fator de correção $n/(n - p - 1)$, sendo escrito por:

$$AIC_c = AIC + \frac{2p(p + 1)}{n - p - 1}$$

Onde: n = número de observações; p = número de parâmetros do modelo

Da mesma forma que no AIC, o método seleciona o modelo que apresentar o menor valor para AICc.

Burnham e Anderson (2002) recomendam o uso do AICc na seleção de modelos aplicados em base de dados em que o número de observações seja pequeno, sendo a relação n/p menor do que 40, nos demais casos pode-se utilizar o AIC.

Devido à melhor performance em relação ao AIC para bases de dados menores, o AICc frequentemente é utilizado em detrimento do seu predecessor, como o principal método de seleção de modelos, sendo utilizado no setor florestal quando sua aplicação se justifica em relação à do AIC.

Mate, Johansson e Siteo (2015) utilizaram-se do AICc para selecionar o melhor modelo para o cálculo de volume comercial de madeira de espécies arbóreas de importância em Moçambique.

Nunes e Görgens (2016), realizando o afilamento do fuste de árvores do Cerrado, Floresta Ombrófila Densa e Floresta Estacional Semi-Decidual com a técnica de Redes Neurais Artificiais, utilizaram-se do AICc como o principal critério de seleção, sendo aquele que apresentou por mais vezes o menor valor para o AICc, escolhido para a realização das estimativas dos diâmetros.

2.1.4. Critério de Informação Bayesiano

O Critério de Informação Bayesiano (BIC), também chamado de Critério de Schwarz, proposto por Schwarz (1978), é um critério de seleção de modelos para ser aplicado considerando-se a probabilidade *a posteriori* dos parâmetros do modelo ajustado, sendo assim chamado porque Schwarz deu um argumento Bayesiano para prová-lo (EMILIANO, 2009). Este critério pondera entre o máximo da função log-verossimilhança e o número de parâmetros do modelo, sendo o melhor modelo aquele que apresentar o menor valor de BIC dado por:

$$BIC = -2 \log f(x|\theta) + p \log n$$

Quando o modelo verdadeiro está contido dentro do conjunto dos modelos candidatos, com probabilidade tendendo a 1, o BIC seleciona o modelo verdadeiro; por outro lado, se nenhum dos modelos testados é o modelo verdadeiro, o AIC é superior ao BIC em termos de riscos estatísticos (YANG, 2003).

Juntamente com o AIC, o Critério de Informação Bayesiano é um dos critérios de seleção de modelos mais consolidado na literatura, muito utilizado no setor florestal.

Schröder, Pereira e Hofiço (2013), estudando a comparação de métodos de estimativa de volume total para *Eucalyptus grandis*, sendo um dos métodos empregados o modelo de afilamento de Kozak, Munro e Smith (1969) e, devido à estrutura hierárquica dos dados de afilamento, os autores aleatorizaram os coeficientes do modelo em função da classe diamétrica e mediram a estrutura de covariância com base em cinco métodos, sendo a melhor estrutura de covariância escolhida com base nos valores de BIC e também nos valores de AIC.

Rocha (2018) ao estudar o afilamento e otimização de sortimentos de *Tectona grandis*, e testar diversas funções de afilamento e propor alterações em uma das funções testadas, utilizou-se da técnica de modelagem de efeitos mistos e, para isso selecionou a melhor combinação desses efeitos com os efeitos aleatórios da função original e das posições de alteração propostas utilizando-se do BIC. Por fim, selecionou o melhor modelo pelo uso do BIC por considerar que este penaliza mais os modelos menos parcimoniosos quando comparado ao AIC.

Horle *et al.* (2010), também utilizaram o BIC como um dos critérios de avaliação do modelo linear do polinômio de 5º grau de Schöepfer modificado e de um modelo não linear, o modelo logístico modificado para consequente escolha do melhor modelo para realizar o afilamento florestal de *Pinus oocarpa* na região do Triângulo Mineiro, Brasil. Além do BIC, utilizaram também o AIC como um dos critérios de seleção dos modelos.

2.1.5. Outros critérios de seleção de modelos

Outros critérios de informação presentes na literatura, porém menos utilizados do que AIC e BIC podem ser citados, como o Critério *Bridge* (BC), proposto por Ding, Tarokh e Yang (2018) é um novo critério de seleção que realiza a ligação entre AIC e BIC, de maneira assintótica. Dado por:

$$BC = -2 \log f(x|\theta) + c(1 + 2^{-1} + \dots + p^{-1})$$

Onde: $f(x|\theta)$ = máxima verossimilhança da função; $c = n^{2/3}$ (sugerido pelos autores); p = número de parâmetros do modelo.

Esse método seleciona o modelo que obtiver o menor valor de BC dentre todos os modelos candidatos cujo número de parâmetros não seja maior do que o número de parâmetros do modelo selecionado pelo método AIC. O BC provou ter performance similar ao AIC em situações não-paramétricas e similar ao BIC em situações paramétricas.

A Validação-Cruzada (CV) é uma classe de método de seleção de modelo que é amplamente utilizada em técnicas de *machine learning* (ou aprendizado de máquina). Esse método não tem como premissa o fato de que os modelos candidatos sejam paramétricos e funciona desde que os dados sejam permutáveis e se possa avaliar o desempenho preditivo baseado em alguma medida (DING, TAROK e YANG, 2018).

A CV primeiramente aleatoriza a base de dados e a divide em duas partes, sendo uma parte chamada de base de treinamento n_t , onde $1 \geq n_t \geq n - 1$; e a outra de base de validação $n_v = n - n_t$. Sendo assim, cada modelo é treinado na base inicial e validado na segunda base. Esse procedimento é repetido algumas vezes de forma independente, cada repetição com um diferente conjunto de dados de validação, de forma a reduzir a variabilidade causada pela separação da base. Por fim, o modelo com a menor variação de perda pela validação é selecionado e utilizado para predições futuras.

2.1. INFERÊNCIA MULTIMODELOS

Yang (2003) afirma que todos os critérios de seleção, mesmo os mais consolidados, como AIC e BIC, não possuem boas performances todo o tempo. O mesmo autor aponta que o AIC possui melhor performance à medida que os erros de aproximação dos modelos candidatos decrescem lentamente em relação ao tamanho da amostra e da sua variância e propõe, então, um questionamento: é possível combinar as qualidades do AIC e BIC?

Um cenário ideal seria aquele onde não houvesse incerteza na seleção de um modelo, que houvesse apenas um modelo disponível e coubesse ao analista dos dados apenas escolher o melhor método para ajustar esse modelo, de acordo com a base de dados disponível. Entretanto, uma base de dados é sempre muito variável, influenciada por inúmeros fatores e covariáveis, e por isso, a seleção de modelos é um processo importante na realização da inferência (LINK e BARKER, 2006).

Ao selecionar um conjunto de modelos para estimar uma variável, cada modelo fornecerá uma estimativa dessa variável. Se um dos modelos for muito superior aos demais, então a inferência poderá ser feita, condicionalmente, pelo modelo selecionado. Porém, comumente observam-se situações onde nenhum modelo se sobressai como sendo claramente superior aos demais modelos testados, dessa forma é arriscado basear-se apenas no modelo selecionado (BURHNAM e ANDERSON, 2002).

Breiman (1996) apontou que uma leve perturbação dos dados pode resultar na seleção de um modelo totalmente diferente e, como consequência disso, estimadores de funções de regressão baseados na seleção de modelos frequentemente apresentam-se com uma variância desnecessariamente grande.

Akaike realizou diversos estudos onde sugeriu o uso de pesos para o ranqueamento de modelos, através do critério conhecido como *Akaike Weights* (Pesos Akaike) e com isso criou uma linha de pensamento que norteou as futuras pesquisas para que uma inferência com base em mais de um modelo fosse possível (BURNHAM e ANDERSON, 2002).

Da mesma forma, do ponto de vista Bayesiano, o modelo médio (*model averaging*) é uma abordagem que reduz a variabilidade na seleção de modelos (DRAPER, 1995; GEORGE e MCCULLOCH, 1997).

Buckland, Burnham e Augustin (1997) propuseram um método para conferir pesos aos modelos de acordo com os valores dos critérios de seleção de modelos. As técnicas de validação cruzada e *bootstrapping* já foram usadas também para combinar de forma linear diferentes estimadores com a intenção de melhorar a acurácia encontrando a melhor combinação linear (WOLPERT, 1992).

Yang (2003) afirma que existem duas diferentes direções na mixagem de modelos: uma é a combinação para adaptação (YANG, 2001), que pretende capturar a melhor performance dentre os modelos candidatos; a outra é a combinação para melhoria (JUDITSKY e NEMIROVSKI, 2000; YANG, 2004) que pretende ter uma performance superior a qualquer modelo candidato.

Yang (2003), após afirmar que AIC e BIC não possuem boas performances todo o tempo, propõe então um algoritmo, ARM (*adaptive regression by mixing*) que realiza a combinação dos pontos fortes de AIC e BIC, de forma a conferir pesos aos modelos envolvidos no processo de seleção e, assim, gerar um modelo final composto pela predição de todos os modelos, multiplicado pelos respectivos pesos.

2.2. EQUAÇÕES DE VOLUME

O volume de uma árvore apresenta-se como uma das informações de maior importância para o conhecimento do potencial disponível em um povoamento, pois o volume individual fornece subsídios para a avaliação do estoque de madeira dos mesmos e ainda fornece dados para a análise do potencial produtivo (THOMAS *et al.*, 2006).

No entanto, o volume é caracterizado por não ser facilmente mensurável, assim é necessário aplicar técnicas para a realização de estimativas. Nesse sentido, ao longo da história diversos modelos estatísticos foram desenvolvidos e testados para a obtenção do volume do povoamento (MACHADO, CONCEIÇÃO e FIGUEIREDO-FILHO, 2002).

Apesar da eficiência de alguns modelos, nem sempre estes ajustam-se a toda e qualquer espécie, região, regime de manejo e às condições estudadas, portanto recomenda-se ajustar e testar vários modelos para, através de análises e critérios estatísticos, selecionar o que melhor se identifica em cada situação (MIGUEL *et al.*, 2014).

Os modelos volumétricos utilizam a variável dependente volume, advinda do processo de cubagem rigorosa, correlacionada com variáveis independentes de fácil mensuração na floresta, como diâmetro a 1,30 m de altura, conhecido como diâmetro à altura do peito (DAP) e altura total ou comercial (MACHADO, CONCEIÇÃO e FIGUEIREDO-FILHO, 2002).

Obtidas as equações, o volume individual de cada árvore é estimado e, quando estas apresentam bons ajustes, são capazes de captar as variações do povoamento e com base no procedimento de amostragem realizado em campo, utilizando unidades amostrais (parcelas), chega-se a estimativa de volume por árvore, por parcelas, por hectare e para o povoamento (MIGUEL *et al.*, 2014).

Estudos detalhados visando o ajuste, comparação e seleção de modelos volumétricos para diferentes espécies, espaçamentos e regiões vem sendo realizados desde a metade do século passado com o intuito de melhorar os estimadores da variável de interesse (volume), bem como facilitar a quantificação desta variável para produtores florestais no momento da comercialização de suas áreas (GOLDING e HALL, 1961; PAULA NETO, BRANDI e RIBEIRO, 1977; MACHADO, CONCEIÇÃO e FIGUEIREDO-FILHO, 2002; PELISSARI, LASSANOVA e DRESCHER, 2011; SCHRÖDER *et al.*, 2013; MIGUEL *et al.*, 2014).

2.3. AFILAMENTO DO TRONCO

Um empreendimento de base florestal, para não ficar sujeito às flutuações de preço de mercado de apenas um produto, pode optar por atender à demanda dos múltiplos produtos da madeira, como laminação, serraria, painéis, celulose, bobinas de papel, carvão e biomassa, por exemplo.

Para que a quantificação do estoque de madeira das indústrias de base florestal seja realizada de maneira acurada, são necessários frequentes estudos ligados ao volume da árvore, de forma individual ou do povoamento (MÔRA, 2015), tornando o uso de modelos matemáticos capazes de representar com acuracidade os perfis do fuste ferramenta de grande importância para os empreendimentos florestais (SOUZA, 2007).

A forma do fuste de uma árvore é definida pela taxa de decréscimo do diâmetro ao longo do tronco, sendo esta relação conhecida também como afilamento do fuste (*taper*), que é a razão fundamental de variação na forma e no volume destes fustes (HUSCH; BEERS; KERSHAW, 2003).

O afilamento do tronco é um dos vários componentes necessários em um sistema de planejamento do manejo florestal moderno (HEIDARSSON e PUKKALA, 2011) e, ao utilizar o DAP, a altura total e diferentes alturas ao longo do tronco da árvore, permite estimar o diâmetro do tronco a qualquer altura, volume total do tronco, volume comercial e altura comercial para qualquer diâmetro a qualquer altura acima do toco e os volumes individuais em qualquer parte, em qualquer altura a partir do solo (KOZAK, 2004).

Dentro do conceito de usos múltiplos da madeira, as equações de afilamento são a única alternativa para o planejamento e quantificação do estoque, devido ao fato de ser o único método para estimar o volume entre quaisquer porções da árvore, processo denominado *sortimento*, maximizando o aproveitamento do estoque florestal.

O estudo do *sortimento* de florestas de produção (nativas ou plantadas) é uma das ferramentas importantes no planejamento da utilização dos recursos florestais e na avaliação do

manejo empregado, em razão de permitir a quantificação da matéria-prima, classificando-a de acordo com as dimensões e qualidade, com relação aos múltiplos usos da madeira (MÜLLER, FINGER e SCHNEIDER, 2005).

Na literatura foram propostas inúmeras funções de afilamento com diferentes formas e graus de complexidade. Dentre esses modelos existem tanto os que se utilizam de abordagens paramétricas quanto os que se utilizam de abordagens não paramétricas. O grupo de modelos de afilamento que se utilizam de abordagens paramétricas inclui várias funções lineares e não lineares capazes de estimar os diâmetros a qualquer ponto do fuste em função do DAP e, por vezes, outros preditores, variando desde polinômios simples até modelos de regressão não linear e multivariada (ROJO *et al.*, 2005).

Ao longo dos anos, diversas formas de classificação dos tipos de modelos de afilamento foram propostas, sendo as mais recentes baseadas na separação de acordo com a função matemática utilizada no modelo, como polinomial simples, polinomial segmentado, potencial, exponencial, forma variável, trigonométricas e modelos logarítmicos (CASTELO-DORADO e ÁLVAREZ, 2000; NOVO, ROJO e ÁLVAREZ-GONZÁLEZ, 2003). Uma outra forma de separação dos tipos de modelos, também baseando-se no tipo de função matemática utilizada é a proposta por Diéguez-Aranda *et al.* (2006), que os separa em polinomial simples (não-segmentados), segmentados, e os modelos de forma variável.

2.3.1. Funções de afilamento não-segmentadas

Os primeiros modelos de afilamento a serem implementados foram os modelos mais simples, os não segmentados, que se utilizam de uma única equação que descreve a mudança de diâmetro do solo ao topo da árvore, sendo assim, esses modelos são os menos flexíveis, pois descrevem a árvore como se possuísse uma única forma geométrica, isso pode incidir em grandes problemas de estimativas, principalmente nas regiões próximas à base ou ao ápice do fuste (MÔRA, 2015).

Os primeiros modelos de afilamento desenvolvidos e utilizados foram os polinômios de pequena ordem (ROJO *et al.*, 2005), porém o tronco de uma árvore não é um sólido definido, sendo formado por várias formas geométricas (MÔRA, 2015). A região inferior do tronco de uma árvore é geralmente assumida como neilóide, a porção mediana é assumida como parabolóide, já a região apical é assumida como um tronco de cone (HUSCH; MILLER; BEERS, 1982),

A primeira equação de afilamento conhecida foi proposta por Höjer em 1903 (BEHRE, 1923), esta equação foi desenvolvida para expressar o diâmetro em qualquer ponto do tronco de *Picea abies*.

Trabalhando com a função proposta por Höjer, Jonson (1910 apud BEHRE, 1923), ao comparar suas estimativas com dados provenientes de medições reais do afilamento de *Pinus sylvestris*, encontrou elevada tendenciosidade dessa equação nas estimativas do ápice das árvores e tentou corrigir essa inclinação com o uso de uma “constante biológica”, alterando a função proposta por Höjer.

Behre (1923), também estudando a função de Höjer na tentativa de propor a introdução de um novo parâmetro ou termo, propôs uma nova equação para descrever o afilamento do fuste, através de dados provenientes de uma floresta de *Pinus ponderosa*. Essa equação proposta descrevia o fuste de forma hiperbólica, estimando a forma média do fuste das árvores de maneira mais consistente.

Kozak, Munro e Smith (1969), trabalhando com *Picea sitchensis* e *Cedrela fissilis*, desenvolveram 3 equações para descrever o afilamento dessas espécies com base em funções parabólicas, partindo-se do pressuposto que a árvore é um parabolóide quadrático. Dezenove espécies de importância para a região da Columbia Britânica, no Canadá foram testadas e, para algumas das espécies verificou-se um problema em uma das equações propostas, pois apresentava sérias tendências nas porções mais altas das árvores, as vezes resultando em diâmetros negativos. Dessa forma, propuseram uma nova fórmula que garantiria que quando a altura de medição fosse igual à altura total da árvore, o diâmetro medido pela equação deveria ser necessariamente igual a zero.

O maior desafio dos modelos de afilamento ao longo dos anos sempre foi de se manterem acurados mesmo nas estimativas dos diâmetros mais próximos à linha do solo, sendo que a maioria das equações não-segmentadas desenvolvidas com base em polinômios de baixo grau apresentam vieses significativos para essas estimativas (Bi, 2000; Kozak, 1988; ROJO *et al.*, 2005). Bi (2000) reforça o fato de que essas equações são incapazes de levar em conta as diferentes formas que constituem o fuste das árvores.

Porém, Kozak (1988) afirma que essas equações não são de todo ruins e que possuem pontos vantajosos, que seriam relacionados à facilidade de ajuste, à fácil integração para o cálculo do volume, e à facilidade de se isolar variáveis do modelo para o cálculo da altura comercial dado um determinado diâmetro, por exemplo.

Devido a esses pontos negativos observados na maioria das equações não-segmentadas mais simples, diversos estudos foram desenvolvidos com polinômios de grau mais elevado com

o intuito de se obter estimativas mais precisas nessas regiões, como os polinômios de Schöepfer (1966), Goulding e Murray (1976) e Bruce, Curtis e Vancoevering (1968).

Alguns autores optaram por fazer suas alterações nos princípios que regem suas funções, ou mesmo nas variáveis independentes. Por exemplo, Benhabrim e Gavaland (2003), estudando o afilamento de dois híbridos do gênero *Populus*, em Orléans, na França, propuseram a utilização do diâmetro na linha do solo ao invés do DAP. Os autores apresentaram uma equação que possibilita o cálculo de maneira simplificada do diâmetro dada uma altura, mas também da altura dado um diâmetro.

2.3.2. Funções de afilamento segmentadas

A maioria dos modelos de afilamento desenvolvidos até hoje são baseados em funções não-segmentadas, porém, algumas vezes esses modelos geram previsões insatisfatórias (MÔRA, 2015).

Max e Burkhart (1976), baseando-se na afirmação de que as árvores não são formadas por um único sólido geométrico, constataram a necessidade de três modelos para descrever o afilamento do fuste, um modelo para cada segmento da árvore. Posteriormente, estes modelos foram unidos para formar um único modelo, introduzindo o conceito de modelos de regressão polinomial segmentada, onde diferentes equações eram combinadas para representarem várias partes do fuste, fazendo a junção dessas equações através de pontos de inflexão.

Os modelos segmentados foram desenvolvidos com o intuito de melhorar as estimativas ao longo do tronco, corrigindo principalmente os efeitos tendenciosos na base e no ápice do fuste (MÔRA, 2015).

Com o advento dos computadores e sua conseqüente introdução nas pesquisas florestais, por volta da década de 1960, além do aumento do número de programas disponíveis para a realização de cálculos matemáticos (Muhairwe, 1993), o caminho para o desenvolvimento de equações cada vez mais complexas foi aberto.

Sendo assim, outros pesquisadores do setor florestal desenvolveram funções segmentadas, dentre eles cita-se os de maior importância como os modelos propostos por Demaerschalk e Kozak (1977), Cao, Burkhart e Max (1980), Parresol, Hotvedt e Cao (1987) e Clark *et al.* (1991).

Dentre as funções segmentadas citadas acima, a proposta por Clark *et al.* (1991) possui maior destaque. Até os dias de hoje, em estudos onde diversas metodologias inovadoras e complexas estão disponíveis, quando pressupostos e condições da equação de Clark *et al.* (1991) são satisfeitas, esta se comporta como a melhor ou uma das melhores equações para o

processamento do afilamento (LI e WEISKITTEL, 2010; ÖZÇELIK e CRECENTE-CAMPO, 2016). Esta equação se utiliza do valor do diâmetro mensurado a 5,30 m de altura para o cálculo da forma do fuste. Quando existe a possibilidade de se mensurar esse diâmetro a 5,30 m, o modelo de Clark *et al.* (1991) se comporta de maneira muito superior com estimativas muito precisas (LI e WEISKITTEL, 2010), porém, não são todos os casos onde essa medição é possível de ser realizada. Nestes casos, quando essa medição precisa ser estimada por algum outro método, sua precisão cai significativamente.

As funções segmentadas possuem uma vantagem em relação às não-segmentadas por serem capazes de prever os diâmetros com menor viés na maioria das partes do fuste (KOZAK, 1988), proporcionando melhores ajustes do perfil do tronco em comparação aos modelos não-segmentados, principalmente no cálculo do volume na base do fuste (CAO, BURKHART e MAX, 1980).

A principal desvantagem de uma equação de afilamento segmentada, todavia, é a necessidade dos pontos de inflexão. Esses pontos são considerados como restrições do modelo, sendo determinantes para o bom ajuste dos mesmos, porém são difíceis de serem estimados, dificultando o ajuste e a utilização destes modelos.

Além disso, uma outra desvantagem que pode ser considerada é a de que para esses modelos, as equações que se obtém altura comercial ou o processo de integração para se chegar no volume comercial são difíceis de se obter ou até mesmo inexistentes (MÔRA, 2015).

O conceito de equações segmentadas divide a árvore em duas ou mais formas distintas, utilizando-se do ponto de inflexão para trocar de uma equação para outra, ou seja, mudar a forma do fuste daquele ponto em diante. Entretanto as árvores não podem ser facilmente definidas por uma única equação de formulação simples, pois não são constituídas apenas por 3 formas, como a maioria das equações segmentadas as representam, mas sim assumem um número infinito de formas (GROSENBAUGH, 1966).

Kozak (1988) e Newnham (1988) apontam que, diferentemente do que dizem as equações segmentadas, a forma do fuste das árvores é contínua e muda suavemente ao longo do tronco, não havendo um ponto específico, mas sim uma região, onde se observa a troca entre as infinitas formas geométricas que essas árvores assumem.

2.3.3. Funções de afilamento de forma variável

Buscando aprimorar o desempenho das funções de afilamento, baseando-se no princípio da troca suave e contínua da forma da árvore ao longo do fuste, e na tentativa de facilitar o ajuste das equações quanto à necessidade de pontos de inflexão, Kozak (1988) e Newnham

(1988) introduziram o conceito de modelos de afilamento de forma variável. Estes modelos, também chamados de modelos de expoente variável, utilizam-se de uma única função contínua para descrever a forma das árvores, assim como os modelos de regressão não-segmentados, mas que, de forma implícita, dividem a árvore em três ou mais segmentos, representando as infinitas formas do fuste. Estes modelos não necessitam do auxílio dos pontos de inflexão, pois a mudança de forma é ocasionada pelos expoentes variáveis ao longo do tronco, representando assim o neilóide, parabolóide, tronco de cone e todas as demais formas intermediárias que a árvore possa apresentar.

O expoente das funções de afilamento de forma variável funciona como uma função própria de potência, criando uma forma contínua no tronco, semelhante aos modelos não-segmentados, que descreviam o fuste nas diferentes porções da árvore (MÔRA, 2015). Essa abordagem em geral, quando comparada com a abordagem segmentada, possui menor viés e maior acurácia ao realizar as previsões do afilamento (NEWNHAM, 1988; KOZAK, 1988).

Após ser utilizada por quase uma década sem apresentar problemas sérios, a equação de afilamento de Kozak (1988) demonstrou não ser capaz de estimar altura e/ou volumes comerciais sem apresentar viés para algumas espécies de árvores da Columbia Britânica (KOZAK, 1997). Esse erro aconteceu durante a modelagem das maiores árvores de *Thuja plicata* e *Tsuga heterophylla*, espécies da região do Rio Columbia. Além disso a equação apresentou tendências a superestimar os maiores volumes e subestimar os menores volumes.

Kozak (1997) observou a existência de multicolinearidade entre algumas das variáveis independentes da equação proposta pelo autor em 1988. Além disso, como os dados de afilamento são compostos de diversas observações de uma mesma árvore, existe autocorrelação, ou seja, ocorre a violação da suposição de erros independentes, ainda segundo Kozak (1997) essa afirmação é verdadeira para toda e qualquer equação de afilamento desenvolvida e publicada até aquele momento. Sendo assim, Kozak (1997) propôs o uso de duas novas equações, propostas pelo mesmo autor em 1994 e 1995, que prometem reduzir os efeitos negativos da primeira equação de forma variável de Kozak (1988) e também da proposta por Newnham (1988).

Muhairwe (1999) aponta que até então existiam poucos estudos voltados ao afilamento de árvores nativas da Austrália, principalmente para as florestas nativas de *Eucalyptus*. Diante dessa situação, o estudo do afilamento de espécies de grande importância para a economia da região de New South Wales, na Nova Zelândia como *Eucalyptus pilularis* e *Eucalyptus grandis*, foi realizado propondo duas equações, uma para melhor modelar o afilamento de *E. pilularis* e outra para melhor modelar o afilamento de *E. grandis*.

Bi (2000) desenvolveu uma equação de forma variável com base em equações de volume que se utilizam de princípios trigonométricos para estimar o afilamento de diversas espécies de *Eucalyptus*. O modelo proposto demonstrou ser superior aos propostos por Kozak (1988) e Newnham (1988) para o processamento do afilamento das espécies estudadas.

Kozak (2004) publicou o artigo “*My last words in taper equations*”, ou seja, minhas palavras finais em equações de afilamento. Com o objetivo de resolver os problemas de estimação da equação proposta pelo autor em 1988, e de multicolinearidade da equação também proposta pelo autor em 1994. O autor apresenta dois novos modelos, considerados como bons substitutos dos modelos anteriores, sendo o modelo dois o que obtém as melhores estimativas.

Apesar de muito estudados, segundo Li e Weiskittel (2010), os desempenhos das equações de afilamento do tronco são frequentemente comparados a um número limitado de outros modelos de formatos alternativos, além disso essas equações foram desenvolvidas para certas espécies, fazendo com que muitas vezes, os parâmetros propostos sejam específicos para o comportamento do perfil do fuste dessas espécies.

Além disso, as empresas de base florestal, por basearem-se na agregação de renda e maximização dos lucros, buscam estimativas cada vez mais acuradas do volume comercial de madeira, tanto das árvores, individualmente, como também por área ou para todo o povoamento (MÔRA, 2015).

Sendo assim, faz-se necessário e entende-se como benéfico o estudo aprofundado das características dessas equações de afilamento, de forma a estender o uso a outras espécies, além daquelas para as quais foram originalmente desenvolvidas.

3. MATERIAL E MÉTODOS

3.1. OBTENÇÃO DAS BASES DE DADOS

3.1.1. Equações de volume

Os dados para ajuste das equações volumétricas foram coletados em dois plantios clonais de híbridos de *Eucalyptus* sp. ambos localizados no sul do estado de São Paulo, com idade aproximada de sete anos, chamados a seguir de Fazenda A e Fazenda B. Os pontos de coleta dos dados no interior da fazenda foram selecionados de forma aleatória e, nesses pontos, foram abatidas 60 árvores na Fazenda A e 50 árvores na Fazenda B para a realização da cubagem rigorosa. Em ambas as fazendas os diâmetros da cubagem rigorosa foram mensurados nas alturas 0,1; 0,5; 1,5; 2; 3; 4 m e assim sucessivamente, de 1 em 1 metro, até a altura total da árvore. Os volumes com casca das árvores para ajuste e comparação com as equações volumétricas foi obtido pelo método de Smalian.

3.1.2. Equações de afilamento

Os dados para ajuste das equações de afilamento foram coletados em plantios clonais de híbridos de *Eucalyptus* sp., localizados no Estado do Mato Grosso do Sul, no município de Água Clara, com idade aproximada de sete anos. Os pontos de coleta de dados no interior da fazenda foram selecionados de forma aleatória, sendo abatidas 75 árvores nesses pontos. Essas 75 árvores foram distribuídas igualmente em três diferentes materiais genéticos (A, B e C). Os indivíduos amostrados foram abatidos e cubados rigorosamente realizando-se a mensuração dos diâmetros nas alturas 0,2; 0,7; 1,3; 2; 3; 4 m e assim sucessivamente, de 1 em 1 metro, até atingir o diâmetro comercial mínimo de 5 cm com casca.

A coleta dos diâmetros nas alturas de medição foi realizada por meio de fotografias digitais de discos de madeira retirados nas alturas de medição, as quais foram processadas com o auxílio do *software* R versão 3.6.0 (R CORE TEAM, 2019) e com o uso da técnica de Redes Neurais Artificiais (RNA) conforme Thiersch *et al.* (2020).

3.1.2.1. Obtenção dos dados de volume real

Para a obtenção do volume real utilizou-se de um xilômetro, que consiste em um tubo cilíndrico que contém água em seu interior e um tubo menor, em formato de L, acoplado à região média do cilindro, com uma parte graduada, geralmente em vidro, de forma a se poder observar o deslocamento do nível de água na região graduada, assim, ao mergulhar-se as toras de madeira no cilindro, a relação entre o volume de água deslocado e o volume da tora segue a proporção 1:1. Na medição com o xilômetro para utilização com os modelos de afilamento do

fuste, as árvores foram previamente seccionadas e descascadas com o intuito de determinar o volume real sem casca de cada seção.

3.2. PROCESSAMENTO DOS DADOS

3.2.1. Definições

Considere-se as observações $(Y_i, X_i), i = 1, \dots, n$, onde $X_i = (X_{i1}, \dots, X_{ij})$ é a variável explicativa de dimensão j e Y_i é a variável de resposta. Assume-se que $(Y_i, X_i)_{i=1}^n$ sejam i.i.d. cópias de um par aleatório (Y, X) . O objetivo é estimar a relação entre a variável resposta e a explicativa, dado que:

$$Y = f(X) + \varepsilon$$

Onde: $f(X)$ é a verdadeira função de regressão que explica as variações da variável Y e o erro aleatório ε é independente de X e possui distribuição normal com variância (σ^2) desconhecida.

Para estimar f , um conjunto j de k plausíveis modelos é considerado:

$$Y = f_k(X, \theta_k) + \varepsilon$$

Onde θ_k é o parâmetro a ser estimado (geralmente vetorial). Para cada modelo diferentes métodos podem ser utilizados para estimar θ_k .

3.2.2. Inferência Multimodelos – Algoritmo

3.2.2.1. Algoritmo ARM para predição volumétrica

O algoritmo ARM (*Adaptive Regression by Mixing* – Regressão Adaptativa por Mixagem) foi proposto inicialmente por Yang (2001) como método de realização de inferência estatística utilizando-se de vários modelos. Ao invés de selecionar-se o melhor modelo através dos critérios de informação como AIC ou BIC, realizando a inferência apenas por meio do modelo selecionado, o algoritmo realiza o cálculo de pesos através do uso de uma medida de discrepância e define que a inferência deva ser realizada combinando múltiplos modelos de acordo com seus respectivos pesos.

Yang (2003) demonstrou amplamente a comparação entre a seleção e a mixagem de modelos e propôs uma alteração na fórmula do cálculo dos pesos para o algoritmo ARM. Sendo assim, este trabalho utilizou o método proposto e atualizado pelo autor em 2003.

Existem duas etapas principais envolvidas no algoritmo. Para a primeira etapa, metade da amostra é utilizada para o ajuste dos modelos, ou seja, para estimar θ_k de cada um deles; na segunda etapa a metade restante da amostra é utilizada para a predição baseada nos modelos

ajustados, então é calculada uma medida de discrepância entre os valores observados e os valores preditos, possibilitando o cálculo dos pesos para cada modelo.

Dada uma amostra de tamanho n com observações $Z^n = (Y_i, X_i)_{i=1}^n$, o algoritmo ARM é dado por:

- Passo 1. Separar a base de dados (Z) em duas partes: $Z^{(1)} = (X_i, Y_i)_{i=1}^{n/2}$ e $Z^{(2)} = (X_i, Y_i)_{i=n/2+1}^n$.
- Passo 2. Estimar θ_k por $\hat{\theta}_k = \hat{\theta}_{k,n/2}$ por um método de mínimos quadrados, baseando-se em $Z^{(1)}$. Obter, a variância da função $\sigma^2(X)$, por $\hat{\sigma}_k^2 = \hat{\sigma}_{k,n/2}^2$ novamente baseando-se apenas em $Z^{(1)}$.
- Passo 3. Obter as acurácias dos modelos usando a segunda metade da base de dados $Z^{(2)}$. Para cada k , para $n/2 + 1 \leq i \leq n$, prever Y_i por $f_k(X_i, \hat{\theta}_k)$.
Calcular a medida de discrepância $D_k = \sum_{i=n/2+1}^n (Y_i - f_k(X_i, \hat{\theta}_k))^2$.
- Passo 4. Calcule o peso para cada modelo k :

$$W_k = \frac{(\hat{\sigma}_k)^{-n/2} \exp(\hat{\sigma}_k^2 D_k/2)}{\sum_{j=1}^K (\hat{\sigma}_j)^{-n/2} \exp(\hat{\sigma}_j^2 D_j/2)}$$

Note que: $\sum_{k=1}^K W_k = 1$.

- Passo 5. Permutar aleatoriamente a ordem dos dados (Yang (2003) recomenda 250 permutações). Repetir os 4 passos anteriores para cada aleatorização e, sendo $W_{k,r}$, $k = 1, \dots, K$ o peso do modelo k computado na r -ésima vez para $1 \leq r \leq M$. Seja $\hat{W}_k = 1/M \sum_{r=1}^M W_{k,r}$ o estimador final de f é dado por:
 $\hat{f}_n(X) = \sum_{k=1}^K \hat{W}_k f_k(X, \hat{\theta}_{k,n})$.

Este algoritmo, por trabalhar o cálculo dos pesos baseando-se nos erros de cada modelo, não possui restrição para o cálculo da variável de interesse final. Dessa forma, permite o uso de modelos lineares, não lineares e logarítmicos simultaneamente. Dada essa característica o algoritmo permite também que o cálculo do volume de um povoamento, por exemplo, seja feito através do uso de equações volumétricas, de afilamento não segmentadas, segmentadas, de forma variável e potências fracionárias.

Neste estudo foi adicionado um passo inicial (Passo 0) que consiste na estratificação da base de dados em três classes diamétricas todas com o mesmo número de amostras contendo, respectivamente, a terça parte dos dados com os menores DAP, a terça parte com os DAP médios e a terça parte com os maiores DAP. Portanto, cada porção de dados respectiva a uma

classe diamétrica consistiu na base de dados para a aplicação do algoritmo ARM, sendo assim a sequência de passos 1 a 5 foi repetida 3 vezes, uma para cada classe diamétrica.

Isso faz com que os pesos de cada modelo sejam calculados de acordo com a classe diamétrica, sendo assim o volume final será composto conforme o tamanho das árvores.

- Passo 0: Ordenar a base de dados em ordem crescente de DAP, e separar em 3 partes: $Z^{(a)} = (X_i, Y_i)_{i=1}^{n/3}$, $Z^{(b)} = (X_i, Y_i)_{i=n/3+1}^{2n/3}$ e $Z^{(c)} = (X_i, Y_i)_{i=2n/3+1}^n$.

Sendo assim, o Passo 1 também deve ser alterado para a nomenclatura correta, sejam três amostras ($Z^{(a)}$, $Z^{(b)}$ e $Z^{(c)}$) de mesmo tamanho n , todas com observações $Z^n = (Y_i, X_i)_{i=1}^n$, o novo Passo 1 do algoritmo ARM é dado por:

- Passo 1: Separar cada conjunto de dados ($Z^{(a)}$, $Z^{(b)}$ e $Z^{(c)}$) em duas partes: $Z^{(1)} = (X_i, Y_i)_{i=1}^{n/2}$ e $Z^{(2)} = (X_i, Y_i)_{i=n/2+1}^n$.

3.2.2.2. Algoritmo MTE (*Mixed Taper Equations*)

O algoritmo ARM, ao propor a mixagem de modelos, promove, em tese, aumento na acurácia, em comparação com a seleção, pois a tendência é de que o resultado final se apresente no mínimo igual ao do modelo selecionado por AIC ou BIC (YANG, 2001; YANG, 2003).

Para o caso do afilamento do fuste, como este possui a característica de estimar infinitos diâmetros ao longo do fuste, altura a qualquer diâmetro e volume entre quaisquer segmentos do tronco, a forma de se realizar o uso do algoritmo pode ser diferente, visto que o mesmo propõe um peso único para toda a base dos dados, portanto, independentemente da altura da árvore em que se estiver pretendendo calcular alguma informação, o peso estipulado para o modelo será o mesmo.

Sendo assim, propõe-se uma abordagem diferente, a qual foi denominada de algoritmo MTE (*Mixed Taper Equations*, do inglês, Equações de Afilamento Mixadas), particionando os dados árvore a árvore em regiões definidas, em pontos fixos ao longo do fuste, para que o algoritmo então calcule diferentes pesos em cada fração da altura para todos os modelos. Dessa forma, espera-se reduzir o impacto dos erros de estimativa de cada modelo, visto que diferentes modelos apresentam diferentes comportamentos, podendo subestimar ou superestimar os dados de acordo com a região do fuste em que realizam as estimativas.

Na literatura é comum verificar que diferentes modelos de afilamento tendem a estimar os perfis do fuste de maneiras distintas, visto que determinados modelos erram mais na região da altura mais próxima ao solo, outros na região do ápice das árvores enquanto outros possuem tendência a errar mais nas regiões intermediárias da árvore (ASSIS, 2002; DIÉGUEZ-ARANDA *et al.*, 2006; LI e WEISKITTEL, 2010; MÔRA, 2015; ÖZÇELIK e CRECENTE CAMPO, 2016).

Portanto, reconhecendo essa característica dos modelos de afilamento, o algoritmo MTE baseia-se na escolha de pesos dentre os melhores modelos para as estimativas dos diâmetros em cada porção da árvore, aproveitando o que cada modelo, em sua particularidade, oferece de melhor, assim, foram então determinados pesos para cada modelo ao longo do fuste nas alturas absolutas provenientes das medições da cubagem para cada base de dados.

Dada uma amostra de n árvores com observações $Z^n = (Y_i, X_i)_{i=1}^n$, o algoritmo MTE é dado por:

- Passo 1. Separar a base de dados em duas partes iguais de acordo com o número de árvores, mantendo todas as medições de uma mesma árvore em um mesmo estrato, sendo: $Z^{(1)} = (X_i, Y_i)_{i=1}^{n/2}$ e $Z^{(2)} = (X_i, Y_i)_{i=n/2+1}^n$.
- Passo 2. Estimar θ_k por $\hat{\theta}_k = \hat{\theta}_{k,n/2}$ por um método de mínimos quadrados, baseando-se em $Z^{(1)}$.
- Passo 3. Estratificar $Z^{(1)}$ em n_{MAX_1} partes, sendo n_{MAX_1} o número de estratos definidos pelas alturas relativas (h_i) sendo, por exemplo: $h_i = 0,1; 0,7; 1,3; 2; 3; \dots; h_{n_{MAX}}$ de forma que: $Z'^{(1)} = (X_i, Y_i)_{i=1}^{h_i=0,1}$; $Z'^{(2)} = (X_i, Y_i)_{i=2}^{h_i=0,7}; \dots; Z'^{(h)} = (X_i, Y_i)_{i=n_{MAX}}^{h_i=h_{n_{MAX}}}$.
- Passo 4. Obter, a variância da função $\sigma^2(X)$, por:

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^n \sigma_{k,Z'^{(1)}}^2 + \sigma_{k,Z'^{(2)}}^2 + \dots + \sigma_{k,Z'^{(h)}}^2}{n_{MAX}}$$
- Passo 5. Estratificar $Z^{(2)}$ em n_{MAX_2} partes, sendo n_{MAX_2} o número de estratos definidos pelas alturas relativas (h_i) sendo, por exemplo: $h_i = 0,1; 0,7; 1,3; 2; 3; \dots; h_{n_{MAX}}$ de forma que: $Z''^{(1)} = (X_i, Y_i)_{i=1}^{h_i=0,1}$; $Z''^{(2)} = (X_i, Y_i)_{i=2}^{h_i=0,7}; \dots; Z''^{(h)} = (X_i, Y_i)_{i=n_{MAX}}^{h_i=h_{n_{MAX}}}$.
- Passo 6. Obter as acurácias dos modelos cada porção da segunda metade da base de dados $Z''^{(i)}$. Para cada modelo k , para $n_{Z''}/2 + 1 \leq i \leq n_{Z''}$, predizer

Y_i por $f_k(X_i, \hat{\theta}_k)$. Calcular a medida de discrepância por $D_k = \sum_{i=n/2+1}^n (Y_i - f_k(X_i, \hat{\theta}_k))^2$.

- Passo 7. Calcular o peso para cada modelo k , para cada porção da altura, $Z''^{(i)}$, por:

$$W_k = \frac{(\hat{\sigma}_k)^{-n/2} \exp(\hat{\sigma}_k^2 D_k/2)}{\sum_{j=1}^K (\hat{\sigma}_j)^{-n/2} \exp(\hat{\sigma}_j^2 D_j/2)}$$

Note que: $\sum_{k=1}^K W_k = 1$.

- Passo 8. Permutar aleatoriamente a ordem dos dados (100 permutações). Repetir as 7 etapas e, sendo $W_{k,r}$, $k = 1, \dots, K$ o peso do modelo k computado na r -ésima vez para $1 \leq r \leq M$, seja $\hat{W}_k = 1/M \sum_{r=1}^M W_{k,r}$ o estimador de f é dado por: $\hat{f}_{n_z}(X) = \sum_{k=1}^K \hat{W}_k f_k(X, \hat{\theta}_{k,n_z})$.
- Passo 9. Reconstituir o volume a partir da integral de cada modelo de afilamento de acordo com seus respectivos pesos. Sendo W_{kh_i} , $k = 1, \dots, K$ o peso do modelo k computado na h_i -ésima altura, para $h_i = 0,1; 0,7; 1,3; 2; 3; \dots; h_{n_{MAX}}$, os limites de cada integral serão definidos pelas alturas h_i e o peso dos modelos utilizado para o cálculo do volume final será o referente ao limite superior.

3.2.3. Critérios estatísticos utilizados

Os desempenhos dos algoritmos ARM e MTE foram comparados com os desempenhos dos modelos volumétricos e de afilamento ajustados, respectivamente, de forma a afirmar a eficácia dos métodos propostos. Os critérios utilizados foram:

A raiz do erro quadrático médio em porcentagem (RMSE), utilizado para o julgamento da performance do ajuste de equações de regressão, onde a melhor equação é a qual se observa o menor valor para tal medida de precisão. O “*Mean Absolute Bias*” (MAB), como uma forma de medir o viés segundo Kozak (2004), calculado como a diferença entre o verdadeiro valor do parâmetro e o valor estimado pela equação, indicando tendenciosidade na equação testada e, portanto, quanto menor o valor, menor a tendência total apresentada pela equação. Também foi realizada a análise do comportamento gráfico da dispersão dos resíduos para os modelos ajustados e para a inferência produzida pelos algoritmos. Para a comparação do algoritmo ARM com as equações volumétricas apenas o RMSE% foi utilizado.

$$RMSE\% = \frac{\sqrt{\frac{\sum (Y_i - f_k(X_i, \hat{\theta}_k))^2}{n - p - 1}}}{\bar{Y}} \times 100$$

$$MAB = \frac{\sum (|Y_i - f_k(X_i, \hat{\theta}_k)|)}{n}$$

Onde: n = número de observações, p = número de parâmetros do modelo; Y_i = valor observado, $f_k(X_i, \hat{\theta}_k)$ = valor estimado e \bar{Y} = média observada.

3.2.4. Modelos volumétricos testados

Inicialmente para testar a eficácia do método proposto pelo algoritmo ARM em relação aos critérios de seleção, foram utilizados modelos volumétricos que realizam uma única estimativa de volume para cada árvore amostrada. Quatro modelos volumétricos foram escolhidos (Tabela 1) (LOETSCH, ZÖHRER e HALLER, 1973, apud COELHO, 2016). O ajuste e aplicação dos modelos volumétricos, assim como a implementação do algoritmo ARM foram realizados com o *software* estatístico R versão 3.6.0 (R CORE TEAM, 2019).

Tabela 1. Modelos volumétricos selecionados para o ajuste e aplicação com o algoritmo ARM

Nº	Equação	Nome	Referência
(1)	$\ln(v) = \beta_0 + \beta_1 \ln(d) + \beta_2 \ln(h) + \varepsilon$	Schumacher e Hall logarítmica	Schumacher e Hall (1933)
(2)	$\ln(v) = \beta_0 + \beta_1 \ln(d^2 h) + \varepsilon$	Spurr logarítmica	Spurr (1952)
(3)	$v = \beta_0 d^{\beta_1} h^{\beta_2} + \varepsilon$	Schumacher e Hall	Schumacher e Hall (1933)
(4)	$v = \beta_0 + \beta_1 d^2 h$	Spurr	Spurr (1952)

Sendo: v = volume; h = altura total da árvore; d = DAP; β_i = parâmetros a serem estimados; \ln = logaritmo neperiano, ε = erro aleatório.

3.2.5. Modelos de afilamento testados

Para a aplicação e desenvolvimento do algoritmo MTE foram selecionados os modelos de afilamento de forma variável (Tabela 2) de Muhairwe (1999) e Kozak (2004), o modelo segmentado de Max e Burkhart (1976) e o modelo não segmentado de Schöpfer (1966). Toda a aplicação do algoritmo MTE bem como o ajuste das equações de afilamento e utilização no algoritmo foram implementados e desenvolvidos dentro do ambiente do *software* estatístico R versão 3.6.0 (R CORE TEAM, 2019).

Tabela 2. Modelos de afilamentos selecionados para o ajuste e aplicação com o algoritmo MTE

Nº	Equação	Referência
(5)	$d_i = \alpha_0 d^{\alpha_1} \alpha_2 d \left(1 - \sqrt{\frac{hi}{h}} \right)^c$ <p>Sendo:</p> $C = \beta_1 \left(\frac{hi}{h} \right) + \beta_2 \left(\frac{hi}{h} \right)^2 + \beta_3 / \left(\frac{hi}{h} \right) + \beta_4 \left(\frac{hi}{h} \right)^3 + \beta_5 d + \beta_6 \left(\frac{d}{h} \right)$	Muhairwe (1999)
(6)	$d_i = (\alpha_0 d^{\alpha_1} h^{\alpha_2}) m^c$ <p>Sendo:</p> $C = \beta_1 \left(\frac{hi}{h} \right)^4 + \beta_2 \left(\frac{1}{e^{d/h}} \right) + \beta_3 m^{0,1} + \beta_4 \frac{1}{d} + \beta_5 h^{(1-(hi/h)^{1/3})} + \beta_6 m$ $m = \frac{1 - \left(\frac{hi}{h} \right)^{1/3}}{1 - \left(\frac{1,3}{h} \right)^{1/3}}$	Kozak (2004)
(7)	$d_i = d \left[\beta_1 + \left(\frac{hi}{h} - 1 \right) + \beta_2 \left(\left(\frac{hi}{h} \right)^2 - 1 \right) + \beta_3 \left(\alpha_1 - \frac{hi}{h} \right)^2 I_1 + \beta_4 \left(\alpha_2 - \frac{hi}{h} \right)^2 I_2 \right]^{0,5}$ <p>Sendo: $I_i = 1$ se $\alpha_i \geq X$, caso contrário: $I_i = 0$</p>	Max e Burkhart (1976)
(8)	$d_i = d \left[\beta_0 + \beta_1 \left(\frac{hi}{h} \right) + \beta_2 \left(\frac{hi}{h} \right)^2 + \beta_3 \left(\frac{hi}{h} \right)^3 + \beta_4 \left(\frac{hi}{h} \right)^4 + \beta_5 \left(\frac{hi}{h} \right)^5 \right]$	Schöepfer (1966)

Sendo: h_i = altura relativa à seção; h = altura total da árvore; d_i = diâmetro nas alturas h_i (m); d = DAP; β_i e α_i = parâmetros da regressão.

4. RESULTADOS E DISCUSSÃO

4.1. EQUAÇÕES VOLUMÉTRICAS

O ajuste das equações volumétricas e utilização do algoritmo ARM modificado possibilitou o cálculo de pesos referente à porção da base de dados que aquele modelo foi capaz de representar. A introdução do passo 0 no algoritmo ARM para o cálculo de volume das árvores através das equações de volume permitiu aumentar a acurácia do algoritmo, conferindo resultados mais próximos dos dados coletados. Os valores obtidos para as medidas de precisão indicaram superioridade do algoritmo em estimar os volumes totais por árvore (Tabela 3).

Dentre os quatro modelos volumétricos ajustados, o modelo de Schumacher e Hall (equação 3) obteve os melhores resultados com RMSE de 3,9962% para a Fazenda A e de 3,9703% para a Fazenda B (Tabela 3).

O algoritmo ARM apresentou os melhores valores dentre todos os métodos testados em ambas as bases de dados, com RMSE de 3,5317% para a Fazenda A e de 3,8653% para a Fazenda B (Tabela 3).

Tabela 3. Desempenho das equações volumétricas em comparação ao algoritmo ARM

Equação	Fazenda	RMSE (%)	Peso classe 1	Peso classe 2	Peso classe 3
1	A	4,9211	0,1900565	0,2755241	0,3674594
2	A	4,9407	0,3305679	0,2942122	0,1447471
3	A	3,9962	0,2474532	0,1612552	0,3415394
4	A	5,0925	0,2319225	0,2690085	0,1462542
ARM	A	3,5317	-	-	-
1	B	4,5460	0,2630831	0,1766977	0,2173714
2	B	4,9772	0,3133149	0,3366447	0,2080327
3	B	3,9703	0,0984500	0,3052639	0,3497904
4	B	5,2647	0,3251520	0,1813937	0,2248055
ARM	B	3,8653	-	-	-

Onde: RMSE = Raiz do Erro Quadrático Médio; ARM = *adaptive regression by mixing* (Algoritmo)

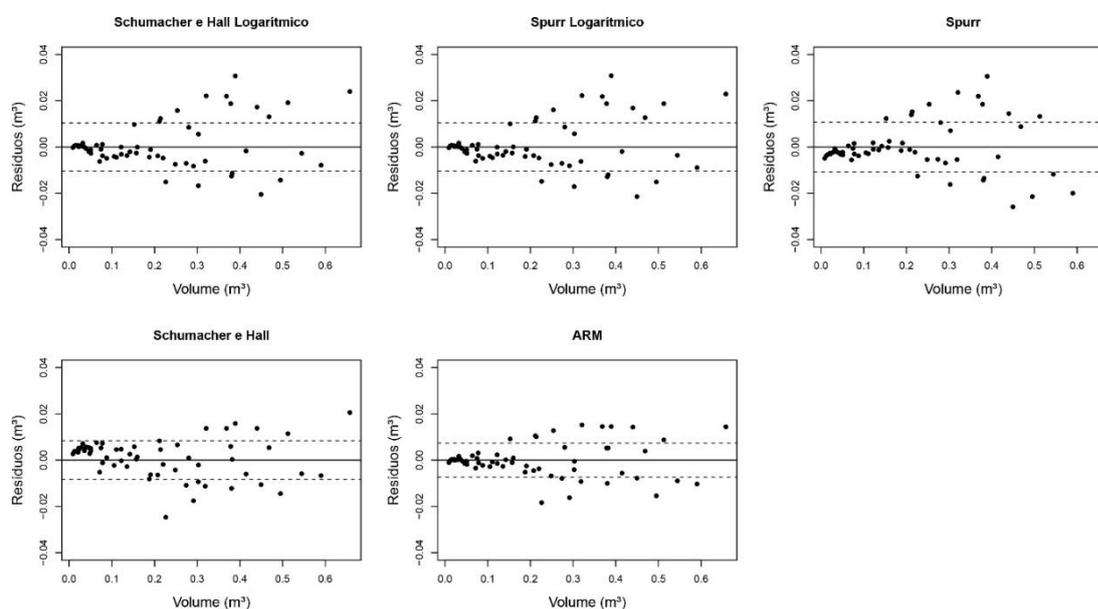
No geral, todos os modelos apresentam bons resultados na média, portanto ao comparar o desempenho dos modelos volumétricos com o algoritmo ARM apenas por medidas de precisão como o RMSE observa-se que a diferença total pode não parecer significativa.

No inventário florestal, a análise gráfica da dispersão dos resíduos possui grande relevância para a seleção do melhor modelo para realização da inferência. Modelos que apresentam comportamentos menos tendenciosos são mais confiáveis, pois, no geral, conferem resultados mais próximos a realidade em número maior de situações.

Como método ARM utiliza combinação dos modelos, o resultado obtido pelo algoritmo tende a apresentar comportamento livre de tendenciosidade, portanto, ao observar o gráfico de dispersão dos resíduos de cada um dos métodos (Figuras 1 e 2) observa-se a capacidade do algoritmo de extrair o melhor de cada modelo para compor um resultado final mais acurado.

A dispersão dos resíduos para os dados provenientes da Fazenda A (Figura 1) e da Fazenda B (Figura 2), demonstram uma melhoria nítida no cálculo dos volumes pelo algoritmo ARM, principalmente com relação à redução da heterogeneidade de variância e comportamentos tendenciosos apresentados pelos modelos ajustados.

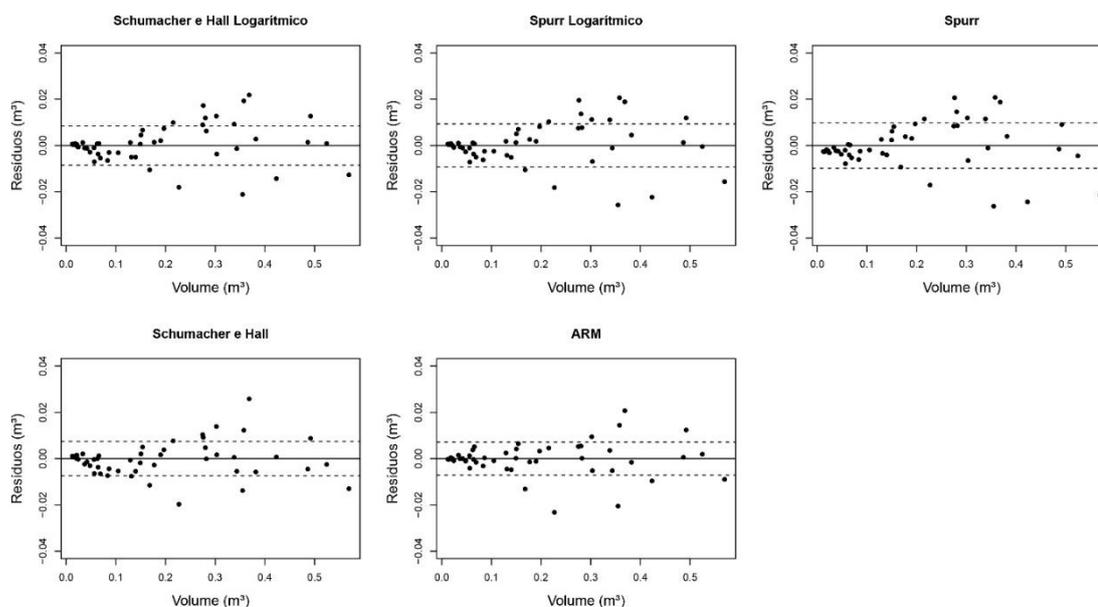
Figura 1. Dispersão dos resíduos (m^3) para a estimativa do volume total das árvores da Fazenda A pelos modelos volumétricos em comparação com a técnica de Inferência Multimodelos (ARM). A linha pontilhada representa o RMSE de cada modelo.



Em ambos os casos é possível observar a dispersão mais concentrada dos resíduos das estimativas realizadas pelo algoritmo e que a adição do passo 0 (aplicação do algoritmo em três etapas de acordo com a classe diamétrica) permitiu ao algoritmo inferir os volumes de acordo com o melhor modelo para cada classe diamétrica, visto que no geral o melhor modelo foi o de Schumacher e Hall, porém este modelo apresentou subestimativa de todas as árvores com volume até $0,05 m^3$, o que não acontece com o algoritmo que, além de não subestimar esses

volumes, ainda apresenta menores resíduos para as árvores até $1,5 \text{ m}^3$, mantendo-se com dispersão concentrada até as maiores árvores.

Figura 2. Dispersão dos resíduos (m^3) para a estimativa do volume total das árvores da Fazenda B pelos modelos volumétricos em comparação com a técnica de Inferência Multimodelos (ARM). A linha pontilhada representa o RMSE de cada modelo.



4.2. EQUAÇÕES DE AFILAMENTO

Dentre os modelos de afilamento ajustados para o material genético A, os modelos de forma variável apresentaram os melhores resultados, com o modelo de Muhairwe (1999) (equação 5) superior para as medidas de precisão, com RMSE de 3,26% e MAB de 0,31963 (Tabela 4), já o modelo não segmentado de Schöpfer (1966) (equação 8) apresentou os piores resultados, com RMSE de 4,14% e MAB de 0,41638 (Tabela 4).

Para o material genético B o modelo de Kozak (2004) (equação 6) obteve os melhores resultados para as medidas de precisão dentre os modelos de afilamento utilizados, com RMSE de 2,47% e MAB de 0,2312, enquanto que o modelo de Schöpfer (1966) (equação 8) foi o único que apresentou RMSE superior a 3% e MAB superior a 0,3, com 3,20% e 0,3110 respectivamente (Tabela 4).

Para o material genético C o modelo de Kozak (2004) (equação 6) também obteve os melhores resultados para as medidas de precisão com RMSE de 2,67% e MAB de 0,2612 e o modelo de Schöpfer (1966) (equação 8) novamente apresentou-se como o único com RMSE superior a 3% e MAB superior a 0,3, com 3,67% e 0,3552 respectivamente.

Tabela 4. Medidas de precisão para os modelos de afilamento em comparação com o algoritmo MTE para as estimativas dos diâmetros ao longo do fuste

Equação	Mat. Gen.	RMSE (%)	MAB	Peso (médio)
5	A	3,2611	0,3196294	0,2841045
6	A	3,3848	0,3266138	0,2704624
7	A	4,0239	0,4034861	0,2697508
8	A	4,1434	0,4163791	0,1756823
MTE	A	3,3492	0,3322179	-
5	B	2,4927	0,2358029	0,2354774
6	B	2,4667	0,2311980	0,1913323
7	B	2,9415	0,2812550	0,3389594
8	B	3,2014	0,3110326	0,2342309
MTE	B	2,5711	0,2471534	-
5	C	2,7652	0,2773106	0,2367030
6	C	2,6670	0,2612540	0,3022816
7	C	2,9129	0,2869845	0,3397783
8	C	3,6680	0,3552616	0,1212371
MTE	C	2,6001	0,2542880	-

Onde: Mat. Gen. = Material Genético; RMSE = Raiz do Erro Quadrático Médio; MAB = *Mean absolute Bias* (viés); MTE = *Mixed Taper Equations* (Algoritmo)

Para as três bases de dados testadas o algoritmo MTE apresentou valores intermediários estando sempre entre as segunda e terceira melhores performances tanto para RMSE quanto para MAB. Entretanto, assim como ocorreu para os modelos volumétricos, todos os modelos testados apresentaram bons resultados na média. Portanto, os resultados de RMSE e MAB apenas apresentam informações médias da performance de cada caso (modelos (Tabela 2) e algoritmo MTE), não significando que um método seja mais ou menos eficiente que outro. O objetivo do algoritmo MTE é aumentar a acurácia das estimativas de diâmetro e volume pelas equações de afilamento.

Deve-se observar que, conforme muda-se a base de dados os pesos médios de cada modelo de afilamento variam, tendo o modelo de Max e Burkhart (1976) obtido peso de 0,2697 para o material genético A, apresentando-se apenas como o terceiro modelo de maior peso, enquanto que para os materiais genéticos B e C o mesmo modelo apresentou pesos de 0,3389 e 0,3397, respectivamente, sendo o modelo com os maiores pesos médios em cada uma das bases.

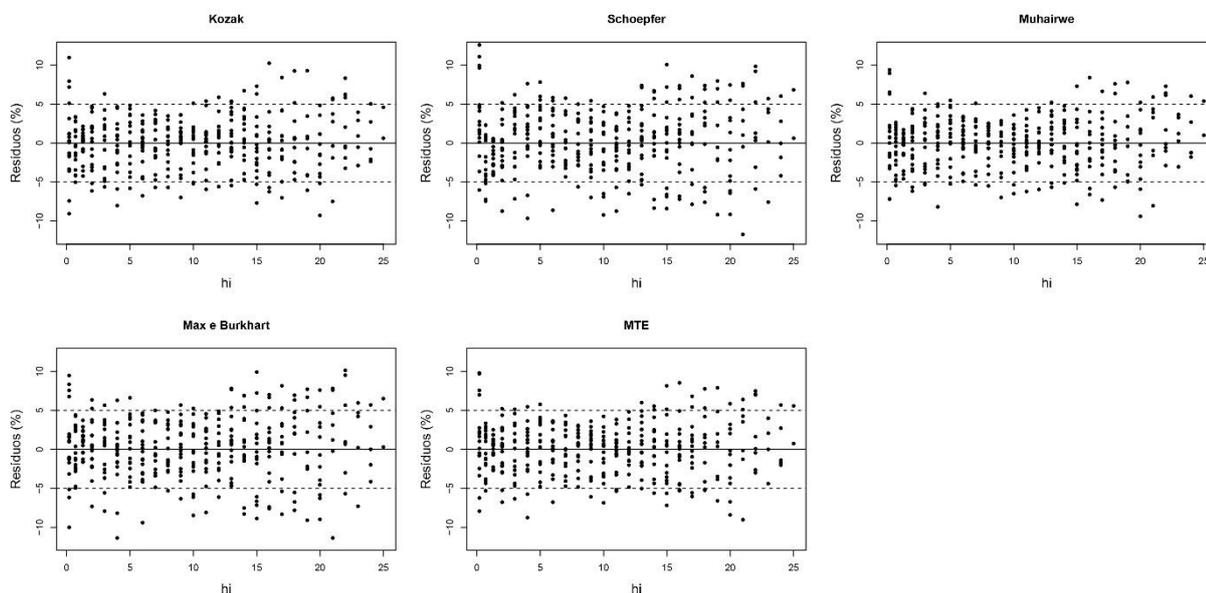
Ao analisar o modelo de Kozak (2004) observou-se que este teve grandes variações conforme o material genético, tendo apresentado peso de 0,2704 para o material genético A, sendo o segundo modelo com maior peso. Já para o caso do material genético B o modelo de Kozak (2004) obteve peso 0,1913, sendo o modelo com menor peso médio dentre todos testados e para o caso do material genético C esse modelo obteve o segundo maior peso com 0,3022. Isso demonstra que mesmo o melhor modelo perante os critérios de seleção nem sempre obterá o melhor resultado árvore a árvore, como é o caso do material genético B, que teve o modelo de Kozak (2004) como o melhor modelo segundo as medidas de precisão mas com o menor peso dentre todos os modelos.

Conforme mencionado, a distribuição dos resíduos possui grande relevância nos inventários florestais ao selecionar-se o melhor modelo, dessa maneira, a eficiência dos modelos de afilamento em comparação com o algoritmo MTE também foi feita por meio de análises gráficas.

A dispersão dos resíduos para os dados de estimativa dos diâmetros para os materiais genéticos A (Figura 3), B (Figura 4) e C (Figura 5), demonstra que o algoritmo MTE foi capaz de produzir resultados mais estáveis ao longo do fuste, com dispersão mais concentrada dos resíduos. No geral todos os modelos de afilamento testados tenderam a apresentar dispersão senoidal dos resíduos para as estimativas de diâmetro (Figuras 3, 4 e 5), porém, apenas para o caso do material genético A (Figura 3) que esse comportamento pode ser observado, de forma discreta, na dispersão dos resíduos do algoritmo MTE, nos demais casos o algoritmo apresentou comportamento estável, não mantendo o comportamento senoidal presente na maioria dos modelos. Isso se deve ao fato da estimativa realizada pelo algoritmo ser a composição de todos os modelos, o que faz com que esse comportamento senoidal seja suavizado.

Para o caso do material genético A (Figura 3) os modelos de Kozak (2004) e Muhairwe (1999) foram os principais responsáveis pela dispersão concentrada dos resíduos apresentada pelo algoritmo MTE. Porém observa-se que em alguns casos, principalmente nas regiões mais basais das árvores (onde se encontram as maiores fontes de erro de estimativa dos diâmetros devido à maior variação dos mesmos a partir da linha do solo), os outros dois modelos (Max e Burkhart (1976) e Schöepfer (1966)) apresentaram comportamento oposto, superestimando diâmetros onde os modelos de forma variável estão subestimando. Isso resulta na estimativa intermediária pelo algoritmo, que tende a zerar os resíduos naquele ponto, realizando estimativas pontuais mais assertivas.

Figura 3. Dispersão dos resíduos (%) para a estimativa dos diâmetros das árvores do material genético A pelos modelos de afilamento do fuste em comparação com a técnica de Inferência Multimodelos pelo algoritmo MTE.

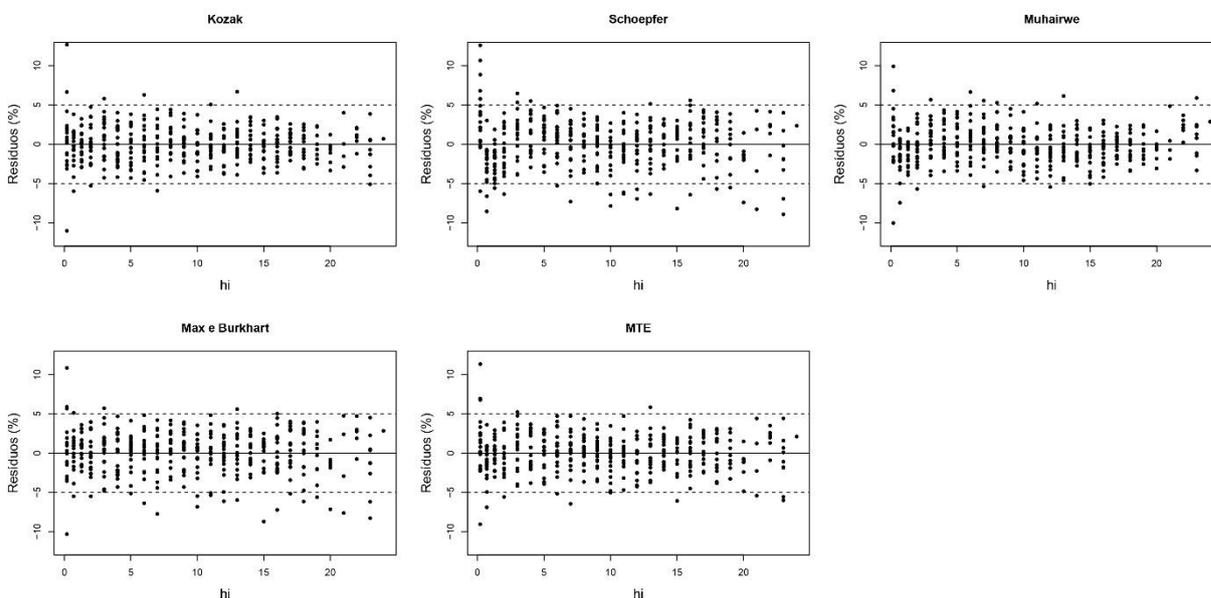


Dentre os modelos de forma variável, o modelo de Kozak (2004) apresentou erros mais dispersos, porém com comportamento mais estável para o material genético A (Figura 3), apresentando subestimativas dos diâmetros entre 5 e 14 m de altura e a partir dos 22 m. Já o modelo de Muhairwe apresentou menores erros de estimativa dos diâmetros ao longo do fuste, porém com menos estabilidade que o modelo de Kozak (2004) apresentando superestimativa destes até os 3 m de altura e subestimativa entre 4 e 11 m e acima dos 18 m de altura (Figura 3). O modelo de Max e Burkhart (1976) tendeu a subestimar todos os diâmetros ao longo do fuste, já o modelo de Schöpfer (1966) subestimou todos os diâmetros acima dos 4 m de altura.

O algoritmo MTE, ao compor as estimativas para o material genético A (Figura 3) através da união dos modelos ajustados, apresentou apenas leve subestimativa dos diâmetros acima de 16 m de altura, causados pelos modelos de Max e Burkhart (1976) e Schöpfer (1966), apresentando-se estável em todo o restante das alturas.

Para o material genético B (Figura 4), o modelo de Kozak (2004) apresentou subestimativa dos diâmetros até 1 m de altura e superestimativa a partir dos 17 m de altura, já o modelo de Muhairwe apresentou superestimativa dos diâmetros até os 3 m de altura e entre 10 e 19 m e subestimativa entre 4 e 8 m e acima dos 21 m de altura. Novamente o modelo de Max e Burkhart (1976) tendeu a subestimar todos os diâmetros ao longo do fuste e o modelo de Schöpfer (1966) superestimou até os 3 m e subestimou todos os diâmetros a partir de 4 m.

Figura 4. Dispersão dos resíduos (%) para a estimativa dos diâmetros das árvores do material genético B pelos modelos de afilamento do fuste em comparação com a técnica de Inferência Multimodelos pelo algoritmo MTE.

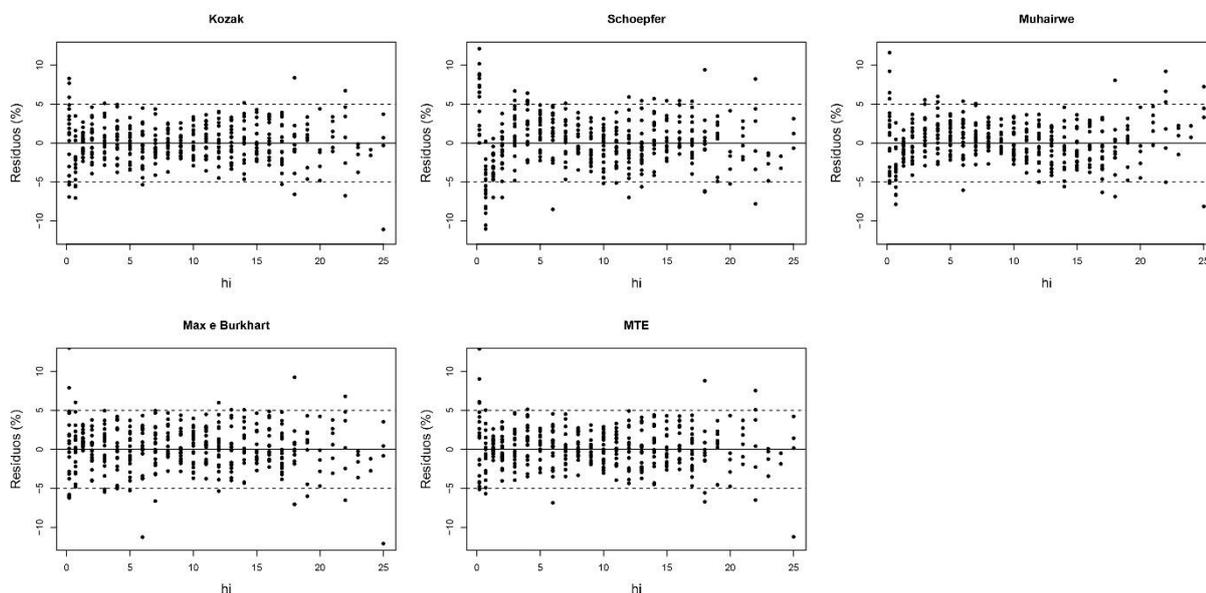


O algoritmo MTE não apresentou superestimativas nem subestimativas visíveis para o material genético B, permanecendo estável ao longo de todo o fuste das árvores, com dispersão mais concentrada do que os melhores modelos e sem apresentar o comportamento senoidal visível no gráfico do modelo de Schöpfer (1966).

Para o material genético C (Figura 5), o modelo de Kozak (2004) apresentou a melhor dispersão de resíduos para os modelos de afilamento, com superestimativa apenas dos diâmetros até 5 m de altura e subestimativa acima dos 15 m. O modelo de Muhairwe (1999) apresentou comportamento senoidal dos resíduos, o que também ocorreu com o modelo de Schöpfer (1966), o que fez com que ambos os modelos apresentassem superestimativas e subestimativas mais evidentes ao longo do fuste, superestimando os diâmetros até 4 m de altura nos dois casos e ainda entre 10 e 17 m no caso do modelo de Muhairwe (1999), além de subestimar entre 4 e 9 m nos dois casos e acima de 19 m no caso do modelo de Muhairwe (1999).

O algoritmo MTE, para o material genético C (Figura 5), não apresentou superestimativa nem subestimativa ao longo do tronco, obtendo a melhor dispersão dos resíduos dentre todos os gráficos. Além disso, novamente observa-se que o comportamento senoidal dos resíduos do modelo de Schöpfer (1966) e também de Muhairwe (1999) não foi herdado pelo algoritmo.

Figura 5. Dispersão dos resíduos (cm) para a estimativa dos diâmetros das árvores do material genético C pelos modelos de afilamento do fuste em comparação com a técnica de Inferência Multimodelos pelo algoritmo MTE. A linha pontilhada representa o RMSE de cada modelo



Os resíduos do algoritmo MTE são compostos da mistura dos resíduos dos demais modelos, portanto espera-se que modelos com comportamentos semelhantes ao longo do fuste gerem um gráfico final para o algoritmo que possua o mesmo comportamento, apresentando então os mesmos erros de estimativas, porém reduzidos pontualmente em cada medição. Em todos os gráficos de dispersão dos resíduos (Figuras 3, 4 e 5) foi possível observar a vantagem de se usar o algoritmo MTE, principalmente onde os modelos apresentam comportamentos diferentes, assim o resultado final do algoritmo foi a perfeita mistura entre as estimativas, o que resultou em dispersões mais concentradas.

Após a integração das equações de afilamento para estimar os volumes das árvores foi possível a obtenção destas estimativas também pelo algoritmo MTE. As performances dos modelos de afilamento bem como a do algoritmo podem ser observadas na Tabela 5.

O algoritmo MTE apresentou os melhores resultados para as medidas de precisão para as estimativas dos volumes das árvores dos materiais genéticos A e C, com RMSE de 4,11% e 4,91% e MAB 0,00535 e 0,00517, respectivamente, sendo o único método que obteve RMSE inferior a 5% para a estimativa de volume das árvores do material genético C. Já para as estimativas dos volumes para o material genético B, na média, o algoritmo MTE apresentou-se com o segundo melhor RMSE (8,97%) e terceiro melhor MAB (0,00701) (Tabela 5).

Tabela 5. Medidas de precisão para os modelos de afilamento em comparação com o algoritmo MTE para as estimativas dos volumes das árvores

Equação	Mat. Gen.	RMSE	MAB	Peso (médio)
5	A	4,4925	0,0054677	0,2841045
6	A	4,6108	0,0055171	0,2704624
7	A	4,9360	0,0057976	0,2697508
8	A	4,9116	0,0057422	0,1756823
MTE	A	4,1098	0,0053538	-
5	B	9,0479	0,0066079	0,2354774
6	B	8,8883	0,0066476	0,1913323
7	B	9,6549	0,0075318	0,3389594
8	B	9,6461	0,0076199	0,2342309
MTE	B	8,9645	0,0070086	-
5	C	5,0639	0,0052600	0,2367030
6	C	5,0256	0,0054486	0,3022816
7	C	5,5751	0,0054607	0,3397783
8	C	5,5298	0,0054113	0,1212371
MTE	C	4,9058	0,0051681	-

Onde: Mat. Gen. = Material Genético; RMSE = Raiz do Erro Quadrático Médio; MAB = *Mean absolute Bias* (viés); MTE = *Mixed Taper Equations* (Algoritmo)

Novamente, os modelos de forma variável apresentaram os melhores resultados dentre os modelos utilizados, tendo o modelo de Muhairwe (1999) (equação 5) sido superior para as medidas de precisão para a estimativa dos volumes do material genético A, com RMSE de 4,49% e MAB de 0,00547. Com performance inferior aos demais modelos, o modelo segmentado de Max e Burkhart (1976) (equação 7) apresentou RMSE de 4,94% e MAB 0,005798.

Para o material genético B o modelo de Kozak (2004) (equação 6) obteve os melhores resultados para as medidas de precisão dentre os modelos de afilamento utilizados, tendo sido o único modelo que apresentou RMSE inferior a 9% com 8,89% e MAB de 0,00665. Já o modelo de Max e Burkhart (1976) (equação 7) apresentou-se como o pior modelo com RMSE de 9,655% e MAB de 0,00753.

Para as árvores do material genético C, dentre os modelos de afilamento testados o modelo de Kozak (2004) (equação 6) mais uma vez obteve os melhores resultados para as

medidas de RMSE de 5,02%, apresentando também MAB de 0,00545. O modelo de Max e Burkhart (1976) (equação 7) apresentou-se novamente com resultados inferiores aos demais, com RMSE de 5,57% e MAB de 0,00546.

Dentre os três materiais genéticos utilizados, para as medidas de precisão das estimativas de volume, o algoritmo MTE apresentou-se como o melhor método em dois destes materiais (A e C), tendo a segunda melhor performance para o outro (B). Porém, quando analisados os gráficos de dispersão de resíduos observou-se o comportamento superior do algoritmo em relação aos modelos utilizados (Figuras 6, 7 e 8), devido sua dispersão mais concentrada e capacidade de compor os resultados de forma a aproveitar o que cada modelo oferece de melhor.

Nos gráficos de dispersão dos resíduos para as estimativas de volume (Figuras 6, 7 e 8) observa-se comportamentos característicos de heterogeneidade de variância para todos os métodos, entretanto isso se deve ao fato da base de dados consistir no volume por seção acumulado, ou seja, conforme caminha-se no fuste no sentido base-topo o volume da seção é somado ao volume anterior.

Para o caso do material genético A (Figura 6) o algoritmo MTE apresentou resultados superiores aos demais modelos, com dispersão mais concentrada e nenhuma tendenciosidade aparente.

Todos modelos apresentaram subestimativa dos volumes até os 10 m de altura, os modelos de Kozak (2004) e Muhairwe (1999) apresentaram, a partir dos 15 m de altura uma pequena superestimativa dos volumes. Os modelos de Max e Burkhart (1976) e de Schöepfer (1966) apresentaram comportamentos parecidos, superestimando uma pequena porção dos volumes e subestimando outra porção, evidenciando dois seguimentos diferentes da nuvem de resíduos. Esses comportamentos refletiram no comportamento da dispersão dos resíduos do algoritmo, que ao compor seus resultados através da mistura dos modelos, não apresentou tendências (Figura 6).

A dispersão dos resíduos para as estimativas dos volumes para as árvores do material genético B pelos modelos de afilamento e pelo algoritmo TEM (Figura 7). Os gráficos demonstram comportamentos distintos dos demais materiais genéticos para todos os modelos, que apresentaram subestimativa dos volumes até os 3 m de altura e superestimativa acima de 10 m de altura. Os resultados apresentados pelo algoritmo foram mais próximos do modelo de Muhairwe (1999) e Kozak (2004), os quais apresentaram os melhores resultados para as medidas de precisão dessas estimativas (Tabela 5).

Figura 6. Dispersão dos resíduos (m^3) para a estimativa do volume por seção das árvores do material genético A pelos modelos de afilamento do fuste em comparação com a técnica de Inferência Multimodelos pelo algoritmo MTE. A linha pontilhada representa o RMSE de cada modelo

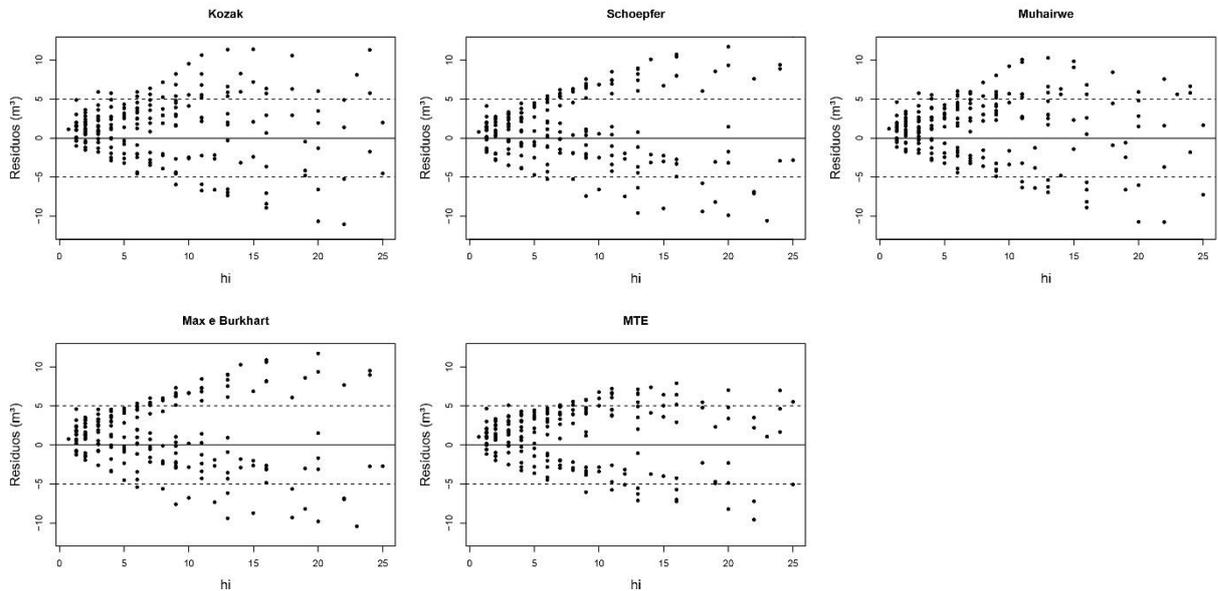
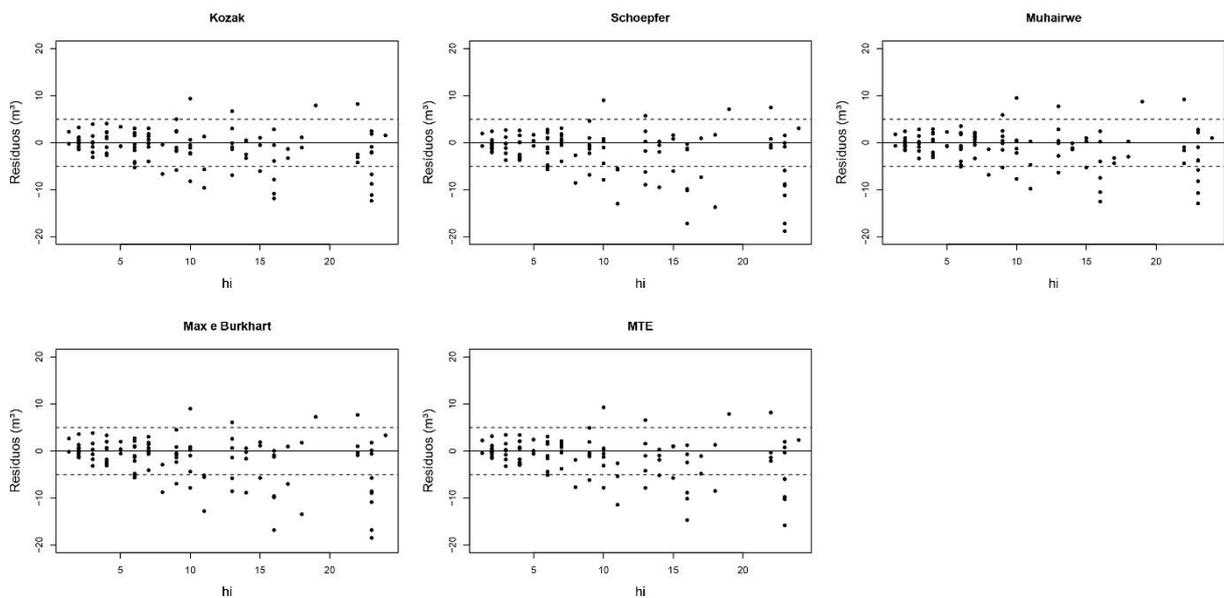


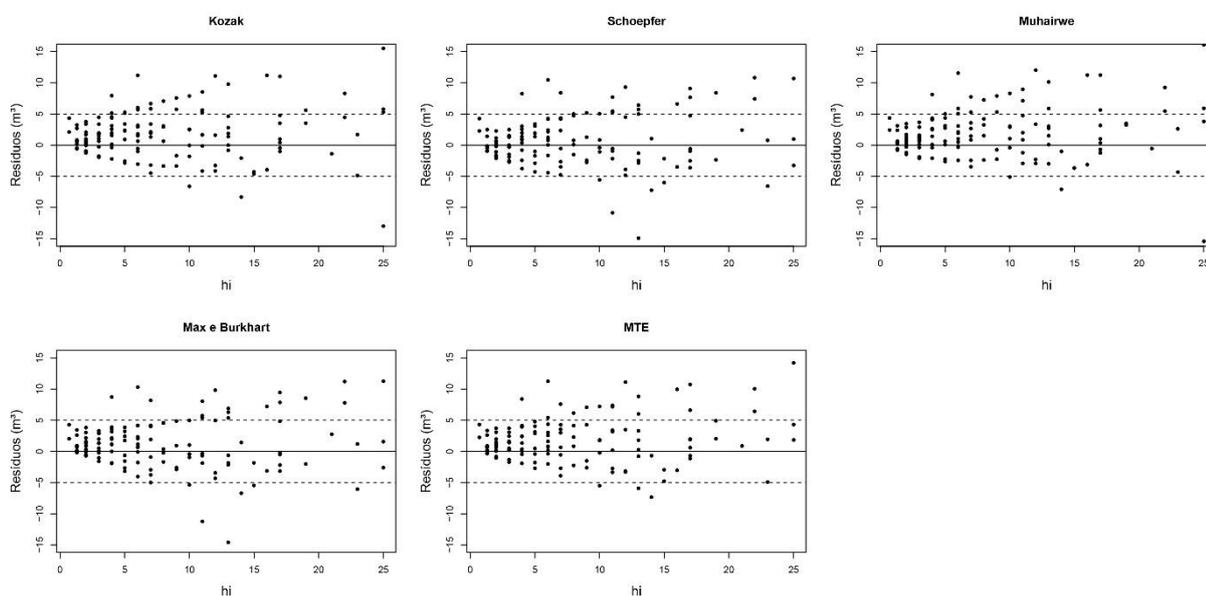
Figura 7. Dispersão dos resíduos (m^3) para a estimativa do volume por seção das árvores do material genético B pelos modelos de afilamento do fuste em comparação com a técnica de Inferência Multimodelos pelo algoritmo MTE. A linha pontilhada representa o RMSE de cada modelo



Para o caso do material genético C (Figura 8) a dispersão dos resíduos para todos os modelos apresentou-se deslocada para resíduos positivos caracterizando subestimativa dos volumes em todas as alturas ao longo do fuste. Entretanto observa-se, assim como nas Figuras 6 e 7, que os pesos dos modelos relativos a cada altura de medição conferem ao algoritmo a característica de não ser fortemente relacionado ao comportamento de nenhum modelo, apresentando assim comportamento distinto devido à mistura dos comportamentos de todos os modelos.

As medidas de precisão apresentam valores das performances de cada modelo e do algoritmo de forma média. Esse tipo de análise acaba por mascarar, por exemplo, erros pontuais mais acentuados devido ao tamanho total da base. Portanto uma boa maneira de constatar a eficiência do algoritmo é através da comparação das árvores isoladamente.

Figura 8. Dispersão dos resíduos (m^3) para a estimativa do volume por seção das árvores do material genético C pelos modelos de afilamento do fuste em comparação com a técnica de Inferência Multimodelos pelo algoritmo MTE. A linha pontilhada representa o RMSE de cada modelo



Considerando-se as medidas de precisão para a estimativa dos volumes (Tabela 5) o algoritmo MTE não obteria o melhor resultado para o material genético B, na média. Porém, ao selecionar aleatoriamente árvores dessa porção da base de dados e analisá-las isoladamente, observou-se que o algoritmo pode não apresentar a melhor estimativa dentre todos os modelos testados para todas as alturas de medição dessas árvores, entretanto, ao comparar o desempenho do algoritmo com cada modelo isoladamente (Figuras 9 a 12), o MTE foi mais preciso na

maioria das estimativas dos diâmetros ao longo do fuste. Essa característica do algoritmo é de extrema utilidade para os inventários florestais, visto que aumenta a acurácia por indivíduo, melhorando as estimativas de quantidade de toras para as estimativas dos múltiplos produtos da madeira, ou mesmo no caso das estimativas de volume de madeira por parcela.

Ao analisar a representação bidimensional da forma do fuste de uma única árvore (Figuras 9 a 12) o algoritmo MTE mostrou que foi capaz de estimar os diâmetros ao longo do fuste de forma mais acurada do que cada um dos modelos testados. As Figuras 9 e 10 comparam as estimativas do algoritmo com os modelos de forma variável, de Kozak (2004) e de Muhairwe (1999), e demonstraram que, mesmo sendo os modelos com as melhores medidas de precisão para todos os materiais genéticos testados, na árvore em questão seus resultados são inferiores até mesmo aos modelos de Max e Burkhart (1976) e Schöepfer (1966) que apresentaram os piores resultados para as medidas de precisão. Analisando ponto a ponto, foi possível constatar que o algoritmo obteve maior precisão do que os modelos de forma variável (Figuras 9 e 10) sendo superior na maioria das alturas de medição.

As Figuras 11 e 12 comparam as estimativas do algoritmo com os modelos de Max e Burkhart (1976) (segmentado) e de Schöepfer (1966) (não segmentado), respectivamente, e são os principais responsáveis pela superioridade do algoritmo em relação aos modelos de forma variável que seriam selecionados pelos métodos de seleção de modelos. Os dois modelos foram superiores ao algoritmo MTE ao estimar os diâmetros de 9 a 13 m de altura, exatamente onde os modelos de forma variável apresentaram maiores erros de estimativa, sendo inferiores ao algoritmo na estimativa dos diâmetros nas demais alturas de medição, regiões onde os modelos de forma variável foram mais assertivos e possibilitaram ao algoritmo estimar essas regiões de forma acurada, fazendo com que o algoritmo fosse superior a todos os modelos de afilamento para as demais alturas de medição, da linha do solo até os 8 m de altura.

Figura 9. Representação bidimensional da forma do fuste de uma árvore (do material genético B) estimada pelo modelo de Muhairwe (1999) sobreposta aos dados reais dos diâmetros coletados através do uso de fotografias digitais segundo Thiersch *et al.* (2020) em comparação com o algoritmo MTE.

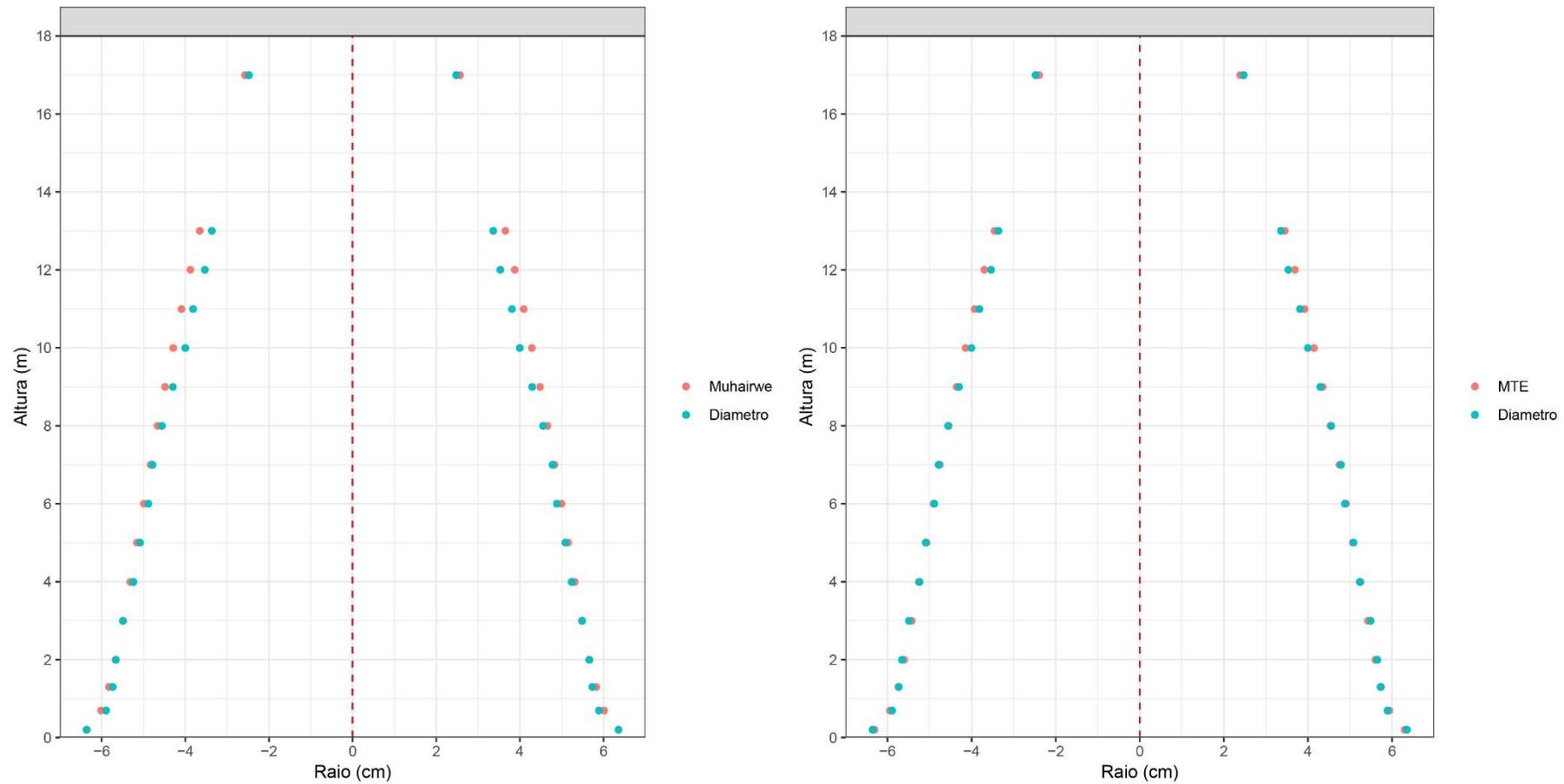


Figura 10. Representação bidimensional da forma do fuste de uma árvore (do material genético B) estimada pelo modelo de Kozak (2004) sobreposta aos dados reais dos diâmetros coletados através do uso de fotografias digitais segundo Thiersch *et al.* (2020) em comparação com o algoritmo MTE.

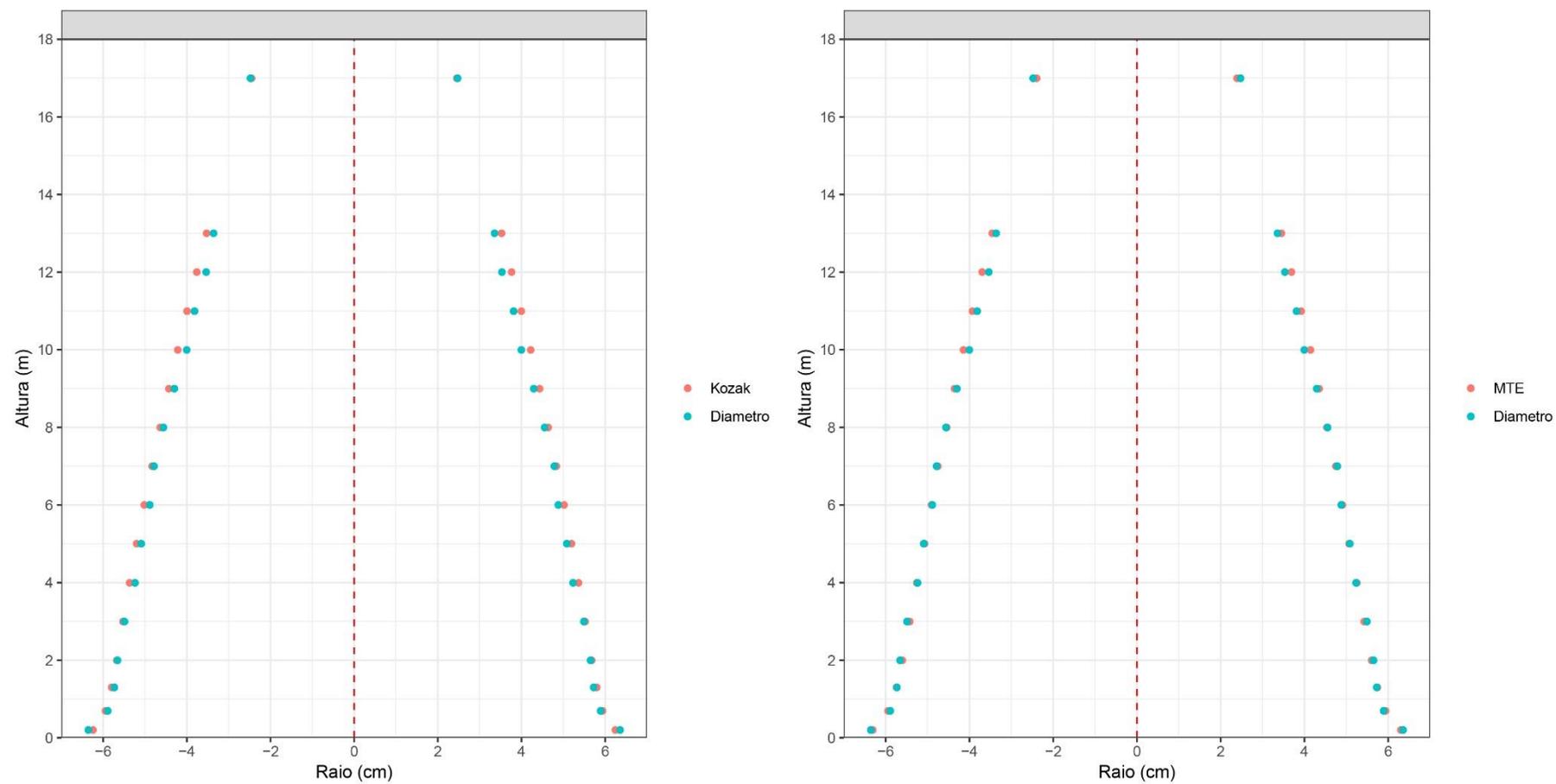


Figura 11. Representação bidimensional da forma do fuste de uma árvore (do material genético B) estimada pelo modelo de Max e Burkhart (1976) sobreposta aos dados reais dos diâmetros coletados através do uso de fotografias digitais segundo Thiersch *et al.* (2020) em comparação com o algoritmo MTE.

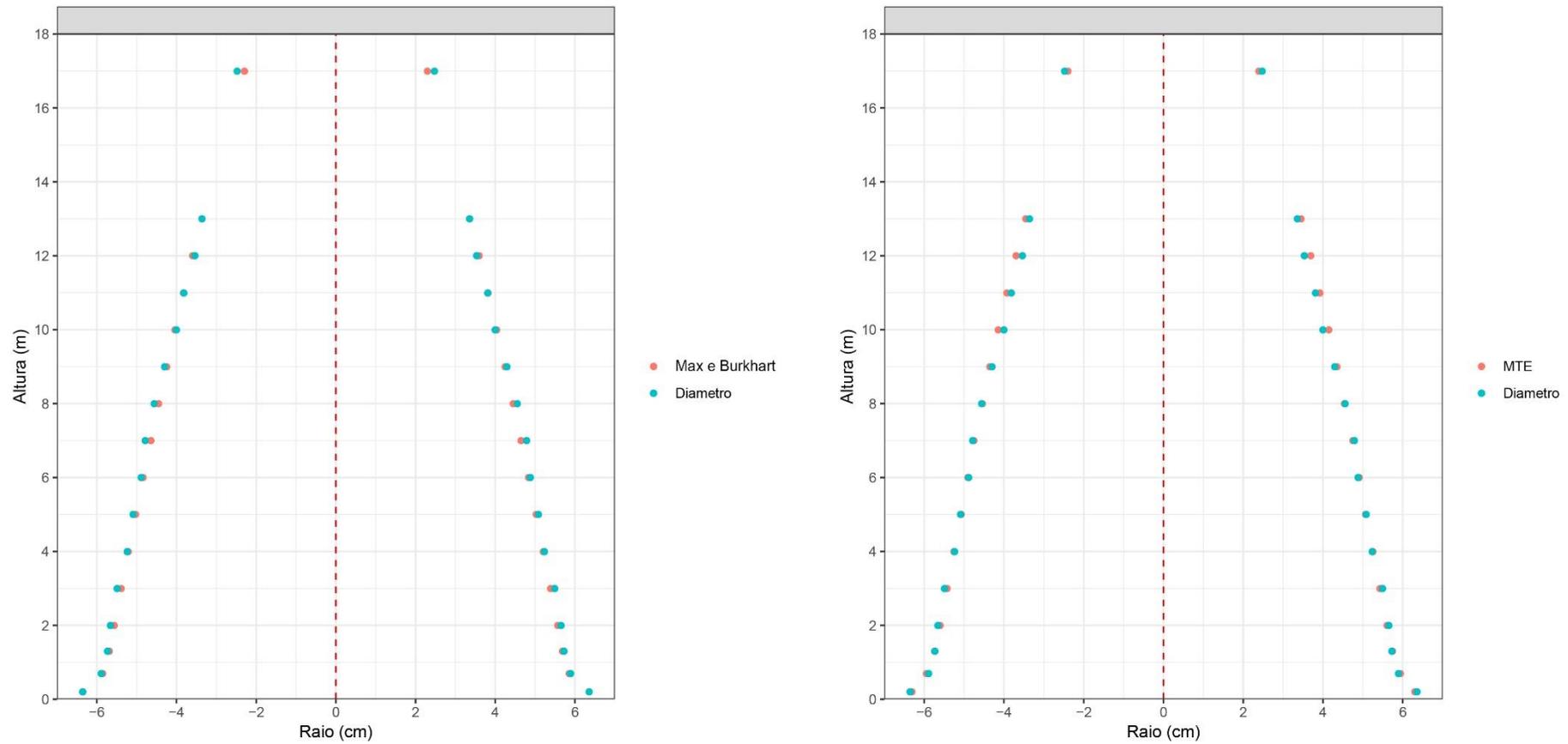
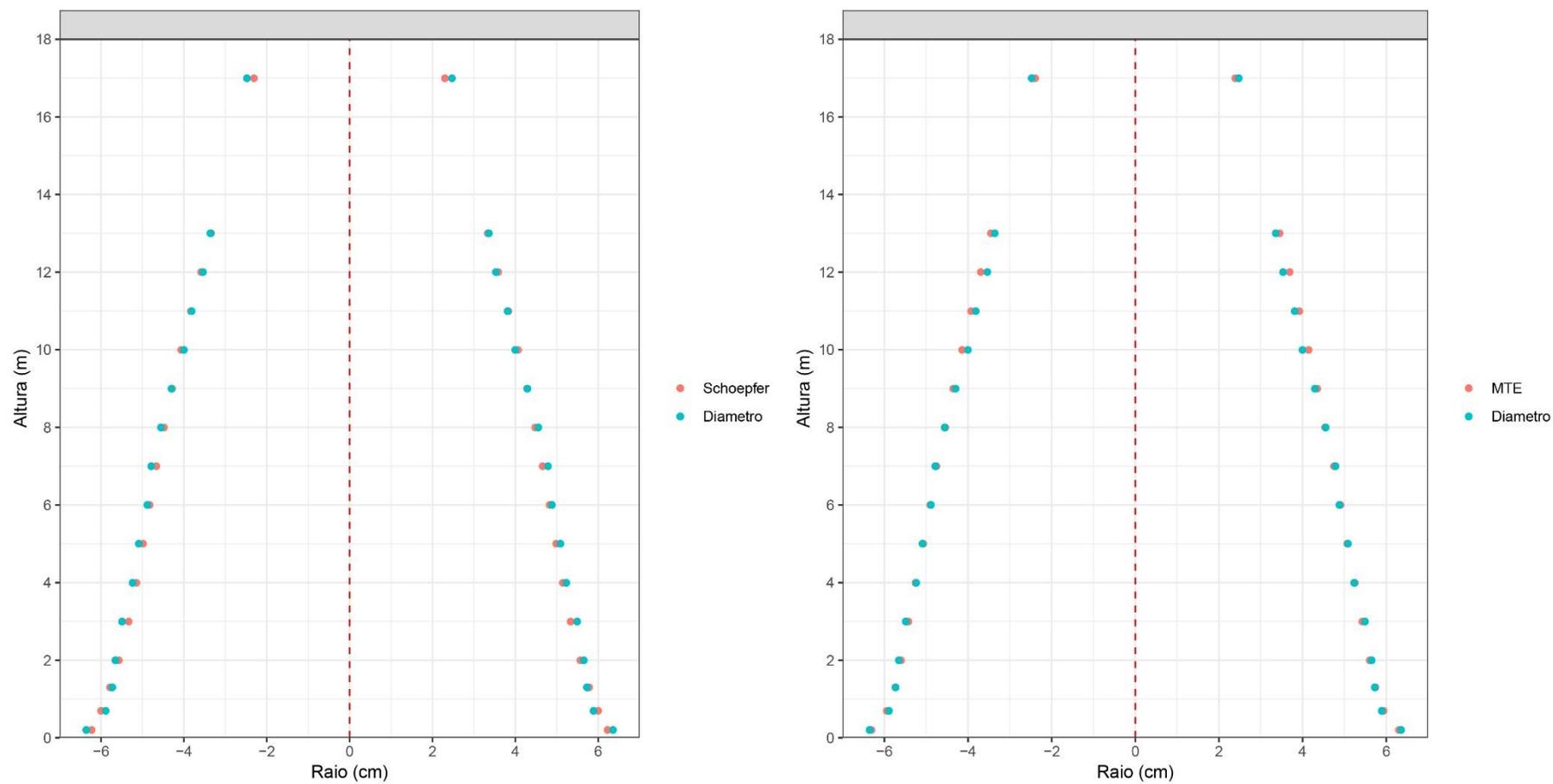


Figura 12. Representação bidimensional da forma do fuste de uma árvore (do material genético B) estimada pelo modelo de Schöpfer (1966) sobreposta aos dados reais dos diâmetros coletados através do uso de fotografias digitais segundo Thiersch *et al.* (2020) em comparação com o algoritmo MTE.



5. CONCLUSÕES

Para as estimativas dos volumes totais das árvores para a Fazenda A e Fazenda B o algoritmo ARM modificado foi superior a todos os modelos volumétricos utilizados.

Constatou-se a eficácia do método que possibilitou obter melhores estimativas pontuais em cada uma das três classes diamétricas definidas pelo algoritmo, como também reduzir a heterogeneidade de variância na dispersão dos resíduos fazendo com que o algoritmo apresentasse resultados superiores ao modelo selecionado pelo RMSE (%).

Para as estimativas de afilamento do fuste de cálculo do volume total por meio das mesmas, analisando-se apenas pelas medidas de precisão, o algoritmo MTE demonstrou ser superior aos modelos de afilamento para os materiais genéticos A e C, porém analisando-se os gráficos de dispersão dos resíduos e as estimativas dos modelos e do algoritmo árvore a árvore, isoladamente, observou-se que o algoritmo realizou o afilamento do fuste de forma mais eficiente do que cada modelo separadamente.

A implementação do algoritmo MTE no *software* R demonstrou que a técnica é viável e de fato conseguiu calcular pesos distintos para os modelos testados ao longo das alturas de medição. Além disso, o método proposto para a estimativa dos volumes se utilizando dos pesos e das próprias equações de afilamento foi capaz de realizar a reconstituição do volume das árvores com resultados árvore a árvore mais acurados do que os modelos individualmente.

REFERÊNCIAS

- AKAIKE, H.; PETROV, B.N.; CSAKI, F. 1998. Information theory and an extension of the maximum likelihood principle, 199–213. IN: PARZEN, E.; TANABE, K.; KITAGAWA, G. eds. **Selected Papers of Hirotugu Akaike**. Springer, New York.
- ASSIS, A L.; SCOLFORO, J. R. S.; MELLO, J. M.; OLIVEIRA, A. D. Avaliação de modelos polinomiais não-segmentados na estimativa de diâmetros e volumes comerciais de *Pinus taeda*. **Ciência Florestal**, v. 12, n. 1, p. 89-107, 2002.
- BEHRE, C. E. Preliminary notes on studies of tree form. **Journal of Forestry**, v. 21, p. 507-511, 1923.
- BENBRAHIM, M.; GAVALAND, A. A new stem taper function for short-rotation poplar. **Scandinavian Journal of Forest Research**, v. 18, p. 377-383, 2003.
- BERHE, L.; ARNOLDSSON, G. Tree taper models for *Cupressus lusitanica* plantations in Ethiopia. **Southern Forests: a Journal of Forest Science**, v. 70 n. 3, p. 193–203, 2008.
- BI, H. Trigonometric variable-form taper equations for Australian Eucalyptus. **Forest Science**, Washington, v. 46, n. 3, p. 397-407, 2000.
- BOZDOGAN, H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. **Psychometrika**, v. 52, p. 345–370, 1987.
- BREIMAN, L. Bagging Predictors. **Machine Learning**, v. 24, n. 2, p. 123-140, 1996.
- BRUCE, D.; CURTIS, R. O.; VANCOEVERING, C. Development of a system of taper and volume tables for Red Alder. **Forest Science**, Washington, v. 14, n. 3, p. 339-350, 1968.
- BUCKLAND, S.; BURNHAM, K.; AUGUSTIN, N. Model Selection: An Integral Part of Inference. **Biometrics**, v. 53, p. 603-618, 1997.
- BURNHAM, K. P.; ANDERSON, D. R. Model selection and Multimodel Inference: A practical Information-Theoretic Approach. **Springer-Verlag**, New York, 2ª Ed., 515 p., 2002.
- CAO, Q. V.; BURKHART, H. E.; MAX, T. A. Evaluating of two methods for cubic-volume prediction of Loblolly Pine to any merchantable limit. **Forest Science**, Washington, v. 26, n. 1, p. 71-80, 1980.
- CARVALHO, S. P. C.; RODRIGUEZ, L. C. E.; CALEGARIO, N.; SAVIAN, T. V.; LIMA, M. P.; SILVA, C. A.; MENDONÇA, A. R.; NICOLETTI, M. F. 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.
- CASTEDO-DORADO, F.; ÁLVAREZ-GONZÁLEZ, J.G. Construcción de una tarifa de cubicación con clasificación de productos para *Pinus radiata* D. Don en Galicia basada en una función de perfil del tronco. **Investigación Agrária: Sistemas y Recursos Forestales**, Madrid, v. 9, n. 2, p. 253–268, 2000.

CLARK III, A.; SOUTER, R. A.; SCHLAEGEL, B. E. Stem profile equations for southern tree species (Research Paper SE-282). **USDA Forest Service**, 113 p., 1991.

COELHO, M. C. B. **Epidometria de Calophyllum brasiliense Camb. em plantios comerciais**. Tese (Doutorado em Ciências Florestais), 100 p., UnB, Brasília-DF, 2016.

DAL BELLO, L. H. A. **Modelagem em Experimentos Mistura-Processo para Otimização de Processos Industriais**. Tese (Doutorado em Engenharia Industrial), PUC-Rio, Rio de Janeiro, 115 p., 2010.

DAVISON, A. C. Biometrika Centenary: Theory and General Methodology. **Biometrika**, 88, pp. 13-52, 2001.

DEMAERSCHALK, J. P.; KOZAK, A. The whole-bole system: a conditional dual-equation system for precise prediction of tree profiles. **Canadian Journal of Forest Research**, Ottawa, v. 7, n. 3, p. 488-497, 1977.

DIÉGUEZ-ARANDA, U.; CASTELO-DORADO, F. ÁLVAREZ-GONZÁLEZ, J. G.; ROJO, A. Compatible taper function for Scots pine plantations in northwestern Spain. **Canadian Journal of Forestry Research**, v. 36, p. 1190-1205, 2006

DING, J.; TAROKH, V.; YANG, Y. "Bridging AIC and BIC: A new criterion for autoregression," **IEEE Transactions on Information Theory**, v. 64, n. 6, p. 4024-4043, 2018.

DRAPER, D. Assessment and propagation of model uncertainty. **Journal of the Royal Statistical Society Series B**, v. 57, p. 45-97, 1995.

EMILIANO, P. C. **Fundamentos e aplicação dos Critérios de Informação: Akaike e Bayesiano**. Dissertação (Mestrado em Estatística e Experimentação Agropecuária), 92 p., UFLA, Lavras-MG, 2009.

EMILIANO, P. C. **Critérios de Informação: como eles se comportam em diferentes modelos?** Tese (Doutorado em Estatística e Experimentação Agropecuária), 193 p., UFLA, Lavras-MG, 2013.

GEORGE, E. I.; MCCULLOCH, R. E. Approaches for Bayesian variable selection. **Statistica Sinica**, v. 7, p. 339-373, 1997.

GOULDING, C. J.; MURRAY, J. C. Polynomial taper equations that are compatible with tree volume equations. **New Zealand Journal of Forestry Science**, Tiergartenstr, v. 5, n. 3, p. 313-322, 1976.

GOLDING, D. L.; HALL, O. F. A Test of precision of cubic-foot tree volume equation on Aspen, Jack Pine and White Spruce. **The Forestry Chronicle**, Ottawa, v. 37, n. 2, p. 123-132, 1961.

- GROSENBAUGH, L. R. Tree Form: definition, interpolation, extrapolation. **The Forestry Chronicle**, v. 42, n. 4, p. 444-457, 1966
- HEIDARSSON, L.; PUKKALA, T. Taper functions for lodgepole pine (*Pinus contorta*) and Siberian larch (*Larix sibirica*) in Iceland. **Icelandic Agricultural Sciences**, v. 24, n. 1, p. 3–11, 2011.
- HORLE, D. C.; MENDONÇA, A. R.; CARVALHO, S. P. C.; CALEGARIO, N. Modelagem não linear do perfil longitudinal de fustes de *Pinus oocarpa*. **Cerne**, Lavras, v.16, n.2, p.177-184, 2010.
- HU, X.; MADDEN, L. V.; EDWARDS, S.; XU, X. Combining models is more likely to give better predictions than single models. **Analytical and Theoretical Plant Pathology**, v. 105, p. 1174-1182, 2015.
- HURVICH, C.M.; TSAI, C.L. Regression and time series model selection in small samples. **Biometrika**, v. 76, p. 297-307, 1989.
- HUSCH, B.; MILLER, C. I.; BEERS, T. W. **Forest mensuration**. 3 ed, New York: Ronald Press Company, 402 p., 1982.
- HUSCH, B.; BEERS, T. W.; KERSHAW JR., J. A. **Forest mensuration**. 4. ed. Hoboken, New Jersey: John Wiley & Sons, 443 p., 2003.
- IBÁ – INDÚSTRIA BRASILEIRA DE ÁRVORES. **Relatório IBÁ 2019**. São Paulo. 2019. Disponível em: <shorturl.at/cfOT1>. Acesso em: 10 jan. de 2020.
- JUDITSKY, A.; NEMIROVSKI, A. Functional aggregation for nonparametric estimation. **The Annals of Statistics**, v. 28, p. 681-712, 2000.
- KOZAK, A. A variable exponent taper equation. **Canadian Journal of Forest Research**, Ottawa, v.18, n.11, p.1363-1368, 1988.
- KOZAK, A. Effects of multicollinearity and autocorrelation on the variable-exponent taper functions. **Canadian Journal of Forest Research**, Ottawa, v. 27, n. 5, p. 619-629, 1997.
- KOZAK, A. My last words on taper equations. **Forestry Chronicle**, Mattawa, v. 80, n. 4, p. 507-515, 2004.
- KOZAK, A.; MUNRO D. D.; SMITH J. H G. Taper functions and their application in forest inventory. **Forestry Chronicle**, Ottawa, v.45, n.4, p.278-283, 1969.
- LI, R.; WEISKITTEL, A. R. Comparison of model forma for estimating stem taper and volume in the primary conifer species of the North American Acadian Region. **Annals of Forest Science**, v. 67, p. 302–317, 2010.
- LINK, W. A.; BARKER, R. J. Model Weights and the foundations of Multimodel Inference. **The Ecology Society of America**, v. 87, n. 10, p. 2626-2635, 2006.

LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. **SAS System for Mixed Models**. Statistical Analysis System Institute, 633 p., 2002.

LOETSCH, F.; ZÖHRER, F.; HALLER, K. E. **Forest Inventory Reinbek: Federal Research Organization for Forestry and Forest Products**, Forest Inventory Section, 1973. 469p. v. 2.

MACHADO, S. A.; CONCEIÇÃO, M. B.; FIGUEIREDO FILHO, A. Modelagem do volume individual para diferentes idades e regimes de desbaste em plantações de Pinus oocarpa. **Revista Ciências Exatas e Naturais**, Guarapuava, v. 4, n. 2, p. 41-50, 2002.

MATE, R.; JOHANSSON, T.; SITO, A. Stem volume equations for valuable timber species in Mozambique. **Journal of Sustainable Forestry**, v. 34, p. 787-806, 2015

MAX, T. A.; BURKHART, H. E. Segmented polynomial regression applied to taper equations. **Forest Science**, Washington, v. 22, n. 3, p. 283-289, 1976.

MIGUEL, E. P.; LEAL, F. A.; ONO, H. A.; LEAL, U. A. S. Modelagem na predição do volume individual em plantio de *Eucalyptus urograndis*. **Revista Brasileira de Biometria**, v. 32, n. 4, p. 584-598, 2014.

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.

MUHAIWE, C. K. **Examination and modelling of tree form and taper over time for interior Lodgepole Pine**. 192f. Thesis (Doctor of Philosophy in The Faculty of Graduate studies Forest Resources Management) - University of British Columbia, British, 1993.

MUHAIWE, C. K. Taper equations for *Eucalyptus pilularis* and *Eucalyptus grandis* for the north coast in New South Wales, Australia. **Forest Ecology and Management**, Amsterdã, v. 113, n. 2-3, p. 251-269, 1999.

MÜLLER, I.; FINGER, C. A. G.; SCHNEIDER, P. R. Forma de tronco e sortimentos de madeira de *Eucalyptus grandis* Hill ex Maiden, na região sudeste do estado do Rio Grande do Sul. **Ciência Florestal**, Santa Maria, v.15, n.3, p.293-305, 2005.

NEWHAM, R. M. A variable-form taper function. **Forestry Canada**, Petawawa National Forestry Institute, Chalk River, Information Report PI-X-083, 31 p., 1988.

NOVO, N.; ROJO, A.; ÁLVAREZ-GONZÁLEZ, J. G. Funciones de perfil del tronco y tarifas de cubicación con clasificación de productos para *Pinus sylvestris* L. en Galicia. **Investigación Agrária: Sistemas y Recursos Forestales**, Madrid, v. 12, n. 1, p. 123–136, 2003.

NUNES, M. H.; GÖRGENS, E. B. Artificial Intelligence Procedures for tree taper estimation within a complex vegetation mosaic in Brazil. **PLoS ONE**, v. 11, n. 5, 16 p., 2016.

ÖZÇELİK, R.; CRECENTE-CAMPO, F. Stem Taper Equations for Estimating Merchantable Volume of Lebanon Cedar Trees in the Taurus Mountains, Southern Turkey. **Forest Science**, v. 62, n. 1, p. 78-91, 2015.

PARRESOL, B. R.; HOTVEDT, J. E.; CAO, Q. V. A volume and taper prediction system for bald cypress. **Canadian Journal of Forest Research**, Ottawa, v.17, n.3, p.250-259, 1987.

PAULA NETO F., BRANDI R. M., RIBEIRO J. C. Teste de aplicação de tabelas volumétricas para estimar a produção e plantações de *Eucalyptus paniculata* Sm. na região de Ipatinga, Minas Gerais. **Revista Árvore**, Viçosa, v. 2, n. 1, p. 154-166, 1977.

PELISSARI, A. L.; LANSSANOVA, L. R.; DRESCHER, R. Modelos volumétricos para Pinus tropicais, em povoamento homogêneo, no Estado de Rondônia. **Pesquisa Florestal Brasileira**, Colombo, v. 31, n. 67, p. 173-181, 2011.

POUDEL, K.P.; CAO, Q.V. Evaluation of methods to predict Weibull parameters for characterizing diameter distributions. **Forest Science**, v. 59, p. 243–252, 2013.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL: <https://www.R-project.org/>

ROCHA, K. J. **Afilamento e otimização de sortimentos de *Tectona grandis* Linn. f. para a maximização do volume e da renda**. Tese (Doutorado em Engenharia Florestal), Universidade Federal de Santa Maria, UFSM, Santa Maria-RS, 207 p., 2018.

ROJO, A.; PERALES, X.; SÁNCHEZ-RODRÍGUEZ, F.; GONZÁLEZ-ALVAREZ J. G.; Von GADOW, K. Stem taper functions for maritime pine (*Pinus pinaster* Ait.) in Galicia (Northwestern Spain). **European Journal of Forest Research**, Dordrecht, v. 124, n. 3, p. 177-186, 2005.

SCHÖEPFER, W. **Automatisierung des Massen, Sorten und Wertberechnung stenender Waldbestände Schriftenreihe Bad**. [S.l.]: Wurt-Forstl. Não paginado. 1966.

SCHRÖDER, T.; HOFIÇO, N. A. S.; ZIMMERMANN, A. P. L.; PEREIRA, L. D.; ROCHA JÚNIOR, D. S.; MEYER, E. A.; FLEIG, F. D. Métodos de estimativa de volume comercial para *Eucalyptus grandis*: especificidades e recomendações. **Pesquisa Florestal Brasileira**, Colombo, v. 33, n. 73, p. 01- 07, 2013.

SCHRÖDER, T.; PEREIRA, L. D.; HOFIÇO, N. A. S. Comparação de métodos de estimativa de volume total para *Eucalyptus grandis* W. Hill ex Maiden. **Floresta e Ambiente**, Seropédica, v. 20, n. 4, p. 480-486, 2013.

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

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

SILVA, N. A. M.; LANA, A. M. Q.; SILVA, F. F.; BERGMANN, J. A. G.; SILVA, M. A.; TORAL, F. L. B. Seleção e classificação multivariada de modelos de crescimento não lineares

para bovinos Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 63, n. 2, p. 364-371, 2011.

SOUZA, C. A. M. **Avaliação de modelos de Taper não-segmentados e segmentados na estimação de altura e volume comercial de fustes de *Eucalyptus* sp.** 94 p. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal do Espírito Santo – Espírito Santo. 2007.

SPURR, S.H. **Forest inventory**. New York: The Ronald Co., 1952.

THIERSCH, C. R.; BERNARDI, L. K.; SANTOS, C. J.; VASCONCELOS, S. L.; THIERSCH, M. F. B. M. Uso de fotografias digitais como alternativa à fita e à suta na cubagem rigorosa. **Scientia Forestalis**, Piracicaba, (in press) 2020.

THOMAS, C.; ANDRADE, C. M.; SCHNEIDER, P. R.; FINGER, C. A. G. Comparação de equações volumétricas ajustadas com dados de cubagem e análise de tronco. **Ciência Florestal**, Santa Maria, v. 16, n. 3, p. 319-327, 2006.

WOLFINGER, R. Covariance structure selection in general mixed models. **Communications in Statistics - Simulation and Computation**, v. 22, n. 4, p. 1079-1106, 1993.

WOLPERT, D. H. Stacked generalization. **Neural Networks**, v. 5, n. 2, p. 241-259, 1992.

YAMAOKA, K.; NAKAGAWA, T.; UNO, T. Statistical moments in pharmacokinetics. **Journal of Pharmacokinetics and Biopharmaceutics**, v. 6, n. 6, p. 547-558, 1978.

YANG, Y. Adaptive regression by mixing. **Journal of the American Statistical Association**, v. 96, n. 454, p. 574-588, 2001.

YANG, Y. Regression with multiple candidate models: selecting or mixing? **Statistica Sinica**, v. 13, p. 783-809, 2003.

YANG, Y. Aggregating regression procedures to improve performance. **Bernoulli**, v. 10, n. 1, p. 25-47, 2004.