TreeDimensional Journal, Vol. 11, e2023037, 1-10, 2023
Original Article | **DOI:** doi.org/10.55746/treed.2023.12.037

Combining Digital Imaging and Artificial Neural Networks for Individual Volumetric Estimation of Non-Straight Trees

Pedro Henrique C. de Sousa[1,*] , Hallefy Junio de Sousa[2,*] , Matheus P. Chagas[1] , Daniel Henrique B. Binoti[3] , Fábio Venturoli[1] , Rafael T. Resende[2,4]

Abstract
Tree volume models have become an important decision-making tool in forest resources management. The objective of this work was to create Artificial Neural Networks (ANN) to fit individual tree models from the Brazilian Savanna (the Cerrado biome), characterized by non-straight and very bifurcated tree stems. Data were collected in two different areas and through different methods: *i*) individual tree volume by digital photographs, and also by tree climbing measurements (both obtained in a non-destructive way); *ii*) tree volumes measured destructively by the traditional rigorous cubing procedure after thinning. Data obtained from climbing measurements were compared with the corresponding tree obtained from digital photographs, resulting in a reasonable operational linear relation of 0.80. Ten ANN structures were created with different combinations of forestry variables. Five ANN structures depicted predictive ability results above 0.80. Then, the results obtained by ANN were compared with those obtained by classic tree volumetric models, once again showing the predictive superiority of ANN. The present study indicates that tree volume obtained from digital photographs, achieved by popular mobiles, can be safely used for the estimation of very-non-straight trees from the Brazilian Savanna and demonstrated the greater predictive ability of ANN in relation to the classic regression-based volumetric models in terms of lower modeling bias.

Keywords
Biometrics — Modeling — Precision forestry — Brazilian Savanna — Individual Tree Volume — ANN

[1] *Universidade Federal de Goiás (UFG), Escola de Agronomia / Setor de Engenharia Florestal, Goiás, Brazil*
[2] *Universidade de Brasília (UnB), Departamento de Engenharia Florestal, Posto Ecológico, 70.297-400, Brasília – DF, Brazil*
[3] *Eldorado Celulose, BR158 road, km 231, 79.641-300, Três Lagoas – MS, Brazil*
[4] *Universidade Federal de Goiás (UFG), Escola de Agronomia / Setor de Melhoramento de Plantas, Goiás, Brazil*
*__Corresponding authors__: pedrohcvsousa@gmail.com; hallefyj.souza@gmail.com

1. Introduction

The demand for new technologies in forestry modeling is continuous. Part of it is due to the technology's needs that meet the estimation of tree variables determining the wood stock in forest stands. The efficiency of timber resource exploitations depends on the knowledge about tree growing and harvesting stock, which can be obtained by measuring and/or estimating the tree's traits (Binoti et al. 2012). Besides that, the ecological influence of functional traits on species persistence in the Brazilian Cerrado biome remains incompletely understood (De Souza et al. 2022). Forest site productivity can be understood as the potential of a particular forest stand to produce aboveground wood volume, referring to the production unit composed of the site and tree stand (Skovsgaard and Vanclay 2008). In that sense, tree volume is an important basis for estimating biomass, carbon storage, and other factors of interest in forests (Liu et al. 2019), helping the management plan development (Kershaw Jr et al. 2017; Liu et al. 2019) and in the definition of policies and public strategies (Scolforo et al. 2015).

Recently, prediction techniques with Artificial Neural Networks (ANN) and satellite images have been worldwide used as an option for tree stand modeling prediction (Almeida et al. 2014). Considering the increase in current computational capacity, ANN has become a more interesting option due to its adaptation to different forest data formats (planted or native) and unexpected variable behaviors (de Azevedo, *et al.*, 2020; Wang et al., 2023). The most used ANN for forest measurements are supervised ones, since the solution to a given problem occurs through the repetition of a set of network training. McCann et al. (2017) explain that the training of an ANN aims to reconstruct an original vector and create a pattern.

Received: 2023/10/17 | **Accepted:** 2023/12/11 | **Online published:** 2023/12/12

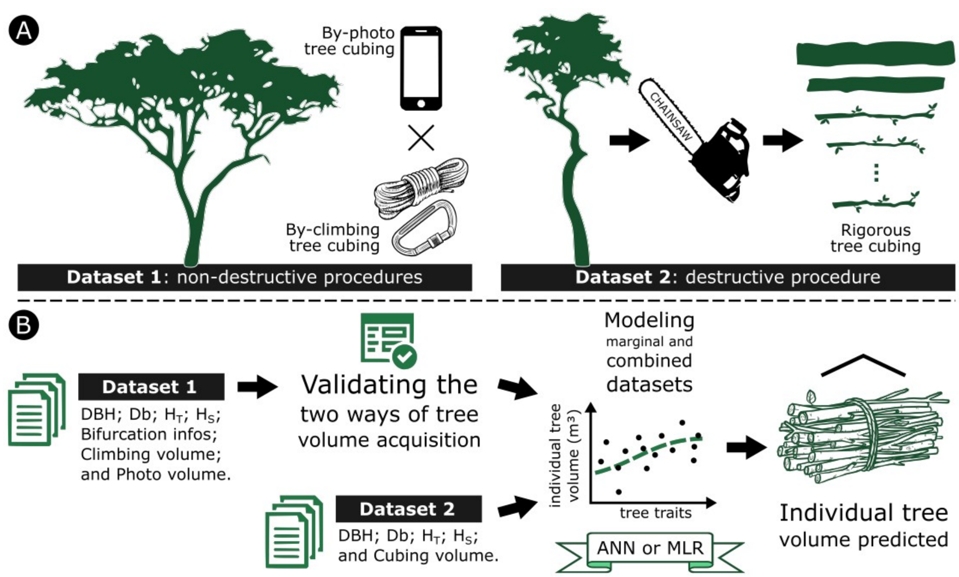

Figure 1. Schematic steps elucidating the Materials and Methods. Part A) depicts the two datasets (G1 and G2 respectively), which in the first, tree volumes were obtained by two non-destructive procedures, one by photo mobiles and the other by climbing; and in the second procedure, tree volumes were obtained by traditional rigorous tree cubing. Part B) depicts the analysis performed, wherein the two datasets were evaluated marginally and combined aiming for volumetric predictions. Please refer to the following for acronyms meanings.

According to Castellanos et al. (2007), the application of neural networks in biometric techniques, forest inventory, and, forest management allows for greater precision in production estimates, allowing for different decision-making. This computational tool used in biometrics and forest inventory is mostly performed with even-aged forest data, improving accuracy in production estimation (Lacerda et al. 2017). In natural forests, parameters such as height and diameter at breast height can be tightly influenced by stem shape (Husch et al. 2002), making volumetric estimation models a more complex task.

In this context, information on volumetric estimates on Savannas biomes is still not completely understood, and this fact is probably due to its great species diversity, high variability among individuals within species and across ages, as well the great variation in the shape of tree stem and the crown (Rezende et al. 2006; Colli et al. 2020). The volumetric estimation models in the Brazilian Savanna commonly use information such as the diameter at breast height (DBH), total height (H_T), or base diameter (Db) through regression models (Rezende et al. 2006; Rufini et al. 2015) and, more recently, through artificial intelligence (Lacerda et al. 2017). In general, its prediction is made by using mathematical tools, such as allometric equations, which are still scarce and, for some of those equations, the error distributions have not yet been reported, making it impossible to assess the bias or determine whether the assumptions of the homoscedasticity and error normality regression analysis were met (Roitman et al. 2018).

According to Nicoletti et al. (2015), destructive measurement makes forest measurement activities longer and more expensive. For the case of individuals with several bifurcations, this becomes an even more onerous job, as it demands longer collection time and a larger number of data to be worked on. As an alternative, non-destructive tree measurements can be interesting, due it is less time-consuming, less labor, and have reasonable reliability. Besides that, especially in biomes with some degree of degradation, and/or featured biomes having sparse plant densities, such as World savannas, it is lawly illegal to thin trees and then measures their biomass. To bypass it, we can observe some non-destructive, but highly expensive, measurement technologies being proposed for this purpose, which is great but something unrealistic for underdeveloped countries belonging to the Savannas (Brady 2020). We can cite some technologies being applied, for example, photogrammetry (Berveglieri et al. 2017), three-dimensional laser scanning (Puliti et al. 2020), and also a method using an electronic theodolite (Liu et al. 2019), however, most of them require high cost and professional training.

Digital-image-based tree measurement has been tested by other researchers (Juujarvi et al., 1998; Dean, 2003; Han e Dong, 2011) with relative accuracy, safeguarding the inherent characteristics of each evaluated forest formation. When combined with techniques such as ANN, can provide a low-cost, expeditious, and safe methodology to obtain forestry variables and fit models for volumetric estimation. This work aims to verify the viability of tree data obtained through the combination of free software usage and digital images of tree individuals in the field, in addition to evaluating the efficiency of using artificial neural networks and classical literature forest volumetric models to adjust volumetric measurement models in an area of native vegetation in the Brazilian Cerrado.

2. Material and Methods

2.1 Sampling sites description

Two dataset sources were used in the elaboration of this work, named G1 and G2, as presented in Table 1. To facilitate the understanding of the procedures adopted in this work, a comprehensive description of this topic was also graphically summarized in Figure 1-A and 1-B.

Table 1. Location, climate classification, parameters, and predominant soil of the two datasets areas

	Location	
	G1	G2
Municipality	Goiânia	Nova Crixás
State / Country	Goiás / Brazil	Goiás / Brazil
Geographic coordinates	16° 35' 42" S; 49° 16' 45" W	14° 05' 52" S; 50° 20' 17" W
Köppen-Geiger climate classification	Aw (Tropical savanna, wet)	Aw (Tropical savanna, wet)
Cerrado phytophysiognomies	Riparian Forest	Transition Zone between Dense Cerrado and Cerradão
Mean annual temperature (ºC)	23.5	26.5
Mean annual precipitation (mm)	1700	1500
Predominant soil	Red-Yellow Latosol	Plinthosol
Source	Cardoso et al. (2014)	Santana et al. (2016)

2.2 Data collections in non-destructive and destructive approaches

Relevant traits of 50 tree individuals were collected, composing the dataset 1 (G1 group), as follows: DBH (Diameter at Breast Height, assumed to be 1.30 m from the ground), and Db (Base diameter), both with the use of measuring tape; the H_T (total height of the individuals) and the H_S (height of the stem before bifurcating) using a *Haglöf EcII* clinometer; and the NS (number of stems at the height of the DBH). When necessary, a four-meter ladder was used to obtain measurements on higher branches. In cases where there was more than one shaft at the height of the DBH, the equivalent diameter, in centimeters, was calculated from the equation:

$$\mathrm{DBH_{EQ}} = \sqrt{\textstyle\sum_{i=1}^{n} \mathrm{DBH}_i^2 / n},$$

where "DBH_i" represents the individual DBH and "n" represents the number of stems. In addition, digital photographs of all these individuals' trees were taken, containing two 50 cm calibration rods (one at the base and the other at the top of the individual), to calibrate the measurements made in the ImageJ software (Schindelin et al. 2012). It's quite easy to use ImageJ, just follow the steps: load the tree photo in the app by doing <File/Open>; select the option <Straight> and set the real calibration rod value; enter the tab <Analyze/Set Scale> and declare the measurements as "Known Distance", which can be in meters or centimeters; after calibration, collect measurements from the whole tree sections by simply clicking and dragging. The results are obtained from the tab <Analyze/Measure>. Figure 2-A illustrates how the photographs and measurements were taken in the ImageJ software.

All those previous described tree individuals had their volume measured twice, besides directly in the field, through the photos and the ImageJ software; they were also measured in a by-climbing non-destructive way, using the Huber' method (de León and Uranga-Valencia 2013), described in the equation:

$$Vf = \textstyle\sum_{i=0}^{n} As \times L,$$

in which Vf represents the individual's final volume, in cubic meters (m³), L represents the section length, in meters (m) and As represents the section area, in square meters (m²), given by:

$$As = \frac{\pi \times (D_{\frac{1}{2}})}{4},$$

which $D_{\frac{1}{2}}$ represents the diameter at half the length of the section, in meters. The results obtained from the two measurements were compared through correlation test and simple linear adjustment, presenting the value of R^2 (Figure 1-B).

Sequentially, volume measurements were conducted on the remaining 63 trees within dataset 2 (G2 location), following a thinning process, utilizing both classical rigorous cubing procedures as outlined by Cerqueira et al., (2021). Although the procedure used by Cerqueira et al., (2021) was applied to eucalyptus, it is relevant to emphasize that it will function technically equally for any other species. Notably, the de Huber's method group underwent measurements without the assistance of photo-

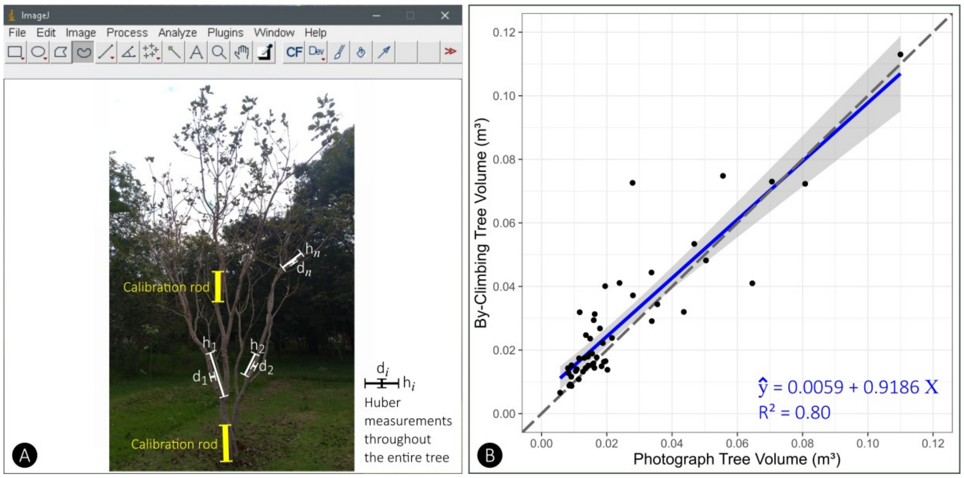

Figure 2. Representation of the volume measurement via digital photography analysis, applying ImageJ software. A) Illustration of photographs; in yellow the representation of the calibration rods of 50 cm each. B) Simple linear regression between the by-climbing volume (ordered axis) and the volume by digital photography (abscissa axis), in cubic meters. The black dashed line is the 45° diagonal line; The blue line is the linear model fit.

graph support. Within this G2 subgroup, comprehensive tree metrics including DBH, Db, H_T, and H_S were also measured. This was done to ensure alignment with the metrics recorded for the initial tree group, as illustrated in Figure 1B.

2.3 Non-straight individual tree volume modeling: ANN × classical regressions

The two groups of data were compared to each other using Multiple Linear Regressions (MLR) and Artificial Neural Networks (ANN) procedures (Figure 2-B). The ANN structures were created by the NeuroForest software (Binoti et al. 2013) to generate volumetric models so that both groups were used for training and for validation of these models, to test the predictive ability between them. The predictive ability $r_{y\hat{y}}$ between the real data and those estimated by the ANNs, in addition to the variance ($S^2_{y\hat{y}}$), standard deviation ($S_{\bar{y}}$) and the RMSE% (the Root Mean of Squared Error, which corresponds to a measure of *bias* between predicted and observed data), a metric used to assess model errors.

The complete set of data was used for modeling through ANN. Here, the 10-fold validation methodology was used, in which the data group was separated into five groups of random individuals (20% of the total sample), to validate the models generated by 80% of the remaining individuals, resulting in three groups with 23 individuals and two groups with 22 individuals. From these results, the predictive ability between real and estimated data, variance, and standard deviation were calculated, besides being submitted to the RMSE%. The models 01 to 09 used with ANN the approach are shown in Table 2.

A 10th ANN model was used only including the G1 trees, due to this dataset having information on bifurcation not completely shown in the G2 dataset. That model (ANN10) incorporates besides DBH_{eq}, Db, H_T, and H_S the number and height of bifurcations.

Table 2. Description of models and variables used in ANNs for volumetric prediction (output variable for all models) and the comparison in the validation process between Groups G1 and G2 (Now using G1 as training group and G2 as validation, G1 → G2, and G2 → G1)

Model	Input	Predictive abilities		RMSE (%)	
	Variable	G1 → G2	G2 → G1	G1 → G2	G2 → G1
ANN01	DBH_{eq}, H_T	0.732	0.692	8.00%	18.30%
ANN02	DBH_{eq}, H_S	0.696	0.863	14.60%	9.90%
ANN03	Db, H_T	0.807	0.435	10.20%	23.60%
ANN04	Db, H_S	0.802	0.862	9.50%	27.60%
ANN05	DBH_{eq}, Db, H_S	0.798	0.759	8.40%	21.60%
ANN06	DBH_{eq}, Db, H_T	0.409	0.782	10.20%	14.20%
ANN07	DBH_{eq}, H_T, H_S	0.729	0.847	7.80%	14.70%
ANN08	Db, H_T, H_S	0.683	0.374	9.00%	25.50%
ANN09	DBH_{eq}, Db, H_T, H_S	0.769	0.721	7.80%	14.70%
$S^2_{y\hat{y}}$		0.015	0.032	-	-
$S_{\bar{y}}$		0.123	0.18	-	-

DBHeq = equivalent diameter; Db = base diameter; H_T = total height; H_S = height of the shaft. G1 and G2 = data groups; ANN01 to ANN09 = architecture of the neural networks used; RMSE = Root Mean Squared Error (%); $S^2_{y\hat{y}}$ = variance; $S_{\bar{y}}$ = standard deviation.

To avoid overfitting, it is advisable to use simpler settings when using ANN, with the fewest possible neurons in the hidden layer. Overfitting can be defined as excessive learning of the information contained in the data offered to the network, making them so well trained on the data set that, in addition to the structural similarity between the variables, they also end up assimilating the error in the relationship. Thus, the use of the network is no longer viable in the entire data sample, as its generalization capacity is compromised (Miranda 2016). In this study, all ANN created are MLP (multilayer perceptron), with 2 neurons in the hidden layer and the algorithm used was backpropagation.

To verify the efficiency of ANNs, the results obtained were compared with results provided by classical models adjusted for forest measurement at R free software (R Core Team, 2015), which also presented the predictive ability between actual and estimated data and the RMSE%. The Multiple Linear Regression (MLR) models here used were chosen according to the adjustments adopted by Ramos et al. (2021), and are presented in Table 3 (models MLR01 – MLR06). Two personalized models predicting tree volumes were also evaluated here (MLR07 and MLR08), after using all predictive tree variables (DBH_{eq}, Db, H_T and H_S) in a computational *stepwise* variable-selection process. The stepwise process was performed at the R pr using the default function of the base package. Note that, using bifurcation tree variables, only the G1 dataset was used (MLR08), due to having this kind of tree predicting variables.

3. Results

The results showed a strong relationship between the volume measurement performed in the field and using ImageJ, which was evidenced by simple linear regression (R^2 = 0.8; $P<0.05$) (Figure 2-B). The analysis of the correlation (Table 2) between groups G1 and G2 showed a great variation between the data of the two groups since the predictive ability had its maximum values in ANN03 and ANN09 (0.766 and 0.743 respectively).

The predictive abilities (i.e., correlation analysis between the actual values and the values generated by the ANNs) using the 10-fold validation can be seen in Figure 3A. ANN03, ANN05, and ANN09 networks showed the best results when compared to the others, reaching 0.851, 0.851, and 0.817, respectively. Through the calculation of the RMSE, it was observed that the errors generated by ANNs are quite small, with their maximum values in models ANN10 and ANN03, as shown in Figure 3-B.

Another type of validation was performed by comparing the performance of ANN models trained at dataset G1 and validated at dataset G2, and also the vice-versa scheme (results opportunely embedded in Table 2). Indeed, although expected worse results than the 10-fold validation, good results were also shown. In that context, the model with the greater predictive abilities was ANN04. Note that, the predictive abilities between these two reciprocal validation schemes have not shown great differences (e.g., on average of all models, 0.714 to G1 $\rightarrow$ G2; and 0.704 to G2 $\rightarrow$ G1), however, the RMSE predicting the G1 subset using of the G2 shown in average non-desirables high values (18.90% to G2 $\rightarrow$ G1 versus 9.50% G1 $\rightarrow$ G2).

Table 4 presents the data referring to the classic models used for comparison with the ANNs. Through the calculation of RMSE, it was observed that the errors generated by ANNs are small, with their maximum values in models ANN04 and ANN07, as shown in Figure 3-B. The maximum predictive ability achieved by the models is 0.952, in MLR08, and 0.951, in MLR07 (Table 4). Despite the high correlation, the calculation of the RMSE demonstrates that the errors generated are still high, reaching a maximum of 54% in the MLR04 model.

Models MRL07 and MRL08 showed similar trends. We observed in MRL07 a slight tendency to overestimate and underestimate the volume at lower values (Fig. $4A_1$). On the other hand, MRL08 presented a slight tendency to underestimate the volume at higher values (Fig. $4A_2$). Despite this, the residuals were mainly scattered near zero. The histograms of the residuals showed a higher frequency in classes close to zero (Fig $4C_1$ and C_2).

4. Discussion

The predictability of any model for unobserved scenarios is commonly described by metrics that compare estimated versus true values (Yang and Huang 2014). As for the values of the correlation coefficient (i.e., predictive ability), we observed at our datasets and by multiple permutation evaluations that: r = 0.10 to 0.30 (assumed *weak*, i.e., always statistical no-significant); r = 0.40 to 0.60 (assumed as *moderate*, i.e., sometimes significant); r = 0.70 to 1.00 (assumed as *strong*, i.e., always significant). Based on this assumption, it can be presumed that the use of data obtained through digital photographs is valid, since the correlation between real (manual measurement) and "digital" volumes had a correlation coefficient of 0.89, and when adjusted in linear regression, presented an R^2 of 0.80, which shows the assertiveness of the regression. Simple estimation of stem diameter and height of isolated trees based on digital image processing has already been presented by other authors (Juujarvi et al. 1998; Zhang and Huang 2009), demonstrating the feasibility of the method, however, the data could be applied to the volume estimation.

Thus, the photographic method signals to secure dendrometry data collection for determining the volume of trees in the Brazilian Savanna, considering the inherent difficulties in the evaluation of multiple species large with variations in shape among individuals. Also, in comparison with other non-destructive methods such as laser scanning and photogrammetry, it proved to be a low-cost, quick-to-apply method that does not require advanced professional training.

Table 3. Multiple Linear Regression (MLR) models and their respective equations were used for fit comparisons.

Model	Model	Equation
MLR01	Meyer*	$v = \beta_0 + \beta_1\ DBH + \beta_2\ \text{DBH}^2 + \beta_3\ \text{DBH}\ H_T + \beta_4\ \text{DBH}^2\ H_T + \varepsilon$
MLR02	Naslund*	$v = \beta_0 + \beta_1\ \text{DBH}^2 + \beta_2\ \text{DBH}^2\ H_T + \beta_3\ \text{DBH Ht}^2 + \beta_4\ H_T^2 + \varepsilon$
MLR03	Schumacher & Hall*	$\ln(v) = \beta_0 + \beta_1\ \ln(\text{DBH}) + \beta_2\ \ln(H_T) + \varepsilon$
MLR04	Husch*	$\ln(v) = \beta_0 + \beta_1\ \ln(\text{DBH}) + \varepsilon$
MLR05	Spurr*	$v = \beta_0 + \beta_1\ \text{DBH}^2 H_T + \varepsilon$
MLR06	Stoate*	$v = \beta_0 + \beta_1\ \text{DBH}^2 + \beta_2\ \text{DBH}^2 H_T + \beta_3\ H_T + \varepsilon$
MLR07	Proposed model 1 (G1 and G2 tree groups)	$v = \beta_0 + \beta_1\ H_{\text{stem}}{}^2 + \beta_2\ \text{DBH}^2\ H_T + \varepsilon$
MLR08	Proposed model 2 (Only G1 tree group)	$v = \beta_0 + \beta_1\ n_{\text{stem}} + \beta_2\ \text{DBH}^2 + \beta_3\ \text{DB}^2 + \varepsilon$

* For further information about these models, please refer to Souza & Soares (2013); $ln(\ )$ = the natural logarithm; v = estimated volume (m³); DBH = diameter at breast height (or DBH_{eq} when there is any bifurcation - cm); H_T = total height of the individual (m); H_{stem} = stem height of the individual (m); n_{stem} = número de bifurcações; β_0 to β_4 = model coefficients.

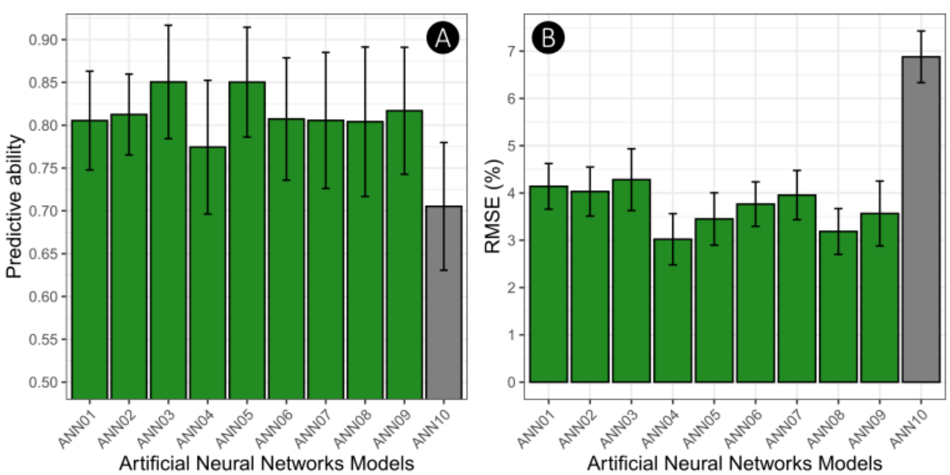

Figure 3. Validation using the 10-fold validation methodology. A) Predictive ability of ANNs by model. B) Bias (or RMSE, in percentage – Root Mean Squared Error) by model between observed and estimated volume. In both charts, the vertical bars of intervals are the standard errors of the means. Green bars are the results depicted using both G1 and G2 datasets (models ANN01 – ANN09); The gray bars are the results depicted using only the G1 fitted model (ANN10).

When the data groups (G1 and G2) were compared to each other, some ANN models had a high predictive ability, while other neural network architectures had worse results. This can be explained by the large variation between tree individuals, even within the same species, considering their physical characteristics such as trunk diameter, crown diameter or even total height (Rezende et al. 2006; Zimbres et al. 2020). Nevertheless, some ANNs had a predictive ability greater than 0.7, considered strong. Analyzing the RMSE (%), there was low bias in certain ANNs, with values of 7.8% (ANN07 and ANN09, G1 $\rightarrow$ G2) and 8.0% (ANN01, G1 G2) while other ANNs showed greater dispersion, reaching values of 27.6% and 25.5% (ANN04 and ANN08, both at G2 $\rightarrow$ G1 validation, respectively). The validation process between G2 predicting G1 provided higher RMSE values, certainly due to the differences in the dendrometry structure between these two sets of data, in addition, G2 is a data with more dispersed ages and higher coefficients of residual variation.

As for the predictive ability of the ANNs created for the complete group of data (using the 10-fold validation scheme), high values were expected and confirmed by our results, always reaching the range above 0.8, except for the ANN04 which reached predictive ability of 0.774. Lacerda et al. (2017) proposed the creation of ANNs for volumetry in the Cerrado separated by species, obtaining an even more accurate result regarding the volumes estimated, reaching predictive ability values of up to 0.85 and 0.93, but with large values of RMSE reaching 42.23% and 54.23% (ANNs created for the species *Terminalia fagifolia* (Combretaceae) and *Heteropterys byrsonimifo-*

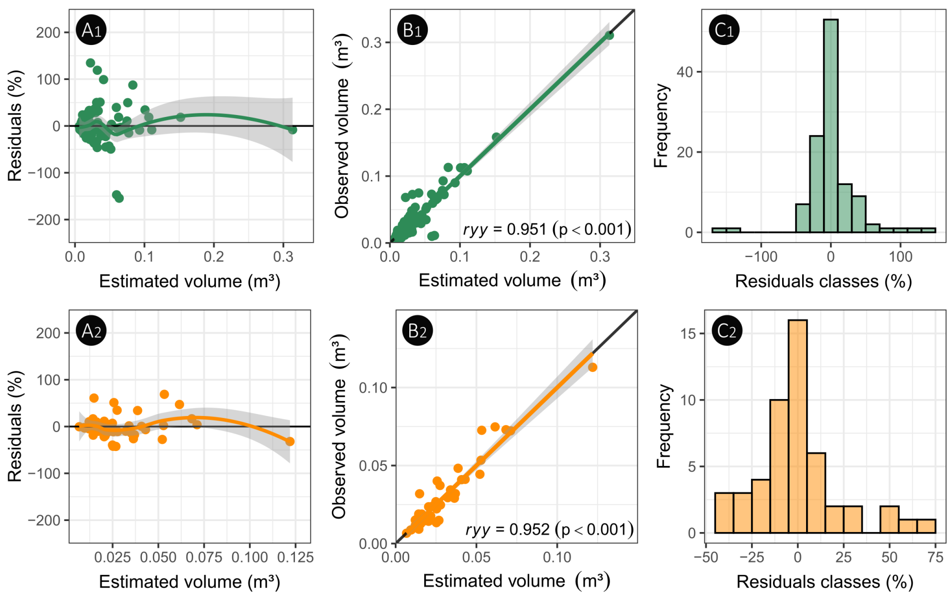

Figure 4. Dispersion of residuals (Ai), the relation between observed x estimated volume (Bi), and distribution of classes of residuals (Ci) for equations. In green (parts A1, B1, and C1) are depicted the proposed model; and in orange (parts A2, B2, and C2) the model with image data only.

lia (Malpighiaceae), respectively), demonstrating the low RMSE values achieved in this work. Ribeiro et al. (2011), adjusting ANNs to predict carbon stock in a clonal *Eucalyptus grandis* plantation, reached efficiency values of the models of up to 0.98 and RMSE of 2.94%, demonstrating the great accuracy achieved by ANNs for situations of planted forests, contrasting with the great challenge presented in the volumetric estimation of very-crooked and bifurcated trees. In this sense, the present study brings relevant contributions through the combination of digital imaging and ANNs for this purpose.

The results obtained by the classic volumetric models demonstrate the efficiency of ANNs for volumetric calculations, although MRL models presented always higher predictive ability, above 0.90, they showed reasonable RMSE values (ranging between 23.32% and 54.09%). Rezende et al. (2006), when comparing volumetric models for individuals from the Cerrado, found R^2 values between 0.94 and 0.98, demonstrating a great fit of these models to the data collected even considering the large variation in shape among individuals from this biome. Also, found errors of up to 30%, which can make the model impracticable for use in different data sources. In this study, the model generated also presented a predictive ability of only 0.386, which corresponds to a moderate strength correlation.

Using Schumacher & Hall and Spurr models, both in their logarithmized forms, for three different regions of Cerrado *sensu stricto*, Rufini et al. (2015) obtained satisfactory results for the volume with bark, reaching R^2 values between 93.21% and 98.74%, demonstrating that these models can be viable. The single-entry Hohenadl-Krenn volume model was tested and selected by Santos et al. (2021) for a dataset of nine representative species of the Cerrado *sensu stricto*, presenting a better distribution of residues and reaching a R^2 value of 99,97 %. While Silva et al. (2009), state that there are no major differences in terms of bias between the results obtained by the Schumacher & Hall model and the ANNs. Furthermore, it states that there is no risk in using the Schumacher & Hall model in its linearized form. This result is also confirmed in this study, since the Schumacher & Hall model (MLR03) obtained only 0.941 of predictive ability, but with RMSE% equal to 52.26%. The personalized models performing in this work (MLR07 and 08) also showed good results.

Going further, Bueno et al. (2020) estimated the parameters for 20 volumetric models for five species of fruit trees native to the Cerrado, including the Meyer, Spurr, and Stoate models, replicated in this study, obtaining R^2 values of 0.68, 0.95 and 0 .68, respectively, and standard error of 0.50, 0.19 and 0.50, respectively, demonstrating acceptable adjustments, with emphasis on the Spurr model, which obtained the best results. Indeed, the floristic variability of the Cerrado is enormous (Ferreira et al. 2021). In the Bueno et al. (2020)' work, the same models achieved results of predictive ability of 0.933, 0.929, and 0.911, and RMSE varying between 40.5%, 41.7%, and 47.2% respectively for the same models (MLR01, MLR05, and MLR06). Such difference in results can be explained by the greater variety of species used in this work, totaling 25 native Cerrado species.

When estimating the total wood volume for an area of Cerrado through linear regression involving the basal

Table 4. Statistics of fit and precision for modeling individual volume (m³). Are shown the Betas, predictive ability and RMSE of the classic regression-based models.

Model	β_i	*Estimate*	*S.E.*	*p-value*	*RMSE*	*RMSE%*	$r_{y\hat{y}}$
MLR01	β_0	-7.49E-03	1.18E-02	0.530	0.0137	40.465	0.933
	β_1	-2.41E-04	3.06E-03	0.930			
	β_2	2.45E-04	1.78E-04	0.170			
	β_3	5.46E-04	1.41E-04	<0.001			
	β_4	-9.25E-06	9.50E-06	0.330			
MLR02	β_0	-5.33E-03	3.26E-03	0.100	0.0129	38.081	0.941
	β_1	5.46E-04	6.28E-05	<0.001			
	β_2	-7.49E-05	1.93E-05	<0.001			
	β_3	1.28E-04	3.56E-05	<0.001			
	β_4	-3.06E-04	2.07E-04	0.142			
MLR03	β_0	-8.2974	0.306	<0.001	0.0177	52.257	0.919
	β_1	1.8194	0.147	<0.001			
	β_2	0.4497	0.133	<0.001			
MLR04	β_0	1.85E-02	1.92E-03	<0.001	0.0183	54.090	0.878
	β_1	2.65E-05	1.37E-06	<0.001			
MLR05	β_0	-1.63E-02	5.02E-03	<0.05	0.0141	41.671	0.929
	β_1	3.43E-04	4.38E-05	<0.001			
	β_2	-5.90E-07	3.35E-06	0.86			
	β_3	3.86E-03	8.07E-04	<0.001			
MLR06	β_0	-7.9332	0.300	<0.001	0.01603	47.189	0.911
	β_1	2.0058	0.142	<0.001			
MLR07	β_0	1.79E-03	1.62E-03	0.273	0.0118	34.940	0.951
	β_1	7.29E-04	9.73E-05	<0.001			
	β_2	2.24E-05	8.10E-07	<0.001			
MLR08	β_0	-1.22E-02	2.42E-03	<0.001	0.0065	23.322	0.952
	β_1	5.44E-03	1.19E-03	<0.001			
	β_2	2.34E-04	3.23E-05	<0.001			
	β_3	6.38E-05	1.54E-05	<0.001			

β_i : β_0 to β_4 = model coefficients.; $r_{y\hat{y}}$ = Predictive Ability; S.E. = Standard Error; RMSE = square root mean error; RMSE% = square root of the mean percent error; MLR01 to MLR06: classic models used; MLR07 and MLR08: proposed models.

area of individuals and vegetation indices Miguel et al. (2015) calculated through satellite images, and standard error of 12.34%, while the ANNs obtained a standard error of 6.01%, suggesting the greater fit of ANNs when compared to models based on linear regression.

From the results obtained, it can be observed that the use of photographs to obtain dendrometric variables is viable, contributing to the establishment of non-destructive methodologies in forest measurement. This one stands out for being a low-cost methodology, since it doesn´t need a lot of manpower and material resources, and with good accuracy compared to others such as photogrammetry and lidar. It is noteworthy that the use of calibration rods is essential for the accuracy of the method, since without them there are no parameters for measurement in the software. It is also highlighted that both ANNs and MLR procedures generate accurate models for volumetric estimation for trees of different species from the Brazilian Savanna, although ANN showed greater potential against MLR in terms of lower bias estimation.

5. Final Considerations

The application of Artificial Neural Networks (ANN) underscores their efficacy in modeling individual tree characteristics within the unique context of the Brazilian Savanna. Particularly suited for the estimation of tree volume, ANN proves especially advantageous in capturing the nuances of non-straight and bifurcated tree stems characteristic of this biome. The emphasis on non-destructive measurements highlights the practical benefits of this approach, showcasing its efficiency in terms of time and labor when compared to traditional destructive methods. Utilizing techniques such as digital photography enables accurate tree volume estimates without causing harm to the trees, exemplifying the utility of modern technologies in environmentally conscious forestry practices.

The dependable estimation derived from digital photographs, captured through widely accessible mobile devices, affirms the feasibility of using this method for estimating volumes in the intricate landscape of very-irregular trees in the Brazilian Savanna. This approach not only proves cost-effective but also democratizes the process, making tree volume estimation accessible to a broader audience. Through rigorous comparison with conventional volumetric models, the study substantiates the superior predictive capabilities of ANN over regression-based volumetric models, underscored by its ability to minimize modeling bias.

Acknowledgments

Authors would thank to Forest Inventory Lab. of EA/UFG (Profloresta); the lab technician Guilherme Murilo for technical and infrastructure support; to agronomist Dr. Francisco José B. Baccarin for data support; and the financial support of *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), and *Fundação de Amparo à Pesquisa do Estado de Goiás* (FAPEG).

Contributors' statement

PHCS: Formal Analysis; Investigation; Methodology; Writing – original draft. **HJS**: Formal Analysis; Investigation; Methodology; Writing – review & editing. **MPC**:

Methodology; Validation; Writing – review & editing. **DHBB**: Software; Formal Analysis. **FV**: Data curation; Project administration; Resources. **RTR**: Supervision; Validation; Visualization; Writing – review & editing.

References

Almeida AQ de, Mello AA de, Neto ALD, Ferraz RC (2014) Empiric relations between dendrometric characteristics of the Brazilian dry forest and Landsat 5 TM data. *Pesqui Agropecuária Bras* 49:306–315

Berveglieri A, Tommaselli A, Liang X, Honkavaara E (2017) Photogrammetric measurement of tree stems from vertical fisheye images. *Scand J For Res* 32:737–747

Binoti DHB., Binoti MLMS., Leite HG (2013) *NeuroForest Star*

Binoti DHB, Binoti MLM da S, Leite HG, et al (2012) Modelagem da distribuição diamétrica em povoamentos de eucalipto submetidos a desbaste utilizando autômatos celulares. *Rev Árvore* 36:931–940

Brady NC (2020) Making agriculture a sustainable industry. In: *Sustainable agricultural systems*. CRC Press, pp 20–32

Bueno LP, Medanha JTR, Rodovalho RS, de Castro VGS (2020) Modelagem volumétrica para frutíferas do cerrado / Volumetric modeling for cerrado fruit. *Mirante* 13:39–54

Castellanos A, Martinez Blanco A, Palencia V (2007) *Applications of radial basis neural networks for area forest.*

Cerqueira CL, Môra R, Tonini H, et al (2021) *Modeling of eucalyptus tree stem taper in mixed production systems*. Embrapa Pecuária Sul-Artigo em periódico indexado

Colli GR, Vieira CR, Dianese JC (2020) Biodiversity and conservation of the Cerrado: recent advances and old challenges. *Biodivers. Conserv.* 29:1465–1475

de Azevedo GB, Tomiazzi HV, Azevedo GT de OS, et al (2020) Multi-volume modeling of Eucalyptus trees using regression and artificial neural networks. *PLoS One* 15:e0238703

de León GC, Uranga-Valencia LP (2013) Theoretical evaluation of Huber and Smalian methods applied to tree stem classical geometries. *Bosque* 34:311–317

Dean C (2003) Calculation of wood volume and stem taper using terrestrial single-image close-range photogrammetry and contemporary software tools. *Silva Fenn* 37:359–380

De Souza HJ, Miguel EP, Resende RT, et al (2022) Effects of functional traits on the spatial distribution and hyperdominance of tree species in the Cerrado biome. *iForest-Biogeosciences* For 15:339

Ferreira INM, Ferreira FG, Miranda S do C de, et al (2021) Floristic and structural aspects of Brazilian Savanna phytophysiognomies in the northern Goiás state, Brazil. *Pesqui Agropecuária Trop* 51:

Han D, Dong H (2011) The study on standing tree measurement based on image processing embedded in a smartphone. In: 2011 International Conference on Image Analysis and Signal Processing. IEEE, pp 252–256

Husch B., Beers TW., Kershaw-Jr JA (2002) Forest mensuration. John Wiley & Sons

Juujarvi J, Heikkonen J, Brandt SS, Lampinen J (1998) Digital-image-based tree measurement for forest inventory. In: Intelligent Robots and Computer Vision XVII: Algorithms, Techniques, and Active Vision. International Society for Optics and Photonics, pp 114–123

Kershaw Jr JA, Ducey MJ, Beers TW, Husch B (2017) Integrating remote sensing in forest inventory. *For Mensurat* 429–454

Lacerda THS, Cabacinha CD, Araújo CA, et al (2017) Artificial neural networks for estimating tree volume in the Brazilian savanna. *Cerne* 23:483–491

Liu J, Feng Z, Mannan A, et al (2019) Comparing Non-Destructive Methods to Estimate Volume of Three Tree Taxa in Beijing, China. *Forests* 10:92

McCann MT, Jin KH, Unser M (2017) Convolutional neural networks for inverse problems in imaging: A review. IEEE *Signal Process Mag* 34:85–95

Miguel EP, Rezende AV, Leal FA, et al (2015) Artificial neural networks for modeling wood volume and aboveground biomass of tall Cerrado using satellite data. *Pesqui Agropecuária Bras* 50:829–839

Miranda JFN de (2016) Regression models and artificial neural networks in the measurement of carbon and woody biomass for a deciduous forest in Central Brazil. Universidade de Brasília, Faculdade de Tecnologia

Nicoletti MF, Batista JLF, Carvalho S de PC, et al (2015) Accuracy of optical dendrometers for determining the volume of standing trees. *Ciência Florest* 25:395–404

Puliti S, Breidenbach J, Astrup R (2020) Estimation of forest growing stock volume with UAV laser scanning data: can it be done without field data? *Remote Sens* 12:1245

Ramos L de O, de Miranda ROV, Soares AAV, et al (2021) WOOD VOLUMETRY OF Tachigali vulgaris PURE PLANTATIONS IN DIFFERENT PLANTING SPACINGS. FLORESTA 51:962–970

Rezende A V, Vale AT do, Sanquetta CR, et al (2006) Comparison of mathematical models to volume, biomass and carbon stock estimation of the woody vegetation of a Cerrado Sensu Stricto in Brasília, DF. Sci For 71:65–76

Ribeiro SC, Fehrmann L, Soares CPB, et al (2011) Above- and belowground biomass in a Brazilian Cerrado. *For Ecol Manage* 262:491–499

Roitman I, Bustamante MMC, Haidar RF, et al (2018) Optimizing biomass estimates of savanna woodland at different spatial scales in the Brazilian Cerrado: Re-evaluating allometric equations and environmental influences. *PLoS One* 13:e0196742

Rufini AL, Scolforo JRS, de Oliveira AD, de Mello JM (2015) Volume equations for the Savannah (Cerrado), in Minas Gerais state. *Cerne* 16:1–11

Schindelin J, Arganda-Carreras I, Frise E, et al (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9:676–682

Scolforo HF, Scolforo JRS, Mello CR, et al (2015) Spatial distribution of aboveground carbon stock of the arboreal vegetation in Brazilian biomes of Savanna, Atlantic Forest and Semi-Arid Woodland. *PLoS One* 10:e0128781

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

Skovsgaard JP and, Vanclay JK (2008) Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. For An *Int J For Res* 81:13–31

Souza AL de, Soares CPB (2013) *Florestas Nativas*, 1st edn. Editora UFV, Viçosa (MG)

Taborda dos Santos A, Dressano Domene V, Póvoa de Mattos P, et al (2021) Equação de volume para espécies de Cerrado em Formosa, GO. Brazilian J *For Res Florest Bras* 41:

Team R (2015) R Development Core Team. R A Lang Environ Stat Comput 55:275–286

Wang F, Sun Y, Jia W, et al (2023) Development of Estimation Models for Individual Tree Aboveground Biomass Based on TLS-Derived Parameters. Forests 14:351

Yang Y, Huang S (2014) Suitability of five cross validation methods for performance evaluation of nonlinear mixed-effects forest models–a case study. *For An Int J For Res* 87:654–662

Zhang J, Huang X (2009) Measuring method of tree height based on digital image processing technology. In: 2009 First International Conference on Information Science and Engineering. IEEE, pp 1327–1331

Zimbres B, Shimbo J, Bustamante M, et al (2020) Savanna vegetation structure in the Brazilian Cerrado allows for the accurate estimation of aboveground biomass using terrestrial laser scanning. *For Ecol Manage* 458:117798

Author Statements

- ✓ No conflicts of interest were declared.
- ✓ All existing funding sources were acknowledged.
- ✓
- ✓ There is no evidence of plagiarism in this article.