SCIENTIA FORESTALIS

ISSN 2318-1222 (online)

ORIGINAL ARTICLE

Biomass modeling in a mixed plantation of *Pinus taeda* L. and *Pinus elliottii* Engelm

Modelagem de biomassa em plantação mista de *Pinus taeda* L. e *Pinus elliottii* Engelm

Renata Reis de Carvalho[1], Jonathan William Trautenmüller[2], Sintia Valerio Kohler[3], Edson Luiz Serpe[1], Ana Paula Dalla Corte[4], Dimas Agostinho da Silva[4], Afonso Figueiredo Filho[1]

[1]Universidade Estadual do Centro-Oeste – UNICENTRO, Iratí, PR, Brasil
[2]Universidade Federal Rural da Amazônia – UFRA, Belém, PA, Brasil
[3]Universidade Federal Rural da Amazônia – UFRA, Parauapebas, PA, Brasil
[4]Universidade Federal do Paraná – UFPR, Curitiba, PR, Brasil

How to cite: Carvalho, R. R., Trautenmüller, J. W., Kohler, S. V., Serpe, E. L., Corte, A. P. D., Silva, D. A., & Figueiredo Filho, A. (2022). Biomass modeling in a mixed plantation of *Pinus taeda* L. and *Pinus elliottii* Engelm. *Scientia Forestalis*, *50*, e3894. https://doi.org/10.18671/scifor.v50.27

Abstract

The objective of this study was to compare three approaches to fit models to estimate forest biomass in a mixed plantation of *Pinus taeda* and *Pinus elliottii* with 16 years of age. The data came from the biomass of 60 trees sampled, 30 trees of *Pinus taeda* and 30 trees of *Pinus elliottii*. The aerial biomass was estimated through the regression analysis (independent adjustment and simultaneous adjustment) and the artificial intelligence method with the nearest neighbor techniques. The models were selected and compared based on the quality of the statistical indicators: adjusted determination coefficient (R^2_{aj}), standard error of the relative estimate ($S_{yx\%}$), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Graphical analysis of residuals (%) and the ranking of the models. In the three approaches there were differences in total and component biomass estimates, being this difference associated with the heterogeneity of species (genetic variability by seminal origin), their components, and the lowest correlation of weight with diameter and height (allometric relationships). The technique of the nearest neighbor did not present satisfactory results, its use being recommended for a larger data base. Simultaneous adjustment was similar to the independent fitting method. However, the simultaneous equation has the advantage that when adding the biomass of the components, the result is compatible with the total biomass, which is more satisfactory for the estimation of total biomass.

Keywords: Nonlinear models; Simulation modeling; Data mining.

Resumo

O objetivo deste estudo foi comparar três abordagens para ajustar modelos para estimar a biomassa aérea de uma plantação mista de *Pinus taeda* e *Pinus elliottii*. com 16 anos de idade. Os dados vieram da biomassa de 60 árvores amostradas, 30 árvores de *Pinus taeda* e 30 árvores de *Pinus elliottii*. A biomassa aérea foi estimada por meio da análise de regressão (ajuste independente e ajuste simultâneo) e do método de inteligência artificial com técnicas do vizinho mais próximo. Os modelos foram selecionados e comparados com base na qualidade dos indicadores estatísticos: coeficiente de determinação ajustado (R^2_{aj}), erro padrão da estimativa relativa ($S_{yx\%}$), Critério de Informação Akaike (AIC), Critério de Informação Bayesiano (BIC), Análise gráfica dos resíduos (%) e ordenação dos modelos. Nas três abordagens houve diferenças nas estimativas de biomassa total e componente, sendo essa diferença associada à heterogeneidade das espécies, seus componentes e a menor correlação de peso com diâmetro e altura. A técnica do vizinho mais próximo não apresentou resultados satisfatórios, sendo seu uso recomendado com uma base de dados

Financial support: None.
Conflict of interest: Nothing to declare.
Corresponding author: jwtraute@gmail.com
Received: 28 April 2022.
Accepted: 4 June 2022.
Editor: Mauro Valdir Schumacher.

maior. O ajuste simultâneo foi semelhante ao método de ajuste independente. No entanto, a equação simultânea tem a vantagem de que ao somar a biomassa dos componentes, o resultado é compatível com a biomassa total, o que é mais satisfatório para a estimativa da biomassa total.

Palavras-chave: Modelos não lineares; Modelagem de simulação; Mineração de dados.

INTRODUCTION

In the last decades, different statistical models and different methodologies have been tested in order to estimate forest biomass in a more practical and efficient way (Parresol, 2001; Chave et al., 2005; Bi et al., 2015; Magalhães, 2015; Zhao et al., 2015; Affleck & Diéguez-Aranda, 2016; Zianis et al., 2016; Djomo & Chimi, 2017; Stas et al., 2017; Oliveira et al., 2017; Behling et al., 2018; Wang et al., 2018), especially when it comes to the extensive areas of forest plantations (Sanquetta et al., 2015a; Behling et al., 2018; Péllico Netto & Behling, 2019). The regression technique has become the most used to generate estimators in the area of forest management, which can be linear or non-linear, in which the equations that estimate biomass are adjusted separately for each component of the tree.

These adjustments independently result in inconsistent estimates between the biomass components and the trees' total biomass (Parresol, 2001; Behling et al., 2018; Trautenmüller et al., 2021). To solve the problem of incompatibility between the estimates, some researchers, in the late 20th and early 21st centuries, applied the simultaneous adjustment (Parresol, 2001; Sanquetta et al., 2015a; Zhao et al., 2015).

Estimates of forest biomass with simultaneous adjustment are a consistent method with good estimates, as it is a technique that ensures the overall quality of the prediction (Sanquetta et al., 2015a; Zhao et al., 2015; Behling et al., 2018). This technique should be applied based on pre-selected models for each component, with total biomass depending on the independent variables of each component's equations (Parresol, 2001; Behling et al., 2018). Data mining is an important technique for generating forest biomass estimators.

The application of data mining with the technique of the nearest neighbor, presents an alternative way to generate biomass estimates. This technique consists in the proximity of the dimensions of the vectors, constituted by the independent variables (DBH and height) which when close tend to belong to the same class. This procedure allows results with greater simplicity, versatility and flexibility during its application (Sanquetta et al., 2013, 2015b), being an alternative method to conventional regression analysis and that does not require compliance with the regression premises.

Thus, for estimates of biomass, carbon and volume, Wojciechowski (2016) created and made the JCarbon software available for web access (www.jcarbon.ufpr.br), which has contributed to the advancement of the technique in forestry and guarantees the practicality in the adjustments and accuracy of the calculations. The new methodologies for estimating forest biomass aim to make the indirect method more accurate, agile and with reduced cost, especially for large forest areas (Weiskittel et al., 2015).

Based on all these reports, the following hypotheses were tested: i) the three procedures (nearest neighbor, independent and simultaneous adjustment) do not differ in relation to the performance of the equations, because all meet the same objective: to minimize the sum of squares of the residues, ii) only the simultaneous adjustment is biologically consistent (component estimates compatible with the total). Therefore, the objective of this research was to apply and compare three different estimator adjustment procedures, regarding the performance and compatibility of the biomass estimates of the tree components and the total biomass for a mixed plantation of *Pinus taeda* L. and *Pinus elliottii* Engelm in southern Paraná State, Brazil.

MATERIAL AND METHODS

The study area belongs to the company REMASA Reflorestadora S.A., established in the municipality of Bituruna, state of Paraná, Brazil. The area is located between the geographic coordinates south latitude 26° 07'00"and west longitude 51° 31' 00". The region's climate is humid subtropical according to the Köppen-Geiger classification, being characterized as Cfb, with a temperature between 16-18 ºC and precipitation between 1600-1900 mm (Alvares et al., 2013), these values are annual averages.

The mixed planting of *Pinus taeda* and *Pinus elliottii*. was carried out in 1998, and started with 1,111 trees ha at 3 m x 3 m spacing, with seedlings. The population underwent thinning at ages 9 and 12 years. Species play an important role in the wood industry with the different uses of their wood (log wood, floors, laminates, cellulose, among others).

The area selected for the present research had 43.5 ha and was 16 years old, with an average of 444 trees per hectare, destined for the final cut of the referred company. The census of this field was carried out, then the definition of the sample to evaluate the biomass was carried out based on the diametric distribution of the field. A total of 60 trees were sampled in the mixed planting, with 30 trees of *Pinus taeda* and 30 of *Pinus elliottii*.

In the modeling of forest biomass, three approaches were used to estimate aerial biomass in a mixed plantation: independent adjustment, simultaneous adjustment and the technique of the nearest neighbor. For this, the components were grouped as follows: needles, branch (live + dead), crown (needles+ branch) and stem (wood + bark).

For approach 1, the independent adjustment was performed separately for needles and branches and only after the selection of the non-linear models was the crown grouping applied in approaches 2 and 3 (simultaneous adjustment and the nearest neighbor technique).

In the independent adjustment, biomass was estimated using four tested nonlinear models, presented in Equation 1 to 4.

$$y = \beta_0 * d^{\beta_1} + \varepsilon_i \quad (1)$$

$$y = \beta_0 * (d^2 h)^{\beta_1} + \varepsilon_i \quad (2)$$

$$y = \beta_0 * d^{\beta_1} * \left(\frac{1}{h}\right)^{\beta_2} + \varepsilon_i \quad (3)$$

$$y = \beta_0 * \left(\frac{1}{d}\right)^{\beta_1} * \left(\frac{1}{h}\right)^{\beta_2} + \varepsilon_i \quad (4)$$

Where: y = Biomass of a given component or total biomass (in kg), d = Diameter at 1.30 m above ground (DBH) (in cm), h = Total height (in m), $\beta_0, \beta_1, \beta_2$ = Model coefficients, ε_i = Random error.

The models were selected based on the quality of the statistical indicators such as: adjusted determination coefficient - $R^2_{aj.}$ (Equation 5), standard error of the absolute estimate - S_{yx} (Equation 6), standard error of the relative estimate - $S_{yx\%}$ (Equation 7), graphic analysis of residues - % (Equation 8), Akaike Information criterion – AIC (Equation 9), Bayesian Information criterion – BIC (Equation 10).

$$R^2_{aj} = 1 - (1 - R^2) * \left(\frac{n-1}{n-p-1}\right) \quad (5)$$

$$S_{yx} = \sqrt{\frac{\Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-p}} \quad (6)$$

$$S_{yx\%} = \frac{S_{yx}}{yi} * 100 \quad (7)$$

$$Residue\% = \frac{y_i - \hat{y}_i}{y_i} * 100 \quad (8)$$

$$AIC = -2\left(\frac{-n}{2} ln\left(\frac{1}{n} SQRes\right)\right) + 2p \quad (9)$$

$$BIC = -2\left(\frac{-n}{2} ln\left(\frac{1}{n} \sum_{i-1}^{n} e_i{}^2\right)\right) + \ln(n)\, p \quad (10)$$

Where: y_i = dry mass per tree observed in the total biomass and per component (kg), $\hat{y_i}$ = dry mass per tree estimated in the total biomass and per component (kg), $\bar{y}$ = average of individual biomass observed (kg), n = number of observations, p = number of independent variables, ln = Neperian logarithm, e = associated error.

In this second approach, simultaneous adjustment was used after identifying the non-linear models with the best performance, with the Nonlinear Seemingly Unrelated Regressions (NSUR) procedure. The models selected for each component of the tree were readjusted simultaneously, using a system of equations (presented in Equation 11 to 15), this methodology was based on the study of Parresol (2001) and applied by Behling et al. (2018).

$$Y_{foliage} = f_1(X_1, \beta_1) + \varepsilon_1 \quad (11)$$

$$Y_{branchs} = f_2(X_2, \beta_2) + \varepsilon_2 \quad (12)$$

$$Y_{crown} = [f_1(X_1, \beta_1) + \varepsilon_3] + [f_2(X_2, \beta_2) + \varepsilon_2] \quad (13)$$

$$Y_{stem} = f_4(X_4, \beta_4) + \varepsilon_3 \quad (14)$$

$$Y_{total} = f_{total}\left(X_{1,}, X_2, X_3, X_4, \beta_1, \beta_2, \beta_3, \beta_4\right) + \varepsilon_{total} \quad (15)$$

Where: Y_n= biomass per tree for a given component (kg), Y_{total} = total tree biomass (kg), X_i = Independent variables (DBH and height), β_i = Model coefficients for each component.

In the third approach, biomass estimation was performed using data mining, which consists of the closest neighbor technique, with independent variables (DBH and height) and dry biomass weight (kg), obtained using the direct method. The software allows to obtain the proximity between the independent variables through different types of distances, weighting and number of neighbors, as presented and applied by Wojciechowski (2016). After calculating all possibilities of distances, weighting and number of neighbors, the best biomass estimates for each component were selected.

For the two initial procedures, the SAS software (Statistical Analysis System, 1993) was used and for the third procedure, the JCarbon software (Wojciechowski, 2016).

The results of the three procedures were compared using the above-mentioned statistical indicators. Then, the approaches were ranked using scores assigned to each statistical indicator, with the lowest score given to the best result. Thus, the technique with the lowest total sum is defined as the technique that performed best for biomass estimates.

RESULTS AND DISCUSSION

Independent adjustment

The best equations adjusted by the independent procedure were presented in Table 1, together with their performance statistics.

Table 1 - Nonlinear equations selected to estimate total above-ground biomass and by component and respective statistics of *Pinus taeda*, *Pinus elliottii* and species combination.

Components	Equation	R^2_{aj}	$S_{yx}(kg)$	$S_{yx}\%$	AIC	BIC
Species combination						
Needles	$y = 0.0729 * d^{2.1390} * \left(\frac{1}{h}\right)^{0.4861}$	0.86	14.04	50.82	479.6	493.5
Branch	$y = 0.0049 * \left(\frac{1}{d}\right)^{-2.4151} * \left(\frac{1}{h}\right)^{-0.4479}$	0.95	25.29	30.20	616.4	630.4
Stem	$y = 0.2394 * (d^2h)^{0.7187}$	0.96	62.32	23.00	654.2	666.2
Total	$y = 0.5878 * d^{1.7837} * \left(\frac{1}{h}\right)^{-0.1175}$	0.98	70.99	18.56	652.7	666.7
Pinus taeda						
Components	Equation	R^2_{aj}	$S_{yx}(kg)$	$S_{yx}\%$	AIC	BIC
Needles	$y = 0.000038 * d^{2.3605} * \left(\frac{1}{h}\right)^{-1.6913}$	0.96	6.74	24.55	193.2	205.8
Branch	$y = 0.00254 * (d^2h)^{1.0560}$	0.97	23.98	21.75	254.0	264.6
Stem	$y = 0.2348 * (d^2h)^{0.7134}$	0.98	53.10	17.44	299.8	310.3
Total	$y = 0.1046 * (d^2h)^{0.8291}$	0.98	64.67	14.62	309.9	320.4
Pinus elliottii						
Components	Equation	R^2_{aj}	$S_{yx}(kg)$	$S_{yx}\%$	AIC	BIC
Needles	$y = 0.00381 * (d^2h)^{0.9324}$	0.86	14.93	53.65	219.3	229.8
Branch	$y = 0.1546 * d^{3.1935} * \left(\frac{1}{h}\right)^{1.6877}$	0.93	21.46	38.15	247.7	260.2
Stem	$y = 0.05980 * (d^2h)^{0.8687}$	0.98	46.64	19.75	279.7	290.2
Total	$y = 0.7359 * \left(\frac{1}{d}\right)^{-2.3482} * \left(\frac{1}{h}\right)^{0.6043}$	0.99	45.00	14.05	331.6	344.1

When analyzing the non-linear models selected for two pine species for biomass estimates, it is noted that the diameter at 1.30 m aboveground (DBH) and height variables were present in all models and the statistical indicators revealed differences between the estimates of the biomass components, mainly in the needles and branches, which presented the standard error of high relative estimate. Thus, the quality of the adjustment was lower for the biomass of the needles of the branches, which may be a consequence of the growth of the two species established in the same plot; or then due to the greater variability in the biomass production of these components. Several authors report the difficulty of modeling the biomass of the tree crown components (Mello & Gonçalves, 2008; Cubas et al., 2016; Behling et al., 2018; Trautenmüller et al., 2019). Trautenmüller et al. (2019) reported that the simple correlations between variables do not represent the real effect of the independent variables on the dependent one; they also mention that for good adjustments of the crown components other variables (allometric crown variables and component biomass) should be included in the modeling.

The adjusted determination coefficients for the models selected for each component ranged from 0.86 to 0.98. Thus, according to the indicators, the best estimates were observed for the stem biomass and total biomass, as they revealed the best predictive capacities and the lowest standard error of relative estimate. The values of AIC and BIC were less for the modeling of the individual species, when compared to the modeling that combined the species (Table 1). The lower value the AIC and BIC improve the quality of the model's fit. According to Vismara (2009), even if all the models do not present good adjustments, still the AIC will indicate the best among them, in the same way as it occurs for the BIC. Therefore, to select the best results for AIC and BIC, the observed values will be compared with the other approaches used in modeling.

Simultaneous adjustment

The results of the simultaneous adjustment for the biomass of *Pinus taeda* and *Pinus elliottii* and the combination of both (*Pinus* spp.) are presented in Table 2.

Table 2 - Simultaneous adjustment for the estimation of total above-ground biomass and by component and respective statistics of *Pinus taeda*, *Pinus elliottii* and species combination.

	Species combination					
Components	Equation	R^2_{aj}	$S_{yx}(kg)$	$S_{yx}\%$	*AIC*	*BIC*
Needles	$y = 0.091377 * d^{2.162162} * \left(\frac{1}{h}\right)^{0.589991}$	0.63	15.12	54.70	315.8	329.8
Branch	$y = 0.008307 * \left(\frac{1}{d}\right)^{-2.47412} * \left(\frac{1}{h}\right)^{-0.20003}$	0.86	26.91	32.13	597.8	611.8
Crown	$y = \left(0.091377\,(d)^{2.162162} * \left(\frac{1}{h}\right)^{0.589991}\right) + \left(0.008307 * \left(\frac{1}{d}\right)^{-2.47412} * \left(\frac{1}{h}\right)^{-0.20003}\right)$	0.88	32.46	29.14	458.8	470.8
Stem	$y = 0.23937 * (d^2h)^{0.718701}$	0.88	61.95	22.86	511.0	523.0
Total	$y = \left(0.091377\,(d)^{2.162162} * \left(\frac{1}{h}\right)^{0.589991}\right) + \left(0.008307 * \left(\frac{1}{d}\right)^{-2.47412} * \left(\frac{1}{h}\right)^{-0.20003}\right) + (0.2394 * (d^2h)^{0.718701})$	0.93	70.92	18.55	476.2	488.2
	Pinus taeda					
Components	Equation	R^2_{aj}	$S_{yx}(kg)$	$S_{yx}\%$	*AIC*	*BIC*
Needles	$y = 0.000029\,(d)^{2.338523} * \left(\frac{1}{h}\right)^{-1.80661}$	0.91	6.51	23.72	103.4	114.0
Branch	$y = 0.00254\,(d^2h)^{1.056016}$	0.91	23.37	21.20	222.3	234.9
Crown	$y = \left(0.000029\,(d)^{2.338523} * \left(\frac{1}{h}\right)^{-1.80661}\right) + (0.00254\,(d^2h)^{1.056017})$	0.92	28.62	20.79	162.9	174.5
Stem	$y = 0.234758\,(d^2h)^{0.713427}$	0.89	52.06	17.09	229.7	240.3
Total	$y = \left(0.000029\,(d)^{2.338523} * \left(\frac{1}{h}\right)^{-1.80661}\right) + (0.00254\,(d^2h)^{1.056017}) + (0.234757\,(d^2h)^{0.713427})$	0.93	65.78	14.87	185.1	196.4
	Pinus elliottii					
Components	Equation	R^2_{aj}	$S_{yx}(kg)$	$S_{yx}\%$	*AIC*	*BIC*
Needles	$y = 0.003804\,(d^2h)^{0.932506}$	0.62	17.27	62.05	147.5	160.0
Branch	$y = 0.261832\,(d)^{3.239444} * \left(\frac{1}{h}\right)^{1.921762}$	0.79	25.09	44.60	217.4	151.7
Crown	$y = (0.003804\,(d^2h)^{0.932506}) + \left(0.261832\,(d)^{3.239444} * \left(\frac{1}{h}\right)^{1.921762}\right)$	0.86	29.61	35.21	182.4	155.8
Stem	$y = 0.322325\,(d)^{1.98903}$	0.94	45.76	19.37	225.1	156.6
Total	$y = (0.003804\,(d^2h)^{0.932506}) + \left(0.261832\,(d)^{3.239444} * \left(\frac{1}{h}\right)^{1.921762}\right) + (0.322325\,(d)^{1.98903})$	0.97	49.07	15.32	196.6	156.1

Simultaneous adjustment by species and their combination performed well, except for the biomass of the needles. However, for the bole and total biomass there was an improvement in the quality of the statistical indicators, as they presented the lowest value for the standard error of the relative estimate and AIC.

As for the independent adjustment, it was found that the best adjustments were obtained in the equations of total biomass, and the shaft and the adjustment with inferior quality was again verified for the crown biomass. In the same region of the present study, in *Pinus elliottii* stands, Sanquetta et al. (2015a) obtained similar results when using simultaneous modeling and corroborated the lower quality of the adjustment for the biomass of needles, branches and crown.

When we adjusted the species together, we have worse statistics for the needles than when we adjusted the species separately (Table 2). It is evident that the genetic variability between species impairs the adjustment statistics, especially if seedlings were used. Another important factor for the low precision of needle estimates may be associated with the lack of crown variables to explain their variability (Trautenmüller et al., 2019).

Canga et al. (2013) also observed in their research the lower quality of simultaneous adjustment for crown biomass. These same authors explained that the variability found in the crown, in addition to being related to the competition between species, may also be the result of the silvicultural management applied in the stand, making one species easier to adapt than the other.

In estimating *Pinus pinaster* biomass with the simultaneous adjustment technique, Viana et al. (2013) stated that the equations with the DBH and the total height of the tree were the ones which yielded the best correlations with the total biomass, using them in all the equations of the adjusted models.

Data mining - nearest neighbor technique

In the technique of the nearest neighbor, among the evaluated distances the ones that stood out were Euclidean, Chebychev and Manhathan, for *Pinus taeda*, *Pinus elliottii* and the combination of species, respectively. The technique presented the best performance with 3 closest neighbors and weighted 1 / d (Table 3).

Table 3 - Estimation of total above-ground biomass and component of *Pinus taeda*, *Pinus elliottii* and species combination by means of the nearest neighbor technique.

Species combination								
Components	Distance	Number of neighbors	Weighting	R^2_{aj}	$S_{yx}(kg)$	$S_{yx}\%$	*AIC*	*BIC*
Needles	Manhathan	3	1/d	0.78	15.34	55.50	292.40	506.77
Branch	Manhathan	3	1/d	0.96	35.07	41.87	380.07	594.43
Crown	Manhathan	3	1/d	0.96	39.53	35.49	392.77	607.13
Stem	Manhathan	3	1/d	0.93	73.00	26.94	457.77	672.14
Total	Manhathan	3	1/d	0.97	87.72	22.94	477.25	691.61
Pinus taeda								
Components	Distance	Number of neighbors	Weighting	R^2_{aj}	$S_{yx}(kg)$	$S_{yx}\%$	*AIC*	*BIC*
Needles	Euclidean	3	1/d	0.85	8.71	31.73	119.99	211.42
Branch	Euclidean	3	1/d	0.88	33.88	30.73	193.37	284.79
Crown	Euclidean	3	1/d	0.88	40.58	29.47	203.10	294.53
Stem	Euclidean	3	1/d	0.83	67.06	22.02	230.23	321.60
Total	Euclidean	3	1/d	0.87	89.35	20.21	245.73	337.15
Pinus elliottii								
Components	Distance	Number of neighbors	Weighting	R^2_{aj}	$S_{yx}(kg)$	$S_{yx}\%$	*AIC*	*BIC*
Needles	Chebychev	3	1/d	0.87	15.79	56.73	146.62	233.69
Branch	Chebychev	3	1/d	0.81	31.26	55.56	182.14	269.20
Crown	Chebychev	3	1/d	0.87	42.78	50.88	198.46	285.53
Stem	Chebychev	3	1/d	0.96	52.81	22.36	209.41	296.47
Total	Chebychev	3	1/d	0.97	56.67	17.70	213.07	300.14

Considering the statistical indicators, the nearest neighbor technique revealed slightly lower adjustments, in relation to the other two approaches used. Estimates of total biomass and stem revealed good adjustments. However, a low performance for needles, branches and crown biomass was noted.

Sanquetta et al. (2015b) used the technique of the nearest neighbor in native species planted in the Atlantic Forest biome and pointed out that the Chebychev distance, weighting 1 / d and 5 neighbors, was the best result in their estimates. These same authors, when comparing this result with the Schumacher-Hall model, found that the technique of nearest neighbor had a significant gain of accuracy with up to 16.5% reduction of the standard error of estimate.

A similar result was found by Sanquetta et al. (2013), where the data mining to the nearest neighbor technique was used to estimate the stock of stored carbon in the biomass of the species *Araucaria angustifolia*, and was found to be as relevant as independent methods. According to these authors, due to the simplicity of application, the technique of the nearest neighbor may be widely used in the forest area.

Comparison of approaches: independent adjustment, simultaneous adjustment and nearest neighbor

Based on the results of the adjustments, the three approaches were compared to identify which methodology performed best in biomass estimates. When comparing the conventional statistical indicators of the three approaches, it was found that the regression technique with the independent adjustment was better in almost all circumstances, with the simultaneous adjustment second, and finally the nearest neighbor technique.

Among all approaches, the technique of the nearest neighbor did not present good estimates and revealed relative standard errors of the estimate superior to the independent and simultaneous adjustments. In the study by Sanquetta et al. (2015b), the technique of the closest neighbor presented the best adjustments when applied to a larger data set.

For combination the two species, the standard error of relative estimate revealed the independent adjustment with the variation of 18.56 to 50.82%, followed by the simultaneous adjustment that varied from 18.55 to 54.70%. In *Pinus elliottii*, the error in the independent adjustment varied from 14.05 to 53.65%; in the simultaneous adjustment the error varied from 15.32 to 62.05%. Among the two species, the smallest variation in error was observed in *Pinus taeda*, from 14.62 to 24.55%.

A result similar to the present study was observed in the *Pinus pinaster* biomass estimate with the simultaneous adjustment, in which Viana et al. (2013) obtained for the total biomass the standard error of relative estimate of 13.7%, and the lowest predictive capacity was observed for the biomass of the needles and branches. These authors reported that the difference in biomass estimates refers to the heterogeneity of the components and the lower correlation of DBH and height. The same occurred in the simultaneous adjustment performed by Canga et al. (2013), when estimating biomass from a *Pinus radiata* stand established in Spain.

In general, from the complementary statistical indicators, there is a greater advantage in the estimates when using the simultaneous adjustment. In some cases, Akaike's information criteria revealed that the nearest neighbor technique was superior to the other approaches for *Pinus elliottii* and for the combination of the two species. According to the Bayesian information criterion, the best estimates were obtained in the simultaneous adjustment.

In Southeast Australia, Bi et al. (2015), after applying different adjustment techniques to estimate biomass in native eucalyptus forests, found that the simultaneous adjustment presented biomass equations with greater precision. In the study of Canga et al. (2013), the simultaneous adjustment, when compared with independent adjustments, showed an improvement in most adjustment statistics. With that, the same authors confirm the accuracy of the simultaneous adjustment, claiming it to be a technique which presents the best performance on the independent models of the literature.

To identify the approach with the best performance, the models were ranked according to the result of the sum of the statistical indicators of all components (Table 4).

Table 4 - Classification of complementary statistical indicators according to the approach for *Pinus taeda*, *Pinus elliottii* and species combination, at age 16 years.

Statistics	Species combination		
	Independent Adjustment	Simultaneous Adjustment	Nearest neighbor
R^2_{aj}	5	14	8
$S_{yx}\%$	6	7	14
AIC	11	10	6
BIC	9	6	12
Total	31	37	40
		Pinus taeda	
R^2_{aj}	4	9	14
$S_{yx}\%$	7	7	14
AIC	12	6	9
BIC	9	5	13
Total	32	27	50
		Pinus elliottii	
R^2_{aj}	5	13	9
$S_{yx}\%$	5	9	13
AIC	12	8	7
BIC	9	5	13
Total	31	35	42

It appears that the independent adjustment presented the best performance for *Pinus elliottii* and species combination. The simultaneous adjustment revealed the best performance for *Pinus taeda* biomass estimates. In general, the technique of the nearest neighbor did not perform well.

Mognon et al. (2014), after performing the comparison of regression using independent adjustment to the nearest neighbor technique; for estimating the total biomass of bamboo plants *Guadua* gender, they found it possible to estimate accurately the biomass by means of the two procedures. However, the authors noted that the independent adjustment with the linear regression method was more satisfactory and, they stressed that the regression technique should be adopted when the technique of the nearest neighbor included a small amount of data.

According to Faceli et al. (2011), the larger the set of data analyzed, the smaller the quadratic error, as it facilitates the classification of close neighbors classified to the same class. Likewise, Sanquetta et al. (2013) compared the biomass estimates using the technique of the nearest neighbor with the set of 50 and 180 data and found better results when using the largest number of samples. Thus, the authors stated that the adjustment can be compromised when the number of data is reduced. Thus, for this study, biomass estimates by nearest neighbor technique could have been more significant with increasing amounts of data.

The simultaneous adjustment, instead of seeking to improve the biomass equations by independent adjustment, seeks to make the results of the adjustments compatible, with the sum of the biomass estimates of the components equal to the total biomass. Thus, the results do not differentiate between themselves; it only makes the simultaneous adjustment advantageous by the compatibility in the partial estimates with the total over the independent adjustment. But they are similar techniques.

Graphical comparison of waste

The graphic distribution of residues comparing the three approaches used in the biomass estimates (kg tree^{-1}) of *Pinus taeda*, *Pinus elliottii* and species combination is shown, respectively, in Figures 1 to 3.

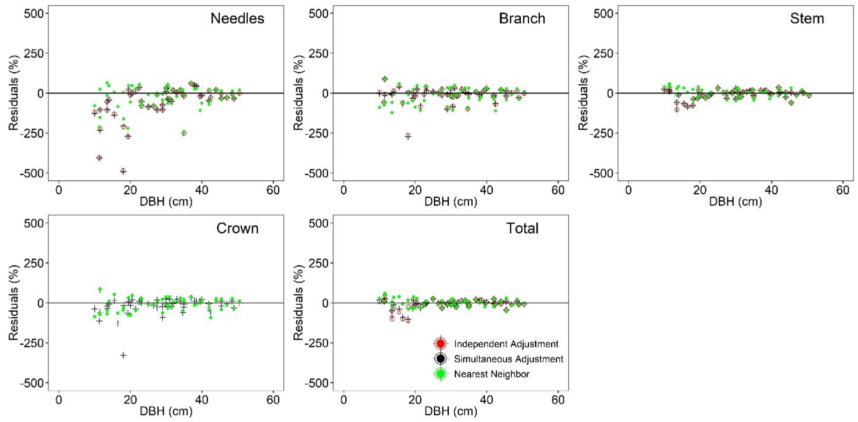

Figure 1 - Graphic comparison of residues (%) of each approach used to estimate the above-ground biomass (kg tree^{-1}) in the mixed planting of *Pinus taeda* and *Pinus elliottii*, at age 16 years, in the city of Bituruna, state of Paraná.

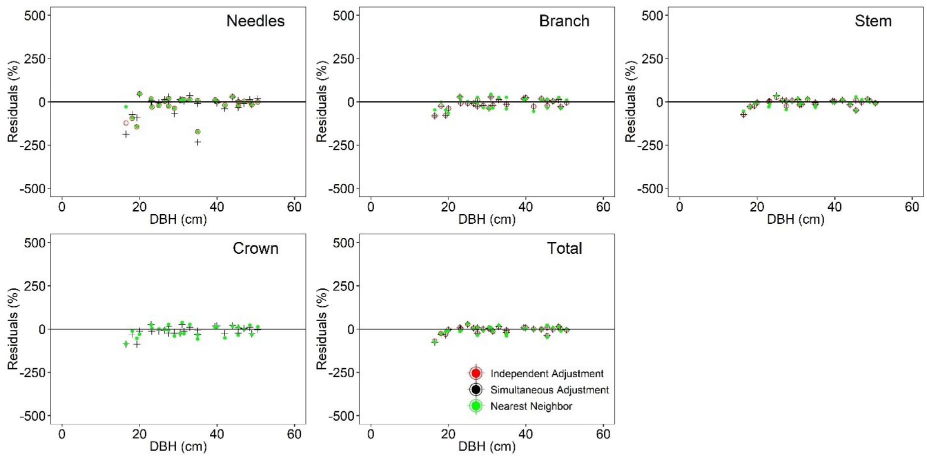

Figure 2 - Graphic comparison of the residues (%) of each approach used to estimate the above-ground biomass (kg tree^{-1}) of *Pinus taeda*, at age 16, in the city of Bituruna, state of Paraná.

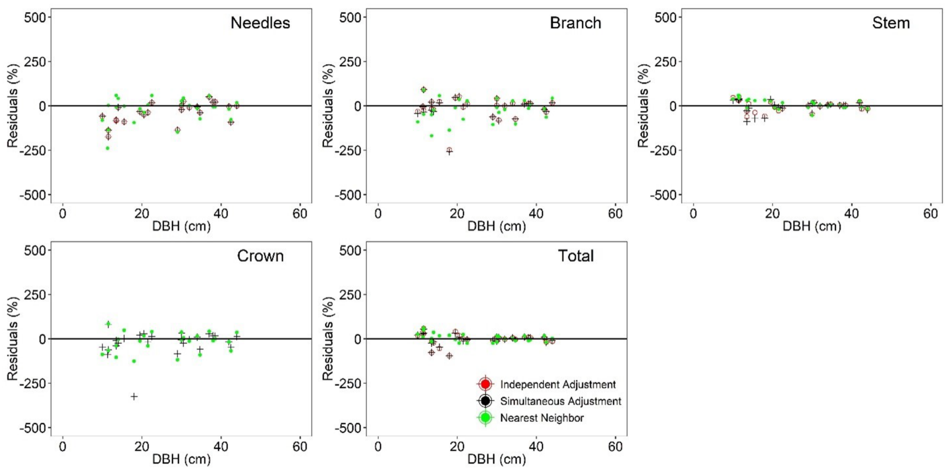

Figure 3 - Graphical comparison of the residues (%) of each approach used to estimate the above-ground biomass (kg tree^{-1}) of *Pinus elliottii* at age 16, in the city of Bituruna, state of Paraná.

In the distribution of residues it was observed that the independent and simultaneous adjustments were similar, and, in general, revealed good adjustments for total biomass and the shaft. Waste of *Pinus elliottii* and species combination presented high amplitude of the deviations, which is associated with the amplitude of variation of the canopy between the trees and the inability of the models to express the dry weight of this variable with the DBH and height. Therefore, the introduction of more variables in the models can result in increased precision; Trautenmüller et al. (2019) mention that the use of variables from canopy will improve biomass estimators.

Graphic analysis of residues also revealed a tendency to overestimate the biomass of the needles and branches. Note an improvement in the distribution of waste to the nearest neighbor technique to estimate the biomass of needles, branches and crown. However, in some cases the technique of the nearest neighbor underestimated the total biomass and the stem.

The technique of the nearest neighbor is still little explored in forest biomass estimates. In the present study, although it has presented satisfactory results, this technique could generate more accurate estimates of forest biomass, provided that a larger sample was used as mentioned by Faceli et al. (2011).

For Sanquetta et al. (2015a), the simultaneous adjustment technique is efficient and its use generates adequate biomass estimates. Furthermore, according to these authors, the main advantage of simultaneous adjustment is the possibility of compatibility between the estimates of total biomass with the biomass of the components. Thus, among the approaches used, it was found that the simultaneous adjustment showed more satisfactory levels for the estimate of total biomass in the mixed *Pinus taeda* and *Pinus elliottii* plantation studied.

CONCLUSION

Among the approaches used to estimate the aerial biomass of mixed planting, the independent adjustment performed similarly to the simultaneous adjustment, making it possible to accurately estimate the forest biomass through the two approaches. However, the simultaneous adjustment has the advantage that when adding the biomass of the components, the result is compatible with the total biomass. On the other hand, the technique of the nearest neighbor did not present satisfactory results.

In the three approaches, the best adjustments were obtained in the estimates of total biomass and stem. Among the species, the worst adjustments were for the *Pinus elliottii* biomass estimate. In the components, the worst adjustments were for the biomass estimation of needles and branches.

ACKNOWLEDGEMENTS

The authors grateful the "Fundação de Amparo a Pesquisa do Estado do Amazonas (FAPEAM)" for the scholarship to the author R. R. Carvalho.

The authors grateful the "Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA)" for the BJT scholarship to the author J. W. Trautenmüller.

REFERENCES

Affleck, D. L. R., & Diéguez-Aranda, U. (2016). Additive nonlinear biomass equations: a likelihood-based approach. *Forest Science*, *62*(2), 129-140. http://dx.doi.org/10.5849/forsci.15-126.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift (Berlin)*, *22*(6), 711-728. http://dx.doi.org/10.1127/0941-2948/2013/0507.

Behling, A., Péllico Netto, S., Sanquetta, C. R., Dalla Corte, A. P., Affleck, D. L. R., Rodrigues, A. L., & Behling, M. (2018). Critical analyses when modeling tree biomass to ensure additivity of its components. *Anais da Academia Brasileira de Ciências*, *90*(2), 1759-1774. PMid:29791562. http://dx.doi.org/10.1590/0001-3765201820170684.

Bi, H., Murphy, S., Volkova, L., Weston, C., Fairman, T., Li, Y., Law, R., Norris, J., Lei, X., & Caccamo, G. (2015). Additive biomass equations based on complete weighing of sample trees for open eucalypt forest species in South-Eastern Australia. *Forest Ecology and Management*, *349*, 106-121. http://dx.doi.org/10.1016/j.foreco.2015.03.007.

Canga, E., Diéguez-Aranda, U., Elias, A. K., & Cámara, A. (2013). Above-ground biomass equations for *Pinus radiata* D. Don in Asturias. *Forest Systems*, *22*(3), 408-415. http://dx.doi.org/10.5424/fs/2013223-04143.

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J. P., Nelson, B. W., Ogawa, H., Puig, H., Riéra, B., & Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, *145*(1), 87-99. PMid:15971085. http://dx.doi.org/10.1007/s00442-005-0100-x.

Cubas, R., Costa, E. A., Zaniz, V. (2016). Carbon contents and modelling of total organic carbon for Pinus taeda l from natural regeneration. *Árvore*, *40*(4), 661-668.

Djomo, A. N., & Chimi, C. D. (2017). Tree allometric equations for estimation of above, below and total biomass in a tropical moist forest: case study with application to remote sensing. *Forest Ecology and Management*, *391*, 184-193. http://dx.doi.org/10.1016/j.foreco.2017.02.022.

Faceli, K., Lorena, A. C., Gama, J., & Carvalho, C. (2011). *Inteligência Artificial: uma abordagem de aprendizado de máquina* (394 p.). Rio de Janeiro: LTC.

Magalhães, T. M. (2015). Allometric equations for estimating belowground biomass of *Androstachys johnsonii* Prain. *Carbon Balance and Management*, *10*(1), 16. PMid:26217395. http://dx.doi.org/10.1186/s13021-015-0027-4.

Mello, S. L. M., Gonçalves, J. L. M. (2008). Equações para estimar a biomassa da parte aérea e do sistema radicular em povoamentos de *Eucalyptus grandis* em sítios com produtividades distintas. *Árvore*, *32*(1), 101-111.

Mognon, F., Dalla Corte, A. P., Sanquetta, C. R., Barreto, T. G., & Wojciechowski, J. (2014). Estimativas de biomassa para plantas de bambu do gênero *Guadua. Ceres*, *61*(6), 900-906. http://dx.doi.org/10.1590/0034-737X201461060003.

Oliveira, N., Rodríguez-Soalleiro, R., Hernández, M. J., Cañellas, I., Sixto, H., & Pérez-Cruzado, C. (2017). Improving biomass estimation in a *Populus* short rotation coppice plantation. *Forest Ecology and Management*, *391*, 194-206. http://dx.doi.org/10.1016/j.foreco.2017.02.020.

Parresol, B. R. (2001). Additivity of nonlinear biomass equations. *Canadian Journal of Forest Research*, *31*(5), 865-878. http://dx.doi.org/10.1139/x00-202.

Péllico Netto, S., & Behling, A. (2019). Additivity of tree biomass components using ratio estimate. *Annals of the Brazilian Academy of Sciences*, *91*(3), e20180272. PMid:31166368. http://dx.doi.org/10.1590/0001-3765201920180272.

Sanquetta, C. R., Wojciechowski, J., Corte, A. P. D., Rodrigues, A. L., & Maas, G. C. B. (2013). On the use of *data mining* for estimating carbon storage in the trees. *Carbon Balance and Management*, *8*(1), 6-9. PMid:23758745. http://dx.doi.org/10.1186/1750-0680-8-6.

Sanquetta, C. R., Behling, A., Corte, A. P. D., Péllico Netto, S., Schikowski, A. B., & Amaral, M. K. (2015a). Simultaneous estimation as alternative to independent modeling of tree biomass. *Annals of Forest Science*, *72*(8), 1099-1112. http://dx.doi.org/10.1007/s13595-015-0497-2.

Sanquetta, C. R., Wojciechowski, J., Corte, A. P. D., Behling, A., Péllico Netto, S., Rodrigues, A. L., & Sanquetta, M. N. I. (2015b). Comparison of data mining and allometric model in estimation of tree biomass. *BMC Bioinformatics*, *16*(247), 247. PMid:26250142. http://dx.doi.org/10.1186/s12859-015-0662-5.

Stas, S. M., Rutishauser, E., Chave, J., Anten, N. P. R., & Laumonier, Y. (2017). Estimating the aboveground biomass in an old secondary forest on limestone in the Moluccas, Indonesia: comparing locally developed versus existing allometric models. *Forest Ecology and Management*, *389*, 27-34. http://dx.doi.org/10.1016/j.foreco.2016.12.010.

Statistical Analysis System – SAS. (1993). *Programa de computador, ambiente VM. Versão 6.08*. Cary: SAS.

Trautenmüller, J. W., Péllico Netto, S., Balbinot, R., Dalla Corte, A. P., & Borella, J. (2019). Path analysis applied to evaluation of biomass estimates in subtropical forests of Brazil. *Floresta*, *49*(3), 587-596. http://dx.doi.org/10.5380/rf.v49i3.60782.

Trautenmüller, J. W., Péllico Netto, S., Balbinot, R., Watzlawick, L. F., Dalla Corte, A. P., Sanquetta, C. R., & Behling, A. (2021). Regression estimators for aboveground biomass and its constituent parts of trees in native southern Brazilian forests. *Ecological Indicators*, *130*, 108025. http://dx.doi.org/10.1016/j.ecolind.2021.108025.

Viana, H., Marques, C. P., & Fonseca, T. (2013). Predição da biomassa aérea da *Pinus pinaster* Aiton por um sistema de equações aditivas integrado no simulador *open source modispinaster. Silva Lusitana*, *21*, 1-9.

Vismara, E. S. (2009). *Mensuração da biomassa e construção de modelos para a construção de equações de biomassa* (Dissertação de mestrado). Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. http://dx.doi.org/10.11606/D.11.2009.tde-18052009-155116.

Wang, X., Zhao, D., Liu, G., Yang, C., & Teskey, R. O. (2018). Additive tree biomass equations for *Betula platyphylla* Suk. plantations in Northeast China. *Annals of Forest Science*, *75*(2), 60. http://dx.doi.org/10.1007/s13595-018-0738-2.

Weiskittel, A. R., Macfarlane, D. W., Radtke, P. J., Affleck, D. L. R., Temesgen, H., Westfall, J. A., Woodall, C. W., & Coulston, J. W. (2015). A call to improve methods for estimating tree biomass for regional and national assessments. *Journal of Forestry*, *113*(4), 414-424. http://dx.doi.org/10.5849/jof.14-091.

Wojciechowski, J. (2016). *JCARBON: software na WEB com Data Mining para estimativas de Carbono* (1. ed., 180 p.). São Paulo: Novas Edições Acadêmicas.

Zhao, D., Kane, M., Markewitz, D., Teskey, R., & Clutter, M. (2015). Additive tree biomass equations for midrotation loblolly pine plantations. *Forest Science*, *61*(4), 613-623. http://dx.doi.org/10.5849/forsci.14-193.

Zianis, D., Spyroglou, G., Tiakas, E., & Radoglou, K. M. (2016). Bayesian and classical models to predict aboveground tree biomass allometry. *Forest Science*, *62*(3), 247-259. http://dx.doi.org/10.5849/forsci.15-045.

Author contributions: RRC: conceptualization, data curation, formal analysis, investigation, methodology, writing – original draf, writing – review & editing; JWT and APDC: formal analysis, validation, visualization, writing – review & editing; DAS and AFF: conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, validation, writing – original draf, writing – review & editing; SVK and ELS: data curation, formal analysis, investigation, methodology, visualization, writing – original draf.