REVISTA FLORESTA
Engenharia Florestal
ISSN 1982-4688
Edição 54
Universidade Federal do Paraná
Setor de Ciências Agrárias

USE OF MACHINE LEARNING, FIXED AND MIXED MODELS FOR VOLUME ESTIMATION IN FLOODPLAIN FOREST IN THE AMAZON ESTUARY

Anthoinny Vittória dos Santos Silva[1*], Rodrigo Galvão Teixeira de Souza[2], Rafael Lucas Figueiredo de Souza[3], Renan Mendes Santos[4], Robson Borges de Lima[2], Jadson Coelho de Abreu[2]

[1*] Federal University of the Jequitinhonha and Mucuri Valleys (UFVJM), Department of Forestry Engineering, Diamantina, Minas Gerais, Brazil - vittoria.sntsilva@gmail.com *
[2]State University of Amapá (UEAP), Department of Forestry Engineering, Macapá, Amapá, Brazil - rodrigosouza55@hotmail.com; robson.lima@ueap.edu.br; jadson.abreu@ueap.edu.br
[3]University of São Paulo (USP), Department of Forestry Engineering, Piracicaba, São Paulo, Brazil - rafaelflorestal55@gmail.com
[4] Forestry Engineer, Macapá, Amapá, Brazil – renansantosfl2016@gmail.com

Received for publication: 01/10/2021 - Accepted for publication: 19/02/2024

Resumo

Uso de aprendizado de máquina, modelos fixos e mistos para estimativa de volume em floresta de várzea no estuário Amazônico. O objetivo deste estudo foi avaliar o desempenho de redes neurais artificiais (RNAs), máquina de vetores de suporte (MVS) e modelos fixos e mistos para estimativa de volume de árvores individuais em uma floresta de várzea do estuário amazônico. O inventário foi efetuado em uma área localizada no distrito de Itatupã - Gurupá, Pará, Brasil. O modelo linear de *Schumacher e Hall* foi empregado para estimativa do volume. Treinou-se 240 RNAs e 32 configurações de MVS, com 4 variações das variáveis de entrada: diâmetro à altura do peito (*D*), altura comercial (h_m), classes diamétricas (*CCD*), classes de altura comercial (CCh_m) e espécie (*ESP*), sendo o volume (*V*) a variável de saída. As RNAs foram treinadas no *software* Neuro 4.0.6, as análises dos modelos de regressão fixo, misto e MVS foram efetuadas no *software* R. Os critérios de informação AIC (*Akaike's Information Criterion*) e BIC (*Bayesian Information Criterion*) foram empregados para os modelos. Para as comparações de todos os modelos se utilizou: coeficiente de correlação ($r_{Y\hat{Y}}$), *bias*, raiz quadrada do erro quadrático médio (REQM) e análise de distribuição dos resíduos. As melhores métricas estatística foram obtidas pela RNA do algoritmo RPROP- com *inputs* $D+h_m+CCD+CCh_m$, com oito neurônios na camada oculta e função de ativação tangente hiperbólica, com 0,9827 de $r_{Y\hat{Y}}$, REQM de 0,2439, bias de 0,0080 e distribuição gráfica residual tendendo a homogeneidade.
Palavras-chave: floresta nativas estuarinas, modelagem computacional, volumetria.

Abstract

The present study aimed to evaluate the performance of artificial neural networks (ANNs), support vector machines (SVM), and fixed and mixed models to estimate the volume of individual trees in an Amazonian estuary floodplain forest. The forest inventory was conducted in an area located in the district of Itatupã - Gurupá, Pará, Brazil. The Schumacher and Hall linear model was used in its fixed and mixed form to estimate the volume. A total of 240 ANNs and 32 SVM configurations were trained, with 4 variations of the input variables: diameter at breast height (*D*), commercial height (h_m), diameter classes (*CCD*), commercial height classes (CCh_m),) and species (*SP*), with the volume (*V*) being the output variable. The ANNs were trained in the Neuro 4.0.6 software program, and the analyses of the fixed and mixed regression models and SVM were performed in the R software program. The AIC (Akaike's Information Criterion) and BIC (Bayesian Information Criterion) information criteria were used for the regression models, while the following were used for comparisons of all models: correlation coefficient ($r_{Y\hat{Y}}$), bias, root mean squared error (RMSE), and residual distribution analysis. The best statistical metrics were obtained by the ANN of the RPROP- algorithm with $D+h_m+CCD+CCh_m$ inputs, with eight neurons in the hidden layer and hyperbolic tangent activation function, with 0.9827 of $r_{Y\hat{Y}}$ RMSE of 0.2439, bias of 0.0080 and residual graphic distribution tending to homogeneity.
Keywords: computational modeling, estuarine native forests, volumetry.

INTRODUCTION

Volume is one of the essential variables in forest data quantification (GAMA *et al.*, 2017). Whether at the individual or stand level, the forest potential of the area, the viability of timber stocks, and the planning of harvesting of commercial species can be defined through volume prediction.

Considering the geometrically irregular shape of tree stems, there are several alternatives for volume determination, both destructive and non-destructive. Rigorous cubing is generally used to obtain parametric values that have higher accuracy in volume calculation (SOUZA *et al.*, 2017). The method involves applying n sections on the stem, where the diameter is successively measured at fixed heights.

Another alternative is volume estimation, conventionally done using volume equations. These equations are developed through a regression model, also known as a fixed-effects model, and use the diameter at breast height (D) and commercial height (h_(m)) of trees as independent variables (MARTINS *et al.*, 2016).

These equations may have lower precision due to data variability, reducing reliability for economic analyses (ABREU *et al.*, 2020). Therefore, mixed models are employed as an alternative for describing reality with precision and accuracy, aligning with the main goal of modeling (GONÇALVES *et al.*, 2016).

However, alternatives that replace regression models are necessary due to the increased complexity in equation adjustments, such as tools provided by artificial intelligence (AI) (CAMPOS *et al.*, 2016). Support Vector Machines (SVM) and Artificial Neural Networks (ANNs) are part of machine learning and have been employed in forestry sciences due to their effectiveness and higher levels of confidence (BINOTI *et al.*, 2016).

Based on the hypothesis that floodplain areas present greater complexities in volume determination and following the premise that AI tools provide feasible predictions with better performance compared to conventional volumetric regression models, the objective of this study was to evaluate the statistical performance of these different alternatives for estimating the volume of individual trees in a floodplain forest of the Amazon estuary.

MATERIAL AND METHODS

Study area

The study area is characterized as an estuarine floodplain forest located in the Itatupã district, belonging to the municipality of Gurupá, Pará, Brazil, situated at coordinates 0°32'54.68"S and 51°15'11.10"W (Figure 1).

According to the Köppen climate classification, the area has an “Ami” climate type with an annual rainfall regime defined by a dry season, although rainfall indices are sufficient to ensure necessary water levels for establishing specific local biodiversity. The average annual temperature is 26°C. Annual precipitation exceeds 2,000 mm, and seasons are virtually non-existent.

The relative humidity is above 80%, with a total absence of dry periods. The geological formation corresponds to the Quaternary period, composed of alluvial units (i.e. recent sedimentary covers) comprising unconsolidated alluvial deposits of varied grain sizes, forming an extensive alluvial plain. The predominant soils in the study region are glazed eutrophic hydromorphic soils (humic and slightly humic) with a silty texture on the margins of the Amazon Islands, becoming clayey as one moves further inland.

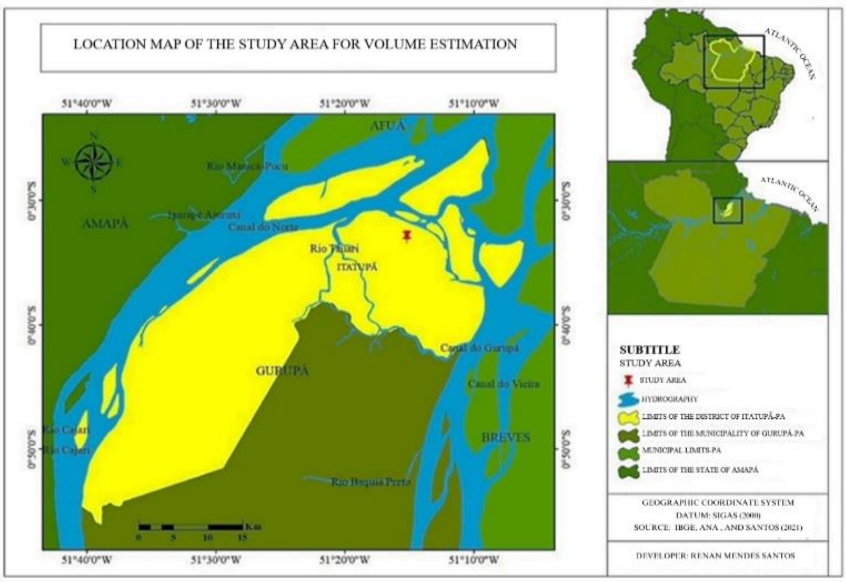

Figure 1. Study map to estimate the volume of a floodplain forest in the Amazon estuary.
Figura 1. Mapa de estudo para estimativa de volume de uma floresta de várzea no estuário amazônico.

Data collection

The forest inventory was conducted through cluster sampling with four subunits of 20x50 m, where the variables diameter at breast height (*D*), and commercial height (h_m) were collected using a measuring tape. In addition to the species variable, diameter classes (*CCD*) and commercial height classes (CCh_m) were calculated using the Sturges method to be used as categorical variables. The volume (*V*) of 100 fallen trees was obtained within and around the clusters using the Smalian method, given by:

$$v = \frac{(g1+g2)}{2}.l \tag{1}$$

Where: v is the volume of the section (m³), g_1 is the cross-sectional area at the base of the log (m²), g_2 is the cross-sectional area at the base of the log (m²), and l is the length of the log.

Linear fixed model and mixed effects

The intrinsically linearized model by Schumacher and Hall (1933) was employed for volume estimation, with its fitting performed through Restricted Maximum Likelihood (REML). The model is described in the fixed form as:

$$LnV = \beta_0 + \beta_1 LnD + \beta_2 Lnh_m + \varepsilon \tag{2}$$

Where: *Ln* is the natural logarithm, V is the volume in m³, *D* is the diameter with bark in cm, h_m is the total height in m, β_0, β_1 and β_2 are the parameters of the model, and ε is the random error.

For the mixed model, the Schumacher and Hall (1933) model was modified by adding random effects and adjusted using the Restricted Maximum Likelihood (REML) method. The model is presented as:

$$LnV = (\beta_0 + \alpha_i) + (\beta_1 + b_{1i})LnD + (\beta_2 + b_{2i})Lnh_m + \varepsilon \tag{3}$$

Where: β_0, β_1 and β_2 are the fixed parameters of the model, a_i is the random asymptote for the i-th species, and b_{1i} and b_{2i} are the random slope coefficients for the i-th species.

Three variables were used as random effects: diameter classes (*CCD*), commercial height classes (CCh_m), and species. The significance of the random effects on the parameters was verified through the Likelihood Ratio Test using the chi-squared test, with 1 degree of freedom and 5% significance level, given by:

$$DM = -2 \ln \left[\frac{(likelihood\ without\ variable)}{(liklihood\ with\ the\ variable)} \right] \tag{4}$$

Where: DM is the significance of the difference.

The *nlme* package (PINHEIRO *et al.*, 2021) of the R software program (R DEVELOPMENT CORE TEAM, 2014) was used for fitting the fixed and mixed models. Moreover, the AIC (Akaike's Information Criterion) and BIC (Bayesian Information Criterion) information criteria were used to compare the models, and the following statistical criteria were used: ($r_{Y\hat{Y}}$), *bias*, root mean square error (RQEM), and analysis of residual distribution.

Artificial Neural Networks (ANNs)

The data for the machine learning models was divided into 70% for training and 30% for validation. A total of 240 artificial neural networks (ANNs) of the multilayer perceptron type were trained in a supervised manner using the Neuro 4.0.6 software program, employing the resilient propagation (RPROP+ and RPROP-) and backpropagation (BACKPROP) training algorithms. The stopping criteria adopted for the ANNs were mean squared error (0.001) or number of cycles (1000). Logistic and hyperbolic tangent activation functions were used, with four input variations (Table 1) and volume (V) as the output. The best model was selected based on the statistical criteria: correlation coefficient ($r_{Y\hat{Y}}$), *bias*, root mean square error (RQEM), and analysis of residual distribution.

Table 1. Input variable settings for artificial neural network and support vector machine models.
Tabela 1. Configurações de variáveis de entrada para os modelos de redes neurais artificiais e máquina de vetores de suporte.

Configurations	*Inputs*	*Output*
1	$D+h_m$	*V*
2	$D+h_m+SP$	*V*
3	$D+h_m+CCD+CCh_m$	*V*
4	$D+h_m+CCD+CCh_m+SP$	*V*

Where: *D* is the diameter at breast height; h_m is commercial height; SP is the species; *CCD* is the diameter class; CCh_m is the commercial height class; *V* is volume.

Support Vector Machine (SVM)

Next, 32 configurations of Support Vector Machine (SVM) were trained using the e1071 package in the R software program (R DEVELOPMENT CORE TEAM, 2014) for four types of Kernel functions: Radial Basis

Function (RBF), linear, polynomial, and sigmoidal, with two error functions, Type I (equation 13) and Type II (equation 17), and with four input variations (Table 1).

Statistical criteria were employed for unbiased evaluation of the best SVM configuration for volumetric estimation: correlation coefficient ($r_{Y\hat{Y}}$), *bias*, root mean square error (RQEM), and analysis of residual distribution.

The Type I error function is described as:

$$Minimize\,\frac{1}{2}.w^T w + C.\sum_{i=1}^{n}\xi_i + C.\sum_{i=1}^{n}\xi_i^* \tag{5}$$

Subject to the following restrictions:

$$w^T.\phi.(x_i\ ) + b - y_i \leq \varepsilon + \xi_i^* \tag{6}$$

$$y_i - w^T.\phi(x_i\ ) - b\ \leq \varepsilon + \xi_i^* \tag{7}$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots N \tag{8}$$

Where: *w* is the coefficient vector, *C* is the penalty parameter of than an operator and ξ^* are variables representing the error space above and below the *ε-tube*, *i* are the cases to be trained, *N* is the total number of these cases to be trained, $\phi.(x_i\ )$ is the chosen kernel function, *b* is the bias, y_i is the output data, and ε is the maximum tolerated error.

The function of the Type II error is demonstrated as:

$$Minimize\,\tfrac{1}{2}.w^T w - C(v.\varepsilon + \tfrac{1}{2}.\ \textstyle\sum_{i=1}^{n}(\xi_i + \xi_i^*)) \tag{9}$$

Where: v is the parameter that regulates the number of support vectors.

Subject to the constraints:

$$(w^T.\phi.(x_i\ ) + b) - y_i \leq \varepsilon + \xi_i^* \tag{10}$$

$$y_i - (w^T.\phi(x_i\ ) + b)\ \leq \varepsilon + \xi_i^* \tag{11}$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots N\ \ , \varepsilon \geq 0 \tag{12}$$

RESULTS

Data description

A total of 100 sample trees were used in the analyses, totaling 23 species categorized into 12 families. Among these, the species *Vochysia tucanorum* Mart. (Cinzeira), *Carapa guianensis* Aubl. (Andiroba), *Swartzia schomburgkii* Benth. (Pitaica), and *Alantoma decandra* (Ducke) Samori, Ya Y. Huang & Prance (Cerú), from the Vochysiaceae, Meliaceae, Fabaceae, and Lecythidaceae families, respectively, showed higher frequencies. These species are common in estuarine floodplain areas with timber potential.

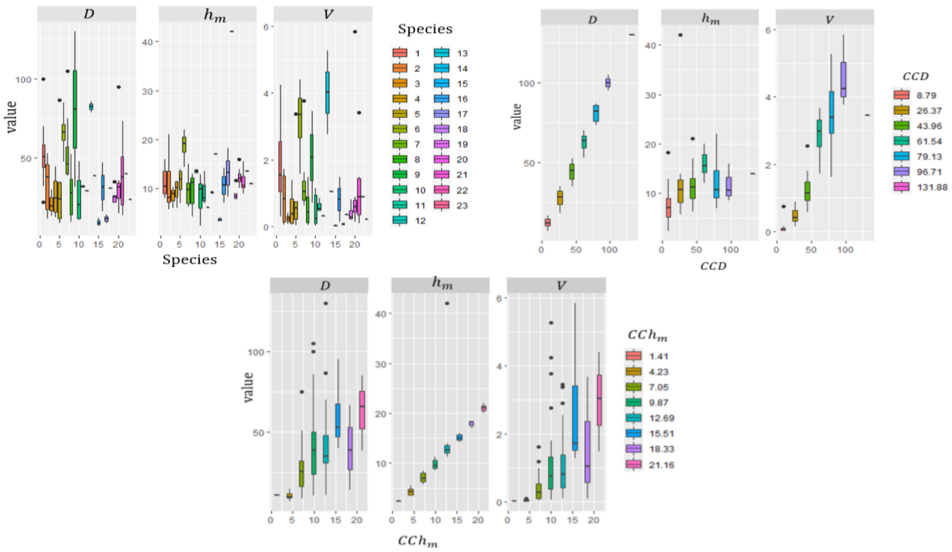

Figure 2. Boxplot of diameter at breast height (*D*), commercial height (h_m), volume data with species, diameter classes (*CCD*), and commercial height classes (CCh_m).

Figura 2.Boxplot dos dados de diâmetro à altura do peito (D), altura comercial (h_m) e volume em relação as espécies, as classes de diâmetro (CCD) e as classes de altura comercial (CCh_m).

The relationship between species, diameter at breast height (D), commercial height (h_m) and volume (V) can be observed in Figure 2, where great heterogeneity in the data is noted, along with scarce participation of some species, and the presence of outliers.

Seven classes of D were generated, where the diameters at breast height ranged from 7.0 cm to 130 cm and commercial heights ranged from 2.35 m to 42 m. Then, the relationships between the D and h_m classes with the studied dendrometric variables can be seen in Figure 3. Unlike the species, the classes present greater homogeneity of the data, although h_m show a greater presence of outliers.

Fixed and Mixed Effects Linear Model

The parameters obtained for the Schumacher and Hall (1933) model were significant with p-value<0.05, resulting in the equation:

$$LnV = -8.49 + 1.89.LnD + 0.58.Lnh_m + \varepsilon \qquad (13)$$

The statistical criteria obtained were a correlation of 88.5%, root mean square error (RQEM) of 69.86%, and bias of 0.24%. Considering the heterogeneity found in a native floodplain forest as a parametric method, the values resulting from the obtained equation are satisfactory.

Although, Schumacher and Hall's (1933) model in its fixed form yielded satisfactory results, upon inclusion of random effect variables (Species, CCD, CCh_m) in transforming the model into mixed form, it was observed that the random effects were not significant to the model, as verified through the Likelihood Ratio Test using the chi-squared test with 1 degree of freedom and 5% significance level.

The AIC and BIC information criteria confirmed better performance of the model in its fixed form with values of 90.89 and 101.31, respectively, as opposed to the values of 96.89 and 115.13 for the mixed effects model. It is observed in Figure 3 that the residuals show a tendency towards homogeneity around the mean line, with slight underestimations and overestimations.

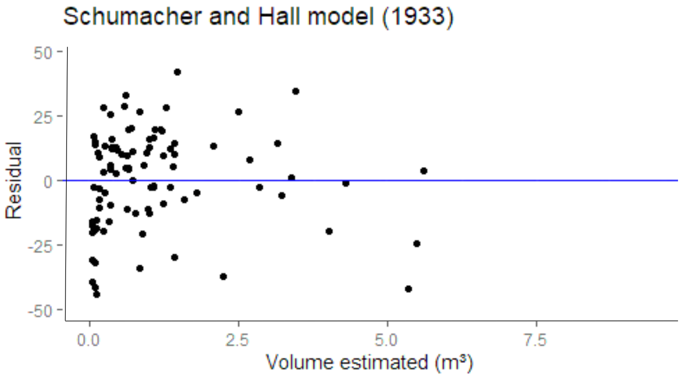

Figure 3. Residual graphical distribution of the Schumacher and Hall model (1933) for volume estimation.
Figura 3. Distribuição gráfica residual do modelo de Schumacher e Hall (1933) para estimativa de volume.

Artificial Neural Networks (ANNs)

The best models among the 240 ANNs trained were selected among the four input configurations (inputs) for the three algorithms used and the two activation functions, as shown in Table 2.

The validation step demonstrates the real performance of the algorithm, whether it achieved satisfactory generalization, meaning if there was learning through unknown data. It is observed in Table 2 that the best ANN model was with the resilient propagation algorithm without backpropagation (RPROP-), employing the hyperbolic tangent activation function and the input variables D+h_m +CCD+CCh_m.

The aforementioned model obtained the following results for statistical metrics in the training stage: 0.9642 for $r_{y\hat{y}}$, 0.3108 for RQEM, and -0.0089 for *bias*. This configuration showed improvement in validation data.

It is noted that the categorical variable species was not significant as a categorical variable for the ANN models. Although diameter and commercial height classes did not positively influence the model as a variable in the BACKPROP algorithm, RPROP with and without backpropagation favored better performances.

All algorithms with only the variables D and h_m, and emphasizing RPROP+, showed statistical metrics that demonstrate the potential of these models in optimizing forest activities.

Table 2. Statistical criteria used in artificial neural network (ANN) models for validation data.
Tabela 2. Critérios estatísticos empregados nos modelos de redes neurais artificiais (RNA) para os dados de validação.

ALGORITHM	CONF.	F. A	NN	$r_{Y\hat{Y}}$	RQEM	BIAS
BACKPROP	1	LOG	8	0.9743	0.2939	-0.0204
BACKPROP	2	LOG	8	0.9306	0.4287	0.0917
BACKPROP	3	LOG	8	0.9218	0.4393	-0.0050
BACKPROP	4	TANH	8	0.9457	0.4344	0.2111
RPROP-	1	TANH	8	0.9716	0.3444	0.1904
RPROP-	2	LOG	8	0.9554	0.2391	-0.0406
RPROP-	3	TANH	8	0.9827	0.2439	0.0080
RPROP-	4	TANH	8	0.9738	0.2985	-0.0301
RPROP+	1	LOG	8	0.9800	0.2440	0.0036
RPROP+	2	LOG	8	0.9655	0.3492	-0.0928
RPROP+	3	LOG	8	0.9692	0.3645	0.0338
RPROP+	4	LOG	8	0.9451	0.2655	0.0011

Where: Conf. are the input variable settings; F.A is the activation function; NN is the number of neurons in the intermediate layer.

In Figure 4, it is observed that the model showed residual behavior with a tendency towards homogeneity around the mean line for both training and validation data, however slight underestimations and overestimations are observed.

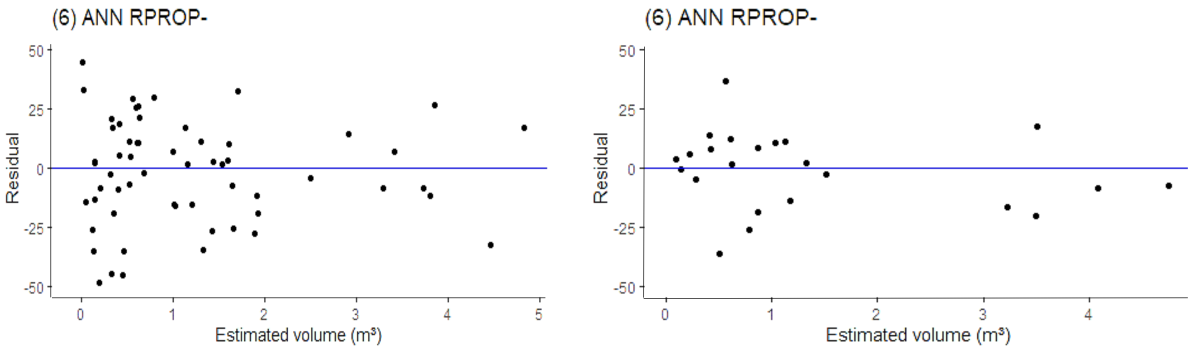

Figure 4. Residual graphical distribution for the best artificial neural network (ANN) model for volume estimation for training and validation data.
Figura 4. Distribuição gráfica residual para o melhore modelo de rede neural artificial (RNA) para estimativa de volume para os dados de treino e validação.

Support Vector Machines (SVM)

Among the 32 trained SVM configurations, four models were chosen from the eight models for the four input variable configurations. The results for the validation data are observed in Table 3, and in contrast to the other machine learning algorithms used in this study, which are ANNs, for SVMs, the model configuration which includes all five variables ($D+h_m+ESP+CCD+CCh_m$) obtained the best results employing the Type II RBF function.

However, it is worth emphasizing that for the most simplified configuration using only D+ h_m for RBF II, as shown in Table 3, the statistical criteria presented good values for the validation data, demonstrating the potential of using this simplified configuration for volumetric estimation.

It is observed in Figure 5 that the best SVM model for the training and validation data showed a tendency towards more homogeneous behavior of the residuals, with slight overestimations and underestimations in the initial volumes.

Table 3. Statistical criteria used in support vector machine (SVM) models for validation data.
Tabela 3. Critérios estatísticos empregados nos modelos de máquinas de vetores de suporte (MVS) para os dados de validação.

CONFIGURATION	F. K.	F. E	$r_{Y\hat{Y}}$	RQEM	BIAS
1	Linear	II	0.9051	0.5930	0.1209
1	Polynomial	II	0.7080	1.3222	0.3099
1	RBF	II	0.9663	0.3807	0.0805
1	Sigmoidal	II	0.9056	0.5567	0.0402
2	Linear	II	0.8963	0.6032	0.0881
2	Polynomial	II	0.6906	1.3816	0.2988
2	RBF	II	0.9536	0.4283	0.0689
2	Sigmoidal	II	0.8966	0.5892	0.0675
3	Linear	II	0.9076	0.5800	0.1243
3	Polynomial	II	0.7333	1.0705	0.2503
3	RBF	II	0.9663	0.3417	0.0316
3	Sygmoidal	II	0.9113	0.5424	0.0682
4	Linear	II	0.9267	0.4520	-0.0451
4	Polynomial	I	0.7193	0.8469	-0.0893
4	RBF	II	0.9673	0.3113	-0.0383
4	Sigmoidal	II	0.9212	0.4846	-0.0582

Where: F.K is the Kernel function; F.E is the error function.

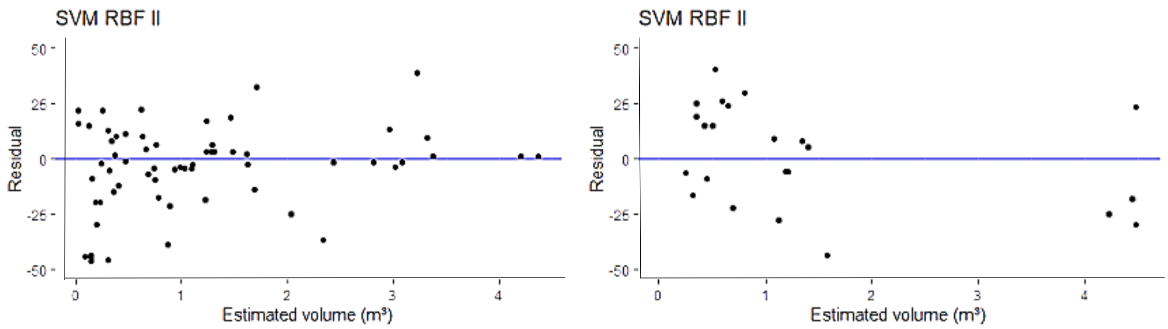

Figure 5. Residual graphical distribution for the best support vector machine (SVM) model for volume estimation for training and validation data.
Figura 5. Distribuição gráfica residual para o melhore modelo de máquina de vetor de suporte (MVS) para estimativa de volume para os dados de treino e validação.

DISCUSSION

The Schumacher and Hall (1933) model is a mathematical expression that effectively captures the functional relationship between volume and the variables diameter at breast height (D) and height (H). Its mathematical characteristics favor its widespread application in volumetric estimates of forest species (BIAZATTI *et al.*, 2020; GAMA *et al.*, 2017; FERNANDES *et al.*, 2017). It utilizes 100% of the data for modeling compared to artificial intelligence, which requires training and validation (ABREU *et al.*, 2020).

A smaller number of sample trees are accepted for regression models (FREITAS; ANDRADE, 2017). However, the heterogeneity in the number of species in the study database influenced the mixed-effects modeling, where the D and H classes as random effects did not improve the model.

Although machine learning models offer greater possibilities for input variable configurations, using only D and H commercial height was sufficient for the tested models. ANNs and SVMs demonstrated superior performance in volume estimation compared to regression methods (BINOITI *et al.*, 2016).

SVMs operate on smaller sample levels, offering flexibility and adaptability for more heterogeneous data (ZHANG; SUN, 2020). Consequently, their mathematics favors the values obtained in volumetric estimates in both the complexity of the typology studied and in other forest areas (LAFETÁ *et al.*, 2020; CUNHA NETO *et al.*, 2021).

Data normalization is necessary for some RNA datasets, i.e., homogenization of variable magnitudes in preprocessing (GORGENS *et al.*, 2010). In this study with highly heterogeneous variables, programming in R software resulted in difficulties in normalizing the data scale. The best generalization occurred with the ANN, with the normalization facilitated by the Neuro software program.

Obtaining the best RPROP- ANN configuration with eight neurons in just one hidden layer of a multi-layered network satisfies the universal approximation theorem (CYBENKO, 1989). This demonstrates that an RNA with a single hidden layer is capable of accurately approximating any continuous function (WINKLER; LE, 2016).

The use of the hyperbolic tangent activation function favored better convergence for the artificial neural network due to its simple operation in the range from -1 to 1. This function tends to produce outputs with averages close to zero, as it allows inputs to be normalized within the interval (LECUN *et al.*, 2012).

The RPROP algorithm has the characteristic of favoring better training (BINOTI *et al.*, 2014; MARTINS *et al.*, 2016). It operates more directly in changing synaptic weights based on local gradient data, and it updates the value for each weight individually (SAPUTRA *et al.*, 2017). The RPROP- variation allows the algorithm to jump over local minima numerous times. If there is a parameter rollback, inhibition of the last iteration occurs to favor a lower value for the parameter (GÜNTHER; FRITSCH, 2010).

Rigorous cubic measurement in inventory activities in estuarine floodplain forests is a costly activity. Impediments arise due to the complex characteristics of the area, such as difficult access and being regions which are prone to flooding. Studies in this typology are scarce, so providing more advanced metrics for obtaining more optimized volume is important. Volume is extremely important for forest conservation planning and decision-making (LANSSANOVA *et al.*, 2018).

CONCLUSIONS

- Studies of different volume estimation alternatives in floodplain forests are important due to the increased complexity of data acquisition and the scarcity of studies in this typology;
- Random effects were not significant in the mixed model;
- Although the Schumacher and Hall model presents inferior metrics to machine learning models for volumetric estimation, as it offers less flexibility regarding input variables, it can still be recommended due to its easy application and use in natural forest inventories;
- The species categorical variable only showed significance in the SVM RBF II model with all five input variables;
- Diametric and commercial height classes were significant as categorical variables for ANNs and SVMs;
- SVMs appear to be a potential model for volume estimation; however, the best statistical metrics were obtained by the RPROP- algorithm ANN with inputs $D+h_m+CCD+CCh_m$, with eight neurons in the hidden layer and hyperbolic tangent activation function.

REFERENCES

ABREU, J. C. D.; SOARES, C. P. B.; LEITE, H. G.; BINOTI, D. H. B.; SILVA, G. F. D. Alternatives to estimate the volume of individual trees in forest formations in the state of Minas Gerais, Brazil. **Cerne**, Lavras, v.26, p.393-402, 2020.

BIAZATTI, S. C.; SCCOTI, M. S. V.; DE BRITO JÚNIOR, J. F.; MÔRA, R.; CARDOZO, G. H. Eficiência de métodos para estimativa volumétrica de espécies nativas na Amazônia Ocidental. **Nativa**, Mato Grosso, v.8, n.2, p.253-259, 2020.

BINOTI, D. H. B.; BINOTI, M. L. M. S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Revista Ciência da Madeira (Brazilian Journal of Wood Science)**, Pelotas, v.5, n.1, p.58-67, 2014.

BINOTI, D.H.B.; BINOTI, M.L.M.S.; LEITE, H.G.; ANDRADE, A.V.; NOGUEIRA, G.S.; ROMARCO, M.L.; PINTANGUI, C.G. Support vector machine to estimate volume of eucalypt trees**. Revista Árvore**, Viçosa, v.40, n.4, p.689-693, 2016.

CAMPOS, B.P.F.; SILVA, G.F.; BINOTI, D.H.B.; MENDONÇA, A.R.; LEITE, H.G. Predição da altura total de árvores em plantios de diferentes espécies. **Pesquisa Florestal Brasileira**, Colombo, v.36, n.88, p.375-385, 2016.

CUNHA NETO, E. M.; MOURA, M. M.; ARAÚJO, E. C. G.; SANTANA, G. M.; DALLA CORTE, A. P.; SANQUETTA, C. R. Aprendizado de máquina e regressão linear na estimativa do volume de *Eucalyptus* na Amazônia Oriental. **BIOFIX Scientific Journal**, Paraná, v.6, n.1, p.08-14, 2021.

CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of control, signals, and systems**, [*S.l.*] v. 2, n. 4, p. 303-314, 1989.

FERNANDES, A. M. V., GAMA, J. R. V., RODE, R., & MELO, L. DE O. Equações volumétricas para *Carapa guianensis* Aubl. e *Swietenia macrophylla* king em sistema silvipastoril na amazônia. **Nativa**, Mato Grosso, v.5, n.1, p.73-77, 2017.

FREITAS, B. C.; DE ANDRADE, V. C. L. Bancos de dados de cubagem rigorosa de árvores-amostra de Corymbia citriodora empregados no ajuste de modelos volumétricos regionais. **Advances in Forestry Science**, Mato Grosso, v.4, n.2, p.93-97, 2017.

GAMA, J.R.V.; SOUZA, A.L.; VIEIRA, D.S.; LEITE, H.G. Equações de volume para uma floresta ombrófila aberta, município de Codó, estado do Maranhão. **Revista Brasileira de Ciências Agrária**s, Pernambuco, v. 12, n. 4, p. 535-542, 2017.

GONÇALVES, J.C. CARVALHO, S.P.C.; OLIVEIRA, A.D.; GOMIDE, L.R. Comparação dos modelos prognósticos de Clutter e da função logística. **Pesquisa Florestal Brasileira**, Colombo, v. 36, n. 87, p. 311-317, 2016.

GORGENS, E. B.; LEITE, H. G.; SANTOS, H. D. N.; GLERIANI, J. M. Estimação do volume de árvores utilizando redes neurais artificiais. **Revista Árvore**, Viçosa, v.33, p.1141-1147, 2009.

GÜNTHER, F.; FRITSCH, S. **Neuralnet:** Training of Neural Networks. The R Journal, 2010.

LAFETÁ, B. O.; SILVA, F. F.; SANTOS, M. A.; PIMENTA, I. A.; FONTAN, I. D. C. I.; FONSECA, N. R.; SARTORI, C. J. Modelagem morfométrica de *Licania tomentosa* (Benth.) por regressão logística e máquinas vetor de suporte. **Scientia Plena**, Sergipe, v.16, n.6, 2020.

LANSSANOVA, L.R.; SILVA, F.A.; SCHONS, C.T.; PEREIRA, A.C.S. Comparação entre diferentes métodos para estimativa volumétrica de espécies comerciais da Amazônia. **BIOFIX Scientific Journal**, Paraná, v.3, n.1, p.109-1145, 2018.

LECUN, Y.A.; BOTTOU, L.; ORR, G.B.; MULLER, K. Efficient backprop. **Neural networks:** Tricks of the trade. Springer, p. 9-48, 2012.

MARTINS, A.P.M.; CARVALHO, D.E.; SCHMIDT, L.N.; GORENSTEIN, M.R.; CYRSNEIROS, V.C. Eficiência de métodos para estimativa do volume total de Corymbia citriodora. **Sciencias Agrarias**. Paraná, v.15, n.2, p.178-184, 2016.

R DEVELOPMENT CORE TEAM. **R:** A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2014. Disponível em: <http://www.R-project.org/>

SAPUTRA, W.; ZARLIS, M.; SEMBIRING, R. W.; HARTAMA, D. Analysis of the resilient algorithm on artificial neural network backpropagation. **Journal of Physics: Conference Series,** Sumatera Utara, v.930, n.1, p.012035. IOP Publishing, 2017.

SCHUMACHER, F. X.; HALL, F. D. S. Logarithmic expression of timber-tree volume. **Journal of Agriculture Research,** [*S.l.*], v.47, n.9, p.719-734, 1933.

SOUZA, H.S.; DRESCHER, R.; VENDRUSCOLO, D.G.S.; MOURA, J.P.V.M.; SIQUEIRA, T.A.S.; MAMORÉ, F.M.D. Comparação de métodos de cubagem para eucalipto. **Revista Brasileira de Biometria**, Lavras, v.35, n.1, p.15-26, 2017.

WINKLER, D. A.; LE, T. C. Performance of deep and shallow neural networks, the universal approximation theorem, activity cliffs, and QSAR. **Molecular informatics**, Berlin, v. 36, n. 1-2, p. 1600118, 2017.

ZHANG, F.; O'DONNELL, L. J. Support vector regression. **Machine Learning**. Academic Press, [*S.l.*], p.123-140, 2020.