Multi-volume modeling of *Eucalyptus* trees using regression and artificial neural networks

Gileno Brito de Azevedo¹, Heitor Vicensotto Tomiazzi¹, Glaucia Taís de Oliveira Sousa Azevedo¹, Larissa Pereira Ribeiro Teodoro¹, Paulo Eduardo Teodoro°¹ *, Marcos Talvani Pereira de Souza¹, Tays Silva Batista¹, Humberto de Jesus Eufrade-Junior², Saulo Filipe Sebastião Guerra²

¹ Federal University of Mato Grosso do Sul (UFMS), Chapadão do Sul, MS, Brazil, ² College of Agricultural Sciences (FCA), Sao Paulo State University (UNESP), Botucatu, SP, Brazil

* eduteodoro@hotmail.com

Abstract

The stem volume of commercial trees is an important variable that assists in decision making and economic analysis in forest management. Wood from forest plantations can be used for several purposes, which makes estimating multi-volumes for the same tree a necessary task. Defining its exploitation and use potential, such as the total and merchantable volumes (up to a minimum diameter of interest), with or without bark, is a possible work. The goal of this study was to use different strategies to model multi-volumes of the stem of eucalyptus trees. The data came from rigorous scaling of 460 felled trees stems from four eucalyptus clones in high forest and coppice regimes. The diameters were measured at different heights, with the volume of the sections obtained by the Smalian method. Data were randomly separated into fit and validation data. The single multi-volume model, volume-specific models, and the training of artificial neural networks (ANNs) were fitted. The evaluation criteria of the models were: coefficient of determination, root mean square error, mean absolute error, mean bias error, as well as graphical analysis of observed and estimated values and distribution of residuals. Additionally, the t-test (α = 0.05) was performed between the volume obtained in the rigorous scaling and estimated by each strategy with the validation data. Results showed that the strategies used to model different tree stem volumes are efficient. The actual and estimated volumes showed no differences. The multi-volume model had the most considerable advantage in volume estimation practicality, while the volume-specific models were more efficient in the accuracy of estimates. Given the conditions of this study, the ANNs are more suitable than the regression models in the estimation of multi-volumes of eucalyptus trees, revealing greater accuracy and practicality.

Introduction

Currently, forest plantations in Brazil occupy about 9.85 million hectares of the planted area [1], being 6.73 certified by FSC and/or PEFC, with revenues that reach 1.1% of the gross domestic product (GDP), corresponding to U$ 18.5 billion [2]. The forest sector is increasingly...
consolidated, and along with its growth, the need to know the characteristics of forest plantations and their growth stocks increases. In Brazil, eucalyptus wood is used for several purposes, such as the production of firewood, coal, cellulose, wood panels, stakes, posts, wood for sawmills, among others. For each purpose, the wood must have specific dimensions. Thus, in plantations where wood is intended for multiple uses, using appropriate techniques to accurately estimate the sustainable production potential of wood for different purposes is paramount.

Stem volume is one of the essential measures to know the wood potential of the forest plantations, considering that this variable provides subsidies for the evaluation of the wood stock and the analysis of the sustainable yield potential of the forests [3–7]. In Brazil, the most commonly used procedure for obtaining tree volume in forest plantation is the volumetric equations. The Schumacher and Hall [8] model has been the primary model [6, 9, 10], besides the Taper equations in some cases [10, 11]. These equations allow the estimation of volume (total, or even a minimum diameter of interest), which is difficult to be directly obtained from variables that are more easily measured in the field, such as diameter at breast height and the total height of the trees [6, 12].

Despite the efficiency of volumetric equations in estimating wood volume [13, 14], they may have limited properties for some situations. For instance, when the goal is to estimate more than one volume for trees, such as total and merchantable volumes (up to a minimum diameter of interest), with or without bark, these volumetric equations are not efficient. With the ever-changing market conditions, the accurate estimate of tree volumes to different upper stem merchantability limits is paramount [4]. Nevertheless would require several equations, one for each volume of interest [15–17]. In turn, the Taper equations allow estimating the diameter at any stem height, the commercial height that occurs in a specific diameter, and the volume between stem sections. These estimates are adequate when the goal is to quantify multi-products of the tree [11, 18, 19]. A disadvantage of Taper equations, though, is the impossibility of estimating volumes with and without bark simultaneously. This information is desired, depending on the purpose of production and management.

In this sense, multi-volume models were developed to estimate different volumes, using a single equation [16, 20–22]. The multi-volume model developed by Leite et al. [20] was based on the expansion of the Schumacher and Hall [8] model and consisted of a volumetric model to simultaneously estimate the total and merchantable volumes (in different tree diameters), with and without bark. Adjusting these models can result in more available volume estimation keeping the accuracy, compared with traditional volumetric models, and maintaining compatibility properties between the total volume and the other parts [16, 20, 22, 23].

In general, estimators of volumetric and multi-volume models are obtained by regression techniques [6, 14, 17, 20, 24]. However, Artificial Neural Networks (ANNs) is an alternative that has been widespread in forest management to obtain tree volume estimates [5, 16, 25–33]. ANNs training provides linear and non-linear models based on mathematical estimates obtained identically to the logic of human reasoning [34].

Artificial intelligence tools have been increasingly adopted over the past 20 years to overcome problems related to the lack of conventional statistical assumptions, where foresters need to deal with noisy multidimensional data that are strongly non-linear [25, 35]. RNAs provide a specific approach for developing predictive models, offering a powerful method to analyze complex relationships between variables, without having to make assumptions about the data, which creates significant advantages over the use of conventional regression models [35]. Therefore several studies have shown advantages in employing ANNs in forestry measurement [26, 36–40].

Leite et al. [20] proposed a strategy for estimating multiple volumes of the stem that allows considering the presence or not of bark and different minimum commercial diameters
predictor variables. However, to this day, the present study is the first one to address the efficiency of ANNs to predict multi-volumes of the stem. This work aims to use different strategies to model multi-volumes of the stem of eucalyptus trees. The hypotheses of this study were: (1) the multi-volume single model proposed by Leite et al. [20] provides similar estimates to volume-specifics models, proposed by Schumacher and Hall [8], separately adjusted for each volume of interest (total and a merchantable volume, with and without bark); (2) the ANNs are efficient to estimate multi-volumes of the stem.

**Material and methods**

**Study area**

Data were obtained from four eucalyptus clones (AEC0144—*Eucalyptus urophylla*, AEC0224—*Eucalyptus urophylla*, VM01—*Eucalyptus urophylla × Eucalyptus camaldulensis* and H77—*Eucalyptus urophylla × Eucalyptus grandis*), grown at an average planting spacing of 3 x 3 m, in the high forest and coppice scheme (first and second rotation). In Brazil, these clones are used for several purposes. The plantations are located in Ribas do Rio Pardo, State of Mato Grosso do Sul, Brazil, at an average elevation of 415 m, in the geographic coordinates 20°27’15” S and 53°19’60” W. The climate is classified, according to Köppen, as Aw tropical, with annual temperature and rainfall averages of 24.1˚C and 1425 mm, respectively [41]. The silvicultural and dendrometric characteristics of the plantations are shown in Table 1.

**Data obtaining**

The data came from rigorous scaling on a 460 felled tree stems from 388 trees/stumps. Bark diameters (cm) were measured with a bevel gauge, at different tree heights: cutting height (±0.15 m); 0.4; 0.7; 1.0; 1.3; and 2.0 m, counted from ground level. From this point on, measurements were taken every two meters to the minimum commercial diameter with a 5 cm bark. The heights (m) at which the bark diameters of 10 cm, 15 cm, and 20 cm occurred were also obtained. The bark thickness (cm) was measured in a slice of the bark using a bevel gauge in each position of the stem that had the bark diameter measured. The diameter without bark (cm) was obtained by subtracting the bark diameter by multiplying twice the bark thickness.

The volumes between the diameters measured at the stem (sections), with and without bark, were obtained by the Smalian method. The volume of the stem tip was obtained by multiplying half of the sectional area in the 5 cm diameter by the tip length. The total volume of each tree, with and without bark, was obtained by summing the volume of the sections with

### Table 1. Average silvicultural and dendrometric characteristics of the evaluated plantings.

| Clone          | AEC0144 | AEC0144 | AEC0224 | AEC0224 | VM01 | VM01 | H77 | H77 |
|----------------|---------|---------|---------|---------|------|------|-----|-----|
| Rotation       | 1st     | 2nd     | 1st     | 2nd     | 1st  | 2nd  | 1st | 2nd |
| Planting age (years) | 6.8     | 13.6    | 6.6     | 9.5     | 6.3  | 12.7 | 14.5| 13.5|
| Sprouting conduction age (years) | -       | 6.0     | -       | 3.1     | -    | 5.7  | -   | 5.9 |
| DBH (cm)*     | 17.8±2.9| 14.6±3.5| 18.8±3.0| 13.7±2.6| 16.7±2.9| 13.4±5.3| 20.6±5.3| 16.4±4.8|
| Height (m)*   | 23.5±2.9| 23.7±4.0| 24.1±4.0| 18.3±2.5| 19.9±2.9| 19.1±5.7| 27.9±3.0| 19.2±3.9|
| N. stems by hectare* | 987±42  | 1703±93 | 933±30  | 1088±9 | 1049±59 | 1635±73 | 976±57  | 1067±212|
| N. stems by stump* | 1.00±0.00| 1.56±0.09 | 1.00±0.00 | 1.16±0.08 | 1.00±0.00 | 1.59±0.04 | 1.00±0.00 | 1.39±0.10|
| Basal area (m² ha⁻¹)* | 25.3±1.2 | 30.2±1.8 | 26.6±0.4 | 16.6±1.0 | 23.7±0.8 | 26.8±1.1 | 34.8±4.5 | 24.3±3.7|
| Volume (m³ ha⁻¹)* | 279.9±12.1 | 369.0±23.0 | 285.6±6.1 | 144.6±8.5 | 210.3±3.5 | 235.3±20.9 | 388.0±37.9 | 208.6±29.4|

*Mean ± standard deviation, derived from the measurement of five plots with ±500 m² in each planting.

https://doi.org/10.1371/journal.pone.0238703.t001
the volume of the tip. Commercial volumes with and without bark were also obtained from each tree to commercial bark diameters of 5 cm, 10 cm, 15 cm, and 20 cm, provided that the stem had a minimum log length of 1.0 m or more (Fig 1). Thus, up to 10 volumes were obtained for each tree, limited by the diameter of the trees.

Data analysis

Initially, the 460 tree stems rigorous scaling were randomly divided into two datasets, one intended for fitting regression models and ANN training (70%) and another for validation (30%). The first was composed of the different volumes obtained in 323 tree stems, totaling 2470 observed volumes. The validation was composed of volumes obtained in 137 tree stems, totaling 1048 observed volumes. Boxplots of the variables used in the fit and validation steps are shown in Fig 2. The data file used in each stage (training and validation) is available as Support Information.

These data were used for the modeling of the multiple volumes of the stem by different strategies: Schumacher and Hall [8] model fitting, separately adjusted for each volume of interest (total and a merchantable volume, with and without bark); multi-volume single model fitting proposed by Leite et al. [20]; and ANNs training to estimate multi-volumes of the stem.

Schumacher and Hall model

The Schumacher and Hall model [8] was fitted for each of the ten volumes separately, with the support of the R software, package nlme [42]. Among the many existing models for expressing wood volume as a function of diameter and height. The model proposed by Schumacher and Hall [8] is one of the most widespread in the forestry area due to its statistical properties since it almost always results in unbiased estimates [29, 37, 43].
The functional form of the Schumacher and Hall model is given by:

\[ V = \beta_0 \cdot DBH^{\beta_1} \cdot H^{\beta_2} + \varepsilon \]  

Wherein: \( V \) = volume (m\(^3\)); \( \beta_i \) = regression coefficient; \( DBH \) = diameter obtained at 1.3 m height from the ground level (cm); \( H \) = total height (m); \( \varepsilon \) = random error.

**Multi-volume model**

The fitted multi-volume model was proposed by Leite et al. [20], with the support of the R software, package nlme [42]. Whose functional form is:

\[ V = \beta_0 \cdot DBH^{\beta_1} \cdot H^{\beta_2} \cdot \exp(\beta_3 \cdot TX) \cdot DBH^{\beta_4} + \varepsilon \]  

Wherein: \( \beta_i \) = regression coefficients; \( DBH \) = diameter at breast height of 1.3 m ground level (cm); \( H \) = total height (m); \( TX \) = binary variable, which takes values 0 and 1; \( d \) = commercial diameter (cm); \( \varepsilon \) = random error. If \( TX = 0 \), the equation gives the volume with bark; if \( TX = 1 \), it provides the volume without bark. If \( d = 0 \), the adjusted equation provides the total volume. If \( d \) assumes any other value, the equation provides the volume up to the stipulated diameter.

Table 2 shows how the data were organized to fit the multi-volume model is shown.

**ANNs training**

The ANNs training was performed using the Intelligent Problem Solver tool from the Statistica 7.0 software [44]. This tool allows optimizing the ANN architecture by automatically setting the best number of neurons in the hidden layer and the best activation functions of the hidden and output layers, choosing the one with the least possible error, and is widely used by the scientific community [45–47].
For this reason, we used the most popular neural network layout, the Multilayer Perceptron (MLP). The main algorithm for MLP training is backpropagation (BP), which has been described in detail by Goh [48]. BP uses the momentum term to control the ANN learning rate. This momentum causes weight changes to be affected by the size of the previous weight changes that are used to avoid overfitting. The learning rate tells the network how slowly to progress. The weights are updated by a fraction of the calculated error each time to prevent the network from making large swings about the best values without ever getting it right [34]. The training of the MLP network by BP involves three stages [25, 49]: (i) the feedforward of the input training pattern, (ii) the calculation and backpropagation of the associated error, and (iii) the adjustment of the weights.

In this work, 1,000 MLP networks with one hidden layer containing a maximum of ten neurons were trained. The backpropagation training algorithm was used, and the logistic function performed the activations of the hidden and output layers. The input variables were DBH, H, TX, and d (previously defined), with data organized similarly to that adopted for multiple volumes (Table 2). The software retained the five best performing training ANNs.

### Assessing the fit quality of the regression and ANN training models

The quality of the estimates generated by the different strategies was analyzed by the following criteria: coefficient of determination—$R^2$ (Eq 3), root mean square error—RMSE(%) (Eq 4), mean absolute error—MAE(%) (Eq 5), mean bias error—MBE(%) (Eq 6), residual (%) (Eq 7), and graphical analysis for observed and estimated values and distribution of residuals [14, 35, 40, 50]. These analyses were performed using the R software [51] and Microsoft Excel.

Table 2. Data used for fitting the multi-volume model.

| V    | DBH | H  | TX | d |
|------|-----|----|----|---|
| TVb1 | DBH1| H1 | 0  | 0 |
| TVw1 | DBH1| H1 | 1  | 0 |
| V5b1 | DBH1| H1 | 0  | 5 |
| V5w1 | DBH1| H1 | 1  | 5 |
| V10b1| DBH1| H1 | 0  | 10|
| V10w1| DBH1| H1 | 1  | 10|
| V15b1| DBH1| H1 | 0  | 15|
| V15w1| DBH1| H1 | 1  | 15|
| V20b1| DBH1| H1 | 0  | 20|
| V20w1| DBH1| H1 | 1  | 20|
| ...  | ... | ...| ...| ...|
| V20w_n| DBH_n| H_n| 1  | 20|

$V = \text{volume (m}^3\text{)}$; DBH = diameter obtained at 1.3 m height from the ground level (cm); H = total height (m); TX = binary variable, which takes values 0 and 1; d = commercial diameter (cm); n = number of tree stem; TVb to V20w = Multi-volumes specified in Fig 1.

For this reason, we used the most popular neural network layout, the Multilayer Perceptron (MLP). The main algorithm for MLP training is backpropagation (BP), which has been described in detail by Goh [48]. BP uses the momentum term to control the ANN learning rate. This momentum causes weight changes to be affected by the size of the previous weight changes that are used to avoid overfitting. The learning rate tells the network how slowly to progress. The weights are updated by a fraction of the calculated error each time to prevent the network from making large swings about the best values without ever getting it right [34]. The training of the MLP network by BP involves three stages [25, 49]: (i) the feedforward of the input training pattern, (ii) the calculation and backpropagation of the associated error, and (iii) the adjustment of the weights.

In this work, 1,000 MLP networks with one hidden layer containing a maximum of ten neurons were trained. The backpropagation training algorithm was used, and the logistic function performed the activations of the hidden and output layers. The input variables were DBH, H, TX, and d (previously defined), with data organized similarly to that adopted for multiple volumes (Table 2). The software retained the five best performing training ANNs.
\[ MAE(\%) = \frac{100 \sum_{i=1}^{n} |y_i - \hat{y}_i|}{\bar{y}} \]  
\[ MBE(\%) = \frac{100 \sum_{i=1}^{n} (y_i - \hat{y}_i)}{\bar{y}} \]  
\[ \text{Residual}(\%) = \frac{(y_i - \hat{y}_i)}{y_i} \times 100 \]

Wherein: \( y_i \) = observed value of the \( i \)th variable; \( \hat{y}_i \) = estimated value of the \( i \)th variable; \( \bar{y} \) = observed mean of the variable; \( n \) = sample size.

**Validating equations and ANNs**

For validation of the generated equations and trained ANNs, estimates of multi-volumes were generated using the unused dataset for model fit and ANNs training. The quality of the estimates in the validation step was analyzed by the same criteria adopted for assessing the fit and training quality of the ANNs.

Additionally, to confirm the hypotheses tested, the t-test (\( \alpha = 0.05 \)) was performed between the volume obtained in the rigorous scaling and the estimated volume by each strategy with the validation data. In the case of ANNs, the t-test was performed for the best performing ANN in the validation step. This analysis was performed on Rbio software [52].

**Results**

**Schumacher and Hall equations**

Performance on volume estimates was variable by fitting the Schumacher and Hall volumetric model for each volume separately (Table 3, S1 and S2 Figs). The best estimates were obtained for total and commercial volume up to 10 cm diameter, which presented the best statistics.

| V      | \( \beta_0 \) | \( \beta_1 \) | \( \beta_2 \) | Fit       | Validation |
|--------|----------------|----------------|----------------|------------|------------|
|        | \( R^2 \)      | RSME (%)       | MAD (%)        | MBE (%)    | \( R^2 \)  | RSME (%)       | MAD (%)        | MBE (%)    |
| TVb    | 6.85 \times 10^{-5} | 1.4120         | 1.3260         | 0.970      | 10.66      | 7.29        | -0.54         | 0.971      | 10.34      | 6.89        | -0.15       |
| TVw    | 4.64 \times 10^{-5} | 1.4161         | 1.4014         | 0.969      | 11.10      | 7.72        | -0.54         | 0.971      | 10.68      | 7.18        | -0.18       |
| V5b    | 5.83 \times 10^{-5} | 1.4330         | 1.3520         | 0.969      | 11.14      | 7.78        | -0.72         | 0.970      | 10.77      | 7.36        | -0.02       |
| V5w    | 4.01 \times 10^{-5} | 1.4366         | 1.4245         | 0.968      | 11.53      | 8.18        | -0.71         | 0.970      | 11.04      | 7.56        | -0.28       |
| V10b   | 2.20 \times 10^{-5} | 1.6520         | 1.4200         | 0.947      | 14.41      | 10.71       | -1.19         | 0.949      | 14.62      | 10.78       | -1.09       |
| V10w   | 1.56 \times 10^{-5} | 1.6480         | 1.4910         | 0.946      | 14.66      | 10.96       | -1.18         | 0.951      | 14.66      | 10.80       | -1.07       |
| V15b   | 9.01 \times 10^{-7} | 2.4120         | 1.5770         | 0.873      | 25.38      | 20.17       | -2.15         | 0.869      | 25.33      | 19.68       | 0.14        |
| V15w   | 6.43 \times 10^{-7} | 2.4130         | 1.6420         | 0.873      | 25.76      | 20.51       | -2.16         | 0.874      | 25.38      | 19.68       | 0.21        |
| V20b   | 1.31 \times 10^{-8} | 4.1830         | 0.9340         | 0.747      | 34.34      | 27.81       | -1.34         | 0.768      | 38.56      | 27.85       | -1.96       |
| V20w   | 7.62 \times 10^{-9} | 4.2220         | 1.0230         | 0.746      | 35.02      | 28.37       | -1.34         | 0.774      | 38.89      | 27.81       | -1.76       |

Wherein: \( V \) = Volume; TVb to V20w = Multi-volumes specified in Fig 1; \( \beta_0, \beta_1, \beta_2 \) = model coefficients; \( R^2 \) = coefficient of determination; RMSE(\%) = root mean square error; MAE(\%) = mean absolut error; MBE(\%) = mean bias error.

https://doi.org/10.1371/journal.pone.0238703.t003
both in the model fitting and validation. Overall, the best quality of estimates was proportional to the decrease in the minimum commercial diameter.

The Schumacher and Hall equations demonstrated good performance in the graphical analysis of the multi-volumes jointly assessed in the fit and validation steps. The estimates showed a good distribution of residuals and were not biased, except for volumes smaller than 0.15 m³, which were overestimated (Fig 3A–3C). The recalculated fit and validation statistics for the multi-volumes jointly were adequate, with $R^2 = 0.947$, RMSE(%) = 15.29, MAE(%) = 11.10 and MBE(%) = -1.26 for fit, and $R^2 = 0.948$, RMSE(%) = 15.17, MAE(%) = 11.70, and MBE(%) = -0.02 for validation. These figures indicate that the strategy used was efficient in estimating multi-volumes, and this is supported by the higher distribution of residuals around zero for each multi-volume in the validation (Fig 3D). The distribution pattern is more influenced by the commercial diameters adopted than by the presence or not of bark in the stem. The

Fig 3. (A) Correlation between observed and estimated volume, (B) residual distribution, and (C) residual histogram for the fit and validation steps from the Schumacher and Hall equations. (D) Boxplot of residues for multi-volumes separately at the validation step. TVb to V20w = Multi-volumes specified in Fig 1.  
https://doi.org/10.1371/journal.pone.0238703.g003
volumes for the commercial diameters of 15 and 20 cm showed higher dispersion and a slight tendency to overestimate the values in the validation.

**Multi-volume equation**

Eq 8, from the multi-volume model fitting (Eq 2), performed well for the estimation of the total and commercial volumes (up to different minimum diameters), with and without bark (R² = 0.932; RMSE(%) = 17.26; MAE(%) = 12.28; MBE(%) = 0.41). When applying the equation to validation data, a behavior similar to that observed in the fit step was verified (R² = 0.938; RMSE(%) = 16.56; MAE(%) = 11.83; MBE(%) = 1.81), indicating that the equation is adequate to estimate the volume of parts of the eucalyptus stem through a single equation.

\[
V = 0.0001486 \cdot DBH^{1.0070} \cdot H^{1.3919} \left[ \exp\left(\frac{-2.5488}{DAP}\right) \right] \left[ 1 - \left(\frac{d}{DAP}\right)^{1.0+0.1529d} \right]
\]  

(8)

Overall, the equation provided good estimates in the fit and validation steps without bias in the estimates (Fig 4A–4C). At validation, the boxplot of the residuals for multi-volumes separately shows a higher distribution of residuals around zero (Fig 4D), which indicates that this strategy was also efficient. As for the Schumacher and Hall equations, the pattern of this distribution is more influenced by different commercial diameters adopted than by the presence or not of bark in the stem. A slight bias in overestimating total volumes and a higher bias in underestimating volumes for the 20 cm commercial diameter were observed. This fact revealed a higher dispersion between the observed and estimated values.

**ANNs**

ANNs have also been shown to be an efficient strategy for estimating multi-volumes of eucalyptus trees. The best performing ANNs had variable architecture in the number of neurons in the input layer (3–4) and the intermediate layer (4–7) (Table 4). The weight of the input variable followed the order: DBH > d > H > TX, where the TX variable does not influence the volume estimates in ANN 1. The exclusion of this variable from the input layer impaired the quality of the estimates generated by this network (S3, S4, S5 and S6 Figs), causing an increased in RSME (%) and MAD (%). Overall, all ANNs showed superior performance in the validation.

The ANN 3 presented statistics slightly better than the others. This network provided unbiased estimates (Fig 5A–5C) and less dispersion around zero for all volumes when compared with the estimates obtained by Schumacher and Hall and multi-volume equations (Figs 3 and 4), supporting the statistics obtained. In the validation, despite the ANN 3 superiority, the volume estimates for 20 cm diameters were less accurate than the volume up to the other assessed diameters, with a higher dispersion between the observed and estimated values (Fig 5D).

Regarding the other strategies, ANN 3 provided better graphic patterns.

**Efficiency of strategies in estimating multi-volumes**

Test t demonstrated that there were no significant differences between the actual volumes obtained in rigorous scaling and those estimated by the different strategies adopted in the validation (Fig 6). Overall, the standard deviation for the estimated values is similar to those observed for the actual volumes. This finding reveals that the estimated volumes are consistent with the actual volumes.
Discussion

These findings demonstrate that the strategies used for modeling different tree stem volumes (Schumacher and Hall volume-specific models, multi-volume model, and ANNs) were efficient. No significant differences were detected between the estimated volumes and the actual volume (Fig 6), and all strategies showed high accuracy and adequate graphic patterns. However, for larger commercial diameters, the estimates proved to be less accurate, with higher dispersion between observed and estimated values. One possible reason for this result is the higher irregularity in the form of the trunk base [6, 20, 53], which increases the natural volume variability in this portion of the stem and may weaken the relationship between the independent variable and the dependent variables. This fact impairs the predictive ability of the models. Therefore, when estimating these volumes, one should pay close attention to possible inconsistencies in the estimates, regardless of the modeling strategies evaluated.
Each of the strategies for estimating volumes has advantages and disadvantages in some situations. Adjusting a Schumacher and Hall volume-specific equation for each volume of interest separately improved the evaluated statistics and graphical standards when compared with the single multi-volume model proposed by Leite et al. [20]. However, this strategy would require the use of several equations (ten in this study), which can make it challenging to plan a forest enterprise, due to the cost and time spent adjusting the equations separately [16]. Thus, the forester should consider the benefit that would have more weight in forest planning, either greater accuracy or greater practicality.

In this sense, the ANNs proved to be useful for estimating the multi-volumes of eucalyptus trees simultaneously. The better performance verified for the ANN 3 in the validation (Table 4 and Fig 5) can be explained by the fact that they have simpler topologies between networks, with smaller numbers of neurons in the hidden layer. A reduced number of neurons in the hidden layer avoids the overfitting problem, which consists of over-learning the information contained in the data offered to ANNs [28]. ANNs can suffer from over-fitting, but the selection of suitable architecture may avoid this problem by using training and testing data sets [33]. Simpler topologies also facilitate the search and optimization of the configuration for a given task.

ANNs stood out for their higher accuracy in estimates when compared with the Schumacher and Hall volume-specific and multi-volume equations, although no statistical differences on average were detected. Regarding the setting of a Schumacher and Hall volume-specific model for each desired volume, the ANNs reduced the time required to obtain the estimates and the number of equations required. They also provided similar or superior accuracy to that found for the volumetric equations. In practice, accuracy, convenience, and cost-effectiveness play a crucial role when deciding which methods to utilize for forest inventory [33].

The ANNs are the best alternative to be used by companies or forestry enterprises due to their high potential to replace traditional volumetric models. This method maintains the precision in volume estimates and is convenient and efficient in obtaining results. It can help reduce inventory costs and time to make estimates available [30, 54, 55]. These results were already expected due to the several advantages shown by the ANNs, such as their massive and parallel-distributed structure (layers), the ability to learn and generalize, which enables them to solve complex problems and the fault and noise tolerance. In addition, no need to assume an underlying data distribution, as is usually done in statistical modeling; the possibility of modeling

| ANN | Arch | Input variables* | Training | Validation |
|-----|------|-----------------|----------|------------|
|     |      |                 |          |            |
|     |      |                 | $R^2$ (%) | RSME (%)  | MAD (%)   | MBE (%)  | $R^2$ (%) | RSME (%) | MAD (%)   | MBE (%)  |
| 1   | 3-5-1 | DBH, d, H       | 0.946    | 15.33     | 11.09     | 0.01     | 0.946     | 15.35     | 10.93     | 0.46     |
| 2   | 4-4-1 | DBH, d, H, TX   | 0.957    | 13.76     | 9.64      | -0.11    | 0.957     | 13.68     | 9.22      | 0.38     |
| 3   | 4-5-1 | DBH, d, H, TX   | 0.958    | 13.47     | 9.41      | -0.25    | 0.960     | 13.24     | 8.91      | 0.41     |
| 4   | 4-6-1 | DBH, d, H, TX   | 0.958    | 13.49     | 9.48      | -0.20    | 0.959     | 13.40     | 8.96      | 0.38     |
| 5   | 4-7-1 | DBH, d, H, TX   | 0.958    | 13.52     | 9.54      | -0.05    | 0.958     | 13.51     | 9.17      | 0.33     |

$n = 10$; $R^2 = \text{coefficient of determination};$ RMSE(%) = root mean square error; MAE(%) = mean absolut error; MBE(%) = mean bias error.

https://doi.org/10.1371/journal.pone.0238703.t004

Table 4. Characteristics of the best performing artificial neural networks (ANNs) and statistics obtained in the training and validation to estimate multi-volumes of eucalyptus trees.
several variables and their non-linear relationships; the possibility of modeling using categorical variables, besides quantitative variables; and neurobiological analogy [33–35, 55].

Therefore, among the multi-volume modeling strategies evaluated here, the use of ANNs is the most promising. Reliable and accurate volume estimating is essential for the forestry company to make a correct assessment of the wood stock and to analyze the productive potential of a forest for multiple uses [4, 7, 56]. The optimization of the techniques used to obtain volumetric estimates has been increasingly frequent in sustainable forest management. Thus, the development of methods to estimate multi-volumes along the tree stem is necessary by combining precise and practical strategies.

However, the application of ANNs to practical forestry is still immature [33]. Using artificial intelligence (AI) techniques demands much training time and can easily incur data overfitting. Furthermore, the most critical decision support systems in forestry are not yet able to
handle AI [35, 57]. The adoption of multi-volume ANNs in wood inventories and management plans in other regions of Brazil for other species and silvicultural conditions is made possible by the continuous training of new ANNs as new data are obtained. Depending on the nature of the data (broader scale), a single ANN may be used with efficiency and accuracy for different regions. Therefore, multi-volume ANNs should be the subject of future research assessing other experimental conditions and species/clones, different ANN architectures, or including other quantitative variables and categorical (qualitative) variables.

Conclusion

The multi-volume model had the most considerable advantage in volume estimation practicality, while the volume-specific models were more efficient in the accuracy of estimates. Under the conditions tested here, the ANNs are more suitable than regression models in the estimation of multi-volumes of eucalyptus trees, revealing higher accuracy and practicality.
Supporting information

S1 Fig. Correlation between observed and estimated volume, residual distribution, and residual histogram for the fit and validation steps from the Schumacher and Hall equations. Boxplot of residues for multi-volumes separately at the validation step. TVb to V10w = Multi-volumes specified in Fig 1. (JPG)

S2 Fig. Correlation between observed and estimated volume, residual distribution, and residual histogram for the fit and validation steps from the Schumacher and Hall equations. Boxplot of residues for multi-volumes separately at the validation step. V15b to V20w = Multi-volumes specified in Fig 1. (JPG)

S3 Fig. Correlation between observed and estimated volume, residual distribution, residual histogram for fit and validation steps and boxplot of residues for multi-volumes separately in the validation step from the ANN1. TVb to V20w = Multi-volumes specified in Fig 1. (JPG)

S4 Fig. Correlation between observed and estimated volume, residual distribution, residual histogram for fit and validation steps and boxplot of residues for multi-volumes separately in the validation step from the ANN2. TVb to V20w = Multi-volumes specified in Fig 1. (JPG)

S5 Fig. Correlation between observed and estimated volume, residual distribution, residual histogram for fit and validation steps and boxplot of residues for multi-volumes separately in the validation step from the ANN4. TVb to V20w = Multi-volumes specified in Fig 1. (JPG)

S6 Fig. Correlation between observed and estimated volume, residual distribution, residual histogram for fit and validation steps and boxplot of residues for multi-volumes separately in the validation step from the ANN5. TVb to V20w = Multi-volumes specified in Fig 1. (JPG)

S1 Table. Data file used in training stage for fit of models used in this manuscript. (XLSX)

S2 Table. Data file used in validation stage for fit of models used in this manuscript. (XLSX)

Acknowledgments

The authors thank Âmbar Energia Ltda.—P&D ANEEL Program—for the financial support to carry out data collection in the field; UFMS (Federal University of Mato Grosso do Sul), CAPES—Financial Code 001 (Brazilian Federal Agency for Support and Evaluation of Graduate Education) and CNPq (National Council for Scientific and Technological Development) for their financial support.
Author Contributions

Conceptualization: Heitor Vicensotto Tomiazzi, Glauce Taís de Oliveira Sousa Azevedo.

Data curation: Gileno Brito de Azevedo.

Formal analysis: Gileno Brito de Azevedo, Glauce Taís de Oliveira Sousa Azevedo, Larissa Pereira Ribeiro Teodoro, Paulo Eduardo Teodoro.

Methodology: Heitor Vicensotto Tomiazzi.

Software: Paulo Eduardo Teodoro.

Supervision: Gileno Brito de Azevedo.

Writing – original draft: Gileno Brito de Azevedo, Heitor Vicensotto Tomiazzi, Glauce Taís de Oliveira Sousa Azevedo.

Writing – review & editing: Gileno Brito de Azevedo, Larissa Pereira Ribeiro Teodoro, Paulo Eduardo Teodoro, Marcos Talvani Pereira de Souza, Tays Silva Batista, Humberto de Jesus Eufrade-Junior, Saulo Philipe Sebastião Guerra.

References

1. MAPA/SFB. Brazil. Ministry of Agriculture, Livestock, and Food Supply. Brazilian forest at a glance. Ministry of Agriculture, Livestock, and Food Supply. Brazilian Forest Service; 2019.

2. IBÁ. Relatório 2019. 2019. Available: https://iba.org/datafiles/publicacoes/relatorios/iba-relatorioannual2019.pdf

3. Azevedo GB, Sousa GTO, Barreto PAB, Conceição Júnior V. Estimativas volumétricas em povoaamentos de eucalipto sob regime de alto fuste e talhadia no sudoeste da Bahia. Pesqui Florest Bras. 2011; 31: 309–318. https://doi.org/10.4336/2011.pfb.31.68.309

4. Özçelik R, Brooks JR. Compatible volume and taper models for economically important tree species of Turkey. Ann For Sci. 2012; 69: 105–118. https://doi.org/10.1007/s13595-011-0137-4

5. Binoti DHB, Binoti MLM da S, Leite HG. Configuração de redes neurais artificiais para estimativa do volume de árvores. Ciência da Madeira. 2014; 5: 58–67. https://doi.org/10.15210/cmad.v5i1.4067

6. Miguel E, Péllico Netto S, Azevedo G, Azevedo G, Rezende A, Pereira R. Alternative methods of scaling Eucalyptus urophylla trees in forest stands: compatibility and accuracy of volume equations. iForest—Biogeosciences For. 2018; 11: 275–283. https://doi.org/10.3832/ifor2155-011

7. Liu J, Wang X, Wang T. Classification of tree species and stock volume estimation in ground forest images using Deep Learning. Comput Electron Agric. 2019; 166: 105012. https://doi.org/10.1016/j.compag.2019.105012

8. Schumacher FX, Hall FS. Logarithmic expression of timber-tree volume. J Agric Res. 1933; 47: 719–734.

9. Andrade VCL. Modelos volumétricos de dupla entrada para aplicar em povoaamentos florestais brasileiros. Floresta e Ambient. 2017; 24: 1–10. https://doi.org/10.1590/2179-8087.135415

10. Oliveira MLR de Leite HG, Garcia SLR Campos JCC, Soares CPB Santana RC. Estimação do volume de árvores de clones de eucalipto pelo método da similaridade de perfilis. Rev Árvore. 2009; 33: 133–141. https://doi.org/10.1590/S0100-67622009000100014

11. Silva F da, Corte APD, Sanquetta CR. Equações de afilamento para descrever o volume total do fuste de Pinus caribaea var. hondurensis na região do Triângulo Mineiro. Sci Agrar Parana. 2016; 15: 367–376. https://doi.org/10.1016/j.sagr.2016.07.007

12. Figueiredo EO, D’Oliveira MVN, Braz EM, de Almeida Papa D, Fearnside PM. LIDAR-based estimation of bole biomass for precision management of an Amazonian forest: Comparisons of ground-based and remotely sensed estimates. Remote Sens Environ. 2016; 187: 281–293. https://doi.org/10.1016/j.rse.2016.07.026

13. Martins APM, Carvalho DE, Schmidt LN, Gorenstein MR, Cysneiros VC. Eficiência de métodos para estimativa do volume total de Corymbia citriodora. Sci Agrar Paraana. 2016; 15: 178–184. https://doi.org/10.1818/sap.v15i2.13203

14. Khan MN, Shill MC, Azad MS, Sadath MN, Feroz SM, Mollick AS. Allometric relationships of stem volume and stand level carbon stocks at varying stand density in Swietenia macrophylla King plantations, Bangladesh. For Ecol Manage. 2018; 430: 639–648. https://doi.org/10.1016/j.foreco.2018.09.002
15. Miguel EP, Leal FA, Ono HA, Leal UAS. Modelagem na predição do volume individual em plantio de *Eucalyptus urograndis*. Rev Bras Biometria. 2014; 32: 584–598.

16. Ribeiro RB da S, Gama JRV, Souza AL de, Leite HG, Soares CPB, Silva GF da. Métodos para estimar o volume de fustes e galhos na Floresta Nacional do Tapajós. Rev Árvore. 2016; 40: 81–88. https://doi.org/10.1590/0100-67622016000100009

17. Corral-Rivas J, Vega-Nieva D, Rodríguez-Solís J, López-Sánchez C, Wehenkel C, Vargas-Larreta B, et al. Compatible system for predicting total and merchantable stem volume over and under bark, branch volume and whole-tree volume of pine species. Forests. 2017; 8: 417. https://doi.org/10.3390/f18100417

18. Arias-Rodríguez M, Castedo-Dorado F, Cámara-Obregón A, Diéguez-Arandia U. Fitting and Calibrating a Multilevel Mixed-Effects Stem Taper Model for Maritime Pine in NW Spain. Reigosa M, editor. PLoS One. 2015; 10: e0143521. https://doi.org/10.1371/journal.pone.0143521 PMID: 26630156

19. Tang X, Pérez-Cruzo C, Fehrmann L, Álvarez-González JG, Lu Y, Kleinn C. Development of a Compatible Taper Function and Stand-Level Merchantable Volume Model for Chinese Fir Plantations. Wu R, editor. PLoS One. 2016; 11: e0147610. https://doi.org/10.1371/journal.pone.0147610 PMID: 26799399

20. Leite HG, Guimarães DP, Campos JCC. Descrição de um modelo para estimar múltiplos volumes de árvores. Rev Árvore. 1995; 19: 65–79.

21. Silva GF, Campos JCC, Souza AL, Leite HG. Uso de métodos estatísticos para comparar alternativas de estimativa do volume comercial. Rev Árvore. 1997; 21: 59–70.

22. Campos JC, Leite HG, Silva GF, Soares CPB, Carneiro JA. Estimação de volumes de tronco e da copa de árvores de povoamentos mistos. Rev Árvore. 2001; 25: 223–230.

23. Silva GF, Campos JCC, Leite HG. Alternativas para estimar o volume comercial em árvores de eucalipto. Rev Árvore. 1996; 20: 467–481.

24. Sanquetta CR, Dolci M, Corte APD, Sanquetta MNI, Pelissari AL. Estimativa de volumes de *Araucaria angustifolia* (Bertol.) O. Kuntze por fatores de forma em classes diamétricas e modelos de regressão. Enciclópédia Biosf. 2016; 13: 588–597.

25. Socha J, Netzel P, Cywicka D. Stem Taper Approximation by Artificial Neural Network and a Regression Set Models. Forests. 2020; 11: 79. https://doi.org/10.3390/f11010079

26. Silva MLM, Binoti DHB, Gierrani JM, Leite HG. Ajuste do modelo de Schumacher e Hall e aplicação de redes neurais artificiais para estimar volume de árvores de eucalipto. Rev Árvore. 2009; 33: 1133–1139.

27. Cordeiro MA, Pereira NN de J, Binoti DHB, Binoti MLM da S, Leite HG. Estimativa do volume de *Acacia mangium* utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. Pesqui Florest Bras. 2015; 35: 255. https://doi.org/10.4336/2015.pfb.35.83.596

28. Leal FA, Miguel EP, Matricardi EAT, Pereira RS. Redes neurais artificiais na estimativa de volume em um plantio de eucalipto em função de fotografias hemisféricas e número de árvores. Rev Bras Biometria. 2015; 33: 233–249.

29. Schikowski AB, Corte APD, Ruza MS, Sanquetta CR, Montaño RANR. Modeling of stem form and volume through machine learning. An Acad Bras Cienc. 2018; 90: 3389–3401. https://doi.org/10.1590/0011-3765201800100009

30. Tavares Júnior I, Rocha J, Ebling Á, Chaves A, Zanuncio J, Farias A, et al. Artificial neural networks and linear regression reduce sample intensity to predict the commercial volume of *Eucalyptus clonae*. Forests. 2019; 10: 268. https://doi.org/10.3390/f10030268

31. Silva JPM, Silva MLM, Silva EF, Silva GF, Mendonça AR, Cabacinha CD, et al. Computational techniques applied to volume and biomass estimation of trees in Brazilian savanna. J Environ Manage. 2019; 249: 109368. https://doi.org/10.1016/j.jenvman.2019.109368 PMID: 31421480

32. Çatal Y, Sapiloğlu K. Comparison of adaptive neuro-fuzzy inference system, artificial neural networks and non-linear regression for bark volume estimation in brutian pine (*Pinus brutia* Ten.). Appl Ecol Environ Res. 2018; 16: 2015–2027. https://doi.org/10.15666/aeer/1602_20152027

33. Özçelik R, Diamantopoulou MJ, Brooks JR, Wiart H V. Estimating tree bole volume using artificial neural network models for four species in Turkey. J Environ Manage. 2010; 91: 742–753. https://doi.org/10.1016/j.jenvman.2009.10.002 PMID: 20180241

34. Haykin S. Neural Network: A Comprehensive Foundation. 2nd ed. Ontario Canada: McMaster University; 2004.

35. Nunes MH, Gorgens EB. Artificial Intelligence Procedures for Tree Taper Estimation within a Complex Vegetation Mosaic in Brazil. Hernandez Montoya AR, editor. PLoS One. 2016; 11: e0154738. https://doi.org/10.1371/journal.pone.0154738 PMID: 27187074
36. Gorgens EB, Leite HG, Santos HN, Gleriani JM. Estimação do volume de árvores utilizando redes neurais artificiais. Rev Árvore. 2009; 33: 1141–1147.
37. Dolácio CJF, de Oliveira TWG, Oliveira RS, Cerqueira CL, Costa LRR. Different approaches for modeling Swietenia macrophylla commercial volume in an Amazon agroforestry system. Agrofor Syst. 2019. https://doi.org/10.1007/s10457-019-00468-3
38. Leite HG, Binoti DHB, Guimarães DP, Silva MLM, Garcia SLR. Avaliação do ajuste das funções Weibull e hiperbólica a dados de povoamentos de eucalipto submetidos a desbaste. Rev Árvore. 2010; 34: 305–311.
39. Freitas ECS de, Paiva HN de, Neves JCL, Marcatti GE, Leite HG. Modeling of eucalypt productivity with artificial neural networks. Ind Crops Prod. 2020; 146: 112149. https://doi.org/10.1016/j.indcrop.2020.112149
40. Vieira GC, de Mendonça AR, da Silva GF, Zanetti SS, da Silva MM, dos Santos AR. Prognoses of diameter and height of eucalyptus trees using artificial intelligence. Sci Total Environ. 2018; 619–620: 1473–1481. https://doi.org/10.1016/j.scitotenv.2017.11.138 PMID: 29734623
41. Climate-D ata.ORG. Clima: Ribas do Rio Pardo. 2019 [cited 28 Aug 2019]. Available: https://pt.climatedata.org/am erica-do-sul/brasil/mato-grosso-do-sul/ribas-do-rio-pardo-43544/
42. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–148. 2020. p. 336. Available: https://svn.r-project.org/R-packages/trunk/nlme
43. Campos JCC, Leite HG. Mensuração florestal: perguntas e respostas. 5th ed. Viçosa-MG: Editora UFV; 2017.
44. Statsoft I. Statistica (data analysis software system), version 7.0. 2007.
45. Binoti MLM da S, Leite HG, Binoti DHB, Gleriani JM. Prognoses em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. CERNE. 2015; 21: 97–105. https://doi.org/10.1590/01047760201521011153
46. Vale A do, Miguel EP, Moreira AC de O, Lima CM, Orellana BBMA, Fortes MM, et al. Artificial neural networks in predicting energy density of Bambusa vulgaris in Brazil. African J Agric Res. 2017; 12: 856–862. https://doi.org/10.5897/AJAR2016.12083
47. Gitoee A, Faridi A, Franco J. Mathematical models for response to amino acids: estimating the response of broiler chickens to branched-chain amino acids using support vector regression and neural network models. Neural Comput Appl. 2018; 30: 2499–2508. https://doi.org/10.1007/s00521-017-2842-x
48. Goh ATC. Back-propagation neural networks for modeling complex systems. Artif Intell Eng. 1995; 9: 143–151. https://doi.org/10.1016/0954-1810(94)00011-S
49. Whittington JCR, Bogacz R. Theories of Error Back-Propag ation in the Brain. Trends Cogn Sci. 2019; 23: 235–250. https://doi.org/10.1016/j.tics.2018.12.005 PMID: 30704969
50. Zeng W, Zhang L, Chen X, Cheng Z, Ma K, Li Z. Construction of compatible and additive individual-tree biomass models for Pinus tabulaeformis in China. Can J For Res. 2017; 47: 467–475. https://doi.org/10.1139/cjfr-2016-0342
51. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available: https://www.r-project.org/.
52. Bhering LL. Rbio: A tool for biometric and statistical analysis using the R platform. Crop Breed Appl Biotechnol. 2017; 17: 187–190.
53. Téo SJ, Marcon A, Ehlers T, Bianchi JC, Peloso A, Nava PR, et al. Modelos de afilamento para Pinus elliottii em diferentes idades, na região de Caçador, SC. Floresta. 2013; 43: 439. https://doi.org/10.5380/f.v43i3.30320
54. Binoti DHB, Binoti MLM., Leite HG, Silva A. Redução dos custos em inventário de povoamentos equiaênios. Rev Bras Ciências Agrárias. 2013; 8: 125–129. https://doi.org/10.5039/agraria.v8i1a2209
55. Chiarello F, Steiner MTA, Oliveira EB, Arce JE, Ferreira JC. Artificial neural networks applied in forest biometrics and modeling: state of the art (January/2007 to July/2018). Cerne. 2019; 25: 140–155. https://doi.org/10.1590/01047760201925022626
56. Ribeiro RB da S, Gama JRV, Melo L de O. Seccionamento para cubagem e escolha de equações de volume para a Floresta Nacional do Tapajós. Cerne. 2014; 20: 605–612. https://doi.org/10.1590/01047760201420041400
57. Aertsens W, Kint V, van Orshoven J, Özkan K, Muys B. Comparison and ranking of different modelling techniques for prediction of site index mountain forests. Ecol Model. 2010; 221: 1119–1130. https://doi.org/10.1016/j.ecolmodel.2010.01.007