

Article

Machine Learning: Volume and Biomass Estimates of Commercial Trees in the Amazon Forest

Samuel José Silva Soares da Rocha ^{1,*}, Flora Magdaline Benitez Romero ², Carlos Moreira Miquelino Eleto Torres ³, Laércio Antônio Gonçalves Jacovine ³, Sabina Cerruto Ribeiro ⁴, Paulo Henrique Villanova ³, Bruno Leão Said Schettini ³, Vicente Toledo Machado de Moraes Junior ⁵, Leonardo Pequeno Reis ⁶, Maria Paula Miranda Xavier Rufino ³, Indira Bifano Comini ³, Ivaldo da Silva Tavares Júnior ³ and Águida Beatriz Traváglia Viana ³

- ¹ Departamento de Ciências Florestais, Universidade Federal de Lavras, Lavras 37200-900, MG, Brazil
² Instituto Nacional de Pesquisas da Amazônia—INPA, Manaus 69067-375, AM, Brazil
³ Departamento de Engenharia Florestal, Universidade Federal de Viçosa, Viçosa 36570-900, MG, Brazil
⁴ Centro de Ciências Biológicas e da Natureza, Universidade Federal do Acre (UFAC), Campus Universitário BR 364, Km 04, Distrito Industrial, Rio Branco 69920-900, AC, Brazil
⁵ Brandt Meio Ambiente LTDA, Alameda do Ingá, 89, Vale do Sereno, Nova Lima 34006-042, MG, Brazil
⁶ Instituto de Desenvolvimento Sustentável Mamirauá, Tefé 69553-225, AM, Brazil
* Correspondence: samuel.rocha@ufla.br

Abstract: Accurate estimation of the volume and above-ground biomass of exploitable trees by the practice of selective logging is essential for the elaboration of a sustainable management plan. The objective of this study is to develop machine learning models capable of estimating the volume and biomass of commercial trees in the Southwestern Amazon, based on dendrometric, climatic and topographic characteristics. The study was carried out in the municipality of Porto Acre, Acre state, Brazil. The volume and biomass of sample trees were determined using dendrometric, climatic and topographic variables. The Boruta algorithm was applied to select the best set of variables. Support Vector Machines (SVM), Artificial Neural Networks (ANN), Random Forests (RF) and the Generalized Linear Model (GLM) were the machine learning methods evaluated. In general, the evaluated methods showed a satisfactory generalization power. The results showed that the volume and biomass predictions of commercial trees in the Amazon rainforest differed between the techniques ($p < 0.05$). ANNs showed the best performance in predicting the volume and biomass of commercial trees, with the highest $r_{y\hat{y}}$ and the lowest RSME and MAE. Thus, machine learning methods such as SVM, ANN, RF and GLM are shown to be useful and efficient tools for estimating the volume and biomass of commercial trees in the Amazon rainforest. These methods can be useful tools to improve the accuracy of estimates in forest management plans.

Keywords: biome allometry; dense rainforest; models; artificial intelligence; biometry



Citation: Rocha, S.J.S.S.d.; Romero, F.M.B.; Torres, C.M.M.E.; Jacovine, L.A.G.; Ribeiro, S.C.; Villanova, P.H.; Schettini, B.L.S.; Junior, V.T.M.d.M.; Reis, L.P.; Rufino, M.P.M.X.; et al. Machine Learning: Volume and Biomass Estimates of Commercial Trees in the Amazon Forest. *Sustainability* **2023**, *15*, 9452. <https://doi.org/10.3390/su15129452>

Academic Editors: Nancai Pei, Sven Günter and Steve W. Lyon

Received: 28 February 2023

Revised: 17 May 2023

Accepted: 9 June 2023

Published: 12 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Machine learning is a rapidly growing area of study in artificial intelligence and is useful in forest modeling because of its potential to produce better models than traditional data modeling approaches [1–3], particularly allometric equations [4–15]. The applications of this computational intelligence technique in the forestry sector have gained great relevance [16–18]. Models have already been efficiently tested to estimate tree growth [19], biomass and carbon [20–22], map species richness and composition [23], predict tree diameter and height [24–26], map tropical forest structure [27] and assess forest quality parameters [28].

The application of these models to estimate tree volume and biomass can be promising for obtaining accurate estimates of these tree parameters. Machine learning algorithms are capable of processing vast amounts of data from various sources, such as remote

sensing and ground-based measurements, to generate accurate estimates of tree volume and biomass. This information is critical for forest management, conservation and carbon accounting efforts in the Amazon.

The Brazilian Amazon rainforest is the largest remnant of the forest [29] and has an important role in the carbon cycle, storing around 150–200 Pg of carbon in biomass and living things [30]. The biome is home to more than a third of all the diversity of neotropical plants [31] and 6700–16,000 tree species [32,33].

Sustainable forest management is critical for ensuring the long-term health and productivity of the forest while meeting the needs of local communities and supporting economic development [34–36]. Accurate estimates of tree volume and biomass are essential for planning and implementing sustainable forest management practices [37,38] such as selective logging and forest restoration [5].

Machine learning (ML) methods have been increasingly used for biomass estimation, as they can handle large datasets, complex relationships and nonlinear patterns in ecological systems. Algorithms such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), Random Forests (RF) and the Generalized Linear Model (GLM) are often used in these studies, seeking more accurate volume and biomass estimates [21,39–42]. For example, researchers have used Random Forests to estimate aboveground biomass using airborne spectral indices [43], and also used Neural Networks and support-vector regression to estimate forest biomass using climate data [44].

This superiority in the generation of estimates is associated with a lower number of assumptions about data and processes [45], which allows the generation of better prediction results, in view of the complex relationships of forest dynamics. In the above context, this study aims to develop machine learning models that are capable of estimating the volume and biomass of commercial trees in the Southwestern Amazon, based on dendrometric, climatic and topographic characteristics. The research questions of this study are: (i) Are the machine learning methods evaluated efficient for estimating the volume and biomass of commercial trees? (ii) What is the best method to estimate the volume of commercial trees? and (iii) What is the best method to estimate the biomass of commercial trees?

To achieve the objectives of this study, a systematic approach was followed in this study. Firstly, relevant literature was reviewed to identify the key variables affecting the volume and biomass of commercial trees in the Southwestern Amazon. Secondly, an extensive dataset was collected, which included dendrometric, climatic and topographic variables, as well as volume and biomass measurements of commercial trees. Thirdly, machine learning models were developed and trained using the collected dataset. Fourthly, the performance of different machine learning methods was evaluated and compared in terms of accuracy and precision. Finally, the best-performing machine learning models were selected, and the results were discussed in light of the research questions.

2. Materials and Methods

2.1. Characterization of the Study Area

The study was carried out at Antimary Farm I and II, located in the Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil (Figure 1). The area under sustainable management comprises 1253.02 ha. The region's vegetation is classified as "terra firme"—a forest with solid ground—and wetland rainforest [46]. The climate of the region is of the Am type, according to the Köppen classification [47]. The study area presents two types of soil, Red Argisol and Dystrophic Red Yellow Latosol [48]. The topography is predominantly flat, with a slope of around 5%. The altimetry varies between 220 and 300 m above sea level.

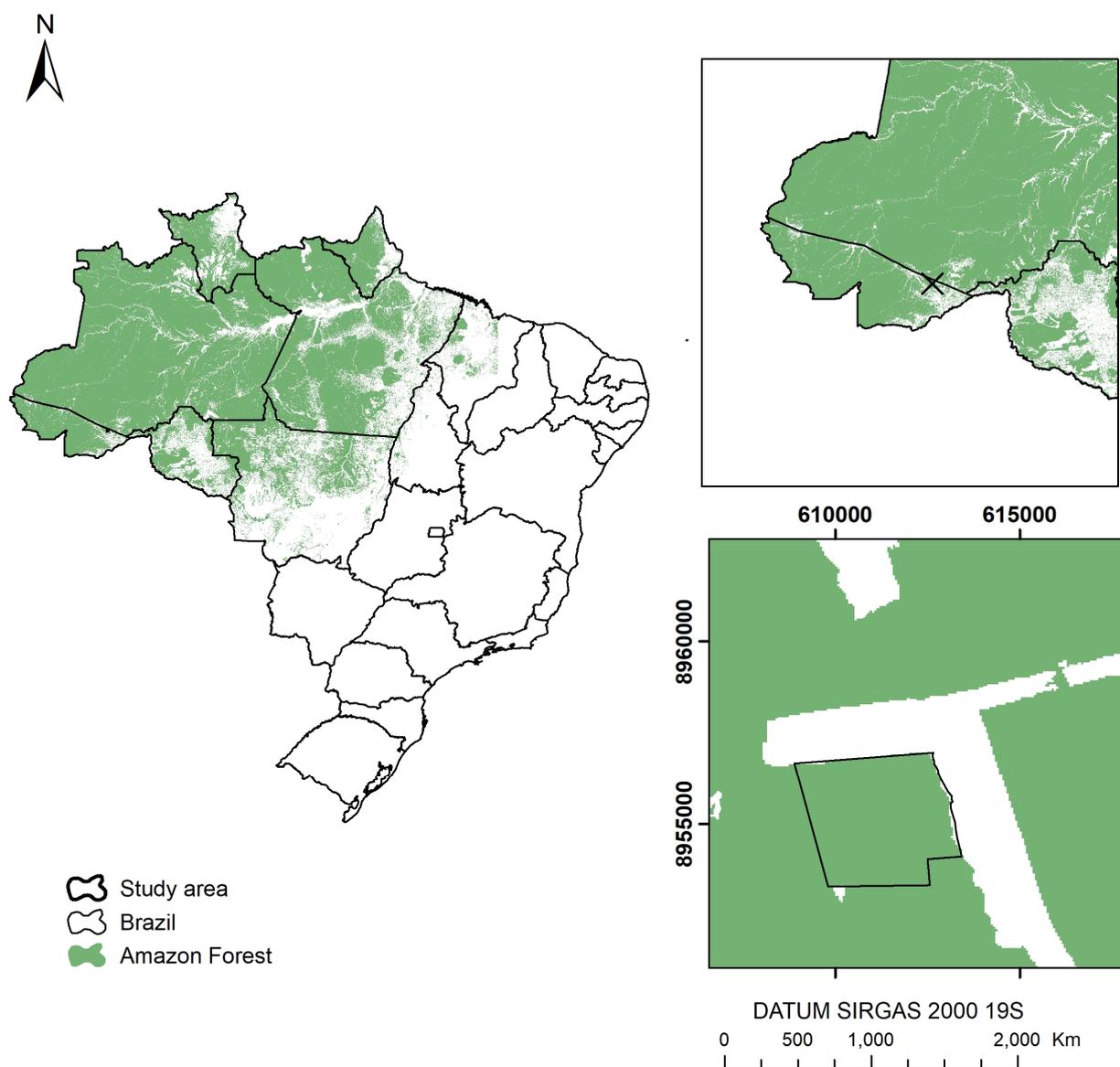


Figure 1. Location of the study area in the Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil.

A census was conducted in the exploitable area in May 2015 and the Sustainable Forest Management Plan (SFMP) was approved in 2016 by the Acre Environment Institute (Instituto de Meio Ambiente do Acre, IMAC).

2.2. Determination of Volume and Biomass Stocks

Sample trees were selected based on their density and basal area, obtained from information from the census provided by the company responsible for the management, in which all trees of a commercial interest with a diameter at breast height (DBH; 1.30 m) \geq 50 cm were measured. Eighteen species of the highest importance and value were selected and distributed in 214 individuals [49] (Table 1).

Table 1. Number of trees and basic wood density of commercial tree species present in Southwestern Amazon in the municipality of Porto Acre, Acre, Brazil.

SN	F	N	Bd
<i>Albizia niopoides</i> (Spruce ex Benth.) Burkart	Fabaceae Lindl.	7	0.64 ± 0.03
<i>Apuleia leiocarpa</i> (Vogel) J.F.Macbr.	Fabaceae Lindl.	13	0.77 ± 0.03
<i>Astronium lecointei</i> Ducke	Anacardiaceae R.Br.	6	0.82 ± 0.05
<i>Barnebydendron riedelii</i> (Tul.) J.H.Kirkbr.	Fabaceae Lindl.	5	0.57 ± 0.03
<i>Buchenavia tetraphylla</i> (Aubl.) R.A.Howard	Combretaceae R.Br.	9	0.69 ± 0.04
<i>Castilla ulei</i> Warb.	Moraceae Gaudich.	37	0.41 ± 0.04
<i>Cedrela odorata</i> L.	Meliaceae A.Juss.	8	0.43 ± 0.04
<i>Ceiba pentandra</i> (L.) Gaertn.	Malvaceae Juss.	4	0.29 ± 0.03
<i>Ceiba samauma</i> (Mart.) K.Schum.	Malvaceae Juss.	22	0.51 ± 0.05
<i>Copaifera multijuga</i> Hayne	Fabaceae Lindl.	6	0.52 ± 0.05
<i>Dipteryx odorata</i> (Aubl.) Willd.	Fabaceae Lindl.	11	0.80 ± 0.04
<i>Eschweilera bracteosa</i> (Poepp. ex O.Berg) Miers	Lecythidaceae A.Rich.	15	0.65 ± 0.05
<i>Eschweilera grandiflora</i> (Aubl.) Sandwith	Lecythidaceae A.Rich.	13	0.73 ± 0.03
<i>Handroanthus serratifolius</i> (Vahl) S.Grose	Bignoniaceae Juss.	8	0.82 ± 0.04
<i>Hura crepitans</i> L.	Euphorbiaceae Juss.	6	0.36 ± 0.06
<i>Hymenaea courbaril</i> L.	Fabaceae Lindl.	8	0.76 ± 0.04
<i>Parkia paraensis</i> Ducke	Fabaceae Lindl.	20	0.46 ± 0.06
<i>Schizolobium parahyba</i> var. <i>amazonicum</i> (Huber ex Ducke) Barneby	Fabaceae Lindl.	16	0.48 ± 0.08
$\bar{X} \pm CI$		11.89 ± 8.12	0.59 ± 0.17

Where: SN = Scientific name; F = family; N = number of individuals; Bd = Basic wood density, in $g\ cm^{-3}$; \bar{X} = Mean; CI = confidence interval.

The volume of the selected individuals was determined by strict cubing using the method of Smalian [50]. Wood disks from the base of the logs were collected to determine the basic wood density according to the ABNT standard (2003) [51]. Biomass was calculated by multiplying the volume and the basic density of the wood [8]. The average basic wood density of the trees analyzed was $0.59\ g\ cm^{-3}$ (Table 1).

2.3. Predictor Variables for Modeling

For the modeling, dendrometric, qualitative, climatic and topographic variables (predictor variables) were used to estimate the volume and biomass stocks (response variables). These variables have been shown to be important drivers of vegetation patterns in many ecosystems including the Amazon. Climatic variables such as temperature and precipitation can affect plant growth and distribution, while topographic variables such as elevation and slope can influence factors such as soil moisture and nutrient availability. The dendrometric variables used were DBH, commercial height (Ch) and basic wood density (Bd) [49]. The qualitative variables were the species and the family of the individuals [49]. The bioclimatic variables used (Bio 1-19) are derived from the monthly values of temperature and precipitation and were obtained from the WorldClim—Global Climate Data database [52], with a spatial resolution of approximately $1\ km^2$. The Bio 5 climatic variable was not used because it did not show variability in the study area. The topographic variable used was altitude (Table 2).

Table 2. Predictor variables used in the modeling of volume and biomass in the Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil.

Variable	Min	1° Quartil	Median	Mean	3° Quartil	Max	SD
DBH	50.38	64.78	75.44	79.60	89.52	149.92	20.14
Ch	7.30	11.71	14.20	14.82	17.87	25.40	4.03
Bd	0.29	0.43	14.20	0.57	0.73	0.82	0.15
Bio1	24.86	24.88	24.88	24.88	24.89	24.92	0.01
Bio2	11.45	11.51	11.52	11.51	11.52	11.53	0.01

Table 2. Cont.

Variable	Min	1° Quartil	Median	Mean	3° Quartil	Max	SD
Bio3	81.21	81.62	81.68	81.65	81.68	82.02	0.09
Bio4	82.28	83.38	83.42	83.49	83.63	84.39	0.25
Bio6	17.30	17.30	17.30	17.30	17.30	17.40	0.01
Bio7	14.00	14.10	14.10	14.10	14.10	14.10	0.01
Bio8	25.32	25.32	25.33	25.33	25.35	25.38	0.02
Bio9	23.63	23.67	23.67	23.67	23.68	23.73	0.02
Bio10	25.55	25.57	25.57	25.57	25.58	25.60	0.01
Bio11	23.63	23.67	23.67	23.67	23.68	23.73	0.02
Bio12	1830.00	1834.00	1836.00	1836.31	1839.00	1853.00	3.98
Bio13	250.00	251.00	252.00	251.61	252.00	254.00	1.03
Bio14	40.00	40.00	41.00	40.70	41.00	41.00	0.46
Bio15	51.47	51.58	51.77	51.74	51.77	52.05	0.17
Bio16	735.00	736.00	738.00	737.61	738.00	744.00	2.09
Bio17	154.00	155.00	155.00	155.29	156.00	158.00	0.91
Bio18	568.00	570.00	570.00	570.27	571.00	576.00	1.69
Bio19	154.00	155.00	155.00	156.82	156.00	198.00	8.01
Alt	151.05	164.99	173.16	175.08	183.08	248.87	13.75

Where: Min = minimum value; Max = maximum value; SD = standard deviation; DBH = diameter at breast height, in cm; Ch = commercial height, in m; Db = basic wood density; in g cm^{-3} ; Bio1 = annual mean temperature ($^{\circ}\text{C}$); Bio2 = mean diurnal range ($^{\circ}\text{C}$); Bio3 = Isothermality (%); Bio4 = temperature seasonality; Bio5 = max temperature of warmest month ($^{\circ}\text{C}$); Bio6 = min temperature of coldest month ($^{\circ}\text{C}$); Bio7 = temperature annual range ($^{\circ}\text{C}$); Bio8 = mean temperature of wettest quarter ($^{\circ}\text{C}$); Bio9 = mean temperature of driest quarter ($^{\circ}\text{C}$); Bio10 = mean temperature of warmest quarter ($^{\circ}\text{C}$); Bio11 = mean temperature of coldest quarter ($^{\circ}\text{C}$); Bio12 = annual precipitation (mm); Bio13 = precipitation of wettest month (mm); Bio14 = precipitation of driest month (mm); Bio15 = precipitation seasonality (Coefficient of Variation) (mm); Bio16 = precipitation of wettest quarter (mm); Bio17 = precipitation of driest quarter (mm); Bio18 = precipitation of warmest quarter (mm); Bio19 = precipitation of coldest quarter (mm); Alt = altitude.

The Boruta algorithm [53] was applied to select the best set of predictor variables to estimate volume and biomass. Boruta is a feature selection algorithm that was specifically designed to work with random forest models. The algorithm identifies the most relevant variables in a dataset by comparing the importance of each variable to that of randomly generated shadow variables. The shadow variables are created by permuting the values of each variable so that they are no longer associated with the response variable. The importance of each variable is then assessed by comparing the accuracy of a random forest model trained with the original variables to that of a model trained with the shadow variables. This algorithm iteratively removes the resources that are proven by a statistical test to be less relevant than random probes [53]. The Boruta R Software Package was used.

The quantitative variables were standardized to accelerate the convergence rate and reduce the iteration process in training:

$$Z_i = (x_i - \bar{x}) / \sigma$$

where:

Z_i —standardized value of the i -th observation;

x_i —value of the i -th observation;

\bar{x} —average of the observed values;

σ —standard deviation;

The scale function of R Software was used in this step.

2.4. Model Evaluation

The models tested to estimate the volume and biomass were: Support Vector Machines (SVM), Artificial Neural Networks (ANN), Random Forests (RF) and the Generalized Linear Model (GLM).

The trained ANNs were of the multilayer perceptron (MLP) type. The typical MLP architecture consists of an input layer containing the predictor variables, one or more hidden layers and an output layer containing the predicted variable. The activation function used was logistics. The training algorithms used were resilient propagation. The ANNs were implemented with the MLP function of the “RSNNS” Package in R.

The function svm of the “e1071” Package on R was used for training SVMs. The Kernel function was of the linear type. The random Forest function of the package of the same name in R was used for RF training. The glm function and link function of identity type and Gaussian family were used for GLM.

The performance of the models in the estimation of volume and biomass was assessed using the k-fold cross-validation method, with the data divided into 5 folds (4 for adjustments/training and 1 for validation). The k-fold cross-validation method is a technique used in machine learning to assess the performance of a model. In this method, the data is divided into k equally sized subsets or “folds”. The model is then trained on k-1 folds and validated on the remaining fold. This process is repeated k times, with each fold being used as the validation set once. The results are averaged over the k runs to obtain an estimate of the model’s performance. At each adjustment/training of the 5 folds, the metrics of Root Mean Square Error—RMSE (Equation (1)) and mean absolute error—MAE (Equation (2)) were calculated. This process was repeated 50 times, obtaining the average of the metrics for comparison of all models. The data were selected randomly in each of the 50 repetitions, resulting in different data sets, for greater robustness of the evaluation.

$$\text{RMSE} = \sqrt{\frac{1}{R} \sum_{r=1}^R \frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{n}} \quad (1)$$

$$\text{MAE} = \frac{1}{R} \sum_{r=1}^R \frac{|\sum_{i=1}^n (X_i - \hat{X}_i)|}{n} \quad (2)$$

where:

R —number of repetitions (50);

n —number of observations;

X_i —observed variable from the i -th tree, in m;

\hat{X}_i —estimated variable of the i -th tree, in m.

The averages of RMSE and MAE of each method in each repetition were ranked with weight assignments from 1 to 4, with 1 for the lowest value and 4 for the highest value. The weight p_{ij} assigned to the model m_j for the mean of RMSE was added to the weight p_{ij} assigned to the same model m_j for the mean of MAE, with $i = 1, 2, \dots, 50$. With the result of these sums, the values were submitted to the Friedman–Nemenyi test, at the 5% significance level (Equation (3)).

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \quad (3)$$

where:

CD —critical difference;

q_α —critical value calculated based on Studentized interval statistics (Harter, 1960) divided by $\sqrt{2}$;

k —number of algorithms being compared;

N —number of data sets.

3. Results

3.1. Selection of Variables

The applied variable selection procedure allowed the choice of the best model based on the ideal subset of variables. Figure 2a shows the variables that are most important for modeling tree volume, with the size of the bar representing the importance of the variable

and the green color of the barplots indicating which variables have been selected. The figure demonstrates that the diameter at breast height (DBH) and commercial height (Ch) are significant predictors of tree volume. This relationship is expected, as DBH and Ch are common metrics used to estimate tree volume (Figure 2). On the other hand, Figure 2b displays the addition of the wood basic density variable for biomass prediction. The authors note that this finding is consistent with previous research, which has shown that denser trees tend to have higher biomass. The bioclimatic variables and the topographic variable altitude were not considered significant for the modeling of volume and biomass based on the Boruta algorithm (Figure 2).

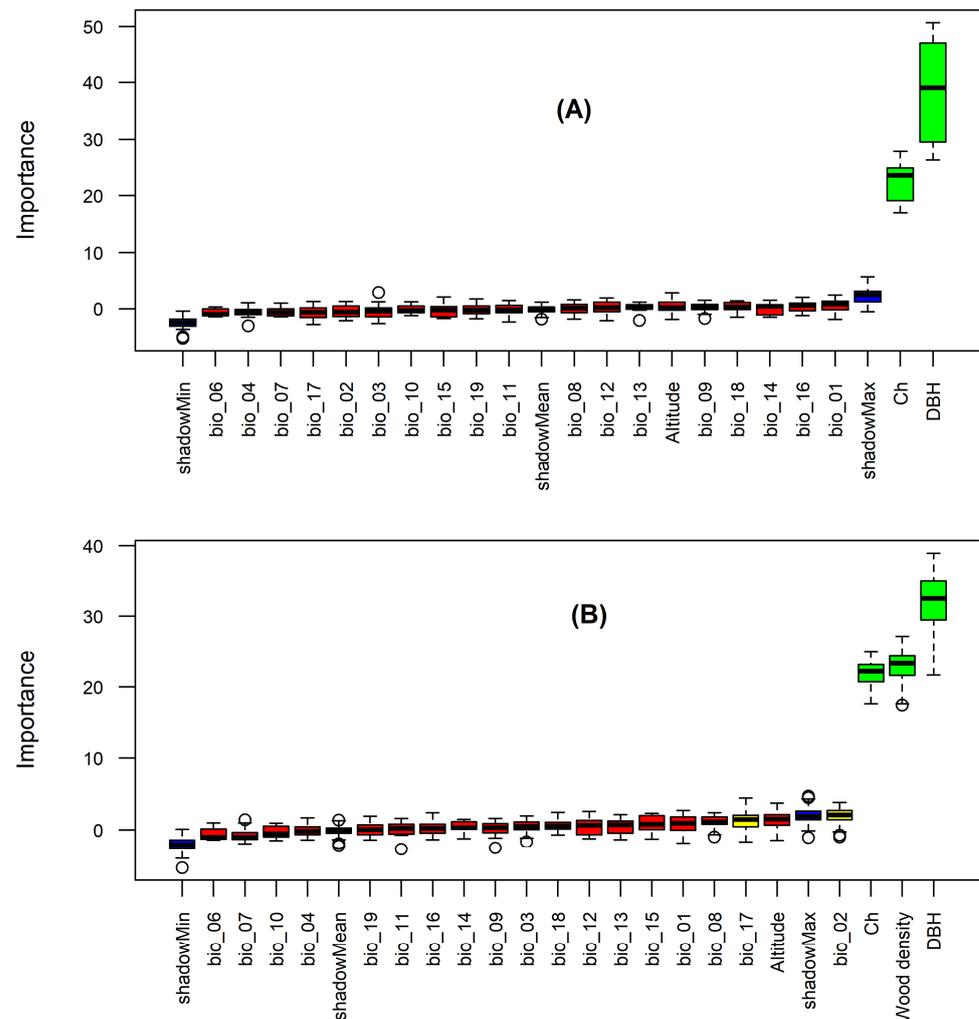


Figure 2. Most important variables for modeling the volume (A) and biomass (B) of commercial trees in Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil.

3.2. Selection of Models

Figure 3 shows the relationship between the observed and predicted values of volume and biomass for each evaluated technique. The blue line represents the tendency of the data. In general, the evaluated models showed a satisfactory generalization power, indicated by similar precision results between the observed and estimated data in the validation for all variables studied (Figure 3). Pearson's correlation coefficient ($r_{y\hat{y}}$) between the estimated and observed volume and biomass data was greater than 0.85 in all the machine learning models applied.

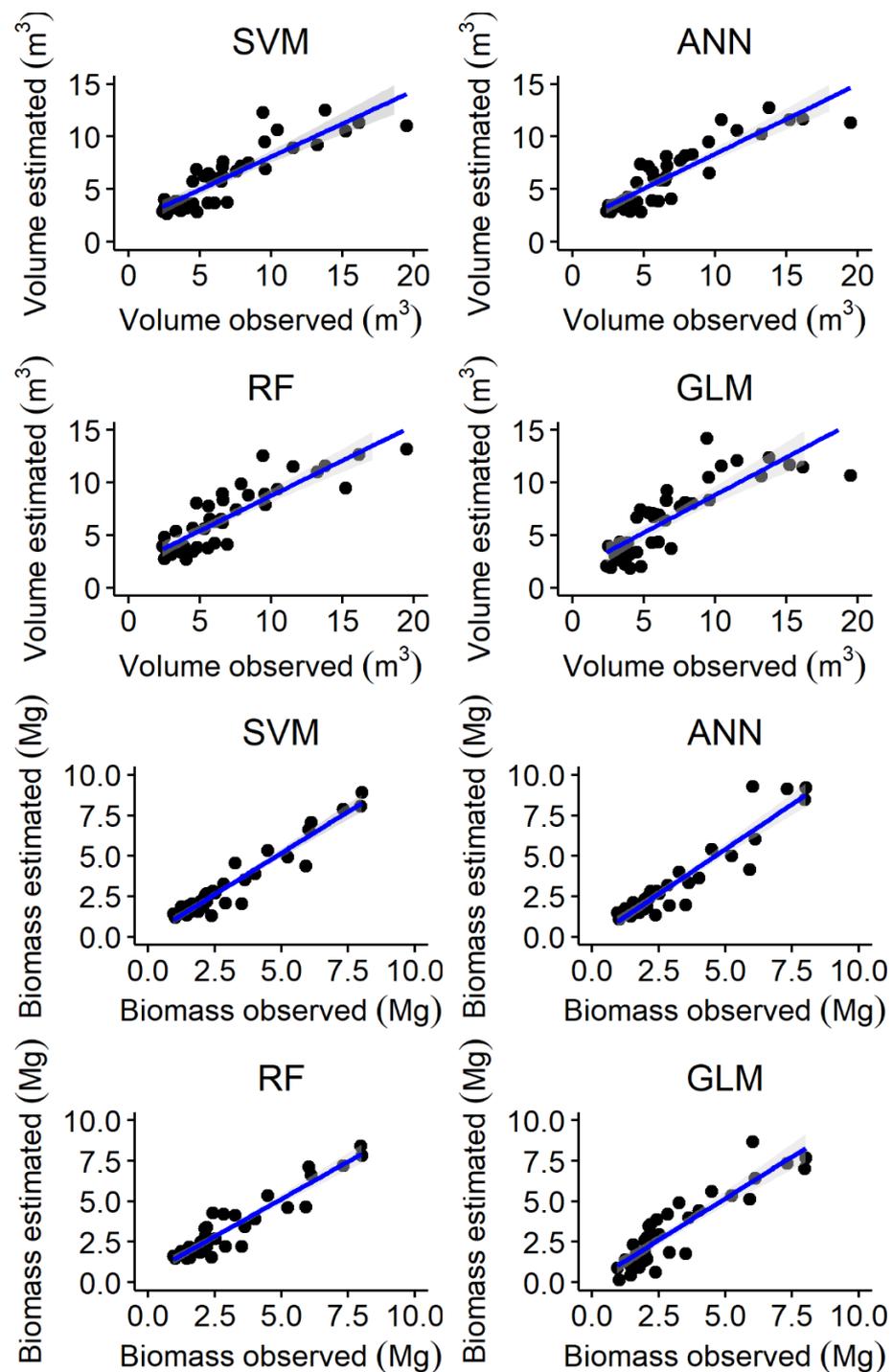


Figure 3. Observed and estimated values of volume and biomass by the different machine learning models, SVM, ANN, RF and GLM tested.

GLM showed the best performance to estimate the volume of commercial trees, with the highest $r_{y\hat{y}}$ and the lowest RSME and MAE for all repetitions (Table 3). RF had a $r_{y\hat{y}}$ close to the GLM model and the second-best performance for predicting volume. The ANN algorithm showed moderate performance, and the SVM had the worst performance for predicting the volume of commercial trees in the Amazon.

Table 3. Statistics of the tested machine learning models (SVM, ANN, RF and GLM) for modeling volume and biomass of commercial trees in the Southwestern Amazon in the municipality of Porto Acre, Acre, Brazil.

Variable	Model	RMSE	MAE	$r_{y\hat{y}}$
Volume	SVM	1.93 ± 0.54	1.19 ± 0.23	0.89 ± 0.04
	ANN	1.67 ± 0.36	1.13 ± 0.19	0.91 ± 0.04
	RF	1.82 ± 0.41	1.24 ± 0.20	0.90 ± 0.04
	GLM	1.82 ± 0.33	1.30 ± 0.18	0.89 ± 0.04
Biomass	SVM	1.15 ± 0.33	0.67 ± 0.15	0.92 ± 0.03
	ANN	1.10 ± 0.27	0.69 ± 0.13	0.92 ± 0.03
	RF	1.19 ± 0.31	0.76 ± 0.14	0.91 ± 0.03
	GLM	1.35 ± 0.27	0.97 ± 0.12	0.88 ± 0.04

Where: RMSE: Root Mean Square Error; MAE: mean absolute error; SVM: Support Vector Machine; ANN: Artificial Neural Networks; RF: Random Forest; GLM: Generalized Linear Model.

ANN showed the best performance for predicting biomass, with the highest $r_{y\hat{y}}$ and the lowest RMSE and MAE for all repetitions. RF also had the second-best performance for predicting biomass. SVM had the worst performance for the prediction of commercial tree biomass in the Southwestern Amazon (Table 3).

The means of RMSE and MAE varied over the repetitions for each technique. The RMSE and MAE averages of ANN and GLM showed the lowest values when estimating the volume of trees (Figure 3). ANN and RF showed the lowest RMSE and MAE over 50 repetitions in the cross-validation when estimating the biomass (Figure 4). SVM showed greater instability in the values of RMSE and MAE of the cross-validation and the highest values of RMSE and MAE for all variables evaluated in the present study.

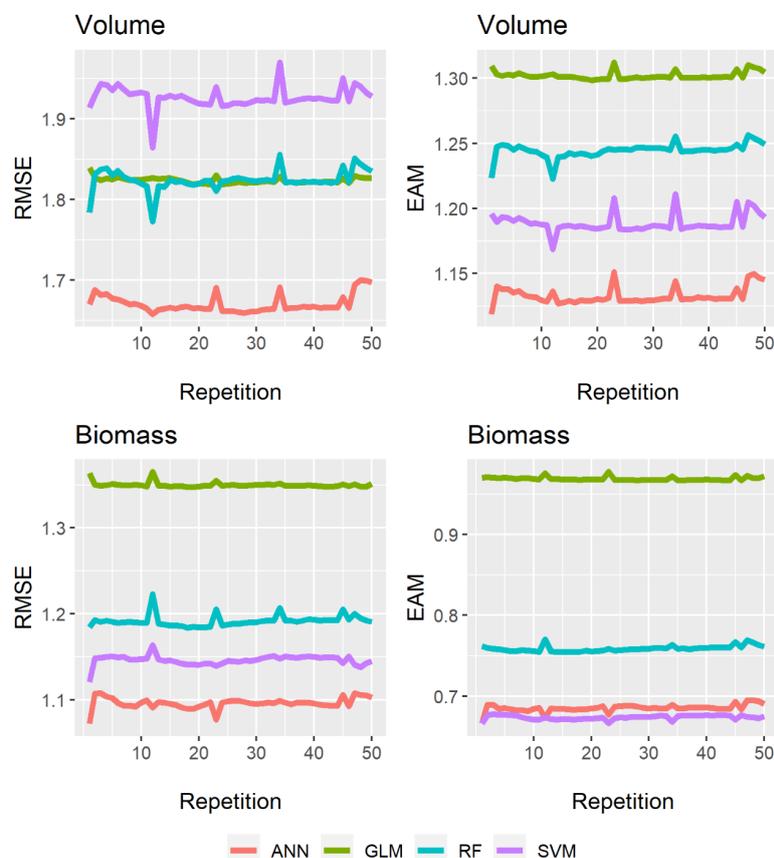


Figure 4. Root Mean Square Error (RMSE) of the machine learning models SVM, ANN, RF and GLM

in the modeling of the volume and biomass of commercial trees in the Southwestern Amazon in the municipality of Porto Acre, Acre, Brazil.

The Friedman test with the means of cross-validation RMSE showed that the predictions of the volume and biomass of commercial trees in the Amazon differed between the techniques ($p < 0.05$). Thus, the hypothesis that at least one average of one of the techniques differs from the others was accepted. The Nemenyi test indicated that the difference between the GLM model and the other techniques was greater than the calculated critical difference (CD) when estimating the volume of the trees. The calculated critical difference (CD) of the ANN was greater than the other machine learning techniques evaluated when estimating the biomass (Figure 5).

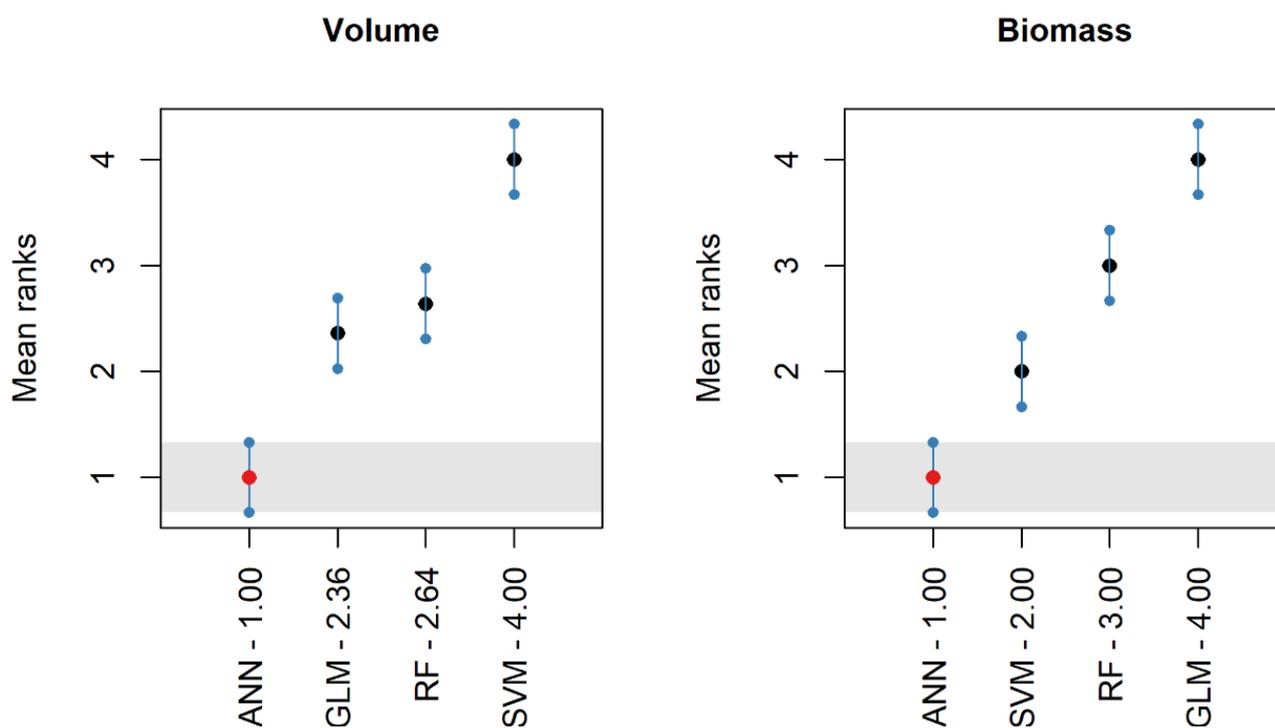


Figure 5. Nemenyi test in the cross-validation of the estimates of volume and biomass of commercial trees in the Southwestern Amazon in the municipality of Porto Acre, Acre, Brazil.

4. Discussion

4.1. Selection of Variables

The Boruta variable selection method is the preferred algorithm among the variable selection methods because it has a high computational efficiency for working on data sets with many predictors. The use of the Boruta algorithm for variable selection offers several advantages including model independence, robustness against multicollinearity, the ability to handle variables of different types, identification of irrelevant variables and statistical evaluation of variable importance [54]. One of the main advantages of Boruta is that it can handle noisy data and correlations between variables, which is particularly useful in real-world problems.

In this study, the method indicated that the diameter and height are fundamental to explaining the allometric attributes of the trees. The stem diameter is a good predictor [7,55,56] and this is an important advantage for practical use. However, the integration of tree height significantly reduces uncertainties [5,10,57]. The diameter of a tree is a critical variable, as it has a strong relationship with tree volume. The height of a tree is also important, as it can affect the distribution of the crown, which impacts the volume of the tree.

Besides that, the inclusion of the wood basic density variable is an important predictor for biomass. This variable, with the diameter of the trunk, the height of the tree and the type of forest (dry, moist or wet), are the most important predictors of the biomass [8]. The inclusion and combination of both provide better quality in adjustment and estimates [8,10].

Our analysis also showed that the bioclimatic and topographic variables were not significant for estimating volume and biomass in commercial trees in the Amazon. This is due to the low variability of this information in the area, in view of the uniform distribution of these characteristics in the study region. However, considering the role of climate in predicting forest attributes can provide more accurate estimates [5,58], since diameter-height relationships in trees depend on a series of physiological and environmental factors [59,60]. The maximum and minimum temperature, precipitation seasonality and degree of solar radiation have strong correlations with biomass [61,62].

4.2. Model Performance

In general, the machine learning models accurately estimated the production attributes of the evaluated commercial trees in the Southwestern Amazon. A major advantage of using machine learning methods over traditional models is their applicability to any number of variables [63]. This method is a very valuable procedure for working with data sets in large-scale databases [64,65] because it can manipulate continuous, categorical and binary data [66] and is able to adapt to complex and non-linear relationships between variables, in addition to dealing with interaction effects between them [67]. These models can encompass several types of information and it is possible to work with a single model for different situations.

The Friedman test confirms ($p < 0.05$) that the ANN model gave the best estimates of the volume and biomass of commercial trees in the Southwestern Amazon. The RMSE and MAE corroborate this statement, showing small differences in training and validation. ANN is considered an important non-parametric algorithm for estimating the biophysical parameters of the forest [21]. Neural networks can implicitly detect any complex nonlinear relationships between independent and dependent variables [68]. In contrast to conventional parametric approaches, ANN does not require any assumptions about the statistical distribution of the data.

RF presented intermediate results for estimating the volume and biomass variables. This method produces the most accurate and stable predictions [69]. This algorithm has been widely applied in ecological studies, as it can work with complex data analyses [67]. In addition, RF has been considered one of the best methods of classification and regression due to its high precision for estimation results, high calculation speed, robustness and the ability to predict important variables [70]. Decision tree-based algorithms are easy to apply since fewer parameters need to be estimated. Therefore, they have a high degree of automation [71].

SVM presented less precision in the volume estimates compared to the RMSE and MAE values. SVMs have the inconvenience of a delicate and computationally expensive hyperparameter adjustment. In addition, results for SVM compared to other methods showed only average accuracy. ANNs and RF generally produce better results than SVMs for regression tasks. The simple statistical procedures and the set methods were very competitive for classification [72].

GLM showed less precision in the biomass estimates. This may be related to the link function used. GLMs have the characteristic of being able to choose the residual distribution family, which is important in the case of non-parametric models such as those that follow a Poisson distribution and negative binomial errors [73]. The choice of one function over another may explain the low performance.

As forest volume and biomass are important for forest management, global change monitoring and the modeling of forest productivity, there is a need for reliable methods of assessing and monitoring forest production [21]. The results presented here suggest a

new alternative to predict these forest attributes, providing valuable information for forest managers and for the formation of public policies.

Machine learning models offer a powerful solution to the challenges posed by high variability and complexity in the Amazon Forest ecosystem, enabling accurate estimations of tree volume and biomass. This is crucial for effective forest management and decision-making regarding resource utilization and conservation efforts in the face of deforestation and land-use change.

A limitation of our study is the limited spatial variability of the climatic and topographic data used. However, future research utilizing larger and more comprehensive datasets holds the potential to further enhance the estimation of biomass and volume by incorporating a broader range of spatially diverse environmental variables such as soil characteristics and vegetation indices, leading to improved accuracy and precision. Machine learning's advantage lies in its ability to handle complex and high-dimensional variables that are difficult to incorporate into traditional models. By leveraging machine learning algorithms, researchers can easily include a wide range of variables with intricate relationships, resulting in more accurate and comprehensive estimations.

Overall, while machine learning offers promising opportunities for predicting tree volume and biomass in the Amazon, there are several challenges that need to be addressed to ensure that these models are accurate and reliable. These challenges include selecting appropriate features, choosing the best algorithm and optimizing the model parameters to achieve the best performance. Another challenge is the limited availability and quality of training data, particularly in remote areas of the Amazon. Accurate predictions require large and representative datasets that capture the diversity of tree species and environmental conditions in the region.

5. Conclusions

The tested machine learning methods (SVM, ANN, RF and GLM) are useful and efficient tools for estimating the volume and biomass of commercial trees in the Southwestern Amazon. This study represents a new approach to estimating these attributes linked to forest production. ANN is the most suitable for estimating the volume and biomass of commercial trees.

Further research is needed to improve the accuracy of the model and to test its applicability to other tree species and forest types. To achieve this, new techniques, additional variables and larger and more diverse datasets should be explored, along with efforts to minimize biases and uncertainties in the ground-truth measurements. These efforts may lead to more accurate and reliable estimates of tree volume and biomass, which are crucial for sustainable forest management and conservation.

Author Contributions: Data collection, writing, methodology, formal analysis, figures—S.J.S.S.d.R. and F.M.B.R.; data collection, review, editing, supervision—C.M.M.E.T., L.A.G.J., S.C.R., P.H.V., B.L.S.S. and V.T.M.d.M.J.; review, editing, supervision—L.P.R., I.B.C., I.d.S.T.J., Á.B.T.V. and M.P.M.X.R. All authors have read and agreed to the published version of the manuscript.

Funding: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES) (88882.437301/2019-01; 88881.311787/2018-01 and 88887.319055/2019-00), Conselho Nacional de Desenvolvimento Científico e Tecnológico—Brazil (CNPq) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais—Brazil (FAPEMIG).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The supporting datasets are stored within the institutions of the authors and available on demand.

Acknowledgments: This research was supported logistically by Empresa Fox Laminados Ltd.a. We thank the company's owner, Antônio Aparecido Barlati and his team for their help in the planning, harvesting and sawmill stages of the study. We thank the Laboratory of Wood Anatomy of the Federal

University of Acre (headed by Moisés Silveira Lobão) and the Laboratory of Wood Panels and Energy of the Federal University of Viçosa (headed by Angelica de Cássia Oliveira Carneiro). This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES)—CAPES/PrInt [88881.311787/2018-01 and 88887.319055/2019-00].

Conflicts of Interest: The authors declare no conflict of interest. Supporting entities had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Gleason, C.J.; Im, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91. [[CrossRef](#)]
- Jachowski, N.R.A.; Quak, M.S.Y.; Friess, D.A.; Duangnamon, D.; Webb, E.L.; Ziegler, A.D. Mangrove biomass estimation in Southwest Thailand using machine learning. *Appl. Geogr.* **2013**, *45*, 311–321. [[CrossRef](#)]
- Zhao, K.; Popescu, S.; Meng, X.; Pang, Y.; Agca, M. Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sens. Environ.* **2011**, *115*, 1978–1996. [[CrossRef](#)]
- Goodman, R.C.; Phillips, O.L.; del Castillo Torres, D.; Freitas, L.; Cortese, S.T.; Monteagudo, A.; Baker, T.R. Amazon palm biomass and allometry. *For. Ecol. Manag.* **2013**, *310*, 994–1004. [[CrossRef](#)]
- Chave, J.; Réjou-Méchain, M.; Búrquez, A.; Chidumayo, E.; Colgan, M.S.; Delitti, W.B.; Duque, A.; Eid, T.; Fearnside, P.M.; Goodman, R.C.; et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Change Biol.* **2014**, *20*, 3177–3190. [[CrossRef](#)]
- Alvarez, E.; Duque, A.; Saldarriaga, J.; Cabrera, K.; Salas, G.D.L.; del Valle, I.; Lema, A.; Moreno, F.; Orrego, S.; Rodríguez, L. Tree above-ground biomass allometries for carbon stocks estimation in the natural forests of Colombia. *For. Ecol. Manag.* **2012**, *267*, 297–308. [[CrossRef](#)]
- Chambers, J.Q.; dos Santos, J.; Ribeiro, R.J.; Higuchi, N. Tree damage, allometric relationships, and above-ground net primary production in central Amazon forest. *For. Ecol. Manag.* **2001**, *152*, 73–84. [[CrossRef](#)]
- Chave, J.; Andalo, C.; Brown, S.; Cairns, M.A.; Chambers, J.Q.; Eamus, D.; Fölster, H.; Fromard, F.; Higuchi, N.; Kira, T.; et al. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* **2005**, *145*, 87–99. [[CrossRef](#)]
- Cummings, D.; Kauffman, J.B.; Perry, D.A.; Hughes, R.F. Aboveground biomass and structure of rainforests in the southwestern Brazilian Amazon. *For. Ecol. Manag.* **2002**, *163*, 293–307. [[CrossRef](#)]
- Goodman, R.C.; Phillips, O.L.; Baker, T.R. The importance of crown dimensions to improve tropical tree biomass estimates. *Ecol. Appl.* **2014**, *24*, 680–698. [[CrossRef](#)]
- Keller, M.; Palace, M.; Hurr, G. Biomass estimation in the Tapajos National Forest, Brazil: Examination of sampling and allometric uncertainties. *For. Ecol. Manag.* **2001**, *154*, 371–382. [[CrossRef](#)]
- Lima, A.J.N.; Suwa, R.; de Mello Ribeiro, G.H.P.; Kajimoto, T.; Dos Santos, J.; Da Silva, R.P.; De Souza, C.A.S.; De Barros, P.C.; Noguchi, H.; Ishizuka, M. Allometric models for estimating above- and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil. *For. Ecol. Manag.* **2012**, *277*, 163–172. [[CrossRef](#)]
- Nelson, B.W.; Mesquita, R.; Pereira, J.L.G.; de Souza, S.G.A.; Batista, G.T.; Couto, L.B. Allometric regressions for improved estimate of secondary forest biomass in the central Amazon. *For. Ecol. Manag.* **1999**, *117*, 149–167. [[CrossRef](#)]
- Nogueira, E.M.; Fearnside, P.M.; Nelson, B.W.; Barbosa, R.I.; Keizer, E.W.H. Estimates of forest biomass in the Brazilian Amazon: New allometric equations and adjustments to biomass from wood-volume inventories. *For. Ecol. Manag.* **2008**, *256*, 1853–1867. [[CrossRef](#)]
- Vieilledent, G.; Vaudry, R.; Andriamanohisoa, S.F.D.; Rakotonarivo, O.S.; Randrianasolo, H.Z.; Razafindrabe, H.N.; Rakotoarivony, C.B.; Ebeling, J.; Rasamoelina, M. A universal approach to estimate biomass and carbon stock in tropical forests using generic allometric models. *Ecol. Appl.* **2012**, *22*, 572–583. [[CrossRef](#)] [[PubMed](#)]
- Reis, L.P.; de Souza, A.L.; dos Reis, P.C.M.; Mazzei, L.; Soares, C.P.B.; Torres, C.M.M.E.; da Silva, L.F.; Ruschel, A.R.; Rêgo, L.J.S.; Leite, H.G. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. *Ecol. Eng.* **2018**, *112*, 140–147. [[CrossRef](#)]
- da Rocha, S.J.S.S.; Torres, C.M.M.E.; Jacovine, L.A.G.; Leite, H.G.; Gelcer, E.M.; Neves, K.M.; Schettini, B.L.S.; Villanova, P.H.; da Silva, L.F.; Reis, L.P.; et al. Artificial neural networks: Modeling tree survival and mortality in the Atlantic Forest biome in Brazil. *Sci. Total Environ.* **2018**, *645*, 655–661. [[CrossRef](#)]
- Júnior, I.d.S.T.; Torres, C.M.M.E.; Leite, H.G.; de Castro, N.L.M.; Soares, C.P.B.; Castro, R.V.O.; Farias, A.A. Machine learning: Modeling increment in diameter of individual trees on Atlantic Forest fragments. *Ecol. Indic.* **2020**, *117*, 106685. [[CrossRef](#)]
- Reis, L.P.; de Souza, A.L.; Mazzei, L.; dos Reis, P.C.M.; Leite, H.G.; Soares, C.P.B.; Torres, C.M.M.E.; da Silva, L.F.; Ruschel, A.R. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. *For. Ecol. Manag.* **2016**, *382*, 161–167. [[CrossRef](#)]
- Corona-Núñez, R.; Mendoza-Ponce, A.; López-Martínez, R. Model selection changes the spatial heterogeneity and total potential carbon in a tropical dry forest. *For. Ecol. Manag.* **2017**, *405*, 69–80. [[CrossRef](#)]

21. Nandy, S.; Singh, R.; Ghosh, S.; Watham, T.; Kushwaha, S.P.S.; Kumar, A.S.; Dadhwal, V.K. Neural network-based modelling for forest biomass assessment. *Carbon Manag.* **2017**, *8*, 305–317. [[CrossRef](#)]
22. Santi, E.; Paloscia, S.; Pettinato, S.; Fontanelli, G.; Mura, M.; Zolli, C.; Maselli, F.; Chiesi, M.; Bottai, L.; Chirici, G. The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas. *Remote Sens. Environ.* **2017**, *200*, 63–73. [[CrossRef](#)]
23. Foody, G.M.; Cutler, M.E. Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. *Ecol. Model.* **2006**, *195*, 37–42. [[CrossRef](#)]
24. Diamantopoulou, M.J.; Özçelik, R.; Crecente-Campo, F.; Eler, Ü. Estimation of Weibull function parameters for modelling tree diameter distribution using least squares and artificial neural networks methods. *Biosyst. Eng.* **2015**, *133*, 33–45. [[CrossRef](#)]
25. Diamantopoulou, M.J.; Özçelik, R. Evaluation of different modeling approaches for total tree-height estimation in Mediterranean Region of Turkey. *For. Syst.* **2012**, *21*, 383. [[CrossRef](#)]
26. Vieira, G.C.; de Mendonça, A.R.; da Silva, G.F.; Zanetti, S.S.; da Silva, M.M.; dos Santos, A.R. Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. *Sci. Total Environ.* **2018**, 619–620, 1473–1481. [[CrossRef](#)]
27. Ingram, J.C.; Whittaker, R.J.; Dawson, T.P. Tree Structure and Diversity in Human-Impacted Littoral Forests, Madagascar. *Environ. Manag.* **2005**, *35*, 779–798. [[CrossRef](#)]
28. Zhao, Q.; Yu, S.; Zhao, F.; Tian, L.; Zhao, Z. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. *For. Ecol. Manag.* **2019**, *434*, 224–234. [[CrossRef](#)]
29. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-resolution global maps of 21st-century forest cover change. *Science* **2013**, *342*, 850–853. [[CrossRef](#)]
30. Brienen, R.J.W.; Phillips, O.L.; Feldpausch, T.R.; Gloor, E.; Baker, T.R.; Lloyd, J.; Lopez-Gonzalez, G.; Monteagudo-Mendoza, A.; Malhi, Y.; Lewis, S.L.; et al. Long-term decline of the Amazon carbon sink. *Nature* **2015**, *519*, 344–348. [[CrossRef](#)]
31. Antonelli, A.; Sanmartín, I. Why are there so many plant species in the Neotropics? *Taxon* **2011**, *60*, 403–414. [[CrossRef](#)]
32. Ter Steege, H.; Pitman, N.C.A.; Sabatier, D.; Baraloto, C.; Salomão, R.P.; Guevara, J.E.; Phillips, O.L.; Castilho, C.V.; Magnusson, W.E.; Molino, J.-F.; et al. Hyperdominance in the Amazonian Tree Flora. *Science* **2013**, *342*, 1243092. [[CrossRef](#)]
33. Cardoso, D.; Särkinen, T.; Alexander, S.; Amorim, A.M.; Bittrich, V.; Celis, M.; Daly, D.C.; Fiaschi, P.; Funk, V.A.; Giacomini, L.L.; et al. Amazon plant diversity revealed by a taxonomically verified species list. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10695–10700. [[CrossRef](#)] [[PubMed](#)]
34. Andrade, V.H.F.; Machado, S.D.A.; Filho, A.F.; Botosso, P.C.; Miranda, B.P.; Schöngart, J. Growth models for two commercial tree species in upland forests of the Southern Brazilian Amazon. *For. Ecol. Manag.* **2019**, *438*, 215–223. [[CrossRef](#)]
35. Luize, B.G.; Magalhães, J.L.L.; Queiroz, H.; Lopes, M.A.; Venticinque, E.M.; Novo, E.M.L.D.M.; Silva, T.S.F. The tree species pool of Amazonian wetland forests: Which species can assemble in periodically waterlogged habitats? *PLoS ONE* **2018**, *13*, e0198130. [[CrossRef](#)] [[PubMed](#)]
36. ter Steege, H.; Vaessen, R.W.; Cárdenas-López, D.; Sabatier, D.; Antonelli, A.; de Oliveira, S.M.; Pitman, N.C.A.; Jørgensen, P.M.; Salomão, R.P. The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Sci. Rep.* **2016**, *6*, 29549. [[CrossRef](#)]
37. Fortini, L.B. Integrated models show a transient opportunity for sustainable management by tropical forest dwellers. *For. Ecol. Manag.* **2019**, *438*, 233–242. [[CrossRef](#)]
38. Putz, F.; Sist, P.; Fredericksen, T.; Dykstra, D. Reduced-impact logging: Challenges and opportunities. *For. Ecol. Manag.* **2008**, *256*, 1427–1433. [[CrossRef](#)]
39. Dang, A.T.N.; Nandy, S.; Srinet, R.; Luong, N.V.; Ghosh, S.; Kumar, A.S. Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam. *Ecol. Inform.* **2018**, *50*, 24–32. [[CrossRef](#)]
40. Ye, Q.; Yu, S.; Liu, J.; Zhao, Q.; Zhao, Z. Aboveground biomass estimation of black locust planted forests with aspect variable using machine learning regression algorithms. *Ecol. Indic.* **2021**, *129*, 107948. [[CrossRef](#)]
41. Zarzosa, P.S.; Navarro-Cerrillo, R.M.; Mc Cubbin, E.P.; Cruz, G.; Lopez, M. Biomass estimation models for four priority Prosopis species: Tools required for forestry management in overexploited arid ecosystems. *J. Arid. Environ.* **2023**, *209*, 104904. [[CrossRef](#)]
42. Jiang, X.; Li, G.; Lu, D.; Moran, E.; Batistella, M. Modeling Forest Aboveground Carbon Density in the Brazilian Amazon with Integration of MODIS and Airborne LiDAR Data. *Remote Sens.* **2020**, *12*, 3330. [[CrossRef](#)]
43. Zhu, Y.; Feng, Z.; Lu, J.; Liu, J. Estimation of Forest Biomass in Beijing (China) Using Multisource Remote Sensing and Forest Inventory Data. *Forests* **2020**, *11*, 163. [[CrossRef](#)]
44. Prăvălie, R.; Niculiță, M.; Roșca, B.; Patriche, C.; Dumitrașcu, M.; Marin, G.; Nita, I.-A.; Bandoc, G.; Birsan, M.-V. Modelling forest biomass dynamics in relation to climate change in Romania using complex data and machine learning algorithms. *Stoch. Environ. Res. Risk Assess.* **2023**, *37*, 1669–1695. [[CrossRef](#)]
45. Diamantopoulou, M.J.; Milios, E. Modelling total volume of dominant pine trees in reforestations via multivariate analysis and artificial neural network models. *Biosyst. Eng.* **2010**, *105*, 306–315. [[CrossRef](#)]
46. Alencar, J.D.C.; De Almeida, R.A.; Fernandes, N.P. Fenologia de espécies florestais em floresta tropical úmida de terra firme na Amazônia Central. *Acta Amaz.* **1979**, *9*, 163–199. [[CrossRef](#)]
47. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; Moraes, G.J.L.; Sparovek, G. Köppen's climate classification map for Brazil. *Meteorol. Z.* **2013**, *22*, 711–728. [[CrossRef](#)]
48. Governo do Estado do Acre Zoneamento Ecológico-Econômico. *Fase III: Escala 1:250.000*; Semapi: Rio Branco-Acre, Brazil, 2021.

49. Romero, F.M.B.; Jacovine, L.A.G.; Ribeiro, S.C.; Torres, C.M.M.E.; da Silva, L.F.; Gaspar, R.d.O.; da Rocha, S.J.S.S.; Staudhammer, C.L.; Fearnside, P.M. Allometric Equations for Volume, Biomass, and Carbon in Commercial Stems Harvested in a Managed Forest in the Southwestern Amazon: A Case Study. *Forests* **2020**, *11*, 874. [\[CrossRef\]](#)
50. Husch, B.; Beers, T.W.; Kershaw, J.A., Jr. *Forest Mensuration*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016.
51. Associação Brasileira de Normas Técnicas. 11941: *Madeira—Determinação Da Densidade Básica*; Associação Brasileira de Normas Técnicas: São Paulo, Brazil, 2003.
52. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [\[CrossRef\]](#)
53. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [\[CrossRef\]](#)
54. Speiser, J.L.; Miller, M.E.; Tooze, J.; Ip, E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **2019**, *134*, 93–101. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Kuyah, S.; Dietz, J.; Muthuri, C.; Jamnadass, R.; Mwangi, P.; Coe, R.; Neufeldt, H. Allometric equations for estimating biomass in agricultural landscapes: I. Aboveground biomass. *Agric. Ecosyst. Environ.* **2012**, *158*, 216–224. [\[CrossRef\]](#)
56. Yuen, J.Q.; Fung, T.; Ziegler, A.D. Review of allometric equations for major land covers in SE Asia: Uncertainty and implications for above- and below-ground carbon estimates. *For. Ecol. Manag.* **2016**, *360*, 323–340. [\[CrossRef\]](#)
57. Rutishauser, E.; Noor'an, F.; Laumonier, Y.; Halperin, J.; Ie, R.; Hergoualc'h, K.; Verhot, L. Generic allometric models including height best estimate forest biomass and carbon stocks in Indonesia. *For. Ecol. Manag.* **2013**, *307*, 219–225. [\[CrossRef\]](#)
58. Feldpausch, T.R.; Banin, L.; Phillips, O.L.; Baker, T.R.; Lewis, S.L.; Quesada, C.A.; Affum-Baffoe, K.; Arets, E.J.M.M.; Berry, N.J.; Bird, M.; et al. Height-diameter allometry of tropical forest trees. *Biogeosciences* **2011**, *8*, 1081–1106. [\[CrossRef\]](#)
59. Lines, E.R.; Zavala, M.A.; Purves, D.W.; Coomes, D.A. Predictable changes in aboveground allometry of trees along gradients of temperature, aridity and competition. *Glob. Ecol. Biogeogr.* **2012**, *21*, 1017–1028. [\[CrossRef\]](#)
60. Marshall, A.; Willcock, S.; Platts, P.; Lovett, J.; Balmford, A.; Burgess, N.; Latham, J.; Munishi, P.; Salter, R.; Shirima, D.; et al. Measuring and modelling above-ground carbon and tree allometry along a tropical elevation gradient. *Biol. Conserv.* **2012**, *154*, 20–33. [\[CrossRef\]](#)
61. Banin, L.; Feldpausch, T.R.; Phillips, O.L.; Baker, T.R.; Lloyd, J.; Affum-Baffoe, K.; Arets, E.J.M.M.; Berry, N.J.; Bradford, M.; Brienen, R.J.W.; et al. What controls tropical forest architecture? Testing environmental, structural and floristic drivers. *Glob. Ecol. Biogeogr.* **2012**, *21*, 1179–1190. [\[CrossRef\]](#)
62. Taylor, P.G.; Cleveland, C.C.; Soper, F.; Wieder, W.R.; Dobrowski, S.; Doughty, C.E.; Townsend, A.R. Greater stem growth, woody allocation, and aboveground biomass in Paleotropical forests than in Neotropical forests. *Ecology* **2019**, *100*, e02589. [\[CrossRef\]](#)
63. Loh, W.-Y. Classification and regression trees. *WIREs Data Min. Knowl. Discov.* **2011**, *1*, 14–23. [\[CrossRef\]](#)
64. Abdel-Rahman, E.M.; Mutanga, O.; Adam, E.; Ismail, R. Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 48–59. [\[CrossRef\]](#)
65. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [\[CrossRef\]](#)
66. Ali, J.; Khan, R.; Ahmad, N.; Maqsood, I. Random Forests and Decision Trees. *Int. J. Comput. Sci. Issues* **2012**, *9*, 272–278.
67. Reise, J.; Kukulka, F.; Flade, M.; Winter, S. Characterising the richness and diversity of forest bird species using National Forest Inventory data in Germany. *For. Ecol. Manag.* **2019**, *432*, 799–811. [\[CrossRef\]](#)
68. Lazri, M.; Ameur, S. Combination of support vector machine, artificial neural network and random forest for improving the classification of convective and stratiform rain using spectral features of SEVIRI data. *Atmos. Res.* **2018**, *203*, 118–129. [\[CrossRef\]](#)
69. Sun, T.; Chen, F.; Zhong, L.; Liu, W.; Wang, Y. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geol. Rev.* **2019**, *109*, 26–49. [\[CrossRef\]](#)
70. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
71. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M.J.O.G.R. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [\[CrossRef\]](#)
72. Meyer, D.; Leisch, F.; Hornik, K. The support vector machine under test. *Neurocomputing* **2003**, *55*, 169–186. [\[CrossRef\]](#)
73. Lopatin, J.; Dolos, K.; Hernández, H.; Galleguillos, M.; Fassnacht, F. Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. *Remote Sens. Environ.* **2016**, *173*, 200–210. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.