

REDES NEURAIS PARA ESTIMATIVA VOLUMÉTRICA DE CLONES DE EUCALYPTUS SPP. NO PÓLO GESSEIRO DO ARARIPE

Syntia Regina Rodrigues de SOUZA¹
José Antônio Aleixo da SILVA²
Tiago Alessandro Espínola FERREIRA³
Oourou Ganni Mariel GUERA²

- RESUMO: O objetivo deste trabalho foi estimar os volumes de clones de *Eucalyptus* spp., no Polo Gesseiro do Araripe, que é responsável por 97% da produção nacional de gesso, empregando a metodologia de Redes Neurais Artificiais (RNAs) comparando-a com os modelos volumétricos de Schumacher e Hall e de Spurr (variável combinada), bem como verificar a eficiência das estimativas usando diferentes tamanhos de amostras e avaliar a contribuição de uma variável categórica na estimativa. Os dados são referentes a um experimento implantado na Estação Experimental do Instituto Agrônomo de Pernambuco, onde foram testados 15 clones de *Eucalyptus* spp. plantados em 2002, e com o corte final em 2009. A qualidade do ajuste dos modelos foi avaliada com base ao coeficiente de determinação ajustado (R^2_{aj}), a raiz quadrada do erro médio percentual (RMSE%), o erro padrão da estimativa (S_{yx} %) e a análise dos gráficos de dispersão dos resíduos. Os resultados obtidos no trabalho demonstraram que todas as modelagens são adequadas. Observou-se que a eficiência dos ajustes depende do tamanho da amostra e da variância. A adição de uma variável categórica na RNA não apresentou diferenças perceptíveis, sendo desnecessária para a estimativa do volume.
- PALAVRAS-CHAVE: Manejo florestal; modelos volumétricos; redes neurais artificiais.

1 Introdução

O local de maior produção de gesso no Brasil, está situado no Nordeste, no estado de Pernambuco, com destaque mundial por suas reservas de gipsita de alta qualidade e de condições favoráveis à exploração. Localiza-se na região do Araripe, fronteira dos estados do Piauí, Ceará e Pernambuco (SOBRINHO *et al.*, 2003) no arranjo produtivo local (APL) denominado de o “Polo Gesseiro do Araripe”, sendo responsável por 97% da produção nacional de gesso. A principal fonte de energia para o processo de calcinação da gipsita, matéria prima para o gesso, que é submetida a altas temperaturas, é a lenha proveniente da Caatinga, vegetação nativa predominante na região (SILVA, 2008/2009).

¹ Universidade Federal Rural de Pernambuco – UFRPE, Rua Teixeira Coimbra, nº 300, São Lourenço da Mata, CEP: 54720-090, PE, Brasil. E-mail: syntia.regina@gmail.com

² Universidade Federal Rural de Pernambuco – UFRPE, Departamento de Ciência Florestal, Rua Manoel de Medeiros, s/n, Dois Irmãos, CEP: 52.171-900 – Recife, PE, Brasil. E-mail: jaaleixo@uol.com.br; gueraforest@gmail.com

³ Universidade Federal Rural de Pernambuco – UFRPE, Departamento de Estatística e Informática, Rua Manoel de Medeiros, s/n, Dois Irmãos, CEP: 52.171-900, Recife, PE, Brasil. E-mail: taef.first@gmail.com

Devido aos altos custos de outras fontes de energia, elevar a produção de gesso implica em aumentar o desmatamento da Caatinga. Uma solução econômica e ambiental é a implantação e o manejo sustentado de povoamento de espécies nativas ou o reflorestamento com espécies florestais exóticas de rápido crescimento.

Diante disto, as fontes energéticas renováveis foram consideradas como opções viáveis e uma das ações realizadas para concretizar essas opções foi o estabelecimento de experimentos com clones de *Eucalyptus* spp. em 2002. Esse gênero foi escolhido pela sua diversidade, seu rápido crescimento, diversas opções de aproveitamento de sua madeira e fácil adaptação a diversas condições ambientais.

Existem inúmeros métodos utilizados para estimar o volume de madeira em povoamentos florestais. Os métodos tradicionais usam principalmente os modelos volumétricos ajustados por meio de análise de regressão linear e não linear. O uso da inteligência artificial por meio de redes neurais artificiais (RNA) na pesquisa florestal tem apresentando resultados satisfatórios e sendo útil na tomada de decisões (LIMA *et al.*, 2014). Há muitos estudos que comprovam a eficiência da aplicação da RNA na área florestal. Binoti *et al.* (2014), concluíram que a estimação do volume de árvores por RNA pode ser feita com configurações diferentes, também concluíram que redes neurais simples podem ser utilizadas com eficiência para estimação do volume de árvores. Leal *et al.* (2015) comparando desempenho de redes neurais com o método de regressão linear na estimativa de volume de *Eucalyptus urophylla* concluíram que os métodos testados foram satisfatórios, porém que as RNAs foram superiores no desempenho. Campos *et al.* (2017) avaliando a capacidade de uma RNA descrever o perfil do fuste de árvores de diferentes gêneros e espécies, em distintas condições de crescimento, concluíram que apenas uma RNA descreve o perfil com semelhante eficiência aos modelos de regressão tradicionais. Martins *et al.* (2017) compararam o desempenho da RNA com funções de afilamento e também avaliaram a contribuição da idade na estimativa, a RNA forneceu a melhor estimativa dentre os modelos avaliados com destaque para a RNA com adição da variável idade.

O objetivo deste trabalho foi estimar o volume de clones de *Eucalyptus* spp. no Polo Gesseiro do Araripe utilizando RNAs e comparar as mesmas com os modelos volumétricos de Schumacher e Hall (SCHUMACHER e HALL, 1933) e Spurr (1952), bem como verificar a eficiência da estimativa usando diferentes tamanhos de amostras e avaliar a contribuição de uma variável categórica na estimativa volumétrica.

2 Material e métodos

A introdução de clones de eucaliptos no Pólo Gesseiro de Araripe, que corresponde à 18% da área total do estado de Pernambuco, iniciou no ano de 2002, com o projeto denominado “Experimentação Florestal no Pólo Gesseiro do Araripe” financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Foram testados 15 clones de *Eucalyptus* spp., doados pela Companhia Suzano, na Estação Experimental do Instituto Agrônomo de Pernambuco (IPA). A referida estação está situada na Chapada do Araripe, município de Araripina, mais especificamente na coordenada geográfica 07°27'37''S e 40°24'36''W e altitude 831m (SILVA, 2008/2009).

Os tratamentos (clones) foram mensurados a cada seis meses, de março de 2002 a setembro de 2009. Foi realizada a cubagem rigorosa de 1.862 árvores, distribuídas por 15 tipos de clones de *Eucalyptus* spp., nas quais as toras foram divididas em seções, medida as

circunferências das extremidades de cada seção e o volume total foi calculado pelo método de Smalian, ou seja, a soma dos volumes seccionados (Tabela 1).

Tabela 1 - Descrição dos tratamentos (tipos de clones) usados no experimento

Tratamento	Descrição do NAC	NAC	%
T01	Híbrido de <i>Eucalyptus tereticornis</i> (cn)	147	8,20%
T02	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)	156	8,70%
T03	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)	137	7,20%
T04	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)	151	8,70%
T05	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)	104	4,70%
T06	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> x <i>E. pellita</i> (pc)	135	6,80%
T07	Híbrido de <i>E. urophylla</i> (cn)	126	7,10%
T08	Híbrido de <i>E. brassiana</i> (cn)	113	6,00%
T09	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)	149	8,10%
T10	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)	148	7,90%
T11	Híbrido de <i>E. urophylla</i> (cn)	105	5,10%
T12	Híbrido de <i>E. brassiana</i> (cn)	69	3,20%
T13	Híbrido de <i>E. brassiana</i> (cn)	121	6,60%
T14	Híbrido de <i>E. urophylla</i> (cn)	134	8,10%
T15	Híbrido de <i>E. brassiana</i> (cn)	67	3,30%
Total		1862	100%

Em que: pc = polinização controlada, cn = cruzamento natural e NAC = número de árvores cubadas.

Para estimar o volume de *Eucalyptus* spp. foram utilizados os modelos Schumacher e Hall (1933) e Spurr (1952), também conhecido como modelos da variável combinada. As variáveis independentes foram a altura total da árvore (Ht), em metros e diâmetro à altura do peito (DAP, 1,30 m do solo), conforme a relação funcional

$$V = f(DAP, Ht). \quad (1)$$

Os parâmetros foram estimados por meio do Método de Mínimos Quadrados Ordinários. Além disto, optou-se por utilizar as equações em suas formas lineares aplicando a transformação logarítmica no modelo Schumacher e Hall, a fim de corrigir eventuais problemas de heterocedasticidade de variância e de normalidade dos erros. O modelo de Spurr é linear nos parâmetros, basta considerarmos a expressão $DAP^2H = X$ e o modelo se resume a um clássico modelo de regressão linear simples.

Para a aplicação das RNAs, foram consideradas as variáveis supracitadas, como entrada. Como funções de ativação das camadas oculta e de saída das RNAs foram testadas as funções: exponencial, identidade, tangente hiperbólica e sigmoide. Foram treinadas 20 redes, sendo avaliado o desempenho das cinco melhores redes para posterior comparação entre os modelos. O critério de seleção das redes se baseou nos índices de treinamento da rede, fornecido pelo *software* na fase de treinamento e no coeficiente de determinação ajustado ($R^2_{aj.}$).

As RNAs treinadas foram do tipo *perceptrons* de múltiplas camadas, conhecidas como MLP, de forma supervisionada. O algoritmo empregado foi a retropropagação do erro (*Error Back propagation*), sendo que a regra de aprendizagem minimiza o erro médio quadrático. Esta forma de aprendizagem ajusta os pesos das conexões entre os neurônios da rede de acordo com o erro e tem por objetivo encontrar um conjunto de pesos e polarizações que reduzam a função erro (HAYKIN, 2001).

A comparação das estimativas das redes neurais foi por meio de dois modelos muito utilizados para estimar o volume de árvores na ciência florestal, são o logarítmico de Schumacher e Hall (1933) (equação 2) e o da variável combinada Spurr (1952) (equação 3).

$$V_i = \beta_0 \text{DAP}_i^{\beta_1} \text{Ht}_i^{\beta_2} \varepsilon_i \quad (2)$$

$$V_i = \beta_0 + \beta_1 \text{DAP}_i^2 \text{Ht}_i + \varepsilon_i \quad (3)$$

Os critérios de avaliação de desempenho das RNAs e dos modelos foram: o coeficiente de determinação ajustado ($R^2_{aj.}$), calculado por meio das expressões 04 e 05; a raiz quadrada do erro médio percentual (RMSE) (equação 6); o erro padrão da estimativa (S_{yx}) (equação 7); a análise visual dos gráficos de dispersão dos resíduos por volume (m^3) e histograma de frequência de resíduos. O teste de Kolmogorov-Smirnov (KS) foi utilizado para descrever o quanto os valores observados e estimados são próximos (valores menores são desejáveis). As RNAs escolhidas foram as que apresentaram maiores coeficientes de correlação e determinação e menores RMSE e S_{yx} , e obtiveram gráficos de dispersão dos resíduos e histograma de resíduos não tendenciosos, com maior ocorrência de observações em torno de zero.

$$R^2 = \frac{SQR}{SQT} \quad (4)$$

$$R^2_{aj} = 1 - (1 - R^2) \left(\frac{n-1}{n-k} \right) \quad (5)$$

Nas equações (4) e (5) k : número de coeficientes da equação e n : número de árvores.

A interpretação do coeficiente de determinação é a proporção de variação total do volume que é explicado pelo modelo. Seu resultado pode ser obtido pela razão da soma de quadrados da regressão (SQR) e a soma de quadrados total (SQT).

$$RMSE(\%) = 100 \sqrt{\left(\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \right)} / \bar{Y} \quad (6)$$

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

Nas equações (6) e (7) temos que Y : volume total observado, \hat{Y} : volume total estimado, \bar{Y} : média dos volumes totais observados e n : número total de árvores.

A análise dos gráficos de dispersão dos resíduos dá uma ideia de quanto às estimativas pontuais estão próximas dos valores observados.

Para verificar a eficiência do uso de RNA quando comparada com modelos tradicionais ajustados ou análise de regressão, usou-se diferentes tamanhos de amostras nas quais foram treinadas as RNAs, considerados dois tipos de treinamento, o primeiro das variáveis supracitadas, e o segundo com a inclusão de uma variável categórica, tipo de clone, como entrada, pois uma característica importante das RNAs, é o fato de não ser necessário ter conhecimentos prévios sobre as relações das variáveis utilizadas no ajuste do modelo ao final foi ajustado o modelo de regressão que melhor se ajustou aos dados, o de Schumacher e Hall.

A metodologia empregada neste objetivo foi a mesma empregada no modelo geral. Para o mesmo, foram utilizados os seguintes tamanhos de amostra sorteadas aleatoriamente entre 1862 árvores (Tabela 2).

Tabela 2 - Definição das amostras e erros amostrais da base de dados do experimento com *Eucalyptus spp*

Amostra	Quantidade	E _A %
Amostra 1	25 árvores	28,27%
Amostra 2	50 árvores	15,44%
Amostra 3	100 árvores	9,65%
Amostra 4	200 árvores	7,57%
Amostra 5	400 árvores	5,44%
Amostra 6	800 árvores	3,80%
Amostra 7	1600 árvores	2,66%
Base Completa	1862 árvores	-

O Erro amostral (E_A %) de cada amostra é dado pela formula abaixo:

$$E_A \% = \left(\frac{t_{\alpha/2} \cdot S_{\bar{Y}}}{\bar{X}} \right) \cdot 100 \quad (8)$$

A avaliação dos desempenhos das redes neurais artificiais e do modelo de Schumacher e Hall considerando os diferentes tamanhos amostrais foi realizada com base nos mesmos critérios utilizados para modelagem da base de dados completa.

3 Resultados e discussão

O perfil métrico da amostra de *Eucalyptus* spp. pode ser observado na Tabela 3, na qual são demonstradas as amplitudes do volume, DAP e Ht. Vale ressaltar que as medidas apresentaram leve desvio comparado a magnitude das observações pontuais em torno da média.

Tabela 3 - Estatísticas descritivas das variáveis contínuas dos indivíduos de *Eucalyptus* spp. utilizados no estudo. Em que: DAP (cm) é o diâmetro a 1,30m do solo e Ht (m) é a altura total da árvore

Estatística Descritiva	Variáveis		
	Volume (m ³)	DAP (cm)	Ht (m)
Mínimo	0,01	4,50	7,29
Máximo	0,38	25,97	21,85
Mediana	0,07	11,40	15,53
Média	0,09	11,97	15,52
Desvio	0,05	2,76	1,91

A Tabela 4 apresenta a estimativa dos modelos propostos para comparação dos resultados na predição de volume com a técnica de redes neurais artificiais.

Tabela 4 - Estatísticas do ajuste dos modelos volumétricos de *Eucalyptus* spp

Modelo	Parâmetro	Estimativa	Erro Padrão	T	p-valor
Schumacher e-Hall	β_0	0,000062	< 0,0000001	<0.001	<0.001
	β_1	1,682530	0,013684	<0.001	<0.001
	β_2	1,094109	0,028670	<0.001	<0.001
Spurr	β_0	0,011243	0,000471	<0.001	<0.001
	β_1	0,000031	< 0,0000001	<0.001	<0.001

Ao estudar os ajustes dos modelos de Schumacher e Hall e Spurr, como mostra a Tabela 4, o p-valor < 0,001 confirma que ambos os modelos podem ser utilizados. Sendo assim, o ajuste pode ser aceito:

$$\hat{V}_i = 0,000062 \times DAP_i^{1,682530} \times Ht_i^{1,094109}, \text{ para o modelo de Schumacher e Hall e}$$

$$\hat{V}_i = 0,011243 + 0,000031 \times DAP^2 Ht_i, \text{ para o modelo de Spurr.}$$

A Tabela 5 e a Figura 1 (gráficos dos resíduos) permitem realizar comparações entre os modelos de regressão e as RNA selecionadas com melhor capacidade em estimar o volume de árvores de eucalipto.

Tabela 5 - Estatísticas de ajuste para o volume de *Eucalyptus* spp. Em que: R²aj. (Coeficiente de determinação ajustado), RMSE (%) (raiz quadrada do erro médio percentual), Syx (%) (erro padrão da estimativa) e KS (teste de Kolmogorv-Smirnov)

Validação	Schumacher e Hall		RNA 1	RNA 2	RNA 3	RNA 4	RNA 5
	Spurr						
R ²	0,94926	0,95542	0,97721	0,97678	0,97709	0,97673	0,97703
R ² aj.	0,94920	0,95537	0,97718	0,97675	0,97706	0,97670	0,97700
RMSE%	3,59506	3,36966	3,30220	3,32110	3,29317	3,32326	3,30950
Sxy (%)	12,26156	11,46964	11,24014	11,30447	11,20939	11,31184	11,24096
KS%	11,10000	10,00000	8,40000	8,60000	8,30000	8,60000	8,40000

A Tabela 6 apresenta as arquiteturas das RNAs de acordo com o número de neurônios das camadas de entrada, intermediária ou oculta e camada de saída de cada rede neural avaliada e as funções de ativação das camadas oculta e de saída. Observa-se que a RNA 1 além de apresentar um bom desempenho foi utilizada uma configuração de neurônios na camada oculta mais simples para aplicação das RNAs.

Tabela 6 - Definição das RNAs

RNAs	Arquitetura	Função de Ativação
RNA 1	MLP 2-5-1	(Tangente Hiperbólica - Identidade)
RNA 2	MLP 2-6-1	(Sigmoidal - Identidade)
RNA 3	MLP 2-3-1	(Exponencial - Exponencial)
RNA 4	MLP 2-3-1	(Exponencial - Identidade)
RNA 5	MLP 2-7-1	(Sigmoidal - Exponencial)

Os gráficos da Figura 1 demonstram a distribuição dos resíduos por meio da dispersão e o histograma de frequência dos modelos volumétricos utilizados e da RNA 1, a melhor selecionada.

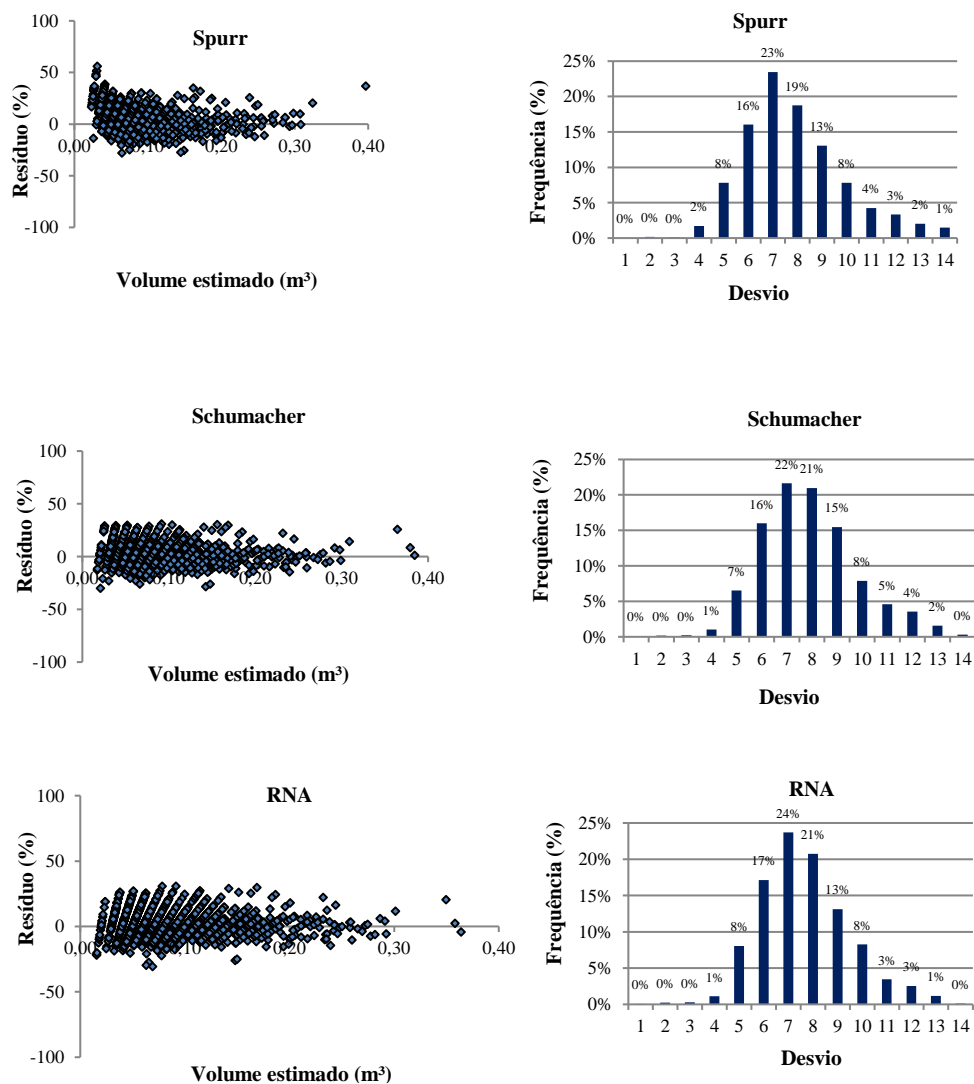


Figura 1 - Gráficos de dispersão dos resíduos e histograma de frequência de resíduos evidenciando o comportamento dos desvios encontrados entre volume observado e volume estimado.

Ao observar os resultados apresentados pelo modelo de Schumacher e Hall, nota-se um $R^2_{aj} = 95,5\%$ e um $RMSE\% = 3,36\%$, que evidencia bom ajuste do modelo. O modelo de Spurr apresentou resultados próximos $R^2_{aj} = 94,9\%$ e $RMSE = 3,59\%$. De acordo com estatísticas apresentadas pelas RNAs selecionadas, pode-se observar coeficientes de determinação ajustado com valores superiores aos modelos de regressão, em torno de 97%,

e RMSE% em torno de 3%. Ao analisar visualmente os gráficos de dispersão dos resíduos e histograma de frequência dos resíduos para os modelos de regressão e RNA, pode-se notar que as estimativas pontuais estão próximas dos valores observados, não apresentando distribuições tendenciosas. Além disso, o valor do Syx% apresenta um bom desempenho.

Quanto a análise do teste K-S, esta apresenta uma boa concordância entre o valor estimado e o valor real nos modelos, assim, aceita-se a hipótese nula, o que indica que todos os métodos testados são válidos para estimar volume de eucalipto.

Resultados satisfatórios nas estimativas das RNAs foram encontrados com configurações de apenas 3 a 7 neurônios na camada oculta, comprovando que não são necessárias estruturas muito complexas para estimação do volume de árvores. Esses resultados estão em correspondência com Rodrigues et al. (2010) cujos resultados indicaram que três neurônios são suficientes para a estimação precisa do volume comercial em plantios de *Schizolobium amazonicum* (Huber) Ducke, da empresa Rio Concrem, estado do Pará. De acordo com Bullinaria (2014), recomenda-se, então, arquitetura de neurônios mais leve para aplicação das RNAs. A escolha da estimativa com configurações com menor número de neurônios na camada oculta tem como objetivo evitar o *overfitting* (sobre ajuste) que consiste na memorização dos dados treinados, comprometendo a capacidade de generalização (BRAGA *et al.*, 2000; SILVA *et al.*, 2010).

As funções de ativação testadas: tangente hiperbólica, exponencial, identidade e sigmoidal, apresentaram resultados satisfatórios para estimação do volume, tanto na camada oculta quanto na camada de saída.

Outro objetivo do estudo é verificar se há mudanças na performance em relação ao tamanho da amostra, sendo assim, consideramos sete (7) amostras aleatórias, além da base completa, para avaliar o desempenho entre o melhor modelo de regressão ajustado, Schumacher e Hall, a rede neural artificial selecionada com melhor capacidade em estimar o volume de árvores de eucalipto e a RNA considerando a adição da variável categórica, o tratamento (tipo de clone).

Foram comparadas, por amostra, as estatísticas das estimativas (Tabela 7), do melhor modelo de regressão ajustado, Schumacher e Hall e os dois melhores modelos de RNAs selecionados, um com a adição da variável categórica (tipo de clone) e o outro considerando as mesmas variáveis do modelo de regressão.

Tabela 7 - Estatísticas das RNAs para o volume de *Eucalyptus spp.* considerando como entrada a variável categórica ‘Tratamento’ (tipo de clone de *Eucalyptus*)

Estatística	Amostra	Schumacher e Hall		
		RNA com Clone	RNA	
R²_{aj}	25	0,99917	0,99413	0,98808
	50	0,99724	0,98403	0,95968
	100	0,98936	0,9862	0,96756
	200	0,99183	0,98378	0,97256
	400	0,99025	0,98379	0,97351
	800	0,98267	0,97838	0,95741
	1600	0,98455	0,98105	0,95484
	1862	0,98050	0,97718	0,95537
RMSE (%)	25	1,02867	2,18045	2,22681
	50	1,35342	2,79569	3,16725
	100	2,18120	2,41915	2,49448
	200	2,19968	2,60782	3,03082
	400	2,28010	2,63863	2,64500
	800	3,00855	3,26196	3,31740
	1600	3,05747	3,34009	3,38756
	1862	3,01761	3,30220	3,36967
S_{yx} (%)	25	3,24001	6,86776	7,48992
	50	4,52146	9,33976	10,93522
	100	7,63573	8,46871	8,83714
	200	7,56937	8,97384	9,01809
	400	7,76382	8,98462	9,05536
	800	10,2546	11,1183	11,3236
	1600	10,3946	11,3554	11,5003
	1862	10,2715	11,2401	11,4696

A Tabela 8 apresenta as arquiteturas das RNAs considerando o número de neurônios nas camadas de entrada, intermediária ou oculta e camada de saída de cada RNA avaliada e as funções de ativação das camadas oculta e de saída para o estudo das RNAs que tiveram a variável categórica incrementada para estimativa do volume de *Eucalyptus spp.* Observou-se que a RNA apresentou um bom desempenho, porém, o número de neurônios aumentou devido a entrada de uma variável categórica. Binoti et al. (2014) avaliando a configuração de RNA para estimação do volume de árvores, observou que não havia diferença perceptíveis entre os resultados das RNAs de pouca complexidade na camada oculta e as RNAs de maior complexidade, ou seja, maior número de neurônios na camada oculta, justificando a utilização da RNA com baixa complexidade evitando o sobre ajuste.

O uso de variável categórica na rede neural não é um problema para essa técnica e essa característica é uma vantagem para estimar volume de *Eucalyptus spp.* de diferentes tipos de clones de forma rápida se comparado com a necessidade de avaliar equações volumétricas por tratamento (tipo de clone). Porém, um classificador mais complexo que o

necessário pode gerar o *overfitting* (sobre ajuste) que tem como consequência básica dois problemas, o custo computacional dos processos e menor poder de generalização. (CAMPOS, 2000). Os resultados obtidos neste trabalho demonstraram que todas as técnicas são adequadas para a estimativa de volume de *Eucalyptus* spp. Além disto, não há diferenças perceptíveis no resultado das estimativas quando adicionado a variável tipo de clone, como esta variável não demonstrou ser necessária para o objetivo deste trabalho não justifica a utilização da RNA com a adição da variável categórica para a estimativa de volume de *Eucalyptus* spp.

Tabela 8 - Definição das RNAs por amostra com e sem a variável categórica “Clone”.

RNAs	RNA com Clone		RNA sem Clone	
	Arquitetura	Função de Ativação	Arquitetura	Função de Ativação
RNA Amostra 1	MLP 17-12-1	(Sigmoial - Identidade)	MLP 2-7-1	(Exponencial - Identidade)
RNA Amostra 2	MLP 17-15-1	(Exponencial - Identidade)	MLP 2-7-1	(Tangente Hiperbólica - Identidade)
RNA Amostra 3	MLP 17-9-1	(Exponencial - Identidade)	MLP 2-4-1	(Sigmoial- Identidade)
RNA Amostra 4	MLP 17-14-1	(Exponencial - Identidade)	MLP 2-3-1	(Exponencial - Identidade)
RNA Amostra 5	MLP 17-12-1	(Exponencial - Identidade)	MLP 2-4-1	(Tangente Hiperbólica - Identidade)
RNA Amostra 6	MLP 17-12-1	(Exponencial - Identidade)	MLP 2-8-1	(Tangente Hiperbólica - Identidade)
RNA Amostra 7	MLP 17-5-1	(Sigmoial- Identidade)	MLP 2-3-1	(Sigmoial- Identidade)
RNA Base Completa	MLP 17-9-1	(Exponencial - Sigmoial)	MLP 2-5-1	(Tangente Hiperbólica - Identidade)

Ao adicionar uma variável categórica (tratamento) na modelagem os resultados das RNAs foram melhorados. De acordo com estatísticas apresentadas pelas RNA treinadas e selecionadas, pode-se observar coeficientes de correlação ajustado com valores entre 98% e 99%, e RMSE entre 1% e 3%. Ao analisar visualmente os gráficos de dispersão dos resíduos, pode-se notar que houve uma melhora nos gráficos de dispersão das RNAs que não consideram a variável categórica na modelagem e também do modelo de regressão, sendo que as estimativas foram pontuais próximas dos valores observados, e não apresentaram distribuições tendenciosas.

Os resultados foram satisfatórios no uso das RNAs com a adição da variável categórica, porém as configurações na camada oculta aumentaram para 5 a 15 neurônios.

Este aumento na configuração da camada oculta pode comprometer a habilidade de generalização da RNA, sendo assim, é importante avaliar a necessidade dessa variável para a estimação de volume de eucaliptos.

Com o interesse de comparar o performance do modelo e das RNAs em relação ao tamanho da base de dados, foram utilizados os gráficos de linha (Figura 2).

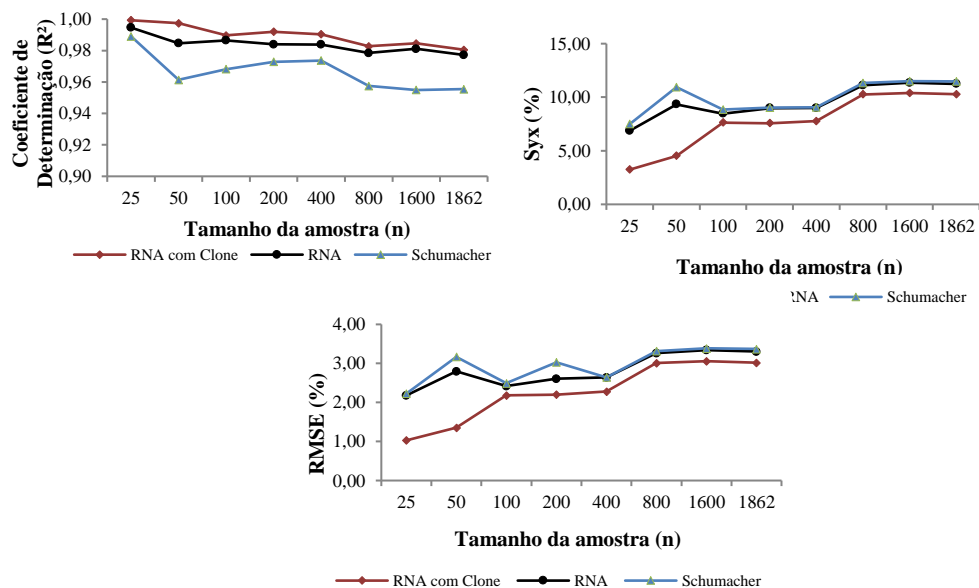


Figura 2 - Gráficos de Coeficiente de determinação (R^2_{aj}), Erro padrão da estimativa (Syx%) e Raiz quadrada do erro médio (RMSE%) por amostra dos modelos analisados.

A avaliação da qualidade do ajuste dos modelos ao longo do tamanho da amostra utilizado para seu ajuste foi realizada por meio dos três indicadores contemplados na Figura 2. A análise do R^2_{aj} favorece as RNAs com a adição da variável categórica (tipo de clone). Os dois modelos de RNAs (com clone e sem clone) foram superiores ao modelo de Schumacher e Hall. A superioridade das RNAs com relação ao modelo de Schumacher e Hall e seu melhor desempenho com a inclusão da variável tipo de clone, foi confirmada com a análise do Erro padrão da estimativa (Syx%) e da Raiz quadrada do erro médio (RMSE), onde observa-se uma leve superioridade na RNA com a inclusão da variável categórica, tipo de clone, em seguida a RNA sem a variável categórica e por último o modelo de Schumacher e Hall. Este resultado coincide com os de Silva *et al.* (2009). Para Hayklin, 2001; Silva *et al.*, 2009, esse resultado é de grande auxílio ao pesquisador. Com a possibilidade dos modelos de RNAs permitir a adição da variável categórica, tipo de clone, não seria necessário o ajuste dos modelos volumétricos para cada tipo de clone já que o ajuste da RNA é representativo para todos os clones. O pesquisador precisa avaliar qual o objetivo principal do estudo para analisar se a leve superioridade nos resultados com a inclusão da variável, tipo de clone, justifica a escolha do modelo.

Quanto a análise da influência do tamanho da amostra sobre os indicadores da performance de ajuste dos modelos, aqui avaliados, os resultados revelam que os indicadores não dependem apenas do tamanho da amostra, mas também da variância. A Tabela 2 indica que quanto menor a amostra maior o erro amostral consequentemente os resultados obtidos, que indicam melhor desempenho em amostras pequenas podem não ser consistente. Não se observa uma tendência de aumento do R^2_{aj} como esperado, nem uma tendência de diminuição da RMSE% e do Syx% com o aumento do tamanho da amostra. Os valores altos do R^2_{aj} nas pequenas amostras com seus correspondentes valores altos de Syx% e RMSE% induz o perigo da utilização de tamanho de amostras pequenas para modelagem de dados que serão usados na tomada de decisão. Assim, ressalta-se que a utilização de base de dados maior permite abranger a maior variabilidade possível e extrapolações mais confiáveis.

O uso de RNA para estimação de volume permite rapidez nos resultados e bom desempenho. Além disso, a adição de variáveis categóricas ao modelo não é um problema para o treinamento das RNAs, sendo esta uma característica que se destaca na metodologia. Entretanto, neste estudo a inclusão da variável categórica tipo de clone não proporcionou aumento de precisão na estimativa dos volumes de clones de *Eucalyptus* spp.

4 Conclusões

Conclui-se que todas as técnicas foram eficientes para estimar volume de árvores de *Eucalyptus* spp. O modelo de regressão que melhor se ajustou aos dados foi o Schumacher e Hall. A modelagem por Redes Neurais Artificiais (RNA) apresentou um melhor desempenho comparado aos modelos tradicionais. A RNA com a adição de uma variável categórica (tipo de clone) na modelagem não apresenta diferenças perceptíveis não demonstrando ser necessária para a estimação do volume. Quanto à influência do tamanho da base de dados os resultados revelam que os indicadores não dependem apenas do tamanho da amostra, mas também da variância, amostras pequenas tendem a ter um erro amostral maior consequentemente resultados menos confiáveis.

Agradecimentos

A primeira autora agradece a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – Brasil, pela concessão da bolsa de estudos. O segundo autor agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – Brasil, pelo incentivo ao projeto de pesquisa “MÓDULO DE EXPERIMENTAÇÃO FLORESTAL PARA O PÓLO GESSEIRO DO ARARIPE/PE”, que é coordenado pelo grupo de pesquisa do Laboratório de Biometria e Manejo Florestal, do Departamento de Ciência Florestal, da Universidade Federal Rural de Pernambuco, em parceria com o Instituto Agrônomo de Pernambuco (IPA). Os autores agradecem aos revisores e editores pelos comentários e sugestões.

SOUZA, S. R. R.; SILVA, J. A. A.; GUERA, O. G. M.; FERREIRA, T. A. E. Neural networks for volumetric estimation of Eucalyptus spp clones in gypsum pole of Araripe. *Rev. Bras. Biom.*, Lavras, v.36, n.3, p.715-729, 2018.

- **ABSTRACT:** *The objective of this study was to estimate the volume of Eucalyptus spp clones (genus of rapid growth) in the Araripe Gypsum Pole, responsible for 97% of the national production of gypsum, employing the methodology of Artificial Neural Networks (ANNs) and comparing it with the volumetric models of Schumacher and Hall and Spurr and also verify the efficiency of the estimation using different sample sizes and evaluate the contribution of a categorical variable in the estimation. Data came from an experiment implanted in the Experimental Station of the Agronomic Institute of Pernambuco, where was tested 15 clones of Eucalyptus spp planted in 2002, with final cut in 2009. It was also valued the adjustment of the best models for sample size. The goodness of fit of the models was evaluated based on: the adjusted coefficient of determination (R^2_{aj}), square root of the percent mean error (RMSE%), standard error estimate (S_{yx} %) and an analysis graphic of the residues. The results obtained in the study showed that all modeling was adequate and it was observed that the efficiency of the adjustments depends not only on the sample size, but also on the variance, and that the addition of a categorical variable in the ANNs does not show any perceptible differences, necessary for volume estimation.*
- **KEYWORDS:** *Forest management; volumetric model; artificial neural networks.*

Referências

BINOTI, D. H. B.; BINOTI, M. L. M. S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. *Ciência da Madeira*, v.5, n.1, p.58-67, 2014.

BRAGA, A. P.; CARVALHO A. P. L. F.; LUDEMIR, T. B. *Redes Neurais Artificiais: Teoria e Aplicações*. LTC - Livros Técnicos e Científicos Editora S.A., 2000. 262p.

BULLINARIA, J. A. *Introduction to Neural Computation*. Notas de aula. Disponível em: <<http://www.cs.bham.ac.uk/~jxb/inc.html>>. (Acesso em 01/10/2014).

CAMPOS, B. P. F.; SILVA, G. F.; BINOTI, D. H. B.; MENDONÇA, A. R.; LEITE, H. G. Descrição do perfil do tronco de árvores em plantios de diferentes espécies por meio de redes neurais artificiais. *Pesquisa Florestal Brasileira*, Colombo, v.37, n.90, p.99-107, abr./jun. 2017.

CAMPOS, T. E. *Técnicas de seleção de atributos e de classificação para reconhecimento de faces. Exame de qualificação do projeto de extensão*. 2000, disponível em: <<http://www.vision.ime.usp.br/~teo/publications/qualificacao/node8.html>>. (Acesso em 23/10/2017).

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

LEAL, F. A.; MIGUEL, E. P.; MATRICARDI, E. A. T.; PEREIRA, R. S. Redes neurais na estimativa de volume em um plantio de eucalipto em função de fotografias hemisféricas e números de árvores. *Revista Brasileira de Biometria*, v.33, n.2, p.233-249, 2015.

LIMA, R. B.; APARÍCIO, P. S.; FERREIRA, R. L. C.; SILVA, W. C.; GUEDES, M. C.; OLIVEIRA, C. P.; SILVA, D. A. S.; BATISTA, A. P. B. *Volumetria e classificação da capacidade produtiva para Mora paraensis (Ducke) no estuário amapaense*. Universidade Federal Rural de Pernambuco. Recife, 2014.

MARTINS, A. P.; DEBASTINI, A. B.; PELISSARI, A. L.; MACHADO, S. A.; SANQUETTA, C. R. *Estimativa do Afilamento do Fuste de Araucária Utilizando Técnicas de Inteligência Artificial*. Floresta e Ambiente, 2017. Disponível em: <http://www.scielo.br/pdf/floram/v24/2179-8087-floram-24-e20160234.pdf> (Acesso em: 17 Outubro 2017).

RODRIGUES, E. F.; OLIVEIRA, T. F.; MADRUGA, M. R.; SILVEIRA, A. M. A Um método para determinar o volume comercial do *Schizolobium Amazonicum* (huber) ducke utilizando redes neurais artificiais. *Revista Brasileira de Biometria*, São Paulo, v.28, n.1, p.16-23, 2010.

SCHUMACHER, F. X.; HALL, F. S. Logarithmic expression of timber-tree volume. *Journal of Agricultural Research*, v.47, n.9, p719-734, 1933.

SILVA, J. A. A. Potencialidades de florestas energéticas de eucalyptus no pólo gesseiro do Araripe – Pernambuco. *Anais da Academia Pernambucana de Ciência Agrônômica*. Recife. v. 5 - 6, p. 301 – 319, 2008- 2009.

SILVA, M. L. M.; BINOTI, D. H. B.; GLERIANE, J. M.; LEITE, H. G. Ajuste do modelo de Schumacher e Hall e aplicações de redes neurais artificiais para estimar volume de árvores de Eucalipto. *Revista Árvore*, v. 33, n. 6, p. 1133-1139, 2009.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. *Redes Neurais Artificiais para engenharia e ciências aplicadas*. Artliber, 2010. 397 p.

SOBRINHO, A. C. P. L.; AMARAL, A. J. R.; DANTAS, J. O. C. *Gipsita*. Departamento Nacional de Produção Mineral. 2003, disponível em: <<http://www.dnpm-pe.gov.br/>>. (Acesso em: 01/03/2015).

SPURR, S. H. *Forest inventory*. New York: The Ronald Press Company, 1952. 476p.

Recebido em 27.03.2017

Aprovado após revisão em 08.12.2017