# Accuracy of tree height estimation with model extracted from artificial neural network and new linear and nonlinear models

**Daniel Dantas**[*] , **Luiz Otávio Rodrigues Pinto, Talles Hudson Souza Lacerda, Natielle Gomes Cordeiro and Natalino Calegario**

Departamento de Ciências Florestais, Universidade Federal de Lavras, Campus Universitário, Cx. Postal 3037, 37203-703, Lavras, Minas Gerais, Brazil.
*Author for correspondence. E-mail: dantasdaniel12@yahoo.com.br

**ABSTRACT.** Variable height is commonly used as an input attribute to estimate other variables. Thus, to ensure less susceptibility to errors, it is necessary to obtain the variable height correctly. In addition to *DBH*, hypsometric relationships are influenced by several factors, such as site, age, genetic variation, and silvicultural practices. The inclusion of these factors in hypsometric models can lead to a gain in the quality of the estimates and in the biological realism. The objective of this study was to propose and evaluate the performance of a model extracted from artificial neural network training and of new models to estimate the total heights of eucalyptus trees. The data used in this study originated from temporary forest inventories conducted in eucalyptus stands in Minas Gerais, Brazil. A multilayer perceptron artificial neural network was trained, and a nonlinear equation was extracted from the best-performing network to predict the total heights of trees. New linear and nonlinear hypsometric models were constructed and fit considering variables related to individual trees (DBH) and stands (plot basal area, age and site index). The new hypsometric models proposed in this study showed satisfactory performance and are effective for estimating the total heights of eucalyptus trees, particularly the model extracted from the artificial neural network and the nonlinear model.

**Keywords:** hypsometric relationship; forest inventory; eucalyptus.

## Introduction

The demand for forest products is increasing, especially due to the growing global population. In this sense, the planning and management of forest plantations are increasingly important, and the accurate and effective evaluation of stands is crucial for appropriate management actions that meet the technical and economic goals of firms (Coelho Júnior, Rezende, & Oliveira 2013).

The forestry sector is constantly seeking techniques that facilitate successful timber production in a sustainable way, minimizing losses, optimizing processes and maximizing profits (David et al., 2017). In this context, the measurement procedures are essential elements and deserve attention because they are the basis for determining the volume in an area (Dantas & Oliveira, 2018).

One of the most important pieces of information to determine the potential of a forest in a given region is the variable "volume", the accurate quantification of which is essential in forest management planning. The individual volume serves as a starting point for assessing the wood content in a forest stand and provides support for decisions related to silvicultural practices and timber harvesting and transport. Thus, it is essential that the volume of trees is correctly determined to provide an accurate representation of the sampled population.

Volume estimates can be obtained through a forest inventory, which basically consists of determining a sampling method, allocating plots in the area and obtaining the variables of interest (Campos & Leite, 2017). The diameter can be obtained through measurements at a height of 1.30 m from the base of the tree, which is accurate and easy to perform. Measuring the height is performed indirectly and is among the challenges of forest inventories due to factors such as difficulty visualizing the top of the trees and the time required to complete the measurements (Vendruscolo et al., 2015). These factors, in addition to interfering with the accuracy of the measurements, significantly affect the cost of forest inventories. Therefore, there is a constant search for methodologies that provide accurate estimates while reducing the cost and time of the measurements.

In 1957, Ker and Smith proposed the use of hypsometric relationships, in which a height-diameter curve (hypsometric relationship) can be obtained from the measurement of the diameters (DBH) and the heights of some trees in the plot, which can be used to estimate the heights of the other trees. Several height prediction models have since been proposed and can be found in the literature (Campos & Leite, 2017).

In addition to DBH, hypsometric relationships are influenced by several factors, such as site, age, genetic variation, and silvicultural practices (Mendonça, Carvalho, & Calegario, 2015). The inclusion of these factors in hypsometric models can lead to a gain in the quality of the estimates and in the biological realism. However, the modeling and quantification of the effects of these characteristics on the variable to be estimated make this inclusion difficult because the relationships have nonlinear characteristics or qualitative (categorical) values (Binoti et al., 2018).

With the advancement of computational software and the diffusion of artificial intelligence, artificial neural networks (ANN) have been used as an alternative to hypsometric models for modeling, estimation of variables and forecasting of forest production (Castaño-Santamaría et al., 2013, Martins, Binoto, Leite, Binoti, & Dutram, 2016). An ANN is a processor composed of simple processing units (artificial neurons) that simulate the neurons in the human brain, which calculate certain functions. These units are distributed in layers and connected to each other by weights that store the experimental knowledge and weigh the inputs of each unit. With this, the acquired knowledge becomes available for use (Silva et al., 2018).

The most attractive characteristics of ANNs are the ability to learn and generalize information. In other words, ANNs can, through a learned example, generalize the knowledge assimilated to a set of unknown data (Lacerda, Cabacinha, Araújo Júnior, Maia, & Lacerda, 2017; Silva Junior, 2018). Another interesting characteristic of ANNs is the ability to extract nonexplicit characteristics from a set of information that is provided as examples (Özçelik, Diamantopoulou, Crecente-Campo, & Eler, 2013). In addition, during learning, networks can identify the relationship between the input and output attributes (Martins et al., 2016), in addition to having as an advantage the inclusion of categorical variables in the training (Vendruscolo et al., 2015).

One aspect that should be considered, with the adoption of ANNs as a modeling tool for forest management, is the possibility of reducing the number of measurements needed to train the networks without loss in the quality of the estimates. This would lead to a decrease in the time and cost of forest inventories. For this, studies are needed to provide support to the manager in the processing of forest inventory data using ANNs. Therefore, the objective of this study is to propose and evaluate the performance of a model extracted from ANN training and of new linear and nonlinear models to estimate the total height of eucalyptus trees.

## Material and methods

The data used in this study were obtained from temporary forest inventories conducted in eucalyptus stands in a 1,090-ha area located in Minas Gerais State, Brazil. The dataset comprised a total of 4,229 individuals distributed in 347 plots, and the following numerical variables were considered: age (months), height (meters), diameter at breast height (centimeters), plot basal area (m² ha⁻¹), and site index (m) (Table 1).

**Table 1.** Descriptive statistics of the forest inventory data of *Eucalyptus* spp. used, where H is the height in meters; DBH is the diameter, in centimeters, at 1.30 m above the ground; A is age, in months; G is the basal area, in $m^2 ha^{-1}$; and S is the site index, in meters.

| Variable | Minimum | Mean | Maximum | Standard Deviation | CV% |
|---|---|---|---|---|---|
| H | 5.00 | 21.26 | 31.50 | 3.94 | 18.54 |
| DBH | 5.00 | 14.51 | 26.20 | 2.97 | 20.44 |
| A | 36 | 59 | 78 | 9 | 15.93 |
| G | 6.05 | 17.28 | 34.21 | 5.11 | 29.59 |
| S | 12.50 | 25.84 | 32.50 | 3.27 | 12.67 |

## Training and validation of artificial neural networks

The ANNs were trained considering four input variables (DBH, age, site index, and basal area) and one output variable (height). The training process consisted of adjusting their weights through a learning algorithm that extracted characteristics from the data provided and aimed to generate an ANN that performed the task of interest (Binoti, Binoti, & Leite, 2014). The training was performed in R, version 3.4.1, using the neuralnet package (Günther & Fritsch, 2010).

The trained ANNs were multilayer perceptron (MLP) networks composed of an input layer, an intermediate layer and an output layer. The algorithm used was resilient backpropagation, in which the learning rate was automatically defined by the neuralnet package, with values ranging from 0.01 to 1.12.

The number of neurons in the intermediate layer was selected using *k-fold*. This methodology randomly subdivides the database into *k* subgroups (Ali & Pazzani, 1997; Cigizoglu & Kisi, 2006). The *k* value was 10 subgroups, at a ratio of 90% for training and 10% for testing (Dantas et al., 2020), applying cross-validation. Different numbers of neurons in the range of 1 to 20 were tested.

The activation function used was logistic (or sigmoidal), with a range from 0 to 1, which implies limiting the amplitude of the outputs and inputs. For this reason, the data were normalized by transforming the values of each variable into values between 0 and 1. The linear normalization was obtained by equation (1) (Soares, Flôres, Cabacinha, Carrijo, & Veiga, 2011), which considers the minimum and maximum values of each variable in the transformation of the values, maintaining the original data distribution (Valença, 2010).

$$x' = \frac{(x - x_{min}) * (b - a)}{(x_{max} - x_{min})} + a \tag{1}$$

where $x'$: normalized value; $x$: original value; $x_{min}$: minimum value of the variable; $x_{max}$: maximum value of the variable; $a$: lower limit of the normalization range; and $b$: upper limit of the normalization range.

Data were randomly divided into two groups: 70% for training and 30% for validation. Among the data used for ANN training, 70% were used in the training phase and 30% in the testing phase.

The criterion for stopping the ANN training process was a maximum number of 100,000 cycles, or a mean square error lower than 1%, and the training ended when one of the criteria was met. At the end of the training, the best ANN was selected based on the lowest mean square error.

Figure 1 illustrates the ANN architecture with the lowest error among those evaluated, composed of five neurons in the hidden layer.
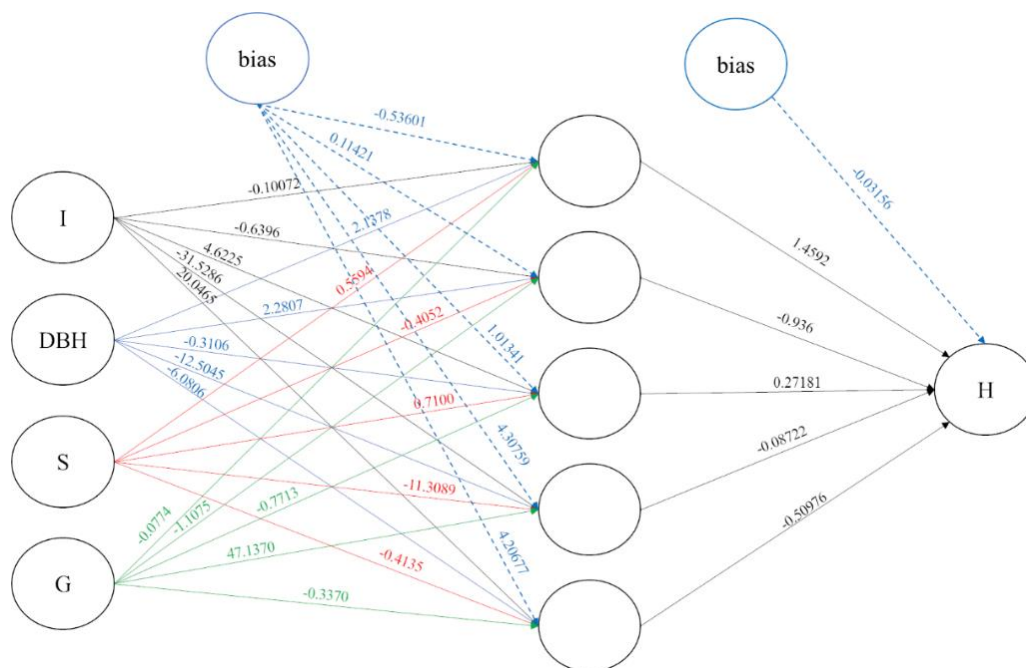


**Figure 1.** ANN architecture. Where H is the height in meters, DBH is the diameter at breast height in centimeters, I is age in months, G is the plot basal area in square meters, and S is the site index in meters.

A nonlinear equation was extracted from the ANN to predict the total height of trees. For this purpose, a system of equations with coefficients resulting from the weights generated by neurons in the neural network was generated. This system was used to predict the height of the trees that made up the validation dataset.

Model (2) expresses the relationship between the hidden layer and the response variable, where $\beta_0$ is the bias, and the other coefficients are the weights related to each neuron. Model (3) represents the activation function used in each neuron of the hidden layer, derived from the logistic model. Finally, model (4) is the result of the relationship between the input variables and the respective hidden layer neurons, generating a model for each neuron.

$$H' = \beta_0 + \beta_1 * z_1 + \beta_2 * z_2 + \beta_3 * z_3 + \beta_4 * z_4 + \beta_5 * z_5 + \varepsilon \tag{2}$$

$$z_n = \left[\frac{1}{1+e^{-w_i}}\right] \tag{3}$$

$$w_i = \beta_{0.n} + \beta_{1.n} * A_i' + \beta_{2.n} * DBH_i' + \beta_{3.n} * S_{ij}' + \beta_{4.n} * G_i' \tag{4}$$

where $H'$: normalized height; $\beta_0$: bias; $\beta_n$: coefficient of the model associated with neuron $n$; $\beta_{k.n}$: coefficient of the model between input variable k and neuron $n$; $z_n$: response of the n-th neuron of the hidden layer; $w_i$: sum of the products between the weights and the inputs; $Ai'$: normalized age of the i-th tree (months); $DBH_i$: normalized diameter at breast height of the i-th tree (cm); $S_{ij}$: site index of the j-th sampling unit containing the i-th tree (m), normalized; $G_i$: normalized basal area of the i-th plot (m² ha⁻¹); and $\varepsilon$: error.

To assess the sensitivity of the variables and observe whether they exhibited biological behavior, the sensitivity analysis proposed by Lek (Lek et al., 1996) was used to analyze the contribution of the input/output variable.

### Fit of the linear and nonlinear models

The data were processed using the statistical software program R, version 3.4.1 (R Development Core Team, 2017). To compare the performance of the ANN and the hypsometric models, the data used to fit the models were the same as those used in the ANN training. Seventy percent of the data were allocated to the fit and 30% to the validation - i.e., the prediction of the total height in a dataset not used in the fit.

To estimate the total height of eucalyptus trees, linear and nonlinear models were constructed and fitted, considering variables related to an individual tree (DBH) and stand (basal area, age and site index), based on the functional relationship presented below:

$$f(H_i) = f(A_i, DBH_i, G_i, S_{ij})$$

where $H_i$: height of the i-th tree (m); $DBH_i$: diameter at breast height of the i-th tree (cm); $A_i$: age of the i-th tree (months); $G_i$: basal area of the i-th plot (m² ha⁻¹); and $S_{ij}$: site index of the j-th sampling unit containing the i-th tree (m).

To construct and fit the models, interactions between the independent variables were tested and multicollinearity was evaluated. For this step, transformation of the variables was necessary to avoid conflict in each variable's influence (positive or negative) on the variability of $H$. The fit of the hypsometric model was also evaluated by analysis of variance (ANOVA) of the regression, parametric analysis and residual analysis. If the dependent variable was transformed using the natural logarithm, the equations were compared by the Furnival index, in which lower values indicated better performance (Furnival 1961).

$$FI = exp\left[\frac{\sum_{i=1}^{n} \ln(Hreal_i)}{n}\right] * S_{YX} \tag{5}$$

where $FI$ is the Furnival index; $Hreal_i$ is the real height of the i-th individual, in meters; $S_{YX}$ is the standard error of the estimate; and $n$ is the number of trees sampled.

The linear (Equation 6) and nonlinear (Equation 7) hypsometric models constructed are shown below:

$$\ln H = \beta_0 + \beta_1 \left(\frac{1}{DBH_i}\right) + \beta_2 \left(\frac{1}{A_i}\right) + \beta_3 \left(\frac{1}{G_i}\right) + \beta_4 \left(\frac{1}{S_{ij}}\right) + \varepsilon \tag{6}$$

$$H = \left[\left(\beta_0 + \beta_1 * S_{ij} + \beta_2 * \left(\frac{1}{A_i}\right)\right) * e^{\beta_3 + \beta_4 * \left(\frac{1}{G_i}\right) + \beta_5 * \left(\frac{1}{A_i} * \frac{1}{DBH_i}\right)}\right] + \varepsilon \tag{7}$$

where $H_i$: height of the i-th tree (m); $\beta_k$: regression coefficients; $DBH_i$: diameter at breast height of the i-th tree (cm); $A_i$: age of the i-th tree (months); $G_i$: basal area of the i-th plot (m² ha⁻¹); $S_{ij}$: site index of the j-th sampling unit containing the i-th tree (m); and $\varepsilon$: random error (m).

### Fit of the parabolic model

The parabolic hypsometric model, in its original (Equation 8) and modified (Equation 9) versions, was used as a reference in this study. The purpose of the modification was to standardize the variables used with the other analyzed techniques.

$$H = \beta_0 + \beta_1 DBH_i + \beta_2 DBH_i^2 + \varepsilon \tag{8}$$

$$H = \beta_0 + \beta_1 DBH_i + \beta_2 A_i + \beta_3 G_i + \beta_4 S_{ij} + \varepsilon \tag{9}$$

where $H_i$: height of the i-th tree (m); $\beta_k$: regression coefficients; $DBH_i$: diameter at breast height of the i-th tree (cm); $A_i$: age of the i-th tree (months); $G_i$: basal area of the i-th plot (m² ha$^{-1}$); $S_{ij}$: site index of the j-th sampling unit containing the i-th tree (m); and $\varepsilon$: random error (m).

### Analysis of estimates generated by the model extracted from ANN and hypsometric models

The quality of the estimates by the models was evaluated based on ANOVA, graphical residual analysis, Pearson correlation, root mean square error (RMSE) (Siipilehto, 2000, Campos & Leite, 2017), the Akaike information criterion (Sakamoto, Ishiguro, & Kitagawa, 1986) (Equation 11) and the Graybill F-test (Graybill, 1976) (Equation 12).

$$RMSE\ (\%) = \left[\sqrt{\sum_{i=1}^{n}\frac{(Yi-\widehat{Yi})^2}{n}}\ /\bar{Y}\right] * 100 \tag{10}$$

$$AIC = -2 * \ln(ml) + 2p \tag{11}$$

where *RMSE*: root mean square error; *Yi*: observed value; *Y^i*: estimated value; *n:* number of cases; Y⁻: mean of observed values; *AIC*: Akaike information criterion; ln: natural logarithm; *ml* = maximum likelihood value; p: number of model parameters.

The Graybill F-test, for comparison between the model predictions and the real values, uses the coefficients generated by equation (13).

$$F\ (H_0) = \frac{\left(\begin{bmatrix}\hat{\beta}_0\\\hat{\beta}_1\end{bmatrix}-\begin{bmatrix}0\\1\end{bmatrix}\right)'\begin{bmatrix}n & \sum x\\\sum x & \sum x^2\end{bmatrix}^{-1}\left(\begin{bmatrix}\hat{\beta}_0\\\hat{\beta}_1\end{bmatrix}-\begin{bmatrix}0\\1\end{bmatrix}\right)}{2*MS_{Res}} \tag{12}$$

$$Y = \beta_0 + \beta_1 x + \varepsilon \tag{13}$$

Where $\beta^{\wedge}_0$ and $\beta^{\wedge}_1$: estimated model parameters (Equation 11); *Y*: estimated variable; $\beta_0$ and $\beta_1$: model parameters; *X*: observed variable; and $\varepsilon$: random error.

# Results

The coefficients of the system of equations extracted from the artificial neural network are presented in Table 2. The coefficients of each neuron of the hidden layer and of the output layer of the ANN are presented.

**Table 2.** Parameters of the artificial neural network.

| Model | β0 | β1 | β2 | β3 | β4 | β5 |
|---|---|---|---|---|---|---|
| ANN | -0.0316 | 1.4592 | -0.9360 | 0.2718 | -0.0872 | -0.5098 |
| Neuron 1 | -0.0536 | -0.1007 | 2.1378 | 0.5594 | -0.0774 | - |
| Neuron 2 | 0.1142 | -0.6396 | 2.2807 | -0.4052 | -1.1075 | - |
| Neuron 3 | 1.0134 | 4.6225 | -0.3106 | 0.7100 | -0.7713 | - |
| Neuron 4 | 4.3076 | -31.5286 | -12.5045 | -11.3089 | 47.1370 | - |
| Neuron 5 | -4.2068 | 20.0465 | -6.0806 | -0.4135 | -0.3370 | - |

Table 3 shows the fitted parameters and the significance of the regression models used in the present study. The coefficients related to the independent variables were significant by Student's t-test at a significance level of 0.05. All coefficients of the nonlinear model (Equation 7) were significant, indicating that all variables contributed to the prediction of height. Among the parameters of the modified parabolic model, all coefficients were significant at a level of 0.05. The coefficients of the parabolic model were significant, except for the coefficient associated with $DBH^2$.

**Table 3.** Synthesis of the adjustments for the regression models with significance level of 0.05.

| Linear regression | | Nonlinear regression | | Modified parabolic | | Parabolic | |
|---|---|---|---|---|---|---|---|
| Variables | Coef | Variables | Coef | Variables | Coef | Variables | Coef |
| - | 4.367* | - | 36.120* | - | -8.681* | - | 6.511* |
| $DBH^{-1}$ | -4.014* | S | 0.607* | DBH | 0.525* | DBH | 1.003* |
| $A^{-1}$ | -28.323* | $A^1$ | -1.314E03* | A | 0.180* | $DBH^2$ | 1.10E-03$^{ns}$ |
| $G^{-1}$ | -0.749* | - | -8.439* | G | 0.046$^*$ | - | - |
| $S^{-1}$ | -12.126* | $G^{-1}$ | -10.340* | S | 0.421* | - | - |
| - | - | $(DBH^{-1})(I^{-1})$ | 272.900* | - | - | - | - |

where DBH: diameter at 1.30 m above ground (cm); A: age (m); G: basal area (m² ha$^{-1}$); S: site index; *: significance at 0.05; and $^{ns}$: denotes no significance.

Table 4 shows the ANOVA results of the models, where all were significant at a level of 0.05. The parabolic model had the largest mean square of residuals, while the model extracted from the ANN had the smallest. The same behavior was evidenced by the coefficient of determination ($R^2$). According to the Furnival index (FI), the equations presented low values, except for the parabolic equation.

**Table 4.** Analysis of variance of the models evaluated.

| Source | DF | SS | MS | F |
|---|---|---|---|---|
| ANOVA ANN ($R^2$ 85.22; Syx: 1.69 m; FI: 1.69) | | | | |
| Source | DF | SS | MS | F |
| ANN | 25 | 38,847.32 | 1,553.89 | 546.17 |
| Error | 2,339 | 6,654.59 | 2.85 | - |
| TOTAL | 2,364 | 45,501.91 | | |
| ANOVA linear regression ($R^2$ 84.27; Syx: 1.74 m; IF: 1.78) | | | | |
| Source | DF | SS | MS | F |
| Regression | 4 | 38,001.83 | 9,500.46 | 2,989.45 |
| Error | 2,360 | 7,500.08 | 3.18 | - |
| TOTAL | 2,364 | 45,501.91 | | |
| ANOVA nonlinear regression ($R^2$ 83.49; Syx: 1.78 m; FI: 1.57) | | | | |
| Source | DF | SS | MS | F |
| Regression | 6 | 38,187.37 | 6,364.56 | 2,051.75 |
| Error | 2,358 | 7,314.54 | 3.10 | - |
| TOTAL | 2,364 | 45,501.91 | | |
| ANOVA Modified parabolic ($R^2$ 82.73; Syx: 1.82 m; FI: 1.57) | | | | |
| Source | DF | SS | MS | F |
| Regression | 4 | 37,655.71 | 9,413.93 | 2,831.54 |
| Error | 2,360 | 7,846.20 | 3.32 | - |
| TOTAL | 2,364 | 45,501.91 | | |
| ANOVA Parabolic ($R^2$ 60.22; Syx: 2.77 m; FI: 2.47) | | | | |
| Source | DF | SS | MS | F |
| Regression | 2 | 27,417.77 | 13,708.88 | 1,790.54 |
| Error | 2,362 | 18,084.14 | 7.66 | - |
| TOTAL | 2,364 | 45,501.91 | | |

where DF: degrees of freedom; SS: sum of squares; MS: mean squares; F: significance test.

Figure 2 shows the graph resulting from the Lek algorithm, with the mean being the fixed variable for constructing this graph. All variables showed a behavior directly proportional to height.
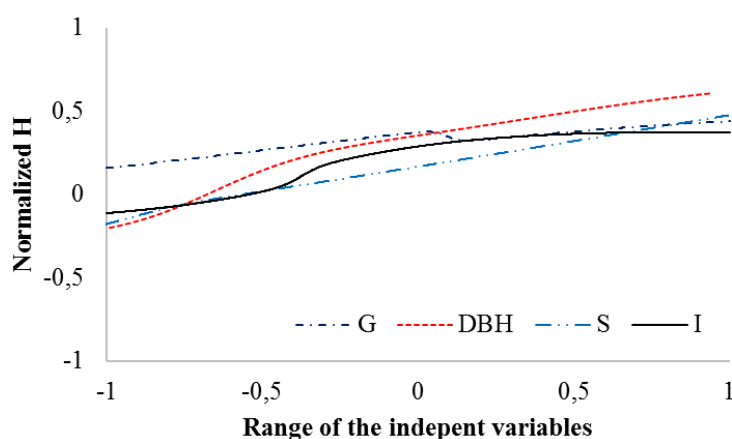


**Figure 2.** Graph resulting from the Lek algorithm for evaluating the behavior of variables as a function of height.

The scatterplots of residuals for the different techniques used are presents in the Figure 3. There was a greater tendency for errors at the higher height values. The best results were obtained by the equation extracted from the ANN, with greater homogeneity of the residuals, and by the nonlinear equation, with lower dispersion of the residuals. The parabolic model resulted in greater dispersion.

Table 5 shows the statistics from the evaluation of the validation quality of the equations. The parabolic equation, used as a reference in this study, presented the worst statistics. According to the analyses, Figure 4 shows the trend of total height estimates by the different techniques evaluated, compared with the average trend line of the predictions. When analyzing the predictions of the models with the 45th line, it is noted that

the prediction techniques had similar behavior, except for the parabolic model predictions, which provided an overestimation for lower heights and an underestimation for greater heights.
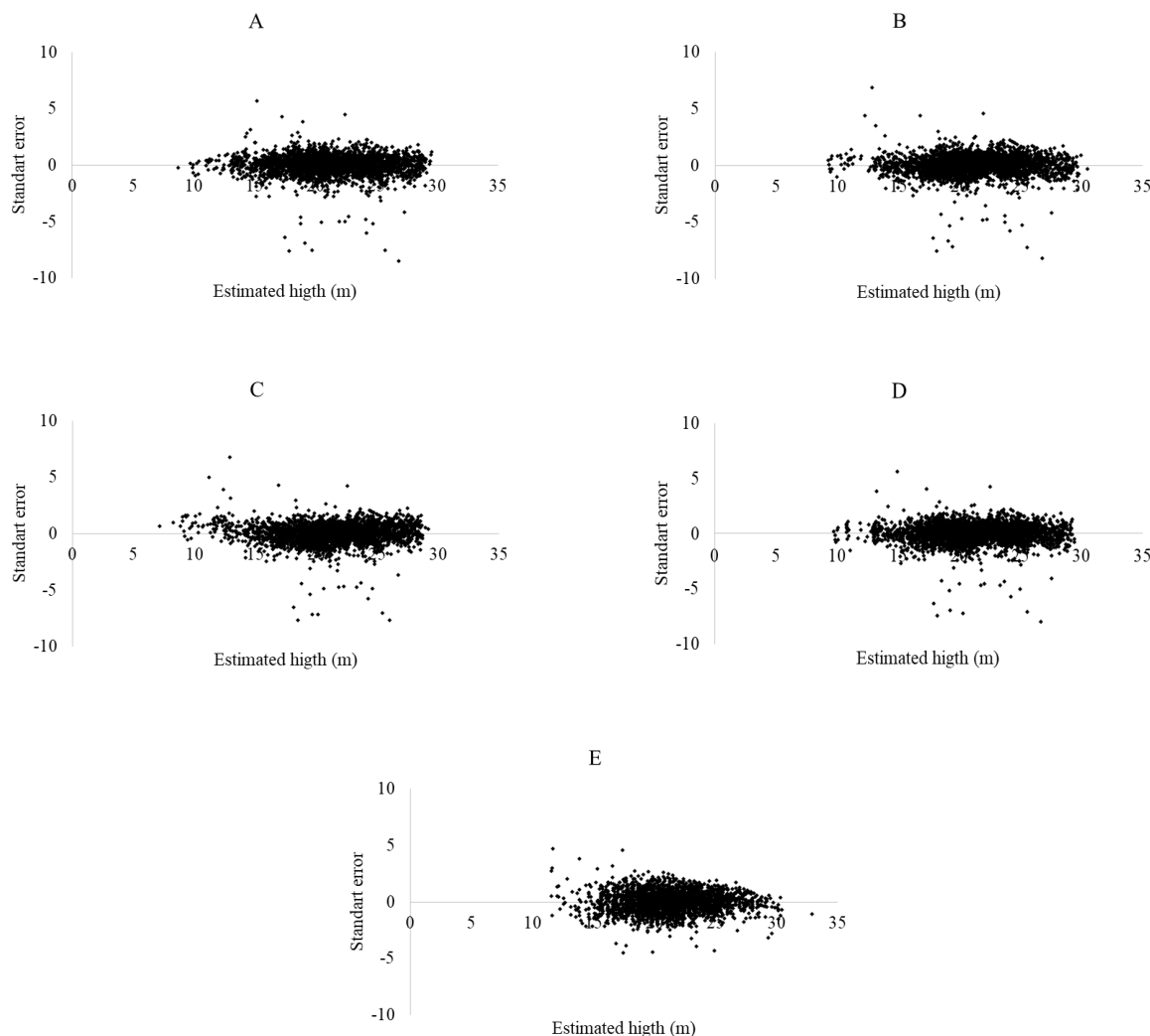


**Figure 3.** Scatterplot of residuals of the fitted equations, (A): ANN, (B): Linear, (C): Nonlinear, (D): Modified parabolic, and (E): Parabolic.

**Table 5.** Equation evaluation criteria.

| | Correlation | RMSE (%) | AIC | Graybill F-Test |
|---|---|---|---|---|
| ANN | 0.93 | 6.95 | 2.19 | 6.45E-04[ns] |
| Linear Regression | 0.92 | 7.25 | 2.39 | 2.09E-02[ns] |
| Nonlinear regression | 0.92 | 7.16 | 2.32 | 2.04E-04[ns] |
| Modified Parabolic | 0.91 | 7.51 | 2.61 | 6.97E-04[ns] |
| Parabolic | 0.79 | 11.66 | 6.15 | 3.03E-03[ns] |

[ns]: not significant at 0.05.

The estimates of the equations were strongly correlated with the observed values, with an error magnitude lower than 10%, except for the parabolic equation. The lower the RMSE, the greater the accuracy of the estimates, and the optimal situation is an RMSE of zero (Mehtätalo, Maltamo, & Kangas, 2006). The RMSE values for the equations proposed in this study (ANN, linear and nonlinear regression) and the modified parabolic equation were similar. The Akaike information criterion for the equations proposed in this study did not show significant differences between them, with the exception of the parabolic equation, according to the criterion proposed by Burnham and Anderson (2002), in which differences smaller than or equal to two do not indicate a significant difference. The Graybill F-test indicated no significant difference between the values observed and estimated by the equations evaluated.

The equation extracted from the ANN and the nonlinear equation achieved the best performance because they presented the best statistics for the evaluated criteria compared with the other criteria.
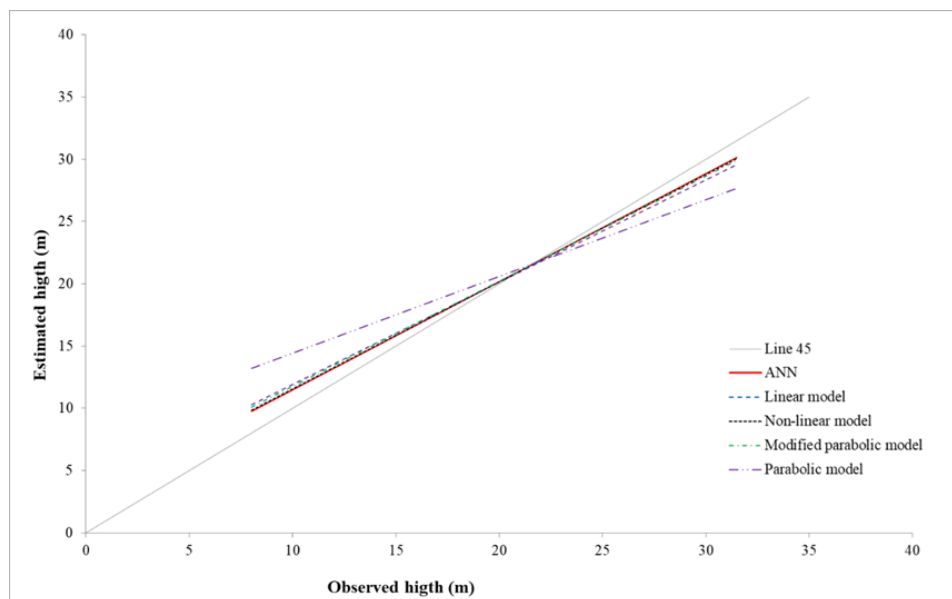
**Figure 4.** Trend plot of the heights observed as a function of the heights estimated by the equations.

## Discussion

ANNs are widely disseminated in the forest sector and have a high ability to describe the behavior of the relationship between dependent and independent variables (Reis et al., 2018; Nandy et al., 2017; Vieira et al., 2018; Dantas et al., 2020; Dantas, Terra, Schorr, & Calegario, 2021). One of the variables used by this technique is total height due to its high cost and difficulty to obtain it in the field. When comparing ANNs with other techniques for height prediction, it is observed that ANNs have good performance. However, discussions of ANNs have not been based on the mathematical relationships between neurons and the studied variables, and such relationships can be considered for obtaining a model and, consequently, its regression statistics, thus allowing comparison with other models.

The coefficients of the equation extracted from the ANN in this study proved to be effective for estimating the total height. This can be attributed to the nonlinear behavior generated by the hyperbolic tangent activation function, which can well describe the biological relationship between height and independent variables. Based on the sensitivity analysis, the variables exhibited biological behavior relative to total height, which is the expected behavior.

Generally, classic hypsometric models encompass only the use of the variable DBH, as in the case of parabolic and Curtis models (Curtis, 1967). Retslaff, Figueiredo Filho, Dias, Bernett, and Figura (2015) reported that the inclusion of other variables, such as age and site index, provides greater accuracy in estimating the height variable, and it is necessary to know the behavior of the effect of the independent variables on the response variable.

To obtain the linear model proposed in this study, the transformations used were the inverse of the independent variables and the natural logarithm of the dependent variable, which contributed to the improvement of the correlation between the variables and generated a linear behavior. Therefore, the variables were effective for predicting height. As the main interaction (1/DBH, 1/A, and 1/G) was significant, it was necessary to break it down to assess the presence of secondary interactions between these variables (6), which allowed a better understanding of its effects on the dependent variable (Preacher, Curran, & Bauer, 2006). The interaction of the site index with the other continuous variables was not significant due to its categorical nature (Aertsen, Kint, Orshoven, Özkan, & Muys, 2010); therefore, it was not considered. However, the site index was effective when included in the model in an isolated manner, as shown in equation (6), helping better position the straight line. The effect of this variable on the linear coefficient was also observed in the model used by Nogueira, Marshall, Leite, and Campos (2015).

Linear models are widely used to model independent variables because they are easy to fit and interpret. However, they do not allow biological interpretation because the existing relationships between the dependent and independent variables are nonlinear (Gevrey, Dimopoulos, & Lek, 2003).

All variables of the fitted linear model showed biological behavior because their related coefficients had negative signs as their inverse was used. Such behavior was also observed in the modified parabolic equation.

With respect to the proposed nonlinear model, the site index and age affect the intercept of the equation. These variables classify the stand according to their hypsometric growth, influencing the location of the curve fitted when categorizing the data. The site index is strictly related to the height of the stands, contributing to the homogeneity of the residuals (Retslaff et al., 2015) and to the prediction of height (Sena, Silva Neto, Oliveira, & Calegario, 2015; Téo, Machado, Figueiredo Filho, & Tomé, 2017). This variable indicates the productive potential of the stand, classifying it by means of the dominant heights. Regarding age, the relationship with the minimum height is directly proportional, where the minimum height describes the position where the equation intercepts the height axis.

According to Silva, Xavier, Rodrigues, and Peternelli (2007), in addition to the linear coefficient of the equation, there is also an effect of the variable age on the slope, given that there is a greater contribution of this variable to the height variation at higher DBH values. Regarding the plot basal area, this variable represents the degree of occupation of the area, indicating competitiveness among individuals. The effect of competitiveness differs according to the sociological position of trees, with a stronger effect on dominated trees and a weaker effect on dominant ones (Bartoszeck, Machado, Figueiredo Filho, & Oliveira, 2004). In this sense, it influences the slope of the fitted curve.

Total height modeling is widely used to support forest inventory and management. Because height has a high measurement cost, reducing the height measurement intensity is a promising alternative given that the prediction model extracted from ANNs requires a smaller sample size for prediction. Furthermore, the techniques used in this study include data from remeasures, allowing the model to better express behavior at different ages.

Although the model has many independent variables, the techniques are easy to apply because the variables are derived from the data commonly collected in forest surveys-namely, DBH and total and dominant height and age. Using these variables, the biological behavior of their interactions was assessed and evaluated.

The statistics presented by the parabolic equation indicate the importance of including variables that enable the height variability to be described, as shown in the models proposed in this study. The model extracted from the ANN and the nonlinear model achieved better performance for accurately describing the biological behavior among the variables.

# Conclusion

The hypsometric models proposed in this study considerably improves the modeling of the total height of eucalyptus trees. The techniques performed satisfactorily, and the model extracted from the artificial neural network and the new nonlinear model more accurately predicted the total height of eucalyptus trees.

# Acknowledgements

# References

Aertsen, W., Kint, V., Orshoven, J. V., Özkan, K., & Muys, B. (2010). Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling*, *221*(8), 1119-1130. DOI: https://doi.org/10.1016/j.ecolmodel.2010.01.007

Ali, K. M., & Pazzani, M. J. (1997). Error reduction through learning multiple descriptions. *Machine Learning*, *24*(3), 173-202. DOI: https://doi.org/10.1007/BF00058611

Bartoszeck, A. C. P. S., Machado, S. A., Figueiredo Filho, A., & Oliveira, E. B. (2004). Dinâmica da relação hipsométrica em função da idade, do sítio e da densidade inicial de povoamentos de bracatinga da Região Metropolitana de Curitiba, PR. *Revista Árvore*, *28*(4), 517-533. DOI: http://doi.org/10.1590/S0100-67622004000400006

Binoti, D. H. B., Binoti, M. L. M. S., & Leite, H. G. (2014). Configuração de redes neurais artificiais para estimação do volume de árvores. *Brazilian Journal of Wood Science*, *5*, 58-67. DOI: https://doi.org/10.12953/2177-6830.v05n01a06

Binoti, D. H. B., Duarte, P. J., Silva, M. L. M., Silva, G. F., Leite, H. G., Mendonça, A. R., ..., Vega, A. E. D. (2018). Estimation of height of eucalyptus trees with neuroevolution of augmenting topologies (NEAT). *Revista Árvore*, *41*(3), 1-8. DOI: https://doi.org/10.1590/1806-90882017000300014

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.). New York, NY: Springer.

Campos, J. C. C., & Leite, H. G. (2017). *Mensuração florestal: perguntas e respostas* (5. ed.). Viçosa, MG: UFV.

Castaño-Santamaría, J., Crecente-Campo, F., Fernández-Martínez, J. L., Barrio-Anta, M., & Obeso, J. R. (2013). Tree height prediction approaches for uneven-aged beech forests in northwestern Spain. *Forest Ecology Management*, *307*, 63-73. DOI: https://doi.org/10.1016/j.foreco.2013.07.014

Cigizoglu, H. K., & Kisi, Ö. (2006). Methods to improve the neural network performance in suspended sediment estimation. *Journal of Hydrology*, *317*(3-4), 221-238. DOI: https://doi.org/10.1016/j.jhydrol.2005.05.019

Coelho Junior, L. M., Rezende, J. L. P., & Oliveira, A. D. (2013). Concentração das exportações mundiais de produtos florestais. *Ciência Florestal*, *23*(4), 691-701. DOI: https://doi.org/10.5902/1980509812353

Curtis, R. 1(967). Height-diameter and height-diameter-age equations for second-growth Douglas-fir. *Forest Science*, *13*(4), 365-375. DOI: https://doi.org/10.1093/forestscience/13.4.365

Dantas, D., & Oliveira, M. L. R. (2018). *Redes neurais artificiais na estimação da altura de árvores*. São Paulo, SP: Novas Edições Acadêmicas.

Dantas, D., Calegario, N., Acerbi Junior, F. W., Carvalho, S. P. C., Isaac Junior, M. A., & Melo, E. A. (2020). Multilevel nonlinear mixed-effects model and machine learning for predicting the volume of *Eucalyptus* spp. trees. *Cerne*, *26*(1), 48-57. DOI: https://doi.org/10.1590/01047760202026012668

Dantas, D., Terra, M. C. N. S., Schorr, L. P. B., & Calegario, N. (2021). Machine learning for carbon stock prediction in a tropical forest in Southeastern Brazil. *Bosque*, *42*(1), 131-140. DOI: http://dx.doi.org/10.4067/S0717-92002021000100131

David, H. C., Arce, J. E., Oliveira, E. B., Netto, S. P., Miranda, R. O. V., & Ebling, A. A. (2017). Economic analysis and revenue optimization in management regimes of *Pinus taeda*. *Revista Ceres*, *64*(3), 222-231. DOI: https://doi.org/10.1590/0034-737x201764030002

Furnival, G. M. (1961). An index for comparing equations used in constructing volume tables. *Forest Science*, *7*(4), 337-341. DOI: https://doi.org/10.1093/forestscience/7.4.337

Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, *160*(3), 249-264. DOI: https://doi.org/10.1016/S0304-3800(02)00257-0

Graybill, F. A. (1976). *Theory and application of the linear model*. Belmont, US: Duxbury Press.

Günther, F., & Fritsch, S. (2010). Neuralnet: Training of neural networks. *The R journal*, *2*(1)**,** 30-38. DOI: https://doi.org/10.32614/RJ-2010-006

Lacerda, T. H. S., Cabacinha, C. D., Araújo Júnior, C. A., Maia, R. D., & Lacerda, K. W. S. (2017). Artificial neural networks for estimating tree volume in the brazilian savanna. *Cerne*, *23*(4), 483-491. DOI: https://doi.org/10.1590/01047760201723042347

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., & Aulagnier, S. (1996). Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling*, *90*(1), 39-52. DOI: https://doi.org/10.1016/0304-3800(95)00142-5.

Martins, E. D. R., Binoto, M. L. M. S., Leite, H. G., Binoti, D. H. B., & Dutram G. C. (2016). Configuração de redes neurais artificiais para estimação da altura total de árvores de eucalipto. *Revista Brasileira de Ciências Agrárias*, *11*(1), 117-123. DOI: https://doi.org/10.5039/agraria.v11i1a5354

Mehtätalo, L., Maltamo, M., & Kangas, A. (2006). The use of quantile trees in the prediction of the diameter distribution of a stand. *Silva Fennica*, *40*(3), 501-516. DOI: https://doi.org/10.14214/sf.333

Mendonça, A. R., Carvalho, S. P. C., & Calegario, N. (2015). modelos hipsométricos generalizados mistos na predição da altura de *Eucalyptus* sp. *Cerne*, *21*(1), 107-115. DOI: http://dx.doi.org/10.1590/01047760201521011191

Nandy, S., Singh, R., Ghosh, S., Watham, T., Kushwaha, S. P. S., Kumar, A. S., & Dadhwal, V. K. (2017). Neural network-based modelling for forest biomass assessment. *Carbon Management, 8*(4), 305-317. DOI: https://doi.org/10.1080/17583004.2017.1357402

Nogueira, G. S., Marshall, P. L., Leite, H. G., & Campos, J. C. C. (2015). Thinning intensity and pruning impacts on Eucalyptus plantations in Brazil. *International Journal of Research*, *2015*, 1-10. DOI: http://dx.doi.org/10.1155/2015/168390

Özçelik, R., Diamantopoulou, M. J., Crecente-Campo, F., & Eler, U. (2013). Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. *Forest Ecology Management, 306*, 52-60. DOI: https://doi.org/10.1016/j.foreco.2013.06.009

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*(4), 437-448. DOI: https://doi.org/10.3102/10769986031004437

R Development Core Team. (2017). *R: a language and environment for statistical computing*. Vienna, AT: R Foundation for Statistical Computing.

Reis, L. P, Souza, A. L., Reis, P. C. M., Mazzei, L., Soares, C. P. B., Torres, C. M. M. E., ... Leite, H. G. (2018). Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. *Ecological Engineering*, *112*, 140-147. DOI: https://doi.org/10.1016/j.ecoleng.2017.12.014

Retslaff, F. A. S., Figueiredo Filho, A., Dias, A. N., Bernett, L. G., & Figura, M. A. (2015). Curvas de sítio e relações hipsométricas para *Eucalyptus grandis* na região dos Campos Gerais, Paraná. *Cerne*, *21*(2), 219-225. DOI: http://dx.doi.org/10.1590/01047760201521021349

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. Dordrecht, NT: D. Reidel.

Sena, A. L. M., Silva Neto, A. J., Oliveira, G. M. V., & Calegario, N. (2015). Linear and nonlinear models with use of covariates for hypsometric relationship of two tropical pine species. *Ciência Florestal*, *25*(4), 969-980. DOI: http://dx.doi.org/10.5902/1980509820651

Siipilehto, J. (2000). A comparison of two parameter prediction methods for stand structure in Finland. *Silva Fennica*, *34*(4), 331-349. DOI: https://doi.org/10.14214/sf.617

Silva, G. F., Xavier, A. C., Rodrigues, F. L., & Peternelli, L. A. (2007). Análise da influência de diferentes tamanhos e composições de amostras no ajuste de uma relação hipsométrica para *Eucalyptus grandis*. *Revista Árvore*, *31*(4), 685-694. DOI: http://doi.org/10.1590/S0100-67622007000400013

Silva, J. P. M., Cabacinha, C. D., Assis, A. L., Monteiro, T. C., Araújo Júnior, C. A, & Maia, R. D. (2018). Redes neurais artificiais para estimar a densidade básica de madeiras do cerrado. *Pesquisa Florestal Brasileira*, *38*, 1-10. DOI: https://doi.org/10.4336/2018.pfb.38e201801656

Silva Junior, E. M., Maia, R. D., & Cabacinha, C. D. (2018). Bee-inspired RBF network for volume estimation of individual trees. *Computers and Electronics in Agriculture*, *152*, 401-408. DOI: https://doi.org/10.1016/j.compag.2018.07.036

Soares, F. A. A. M. N., Flôres, E. L., Cabacinha, C. D., Carrijo, G. A., & Veiga, A. C. P. (2011). Recursive diameter prediction and volume calculation of eucalyptus trees using Multilayer Perceptron Networks. *Computers and Electronics in Agriculture*, *78*(1), 19-27. DOI: https://doi.org/10.1016/j.compag.2011.05.008

Téo, S. J., Machado, S. A., Figueiredo Filho, A., & Tomé, M., (2017). General height-diameter equation with biological attributes for *Pinus taeda* L. stands. *Cerne*, *23*(4), 403-411. DOI: http://doi.org/10.1590/01047760201723042414

Valença, M. (2010). *Fundamentos das redes neurais: exemplos em Java* (2. ed.). Olinda, PE: Livro Rápido.

Vendruscolo, D. G. S., Drescher, R., Souza, H. S., Moura, J. P. V. M., Mamoré, F. M. D., & Siqueira, T. A. S. (2015). Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais artificiais. *Revista Brasileira de Biometria*, *33*(4), 556-569. DOI: http://doi.org/10.13140/RG.2.1.1742.5684

Vieira, G. C. V., Mendonça, A. R., Silva, G. F., Zanetti, S. S., Silva, M. M., & Santos, A. R. (2018). Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. *Science of the Total Environment*, *619-620*, 1473-1481. DOI: https://doi.org/10.1016/j.scitotenv.2017.11.138