THICKNESS ACCURACY OF TEAK BARK BY ARTIFICIAL INTELLIGENCE

Diogo Guido Streck Vendruscolo1, Clebson Lima Cerqueira2, Samuel de Pádua Chaves and Carvalho3, Reginaldo Antonio Medeiros4, Raiane Scandiano da Silva5

1Forest engineer, Alto Boa Vista, Mato Grosso, Brazil – diogoguido@hotmail.com (*AUTHOR FOR CORRESPONDENCE)
2Federal University of Paraná, Postgraduate Program in Forest Engineering, Curitiba, Paraná, Brazil – clebsolima10@hotmail.com
3Federal University of Mato Grosso, Postgraduate Program of Forestry and Environmental Studies, Cuiabá, MT, Brazil – sam.padua@gmail.com
4Federal Institute of Mato Grosso campus Prof. Olegário Baldo, Department of Forest Engineering, Cáceres, MT, Brazil – reginaldo.medeiros@cas.ifmt.edu.br
5Forest engineer, Alto Boa Vista, Mato Grosso, Brazil – raiane.scandian@gmail.com

Received for publication: 24/04/2018 – Accepted for publication: 27/07/2018

INTRODUCTION

Estimates of tree bark thickness are fundamental for forest management, however, the degree of precision is conditioned to the adoption of efficient modeling techniques. The objective of this study was to evaluate and propose a model of artificial neural networks to estimate the thickness of the tree bark of Tectona grandis (Teak). The data originated from the measurement of 68 dominant trees, ranging in age from 6 to 33 years. The thickness of the bark was correlated with variables inherent to the tree, being: diameter in the different positions of the stem (di); diameter at 1.3 m height (dbh); total height (ht); relative height (hi_rel); and age (id). The trained networks were of the multilayer perceptron type, and a linear regression model was adjusted as a comparative support. The accuracy of the estimates was evaluated through statistical indicators and graphical analysis. The results showed a strong correlation between bark thickness and tree diameter, as well as relative height, with values above 0.70. Age also exerted a strong influence on the thickness of the bark of the trees.

The artificial intelligence technique has demonstrated the potential for such application and the model proposed with the input variables: diameter, relative height and age was the one that presented the best statistical performance, and thus was the most suitable for predicting the bark in Teak trees.

Keywords: Tectona grandis, regression, artificial neural networks, forest management.


DOI: 10.5380/lf.v49.n3.59106

*Correspondence*
variables are more requested (STÄNGLE et al., 2016), however, the adoption of a simple linear model can result in less accurate estimates, while more elaborate models describing the bark with the inclusion of other readily available variables are more requested (STÄNGLE et al., 2017). Therefore, the inclusion of predictor variables allows the validation of a bark model for a species in a given region (LI; WEISKITTEL, 2011).

For many conifer species, for example, it has been found that tree thickness can be described by variables inherent to the tree, such as bark diameter, total height, age, and measurement position (LAASASENAHO et al., 2005; LI; WEISKITTEL, 2011; TEWARI; MARISWAMY, 2013; STÄNGLE et al., 2017). In some cases, there is influence of site index on bark thickness, as reported by Božić et al. (2007) for Abies alba. Additionally, genetic variability may also influence the phenotypic development of the bark (KOHNLE et al., 2012) as well as the geographical position (STÄNGLE et al., 2017). Thus, the use of modeling techniques that include a larger number of variables are promising for the field of research.

In the last decades, the computational advance has allowed the manipulation and quick, accessible and with low cost processing of data. In this field, the use of artificial intelligence is already considered a consolidated technique to relate dendrometric variables. Promising results with the use of this technique were found in order to estimate the height of trees (ÖZÇELIK et al., 2013), volume (BINOTTI et al., 2014; MIGUEL et al., 2016), production prognoses, diameter and height (BINOTTI et al., 2015; VIEIRA et al., 2018), and etching (SANQUETTA et al., 2017; MARTINS et al., 2017). Nevertheless, bark estimates using artificial neural networks have limited attention, evidencing the importance of new research on the subject.

This study is based on the hypothesis that artificial intelligence technique is efficient to estimate the bark thickness of Teak trees. Thus, its objective was to evaluate and propose a model to estimate the bark thickness of Tectona grandis trees through artificial neural networks.

MATERIAL AND METHODS

Study area and data

The data were obtained from commercial and experimental plantations with Teak implanted in the middle north and south central regions of the Mato Grosso State, Brazil. A total of 68 dominant trees were sampled, with seminal genetic material and age ranging from 6 to 33 years. In all trees, the values of the bark thickness were obtained in cm (esp_c) by discs collected at different heights of the stem (hi), commonly used in log analyzes (DRESCHER et al., 2016; CHAVES et al., 2016). At all times, fractionations occurred from a minimum height of 0.1 m; 1.3 m; up to a diameter (di) with bark of approximately 5 cm; and with intervals of no more than 3 m.

In each disc, the diameter with and without bark (di) was measured in perpendicular positions using graduated ruler with millimeter precision. Subsequently, the average of diameters with and without bark was obtained. The bark thickness at each position of the tree stem was obtained by means of the average arithmetic. In this way, the database was composed of bark thickness (esp_c), bark diameter in each section (di), diameter at 1.3 m height (dbh) and total height (ht) of each tree.

Data processing

The data obtained from the measurement of the trees were used to train the artificial neural networks (ANN) in order to estimate the bark thickness of Teak trees. The predictor variables were selected by means of linear correlation analysis between esp_c and di, dap, ht, and hi_rel (relative height). In this stage, the variable hi_rel was used to relativize the heights due to the amplitude of variation of the tree ages and, consequently, the amplitude in the heights in which the discs were collected. Subsequently, the data set was randomly subdivided into 70% for ANN training and 30% used for data validation, as suggested by Binoti et al. (2014). In the selection of the input variables, used to train the networks, it was considered only up to the third most correlated with esp_c. Thus, three groups of architectures were evaluated, the first one (G1) was composed of only one predictor variable, the second (G2) of two and the third (G3) of three variables.

Artificial Neural Networks

The neural networks used were the Multilayer Perceptron type, commonly known as MLP, which has a universal capability to approximate functions (HORNIK et al., 1989). The training was oriented to minimize the sum of squares of errors through the backpropagation algorithm (error backpropagation), with a learning rate of 0.01 and the training stop criterion was the minimization of the mean error (0.01) or the maximum of 3,000 times.

Thus, different network architectures were trained with only one hidden layer, then, a hyperbolic tangent transfer function (Eq. 1) was applied in the hidden layer and linear function (Eq. 2) in the output one, as suggested by Vieira et al. (2018). The number of neurons in the input and hidden layers varied from 1 to 3, totaling nine tested architectures (3 groups x 3 architectures), with twenty networks being trained for each one. Training and
validation were performed in R programming language (R CORE TEAM, 2016) through the neuralnet package (FRITSCH; GUENTHER, 2016).

\[
f(x) = \frac{2}{1 + e^{-2x}} - 1
\]
\[
f(x) = x
\]

in which: \( f(x) \) the activation function and \( x \) is the net input of the artificial neuron.

Regression

A simple linear regression model (Eq. 3), commonly used to estimate the bark of Teak trees (TEWARI; MARISWAMY, 2013; NOVAES et al., 2017) was used as the comparative support method.

\[
\ln(esp_{cij}) = \beta_0 + \beta_1 \ln(d_{ij}) + \varepsilon_{ij}
\]

in which: \( \ln \) is the neperian, \( esp_{cij} \) is the bark thickness of the tree "i" at stem position "j" (cm), are the coefficients to be obtained, \( d_{ij} \) is the diameter of the tree "i" at stem position "j", and \( \varepsilon_{ij} \) is a random error of the tree "i" at stem position "j".

The coefficients of the model were obtained by the least squares method. The assumption of normality of errors was verified by the Kolmogorov-Smirnov test, the White's test was used to the homogeneity of variance, and the independence of the residues was tested by the Durbin-Watson method. A significance level of 5% was adopted for both tests.

Statistical evaluation of the proposed methods

In order to evaluate the tree bark estimates obtained by ANN and regression, the Pearson linear correlation coefficient was applied among the observed and estimated values - \( r_{yy} \) (Eq. 4), square root of mean error in percentage - SRME\% (Eq. 5) and graphic analysis.

\[
r_{yy} = \frac{\text{cov}(y, \hat{y})}{\sqrt{S^2(y)S^2(\hat{y})}}
\]

\[
\text{SRME}\% = 100 \frac{\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\bar{y}} \frac{1}{n}
\]

in which: \( \text{cov} \) is covariance, \( y_i \) is the observed variable, \( \hat{y}_i \) is the estimated variable, \( S \) is the variance, \( \bar{y} \) is the mean of the observed variable and \( n \) is the number of observations.

RESULTS

Selection of predictor variables

Among the variables measured, the three that presented the highest Pearson correlation (r) were the diameter (\( d_i \)), relative height (\( h_{i,rel} \)) and age (\( id_i \)), respectively (Figure 1). Also, it has been found that the variables \( d_i \) and \( id_i \) correlate positively with \( esp_c \), while \( h_{i,rel} \) has a negative correlation. Therefore, the training of neural networks was performed using only these variables.

Figura 1. Correlação da espessura de casca (\( esp_c \)) de Tectona grandis com as variáveis diâmetro (\( d_i \)); diâmetro a 1,3 m de altura (\( dap \)); altura total (\( ht \)); altura relativa (\( h_{i,rel} \)) e idade (\( id \)).

Figure 1. Correlation of the shell thickness (\( esp_c \)) of Tectona grandis and the variables diameter (\( d_i \)); diameter at 1.3 m of height (\( dap \)); total height (\( ht \)); relative height (\( h_{i,rel} \)) and age (\( id \)).
1.3 m height (dbh); total height (ht); relative height (hi_rel) and age (id).

Regression modeling

The adjustment of the regression model presented estimates of significant coefficients by the t test (p≤0.05) and resulted in the equation: 

\[ \ln(esp_{cij}) = -2.38397 + 0.62652\ln(di_{ij}) + \varepsilon_{ij}. \]

The accuracy in relation to the square root of the mean error (SRME%) was 21.21% for the adjustment with a small alteration of error in validation (21.61%), being considered values potentially high when dealing with regression analysis. Regarding the correlation between the observed and estimated values (\(r_{yx}\)), similar results were found in both adjustment and validation, with values of 0.853 for adjustment and 0.868 for validation.

In a complementary way, the graphical analyzes (Figure 2) allow to infer about the tendency of overestimation of the bark in smaller diameters. The lines on the histograms presented a larger tail on the left, in addition to the negative residual mean (dashed line) which corroborates with the graphs of Figures 2.A and 2.V. Nonetheless, the curve indicates that the highest residual density is close to zero in adjustment and less than zero (-5.43) in validation.

The normality test of the residues (Kolmogorov-Smirnov) pointed to a normal distribution (p value = 0.201), indicating that the regression technique is robust and consistent to estimate the bark of Teak trees. However, as expected, problems related to the correlation between errors and heterogeneity of residual variance were detected due to the hierarchical structure of the data, with repeated measures in the same tree.

Artificial neural network modeling

Among the networks trained for bark estimation, the accuracy statistics of only one for each group and architecture were evaluated and presented. Based on the statistics, it was observed that, as in regression modeling, errors were considered potentially high in both situations with values higher than 18% and maximum of 21.23% in validation (Table 1). The statistical indicators (\(r_{yx}\) and SRME) did not show large variations between the smaller and higher accuracy network configurations.

Accuracy statistics revealed that the inclusion of predictor variables, relative height (hi_rel) and age (id), provided better performance in network estimates. Moreover, the same behavior was verified in applying the networks in 30% of the data used in validation.

When evaluating the stability of the neural networks in relation to the SRME and \(r_{yx}\) statistics within each group, there was a low error variation, being less than 1% for training and a maximum of 1.08% in validation (Figure 3). The group most sensitive to variation of number of neurons in the hidden layer was G2, which showed an expressive trend of improvements in statistical indicators as the number of neurons increased. G1 networks also
exhibited this trend, but less significantly. Yet, the results for the networks of group 3 show very similar statistics between the different numbers of neurons used in the hidden layer.

Tabela 1. Estatísticas utilizadas de acurácia das redes neurais artificiais para estimar a espessura da casca de Tectonagrandis.

Table 1. Used artificial neural network accuracy statistics to estimate the thickness of the bark of Tectonagrandis.

<table>
<thead>
<tr>
<th>ANN</th>
<th>Group</th>
<th>Architecture</th>
<th>Training</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IL</td>
<td>HL</td>
<td>OL</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>$di$</td>
<td>1</td>
<td>$esp_c$</td>
</tr>
<tr>
<td>2</td>
<td>G1</td>
<td>$di$</td>
<td>2</td>
<td>$esp_c$</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>$di$</td>
<td>3</td>
<td>$esp_c$</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>$di; h_{rel}$</td>
<td>1</td>
<td>$esp_c$</td>
</tr>
<tr>
<td>5</td>
<td>G2</td>
<td>$di; h_{rel}$</td>
<td>2</td>
<td>$esp_c$</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>$di; h_{rel}$</td>
<td>3</td>
<td>$esp_c$</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>$di; h_{rel}; id$</td>
<td>1</td>
<td>$esp_c$</td>
</tr>
<tr>
<td>8</td>
<td>G3</td>
<td>$di; h_{rel}; id$</td>
<td>2</td>
<td>$esp_c$</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>$di; h_{rel}; id$</td>
<td>3</td>
<td>$esp_c$</td>
</tr>
</tbody>
</table>

ANN = artificial neural network; IL = input layer; HL = hidden layer; OL = output layer; $di =$ diameter (cm); $h_{rel} =$ relative height (%); $id =$ age (years); $esp_c =$ bark thickness (cm); $r_{yy} =$ correlation coefficient; SRME = square root of the mean error.

Figura 3. Estabilidade estatística das redes neurais artificiais para estimar a casca de Tectona grandis nos diferentes grupos testados (G1, G2 e G3).

Figure 3. Statistical stability of the artificial neural networks to estimate the thickness of the Tectona grandis bark in the different groups tested (G1, G2 and G3).

The best result for group 1 (G1) networks was obtained for ANN 3 with three neurons in the hidden layer. In addition, the same situation was observed in G2. In G3, the three networks presented good performance with little advantage for architecture with two neurons in the middle layer. The architecture and the respective synaptic weights and bias of the best network of each group is shown in Figure 4.
Based on the graphical analysis of residues, represented by Figure 5, it can be seen that both networks had similar behaviors, with greater dispersion in the smaller diameters. The residual mean (dashed vertical line) indicates a slight tendency to overestimate, since in all cases the values were negative. In validation, the behavior was similar, with relevance for G2 network, which presented the trend of a more symmetrical distribution (Figure 4).
Figure 5. Residue, observed values vs. and histogram for the artificial neural networks of the three groups (G1, G2 and G3) trained to estimate the shell of *Tectona grandis*. T = training; V = validation.
Considering the low differences between statistical indicators verified by the graphical analysis, it is possible to infer that all the suggested network structures can be used to estimate the bark thickness of Teak trees. Comparing group 1 (G1) network with the same input variable (di) used in the regression model, the ANN technique provided a slight reduction of error from 21.21 (regression) to 20.86 (ANN) and an increased correlation from 0.853 (regression) to 0.855 (ANN), and the same behavior was observed in the validation set.

**DISCUSSION**

The bark thickness showed an increasing linear pattern and strong correlation with the diameter (di) of the Teak trees’ log (Figure 1), similar to that observed by Cordero and Kanninen (2003) for the same species in different regions of Costa Rica, Kitikidou *et al.* (2014) for *Pinus halepensis* in northern Greece and Stängle *et al.* (2017) for *Picea abies* in southwestern Germany. Thus, the variable di is widely employed as an independent model for the relationship between the diameter with and without bark at any height of the stem. In this way, the bark thickness can be easily calculated by subtraction once the estimates are obtained (CORDERO; KANNINEN, 2003; LI; WEISKITTEL, 2011; STÄNGLE *et al.*, 2015, 2017).

Additionally, the relative height (hi_rel) of the trees were also shown as an important variable to explain the bark thickness (Figure 1). The regressive variable (hi_rel) also was employed by Laasasenaho *et al.* (2005) in order to shape the thickness of the *Picea abies* Teak. Yet, the authors corroborate with the tendency of tree Teak tapering from the base to the top of the stem. It was opted for the relativization of variables due to the need to standardize the information, once there is a distinct number of samples in trees of different heights. In contrast, it requires additional effort in the field, since it is necessary to compute the relative positions for each tree. It is important to note that age also had a significant effect on the bark thickness, therefore corroborating the studies of Cordero and Kanninen (2003); Laasasenaho *et al.* (2005) and Stängle *et al.* (2017).

In linear regression modeling, the results indicated moderate accuracy because the errors were potentially high. Tewari and Mariswamy (2013) pointed out that this functional relationship may result in adjustments with a lower degree of explanation between the independent and dependent variable. This may occur in virtue of factors, such as geospatial variation, site productivity, and different individual growth rates, that may influence the bark thickness (STÄNGLE *et al.*, 2017). However, although a linear model is very simplistic to explain bark thickness, it can be suggested in cases where low-complexity techniques are desired and does not imply an increase in the number of variables during the forest inventory.

Moreover, artificial intelligence modeling was shown to be slightly superior to regression, even using only the diameter (di) as a predictor variable. The results were satisfactory with simple configurations, demonstrating that networks of complex structures are not necessary to estimate tree bark thickness. Binoti *et al.* (2014) corroborate these findings, the authors evaluated several different ANN configurations in order to estimate the volume of trees and recommended the use of simple architectures. They also pointed out that networks with only one neuron in the hidden layer provide accurate three volume estimatives.

As expected, the increase in the number of input variables resulted in a greater accuracy of the estimates, both in training and in validation of networks. The networks of groups 1 and 2 had a positive answer in regards to accuracy as the number of neurons in the hidden layer increased (Figure 3). Haykin (2001) explains that the definition of an appropriate architecture is not always a simple task and that there are approximate or heuristic solutions that try to estimate these values in relation to the problem to be treated. However, most of the times the number of intermediate neurons in ANNs is defined by the user experience or by previous tests (BINOTTI *et al.*, 2014).

In group 3 (G3), the predictive quality and the statistical stability of the networks were superior to the other groups, pointing out that the sizing of the input layers and numbers of neurons in the hidden and in the output layer was efficient to treat the problem. The gain in accuracy in G3 is related to the inclusion of the variables relative height (hi_rel) and age (id) of the trees in the networks' training. Laasasenaho *et al.* (2005) verified that the bark thickness is more accurately estimated when it is included in the modeling processes the variables age and height of collection of the bark sample in the trees log.

Although limited attention to studies aimed at quantifying the bark of Brazilian Teak trees, the results obtained in this paper are promising. The proposed ANN model can be used for predictions of Teak bark at the regional level, since the effect of geographic features has little influence on the model in regions with similar climate characteristics (JOHNSON; WOOD, 1987; MALONE; LIANG, 2009).

Other researches are prominent and can be developed using neural network models with the inclusion of other numerical variables such as the site index diameter class. In addition, qualitative variables can also be used as the spacing, genetic material, region, among others. Nonetheless, this will lead to ANN models with a greater degree of complexity, but the gain in accuracy will certainly be considerable, which will reduce the sample intensity without losing the degree of accuracy in the estimatives.
CONCLUSION

• The artificial intelligence technique can be used to estimate the bark thickness of the *Tectona grandis* tree.
• The proposed model with the input variables: diameter, relative height and age can be used to support the administration and forest management of the species on a regional scale.
• Considering the low differences between statistical indicators verified by graphical analysis, it is possible to infer that all suggested network structures can be used to estimate the bark thickness of Teak trees.

REFERENCES


