Floresta e Ambiente 2019; 26(Spec No 1): e20180403
https://doi.org/10.1590/2179-8087.040318
ISSN 2179-8087 (online)

Original Article

Forest Management

Allometric Equations to Predict *Pinus palustris* Biomass in the Southeastern United States

Aline Araújo Farias[1], Salvador A. Gezan[2], Melissa Pisaroglo de Carvalho[2], Antonio Carlos Ferraz Filho[3], Carlos Pedro Boechat Soares[1]

[1]Universidade Federal de Viçosa – UFV, Viçosa/MG, Brasil
[2]University of Florida – UF, Gainesville/FL, United States of America
[3]Universidade Federal do Piaui – UFPI, Bom Jesus/PI, Brasil

ABSTRACT

Pinus palustris Mill. ecosystem is considered one of the most threatened of North America. In this context, studies on biomass quantification are fundamental for forest management plans. Thus, the objective of this study was to develop a set of allometric equations to predict total *P. palustris* stump-biomass. Biomass data were collected at different locations in the southeastern United States. A total of 119 allometric equations were fitted from the combination of explanatory variables: diameter at breast height (DBH), height (H), age (I), basal area (G), number of trees per hectare (N), site index (S) and quadratic diameter (Dq). One of the models that presented the lowest residual standard error (Sy.x) and root mean square error (RMSE) was ln(W) = -0.9978+0.7082.(H)+0.1009.ln(H.DBH)-0.5310.(N)-0.0003.ln(Dq). Therefore, the insertion of dendrometric variables characteristic of forest stands has great efficacy in biomass prediction for trees from different sites.

Keywords: forest management, modeling, regression.

1. INTRODUCTION

Longleaf pine (*Pinus palustris* Mill.) ecosystem is considered one of the most threatened in North America. The species previously occupied approximately 37 million hectares of coastal areas, from the state of Virginia to Texas through central Florida, in the southeastern United States. Current estimates suggest that only 2.2% of the original area remains in the country (Boring, 2007).

In Brazil, longleaf pine stands are established in lower proportion when compared to *Pinus ellioti* and *Pinus taeda* stands, mainly due to the growth stagnation in the shoots of trees in the early post-implantation years. However, provenance tests have shown that *Pinus palustris* has higher growth and lower stagnation of shoots, a period known as "grass stage", when planted in adequate locations such as Southern Brazil (Shimizu, 2008).

Measures of above-ground biomass are required to predict site productivity, and growth and yield of trees and stands (Gonzalez-Benecke et al., 2014b). Besides, forest biomass studies allow the quantification of nutrient cycling both for energy purposes and for the carbon credits market. These studies also provide support for the sustainable management of forest resources (Ratuchne et al., 2016).

According to Gonzalez-Benecke et al. (2014b), biomass models are usually based on the stem diameter with bark at breast height (DBH), or DBH and total tree height (H) as independent variables. However, such models are restricted to specific stands and geographic locations. An alternative to improve the accuracy of allometric models is to add stand variables such as age, density, or productivity (Gonzalez-Benecke et al., 2014a).

Few models are found in the literature to predict height, diameter inside bark and volume for longleaf pine, and they are usually based on DBH and H as independent variables (Gonzalez-Benecke et al., 2014a). Only one model to predict height based on stand-level variables such as basal area, stand age, and stand density was found, resulting in a general model (Leduc & Goelz, 2009).

The set of equations presented in this study provides conditions for a better understanding of forest biomass accumulation for one of the oldest species of the American territory, *Pinus palustris* (Boring, 2007). To our knowledge, the set of general biomass models developed in this paper will be the second carried out for the species.

Thus, the objective was to develop a set of allometric equations to predict *Pinus palustris* above-stump biomass. Also, these models can be applied to longleaf pine trees from a wide variety of ages and stands.

2. MATERIAL AND METHODS

The database used in this study was provided by the University of Florida and collected at three different locations: Fort Benning, Camp Lejeune, and Polk County.

The Fort Benning military installation (32.38°N, 84.88°W) is located between the states of Alabama and Georgia. The land use was predominated by agriculture and grazing activities before the installation of the military base. Currently, approximately 84% of the area is reforested, and of these, 15% is composed of pure longleaf pine plantations. The terrain is characterized as predominantly rolling, with elevations between 58 and 225 meters. The climate of the region is humid and mild. The mean annual precipitation measured between 1982 and 2011 in Columbus, Georgia, is 1180 millimeters, and the mean annual temperature for the same period is 18.7 °C (Samuelson et al., 2014).

The Camp Lejeune marine corps base is located in Jacksonville, North Caroline, and the climate is humid subtropical. Camp Lejeune holds ancient longleaf pine lands and their remaining habitats to support legally protected forest species and endangered animals, such as the red-cockaded woodpecker. The Lejuene base conducts studies on alternative soil preparation practices in humid sites, in partnership with the United States Forest Service (USFS), for the establishment of longleaf pine stands (Brockway et al., 2005).

Polk County (27° 57' N, 81° 42' W) is located in the central part of Florida. The prevailing climate in the region is humid subtropical, characterized by hot and humid summers and mild, relatively dry winters. The average annual temperature is around 22 °C, and the average monthly temperature varies from approximately 16 °C in January and 28 °C in August. However, from June to September, the temperature can be above 32 °C (Spechler & Kroening, 2007).

The dataset consisted of measures of 100 longleaf pine trees, obtained from three different sites (Fort Benning, Lejeune and Polk County) including ages ranging from 5 to 87 years. The dataset contained tree-level variables: DBH (diameter at 1.37 meters, cm), H (total height, m) and W (total above-stump biomass, kg), and stand-level variables: A (stand age, years), N (number of trees per hectare, ha^{-1}), BA (stand basal area, $m^2\ ha^{-1}$); Dq (quadratic mean diameter, cm) and S (site index, dominant height at reference age of 50 years, m). The main features of tree-level and stand-level variables are shown in Table 1.

The fitting of equations was performed by the ordinary least squares method (OLS) using the R *software* (R Core Team, 2013). Biomass models were created using independent variables H and DBH, as well as their modified forms: H.DBH2 (product of H by squared DBH), lnH (natural logarithm of H), lnDAP (natural logarithm of DBH), and lnH.DBH (natural logarithm of H by DBH).

The Pearson correlation coefficient (r_{xy}) was used to measure the degree of association among variables added to models. Only variables that showed positive or negative correlation were added to models analyzed in this study. The maximum number of possible combinations was performed among previously mentioned variables, which resulted in 63 models containing tree-level variables.

For generic biomass models, the following stand-level variables were added to traditional models: age (I), basal area (G), number of trees per hectare (N), site index (S), quadratic diameter (Dq), as well as their modified forms: ratio between DBH and Dq (Dp), natural logarithm of age (LnI), basal area (LnG), number of trees per hectare (LnN), site index (LnS), quadratic diameter (LnDq), ratio between DBH and Dq (LnDp), and inverse of age (I^{-1}). Since the combination of these variables would generate thousands of models for analysis, making it impossible to evaluate all of them, a correlation matrix was used to choose 56 models containing stand-level variables.

Models were analyzed by the adjusted determination coefficient (R^2) and residual standard error ($S_{y.x}$) to evaluate the goodness of fit for biomass prediction models. The variance inflation factor (VIF) was calculated to verify the presence of multicollinearity among independent variables. All models with VIF > 10 were eliminated, as proposed by O'Brien (2007).

The validation process was performed by the k-fold cross-validation technique, in which database was randomly divided into 10 blocks: 9 blocks were used to fit biomass models, and then the remaining block was used to predict biomass values. This procedure was repeated 50 times and generated prediction errors for each tree. The root mean square error (RMSE) was used to evaluate the accuracy of models.

3. RESULTS AND DISCUSSION

3.1. Model fitting

Only 12 models resulted in VIF < 10 of the total of 63 models containing H and DBH as explanatory variables. These models are represented by Equations 1-12.

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 (DBH) + \varepsilon_i \quad (1)$$

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 (H.DBH^2) + \varepsilon_i \quad (2)$$

Table 1. Summary statistics of longleaf pine dataset.

Attribute	Units	Mean	Standard deviation	Minimum	Maximum
I	years	28.4	24.2	5	87
DBH	cm	16.0	13.7	0.81	54.3
H	m	11.5	7.7	0.55	30.4
N	$n.ha^{-1}$	924	589.6	50	2150
BA	$m^2\ ha^{-1}$	14.2	8.6	0.21	30.2
Dq	cm	16.9	10.7	1.80	51.1
W	kg	262.3	492.5	0.20	2149.9
SI	m	21.27	4.08	13.74	33.86

I: stand age; DBH: diameter outside-bark at 1.37 meters; H: total tree height; N: number of trees per hectare; BA: stand basal area; Dq: quadratic mean diameter; W: total above-stump biomass; SI: site index, dominant height at reference age of 50 years.

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 \ln(H) + \varepsilon_i \quad (3)$$

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 \ln(DBH) + \varepsilon_i \quad (4)$$

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 \ln(H.DBH) + \varepsilon_i \quad (5)$$

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 (H.DBH^2) + \varepsilon_i \quad (6)$$

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 \ln(H) + \varepsilon_i \quad (7)$$

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 \ln(DBH) + \varepsilon_i \quad (8)$$

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 \ln(H.DBH) + \varepsilon_i \quad (9)$$

$$LnW = \beta_0 + \beta_1 ln(H.DBH^2) + \beta_2 \ln(H) + \varepsilon_i \quad (10)$$

$$LnW = \beta_0 + \beta_1 ln(H.DBH^2) + \beta_2 ln(DBH) + \varepsilon_i \quad (11)$$

$$LnW = \beta_0 + \beta_1 ln(H.DBH^2) + \beta_2 \ln(H.DBH) + \varepsilon_i \quad (12)$$

Some of the models selected by the VIF restriction presented high residual standard error (Table 2) and, consequently, were discarded before the cross-validation process.

The 5th and 9th models presented the best fit according to analyzed precision measurements, and all parameter estimates were significantly different from zero (P <0.001). The 5th model presented $R^2 = 97.67\%$, which means that this model can explain 97.67% of the variation in longleaf pine biomass. Similarly, the 9th model showed a high determination coefficient, indicating that 97.84% of the variation in biomass values are explained by explanatory variables DBH and ln (H.DBH). The residual standard error was similar for both models. When the models analyzed result in similar statistics, a less complex model should be selected. In this case, the 9th model was chosen over the 5th model due to its lower residual standard error.

Models containing stand-level variables as explanatory variables and VIF < 10 are represented by Equations13-26. Models that presented high residual standard error were not used in the cross-validation process (Table 3).

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 ln(H.DBH) + \beta_3 (G) + \beta_4 ln(N) + \beta_5 \ln(D_q) + \varepsilon_i \quad (13)$$

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 ln(H.DBH) + \beta_3 (G) + \beta_4 (N) + \beta_5 ln(D_q) + \varepsilon_i \quad (14)$$

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 ln(H.DBH) + \beta_3 (G) + \beta_4 (N) + \beta_5 \ln(D_q) + \varepsilon_i \quad (15)$$

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 ln(H.DBH) + \beta_3 (N) + \beta_4 ln(D_q) + \varepsilon_i \quad (16)$$

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 ln(H.DBH) + \beta_3 (N) + \beta_4 \ln(D_q) + \varepsilon_i \quad (17)$$

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 ln(H.DBH) + \beta_3 (N) + \varepsilon_i \quad (18)$$

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 ln(H.DBH) + \beta_3 \ln(G) + \varepsilon_i \quad (19)$$

Table 2. Statistics for the 12 biomass models containing tree-level variables and VIF < 10. Best models are highlighted in bold.

Model	$S_{y.x}$	R^2 Adjust.
1	0.7206	0.9155
2	0.6908	0.9223
3	0.5883	0.9437
4	0.3834	0.9761
5	**0.3783**	**0.9767**
6	0.546	0.515
7	0.4624	0.9652
8	0.456	0.9661
9	**0.3641**	**0.9784**
10	0.6999	0.9203
11	0.5266	0.9548
12	0.4417	0.9682

$S_{y.x}$ = Standard error; R^2 Adjust. = adjusted; determination coefficient; VIF = variance; inflation factor.

Table 3. Statistics for the 14 biomass models containing stand-level variables and VIF < 10. Best models are highlighted in bold.

Model	$S_{y.x}$	R^2 Adjust.
13	0.3058	0.9848
14	0.3067	0.9847
15	**0.2898**	**0.9863**
16	**0.2948**	**0.9859**
17	0.3201	0.9833
18	0.3502	0.98
19	0.3264	0.9827
20	0.3363	0.9816
21	0.3529	0.9797
22	0.3084	0.9845
23	0.3191	0.9834
24	0.348	0.9803
25	0.4009	0.9738
26	0.4859	0.9616

$S_{y.x}$ = Standard error; R^2 Adjust. = adjusted determination coefficient; VIF = variance inflation factor.

$$LnW = \beta_0 + \beta_1 (DBH) + \beta_2 ln(H.DBH) + \beta_3 \ln(D_q) + \varepsilon_i \quad (20)$$

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 ln(H.DBH) + \beta_3 (N) + \varepsilon_i \quad (21)$$

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 ln(H.DBH) + \beta_3 ln(G) + \varepsilon_i \quad (22)$$

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 ln(H.DBH) + \beta_3 ln(D_q) + \varepsilon_i \quad (23)$$

$$LnW = \beta_0 + \beta_1 (H) + \beta_2 ln(H.DBH) + \beta_3 (D_p) + \varepsilon_i \quad (24)$$

$$LnW = \beta_0 + \beta_1 ln(H.DBH) + \beta_2 (N) + \beta_3 ln(D_q) + \varepsilon_i \quad (25)$$

$$LnW = \beta_0 + \beta_1 ln(H.DBH) + \beta_2 ln(D_q) + \varepsilon_i \quad (26)$$

The 15th and 16th models were selected after evaluation of precision measurements. The determination coefficient (R^2) for the 15th and 16th models were 98.63% and 98.59% respectively, confirming the ability of models to explain the variation in biomass values. Residual standard errors for the selected models were similar. In this situation, the simplest model should be chosen over the more complex one.

Residual scatterplots against predicted biomass values for the four selected models revealed residuals uniformly distributed around zero, despite the presence of some outliers for Longleaf pine trees with low biomass content (Figure 1).

The presence of outliers for lower biomass values can be explained by the higher frequency of adult trees in stands under study.

3.2. Model validation

Table 4 shows the RMSE values for the 4 selected models after cross-validation.

RMSE was lower when explanatory variables number of trees per hectare (N), quadratic diameter (Dq), and basal area (G) were added to the 5th biomass model.

Cross-validation confirmed the goodness of fit for the 15th and 16th models to predict biomass of Longleaf pine stands located in the southeastern United States, which reinforced the residual standard error results and scatterplots of models.

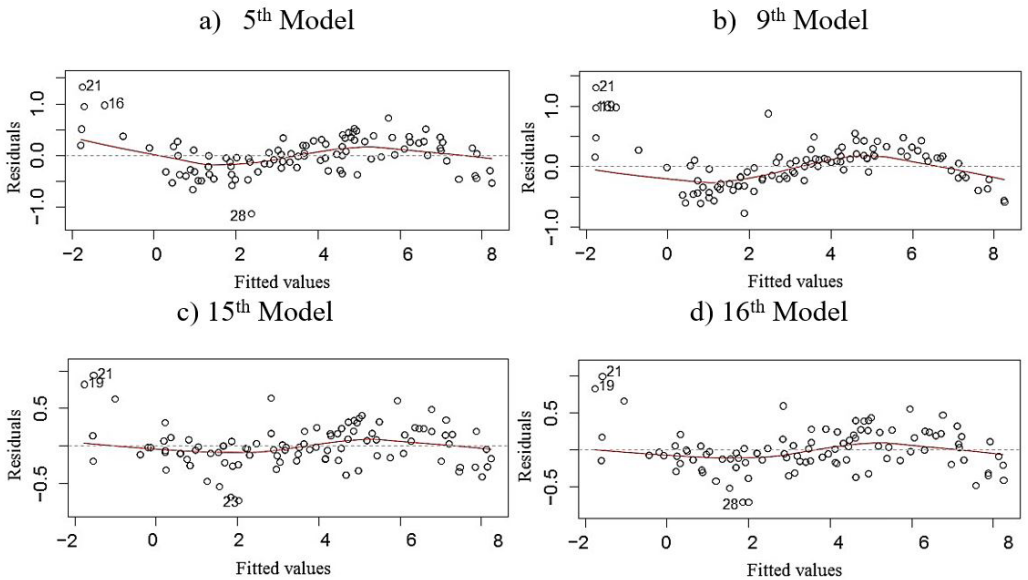

Figure 1. Residual scatterplots for the four selected models containing tree-level and stand-level variables.

Table 4. RMSE (Root Mean Square Error) comparison for the selected models after cross-validation.

Models	RMSE
5th: $ln(W) = -1.8702 + 0.1114.(H) + 0.5818.ln(H.DBH)$	0.2838
9th: $ln(W) = -1.7709 + 0.0544.(DBH) + 0.6273.ln(H.DBH)$	0.2688
15th: $ln(W) = -0.7785 + 0.7124.(H) + 0.0979.ln(H.DBH) - 0.6353.(G) - 0.0004.ln(N) + 0.0122.ln(Dq)$	0.2156
16th: $ln(W) = -0.9978 + 0.7082.(H) + 0.1009.ln(H.DBH) - 0.531.(N) - 0.0003.ln(Dq)$	0.2090

Accurate biomass equations are necessary to manage and monitor longleaf pine stands. The dataset used in this study covered different locations in the southeastern United States, and the models reported here can be applied to a wide variety of ages and stands. However, since there was no information on thinning practices for the stands, the models should be used in conditions of no strong thinning influence, as recommended by Gonzalez-Benecke et al. (2018).

Although the trees used in the modeling process have come from different locations, all stands presented similar features that allowed using multiple linear regression. Linearization of biomass models has been used in many studies (Repola, 2008; Litton & Boone, 2008; Mugasha et al., 2013; Chave et al., 2014), mainly due to its simplicity (Ferraz Filho et al., 2018). Alternatively, other approaches can be used to increase the accuracy of growth predictions of forest stands such as Artificial Neural Networks (Binoti et al., 2013; Reis et al., 2016; Ferraz Filho et al., 2018), support vector machines (Cordeiro et al., 2015; Cosenza et al., 2015) and Quantile Regression (Araújo et al., 2016). However, these methods are computationally intensive and should be used with caution due to potential restrictions on extrapolations (Ferraz Filho et al., 2018).

4. CONCLUSION

The addition of stand-level variables to linear regression models has great efficacy to predict total *Pinus palustris* trees biomass from different sites. The explanatory variables quadratic diameter (Dq), basal area (G) and number of trees.ha^{-1} (N) were crucial to capture stand variation and predict *P. palustris* biomass.

ACKNOWLEDGEMENTS

To CAPES Foundation for funding the scholarship to the first author, and to the University of Florida for providing monitoring data from different sites.

SUBMISSION STATUS

Received: 6 nov., 2018
Accepted: 4 dec., 2018

CORRESPONDENCE TO

Aline Araújo Farias
Departamento de Engenharia Florestal, Universidade Federal de Viçosa – UFV, Avenida Purdue, s/n, Edif. Reinaldo de Jesus Araújo, CEP 36570-900, Viçosa, MG, Brasil
e-mail: araujo.91@hotmail.com

FINANCIAL SUPPORT

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, (Grant/AwardNumber: 88888.041717/2013-00)

REFERENCES

Araújo CAA Jr, Soares CPB, Leite HG. Curvas de índices de local em povoamentos de eucalipto obtidas por regressão quantílica. *Pesquisa Agropecuária Brasileira* 2016; 51(6): 720-727. http://dx.doi.org/10.1590/S0100-204X2016000600003.

Binoti MLMS, Binoti DHB, Leite HG. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. *Revista Árvore* 2013; 37(4): 639-645. http://dx.doi.org/10.1590/S0100-67622013000400007.

Boring LR. The longleaf pine ecosystem: ecology, silviculture, and restoration. *Forest Science* 2007; 53(5): 618.

Brockway DG, Outcalt KW, Tomczak DJ, Johnson EE. *Restoration of longleaf pine ecosystems*. Asheville: US Department of Agriculture, Forest Service; 2005. (General Technical Report Southern Research Station; no. SRS-83).

Chave J, Réjou-Méchain M, Búrquez A, Chidumayo E, Colgan MS, Delitti WB et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology* 2014; 20(10): 3177-3190. http://dx.doi.org/10.1111/gcb.12629. PMid:24817483.

Cordeiro MA, Pereira NNJ, Binoti DHB, Binoti MLMS, Leite HG. Estimativa do volume de Acacia mangium utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. *Pesquisa Florestal Brasileira* 2015; 35(83): 255-261. http://dx.doi.org/10.4336/2015.pfb.35.83.596.

Cosenza DN, Leite HG, Marcatti GE, Binoti DHB, Alcântara AEM, Rode R. Classificação da capacidade produtiva de sítios florestais utilizando máquina de vetor de suporte e rede neural artificial. *Scientia Forestalis* 2015; 43(108): 955-963. http://dx.doi.org/10.18671/scifor.v43n108.19.

Ferraz Filho AC, Mola-Yudego B, Ribeiro A, Scolforo JRS, Loos RA, Scolforo HF. Height-diameter models for Eucalyptus sp. plantations in Brazil. *Cerne* 2018; 24(1): 9-17. http://dx.doi.org/10.1590/01047760201824012466.

Gonzalez-Benecke CA, Gezan SA, Albaugh TJ, Allen HL, Burkhart HE, Fox TR et al. Local and general above-stump biomass functions for loblolly pine and slash pine trees. *Forest Ecology and Management* 2014b; 334: 254-276. http://dx.doi.org/10.1016/j.foreco.2014.09.002.

Gonzalez-Benecke CA, Gezan SA, Martin TA, Cropper WP Jr, Samuelson LJ, Leduc DJ. Individual tree diameter, height, and volume functions for longleaf pine. *Forest Science* 2014a; 60(1): 43-56. http://dx.doi.org/10.5849/forsci.12-074.

Gonzalez-Benecke CA, Zhao D, Samuelson L, Martin T, Leduc D, Jack S. Local and general above-ground biomass functions for pinus palustris trees. *Forests* 2018; 9(6): 310. http://dx.doi.org/10.3390/f9060310.

Leduc D, Goelz J. A height–diameter curve for longleaf pine plantations in the Gulf Coastal Plain. *Southern Journal of Applied Forestry* 2009; 33(4): 164-170.

Litton CM, Boone KJ. Allometric models for predicting aboveground biomass in two widespread woody plants in Hawaii. *Biotropica* 2008; 40(3): 313-320. http://dx.doi.org/10.1111/j.1744-7429.2007.00383.x.

Mugasha WA, Eid T, Bollandsås OM, Malimbwi RE, Chamshama SAO, Zahabu E et al. Allometric models for prediction of above-and belowground biomass of trees in the miombo woodlands of Tanzania. *Forest Ecology and Management* 2013; 310: 87-101. http://dx.doi.org/10.1016/j.foreco.2013.08.003.

O' Brien RM. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity* 2007; 41(5): 673-690. http://dx.doi.org/10.1007/s11135-006-9018-6.

Ratuchne LC, Koehler HS, Watzlawick LF, Sanquetta CR, Schamne PA. Estado da arte na quantificação de biomassa em raízes de formações florestais. *Floresta e Ambiente* 2016; 23(3): 450-462. http://dx.doi.org/10.1590/2179-8087.131515.

R Core Team. *R: a language and environment for statistical computing* [online]. Vienna: The R Foundation; 2013 [cited 2018 Nov 6]. Available from: https://www.r-project.org/

Reis LP, Souza AL, Mazzei L, Reis PCM, Leite HG, Soares CPB et al. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. *Forest Ecology and Management* 2016; 382: 161-167. http://dx.doi.org/10.1016/j.foreco.2016.10.022.

Repola J. Biomass equations for birch in Finland. *Silva Fennica* 2008; 42(4): 605-624. http://dx.doi.org/10.14214/sf.236.

Samuelson LJ, Stokes TA, Butnor JR, Johnsen KH, Gonzalez-Benecke CA, Anderson P et al. Ecosystem carbon stocks in Pinus palustris forests. *Canadian Journal of Forest Research* 2014; 44(5): 476-486. http://dx.doi.org/10.1139/cjfr-2013-0446.

Shimizu JY. *Pinus na silvicultura brasileira*. 1. ed. Colombo: Embrapa Florestas; 2008.

Spechler RM, Kroening SE. *Hydrology of polk county, florida*. 1st ed. Virginia: US Department of the Interior and US Geological Survey; 2007. http://dx.doi.org/10.3133/sir20065320.