ISSN 2395-8618

Artificial Intelligence Models to Estimate Biomass of Tropical Forest Trees

Razer Anthom Nizer Rojas Montaño, Carlos Roberto Sanquetta, Jaime Wojciechowski, Eduardo Mattar, Ana Paula Dalla Corte, and Eduardo Todt

***Abstract*—Artificial Intelligence Models (AI) were tested for aboveground dry biomass estimation of 4,004 trees collected in forests throughout the Tropics, and compared to a classic Allometric model of literature. The data come from various countries, in the Neotropics, Africa, Southeast Asia, and Oceania. Statistical analysis of the data showed that they do not have normal distribution and homocedasticity, which violates the regression assumptions. Examination of bias, precision and accuracy of the Allometric model and the AI models revealed that KNN (K Nearest Neighbors), ANN (Artificial Neural Network) and SVM (Support Vector Machines) have strong estimation power of the biomass of tropical trees, comparable or greater than the linear regression (Allometric model), which is considered the state of the art. It was concluded that AI models can be considered an interesting alternative to the regression technique, especially when the data do not show normality and homoscedasticity, which is the case of biomass of tropical forest trees. In particular, SVM showed better accuracy for data considered. *Index Terms*—machine learning, allometry, carbon, data mining, neural networks, support vector machines**

I. Introduction

Forests are considered important global carbon reservoirs, storing about $296Gt$ of carbon. Carbon concentrations are found in tropical forest of South America and Central Africa, stocking about $120tC.ha^{-1}$, while the world average is $75tC.ha^{-1}$. However, tropical forests have been the main victims of deforestation and degradation [1]. This has led to increase of accumulated emissions of Greenhouse Gases (GHG) by activities using land and forests, of $490 \pm 180GtCO_2$ in 1970 to $680 \pm 300GtCO_2$ 2010 [2].

The largest fraction of carbon stored in the forests of the world is in your living biomass, with $250GtC$ [1], and still there is great uncertainty about these stocks, mainly due to insecurity of estimation on large scale. A complicating factor is that any model to be applied on a larger scale must be based on direct measures, which are complex, costly and destructive [3].

It is essential to develop precise and accurate models of large-scale carbon stocks, but this is not a simple task. There are several variables that interfere in calculations, such as composition and structure of the vegetation, specific information such as density or specific mass of tissue, the carbon content in tissue, the method for calculation of areas and reliability of forest inventory, among others. One of the most important of these factors is the modeling methodology used to estimate the biomass or individual carbon from dendrometric variables, such as the diameter and height of trees.

Allometry is one of the best known indirect methods of estimation of biomass and individual carbon, which is usually materialized via simple input regression models - only with a diameter at breast height (dbh) as an independent variable - double entry, with dbh and height as independent variables - or triple entry, including the density or specific mass of the specie. Another robust methodological alternative, but less flexible, is the application of so-called growth factors. These modeling approaches are widely used and widely found in the literature [4].

Despite the current and widespread use of allometric models to estimate the biomass of trees, literature alerts that regression should respect some basic assumptions, and you should not use it indiscriminately without these assumptions. These assumptions are four: 1. variables must be normally distributed; 2. should have a linear relationship between the dependent and independent variables; 3. variables must be measured reliably; and 4. the variables must have homogeneous variances [5]. Although these assumptions are crucial to give validity to the estimates, they are rarely investigated and/or reported in quantification studies of forest biomass, which represents a risk to the estimation process.

The use of Artificial Intelligence (AI) techniques are a completely different approach to allometry via regression for individual biomass estimation. These techniques offer flexibility, simplicity and versatility, having the potential to estimate forest biomass in a way comparable to allometric classic models usually employed. Perhaps the most important feature in this context is the fact that the AI techniques in principle not require to attend the regression assumptions [6].

AI techniques have been applied in different scientific fields and sectors of human activities. In the last decade several studies have been published on the use in forest science ([7], [8], [9], [10], [11], [12]). One of the standard techniques of AI (machine learning often used in data mining) was recently explored [6], demonstrating its potential in quantification of

Manuscript received on February 12, 2017, accepted for publication on May 7, 2017, published on July 30, 2018.

Felix Freitag and Leandro Navarro are with the Department of Computer Architecture (DAC) at the Universitat Politècnica de Catalunya (UPC), Jordi Girona 1-3, 08034 Barcelona, Spain (e-mail: {felix, leandro}@ac.upc.edu).

The authors are with DINF – Federal University of Paraná (UFPR), Brazil (email: {razer,todt}@inf.ufpr.br), and BIOFIX – Federal University of Paraná (UFPR), Brazil (email: {sanquetta,jaimewo,eduardomattar,anacorte}@ufpr.br).

ISSN 2395-8618

individual tree biomass. Although data mining is a promising technique, some issues became to solve: Are there other AI techniques that can also be applied, or that are better than allometry? Do these techniques respond positively in situations where the data have high dispersion, as in case of biomass species of tropical trees?

This study aims to analyze data of more than 4,000 trees, collected in various regions of the tropics, including the Neotropics, Africa and Southeast Asia, with respect to the assumptions of regression models, and analyze AI techniques K-nearest neighbor, artifical neural networks and support vector machines, to estimate the variable of interest compared with a tipical allometric model.

II. Material and Methods

The data used in this study were provided by Jerome Chave, Director of Research at French Scientific Research French Centre - CNRS, France. These data correspond to 4,004 observations of diameter at breast height in cm (dbh), total height in m (ht), total aboveground dry biomass in kg (b), wood density (basic specific mass, ρ). These data were collected in the following countries: Australia, South Africa, Brazil, Cambodia, Cameroon, Central African Republic, Colombia, Cost Rica, French Guiana, Gabon, Ghana, Guadeloupe, India, Indonesia, Mexico, Madagascar, Malaysia, Mozambique, New Guinea, Peru, Puerto Rico, Tanzania, Venezuela and Zambia.

Data were randomly separated into two parts, 70% for adjustment or training and 30% to validate the estimates.

The basic descriptive statistics for these data were calculated: arithmetic average, standard deviation, coefficient of variation, and maximum and minimum of the variables b, dbh, ht and ρ. It was also investigated the linear correlation between these variables, using Pearson's correlation coefficient. It was held the normality test (Lilliefors and Shapiro-Wilk) of these variables.

Four biomass estimation models (b) were evaluated depending on the variables dbh, ht and ρ, having as a witness the allometric model proposed by [13] and computed by Equation 1.

A. Allometric Model

The allometric model used for comparison was proposed by [13], based on Schumacher-Hall's model [14], and described by Equation 1.

$$\lg(\hat{b}) = \beta_0 + \beta_1 \lg(dbh) + \beta_2 \lg(ht) + \beta_3 \lg(\rho) \qquad (1)$$

where:

- lg : decimal log of previous variables;
- β_0, β_1, β_2, β_3: model coeficients to be adjusted.

B. KNN (K-Nearest Neighbor)

KNN is a non parametric method used in data mining called *instance based learning*, which employs the values of closest neighbors to be estimated. This method applied to biomass estimation is described in more details in [6]. In this work dbh, ht and ρ were applied as proximity variables, to have a direct comparison of its estimated power with allometric model. We used the Euclidean distance (Equation 2), which showed the best results in experiments compared with Squared, Manhattan and Chebychev, and inverse distance weighting [15], [6]. Such choices depend on the performance of adjustment, that is, several simulations were made using the amount of neighbors, types of distance and types of weighting, and we choose the best performance. This procedure is indicated in the study of Bradzil [16], where estimates are made from one to five closer neighbors.

$$d(p,q) = \sqrt{(dbh_p - dbh_q)^2 + (ht_p - ht_q)^2 + (\rho_p - \rho_q)^2} \qquad (2)$$

where:

- p e q : trees variables;
- dbh_p, dbh_q: dbh of tree p and q;
- ht_p, ht_q: total height of tree p and q;
- ρ_p, ρ_q: specific mass of tree p and q.

The method uses a technique known as Cross-Validation where each instance is compared to other from sample, being selected the instance with shorter distance. The biomass estimated for that instance is the biomass of the instance with lowest distance from it. The method allows the use of n nearest neighbors of the tree in question, and the value of the estimated biomass is a balance between the biomass of trees with smaller distances among vectors of weighting by the inverse of the distance ($1/d$). This study employed three nearest neighbors (Equation 3) and weighted by the inverse of distance:

$$\hat{b}_p = \frac{b_1 w_1 + b_2 w_2 + b_3 w_3}{w_1 + w_2 + w_3} \qquad (3)$$

- $\hat{b}_p$: estimated biomass of tree p;
- w_n: $\frac{1}{d(p,q_n)}$ closest trees weighted, from tree p to tree q_n;
- $d(p,q_n)$: distance from tree p to one of three closest trees q_n;
- b_1, b_2, b_3: real biomass of three closest trees, mensured by distance $d(p,q_n)$.

C. Artificial Neural Networks

ANN is a machine learning technique used for various purposes, also recently used to address forestry problems [11]. It is a comutational system composed by simple processing units, highly connected. These units, or neurons, compute mathematical functions and their results are processed together in the network layers. The connections simulate biological synapses and have associated weights to inputs. These weights

ISSN 2395-8618

are adjusted as the whole set is trained, that is, learning acquired knowledge [17].

There are several ANN configurations, the main ones are Multilayer Perceptron (MLP), based on the Radial Basis Function (RBF) and based on Vector Quantization (LVQ) [18]. Here we used the MLP networks, commonly used in studies of forest area [19].

A neuron receives values and returns a result. The input values are weighted, combined (added) and submitted to a mathematical function f_a. Thus, if the vector $x = [x_1, x_2, ...x_m]^t$ is the input of a neuron and the vector $w = [w_1, w_2, ..., w_m]^t$ is the weights applied to each input, the result of the neuron $f'(x)$ is:

$$\begin{aligned} u &= \sum_{i=3}^{m} x_i w_i \\ f'(x) &= f_a(u) \end{aligned} \tag{4}$$

The f_a function is called *activation function* and it can be of various types, the most used are: linear, threshold and sigmoid. In this case we used the sigmoid, the most used according to [18]. Neurons are arranged in one or more layers, and one neuron receives as input the outputs of the previous layer's neurons, and its output is put in the next layer. The input layer receives data to be processed, and the layer that gives the result is called output layer. The other layers are known as hidden.

To solve nonlinearly separable problems we should use one or more hidden layers [20]. In this work we employed Multilayer Perceptron (MLP) with sigmoidal function in their hidden layers. In the case of regression problems, the output can not be discretized, and a decimal value is returned. For the ANN training we used back-propagation algorithm [21]. It consists of two parts: *forward* and *backward*. In phase forward the object is presented to the network, neurons calculate their values to the specific weights and the activation function produces its output value. This is done until the output neurons have their calculated values. The computed result is compared with the expected result and this difference is the error on the network. The error value is then used in step backward to adjust the weights of neurons.

There are several parameters to be configured to find an ANN that gives acceptable estimates and comparable to allometric model. Since the amount of layers, how many neurons in each layer, learning rate (multiplicative value for weights adjustment in the learning process), number of epochs (number of times network is presented to the input data), among others. In this work, the data were entered into multiple networks containing a hidden layer of neurons ranging from 5 to 100. The learning rate was tested between 0.1 and 0.9, with steps of 0.2. The momentum varied between 0.001 and 1. Using a training set with 30% of population. These data are used to verify the mistake of training, in which the increase of error can stop the process without all the expected number of times to run.

D. Support Vector Machines

Support Vector Machines (SVM) is a machine learning technique used in many situations for pattern recognition, obtaining results superior to those achieved by other learning techniques in various situations, such as categorization of texts in image analysis and bioinformatics [22]. The technique is grounded by statistical learning theory, developed by [23], [24]. SVMs can be applied to problems of classification and regression, with potential use in various forestry issues.

Like other methods, several parameters must be set to obtain a SVM model comparable to other tested models. The main parameters are the cost (C), which gives balance between accuracy and complexity of the model, and the kernel function used to design values for a larger, where data have more probability to be linearly separable [25]. The type of kernel function used here was *RBFKernel*, which has the *gamma* parameter, that controls the shape of the peaks when the data is passed to another dimension. Small values indicate pointed peaks, that is, small bias and high variance, which may cause overfitting when the model learns only the entered values and lose the ability to generalize, giving poor results for new entries. Large values result in soft forms, with high bias and low variance and can harm the learning process.

They were tested more than 150 combinations of C and $gamma$, to find the combination with best correlation and residual values. The C parameter was varied from 1,000 to 10,000, initially with steps of 1,000. When a promising region was identified, the steps were reduced to 500, 100, 50 and 10. The $gamma$ value was varied from 0.01 to 0.09 with steps of 0.02, and varied from 0.1 up to 0.9, with steps of 0.2, in each test performed.

E. Quality Assessment of Estimates

The performances of the estimates obtained with the four techniques were evaluated according to three numerical indicators on the average [26], [27], that is, bias, precision and accuracy.

Bias is given by:

$$\bar{e} = \frac{\sum_{i=1}^{n}(\hat{b}_i - b_i)}{n} \tag{5}$$

$$\bar{e}\% = \frac{\bar{e}}{\bar{b}} 100 \tag{6}$$

Precision is given by:

$$s_e = \sqrt{\frac{\sum_{i=1}^{n}(\hat{b}_i - \bar{e} - b_i)^2}{n-1}} \tag{7}$$

$$s_e\% = \frac{s_e}{\bar{b}} 100 \tag{8}$$

Accuracy is given by:

$$m_b = \sqrt{\frac{\sum_{i=1}^{n}(\hat{b}_i - b_i)^2}{n-1}} \tag{9}$$

ISSN 2395-8618

$$m_b\% = \frac{m_b}{\bar{b}}100 \tag{10}$$

where:

– $\hat{b}_i$: estimated biomass (kg);
– b_i: real biomass (kg);
– $\bar{b}$: average of real biomass (kg);
– n: number of observations.

In addition, it was calculated R^2_{adj} and S_{yx} for the allometric model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(b_i - \hat{b}_i)^2}{\sum_{i=1}^{n}(b_i - \bar{b}_i)^2} \tag{11}$$

$$R^2_{adj} = 1 - \frac{n-1}{n-k}(1 - R^2) \tag{12}$$

$$S_{yx} = \sqrt{\frac{\sum_{i=1}^{n}(b_i - \hat{b}_i)^2}{n-k}} \tag{13}$$

where:

– k: number of coefficients of the model.

The closer to zero (in module) is $\bar{e}\%$ the smaller is bias and better is the performance of the model. The smaller is the dispersion expressed by $s_e\%$, the greater is accuracy. The smaller $m_b\%$ the more accurate and closer to the target are the estimates. The interpretation of these indicators can be seen in Figure 1, making an analogy with shots at a target.

Estimates were also appreciated by the linear correlation between the estimated and actual values and the absolute residual distribution graphs (real value - estimated value).

III. Results and Discussion

A. Basic statistics of data

Examining the four variables analyzed (b, dbh, ht and ρ), it appears that none of them presents normality by testing Lilliefors and Shapiro-Wilk, as in their original form as the log-transformed. Therefore, thus it is shown that the first regression assumption [5] is not attended.

The data shows wide dispersion of the variables considered in the modeling (Table I). Correlations between biomass (b) and the independent variables dbh and ht were 0.79 and 0.61, respectively, which were significant to 95% of probability. On the other hand, the correlation ρ was -0.05, which are regarded as null. Therefore, changes in biomass may be significantly explained by their respective variations in diameter and height, but the specific mass alone does not explain the variations in biomass. Considering logarithmic transformation of variables, Pearson's correlation coefficients of biomass were 0.96, 0.86 and 0.07, respectively, which indicates the degree of association of variables increases with such transformation.

The relation of the dependent variable (b) presented curvilinear behavior, with greater dispersion in large trees, that is, large trees showed significant variation in their biomass. Something similar occurs with the height variable, that is, taller trees show greater variation in biomass. There is no direct relationship between density and biomass. When the logarithmic transformations of variables are considered, dbh and ht show a linear relationship with biomass, not occurring with ρ. Thus, it appears that, given the logarithmic model presented in [13], two variables follow the regression presupposition of variable linearity assumption [5], and one not.

The adjustment of allometric model of 2,802 data by the method of ordinary least squares resulted in the following equation:

$$\lg(\hat{b}) = -1.21356 + 2.01484\lg(dbh) + 0.888954\lg(ht) + 0.83138\lg(\rho) \tag{14}$$

With results R^2_{adj} = 0.9730 and S_{yx} = 0.1518.

Given this equation, it was possible to verify the normality and homoscedasticity of its residual. The results of the tests Lilliefors and Shapiro-Wilk stated lack of normality at 95% probability and graphical analysis of residues (Figure 3a) showed absence of heteroscedasticity along the estimate line. However, when the residuals of logarithmic variable are converted into biomass values it is noticed that there is variation of dispersion over adjust line, indicating that they do not behave homogeneously (Figure 3b). Therefore, it is evident that the assumptions of normality and homoscedasticity of residuals are not attended when adjusting allometric model.

B. Avaliation of AI Models

The results of the three AI models, alternatives to the allometric model proposed by [13] showed that KNN, ANN and SVM provide estimates with about the same degree of bias, accuracy and precision (Table II). In training with 70% of the data it was found bias below 10% for all models, which indicates that the estimates do not exhibit pronounced trends over or underestimation. Low bias was also observed in validations, which is a positive aspect, considering that the application of models to independent data to those used in training is consistent and trends free.

Given the high variability of the observed data, considering they refer to occurring trees throughout the Tropics and of different species and sizes, all models showed low precision and accurary indexes, although they have not shown bias, high correlation between the observed values and estimated was detected by the Pearson coefficient. The performance of the models was similar in training base, but SVM model excelled in accuracy (that joins bias and accuracy) which showed the smaller percentage values for both, the set of train data and test data. In testing base, ANN presented better correlation, althogh precision and accuracy are not the best.

Observing the graphic distribution of residuals, it is understood that estimates behave more evenly along the estimation line in comparison with test. It also appears that there is a tendency of heterocedasticity of residual along

ISSN 2395-8618

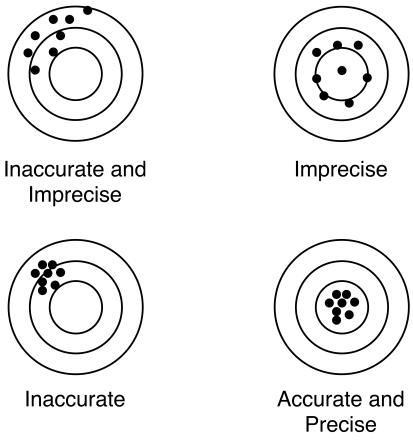


Fig. 1. Precision and Accuracy, where the goal is the center of the target

TABLE I
MEASURES

Variable	Average	Standard deviation	CV%	Minimum	Maximum
Biomass (b in Kg)	1,134.14	3,917.97	345.46	1.23	76,063.52
Diameter at breast height (dbh, in cm)	23.99	24.09	100.41	5.00	212.00
Total height (ht, in m)	16.04	10.77	67.17	1.30	70.70
Specific mass (ρ in $g.cm^{-3}$)	0.63	0.16	25.94	0.09	1.20

TABLE II
RESULTS

Data	Model	Bias%	Precision%	Accuracy%	r
Train (n = 2802)	Allometric	-2.22	100.09	100.10	0.9570
	KNN	1.30	91.13	91.16	0.9479
	ANN	7.75	91.71	92.04	0.9566
	SVM	-3.12	86.43	86.49	0.9619
Test (n = 1202)	Allometric	4.29	84.87	129.71	0.9580
	KNN	6.60	116.69	178.36	0,9092
	ANN	-0.04	133.01	133.01	0.9626
	SVM	-9.86	152.31	152.62	0.9426

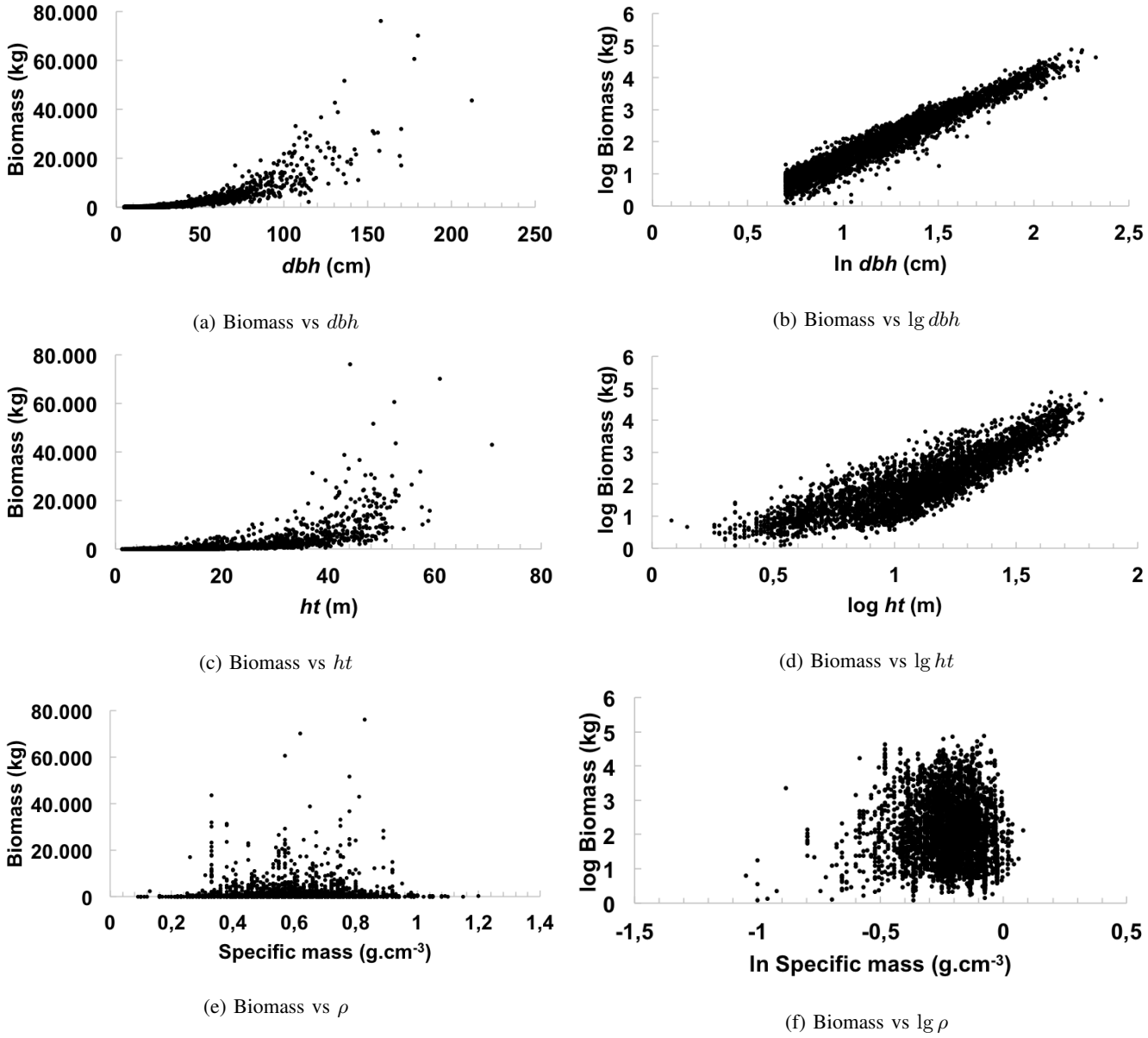
the x-axis, increasing the variability for big trees. In the distributions corresponding to the validation it is more evident, with large dispersion of residual, in addition to detect trend of underestimation in large trees.

Native forests, particularly tropical, have high structural and dimensional variability. The major cause of this variability, especially in their biomass, is the occurrence of large trees. There was a direct linear relationship between the density of trees with dbh above 70 cm and its biomass in tropical forests [28]. The authors evaluated the importance of large trees in the stock biomass of tropical forests and rated intrinsic and extrinsic aspects related to changes that occur in different regions of the tropics. In this study it is evident that the major cause of loss of accuracy and bias is the increased variability in biomass of large trees. Calculating the biomass of trees above 70 cm dbh, it is found that, although few in number compared to smaller trees (about 6%), these individuals represent about 67% of the biomass.

Models to estimate the biomass of trees should seek to reduce uncertainties. These uncertainties are due to the nature of the data and aspects inherent to modeling. In this article we didn't emphasize the intrinsic or extrinsic variations in biomass of trees, but the modeling itself. However, it is important that these variations are clarified, since the accuracy of the models is highly dependent on the behavior of the data, especially the dispersion of biomass in large trees, a fact also reported in this study.

Regarding the models evaluated here, the main concern is that in many cases the conditions or assumptions of regression might not be attended. Data with large variations, such as those analyzed here, can cheat the requirements for application of regression models, as reported in the related literature. The diametric distribution of tropical forests is decreasing, which implies on non-normality of biomass data. The great variability in the biomass production of large dimensions trees implies heterocedasticity. Due to the nature of forest biomass data, these assumptions may lead to uncertainties in the inferences and the estimated potential of regression models

ISSN 2395-8618



(a) Biomass vs dbh

(b) Biomass vs $\lg dbh$

(c) Biomass vs ht

(d) Biomass vs $\lg ht$

(e) Biomass vs ρ

(f) Biomass vs $\lg \rho$

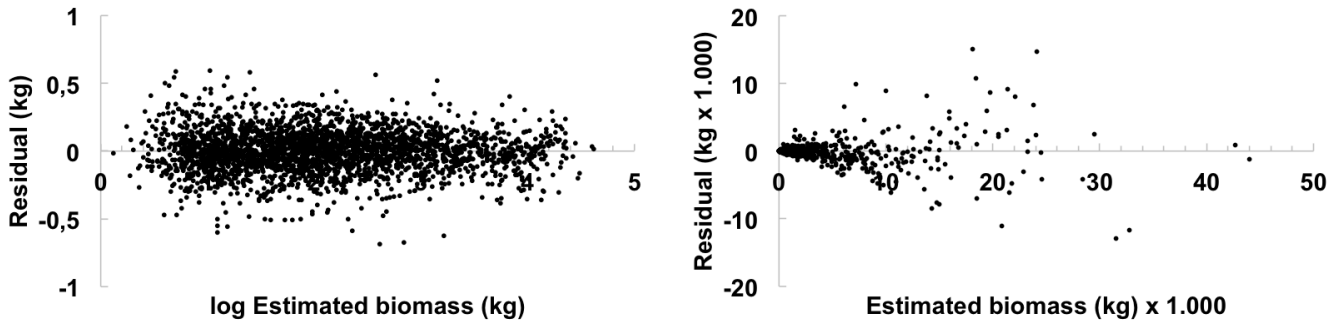
Fig. 2. Relation among biomass and dbh, height and specific mass of 4,004 individual trees at tropical forests.

[6]. Therefore, other modeling approaches need to be sought.

In this study we tested some AI methods, K-nearest neighbor, artificial neural networks and support vector machines, as alternative to allometric ones. They are mathematical and computational methods using different principles of statistical regression. The KNN technique showed comparable accuracy to the allometric model. This technique is used in other areas and its application to estimate carbon in trees was recently introduced [9]. The estimation procedure of this technique is based on the average of the known values of the closest neighbors of a point to estimate, and in this case we analyzed three neighbors, but other options could be considered. Comparative modeling studies to quantify biomass in restoration plantings in the Atlantic indicated that this technique can generate accurate estimates and the setting of the number of neighbors affects the results [6]. The authors concluded that 3 to 5 neighboring enable better performance technique and as the number of neighbors increases no loss of accuracy. The amount of data also influences the performance of this technique in predicting biomass, requiring a large mass of data for the technique to work properly [15].

ANN also provided biomass estimates with the same degree of accuracy and precision that models based on regression. Although the technique already known and reasonably used in forestry, its application in quantification of forest biomass is more restricted to applications of geotechnology, such as remote sensing. Estimates of biomass stocks in a fragment of natural forest with satellite images IKONOS employment,

ISSN 2395-8618



(a) Residual vs $\lg b$

(b) Residual vs b

Fig. 3. Residual of estimation of log decimal variable of biomass with allometric model (Equation 14) and related residual of variable biomass in function of estimates of 2,802 individual trees of tropical forests.

were held recently and with excellent performance [29]. Similar research conducted in tropical forests in Indonesia have also indicated that the estimated biomass ANN applied to Landsat 5 TM satellite images resulted in appropriate and strongly correlated with estimates forest inventory data [30].

SVM is a technique that is poorly explored in forestry, particularly in quantifying forest biomass. Studies on quantification of biomass Juniperus pinchotii in the United States with images derived from its top and SVM as classifier, produced promising results [31]. A review of the different machine learning techniques for applications in estimating biomass pointed to the potential of this technique for this purpose, highlighting its flexibility and ability to properly process large amounts of data is presented in [32].

AI models need to be more widely known and tested in forestry applications. In general, techniques such as KNN, SVM and ANN have the potential development and application in any field, such as forest inventory, forest planning, harvesting and forestry supply systems, etc. These techniques present auspicious prospects in the quantification of biomass and carbon in trees and forests, either as individual estimation strategy or per unit area, with forest inventories or remote sensing data. This potential needs to be further explored, especially when large amounts of data need to be analyzed and interpreted and there are restrictions to the application of conventional techniques such as regression.

IV. Conclusions

In this work, several machine learning algorithms were applied to estimate biomass of tropical forest trees. We trained KNN, SVM and ANN with more than 4,000 trees and results were compared to regression models.

Also, AI methods are presented as a suitable alternative to the state of art allometric techniques, since results obtained here are compared to allometric ones.

The main considerations are:

– Artificial Intelligence Models have strong estimation power for biomass of tropical trees, comparable or superior to regression (allometric model), which is considered state of the art;
– Analysis of a large amount of biomass data tropical forest trees shows that the assumptions underlying the use of regression models are frequently violated and often simply neglected;
– AI models constitute an attractive alternative to the regression technique, especially when the data do not show normality and homoscedasticity, which is the case of biomass of tropical forest trees.

V. Acknowledgements

The authors would like to thank Dr. Jerome Chave for the kind disposition of data used in this research.

References

[1] FAO – Food and Agriculture Organization, “Global forest resources assessments 2015,” FAO, Roma, Itália, Tech. Rep., 2015.
[2] UNFCCC – United Nations Framework Convention on Climate Change, “Fourth assessment report: Summary for policymakers,” Cambridge University, Geneva, Switzerland, Tech. Rep., 2007.
[3] L. F. Watzlawick, F. F. Kirchner, and C. R. Sanquetta, “Estimativa de biomassa e carbono em floresta com araucária utilizando imagens do satélite ikonos ii,” *Ciência Florestal*, vol. 19, pp. 169–181, 2009.
[4] C. R. Sanquetta, A. P. Corte, and F. da Silva, “Biomass expansion factor and root-to-shoot ratio for pinus in brazil,” *Carbon Balance and Management*, vol. 6, no. 1, pp. 1–8, 2011. [Online]. Available: http://dx.doi.org/10.1186/1750-0680-6-6
[5] J. W. Osborne and E. Waters, “Four assumptions of multiple regression that researchers should always test,” *Practical Assessment Research and Evaluation*, vol. 8, pp. 1–8, 2002.
[6] C. R. Sanquetta, J. Wojciechowski, A. P. Dalla Corte, A. Behling, S. Péllico Netto, A. L. Rodrigues, and M. N. I. Sanquetta, “Comparison of data mining and allometric model in estimation of tree biomass,” *BMC Bioinformatics*, vol. 16, no. 1, pp. 1–9, 2015.
[7] M. J. Diamantopoulou, “Predicting fir trees stem diameters using artificial neural network models,” *The Southern African Forestry Journal*, vol. 205, no. 1, pp. 39–44, 2005. [Online]. Available: http://dx.doi.org/10.2989/10295920509505236
[8] H. G. Leite, M. L. M. da Silva, D. H. B. Binoti, L. Fardin, and F. H. Takizawa, “Estimation of inside-bark diameter and heartwood diameter for tectona grandis linn. trees using artificial neural networks,” *European Journal of Forest Research*, vol. 130, no. 2, pp. 263–269, 2010. [Online]. Available: http://dx.doi.org/10.1007/s10342-010-0427-7

ISSN 2395-8618

[9] C. R. Sanquetta, J. Wojciechowski, A. P. D. Corte, A. L. Rodrigues, and G. C. B. Maas, "On the use of data mining for estimating carbon storage in the trees," *Carbon Balance Manag*, vol. 8, pp. 6–6, Jun 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3693975/

[10] R. V. O. Castro, C. P. B. Soares, F. B. Martins, and H. G. Leite, "Crescimento e produção de plantios comerciais de eucalipto estimados por duas categorias de modelos," *Pesquisa Agropecuária Brasileira*, vol. 48, pp. 287 – 295, 03 2013.

[11] A. Imada, "A literature review: Forest management with neural network and artificial intelligence," in *Neural Networks and Artificial Intelligence*, ser. Communications in Computer and Information Science, V. Golovko and A. Imada, Eds. Springer International Publishing, 2014, vol. 440, pp. 9–21.

[12] F. A. A. M. N. Soares, E. L. Flores, C. D. Cabacinha, G. A. Carrijo, and A. C. P. Veiga, "Recursive diameter prediction for calculating merchantable volume of eucalyptus clones using multilayer perceptron," *Neural Computing and Applications*, vol. 22, no. 7, pp. 1407–1418, 2013. [Online]. Available: http://dx.doi.org/10.1007/s00521-012-0823-7

[13] J. Chave, C. Andalo, S. Brown, M. A. Cairns, J. Q. Chambers, D. Eamus, H. Fölster, F. Fromard, N. Higuchi, T. Kira, J.-P. Lescure, B. W. Nelson, H. Ogawa, H. Puig, B. Riéra, and T. Yamakura, "Tree allometry and improved estimation of carbon stocks and balance in tropical forests," *Oecologia*, vol. 145, no. 1, pp. 87–99, 2005. [Online]. Available: http://dx.doi.org/10.1007/s00442-005-0100-x

[14] F. Schumacher and D. Hall, "Logarithmic expression of timber-tree volume," *Journal of Agricultural Research*, vol. 47, no. 9, pp. 719–734, 1933.

[15] F. Mognon, A. P. D. Corte, C. R. Sanquetta, T. G. Barreto, and J. Wojciechowski, "Estimativas de biomassa para plantas de bambu do gênero Guadua," *Revista Ceres*, vol. 61, pp. 900 – 906, 12 2014.

[16] P. B. Brazdil, C. Soares, and J. P. da Costa, "Ranking learning algorithms: Using ibl and meta-learning on accuracy and time results," *Machine Learning*, vol. 50, no. 3, pp. 251–277, 2003. [Online]. Available: http://dx.doi.org/10.1023/A:1021713901879

[17] A. Braga, A. C. Carvalho, and T. B. Ludermir, *Redes Neurais Artificiais: Teoria e aplicações*. LTC Editora, 2007, vol. 2. [Online]. Available: http://www.worldcat.org/isbn/9788521615644

[18] S. Haykin, *Redes neurais: princípios e prática*. Porto Alegre, RS: Bookman, 2001.

[19] M. L. M. da Silva Binoti, "Emprego de redes neurais artificiais em mensuração florestal e manejo florestal," Ph.D. dissertation, Universidade Federal de Viçosa, 2012.

[20] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, pp. 303–314, 1989.

[21] D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group, Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. Cambridge, MA, USA: MIT Press, 1986.

[22] A. C. Lorena and A. C. de Carvalho, "Uma introdução às support vector machines," *Revista de Informática Teórica e Aplicada*, vol. 14, no. 2, pp. 43–67, 2007.

[23] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[24] V. N. Vapnik, *Statistical learning theory*. John Wiley and Sons, 1998.

[25] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Trans. on Electronic Computers*, vol. EC-14, pp. 326–334, 1965.

[26] H. E. Burkhart and M. Tomé, *Modeling Forest Trees and Stands*. Springer Netherlands, 2012.

[27] H. Pretzsch, *Forest Dynamics, Growth and Yield*. Springer-Verlag Berlin Heidelberg, 2009.

[28] J. W. F. Slik, G. Paoli, K. McGuire, I. Amaral, J. Barroso, M. Bastian, L. Blanc, F. Bongers, P. Boundja, C. Clark, M. Collins, G. Dauby, Y. Ding, J.-L. Doucet, E. Eler, L. Ferreira, O. Forshed, G. Fredriksson, J.-F. Gillet, D. Harris, M. Leal, Y. Laumonier, Y. Malhi, A. Mansor, E. Martin, K. Miyamoto, A. Araujo-Murakami, H. Nagamasu, R. Nilus, E. Nurtjahya, A. Oliveira, O. Onrizal, A. Parada-Gutierrez, A. Permana, L. Poorter, J. Poulsen, H. Ramirez-Angulo, J. Reitsma, F. Rovero, A. Rozak, D. Sheil, J. Silva-Espejo, M. Silveira, W. Spironelo, H. ter Steege, T. Stevart, G. E. Navarro-Aguilar, T. Sunderland, E. Suzuki, J. Tang, I. Theilade, G. van der Heijden, J. van Valkenburg, T. Van Do, E. Vilanova, V. Vos, S. Wich, H. Wöll, T. Yoneda, R. Zang, M.-G. Zhang, and N. Zweifel, "Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics," *Global Ecology and Biogeography*, vol. 22, no. 12, pp. 1261–1271, 2013. [Online]. Available: http://dx.doi.org/10.1111/geb.12092

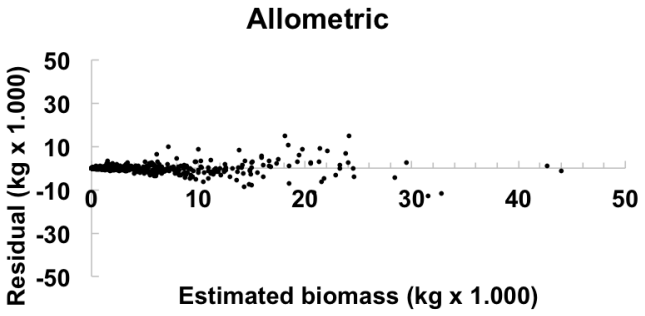
[29] A. S. Ferraz, V. P. Soares, C. P. B. Soares, C. A. A. S. Ribeiro, D. H. B. Binoti, and H. G. Leite, "Estimativa do estoque de biomassa em um fragmento florestal usando imagens orbitais," *Floresta e Ambiente*, vol. 21, pp. 286 – 296, 09 2014.

[30] G. M. Foody, M. E. Cutler, J. McMorrow, D. Pelz, H. Tangki, D. S. Boyd, and I. Douglas, "Mapping the biomass of bornean tropical rain forest from remotely sensed data," *Global Ecology and Biogeography*, vol. 10, no. 4, pp. 379–387, 2001. [Online]. Available: http://www.jstor.org/stable/2665383

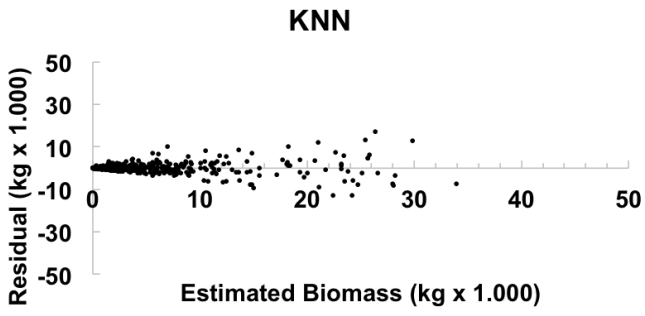
[31] M. Mirik, S. Chaudhuri, B. Surber, S. Ale, and R. J. Ansley, "Evaluating biomass of juniper trees (juniperus pinchotii) from imagery-derived canopy area using the support vector machine classifier," *Advances in Remote Sensing*, vol. 2, pp. 181–192, 2013.

[32] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, "Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data," *Remote Sensing*, vol. 7, pp. 16 398–16 421, Dec. 2015.

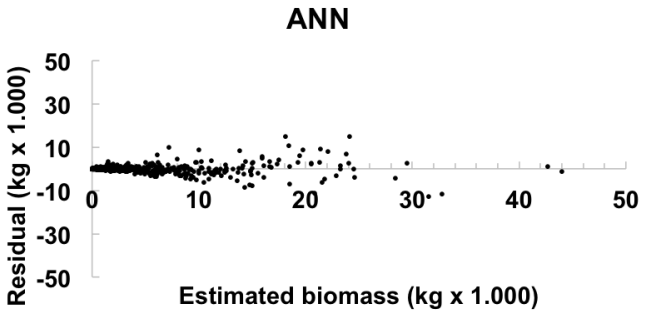
ISSN 2395-8618



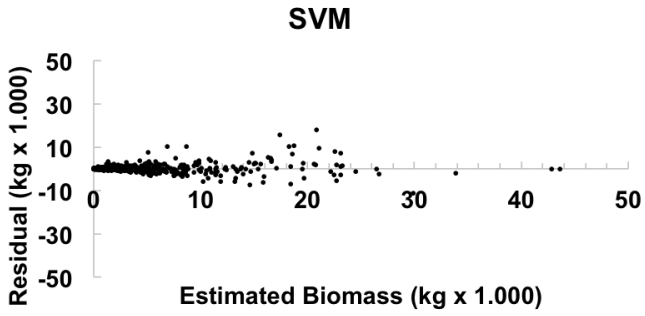
(a) Residual - Training data



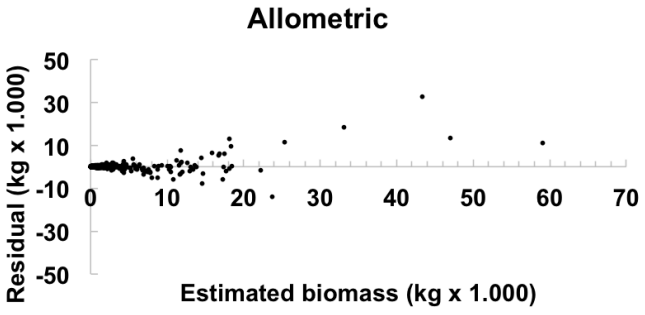
(b) Residual - Training data



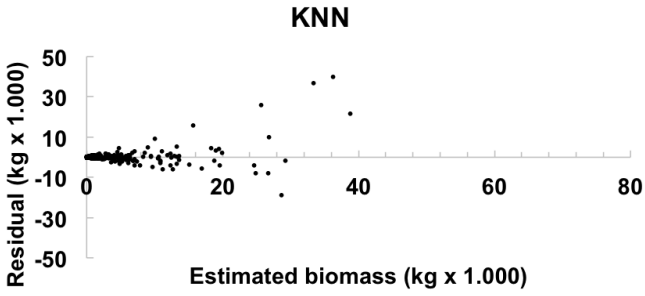
(c) Residual - Training data



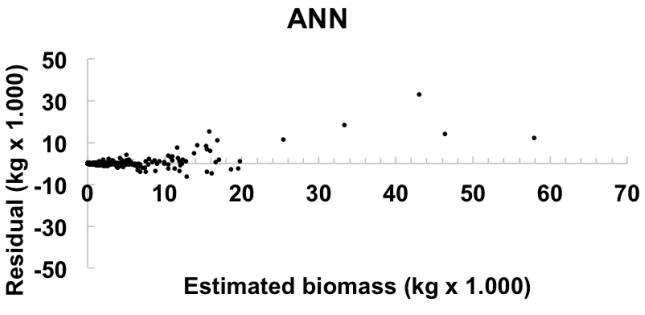
(d) Residual - Training data



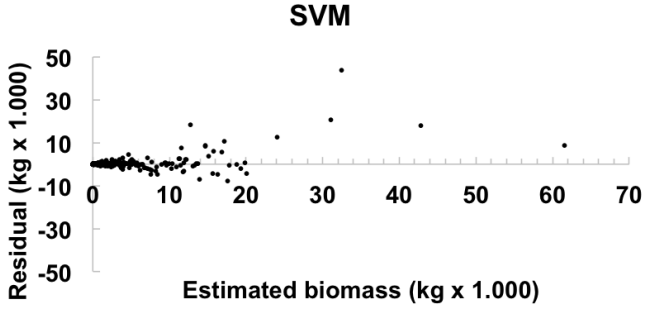
(e) Residual - Test data



(f) Residual - Test data



(g) Residual - Test data



(h) Residual - Test data

Fig. 4. Graphical distribution of residual of four models to estimate dry biomass above ground, *dbh*, total height and specific mass of 4004 individual trees at tropical forests.