# Alternatives to Growth and Yield Prognosis for *Pinus caribaea* var. *caribaea* Barrett & Golfari

Ouorou Ganni Mariel Guera[1] , José Antônio Aleixo da Silva[1] ,
Rinaldo Luiz Caraciolo Ferreira[1] , Daniel Alberto Álvarez Lazo[2] ,
Héctor Barrero Medel[2]

[1]Universidade Federal Rural de Pernambuco (UFRPE), Recife, PE, Brasil
[2]Universidad de Pinar del Río (UPR), Pinar del Río, Cuba

**ABSTRACT**

The objective of this study was to obtain regression equations and artificial neural networks (ANNs) for prediction and prognosis of the yield of *Pinus caribaea* var. *caribaea* Barrett & Golfari. The data used for modeling comes from measuring the variables diameter at breast height (DBH) and total height (Ht) in 550 temporary plots and 14 circular permanent plots with 500 m² in *Pinus caribaea* var. *caribaea* plantations, aged between 3 and 41 years old. In growth prediction, the results indicated Schumacher model as the best fit to the data. On prognosis, the modified Buckman system was better than Clutter's. ANNs presented a similar performance to the Buckman model in volume prognosis, however these were superior for basal area prognosis.

**Keywords:** plantations, nonlinear regression, artificial neural networks.

## 1. INTRODUCTION AND OBJECTIVES

Mathematical models are not new in the forest area and are one of the most important approaches in the study of forest dynamics. In these studies, present estimates (predictions) and future estimates (prognosis) made with modeling techniques, both at the tree and stand level, are essential steps to enable forestry activity planning (Prodan et al., 1997).

Mathematical modeling refers to the development or adjustment of mathematical expressions that describe the behavior of a variable of interest. Regression analysis, a statistical technique whose name is attributed to the British anthropologist Francis Galton (Draper & Smith, 1998), is the most used technique in empirical modeling research, especially when the objective is to describe an existing but hidden relation between a set of independent variables and a dependent variable (Pardoe, 2012).

Equations, the main results of the regression analysis, help forest researchers and managers to forecast future forest yields to select better management options, appropriate silviculture alternatives or to plan forest harvest frequencies and sequences (Burkhart & Tomé, 2012).

When discussing the difference between prediction and prognosis models, it is worth noting that prognoses are performed by regression models in the form of equation systems that estimate the parameters of the function for the projection of production for future ages (Castro et al., 2013) and prediction models can be defined as functions that simply describe the change in the size of an individual (tree) or population (population) over time (age) (Burkhart & Tomé, 2012).

From the perspective of the input variable components for the models, Binoti et al. (2015) assert that prediction is carried out by models that have age as an independent variable, while prognosis is performed by models in which future production is projected as a function of current production among other variables. The errors associated with these prognosis models grow over time, and considering the long horizons of the planning of forest productive processes, making precise forecasts has become the main challenge of forest yield models.

In the last decades, the need for more accurate estimates has led to techniques such as artificial neural networks (ANNs) becoming popular for forest measurement. Due to their effectiveness in understanding complex systems, these modeling techniques are used as alternatives to the adjustment of traditional nonlinear regression models (Özçelik et al., 2017). The ANNs can be defined as mathematical models that have the functioning of the human brain with its biological neural networks as a metaphor (Valença, 2010).

Forest plantation growth and yield modeling using regression analysis were approached in numerous researches, such as the outstanding studies by Schumacher (1939), Buckman (1962) and Clutter (1963). In applied regression analysis, we highlight authors such as Draper & Smith (1998) and Pardoe (2012) whose studies were complementary to specific forest literature. Studies of Ashraf et al. (2013), Castro et al. (2013), Özçelik et al. (2014), are among studies that applied the techniques of ANNs for the same purpose.

Given the above, this study sought to fit regression models and train ANNs for the prediction and prognosis of *Pinus caribaea* var. *caribaea* growth and yield at Macurije forest company, Pinar del Río, Cuba.

## 2. MATERIALS AND METHODS

### 2.1. Geographical location of the study area

This study was carried out in plantations of *Pinus caribaea* var. *caribaea* of a company called Macurije located between the coordinates 22º06' to 22º42' latitude North and 83º48' to 84º23' longitude west, in the most western region of the province of Pinar del Río, Cuba (Figure 1).

### 2.2. Data sources and analysis of sample sufficiency

The database used consisted of 550 temporary plots and 14 circular permanent plots of 500 m² in plantations of *Pinus caribaea* var. *caribaea* with ages ranging from 3 to 41 years old. Temporary plots were collected following a random sampling throughout the company and the permanent plots established and monitored until 2006, distributed in the company's two silvicultural units (Guane and Mantua), and six
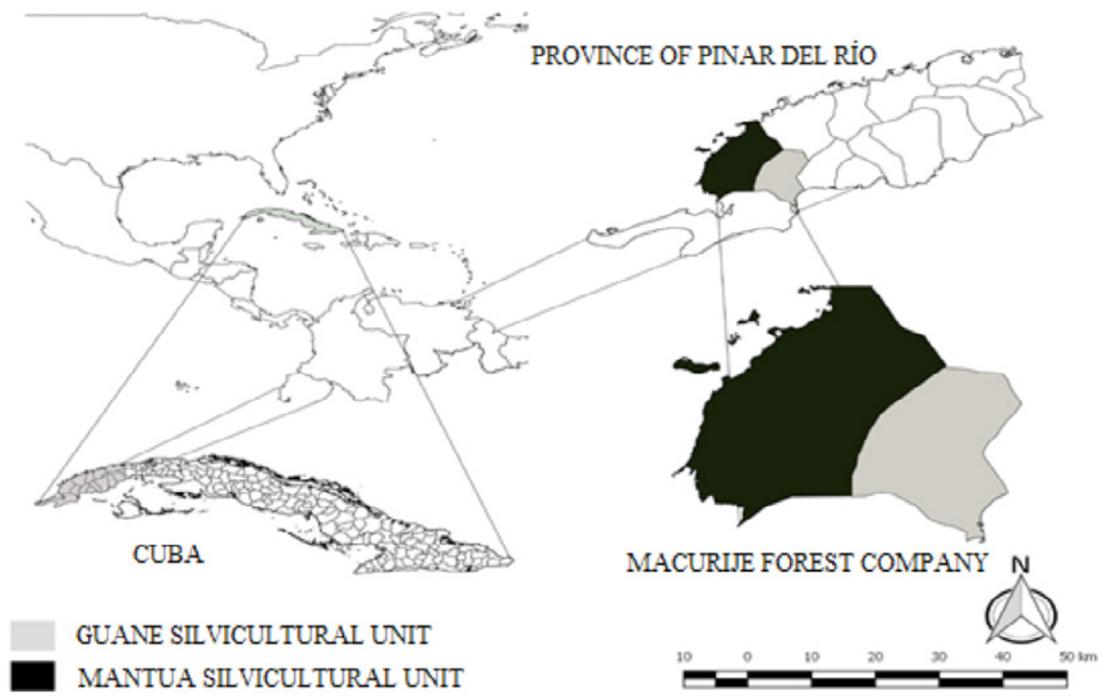
**Figure 1.** Geographic location of Macurije forest company, Pinar del Río, Cuba.

consecutive measurements were made. In the plots, variables age (A), diameter at breast height (DBH) and total height (Ht) were measured, and the yields represented by the variables basal area (G) and volume (V) were calculated.

Sample sufficiency analysis was performed using sampling error, based on the random sampling procedure in an infinite population, with an acceptable error of 10% and a 95% probability level.

### 2.3. Growth and yield models fitted for plantations of Pinus caribaea var. caribaea

The selected growth and yield models (Table 1) were fitted for complete settlement and the one with the best data adherence was adjusted by site class.

For yield prognosis, the models of Clutter (1963) (Equations 1 and 2) and Buckman (1962) modified by Silva et al. (2006) (BMS) (Equations 3 and 4) were fitted.

**Table 1.** Growth models fitted for *Pinus caribaea* var. *caribaea*, Pinar del Río, Cuba.

| № | Authors | Mathematical expressions |
|---|---------|--------------------------|
| 1 | Schumacher (1939) | $Y = e^{\beta_0 + \beta_1\left(\frac{1}{A}\right)} + \varepsilon$ |
| 2 | Korf (1939) | $Y = \beta_0 e^{-\beta_1 A^{-\beta_2}} + \varepsilon$ |
| 3 | Chapman-Richards (Chapman, 1961; Richards, 1959) | $Y = \beta_0 \left(1 - e^{-\beta_1 A}\right)^{\beta_2} + \varepsilon$ |
| 4 | Logistic (Verhulst, 1838) | $Y = \dfrac{\beta_0}{1 + \beta_1 e^{-\beta_2 A}} + \varepsilon$ |
| 5 | Silva-Bailey (Silva, 1986) | $Y = \beta_0 e^{\beta_1 \beta_2^{A}} + \varepsilon$ |

$Y$: volume (m³/ha); $A$: Age (years); $\beta_0$; $\beta_1$; $\beta_2$: parameters to be estimated; $\varepsilon$: random error.

$$LnY_2 = \beta_0 + \beta_1 / A_2 + \beta_2 S + \beta_3 LnG_2 + Ln\varepsilon \qquad (1)$$

$$LnG_2 = LnG_1\left(\frac{A_1}{A_2}\right) + \alpha_0\left(1 - \frac{A_1}{A_2}\right) +$$

$$\alpha_1\left(1 - \frac{A_1}{A_2}\right)S + Ln\varepsilon \qquad (2)$$

$$LnY_2 = \beta_0 + \beta_1 A_2^{-1} + \beta_2 S_1 + \beta_3 LnG_2 + Ln\varepsilon \qquad (3)$$

$$LndG_2 = \beta_4 + \beta_5 S_1 + \beta_6 A_2^{-1} + \beta_7 G_1 + \varepsilon \qquad (4)$$

Where: $Y_2$: expected volume at age $A_2$; $A_1$: current age; $A_2$: future age; $G_1$: current basal area; $G_2$: future basal area; $S_1$: site index; $dG_2$ = increase in basal area from age $A_1$ to age $A_2$; $\beta_i$, $\alpha_i$: parameters to be estimated; $\varepsilon$: random error ~ NID $(0, \sigma^2)$.

## 2.4. Artificial neural networks (ANNs) training for yield prediction and prognosis

There were 100 ANNs of multilayer perceptron (MLP) and radial basis function (RBF) type trained for both growth prediction and yield prognosis and the two-best retained for analysis. The variables and training algorithm used, as well as the activation functions tested, are found in Table 2.

The use of categorical variables is one of the main advantages of ANNs (Martins et al., 2016), dummy variables such as site classes (S) and forest production basic units (FPBUs) (Los Ocujes, Las Cañas, Sábalo, Río Mantua, Macurije) were included in the input set of the ANNs for both growth prediction and yield prognosis. The site classes considered were: SI = (10-13); SII = (13-16); SIII = (16-19); SIV = (19-22) and SV = (22-25).

The dataset was divided into three parts: 50% for training, 25% for test and 25% for cross-validation. The variables were normalized by linear transformation at intervals [0, 1] or [-1, 1] depending on the activation function (Table 2).

## 2.5. Parameters estimation and models (regression and ANNs) selection criteria

The adjustments of the regression models as well as the ANNs training were performed with the application software Statistica 8.0 and SPSS 20.0. The linear models were fitted using the ordinary least squares method (OLS) and nonlinear models with the Levenberg-Marquardt, Gauss-Newton, or Newton-Raphson iterative methods. The prognosis models were fitted with the two-stage least squares method (2SLS) since they were exactly-identified simultaneous equation systems.

The quality of the adjustments was evaluated using the following criteria: adjusted coefficient of determination ($R^2aj$); standard error of estimation (Syx); root mean square error (RMSE) and residuals distribution analyses to verify possible estimation trends in the equations obtained. The assumptions of normality, homoscedasticity and serial autocorrelation of the residuals were also verified by the Kolmogorov-Smirnov, White and Durbin-Watson tests, respectively.

In cases of violation of the first two assumptions, logarithmic transformation was applied. For models that underwent such a transformation, it was necessary to correct the logarithmic discrepancy with the Meyer correction factor as well as recalculate the residual standard error. The problem of the serial autocorrelation of residuals was addressed by the Cochrane-Orcutt method (Cochrane & Orcutt, 1949).

The validation of regression equations and trained ANNs was performed by comparing their estimates with the observed values. The univariate comparisons were performed using the statistical procedure proposed by

**Table 2.** Characteristics of ANNs training for yield prediction and prognosis.

| | Dependent variable (output) | Independent variables (inputs) | Types of ANN | Algorithm | Activation functions |
|---|---|---|---|---|---|
| Growth prediction | V | A, S, FPBU | Multilayer perceptron (MLP) Radial basis function (RBF) | Broyden-Fletcher-Goldfarb-Shanno (BFGS) | Sine, identity, logistic, exponential, hyperbolic tangent |
| Yield prognosis | $V_2$ | $A_1$, $A_2$, S, $G_1$, $G_2$, $V_1$, FPBU | | | |
| | $G_2$ | $A_1$, $A_2$, S, $G_1$, FPBU | | | |

V: estimated volume; $V_1$: current volume; $V_2$: expected volume at age $A_2$; S: site classes; FPBU: forest production basic units; A: age; $A_1$: current age; $A_2$: future age; $G_1$: current basal area; $G_2$: future basal area.

Leite & Oliveira (2002), testing the hypothesis $H_0$: the observed values are equal to the values estimated by the regression equations or the ANNs. This procedure combines Graybill's $F (H_0)$ test, the $t$-test for mean error ( ) and the linear correlation ($r$) between the observed and estimated values.

In order to validate the models (regression equations and ANNs) adjusted for the simultaneous prognosis of production variables (basal area and volume), multivariate comparisons between the observed values and those estimated by the models were performed through the Hotelling $T^2$ test, using the procedure proposed by Balci and Sargent (1982).

## 3. RESULTS AND DISCUSSION

### 3.1. Estimates of the parameters of growth and yield models

The sampling error of 2.19%, corresponding to a pilot sample of 550 plots, was less than the allowable error of 10%, which indicated that this was enough to make the volume estimates with the required precision.

Table 3 shows the estimates of the parameters of each model. All equations resulting from the adjustments indicate rotation ages between 30 and 35 years for the species in the company. The consonance of the rotation ages with those found by Barrero et al. (2011) indicates consistency of the parameter estimates obtained. These results and the high coefficients of determination and smaller standard error of the estimates (Table 3) favored the selection of the Schumacher and Korf equations as the most

adequate for growth prediction in *Pinus caribaea* var. *caribaea* plantations at Macurije forest company.

The Kolmogorov-Smirnov tests indicated that only the residuals of the Schumacher, Logistic and Silva-Bailey models followed a normal distribution ($p$-value > 0.05), a necessary condition for the results of the $t$ and $F$ parametric tests used to test the significance of the models and their respective parameters to be reliable.

The results of the Durbin-Watson test indicated that only the Schumacher model showed uncorrelated residuals. The Chapman-Richards, Silva-Bailey, and Logistic models presented negative serial auto-correlation and Korf's a positive auto-correlation.

The White test results ($p$-value > 0.05), confirmed by the residuals distributions (Figure 2), indicated that only the Schumacher and Korf models met the homoscedasticity assumption. The periodic or sinusoidal distribution of the logistic model residuals indicates its inadequacy for the data. This latter model and Chapman-Richards's model showed a tendency to overestimate smaller volumes.

Site index inclusion in the Schumacher (1939) model for volume prediction by productive capacity generated inconsistent results, opting then for its adjustment by site class. These adjustments allowed for relative control of the site variation source, with good adjustments despite the reduction of sample size per site (Table 4).

The assumption of normality was only observed in the residuals of the last three sites ($p$-value > 0.05), so logarithmic transformation was performed, which was effective in solving the problem. The results of the

**Table 3.** Parameters estimates for growth prediction models fitted for *P. caribaea* plantations.

| Models | Coefficients | | | $R^2_{aj}$ (%) | $S_{yx}$ (%) | Sig. F | TRA |
|---|---|---|---|---|---|---|---|
| | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | | | | |
| Schumacher | 6.92* | −33.57* | - | 98.76 | 1.96 | < 0.0001 | 33.57 |
| Korf | 1055.35* | 31.02* | 0.96* | 98.80 | 1.94 | < 0.0001 | 33.86 |
| Chapman-Richards | 579.05* | 0.06* | 3.37* | 98.70 | 2.09 | < 0.0001 | 33.02 |
| Logistic | 473.09* | 25.43* | 0.14* | 98.20 | 3.57 | < 0.0001 | 32.73 |
| Silva-Bailey | 513.24* | −5.71* | 0.92* | 98.60 | 2.52 | < 0.0001 | 32.21 |

* Significant estimate at 99% confidence by t-test; TRA: technical rotation age.
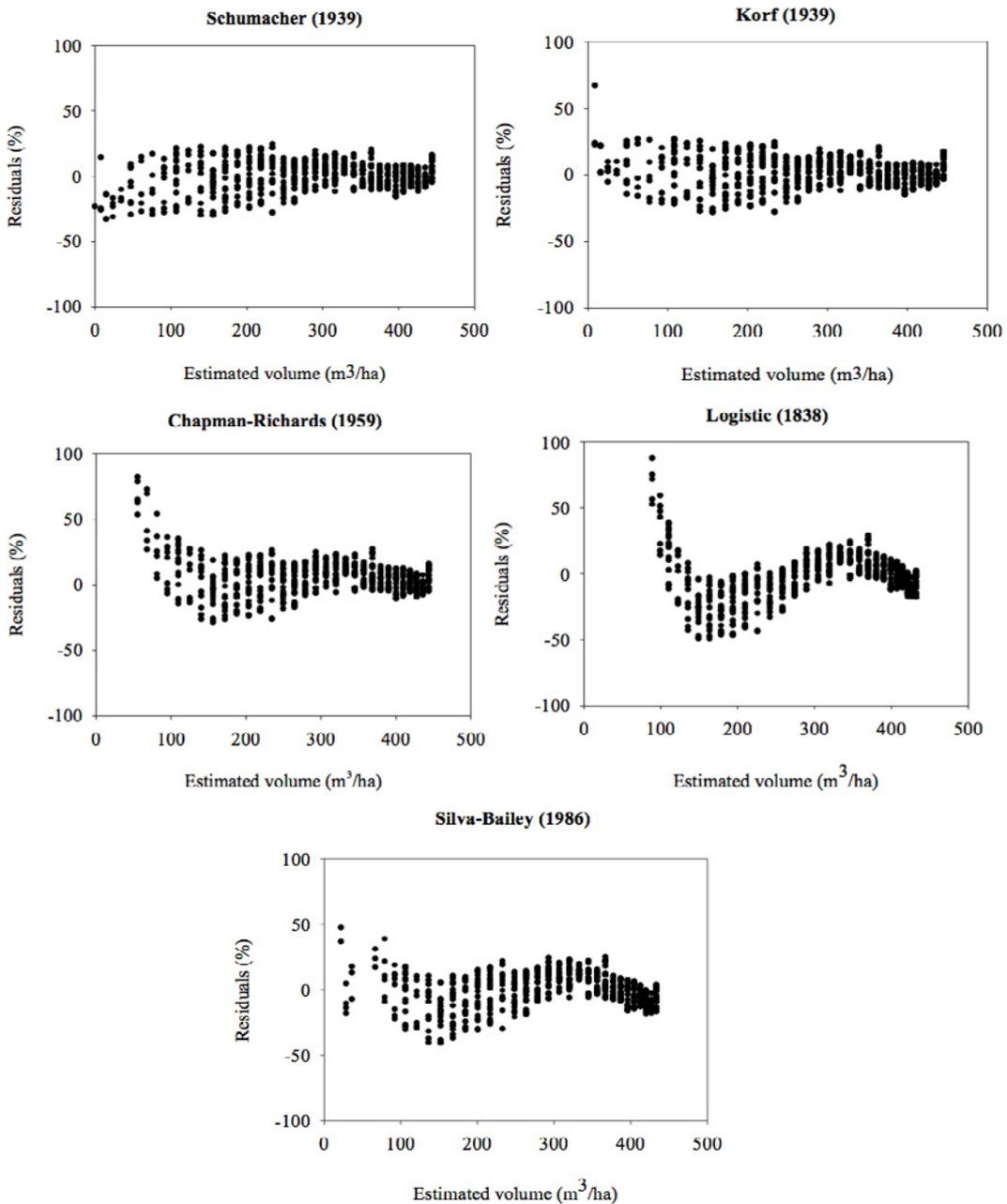
**Figure 2.** Residuals distribution of growth models fitted for *P. caribaea* var. *caribaea*.

Durbin-Watson test indicated the existence of positive serial autocorrelation in the residuals of all models. The application of the Cochrane & Orcutt (1949) procedure has eliminated the problem from the equations that presented good precision and biological consistency (Table 4). The results of the White test (*p*-value > 0.05) indicated compliance with the assumption of homoscedasticity in all equations.

The Schumacher equation indicated a yield of 375.73 m³/ha, corresponding to an MAI (mean annual increment) of 11.05 m³/ha/year. In the estimates obtained from Schumacher equations by site class (Table 4), it is possible to observe that in the case of biological consistency, a reduction of the opposite of the coefficient $\beta_1$ (rotation age) with increase in site quality and tendency to increase productivity in the

same direction occurs. In this sense, MAIs of 6.37 m³/ha/year, 10.96 m³/ha/year, 12.01 m³/ha/year, 12.65 m³/ha/year and 13.21 m³/ha/year were recorded for sites V, IV, III, II and I, respectively.

With the exception of site V, whose productivity was low and similar to that reported by Aldana et al. (2006) for the species in the company's planning (6.50 m³/ha/year), and site I, whose productivity was above 13 m³/ha/year, the MAIs are consonant with the results of Barrero et al. (2011), who found MAIs between 10 m³/ha/year and 12 m³/ha/year. TRAs indicated by the obtained equations (Tables 3 and 4) also correspond to the TRAs between 30 and 35 years found by these authors.

## 3.2. Equations for growth and yield prognosis in *Pinus caribaea* var. *caribaea* plantations

In the Clutter equations (Table 5), the negative signal of the parameter $\beta_1$ estimate indicates the consistency of the volume estimates. On the other hand, the same negative signal in the estimate of parameter $\alpha_1$ ($\alpha_1 = -0.091$), in the basal area projection equation,

indicates that the effect of the site index (S) on the basal area was inconsistent (Table 5). In this case, Campos & Leite (2017) recommend that the S in the term $(1-A_1/A_2)$ S be replaced by $LnG_1$, $(LnG_1)^2$ or $Hd_1$.

The aforementioned substitution did not generate any statistical contribution, so we opted to eliminate this term as recommended by the authors mentioned above and adopted by Dias et al. (2005). The basal area prognosis equation was then reduced in the form presented in Equation 5.

$$LnG_2 = LnG_1\left(\frac{A_1}{A_2}\right) + 3.923\left(1 - \frac{A_1}{A_2}\right);$$

$$\left(R^2 = 95.55\%; RMSE = 1.06\%\right)$$

(5)

Ln: natural logarithm; $A_1$: current age; $A_2$: future age; $G_1$: current basal area; $G_2$: future basal area; RMSE: root mean square error; $R^2$: coefficient of determination.

The minimal changes between the $R^2$ values (from 96.20% to 95.55%) and RMSE (from 0.97% to 1.06%) of both forms of the model indicated that the exclusion of the term did not lead to statistical loss for the initial equation. Thus, the residual distribution of this reduced equation (Figure 3) presented the same problems of the

**Table 4.** Estimates of Schumacher model parameters fitted by site class.

| S | $\hat{\beta}_0^*$ | $\hat{\beta}_1^*$ | $\hat{\beta}_0 = \dfrac{\beta_0^*}{1-\rho}$ | $\hat{\beta}_1 = \hat{\beta}_1^*$ | $R_{aj}^2 (\%)$ | $R_{YX} (\%)$ | Sig. F | TRA |
|---|---|---|---|---|---|---|---|---|
| I | 0.09* | −30.97* | 6.99* | −30.97* | 98.70 | 2.91 | < 0.0001 | 30.97 |
| II | 0.09* | −32.28* | 6.99* | −32.28* | 98.33 | 3.52 | < 0.0001 | 32.28 |
| III | 0.19* | −32.98* | 6.98* | −32.98* | 93.95 | 5.35 | < 0.0001 | 32.98 |
| IV | 0.21* | −34.08* | 6.92* | −34.08* | 98.67 | 3.03 | < 0.0001 | 34.08 |
| V | 0.60* | −34.96* | 6.41* | −34.96* | 96.97 | 4.00 | < 0.0001 | 34.96 |

* Significant estimate at 99% probability; ρ: estimation of autocorrelation; TRA: technical rotation age.

**Table 5.** Parameters estimates for basal area and volume prognosis models.

| Models | PV. | Estimates of parameters | | | | $R^2 (\%)$ | RMSE (%) | Sig. F |
|---|---|---|---|---|---|---|---|---|
| | | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | | | |
| Clutter (1963) | $LnY_2$ | 0.831* | −31.752* | 0.031* | 1.329* | 97.45 | 0.14 | < 0.0001 |
| | $LnG_2$ | $\alpha_0$ | $\alpha_1$ | | | 96.20 | 0.97 | < 0.0001 |
| | | 5.712* | −0.091* | | | | | |
| Buckman modified by Silva et al. (2006) | $LnY_2$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $R^2 (\%)$ | RMSE (%) | Sig. F |
| | | 1.181* | −27.352* | 0.004* | 1.291* | 98.97 | 0.08 | < 0.0001 |
| | $LndG_2$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | 61.65 | 1.99 | < 0.0001 |
| | | −1.496* | 0.061* | −3.859* | 1.045* | | | |

* Significant estimate to 99% confidence; PV.: projected variable; RMSE: root mean square error; Ln: natural logarithm; $Y_2$: expected volume at future age $A_2$; $G_2$: future basal area; $dG_2$: increase in basal area from age $A_1$ to age $A_2$.

initial equation: an overestimation of the lower basal areas and an underestimation of the larger ones, coinciding with the trends observed by Castro et al. (2013).

Regarding the Buckman model modified by Silva (2006) (BMS), the estimates of the parameters related to the variables site index ($S_1$) and basal area ($G_1$) were positive and those related to the reverse of age ($1/A_2$) were negative. This indicates biological consistency of the estimates since the signs of these coefficients assure

that both basal area and volume increase when there is improvement in productive capacity (site index) and/or increase in age (Figure 4).

For comparisons, BMS equations were higher than those of Clutter (1963). Such superiority is evident in the volume projection equations by criteria values such as $R^2$ (98.97 for Buckman versus 97.45 for Clutter), RMSE (0.08 against 0.14), and a non-biased residual distribution for the Buckman model (Figure 3).
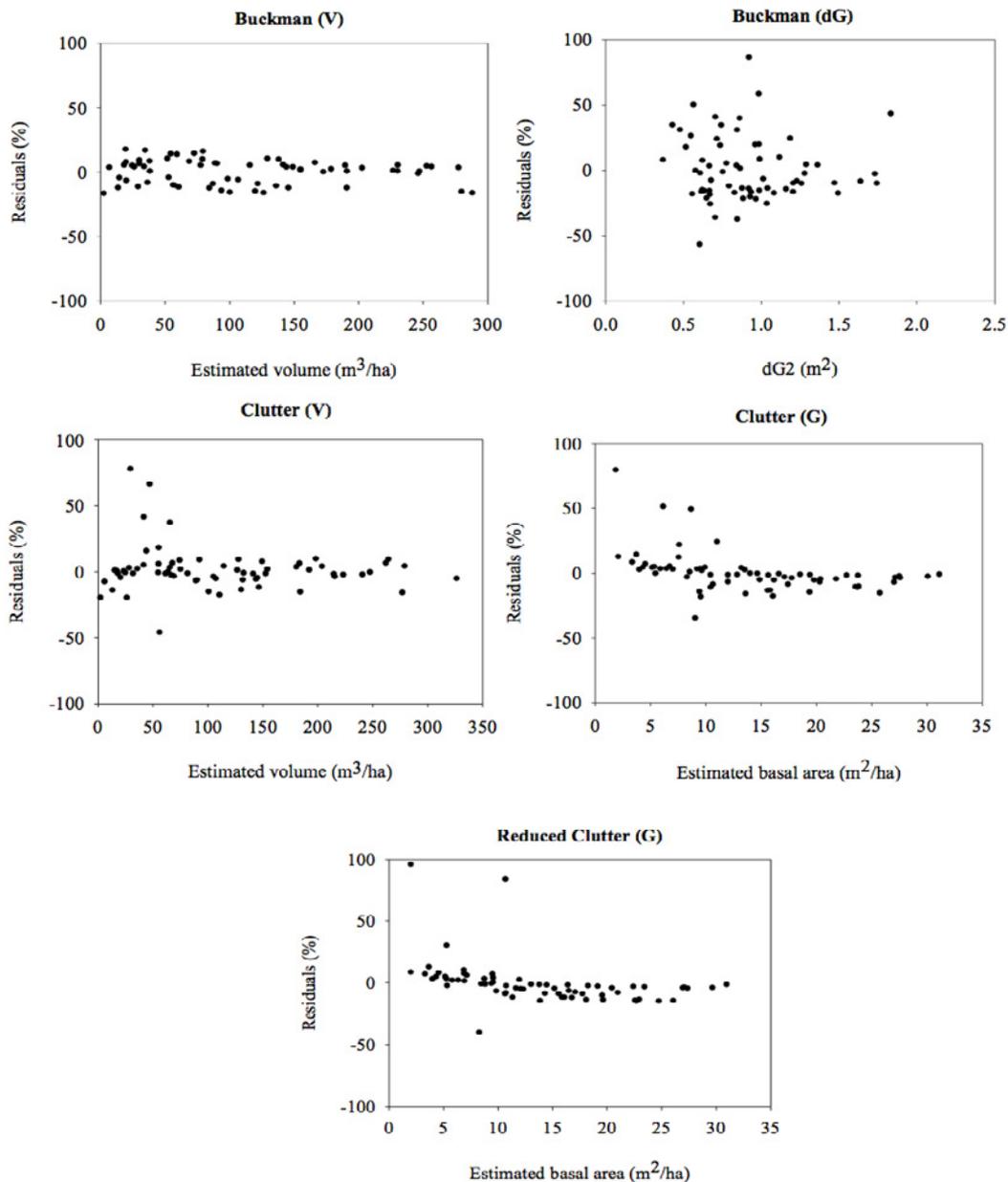


**Figure 3.** Residuals distribution of production prognosis models for *P. caribaea* var. *caribaea*.

Regarding the basal area projection, although the Clutter (1963) model presented higher statistical indicators (Table 5), the tendency to overestimate the smaller basal areas and to underestimate the larger ones is evident as previously pointed out. This tendency in basal area estimates had a marked influence on the volume prognosis whose accuracy was lower in this model.

Concerning the BMS system, the prognoses obtained with the equation of increments in basal areas were not biased (Figure 3).

Other aspects in favor of the Buckman system were the assumptions. The results of the Kolmogorov-Smirnov test indicated that the Buckman system equations satisfied the normality assumption ($p$-value > 0.05) and consequently the results of $F$ and $t$-tests of this model are reliable (Table 5). This is not the case with Clutter's equations, in which this assumption was not met. Regarding the Durbin-Watson test, the results indicate that only the residuals of the Buckman system are relatively free of autocorrelation. Except for the Clutter volume equation, all other equations satisfied the assumption of homoscedasticity, according to the White test results ($p$-value > 0.05).

Simulations of prognoses with Buckman system equations allowed to check their biological realism and the consistency of the estimates obtained (Figure 4). They were observed in these prognoses for rotation ages between 30 and 35 years; yields varying between $V_2$ = 160.439 m³/ha ($G_2$ = 22.46 m²/ha) for site V and $V_2$ = 356.280 m³/ha ($G_2$ = 42.81 m²/ha) for site I (Figure 4), thus indicating a proportionality between production, site, and age. These results are consistent with those of Francis (1992) who reported basal areas between 20 and 60 m²/ha for the species.

### 3.3. Artificial neural networks for yield prognosis for P. caribaea var. caribaea

The results of ANNs training indicated that the neural networks of Multilayer Perceptron (MLP) type with the number of neurons in the hidden layer varying between 5 and 11 were the most efficient in both prediction and prognosis of *Pinus caribaea* var. *caribaea* production in Macurije Forest Company. With respect to volume prediction, inclusion of categorical variables allowed to obtain ANN_P1 with precise and consistent estimates (Table 6 and Figure 5) characterized by yields proportional to site qualities. The technical rotational ages generated by this ANN (Figure 5) were similar to those found with the Schumacher model fitted by site class (Table 4).

The ANNs also provided satisfactory results in prognoses of basal area and volume. Inclusion of dummy variables also improved the generalization capacity of ANNs both in basal area and volume prognosis (Table 6).

Leite & Oliveira (2002) test results (Table 7) indicated that there is no significant difference between the volumes observed and those estimated by the two approaches (ANNs and regression equations). This satisfactory result, evidenced by the excellent values of the ANNs evaluation criteria (Table 6) and the regression models (Table 5), together with the individual (Figures 3 and 6) and comparative (Figure 7) residual
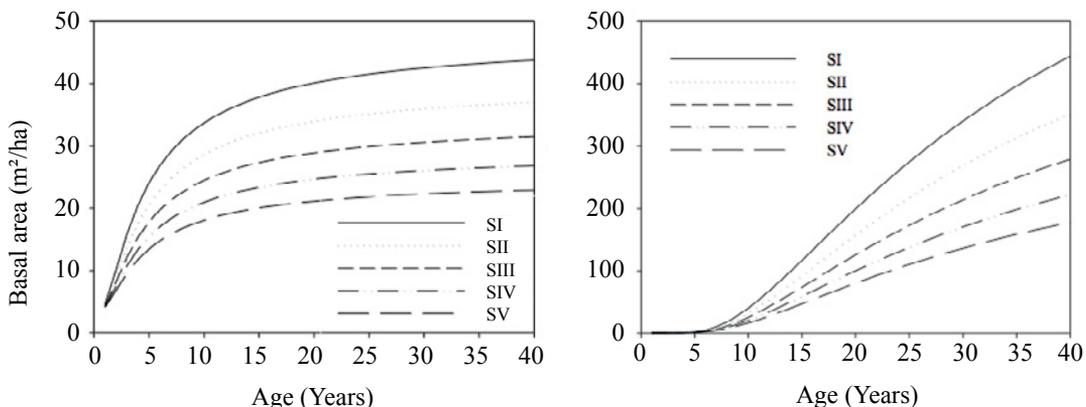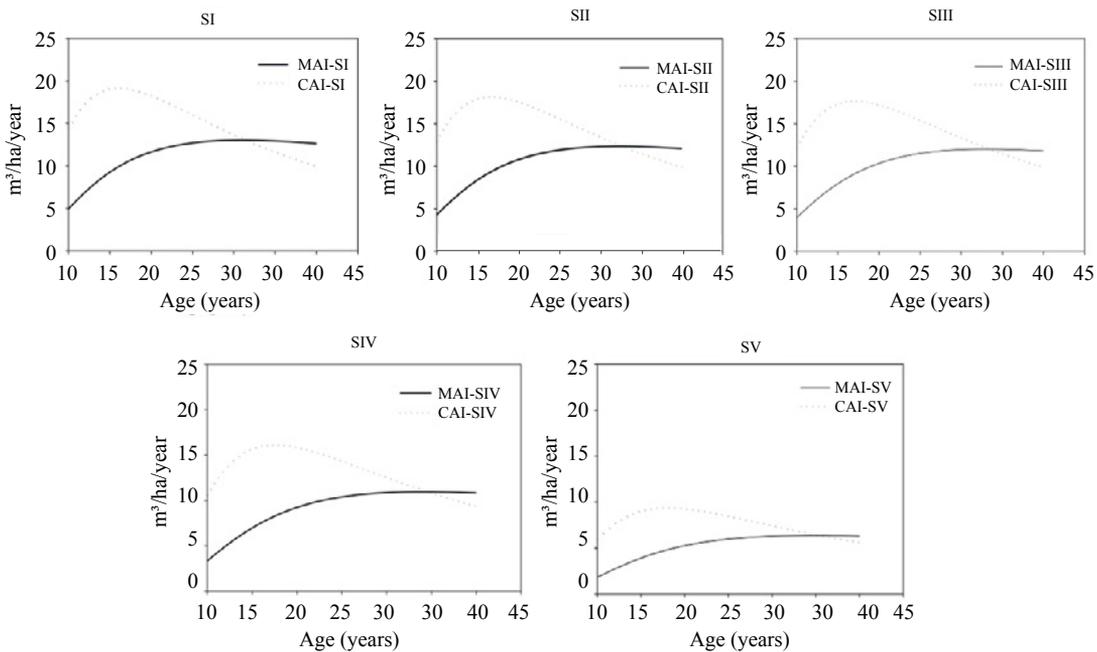


**Figure 4.** Projection of basal area and volume by site class for *P. caribaea* var. *caribaea*.

**Figure 5.** Mean and current annual increments determined by ANN_P1 (MLP 11-8-1).

**Table 6.** ANNs training results for growth prediction and yield prognosis for *Pinus caribaea*.

| Finality | IV | PV | ANNs | Architecture | Activation Functions Hidden | Output | RMSE (%) | R² (%) |
|---|---|---|---|---|---|---|---|---|
| *Volume Prediction* | *A, S, FPBU* | $V$ | ANN_P1 | MLP 11-8-1 | Tanh | Logistic | 0.04 | 97.96 |
| | | | ANN_P2 | MLP 11-7-1 | Tanh | Tanh | 0.05 | 98.96 |
| *Production prognosis* | $A_1, A_2, S, G_1$ | $G_2$ | ANN_1 | MLP 8-9-1 | Tanh | Exponential | 0.57 | 99.21 |
| | | | ANN_2 | MLP 8-9-1 | Identity | Logistic | 0.99 | 99.14 |
| | $A_1, A_2, S, G_1, G_2, V_1$ | $V_2$ | ANN_3 | MLP 10-8-1 | Tanh | Tanh | 1.59 | 98.10 |
| | | | ANN_4 | MLP 10-11-1 | Identity | Tanh | 1.66 | 98.66 |
| | $A_1, A_2, S, G_1, FPBU$ | $G_2$ | ANN_5 | MLP 13-6-1 | Exponential | Logistic | 0.56 | 99.28 |
| | | | ANN_6 | MLP 13-8-1 | Exponential | Logistic | 0.60 | 99.16 |
| | $A_1, A_2, S, G_1, G_2, V_1, FPBU$ | $V_2$ | ANN_7 | MLP 15-9-1 | Exponential | Tanh | 0.42 | 98.78 |
| | | | ANN_8 | MLP 15-5-1 | Logistic | Exponential | 1.10 | 99.68 |

*PV: estimated or projected variable; IV: independent variables.*

distributions, indicated similar performance between both approaches in volume prognosis.

Regarding the basal area prognosis, the results of applying the statistical procedure by Leite & Oliveira (2002) indicated the existence of discrepancy only between the basal areas observed and those estimated by the Buckman equation (Table 7).

In the multivariate comparison, based on both basal area and volume prognoses, the non-significance of the

Hotelling's $T^2$ test ($T^2 = 0.52$; F = 0.26[ns]) between ANN estimates and observed values indicates that there is no difference between them. However, the values estimated by the BMS system differed significantly from those observed ($T^2 = 32.59$, F = 16.17[*]). This difference is likely related to the low performance of this system in basal area prognosis, according to the univariate comparisons.

These results are indicative of the superiority of ANNs in production prognosis and are in agreement
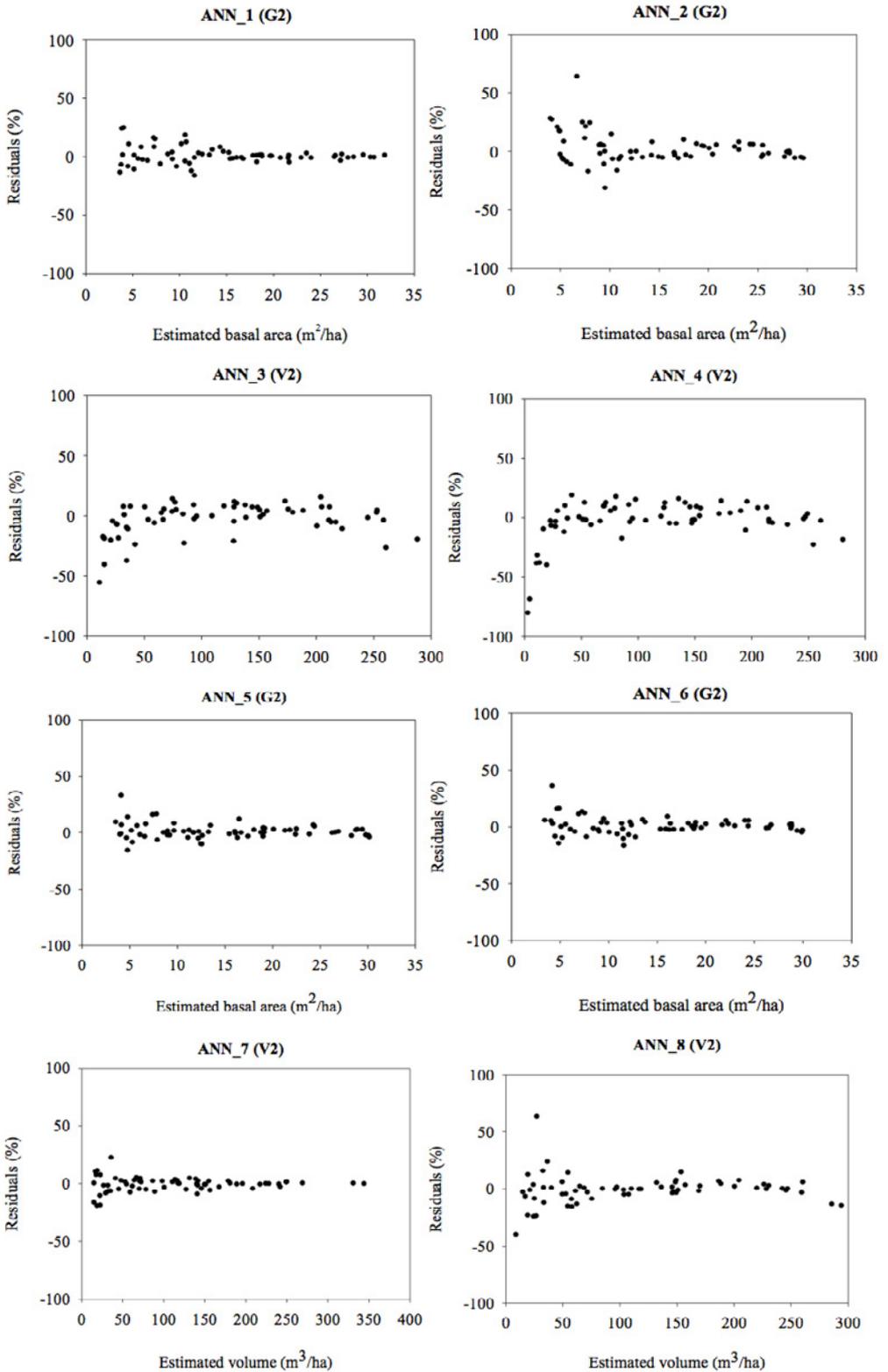
**Figure 6.** Residual distribution of ANNs trained for growth prediction and yield prognosis.
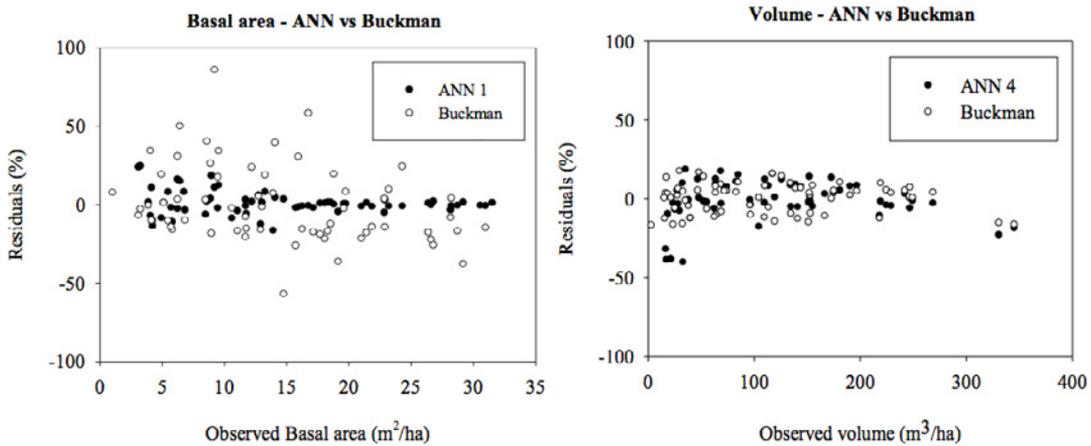
**Figure 7.** Residual distribution of ANNs and Buckman's system modified by Silva (2006).

with Porras (2007) and Ashraf et al. (2013) whose results also pointed to the superiority of ANNs. This superiority can be attributed to exclusive characteristics of ANNs such as fault tolerance, the parallelism of its structure and its greater parsimony in comparison to traditional regression models.

## 4. CONCLUSIONS

The best growth prediction equation for *Pinus caribaea* var. *caribaea* plantations was the one obtained through fitting of the Schumacher model.

The flexibility of ANNs allowed for the inclusion of categorical variables (site index and FPBU) that enabled more accurate predictions, without losing the biological realism of the models and consequently the consistency of the estimates.

In production prognosis, the Buckman model modified by Silva et al. (2006) was higher than the Clutter (1963) model. In volume prognosis, ANNs and Buckman model modified by Silva et al. (2006) performed similarly. This was not the case in basal area prognosis during which ANNs generated more accurate estimates than those of Buckman's equation.

## ACKNOWLEDGEMENTS

**Table 7.** Results obtained by applying the procedure proposed by Leite & Oliveira (2002).

| Variables | Models | $F (H_O)$ | t (e) | r | Mean error (e) | $r_{y_j y_i} \geq 1 - |\bar{e}|$ | Conclusion |
|---|---|---|---|---|---|---|---|
| | BUCKMAN | $1.85^{ns}$ | $1.46^{ns}$ | 0.98 | -1.84 | Yes | $Y_i = Y_j$ |
| $V_2$ | MLP 10-8-1 | $1.77^{ns}$ | $0.97^{ns}$ | 0.98 | -1.34 | Yes | $Y_i = Y_j$ |
| | MLP 15-9-1 | $1.53^{ns}$ | $0.87^{ns}$ | 0.99 | -0.73 | Yes | $Y_i = Y_j$ |
| | BUCKMAN | $1.34^{ns}$ | $2.98^{*}$ | 0.94 | -0.03 | No | $Y_i \neq Y_j$ |
| $G_2$ | MLP 8-9-1 | $1.17^{ns}$ | $1.54^{ns}$ | 0.99 | 0.03 | Yes | $Y_i = Y_j$ |
| | MLP 13-6-1 | $1.03^{ns}$ | $1.24^{ns}$ | 0.99 | 0.02 | Yes | $Y_i = Y_j$ |

* Significant test at 95% probability; ns: non-significant test at 95% probability.

Pinar del Río (Cuba) and the Macurije forestry company (Pinar del Rio/Cuba) for facilitating their databases and making their areas available for the study.

## CORRESPONDENCE TO

**Ouorou Ganni Mariel Guera**

Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros, s/n, CEP 52171-900, Recife, PE, Brasil
e-mail: gueraforest@gmail.com

## REFERENCES

Aldana E, Puentes M, Romero JL. *Proyecto de ordenación de la EFI Macurije*. La Habana: Ministerio de la Agricultura; 2006.

Ashraf MI, Zhao Z, Bourque CPA, Maclean DA, Meng FR. Integrating biophysical controls in forest growth and yield predictions with artificial intelligence technology. *Canadian Journal of Forest Research* 2013; 43(12): 1162-1171. 10.1139/cjfr-2013-0090

Balci O, Sargent RG. Validation of multivariate response models using Hotelling's two-sample $T^2$ test. *Simulation* 1982; 39(6): 185-192. 10.1177/003754978203900602

Barrero MH, Peraza EO, Álvarez LD, Guera M. Determinación del turno de corta para *Pinus caribaea* var. *caribaea* en la Empresa Forestal Integral "Macurije". *Floresta e Ambiente* 2011; 18(1): 109-116. 10.4322/floram.2011.028

Binoti MLMS, Leite HG, Binoti DHB, Gleriani JM. Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. *Cerne* 2015; 21(1): 97-105. 10.1590/01047760201521011153

Buckman RE. *Growth and yield of red pine in Minnesota*. Washington, DC: U.S. Department of Agriculture; 1962.

Burkhart HE, Tomé M. *Modeling forest trees and stands*. Dordrecht: Springer; 2012.

Campos JCC, Leite HG. *Mensuração florestal: perguntas e respostas*. 5. ed. Viçosa: UFV; 2017.

Castro RVO, Soares CPB, Martins FB, Leite HG. Crescimento e produção de plantios comerciais de eucalipto estimados por duas categorias de modelos. *Pesquisa Agropecuária Brasileira* 2013; 48(3): 287-295. 10.1590/S0100-204X2013000300007

Chapman DG. Statistical problems in dynamics of exploited fisheries populations. In: *4th Berkeley Symposium on Mathematical Statistics and Probability*; 1961; Berkeley. Berkeley: University of California Press; 1961. p. 153-168.

Clutter JL. Compatible growth and yield models for lobolly pine. *Forest Science* 1963; 9(3): 354-371. 10.1093/forestscience/9.3.354

Cochrane D, Orcutt GH. Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association* 1949; 44(245): 32-61. 10.2307/2280349

Dias AN, Leite HG, Campos JCC, Couto L, Carvalho AF. Emprego de um modelo de crescimento e produção em povoamentos desbastados de eucalipto. *Revista Árvore* 2005; 29(5): 731-739. 10.1590/S0100-67622005000500008

Draper NR, Smith H. *Applied regression analysis*. 3rd. ed. New York: John Wiley & Sons; 1998.

Francis JK. *Pinus caribaea Morelet. Caribbean pine. Pinaceae. Pine family. SO-ITF-SM-53*. New Orleans: USDA Forest Service; 1992.

Korf V. A mathematical definition of stand volume growth law. *Lesnicka Prace* 1939; 18: 337-379.

Leite HG, Oliveira FHT. Statistical procedure to test identity between analytical methods. *Communications in Soil Science and Plant Analysis* 2002; 33(7-8): 1105-1118. 10.1081/CSS-120003875

Martins ER, Binoti MLMS, Leite HG, Binoti DHB, Dutra GC. Configuração de redes neurais artificiais para estimação do afilamento do fuste de árvores de eucalipto. *Revista Brasileira de Ciências Agrárias* 2016; 11 (1): 33-38. 10.5039/agraria.v11i1a5354

Özçelik R, Diamantopoulou MJ, Brooks JR. The use of tree crown variables in over-bark diameter and volume prediction models. *iForest* 2014; 7: 132-139. 10.3832/ifor0878-007

Özçelik R, Diamantopoulo MJ, Eker M, Gürlevık N. Artificial neural network models: an alternative approach for reliable aboveground pine tree biomass prediction. *Forest Science* 2017; 63(3): 291-302. 10.5849/FS-16-006

Pardoe I. *Applied regression modeling*. 2nd. ed. New Jersey: John Wiley & Sons; 2012.

Porras JC. Growth evaluation of a conifer forest (*Pinus Cooperí* Blanco) using a neural net backpropagation trained with distance independent competition measures. *Computación y Sistemas* 2007 [cited 2019 July 2]; 10(4): 415-427. Available from: https://bit.ly/2KThDkT

Prodan M, Peters R, Cox F, Real P. *Mensura forestal*. 1st. ed. San José: IICA; 1997.

Richards FJ. A flexible growth function for empirical use. *Journal of Experimental Botany* 1959; 10(2): 290-300. 10.1093/jxb/10.2.290

Schumacher FX. A new growth curve and its applications to timber-yield studies. *Journal of Forestry* 1939; 37: 819-820.

Silva AL, Campos JCC, Leite HG, Souza AL, Lopes PF. Growth and yield prediction using the modified Buckman model. *Revista Árvore* 2006; 30(5): 787-793. 10.1590/S0100-67622006000500012

Silva JAA. *Dynamics of stand structure in fertilized slash pine plantations* [dissertation]. Athens: University of Georgia; 1986.

Valença M. *Fundamentos das redes neurais: exemplos em Java*. 2. ed. Olinda: Livro Rápido; 2010.

Verhulst PF. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance Mathématique et Physique* 1838; 10: 113-121.