

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CÂMPUS DE CHAPADÃO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

TAYS SILVA BATISTA

**REDES NEURAIS ARTIFICIAIS PARA QUANTIFICAR O VOLUME
DE MADEIRA EM ESPÉCIES DE EUCALIPTO, COM AMPLIAÇÃO
TEÓRICA DA BASE DE DADOS**

CHAPADÃO DO SUL – MS

2020

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CÂMPUS DE CHAPADÃO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

TAYS SILVA BATISTA

**REDES NEURAIAS ARTIFICIAIS PARA QUANTIFICAR O VOLUME
DE MADEIRA EM ESPÉCIES DE EUCALIPTO, COM AMPLIAÇÃO
TEORICA DA BASE DE DADOS**

Orientador: Prof. Dr. Paulo Eduardo Teodoro

Co-Orientadores: Prof. Dr. Gileno Brito de Azevedo Profa.

Dra. Larissa Pereira Ribeiro Teodoro

Defesa apresentada à Universidade Federal de Mato Grosso do Sul, como requisito para obtenção do título de Mestre em Agronomia, área de concentração: Produção Vegetal.

CHAPADÃO DO SUL – MS

2020

Certificado de aprovação_Tays PDF



Serviço Público Federal
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul



PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

CERTIFICADO DE APROVAÇÃO

DISCENTE: Tays Silva Batista

ORIENTADOR: Paulo Eduardo Teodoro

Redes neurais artificiais para quantificar volume de madeira em espécies de eucalipto, com ampliação teórica da base de dados

Prof(a) Dr (a) Paulo Eduardo Teodoro

Presidente

Prof(a) Dr (a) Glauce Tais de Oliveira Sousa Azevedo

Prof(a) Dr (a) Deborah Nava Soratto

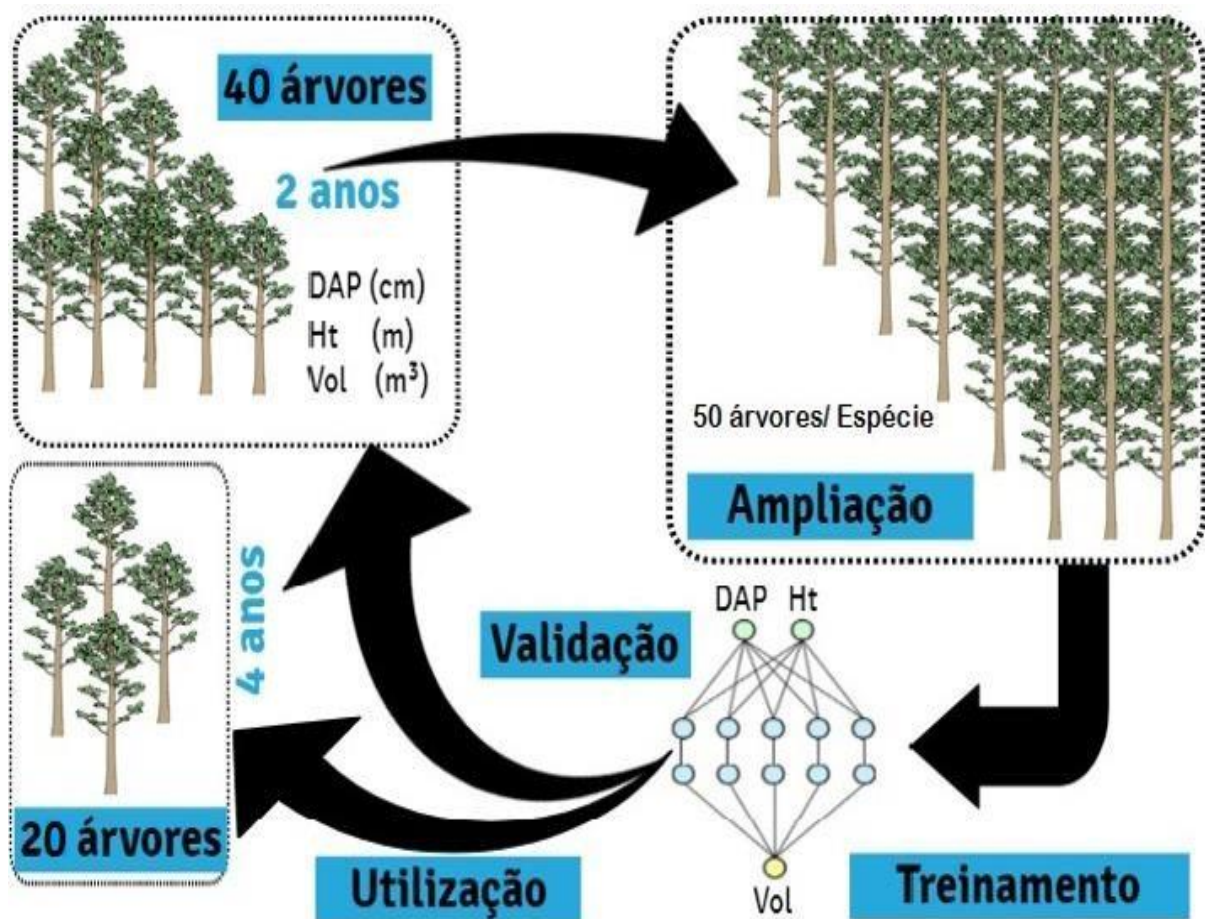
Chapadão do Sul, 27 de fevereiro de 2020.

LISTA DE TABELAS

1. Análise de variância para as variáveis volume, diâmetro a altura do peito e altura total avaliadas aos dois (V1,DAP1 e H1) e quatro anos e meio (V2, DAP2, H2) em espécies de eucalipto.....14
2. Estimativa dos parâmetros estatísticos dos dados originais e simulados para as variáveis volume (V), diâmetro a altura do peito (DAP) e altura total (H) aos dois anos de idade em espécies de eucalipto.....15
3. Parâmetros da equação do modelo Shumacher e Hall para estimativa do volume (V) em espécies de eucalipto, em função do diâmetro a altura do peito (DAP) e altura total (H).....15
4. Coeficiente de correlação de Pearson (r) e raiz do erro quadrático médio (RSME) dos ajustes do modelo de Shumacher e Hall para estimativa do volume (V) em espécies de eucalipto em diferentes etapas.....16
5. Coeficiente de correlação de Pearson (r) e raiz do erro quadrático médio (RSME) e arquiteturas das 10 melhores RNA selecionadas.....17
6. Coeficiente de correlação de Pearson (r) e raiz do erro quadrático médio (RSME) da rede de base radial (RBF) com sete e doze neurônios nas camadas para cada espécie de eucalipto.....18

REDES NEURAIS ARTIFICIAIS PARA QUANTIFICAR O VOLUME DE
MADEIRA EM ESPÉCIES DE EUCALIPTO, COM AMPLIAÇÃO TEÓRICA
DA BASE DE DADOS

RESUMO GRÁFICO



RESUMO

O volume (V) é a variável que melhor representa a produtividade das espécies florestais e vários modelos de regressão são utilizados na sua estimativa. Entretanto, a rede neural artificial (RNA) é reconhecida por sua precisão e capacidade de generalização. A acurácia das redes está associada a qualidade e quantidade de dados no treinamento e validação. O teorema de BoxMüller, gera variáveis aleatórias a partir de dados originais, e fornece um conjunto de dados consistente. Deste modo, essa pesquisa teve como objetivos (i) avaliar a eficiência do teorema de Box-Müller para ampliação de base dados de cubagem de árvores-amostra de eucalipto e usar diferentes topologias de RNA para prever o V de diferentes espécies de eucalipto e (ii) comparar as estimativas com aqueles obtidos ao empregar o modelo de Schumacher e Hall. O delineamento usado foi em blocos ao acaso com quatro repetições, composto pelos seguintes tratamentos: *Corymbia citriodora* e diferentes espécies de *Eucalyptus* spp. Aos dois e quatro anos e meio de idade foi realizado a cubagem de árvores-amostra. E o volume estimado obtido através do modelo de regressão não-linear Shumacher e Hall para cada espécie e comparada com as RNA através das estatísticas: correlação de Pearson (r) e raiz do erro quadrático médio (RSME) nas etapas: treinamento, validação e utilização, com 50, 20 e 4 árvores por espécie, respectivamente. Nas RNA foram testados dois tipos de arquiteturas (*Multi-Layer Perceptron* – MLP e *Radial Basis Function* – RBF). O uso do teorema de Box-Müller foi eficiente na ampliação do número de árvores-amostras na cubagem, para estimativa de volume, e pode ser usado quando houver um número insuficiente de dados. A RNA selecionada foi do tipo RBF, com 7 e 12 neurônios nas camadas ocultas. Comparada ao modelo de Shumacher e Hall, a rede foi superior em termos de r e RSME. Ao prever o volume aos dois anos. Os resultados foram satisfatórios para todas as espécies e para idade futura em certas espécies, com r de 90%.

Palavras-chave: equações de volume; função de base radial; mensuração florestal.

ABSTRACT

Volume (V) is the variable that best represents productivity and several regression models are used in its estimation. However, artificial neural networks (ANN's) are recognized for their precision and generalizability. The accuracy of the networks is associated with the quality and quantity of data in the training and validation. The Box-Müller theorem generates random variables from original data, and provides a consistent data set. Thus, this research aimed to evaluate the efficiency of the Box-Müller theorem to expand the database of eucalyptus sample trees and use different ANN's topologies to predict the V of different eucalyptus species and compare the estimates with those obtained when using the Shumacher and Hall model. The design used was a randomized block with four replications, composed of the following treatments: *Corymbia citriodora* and different species of eucalyptus. At two and four and a half years of age, sample trees were cubed. At two and four and a half years of age, sample trees were cubed. And the estimated volume obtained using the Shumacher and Hall non-linear regression model for each species and compared with the ANN's using Pearson's correlation (r) and root of the mean square error (RSME) in the stages: training, validation and use, with 50, 20 and 4 trees per species. In the ANN's, two types of architectures were tested (Multi-Layer Perceptron – MPL or Radial Basis Function – RBF). The use of the Box-Müller theorem was efficient in increasing the number of sample trees in the cubage, for volume estimation, and can be used when there is an insufficient number of data. The selected ANN was of the RBF type, with 7 and 12 neurons in the hidden layers. Compared to the Shumacher e Hall model, the network was superior in terms of r and RSME. When predicting volume at two years, the results were satisfactory for all species and for future age in certain species, with 90% r.

Keywords: volume equations; radial base function; forest measurement.

SUMÁRIO

1. INTRODUÇÃO.....	09
2. MATERIAL E MÉTODOS.....	10
3. RESULTADOS E DISCUSSÃO.....	14
4. CONCLUSÃO.....	19
5. REFERÊNCIAS.....	20

1 INTRODUÇÃO

O Brasil é o segundo maior produtor de celulose do mundo e líder mundial na produção de aço a partir de carvão vegetal. A atividade florestal é responsável por 6,1% do PIB industrial no país (IBÁ, 2018). Os estados Minas Gerais e Paraná são os que mais contribuem para o crescimento das florestas equiâneas no Brasil, onde predominam espécies de eucaliptos e pinus (SNIF, 2018). Os plantios de eucalipto ocupam uma área de 5,7 milhões de hectares e apresentam uma produtividade de $36 \text{ m}^3 \text{ ha}^{-1} \text{ ano}^{-1}$, considerada uma das maiores do mundo (IBÁ, 2019).

O gênero *Eucalyptus* é o mais cultivado no Brasil, devido seu rápido crescimento, adaptabilidade e diversidade de genótipos para diversos usos da madeira. Dentre as principais espécies de eucalipto cultivadas no Brasil, destacam-se o *E. camaldulensis*, *E. saligna*, *E. grandis* e *C. citriodora* (SHANTHI et al., 2015). A escolha da espécie leva em consideração a finalidade do uso e condições fisiográfica, edáfica e climática, pois cada espécie apresenta características específicas e o seu comportamento pode variar com as condições de solo e clima. O *E. saligna* possui adaptabilidade a locais predispostos a geadas (PALUDZYSZYN FILHO; SANTOS, 2013), *C. citriodora* apresenta alta resistência mecânica na madeira (MORAIS et al., 2010). Os híbridos de *Eucalyptus urophylla* x *Eucalyptus grandis*, são os mais utilizados na produção de celulose, devido as características da madeira e seu elevado potencial produtivo (MAKOUANZI et al., 2017).

A avaliação da produtividade de povoamentos é feita através da mensuração da produção volumétrica, a partir do volume das árvores individuais. Existem alguns métodos para obtenção do volume de árvores, dentre eles, a cubagem de árvores-amostra abatidas, por exemplo, pelo método de Smalian (MIRANDA et al., 2013; MÜLLER et al., 2014; LEAL et al., 2015). Esta cubagem fornece dados de DAP (diâmetro tomada a 1,30 m de altura), altura total e volume. Após essa etapa, são ajustados modelos de regressão, linear ou não-linear para predição do volume por área (MIGUEL et al., 2016).

Os dados obtidos são utilizados para ajuste de modelos volumétricos, como de Shumacher e Hall (1933), de afilamento do fuste, como de Garay (1977). O modelo de Shumacher e Hall ainda é o mais utilizado, por apresentar resultados não tendenciosos, devido à suas propriedades estatísticas, embora este modelo seja eficiente em predizer volume, existem diversas formas de fuste, que torna essencial ajustá-lo para cada espécie, clone, idade, espaçamento e regime de corte.

O uso de redes neurais artificiais (RNA) é uma importante opção para estimar o volume de árvores, por apresentar uma capacidade de aprendizagem através de exemplos (THOMAS et al., 2006). As RNA são sistemas paralelos distribuídos, compostos por unidades de processamento simples (neurônios artificiais). Essas unidades são dispostas em camadas interligadas que armazenam o conhecimento experimental (aprendizagem), e generaliza a informação aprendida e possuem capacidade de auto-organização e de processamento temporal (BRAGA et al., 2000; KAMILARIS; PRENAFETA-BOLDÚ., 2018).

Diversos autores já comprovaram a eficiência das RNA na estimação do volume de árvores (BINOTI et al., 2014; MIGUEL et al., 2016) altura total (VENDRUSCULO et al., 2015; SILVA et al., 2016; VENDRUSCULO et al., 2017), modelagem de crescimento e produção em nível de povoamento, árvores individuais e por classe de diâmetro (BINOTI et al., 2012; BINOTI et al., 2013) e quantificação da casca (DIAMANTOPOULOU, 2005). De acordo com esses autores uma das principais vantagens das RNA é a possibilidade de incluir variáveis categóricas como a espécie e ajustar uma única rede para diferentes espécies.

Mesmo com estas técnicas mais precisas, a cubagem de árvores-amostra pode resultar em custo expressivo, devido ao grande número de árvores a serem mensuradas. Porém, o avanço nas técnicas estatísticas possibilita a ampliação de dados amostrados, mantendo os parâmetros dos dados originais como a média, variância e covariância. Essa técnica denominada teorema de Box-Müller melhora a acurácia no treinamento de RNA e ajuste de equação (SANTOS et al., 2019).

Este estudo teve como objetivos: (i) Avaliar a eficiência do teorema de Box-Müller para ampliação de base dados de cubagem de árvores-amostra de eucalipto. (ii) usar diferentes topologias de RNA para prever volume (V) de diferentes espécies de eucalipto e comparar as estimativas com aqueles obtidos ao empregar o modelo de Schumacher e Hall.

2. MATERIAL E MÉTODOS

Condução do experimento

O experimento foi instalado em janeiro de 2014 na área experimental da Universidade Federal do Mato Grosso do Sul, campus Chapadão do Sul. A altitude é de 820 m. O solo é classificado como latossolo vermelho de textura média. Segundo o sistema de classificação de Köppen, o clima é tropical úmido (Aw) com estação chuvosa de outubro a abril e estação seca entre maio e setembro. A precipitação média varia de 750 a 1.800 mm/ano e a temperatura média anual varia de 20 a 25°C (PEEL et al., 2007).

Todos os requisitos de fertilização foram determinados a partir de análises químicas do solo. Os seguintes resultados foram obtidos: pH (CaCl₂): 4,9; matéria orgânica: 31,5 g dm⁻³; fósforo: 13,6 mg dm⁻³; hidrogênio + alumínio (H + Al): 5,4; potássio: 0,29 cmolc dm⁻³; cálcio: 2,8 cmolc dm⁻³; magnésio: 0,5 cmolc dm⁻³; capacidade de troca catiônica (CTC): 9,0 cmolc dm⁻³; saturação por base: 39,9 %. As proporções de argila, areia e silte foram de 46%, 46% e 8%, respectivamente. Coroamento, capina, controle de formigas e aplicação de herbicidas (glifosato) foram realizados quando necessário.

O delineamento experimental foi o de blocos ao acaso com quatro repetições, com 20 plantas dentro de cada parcela experimental. Os tratamentos foram compostos por quatro espécies de eucalipto (*E. camaldulensis*, *C. citriodora*, *E. saligna*, *E. grandis* e *E. urograndis*).

Variáveis avaliadas

Para cubagem foram medidos com suta os diâmetros nas posições 0,3; 1,30; 2,0 metros, após essa altura, a cada dois metros até altura comercial. As árvores foram abatidas aos dois e quatro anos de idade, com uma e duas árvores por espécie respectivamente, e cubadas pelo método de Smalian. Os volumes das seções foram somados para obter o volume total (equação 1).

$$V = \sum_{i=1}^{n-1} \frac{g_i + g_{(i+1)}}{2} * L \quad (1)$$

em que: V é o volume individual da árvore; g_i é área transversal na posição (altura) i do fuste; L = comprimento da secção; i = posição no fuste, sendo $i = 1$ a base do fuste.

Análises estatísticas

Para verificar a existência de diferença de crescimento entre as espécies de eucalipto, foram realizadas análises de variância para cada variável avaliada em cada medição, de acordo com o modelo estatístico (equação 2).

$$Y_{ij} = \mu + B_j + E_i + \varepsilon_{ij} \quad (2)$$

em que: Y_{ij} é a observação no j -ésimo bloco, avaliada na i -ésima espécie; μ é a média geral dos experimentos; B_j é o efeito fixo do bloco j dentro do ambiente j ; E_i é o efeito da i -ésima espécie de eucalipto considerado como fixo; ε_{ij} é o erro aleatório associado a observação Y_{ij} .

Após verificar diferenças entre as espécies para todas as variáveis, foi realizado o procedimento de ampliação dos dados da primeira medição, visando obter equações e redes com boa capacidade de generalização. O processo de ampliação consistiu em gerar, a partir dos dados reais, uma variável aleatória com distribuição normal com uma matriz de zeros representando a média e variância equivalente a uma matriz sigma ($X \sim N(\phi, \Sigma)$) por meio da transformação linear $X = F'Y$, em que $Y \sim N(\phi, I)$ representava a variável aleatória correspondente aos dados originais e F é obtido pela decomposição espectral de Σ , tal que $\Sigma^{-1} = FF'$. Para ampliar os dados originais, novos valores de Y foram gerados considerando $Y \sim N(\phi, (F')^{-1}X)$. Para isso, utilizou-se o teorema de Box-Müller, em que as variáveis volume (V), diâmetro a altura do peito (DAP) e altura total (H) têm distribuição normal conforme equação 3, segue a mesma para cada variável. Essas análises foram realizadas no software Genes (CRUZ, 2013).

$$x = \sqrt{-2 \ln(RND)} V \cos(2\pi RND) \quad (3)$$

em que x representa a variável a ser utilizada e RND é um número aleatório positivo para cada variável, conservando as médias, variâncias e covariâncias entre as variáveis para cada medição.

Posteriormente, o processo de ajuste dos modelos não-lineares e RNA's ocorreu em três etapas: treinamento, validação e utilização. No treinamento, foram utilizados os 50 valores por espécie gerados pela ampliação; na validação foram utilizadas as oito árvores por espécie da primeira medição; na utilização foram usadas quatro árvores por espécie da segunda medição. Para estimar os volumes foram medidas com fita métrica a circunferência das árvores à 1,30 m do solo e altura total (metros) com hipsômetro Haglof. O ajuste da equação (4) foi feito pelo modelo de Schumacher e Hall, com a função “nls” do pacote nlme do software R (R DEVELOPMENT CORE TEAM, 2015),

$$V = \beta_0 DAP^{\beta_1} H^{\beta_2} + \varepsilon \quad (4)$$

Na construção e treinamento das RNA, foi utilizada a ferramenta *Intelligent Problem Solver*, do STATISTICA 7.0. Essa ferramenta realiza a mineração dos dados, ou seja, normaliza os dados no intervalo entre 0-1, testa diferentes arquiteturas (*Multi-Layer Perceptron* – MLP ou *Radial Basis Function* – RBF) de redes e seleciona as redes que melhor representam os dados.

Nas redes MLP, a camada de entrada foi composta pelas variáveis contínuas DAP, H e a variável categórica espécie; a camada de saída com a variável V. Para a(s) camada(s) intermediária(s) foi utilizada a função de ativação ($f(x) = \text{logística}$ (Equação 5), aplicada a cada neurônio, que utiliza como argumento o produto escalar do vetor de entrada (x) e do vetor de peso (w) associado a esse nodo. Para essas redes foram testadas topologias contendo duas e três camadas contendo entre um e 15 neurônios cada com alto grau de conectividade entre os neurônios, a qual é definida pelos pesos sinápticos.

$$f(x) = \frac{1}{1+e^x} \quad (5)$$

onde: x é um valor binário que representa a ativação do neurônio (1) *versus* não-ativação (0).

Nas redes RBF, a função de ativação tem como argumento a distância entre seus vetores de entrada e de peso, e a arquitetura baseia-se em uma camada de entrada (DAP e H como variáveis contínuas e espécie como variável categórica), uma camada de saída (V) e duas camadas intermediárias. Para composição cada camada intermediária foi testada entre um e 15 neurônios. O treinamento de ambos os tipos de rede, MLP e RBF, é do tipo *feedforward* e ocorre pelo método supervisionado. Portanto, foram testadas 3825 topologias de RNA, composta pela seguinte combinação: MLP duas camadas intermediárias (15x15 possibilidades); MLP três camadas intermediárias (15x15x15 possibilidades) e RBF duas camadas intermediárias (15x15 possibilidades). Foram salvas apenas as 10 RNA com maior correlação linear entre os volumes observados *versus* estimados na etapa de treinamento.

Para comparação do modelo de Schumacher e Hall (1933) e as 10 melhores RNA testadas foram utilizadas para cada etapa (treinamento, validação e utilização) as estatísticas: correlação de Pearson (r - Equação 6) e raiz do erro quadrático médio (RMSE - Equação 7).

$$r = \frac{COV_{(xy)}}{\sqrt{\hat{\sigma}_x^2 \times \hat{\sigma}_y^2}} \quad (6)$$

em que: $COV_{(xy)}$ é a covariância entre os valores observados (X) e estimados (Y); $\hat{\sigma}_x^2$ é a variância dos valores observados; $\hat{\sigma}_y^2$ é a variância dos valores estimados.

$$RMSE(\%) = \frac{100}{\bar{Y}} \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (7)$$

Em que: $\bar{Y}_i$ é a média dos volumes observados; n é o número total de observações.

3. RESULTADOS E DISCUSSÃO

Houve diferença significativa entre as espécies de eucalipto para todas variáveis avaliadas, independentemente da idade (Tabela 1). Resultados similares foram observados por outros autores ao avaliarem o crescimento de diferentes espécies de eucalipto no Cerrado Brasileiro (POERSCH et al., 2017; AZEVEDO et al., 2015; QUIQUI et al., 2001; SILVA et al., 1992). Tais resultados demonstram que a ampliação dos dados deve ser realizada por espécie, por existir diferenças significativas entre as variáveis. Além disso, indicam a necessidade de ajustar o modelo de Schumacher e Hall por espécie e que a espécie deve ser uma variável categórica nas camadas de entrada das RNA.

Tabela 1. Análise de variância para as variáveis volume, diâmetro a altura do peito e altura total avaliadas aos dois (V1, DAP1 e H1) e quatro anos e meio (V2, DAP2, H2) em espécies de eucalipto.

QM							
FV	GL	V1	DAP1	H1	V2	DAP2	H2
Blocos	3	0,0002	3,5312	2,0684	0,0111	25,5104	4,6612
Espécies	4	0,004*	13,5821*	16,0354*	0,04086*	33,2314*	51,0942*
Resíduos	12	0,00014	1,0249	1,2489	0,0048	5,8767	5,2549
CV (%)		12,37	7,36	6,77	31,35	14,06	12,19

*: significativo a 5% de probabilidade pelo teste F; FV: fontes de variação; GL: graus de liberdade; CV: coeficiente de variação;

A ampliação dos dados foi eficaz, como visto na Tabela 2, pois manteve os valores dos parâmetros estatísticos idênticos aos dados originais, com uma variação na amplitude pelo aumento dos dados. Recentemente, esse procedimento foi usado por Silva et al. (2014) e Santos et al. (2019), ao ampliar os dados de variáveis em estudos com RNA. Esta técnica pode ser uma alternativa para treinamento e validação de RNA's nos casos em que um número de variáveis

de entrada é relativamente grande em relação ao número de observações disponíveis, ao invés da usual eliminação das variáveis.

Tabela 2. Estimativa dos parâmetros estatísticos dos dados originais e simulados para as variáveis volume (V), diâmetro a altura do peito (DAP) e altura total (H) aos dois anos de idade em espécies de eucalipto.

Espécie	Variável	Amplitude original	Amplitude simulada	Média original	Média simulada	Variância original	Variância simulada
<i>E.camaldulensis</i>	V	0,0496	0,0679	0,0599	0,0599	0,0003	0,0003
	DAP	4,40	5,68	11,55	11,55	2,0514	2,0514
	H	3,62	5,41	14,4755	14,4755	1,2377	1,2377
<i>C.citriodora</i>	V	0,0411	0,0540	0,0623	0,0623	0,0002	0,0002
	DAP	3,70	4,12	11,775	11,775	1,8107	1,8107
	H	2,19	5,12	14,8063	14,8063	0,6802	0,6802
<i>E. urograndis</i>	V	0,0905	0,1278	0,1469	0,1469	0,0008	0,0008
	DAP	3,30	6,17	15,6875	15,6875	1,4527	1,4527
	H	2,81	6,06	19,0913	19,0913	0,9509	0,9509
<i>E. saligna</i>	V	0,0641	0,1075	0,0884	0,0884	0,0004	0,0004
	DAP	4,60	6,22	13,125	13,125	1,8964	1,8964
	H	3,87	9,09	15,6575	15,6575	2,2895	2,2895
<i>E. grandis</i>	V	0,0834	0,1026	0,1134	0,1134	0,0007	0,0007
	DAP	3,50	4,98	15,05	15,05	1,4429	1,4429
	H	3,25	5,82	18,8513	18,8513	1,6116	1,6116

Tabela 3. Parâmetros da equação do modelo Schumacher e Hall para estimativa do volume (V) em espécies de eucalipto, em função do diâmetro a altura do peito (DAP) e altura total (H).

Espécie	Coeficientes		
	β_1	β_2	β_3
<i>E. camaldulensis</i>	-10,07906414	2,33400318	0,56933421
<i>C. citriodora</i>	-12,14591293	1,74451534	1,74451534
<i>E. urograndis</i>	-10,78347437	-0,21028463	3,19760377
<i>E. saligna</i>	-8,41494700	1,75745174	0,52778998
<i>E. grandis</i>	-1,05752476	1,21162283	2,57682462

Tabela 4. Coeficiente de correlação de Pearson (r) e raiz do erro quadrático médio (RSME) dos ajustes do modelo de Shumacher e Hall para estimativa do volume (V) em espécies de eucalipto em diferentes etapas.

Espécie	Treinamento		Validação		Utilização	
	r	RSME (%)	r	RSME (%)	r	RSME (%)
<i>E. camaldulensis</i>	0,9054	12,2347	0,9409	11,6302	0,9959	15,4573
<i>C. citriodora</i>	0,9584	6,6535	0,9671	5,5487	0,9965	12,2124
<i>E. urograndis</i>	0,8078	11,0498	0,8236	10,0103	0,3543	67,1348
<i>E. saligna</i>	0,9796	4,7403	0,9850	4,4114	0,9095	27,7887
<i>E. grandis</i>	0,8470	12,5538	0,8298	13,3990	0,5809	28,5748

Com base na correlação de Pearson (r) e raiz do erro quadrático médio (RSME) (Tabela 3), o modelo de Shumacher e Hall apresentou um bom desempenho para as espécies *E. camaldulensis*, *C. citriodora* e *E. saligna*. Esses resultados demonstram que esse modelo pode ser utilizado com eficácia para predizer o volume nessas espécies.

Entretanto, as estatísticas de ajuste para estimar o volume do *E. urograndis* e do *E. grandis*, foram inferiores as demais espécies avaliadas em idades futuras, contudo, para idade atual os resultados foram satisfatórios, com 82% de correlação e 10% de RSME. A ampliação dos dados contribuiu para melhorar a precisão do modelo, pois, o número de árvores-amostra cubadas é insuficiente para ajuste de equações. Diversos autores comprovam a eficiência do uso do modelo de Shumacher e Hall em diferentes espécies, mesmo que estas apresentem grande diversidade na forma do fuste, este é o modelo que melhor se ajusta para estimar volume de árvores (LANSSANOVA et al., 2018; MARTINS et al., 2015; LEITE et al., 2011).

As dez melhores topologias das RNA ajustadas, estão presentes na Tabela 5. Dentre estas, cinco redes foram tipo função de base radial (RBF) e cinco do tipo perceptron multicamadas (MPL). É possível observar diferenças nos valores de r para as etapas de treinamento, validação e utilização. Verificou-se que as redes MPL apresentaram maior r na etapa de treinamento. Porém esses valores decaíram nas demais etapas. Uma das explicações para esse problema é a ocorrência de overfitting, ou seja, a rede memoriza tudo sobre os dados de treinamento, não apenas as características específicas mais também os erros. Por ocasionar um erro nas previsões, a rede perde a capacidade de generalização utilizada em dados futuros (BULLINARIA, 2014).

Tabela 5. Coeficiente de correlação de Pearson (r) e raiz do erro quadrático médio (RSME) e arquiteturas das 10 melhores RNA selecionadas.

Rede	Topologia	Treinamento		Validação		Utilização	
		r	RQME (%)	r	RQME (%)	r	RQME (%)
RBF	7-6	0,9507	12,98136	0,9585	11,69368	0,8944	12,94833
RBF	7-8	0,9509	12,95434	0,9588	11,65762	0,9032	12,37527
RBF	7-10	0,9515	12,87976	0,9548	12,20110	0,8808	13,78007
RBF	7-12	0,9532	12,65336	0,9604	11,43152	0,9233	10,98758
RBF	7-15	0,9709	10,02741	0,9569	11,91480	0,6903	22,84997
MLP	7-4-1	0,9656	10,88998	0,9648	10,79509	0,8101	17,56292
MLP	7-5-1	0,9659	10,84161	0,9642	10,88055	0,8247	16,84502
MLP	9-4-1	0,9678	10,53846	0,9627	11,10270	0,8148	17,33286
MLP	15-9-1	0,9692	10,30313	0,9479	13,06749	0,7459	20,51454
MLP	15-10-1	0,9947	4,30909	0,8477	21,76497	0,6824	23,17201

Dessa forma, ao avaliar a capacidade de generalização das dez melhores redes nas etapas de validação e utilização, as redes RBF apresentaram resultados mais acurados comparados com as redes MLP. As redes RBF apresentam como principal vantagem a capacidade de se adaptar aos diferentes conjuntos de dados e produzir estimativas, pois possui arquitetura definida por algoritmos de agrupamento, o que torna mais fácil de trabalhar quando comparada as redes MPL (SILVA JÚNIOR et al., 2018). Resultados similares foram observados por Silva et al. (2009), ao compararem diferentes RNA para prever o volume de eucalipto.

As medidas de exatidão das Tabelas 4 e 6 comprovam que equações de volume ajustadas com dados de árvores relativamente pequenas (menor idade) não devem ser aplicadas para estimar volume de árvores em idades superiores em certas espécies, uma vez que, nesse caso, o

viés nas estimativas pode ser grande. Conforme Campos e Leite (2017) é necessário complementar o banco de dados de árvores-amostra anualmente, sempre que surgirem duas ou mais novas classes de diâmetro, assumindo a amplitude de dois centímetros. Desse modo, conforme os autores o banco de dados é ampliado continuamente com o abate de novas árvores até o final da rotação. O viés ocorreu, também para as redes neurais (Tabela 6).

Portanto, os resultados da Tabela 2 comprovaram a eficiência da ampliação de dados pelo método Box-Müller, mas esta eficiência ocorre apenas para a idade e povoamentos contemplados na amostragem. Este resultado apresenta uma importância econômica relevante, pois o custo de um inventário pode ser reduzindo, ao diminuir o número de árvores abatidas.

A rede selecionada foi a que apresentou maior valor de r e menor RSME na etapa de utilização. Essa rede foi a RBF com sete neurônios na primeira camada intermediária e 12 na segunda. Para detalhar o comportamento dessa rede foram mostradas as estimativas de r e RSME por espécie (Tabela 5). De forma geral, os resultados estratificados por espécie mostram que a rede RBF teve comportamento inferior em termos de RSME ao modelo de Schumacher e Hall (1933) para as espécies *E. citriodora*, *C. grandis*. Porém, a rede RBF apresentou resultados melhores para o *E. urograndis*, *E. camaldulensis* e *E. saligna*.

Tabela 6. Coeficiente de correlação de Pearson (r) e raiz do erro quadrático médio (RSME) da rede de base radial (RBF) com sete e doze neurônios nas camadas ocultas para cada espécie de eucalipto.

Espécie	Treinamento		Validação		Utilização	
	r	RQME (%)	r	RQME (%)	r	RQME (%)
<i>E. camaldulensis</i>	0,8242	16,22404	0,8821	12,38780	0,9928	4,915851
<i>C. citriodora</i>	0,9496	7,288481	0,9621	5,895249	0,9728	12,71652
Clone GG100	0,7343	12,70452	0,7517	11,56505	0,8088	33,69449
<i>E. saligna</i>	0,9541	6,945248	0,9597	6,127334	0,9009	12,42702
<i>E. grandis</i>	0,8472	12,52239	0,8474	11,74719	0,6919	28,72384

Silva Jr et al. (2018) testaram diversas topologias de RNA e observaram que em número de amostras reduzidas, nenhum dos algoritmos se destacaram em termos de qualidade de resultados. Todavia, os modelos clássicos, como o de Schumacher e Hall (1933), não foram influenciados pela variação na quantidade de dados de treinamento. David et al. (2016) e Soares et al. (2011) também demonstraram que os modelos de regressão podem superar as RNA em termos de tendenciosidade. A otimização das RNA é obtida por tentativas sucessivas até gerar resultados satisfatórios, pois o número de neurônios por camada e o número de camadas a ser estabelecidos não se baseiam em nenhum critério, é um processo aleatório (SILVA et al., 2009; GORGENS et al., 2014). Dessa forma, embora tenha sido realizada ampliação dos dados, o tamanho deste pode ter contribuído para o melhor desempenho das RNA.

4. CONCLUSÕES

A ampliação de base de dados de árvores-amostras abatidas para cubagem é eficiente, podendo ser empregada em casos de treinamento de redes neurais artificiais quando houver restrições para a cubagem.

O ajuste do modelo de Shumacher e Hall e treinamento das redes aos dois anos foi satisfatório para todas as espécies, com ampliação dos dados, e para maioria das espécies em idade futura.

A exatidão das estimativas obtidas pelas RNA foi maior do que a obtida com regressão não-linear, ao avaliar todas as espécies.

5. REFERÊNCIAS

- AZEVEDO, L. P. A. et al. Seleção em progênies de *Eucalyptus camaldulensis* em área de cerrado mato-grossense. **Ciência Florestal**, n. 11, v. 45, p. 2001-2006, 2015.
- BINOTI, D. H. B. et al. Modelagem da distribuição diamétrica em povoamentos de eucalipto submetidos a desbastes utilizando autômatos celulares. **Revista Árvore**, v. 36, n. 5, p. 931-939, 2012.
- BINOTI, D. H. B. et al. Redução dos custos em inventário de povoamentos equiâneos. **Revista Brasileira de Ciências Agrárias**, v. 8, n. 1, p.125-129, 2013.
- BINOTI, M. L. M. S. et al. Utilização de redes neurais artificiais para a projeção da distribuição diamétrica de povoamento equiâneos. **Revista Árvore**, v. 38, n. 4, p.747-754, 2014.
- BHERING, L. L. et al. Application of neural networks to predict volume in eucalyptus. **Crop Breeding And Applied Biotechnology**, v. 15, n. 3, p.125-131, 2015.
- BRAGA, A. P. O., Ludemir, T. B., Carvalho, A. P. F. **Redes Neurais Artificiais: Teoria e aplicações**. Ed 2. Rio de Janeiro, RJ: Ltc, 2000. 262 p.
- BULLINARIA, J. A. **Indtroduction to neural computation**. <<http://www.cs.bham.ac.uk/~jxb/inc.html>>. Junho, 2014.
- CAMPOS, B. P. F. et al. Descrição do perfil do tronco de árvores em plantios de diferentes espécies por meio de redes neurais artificiais. **Pesquisa Florestal Brasileira**, v. 37, n. 90, p. 99-107, 2017.
- DAVID, H. C. et al. Strategies for stem measurement sampling: A statistical approach of modelling individual tree volume. **Cerne**, v.22, n. 3, p. 249-260, 2016.
- DIAMANTOPOULOU, M. J. Artificial neural networks as alternative tool in pine bark volume estimation. **Computers and Electronics in Agriculture**, v. 48, p. 235-244, 2005.
- GARAY, L. Tropical forest utilization system. VIII. A taper model for entire stem profile including buttressing. Seattle: Coll. Forest Res., Inst. Forest Prod. Univ. Wash. 64 p. (Contrib. 36), 1979.
- GORGENS, E. B. et al. Influência da Arquitetura na estimativa de volume de árvores individuais por meio de redes neurais artificiais. **Revista Árvore**, v.38, n. 2, p. 289-295, 2014.
- IBÁ, **Industria Brasileira de Árvores**. Relatório IBA 2017. São Paulo. 2017.
- IBÁ, **Industria Brasileira de Árvores**. Relatório IBA 2018. São Paulo. 2018.
- IBÁ, **Industria Brasileira de Árvores**. Relatório IBA 2019. São Paulo. 2019.
- KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. **Computers and Electronics in Agriculture**, v. 147, p. 70-90, 2018.

LANSSANOVA, L. R. et al. Comparação entre diferentes métodos para estimativa volumétrica de espécies comerciais da Amazônia. **Biofix Scientific Journal**, v. 3, n. 1, p. 109-115, 2018.

LEAL, A. J. F. et al. Redes neurais artificiais na predição da produtividade de milho e definição de sítios de manejo diferenciado por meio de atributos do solo. **Bragantia**, v. 74, n. 4, p.436 - 444, 2015.

LEITE, D. R. et al. Análise comparativa entre dois procedimentos para estimativa da variável volume em um povoamento de *Eucalyptus urophylla*, no município de Mineiros, Goiás. **Enciclopédia biosfera**, v. 7, n. 13, p. 1684-1700, 2011.

MAKOUANZI, G. et al. Genetic variability of growth and wood chemical properties in a clonal population of *Eucalyptus urophylla* × *Eucalyptus grandis* in the Congo. **Southern Forests: a Journal of Forest Science**, v. 80, n. 2, p.151-158, 2017.

MARTINS, A. P. A. et al. Eficiência de métodos para estimativa do volume total de *Corymbia citriodora*. **Scientia Agraria Paranaensis**, v. 15, n. 2, p. 178-184, 2015.

MIRANDA, D. L. C. et al. Determinação do volume de toras da espécie *Mezilaurus itaúba* (Meissn.) Taub. **Scientia plena**, v. 9, n. 8, p.1-7, 2013.

MORAIS, E. et al. Variação genética, interação genótipo solo e ganhos na seleção em teste de progênies de *Corymbia citriodora* Hook em Luiz Antônio. **Scientia Forestalis**, v. 38, n. 85, p.11-18, 2010.

MÜLLER, M. D. et al. Equações de altura, volume e afilamento para eucalipto e acácia estabelecidos em sistema silvipastoril. **Revista Floresta**, v. 44, n. 3, p.473-484, 2014.

MIGUEL, E. P. et al. Artificial intelligence tools in predicting the volume of trees within a forest stand. **African Jouenal of Agricultural**, v. 11, n. 21, p. 1914-1923, 2016.

PALUDZYSZYN FILHO, E; SANTOS, P. E. T. Escolha de cultivares de eucaliptos em função do ambiente e do uso. **Colombo: Embrapa**, 2013. 11 p.

PEEL, M. C. et al. Update world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, p. 1633-1644, 2007.

POERSCH, N. L. et al. Influence of climate variables in the initial growth of *Corymbia citriodora* and different species of *Eucalyptus*. **Bioscience Journal**, v. 33, n. 6, p. 1452-1464, 2017.

QUIQUI, E. M. D. et al. Avaliação de espécies e procedências de *Eucalyptus* para o Noroeste do Estado do Paraná. **Acta Scientiarum Agronomy**, v. 23, n. 5, p. 1173-1177, 2001. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statical Computing. ISBN 3-900051-07-0, 2015.

SANTOS, I. G. et al. Selection index as *a priori* information for using artificial neural networks to classify alfalfa genotypes. **Genetics and Molecular Research**, v. 18, n. 2, p. 1-11, 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.

SHANTHI, K. et al. Micropropagation of *Eucalyptus camaldulensis* for the production of rejuvenated stock plants for microcuttings propagation and genetic fidelity assessment. **New Forests**, v. 46, p. 357–371, 2015.

SILVA, H. D. et al. Comportamento silvicultural e aptidão para produção de carvão de cinco espécies de *Eucalyptus*, na região dos cerrados de Minas Gerais. **Boletim de Pesquisa Florestal**, n. 24/25, p. 71-78, 1992.

SILVA, M. L. M. et al. Ajuste do modelo de Shumacher e Hall e aplicação de redes neurais artificiais para estimar volume de árvores de eucalipto. **Revista Árvore**, v.33, n. 6, p. 1133-1139, 2009.

SILVA, L. F. et al. Equações hipsométricas, volumétricas e de crescimento para *Khaya ivorensis* plantada em Pirapora. **Floresta e Ambiente**, v. 23, n. 3, p.362-368, 2016.

SILVA, G. N. et al. Neural networks for predicting breeding values and genetic gains. **Scientia Agricola**, v. 71, n. 6, p. 494-498, 2014.

SILVA Jr, E. M. et al. Bee-inspired RBF network for volume estimation of individual trees. **Computers and Electronics in Agriculture**, v. 152, p. 401-408, 2018.

SNIF. **Sistema de Nacional de Informações Florestais**. Cadeia Produtiva, 2018. Disponível em: <www.florestal.gov.br/snif/producao-florestal/cadeia-produtiva>. Acesso em: 2019.

SOARES, F. A. et al. Recursive diameter prediction and volume calculation of eucalyptus trees using multilayer perceptron networks. **Computers and Electronics in Agriculture**, v. 78, p. 19-27, 2011.

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

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

VENDRUSCOLO, D. G. S. et al. Estimativa da altura de árvores de *Tectona grandis* L.f. utilizando regressão e redes neurais artificiais. **Nativa**, v. 5, n. 1, p.52-58, 2017.