

# Modeling the leaf area of *Ormosia paraensis* Ducke by statistical models and artificial neural networks

Ronaldo Oliveira dos Santos<sup>1\*</sup>, Rubiene Neto Soares<sup>1</sup>, Mickelly Paula Queiroz Pimentel<sup>1</sup>, Jadson Coelho de Abreu<sup>1</sup>, Robson Borges de Lima<sup>1</sup>, and Breno Marques da Silva e Silva<sup>1</sup>

<sup>1</sup>Universidade do Estado do Amapá (UEAP), Av. Presidente Getúlio Vargas, Centro, Macapá, Amapá, Brasil. \*Corresponding author (santosro.ueap@gmail.com).

Received: 14 April 2018; Accepted: 22 September 2018; doi:10.4067/S0718-58392018000400511

# ABSTRACT

The leaf is a very important plant structure because it allows gas exchange and transformation of light energy into chemical energy. The objective of this research study was to test artificial neural networks (ANNs) to estimate the leaf area (LA) of *Ormosia paraensis* Ducke and compare their performance with adjusted statistical models. One hundred forty leaves were selected from seedlings of the species at the leaf age of 100 days after sowing (DAT). The LA was calculated by indirect estimation with ImageJ software, and the linear length (L) and width (W) dimensions were measured with a ruler graduated in centimeters. Afterward, 90 leaves were randomly separated to generate mathematical equations of LA (Y) as a function of linear dimensions (L and W), and 50 leaves to validate the equations. Similarly, 100 networks of the multilayer perceptron (MLP) type were trained with the backpropagation algorithm and the best networks were selected for later validation. The choice of the best equations and ANNs was based on precision and dispersion statistics. The quadratic equation obtained from model (4) demonstrated consistent statistics with the coefficient of determination (R<sup>2</sup>aj = 87.39) and a low result for the standard error of estimate (SEE = 12.07%). For the ANNs, a correlation coefficient (Ryŷ), varying between 0.9316 and 0.9521, was obtained in the training phase, while it ranged from 0.8522 to 0.8825 in the validation phase, generating lower residues. It is concluded that the ANN performance was higher compared with the conventional regression technique.

Key words: Adjustment of equations, artificial intelligence, forest species, ImageJ software.

## INTRODUCTION

The Ormosia paraensis Ducke tree species belongs to the Fabaceae family and is popularly known in Brazil as tento, tenteiro, or olho-de-cabra; it is widely distributed in the Amazon vegetation of dense ombrophilous forest of terra firme and in capoeira vegetation (Silva et al., 2015). This species has been of interest from the economic point of view and is raw material for timber forest products (TFPs), mainly in the furniture industry, as well as for non-timber forest products (NTFPs) in which seeds are used to produce artisanal bio-jewels. In addition, it is important because it is also focuses on the environment where it is being used to recover degraded areas by considering its N fixation potential, which is a vital and intrinsic characteristic of the family it belongs to.

In ecological, agronomic, and ecophysiological studies, knowledge of the forest species leaf area (LA) constitutes relevant information to understand plant responses to environmental factors. For this reason, the leaf is the organ with the highest specificity when performing vital plant functions, mainly to achieve photosynthesis and gas exchange via evapotranspiration. Mota et al. (2014) point out that studies involving LA provides the foundation for understanding the

transpiration rate,  $CO_2$  assimilation rate,  $O_2$  release rate, and plant vigor. Therefore, the development of methodologies seeking to accurately estimate LA, and simultaneously doing it in less time and at a lower cost, are indispensable for companies aiming to optimize production.

Two methods are acknowledged in the literature to estimate LA, namely, destructive and non-destructive methods. In summary, the destructive method is characterized by good accuracy to obtain this variable; however, it has the disadvantage of requiring a considerable amount of time and higher cost to acquire sophisticated equipment for the analysis. In addition, the inconvenience of such a method destroys the LA, which prevents leaf expansion monitoring over time. On the other hand, the second method allows the quick estimation of LA with good precision because it requires the dimensional parameters of the leaf blade without needing to destroy the leaf, thus making it possible to follow the crop cycle (Malagi et al., 2010).

It is possible to estimate LA by measuring leaf blade linear dimensions of length (L) and width (W), thereby establishing a relationship of functionality by statistical regression models. Regression analysis is a statistical tool that can promote reliable estimates by model adjustment and selection: it has been used with an emphasis on solving most forest problems, especially when trying to obtain estimates by biometric relationships (Schneider et al., 2009).

Many regression models have already been successfully used with agronomic species (Cardozo et al., 2014; Zanetti et al., 2017), shrubs (Souza and Habermann, 2014; Carvalho et al., 2017), and native forest trees (Santos, 2016). However, there are few studies using alternative tools such as artificial intelligence (artificial neural networks, ANNs) to estimate the LA of Amazon plants. The ANNs are computational techniques that present a model inspired by the neural structure of intelligent organisms that acquire knowledge through experience, providing a single accurate answer, and which have provided satisfactory results compared with other techniques such as traditional regression models (Gorgens et al., 2009; Binoti et al., 2013).

Thus, modeling the LA of Amazon forest species by statistical models and neural networks emerges as a great opportunity in the field of forest science because such studies of Amazon native forest species have yet to be developed. The literature does not include reports about non-destructive methods to estimate the LA of *O. paraensis*. The proposed hypothesis was that the LA of the species under study could be accurately expressed both by obtaining the best equation and ANN architecture. Therefore, the objective of this research study was to test ANNs to estimate the LA of *O. paraensis* and compare their performance with adjusted regression models.

## MATERIALS AND METHODS

The leaf selection criterion was based on leaf external quality; intact leaves that were free of any apparent damage caused by xylophagous organisms were selected. The leaves of the young plants were collected from *Ormosia paraensis* Ducke seedlings in an experiment entitled "Storage and sanitation of tento olho-de-cabra seeds – *Ormosia paraensis* Ducke" conducted in the Laboratory of Physiology and Forest Seeds of the Universidade do Estado do Amapá (UEAP) in Macapá, Amapá, Brazil.

A total of 140 leaves were collected from the seedlings at the leaf age of 100 days after sowing (DAT) from seeds stored for 8 mo in permeable (paper) and impermeable (plastic) packaging at ambient temperature 20 °C and at -10 °C. These young plants were grown in a greenhouse at UEAP in plastic trays (11 cm × 11 cm × 2.5 cm) containing sand as substrate and which received daily irrigation.

#### Data collection and processing

Once the leaves were obtained, leaf blade L and W were measured with a ruler graduated in centimeters. It should be emphasized that maximum L was considered as the greatest distance between the petiole insertion point in the leaf limbus and the apex end of the leaf and maximum W as the widest dimension perpendicular to the axis of the main vein (Figure 1).

All leaves were numbered and scanned with a flatbed scanner (HP Deskjet 3516 All-in-one; HP Inc., Palo Alto, California, USA). The generated images made it possible to determine LA with ImageJ software (National Institutes of Health [NIH], Bethesda, Maryland, USA; available free at http://rsbweb.nih.gov/ij/) in which the images were calibrated for later measurement. This program detects the leaves and provides the area (cm<sup>2</sup>) and has shown quite satisfactory results compared with the LA integrator (standard method) (Martin et al., 2013).

Figure 1. Measurements of different morphological leaf blade patterns of Ormosia paraensis 100 days after sowing.



#### Data analysis

Descriptive statistics were initially used to verify the behavior of the values obtained for each analyzed variable: LA, L, and W. Data were organized and all the dimensional relationships analyzed by a correlation matrix; the Kolmogorov-Smirnov test ( $\alpha = 0.05$ ) was simultaneously applied to verify normality of data.

Ten theoretical regression models were tested: three linear, one cubic, three quadratic, and three potency regressions; the linear parameters of the leaf blade surface of L and W taken separately or combined  $(L \times W)$  were independent variables (Table 1).

Only 64% of the database was considered for model adjustment and the remaining 36% was used to validate the best two selected equations. For validation, the non-parametric chi-square test ( $\chi^2$ ) was applied at a 99% confidence level. The ordinary least squares (OLS) method was adopted to estimate the model parameters and the t-test was used to test their significance in which coefficients were rejected when p > 0.05. All adjusted statistical procedures were performed with the R software (R Core Team, 2014).

To obtain LA estimates of *O. paraensis*, 100 networks of the multilayer perceptron (MLP) were implemented with a procedure consisting of dividing data into calibration groups (64% of samples), and validation (36%) was performed by the random sampling method, adopting leaf blade L and W as input data and LA as the output variable. Supervised networks trained with the backpropagation algorithm were implemented.

Minimum mean square error (0.0001) was used as a stopping criterion and the number of cycles was set at 3000. Numerical variables were linearly normalized in the range between 0 and 1 (Braga et al., 2007). The activation function

Models	Description	Туре		
(1)	$LA = \beta_0 + \beta_1(L \times W) + \beta_2 \sqrt{L} + \beta_3 W^2 + \varepsilon i$	Quadratic		
(2)	$LA = \beta_0 + \beta_1(L^2 \times W) + \beta_2(W^3 \times L) + \varepsilon i$	Cubic		
(3)	$LA = \beta_0 + \beta_1(\ln(L) \times W) + \beta_2(\ln(W^2)) + \varepsilon i$	Quadratic		
(4)	$LA = \beta_0 + \beta_1 W^2 + \beta_2 L^2 + \beta_3 (L \times W) + \varepsilon i$	Quadratic		
(5)	$LA = \beta_0 + \beta_1 L + \beta_2 W + \varepsilon i$	Linear simple		
(6)	$LA = \beta_1 L + \epsilon i$	Linear modified		
(7)	$LA = \beta_1 W + \epsilon i$	Linear modified		
(8)	$LA = \beta_0(L)^{\beta_1} + \epsilon i$	Potency		
(9)	$LA = \beta_0(W)^{\beta_1} + \epsilon i$	Potency		
(10)	$LA = \beta_0 (L \times W)^{\beta_1} + \varepsilon i$	Potency		

Table	1.9	Stati	stical	l models	tested	for	leaf	area	estimation	on of	1	)rmosia	naraensis
Table	T • r	Juan	Suca	imoucis	icsicu.	IUI	icai	arca	csumatio	UII UI	v	r mosiu	puruensis

LA: Leaf area (cm<sup>2</sup>); L: length (cm); W: width (cm);  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ : model coefficients; ln: Napierian logarithm;  $\epsilon_i$ : random error.

was the hyperbolic tangent for both the intermediate and exit neurons. The neural networks were trained with the NeuroForest 3.3 software (http://www.dapflorestal.com.br).

The choice of the best equations was based on the following statistical criteria: adjusted coefficient of determination  $(R^2aj\%)$ , relative standard error of estimate (SEE%), F value, and Akaike's information criterion (AIC). A ranking was prepared to simplify the choice of the best regression equations. The best ordered equations were analyzed for their graphical distribution of the residues to verify the occurrence of trends in the estimates.

The best networks were selected based on multiple correlations between observed and estimated LA ( $Ry\hat{y}$ ) and the following statistics: bias, root-mean-square error (RMSE), mean absolute deviation (MAD), and residual graphical analysis according to the procedure used by Lafetá et al. (2012) and Binoti et al. (2013). The selected neural networks were applied to 36% of the independent LA data in the validation process (generalization).

An analysis of mean comparison was also performed to test the null hypothesis (H0); there was no significant difference between the LA values estimated by both the best adjusted equation and the best trained neural network compared with the real measured data. Student's t-test was used for paired data (dependent samples) to evaluate this hypothesis at a significance level of  $\alpha = 0.05$  (Lafetá et al., 2012).

## RESULTS

The correlation coefficients for the response variable (LA) were slightly higher with the explanatory variables of leaf blade L and W (Figure 2). On the other hand, a satisfactory correlation was not observed between the other biometric variables measured from *O. paraensis* seedlings. Therefore, precise estimates of LA can be obtained from the functional relationship between L and W because the degree of association between these variables was the highest.

All the analyzed variables exhibited normal behavior according to the Kolmogorov Smirnov (KS) test (LA: KS = 0.0832, p = 0.2890; L: KS = 0.0044, p = 0.1477; W: KS = 0.0753, p = 0.4054), which justified the validity of all subsequent analyses. This behavior is shown by the trend line of the histogram in Figure 2.



Figure 2. Graphical representation of the correlation matrix between morphologic variables of *Ormosia paraensis* seedlings at 100 days after sowing.

BL: Length of the leaf blade (cm); BW: width of the leaf blade (cm); SD: stem diameter (cm); RL: root length (cm); Epic: epicotyl (cm).

The amplitude variation of the sample set for the dependent variable ranged from 6.89 cm<sup>2</sup> (minimum) to 42.12 cm<sup>2</sup> (maximum), with LA mean dimensions of 21.07 cm<sup>2</sup>. Regarding the L and W variables, there was little variation in their morphometric dimensions, as indicated by the low coefficient of variation and standard deviation values (Table 2).

The significance of all the adjusted models was p > 0.05 (Table 3). The linear equations resulting from models (4) and (1) showed the best results, even though they did not report the best adjusted coefficient of determination (R<sup>2</sup>aj). Given the adjustment parameters and statistical analysis of deviations, it was possible to verify low significant differences between the equations of the simple linear model (5) and quadratic models (1 and 4) since both equations showed significant adjusted coefficients of determination (R<sup>2</sup>aj) and low standard errors of estimate (SEE%) (Table 3).

Based on an analysis of the scatter plot of points around the regression line (Figure 3), there was little difference in the residual point distributions between models (4) and (1) in which residual points were homogeneously distributed and unbiased. The percentage residual dispersion for both adjusted equations also followed normality and homogeneity trends, as demonstrated by the histogram I Figure 3; this confirms that the adopted statistical analyses to evaluate the obtained mathematical models were satisfactory to validate these models, considering the studied biological parameter.

This showed that selected equations accurately estimated LA data of the species under study. To verify the efficiency of these equations, the data validation process of 36% of the variables of interest was presented; both equations (4) and (1) obtained chi-square test values below the critical value, namely model (1) -  $\chi^2_{cal} = 28.26 < \chi^2_{tab} = 74.91$ ,  $\alpha = 0.01$  and model (4) -  $\chi^2_{cal} = 27.74 < \chi^2_{tab} = 74.91$ ,  $\alpha = 0.01$ .

Based on the training parameters, three ANNs (10, 27, and 76) were selected because they showed the best accuracy results adjusted to the data for the variable of interest (Table 4). According to these parameters, ANNs can be more accurate than traditional regression models, especially for their excellent correlation values (Ryŷ); this means that the output neuron had satisfactorily produced the best estimate for the LA values, while statistical models had moderate accuracy. There was a low percentage of error when generating the estimated ANNs in accordance with the RMSE result (Table 4), thus indicating greater reliability to generate estimates.

	Dependent variable	Independent variables			
Values	Leaf area (LA)	Length (L)	Width (W)		
	cm <sup>2</sup>	cm	cm		
Minimum	6.8910	3.5001	2.4002		
Maximum	42.1220	10.3002	6.4001		
Mean	21.0701	6.5024	4.1292		
Standard deviation	± 5.9026	± 1.1558	$\pm 0.6115$		
Coefficient of variation (%)	28.0141	17.7750	14.8092		

Table 2. Descriptive statistics for the dataset of the different morphological variables: leaf area (LA), length (L), and width (W) of *Ormosia paraensis* at 100 days after sowing.

Table 3. Estimates of the	e parameters and	precision	statistics for	the adjusted	models	considering t	he 64%	dataset o	of
Ormosia paraensis.									

Models		Coef	ficients		Precision statistics					
	ßo	$\beta_1$	ß <sub>2</sub>	ß3	R²aj%	SEE%	F	AIC	Ranking	
(4)	3.727*	-0.631*	-0.164*	1.324*	87.39	12.07	102.53	140.85	1 <sup>st</sup>	
(1)	$17.171^{*}$	$1.069^{*}$	-6.741*	-0.439*	87.35	12.59	102.33	140.99	$2^{nd}$	
(5)	-13.860*	3.422*	3.163*	_	86.22	12.91	143.60	144.44	3 <sup>rd</sup>	
(10)	$1.256^{*}$	$0.859^{*}$	_	_	87.01	13.05	279.29	145.48	4 <sup>th</sup>	
(2)	$11.198^{*}$	$0.053^{*}$	0.007 <sup>ns</sup>	_	85.41	13.12	137.52	147.41	5 <sup>th</sup>	
(3)	16.368*	$5.005^{*}$	-11.901*	_	82.53	13.87	118.52	157.39	6 <sup>th</sup>	
(8)	$2.210^{*}$	$1.216^{*}$	_	_	87.12	14.92	134.07	169.58	7 <sup>th</sup>	
(6)	_	3.348*	_	_	87.78	15.41	3962	174.38	8 <sup>th</sup>	
(7)	_	5.165*	_	_	88.80	21.18	2055	231.62	9 <sup>th</sup>	
(9)	$4.404^{*}$	$1.109^{*}$	-	-	80.05	21.24	46.066	233.11	$10^{\text{th}}$	

R<sup>2</sup>aj%: Adjusted coefficient of determination; SEE%: relative standard error of estimate; F: value of ANOVA; AIC: Akaike information criterion. \*Linear and angular coefficient differs from zero by t-test at 5% error probability.

nsNonsignificant.



Figure 3. Dispersion and frequency histogram of percentage errors referring to the best two adjusted equations to estimate the leaf area of *Ormosia paraensis* at 100 days after sowing.

Table 4. Accuracy of artificial neural networks (ANNs) constructed to estimate the leaf area of *Ormosia paraensis* at 100 days after sowing.

ANN	Architecture*	Phase	Ryŷ	Bias	RMSE%	MAD
10	MLP 2-5-1	Training	0.9521	-0.5303	10.5124	2.1643
	MLP 2-5-1	Validation	0.8825	-2.2747	12.6798	3.4758
27	MLP 2-6-1	Training	0.9395	-1.6101	11.6772	2.6348
	MLP 2-6-1	Validation	0.8711	-2.6218	14.5399	3.8967
76	MLP 2-3-1	Training	0.9316	-2.4482	13.6772	2.9408
	MLP 2-3-1	Validation	0.8522	-3.6478	15.2616	4.3218

Ryŷ: Multiple correlation coefficient; RMSE: root-mean-square error; MAD: mean absolute deviation.

\*Distribution of neurons in the input, hidden, and output layer, respectively.

# DISCUSSION

Seedling growth showed little morphological variation, even if seedlings came from seed lots stored at different temperatures and with different packaging. This was illustrated by the descriptive statistical values of the analyzed morphometric data (Table 2). These results corroborate Silva et al. (2015), who described the morphoanatomical characteristics of tento in both seedlings and fruits and confirmed low variations in the morphological patterns at different plant phases.

The choice of the best two equations was based on the low values of relative standard error of estimate (SEE%) and the Akaike information criterion (AIC) (Table 3), which in turn are essential parameters to diagnose estimation accuracy reliability of the adjustment quality of the equation. Therefore, the lower the values of these two statistics, the more accurate and reliable the adjusted equation is (Schneider et al., 2009).

According to Rocha et al. (2015), AIC admits the existence of a real model that describes unknown data, which provides a relative measure of the information lost when a certain model is used to describe reality. It can be inferred that when the sum of the error squared increases in the regression variance analysis, AIC also increases; this justifies not selecting models (6) and (7) (Table 3), which both presented excellent values of R<sup>2</sup>aj and F value of ANOVA but also high AIC.

The explanation for this decharacterization of the model adjustment (6 and 7) is probably due to the withdrawal of the regression intercept, which revealed heteroscedastic behavior; this means that it violates the statistical principles of estimates in which the mean values predicted by models are not equal to the true parameter of the response variable.

Similarly, Santos (2016) evaluated the application of nine models to estimate the LA of 14 native tree species; the author observed that three adjusted equations without the intercept showed biased behavior, even though they had excellent R<sup>2</sup>aj and SEE% values. It is emphasized that introducing the intercept constant ( $\beta_0$ ) does not significantly affect estimated LA as demonstrated in other studies (Kandiann et al., 2009; Sousa et al., 2015). It is believed that such discrepancies in the results between studies may be associated with the sample size and morphological pattern between leaves that vary from species to species as well as leaf phytosanitary quality.

Equations (1) and (4) consistently generated reasonably accurate statistics and both equations were able to explain  $R^2aj = 77.35\%$  and 77.39%, respectively, in the variation of seedling LA data, considering the separate insertion of independent variables (L and W) and their product. The  $R^2aj$  value obtained in the present study is inferior to the one mentioned by Nicoletti et al. (2015); these authors found  $R^2aj = 91.62\%$  when modeling LA of *Eucalyptus dunnii* seedlings; however, they obtained SEE = 15.2\%, which is much higher than the values found by models (1) and (4) in the present study. The authors in the abovementioned study consider other seedling morphological variables when modeling the LA of *E. dunnii*, such as stem diameter. Thus, it is understandable that modeling variables that are difficult to measure, such as seedling LA, it is fundamental that researchers consider any variable that is easily measured in the analysis and can be strongly correlated with this variable of interest. This is because simple models that only consider one or two phenotypic characteristics of the seedlings can generate error sensitive estimates and compromise forest production decision-making.

For this reason, one must consider the conditions in which equations have been formulated to be used as a given adjusted equation for the seedling dataset of a given arboreal species (native or planted); this means that their application is recommended for seedlings of the same species, age, and similar climatic characteristics. Otherwise, the estimation of a certain variable of interest can be biased and have low precision.

This result suggests that the generated equations can estimate the LA of *O. paraensis* under the given conditions and that equation (4) is the most recommended because it has the lowest value according to the chi-square test as well as the best precision statistics.

In summary, results demonstrate that the multiple linear equations generated by the product of leaf blade L and W as explanatory variables had the best results compared with the potency models. This finding is corroborated by other studies that have modeled the LA of forest and agronomic species (Antunes et al., 2008; Bianco et al., 2008; Lima et al., 2012; Mota et al., 2014; Carvalho et al., 2017).

Three selected networks exhibited high multiple correlation coefficient  $(Ry\hat{y})$  values for both the training phase and validation; this can infer that the network structure of the MLP type was able to satisfactorily extract the information of the input data, correct the weight assignment errors between the layered connections and generate accurate estimates, as demonstrated by the low values of bias, RMSE, and MAD (Table 4). The RMSE values obtained in the present study were reasonably better than those found by Silva et al. (2008), who trained the network structures (MLP and LMS) to estimate the LA of *Anthurium andreanum* species with RMSE values of 19.09% and 16.57%, respectively.

In this way, the learning of the three networks was proven, and it is suggested that the adopted algorithm exerted a better resolution of estimated LA in the studied species because it intelligently interpreted the application of functionalities between the linear dimensions of independent variables (L and W) with the leaf dependent variable; it created a function that significantly generated smaller prediction errors with high processing capacity. The backpropagation algorithm is also recommended for problem solving of non-linear functions, as reported by Haykin (2001).

Results also demonstrate that the use of ANNs is perfectly feasible and can be an alternative modeling tool to estimate seedling LA s in native forest species. The adjustments obtained by ANNs (10 and 27) were the best, and both had the same distribution of observed LA values, which means that there were no distortions in the estimates. This was observed by the uniform distribution of the concentrated residuals around the error line. In addition, the normality trend of the residuals plotted in the histogram also supported this (Figure 4).

Residual analysis showed that the performance of ANNs was superior to the adopted regression models. However, the use of ANNs during the training phase is noteworthy because they show some limitations. This refers to the quality of network behavior (training time and generalization capacity) in which it can act in the scheduling process (Braga et al., 2007). This means that as the dimension and complexity of task completion increases, the network can memorize the training patterns rather than extract characteristics for generalization.

Indeed, because networks work by the interconnection structure of artificial neurons between layers, the operator must pay attention to the architecture to which the ANNs and database to be used are being configured; they work on the basis of storing information, thus resulting in knowledge learning. Therefore, sensitivity to the loss or failure of a connection may occur if the network is misconfigured or fed with too many variables and improperly collected data.

Corroborating this statement, Binoti et al. (2013) emphasize that the same thing occurs with regression models. The inclusion of many variables in neural networks results in greater modeling complexity because the network will automate the representation of more than one information cell as the basis for the weights in a given learning algorithm to generate a



Figure 4. Dispersion and frequency histogram of percentage errors referring to the best two adjusted networks to estimate the leaf area of *Ormosia paraensis* at 100 days after sowing.

ANN: Artificial neural networks.

response to the proposed problem. Depending on the structure and activation function, it can execute the initially required task with either high or low precision.

The advantage of modeling when it is difficult to measure variables (such as the LA of native arboreal species by ANNs) is simply the flexibility to insert the categorical variable networks, giving the adjustment a greater probability of accuracy in the estimation, which becomes a complex task in traditional regression models. By considering the importance of the LA responsible for the very active photosynthetic performance in the seedling that directly results in its development, it is essential to carry out new research that applies this methodological approach. Moreover, new variables that are categorical or continuous and associated with the production of seedlings in nurseries should be used, so that ANN modeling achieves the highest possible efficiency and satisfactorily portrays the reality of biological variables.

To clarify this statement, Lafetá et al. (2012) studied the application of an MLP type network to estimate specific LA and chlorophyll content in eucalyptus under different spacing: they concluded that the MLP networks were able to learn and generalize the assimilated knowledge for all the measured plants. Therefore, the result of this study corroborates current research, demonstrating that ANNs are an alternative and significantly useful tool to apply to forest seedlings at different ages, with their proven efficiency in other study approaches (Leite et al., 2011; Binoti et al., 2012; Cosenza et al., 2017).

Regarding the paired Student's t-test to compare the two methodologies used in modeling the LA of *O. parensis*, the result showed that there was no difference in the values estimated by the best adjusted regression model, Equation (4), compared with the actual data ( $p = 0.0941 > \alpha = 0.05$ ). This behavior was also observed in the estimates using the best network (ANN 10) compared with the actual data from the control ( $p = 0.2009 > \alpha = 0.05$ ) as well as between the data estimated in both methodologies ( $p = 0.5604 > \alpha = 0.05$ ). Therefore, the null hypothesis in which both analysis techniques are efficient to estimate the studied variable of interest is accepted, highlighting the ANNs that showed better adjustment results.

A similar result was found by Leal et al. (2015), who verified that the performance of ANNs was superior to regression equations in the modeling of dendrometric variables of forest species. On the other hand, Silva et al. (2008) evaluated the efficiency of LA estimation in *Anthurium andreanum* by neural networks and regression models, and concluded that the regression equations obtained by the product of the linear dimensions of the plant leaves exhibited more accurate results compared with ANNs. These studies demonstrate that the two techniques have great potential for modeling both difficult to measure variables and easy to obtain variables and with a relatively low cost for the companies, suggesting that their use be adopted for seedlings cultivated under both field and greenhouse conditions.

# CONCLUSIONS

It is possible to obtain accurate leaf area (LA) estimates of *Ormosia paraensis* using regression models and artificial neural networks (ANNs). For modeling using the traditional regression technique, equations more accurately estimating the variable of interest were generated by models with more than one explanatory variable and with the product of the leaf linear dimensions (length × width).

Results indicated that the performance of the ANNs was superior to the conventional regression technique and can be safely indicated to estimate the LA of other native forest species. Thus, it is recommended to increase the knowledge of this technique to predict foliar areas of agronomic and forest species, considering the advantages of flexibility of simultaneously inserting more explanatory variables into the estimation process.

### REFERENCES

Antunes, W.C., Pompelli, M.F., Carretero, D.M., and Damatta, F.M. 2008. Allometric models for non-destructive leaf area estimation in coffee (*Coffea arabica* and *Coffea canephora*). Annals of Applied Biology 153(1):33-40.

Bianco, S., Carvalho, L.B., e Bianco, M.S. 2008. Estimativa da área foliar de Sida cordifolia e Sida rhombifolia usando dimensões lineares do limbo foliar. Planta Daninha 26(4):807-813.

Binoti, M.L.M.S., Binoti, D.H.B., e Leite, H.G. 2013. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. Revista Árvore 37(4):639-645.

Binoti, D.H.B., Binoti, M.L.M.S., Leite, H.G., e Silva, A. 2012. Redução dos custos em inventário de povoamentos equiâneos utilizando redes neurais artificiais. Agrária 8(1):125-129.

Braga, A.P., Carvalho, A.P.L.F., e Ludermir, T.B. 2007. Redes neurais artificiais: teoria e aplicações. 2ª ed. 248 p. Editora LTC, Livros Técnicos e Científicos, Rio de Janeiro, Brasil.

Cardozo, N.P., Parreira, M.C., Panosso, A.R., e Volpe, C.A. 2014. Modelagem da área foliar de duas cultivares de amendoim em função das dimensões lineares dos folíolos. Bioscience Journal 30(1):101-107.

- Carvalho, J.O., Toebe, M., Tartaglia, F.L., Bandeira, C.T., and Tambara, A.L. 2017. Leaf area estimation from linear measurements in different ages of *Crotalaria juncea* plants. Anais da Academia Brasileira de Ciências 89(3):1851-1868.
- Cosenza, D.N., Soares, A.A.V., Alcântara, A.E.M., Silva, A.A.L., Rode, R., Soares, V.P., et al. 2017. Site classification for eucalypt stands using artificial neural network based on environmental and management features. Cerne 23(3):310-320.
- Gorgens, E.B., Leite, H.G., Santos, H.N., e Gleriani, J.M. 2009. Estimação do volume de árvores utilizando redes neurais artificiais. Revista Árvore 33(6):1141-1147.
- Haykin, S. 2001. Redes neurais: princípios e prática. 900 p. Bookman, Porto Alegre, Rio Grande do Sul, Brasil.
- Kandiann, K., Parthasarathy, U., Krishnamurthy, K.S., Thankamani, C.K., and Srinivasan, V. 2009. Modeling individual leaf area of ginger (*Zingiber officinale* Roscoe) using leaf length and width. Scientia Horticulturae 184(6):532-537.
- Lafetá, B.O., Santana, R.C., Nogueira, G.S., Penido, T.M.A., e Couto, L. 2012. Área foliar específica e clorofila em eucalipto estimadas por redes neurais artificiais. Biomassa & Energia 5(1):67-75.
- Leal, F.A., Miguel, E.P., Matricardi, E.A.T., e Pereira, R.S. 2015. Redes neurais artificiais na estimativa de volume em um plantio de eucalipto em função de fotografias hemisféricas e número de árvores. Revista Brasileira de Biometria 33(2):233-249.
- Leite, H.G., Silva, M.L.M., Binoti, D.H.B., Fardin, L., and Takizawa, F.H. 2011. Estimation of inside-bark diameter and heartwood diameter for *Tectonagrandis* Linn. trees using artificial neural networks. European Journal of Forest Research 130(2):263-269.
- Lima, R.T., Souza, P.J.O.P., Rodrigues, J.C., e Lima, M.J.A. 2012. Modelos para estimativa da área foliar da mangueira utilizando medidas lineares. Revista Brasileira de Fruticultura 34(4):974-980.
- Malagi, G., Citadin, I., Scariot, S., e Reis, L. 2010. Método não destrutivo para determinação da área foliar da videira, cultivar BRS-Violeta. Revista Brasileira de Fruticultura 32(4):1250-1254.
- Martin, T.N., Marchese, J.A., Sousa, A.K.F., Curti, G.L., Fogolari, H., e Cunha, V.S. 2013. Uso do software *ImageJ* na estimativa de área foliar para a cultura do feijão. Interciência 38(12):843-848.
- Mota, C.S., Leite, H.G., e Cano, M.A.O. 2014. Equações para estimar área foliar de folíolos de *Acrocomia aculeta*. Pesquisa Florestal Brasileira 34(79):217-224.
- Nicoletti, M.F., Navroski, M.C., e Júnior, N.S.N. 2015. Modelagem de variáveis morfológicas em mudas de *Eucalyptus dunnii*. Floresta 45(4):809-818.
- R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/.
- Rocha, M.B., Barreto-Garcia, P.A.B., Prado, W.B., de Paula, A., e Júnior, V.C. 2015. Volumetria de *Genipa americana* em plantio homogêneo no Sudoeste da Bahia. Pesquisa Florestal Brasileira 35(84):419-425.
- Santos, J.N.B. 2016. Equações alométricas para estimativa da área foliar de espécies lenhosas de floresta ombrófila de terra baixas. 102 p. MSc Dissertation. Universidade Federal Rural de Pernambuco, Recife, Pernambuco, Brasil.
- Schneider, P.R., Schneider, P.S.P., e Souza, C.A.M. 2009. Análise de regressão aplicada à Engenharia Florestal. 2a ed. 294 p. FACOS Universidade Federal de Santa Maria, Santa Maria, Río Grande del Sur, Brazil.
- Silva, S.H.M.G., Lima, J.D., Bendini, H.D.N., Nomura, E.S., e Moraes, W.D.S. 2008. Estimativa da área foliar do antúrio com o uso de funções de regressão. Ciência Rural 38(7):243-246.
- Silva, B.M.S., Silva, C.O., Môro, F.V., and Vieira, R.D. 2015. Morphoanatomy of fruit, seed and seedling of Ormosia paraensis Ducke. Journal of Seed Science 37(4):192-198.
- Sousa, L.F., Santos, J.G.D., Alexandrino, E., Maurício, R.M., Martins, A.D., e Sousa, J.T.L. 2015. Método prático e eficiente para estimar a área foliar de gramíneas forrageiras tropicais. Revista Archivos de Zootenia 64(245):83-85.
- Souza, M.C., and Habermann, G. 2014. Non-destructive equations to estimate the leaf area of *Styrax pohlli* and *Styrax ferrugineus*. Brazilian Journal of Biology 74(1):222-225.
- Zanetti, S., Pereira, L.F.M., Sartori, M.M.P., and Silva, M.A. 2017. Leaf area estimation of cassava from linear dimensions. Anais da Academia Brasileira de Ciências 89(3):1729-1736.