



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

Lucas Kröhling Bernardi¹, Monica Fabiana Bento Moreira Thiersch¹, Maria Inês Corrêa de Paula Santos¹, Nelson Yoshihiro Nakajima², Cláudio Roberto Thiersch¹

1 Universidade Federal de São Carlos, Sorocaba, SP, Brasil. E-mail: bernardilucas93@gmail.com; monicathiersch@ufscar.br; santosmaria@estudante.ufscar.br; crthiersch@ufscar.br

2 Universidade Federal do Paraná, Curitiba, PR, Brasil. E-mail: nelsonynakajima@gmail.com

Autora correspondente: Monica Fabiana Bento Moreira Thiersch. E-mail: monicathiersch@ufscar.br

RESUMO

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 à subestimativas ou superestimativas, dependendo do modelo em questão e de suas possíveis tendenciosidades. Estudos recentes sugerem a realização da chamada Inferência Multimodelos, que consiste na obtenção do resultado por meio 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. Para uma base de dados clonais de *Eucalyptus* sp., o algoritmo ARM_m foi superior à seleção de modelos e estimou o volume das árvores de forma mais acurada, apresentando RMSE de 3,84% enquanto o modelo selecionado obteve 3,96%, e apresentou melhor dispersão dos resíduos em comparação a todos os modelos.

Palavras-chave: Biometria; mistura de modelos de regressão; técnicas de modelagem

MULTIMODEL INFERENCE IN VOLUMETRIC PREDICTION IN STANDS OF *Eucalyptus* sp.

ABSTRACT

When selecting a model to estimate a variable, the estimation is conditioned on every characteristic of that model, which can lead to underestimation or overestimation, depending on the specific model and its potential biases. Recent studies suggest the use of a technique called Multimodel Inference, which involves obtaining the result through the contribution of two or more mathematical models. This study aimed to propose a new methodology for Multimodel Inference, based on modifying the ARM (*Adaptive Regression by Mixing*) algorithm, applied to forest inventory for estimating wood volume using volumetric equations. For a clonal *Eucalyptus* sp. dataset, the ARM_m algorithm outperformed model selection and provided more accurate tree volume estimation, with an RMSE of 3.84% compared to 3.96% achieved by the selected model. It also showed better dispersion of residuals compared to all other models.

Key words: Biometrics; mixing regression models; modelling techniques

ASSOCIAÇÃO BRASILEIRA DE MENSURAÇÃO FLORESTAL



INTRODUÇÃO

Na biometria florestal o volume é uma variável de grande importância e, portanto, é imprescindível que seja estimado de forma acurada garantindo a continuidade e escala no fornecimento de matéria-prima. Ele é caracterizado por não ser facilmente mensurável de maneira direta. Dessa forma, é necessário aplicar técnicas de modelagem para a realização de estimativas destes volumes. Em geral são testadas várias equações volumétricas para se encontrar o volume e por meio de técnicas de seleção de modelos, é escolhida aquela que será utilizada para as estimativas.

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 (Yang, 2001; Hu *et al.*, 2015), 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 o qual 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 de árvores de *Eucalyptus* sp. por meio do algoritmo ARM para a realização de uma Inferência Multimodelos.

MATERIAL E MÉTODOS

Os dados para ajuste das equações volumétricas foram coletados em um plantio clonal de híbridos de *Eucalyptus* sp. localizado no sul do estado de São Paulo, com idade aproximada de sete anos. Foram abatidas 60 árvores escolhidas de maneira aleatória e os diâmetros ao longo do tronco 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 (1837).

Algoritmo ARM modificado (ARM_m) 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 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.

VI Encontro Brasileiro de Mensuração Florestal

Existem duas etapas principais envolvidas no algoritmo. Para a primeira etapa, metade da amostra é utilizada para o ajuste dos modelos; 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.

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.

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 Yang (2001) e atualizado pelo autor em 2003.

O desempenho do algoritmo ARM_m foi comparado com os desempenhos dos modelos volumétricos por meio da raiz do erro quadrático médio em porcentagem (RMSE).

Os modelos volumétricos utilizados foram o modelo de Schumacher e Hall logarítmico e modelo de Schumacher & Hall (1933), modelo de Spurr logarítmico e modelo de Spurr (1952), modelo de Takata (1958), modelo de Stoate (1945), modelo de Hush (1963) e modelo de Berkhout (1920).

O ajuste e aplicação dos modelos volumétricos, assim como a implementação do algoritmo ARM_m foram realizados com o *software* estatístico R versão 3.6.2 (R Core Team, 2019).

RESULTADOS E DISCUSSÃO

O ajuste das equações volumétricas e utilização do algoritmo ARM_m possibilitou o cálculo de pesos referente à porção da base de dados que aquele modelo foi capaz de representar.

Os valores obtidos para as medidas de precisão indicaram superioridade do algoritmo ARM_m em estimar os volumes totais por árvore, apresentando o melhor valor dentre todos os métodos testados, com RMSE de 3,8405%, como pode ser visto na Tabela 1.

Tabela 1. Desempenho das equações volumétricas em comparação ao algoritmo ARM

Equação	RMSE (%)	Peso classe 1	Peso classe 2	Peso classe 3
1	4,8409	0,1970	0,1440	0,0580
2	4,9054	0,2606	0,1712	0,0665
3	4,0012	0,0155	0,1894	0,3549
4	5,1052	0,1456	0,1787	0,0395
5	4,7577	0,2331	0,1274	0,0361
6	3,9571	0,00023	0,18142	0,44058
7	11,8289	0,14748	0,00288	0,0034
8	8,1229	0,0000612	0,00551	0,00108
ARM	3,8405	-	-	-

Em que: RMSE = raiz do erro quadrático médio; ARM = *adaptive regression by mixing* (Algoritmo); Equação 1: Schumacher & Hall (1933) logarítmica; Equação 2: Spurr (1952) logarítmica; Equação 3: Schumacher & Hall (1933); Equação 4: Spurr (1952); Equação 5: Takata (1958); Equação 6: Stoate (1945); Equação 7: Hush (1963); Equação 8: Berkhout (1920).

O algoritmo ARM_m para o cálculo de volume das árvores através das equações de volume permitiu o cálculo de pesos para cada modelo em cada classe diamétrica, demonstrando as

VI Encontro Brasileiro de Mensuração Florestal

diferentes eficácias de cada um dos modelos. Deve-se observar que, conforme muda-se a classe diamétrica os pesos médios de cada modelo de afilamento variam. Para a amostra utilizada, as equações 2 e 5 obtiveram os maiores pesos para a classe diamétrica 1, mesmo estes não sendo os modelos com os menores RMSE dentre os testados; nenhuma equação se sobressaiu na classe 2; entretanto na classe 3 as equações 3 e 6 se sobressaíram com os maiores pesos, representando somadas aproximadamente 79,5% do peso total. O modelo selecionado pelo RMSE, a equação 6, obteve o segundo menor peso para a classe 1 e o maior peso para a classe diamétrica 3 (Tabela 1).

Além do RMSE, observou-se os resíduos dos modelos ajustados, uma vez que 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 ou que apresentam menores erros de estimativas são mais confiáveis, pois, no geral, conferem resultados mais próximos à realidade em número maior de situações.

A dispersão dos resíduos (Figura 1), demonstrou melhor precisão no cálculo dos volumes pelo algoritmo ARM_m em comparação com os modelos testados. Volume (m^3)

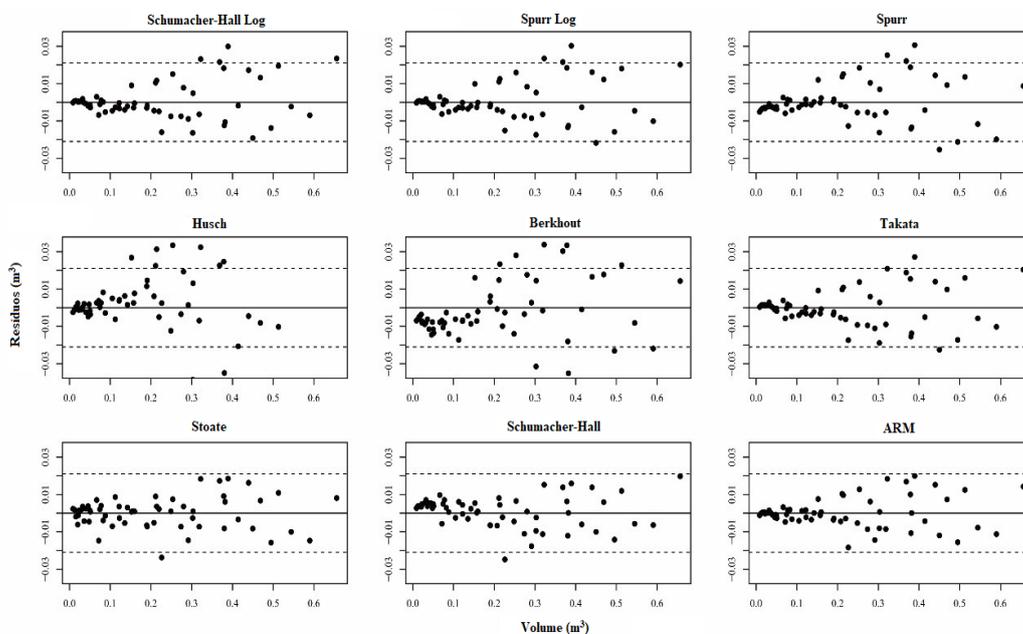


Figura 1. Dispersão dos resíduos (m^3) para a estimativa do volume total das árvores da Fazenda A pelos modelos volumétricos em comparação com a técnica de Inferência Multimodelos (ARM_m). A linha pontilhada representa um RMSE de 10,0%.

Os melhores modelos segundo o RMSE apresentaram subestimativas dos menores volumes, essas subestimativas não são observadas na dispersão dos resíduos do algoritmo. Além disso, observa-se pela Figura 1 que o algoritmo ARM_m apresentou dispersão mais concentrada dos resíduos em relação ao modelo selecionado pelo RMSE (Tabela 1 - Equação 6).

O modelo de Stoaite (1945) (Equação 6), selecionado pelo RMSE, subestimou a maioria das árvores com até $0,1 m^3$ e apresentou dispersão menos concentrada dos resíduos para as árvores de até $0,2 m^3$, em comparação com o algoritmo ARM_m . Quanto aos demais modelos testados, todos foram inferiores ao algoritmo ARM_m . Porém, seus comportamentos distintos em relação ao

VI Encontro Brasileiro de Mensuração Florestal

modelo de Stoate (1945), principalmente para os menores volumes, possibilitou ao algoritmo a estimativa mais acurada ponto a ponto para estes volumes, que apresentou dispersão dos resíduos mais concentrada do que de todos os modelos testados, sem apresentar superestimativa ou subestimativa dos volumes em nenhuma porção da base de dados.

Aplicando a técnica da inferência multimodelos em dois cenários do contexto florestal, utilizando-se do algoritmo ARM para a estimativa de volume e de uma adaptação do algoritmo para o estudo do sortimento, denominando-o de algoritmo MTE - *Mixed Taper Equations* (Equações de afilamento mixadas), Bernardi (2020) observou que o algoritmo ARM foi superior a todos os modelos volumétricos, aumentando a precisão das estimativas de volume para todas as classes de diâmetro. E para o estudo do sortimento o algoritmo de inferência multimodelos MTE também apresentou os melhores resultados para as estimativas tanto de diâmetro ao longo do fuste, bem como de volume total por meio da integração das equações. Ainda para Bernardi (2020), os resultados dos algoritmos de inferência multimodelos foram superiores tanto a nível da amostra como um todo (valores médios) como para os valores por árvore, no caso das equações de afilamento do fuste.

Segundo Burnham & Anderson (2004) os procedimentos teóricos que permitem a inferência multimodelos são simples, tanto em termos de compreensão quanto de cálculo, e, quando utilizados corretamente, fornecem inferências com boas propriedades (por exemplo, em relação ao erro quadrático médio preditivo e à cobertura do intervalo de confiança obtido). Para os autores, o viés e a incerteza na seleção de modelos são questões importantes que merecem uma compreensão mais aprofundada e a Inferência multimodelos é um campo novo do conhecimento, que surge como alternativa à subjetividade da seleção de modelos, no qual ainda é necessário maior entendimento, bem como pesquisas adicionais inovadoras.

CONCLUSÃO

O algoritmo ARM_m foi superior à seleção de modelos apresentando menor valor de RMSE em relação ao modelo selecionado para a Fazenda A e ainda dispersão de resíduos superior a todos os modelos volumétricos testados em todas as classes diamétricas definidas, não apresentando subestimativas ou superestimativas dos volumes.

A implementação do algoritmo ARM_m no *software* R possibilitou obter melhores estimativas pontuais em cada uma das três classes diamétricas definidas pelo algoritmo, realizando uma inferência multimodelos para calcular o volume das árvores de forma mais acurada do que os modelos individualmente. Além disso o algoritmo possibilitou melhorar a dispersão dos resíduos apresentando resultados superiores a todos os modelos testados, inclusive o modelo selecionado pelo RMSE (%).

VI Encontro Brasileiro de Mensuração Florestal

REFERÊNCIAS BIBLIOGRÁFICAS

- Berkhout, A. H. Het meten der boomen in verband met hun aanwas (mit deutscher Zusammenfassung). **Mededelingen van de Landbouwhogeschool Wageningen**, v. 17, n. 1, p.109-225, 1920.
- Bernardi, L. K. **Inferência Multimodelos na Predição de multiprodutos em Povoamentos de *Eucalyptus* sp.** 2020. 65f. Dissertação (Mestrado em Planejamento e Uso de Recursos Renováveis) – Universidade Federal de São Carlos, 2020. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/12384>. Acesso em: 10 Jul. 2023.
- Burnham, K. P.; Anderson, D. R. Understanding AIC and BIC model selection. **Sociological Methods & Research**, v. 33, n. 2, p.261-304, 2004. <https://doi.org/10.1177/0049124104268644>.
- 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, n. 9, p.1174-1182, 2015. <https://doi.org/10.1094/PHYTO-11-14-0315-R>.
- Husch, B. **Forest mensuration and statistics**. New York: The Ronald Press, 1963. 474p.
- R Core Team. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2019.
- Schumacher, F. X.; Hall, F. S. Logarithmic expression of timber-tree volume. **Journal of Agricultural Research**, v. 47, n. 9, p.719-734, 1933.
- Spurr, S. H. **Forest inventory**. New York: The Ronald Co., 1952. 476p.
- Stoate, I. N. The use of a volume equation in pine stands. **Australian Forestry**, v. 9, n.1, p.48-52, 1945. <https://doi.org/10.1080/00049158.1945.10675220>.
- Takata, K. Construction of universal diameter-height curves. **Journal of Japanese Forest Society**, v. 40, n.1, p.1-6, 1958.
- Yang, Y. Adaptive regression by mixing. **Journal of the American Statistical Association**, v. 96, n. 454, p.574-588, 2001. <https://doi.org/10.1198/016214501753168262>.
- Yang, Y. Regression with multiple candidate models: selecting or mixing? **Statistica Sinica**, v. 13, n.1, p. 783-809, 2003. Disponível em: <https://www3.stat.sinica.edu.tw/statistica/oldpdf/A13n313.pdf>. Acesso em: 10 Jul. 2023.