PATH ANALYSIS APPLIED TO EVALUATION OF BIOMASS ESTIMATES IN SUBTROPICAL FORESTS OF BRAZIL

Jonathan William Trautenmüller^{1*}, Sylvio Péllico Netto², Rafaelo Balbinot³, Ana Paula Dalla Corte², Juliane Borella¹

^{1*}Federal University of Paraná (UFPR), Department of Phytotechnie and Phytosanitary, Curitiba, Paraná, Brazil – jwtraute@gmail.com*, borella.juli@gmail.com

²Federal University of Paraná (UFPR), Department of Forest Science, Curitiba, Paraná, Brazil –sylviopelliconetto@gmail.com, anapaulacorte@gmail.

³Federal University of Santa Maria (UFSM), Frederico Westphalen Campus, Department of Forest Engineering, Frederico Westphalen, Rio Grande do Sul, Brazil - rafaelo.balbinot@gmail.com

Received for publication: 31/07/2018 - Accepted for publication: 14/10/2018

Resumo

Análise de trilha aplicada à avaliação de estimativas de biomassa em florestas subtropicais no Brasil. A análise de correlação entre diferentes variáveis nem sempre expressa totalmente o efeito direto real de uma variável sobre outra, especialmente em experimentos florestais. O objetivo do presente estudo foi avaliar as correlações e seus desdobramentos em efeitos diretos e indiretos do diâmetro das árvores, altura e componentes da biomassa na biomassa total acima do solo, utilizando a análise de trilha. Também objetiva definir quais variáveis devem ser incluídas na modelagem de biomassa. Os dados das árvores foram coletados em oito locais localizados no Paraná e no Rio Grande do Sul, onde o diâmetro a 1,30 m acima do solo (*dbh*); altura total (*h*); biomassa de fuste, galhos, folhas e biomassa total acima do solo foram medidas. A correlação de Spearman mostrou que *dbh* (0,93), fuste (0,94) e galhos (0,90) tiveram a maior associação com a biomassa total acima do solo quando os efeitos diretos e indiretos das variáveis foram desdobrados por análise de trilha. Os componentes fuste (0,44) e galhos (0,35) propiciaram maiores efeitos diretos quando comparados com o *dbh* (0,17). Para o componente fuste, galhos e folhas a biomassa total acima do solo apresentou os maiores efeitos diretos de 0,59, 0,74 e 0,36, respectivamente. Assim, esses resultados indicam a conveniência em se incluir pelo menos um componente de biomassa no modelo de regressão de biomassa, juntamente com *dbh* e *h*.

Palavras-chave: correlação de Spearman, componente de biomassa, efeitos diretos e indiretos, florestas nativas.

Abstract

Correlation analysis among different variables do not always fully express the real direct effect of one variable on another, especially in forest experiments. The objective of the present study was to evaluate the correlations among direct/indirect effects of tree diameter, height, and biomass components on tree total aboveground biomass and consequences when using path analysis. We also aimed to define which of the variables should be included in biomass modeling. The tree data were collected from eight sites located in Paraná and Rio Grande do Sul, where the diameter at 1.30 m aboveground (*dbh*), total height (*h*), biomass of stems, branches, leaves and tree total aboveground biomass were measured. Spearman's correlation showed that *dbh* (0.93), stems (0.94), and branches (0.90) had the highest association with total aboveground biomass when the direct and indirect effects of these variables were deconstructed by path analysis. The stem (0.44) and branch (0.35) components provided greater direct effects when compared with *dbh* (0.17). For the component stem, branches, and leaves the total aboveground biomass presented the greatest direct effects of 0.59, 0.74, and 0.36, respectively. Thus, these results indicate the convenience of including at least one biomass component in the biomass regression model, along with *dbh* and height.

Keywords: Spearman correlation, biomass component, direct and indirect effects, native forests.

INTRODUCTION

The study of correlations between variables measured in trees requires a deeper practical implementation in all areas of forestry science. In the modeling of biomass, which uses linear and non-linear regression estimators, such knowledge becomes important to identify which explanatory variables are more correlated with biomass. In this case, to find the variable that has high correlation with the variable biomass, the modeler can progress faster testing all the independent variables and their combinations to reach such aim. These studies apply estimation techniques by regression that normally uses the simple correlation between variables (ZHAO; KANE, 2017; OLIVEIRA *et al.*, 2017; ARAÚJO *et al.*, 2018).

The estimation and interpretation of the magnitude of correlations may, however, result from ambiguities in the modeling process. A high or low association may occur through the influence of a third variable or a set of variables (AZEVEDO *et al.*, 2016). For example, when we want to model the stock of biomass from a forest by means of variables that are easy to obtain, like diameters and heights of trees, the study of correlation does not indicate the relative importance of the direct and indirect effects of these variables on the dependent variable. A simple correlation identifies the magnitude and direction of the association between two or more variables without providing details of cause and effect between a set of variables in relation to the dependent variable (SALLA *et al.*, 2015).

To better understand the factors affecting the correlations between variables, Wright (1921) proposed a method called path analysis, which deconstructs the simple correlations between direct and indirect effects of independent variables on the dependent variable. This method was originally used in plants by Dewey and Lu (1959), and was subsequently applied in the context of various agricultural crops (OLIVOTO *et al.*, 2017). However, forested areas are rarely studied (RESENDE *et al.*, 2016) and no researches are found in the Brazilian literature available for modeling forest biomass. The success of path analysis provides a more detailed study of the association between two variables through the prediction of coefficients that characterize the cause and effect relationship between variables (AZEVEDO *et al.*, 2016).

In the prediction of direct and indirect effects of a set of variables on the dependent variable, it is necessary to estimate the path coefficients that are obtained by linear regression, in which the variables were previously standardized (OLIVOTO *et al.*, 2017). The estimate of coefficients can be adversely influenced by the multicollinearity effects among the variables involved. Multicollinearity occurs when the observed sample of the explanatory variables or their linear combinations present a linear relation or are correlated (GREENE, 2012). When multicollinearity occurs, the variance associated with the path coefficient estimations can reach values that are too high, making them unreliable (MONTGOMERY *et al.*, 2012). In addition, the estimates of the parameters may be unsuitable for the biological phenomenon being studied.

According to Salla *et al.* (2015), to circumvent the adverse effects of multicollinearity, two strategies can be applied: *i*) identification and elimination of the variable responsible for collinearity, and *ii*) use of ridge regression, when the exclusion of the variable is not intended. In their study, the multicollinearity between allometric and biometric variables in the production of native *Plinia cauliflora* (Mart.) Kausel was controlled by eliminating a variable, in this case it was excluded the bark percentage to solve the problem, when applying path analysis to pulp yield as a dependent variable.

The use of correlation coefficients is useful for identifying the degree of association between two variables (SALLA *et al.*, 2015). However, the correlation may not be associated with a measure of cause and effect, and its direct interpretation may result in errors during the selection of variables for the model, because the dependent and an independent variable must have a real direct relationship. When two variables present high correlation, this can be the result of the influence of one or a set of variables (LÚCIO *et al.*, 2013).

The study of correlation is important when quantifying the degree of association and the influence of the independent variables on the dependent variable (TENA *et al.*, 2016). However, the degree of such real influence of the independent variables on the dependent variable is still unknown. To address this gap in knowledge we proposed to deconstruct the direct and indirect effects of independent variables on the dependent variable using path analysis.

Our hypothesis is: "When the direct and indirect effects of variables are deconstructed through path analysis, they can show dissimilarity with the results obtained from simple correlation". Thus, the objective of this paper was to evaluate correlations and their deconstruction into direct and indirect effects by path analysis, using the variables diameter, height, biomass components (i.e. stems, branches and leaves) and the tree total aboveground biomass, and identify the variables that should be included in future forest biomass modeling studies.

MATERIAL AND METHODS

The study was conducted in Paraná and Rio Grande do Sul states of the southern region of Brazil, all located in the Atlantic Forest biome, and included a total of 387 sampled trees. The details are presented in the following:

In the Paraná state (Figure 1), the sites were located in the municipalities of Boa Ventura de São Roque (24°47'34 "S and 51°30'49" W, 775 meters above the sea level, climate Cfb, average annual temperature from 16 to 18 °C, and average annual precipitation from 1,600 to 1,900 mm); Castro (24°57'03"S and 49°51'43"W, 1,023 meters above sea level, climate Cfb, average annual temperature from 16 to 18 °C, and average annual precipitation from 1,300 to 1,600 mm); Coronel Vivida (25°57'31"S and 52°35'21"W, 630 meters above sea level, climate Cfb, average annual temperature from 16 to 18 °C, and average annual temperature from 16 to 18 °C, and average annual precipitation from 1,300 to 1,600 mm); Coronel Vivida (25°57'31"S and 52°35'21"W, 630 meters above sea level, climate Cfb, average annual temperature from 16 to 18 °C, and average annual temperature from 16 to 18 °C, and average annual temperature from 16 to 18 °C, and average annual temperature from 16 to 18 °C, and average annual temperature from 16 to 18 °C, and average annual temperature from 16 to 18 °C, and average annual temperature from 16 to 18 °C, and average annual temperature from 16 to 18 °C, and average annual precipitation between 1,900 and 2,200 mm);

General Carneiro (26°22'19"S and 51°21'32"W, 980 meters above sea level, climate Cfb, the average annual temperature from 14 to 16 °C, and average annual precipitation from 1,600 to 1,900 mm); Reserva do Iguaçú (25°49'02"S and 52°06'27"W, 920 meters above sea level, climate Cfb, average annual temperature from 14 to 16° C, and average annual precipitation from 1,600 to 1,900 mm) and São José das Palmeiras (24°51'54"S and 54°03'43"W, 430 meters above sea level, climate Cfa, average annual temperature from 20 to 22 °C, and average annual precipitation from 1,600 to 1,900 mm, Alvares *et al.*, 2013).

In the state of Rio Grande do Sul, the sites were located in the municipalities of Frederico Westphalen (27°22'03"S and 53°22'18"W, 550 meters above sea level, climate Cfa, average annual temperature from 18 to 20 °C, and average annual precipitation from 1,900 to 2,200 mm) and Iraí (27°13'05"S and 53°17'55"W, 260 meters above sea level, climate Cfa, average annual temperature from 18 to 20 °C, and average annual precipitation from 1,900 to 2,200 mm) and Iraí (27°13'05"S and 53°17'55"W, 260 meters above sea level, climate Cfa, average annual temperature from 18 to 20 °C, and average annual precipitation from 1,900 to 2,200 mm).



Figure 1. Studied sites in the subtropical forests fragments in Paraná (PR) and Rio Grande do Sul (RS), Brazil. Figura 1. Sítios estudados em fragmentos de floresta subtropical no Paraná (PR) e Rio Grande do Sul (RS), Brasil.

The sites referred come from three different phytofisionomies: i) Boa Ventura de São Roque, Castro, Coronel Vivida, Reserva do Iguaçú and General Carneiro sites, which belong to Mixed Tropical Forest, ii) São José das Palmeiras site, which belongs to Semi-decidual Tropical Forest and iii) Frederico Westphalen and Iraí sites, belonging to Decidual Tropical Forest.

In the Boa Ventura de São Roque, Castro, Coronel Vivida, Reserva do Iguaçú and São José das Palmeiras sites a fragment of around 5 ha was chosen in each site and a plot of 0.5 hectares (ha) was randomly selected inside them, which was subdivided into 50 subunits of 100 m². From the data of the forest inventory conducted in these areas, 21, 44, 30, 36 and 27 species with higher importance value were selected, respectively for the Boa Ventura de São Roque, Castro, Coronel Vivida, Reserva do Iguaçú (ZANETTE *et al.*, 2017) and São José das Palmeiras sites, and one tree of each species was sampled for biomass quantification.

In the General Carneiro (ZANETTE *et al.*, 2017), Frederico Westphalen and Iraí sites (BALBINOT *et al.*, 2017) 20, 3 and 4 plots of 144 m² were sampled, respectively. In these plots, all trees with diameter at breast height (*dbh*) greater than 5 cm were sampled. In these plots, all trees were sampled for biomass evaluation, i.e., 138, 33 and 58 trees, respectively.

The tree biomass data were collected in two distinct ways. We first defined a sampling unit of 0.5 ha at each site and subdivided this area into 50 sub-samples. In each individual sub-plot, tree species were identified

and measured to record the overall height (h) and dbh. From the results of this inventory, we have taken the species of greater importance value in the forest, and, for each one, the trees with average diameter were selected for biomass evaluation (ZANETTE *et al.*, 2017). This methodology was applied in Boa Ventura de São Roque, Castro, Coronel Vivida, Foz do Iguaçu, and São José das Palmeiras.

Biomass weight was performed for the components stem, branches and leaves. The stem component (stem wood with bark) was defined as the wood from the stump until the point of morphological inversion. For each individual, the samples were separated and weighed with a scale to obtain the wet weight of each component. The trees from the Frederico Westphalen and Iraí sites were weighted using a dynamometer scale with capacity of 500 kg. The trees from the São José das Palmeiras, Reserva do Iguaçu, Coronel Vivida, Castro, and Boa Aventura do São Roque sites were weighted using a pendulum scale with a capacity of 20 kg and 300 kg. The trees from the site General Carneiro were weighted using an analog platform scale with capacity of 500 kg. Data used in this study were collected with different scales because they came from distinct research projects.

To determine dry biomass of the collected samples, the different components were removed soon after the wet biomass was recorded, and their masses were measured using a digital scale with a precision of 0.1 grams (g). A sample in a disk form was removed at half of the stem height for obtaining the wood moisture content. Branches and leaves subsamples were collected at various points in the crown, including the lower third, middle, and upper regions to obtain their moisture content. All samples were dried in an oven with circulation and renewal of air at $65 \pm 5^{\circ}$ C until a constant weight was obtained. This collection procedure was performed at all sites.

From this data set (dbh, h), tree total aboveground biomass and biomass of components) a descriptive analysis was performed and the Kolmogorov-Smirnov test was applied to verify the goodness of fit to the normal distribution. In some cases, there was no goodness of fit, indicating the suitability of applying Spearman correlation.

The correlation matrix was constructed and the significance of the correlation coefficients was obtained with a *t*-test (p-value <0.01); $t = \hat{\rho} \sqrt{\frac{n-2}{1-\hat{\rho}^2}}$, where $\hat{\rho}$ = estimated correlation coefficient, n = number of ordination pairs, for large n ($n \ge 10$) sample sizes. This statistic presented a *t* distribution with *n*-2 degrees of freedom (SIEGEL; CASTELLAN, 2006).

The analysis of multicollinearity was performed with the correlation matrix of the variables observed for each path and the results were interpreted according to the condition number (NC) (MONTGOMERY *et al.*, 2012). NC is the ratio between the largest and the smallest eigenvalue in the matrix. Then, path analysis was conducted with the tree total aboveground biomass and its components as dependent variables, with each variable at a discrete point in time.

In each scenario, the linear correlation coefficients were deconstructed into direct and indirect effects, in that the systems of a normal equation $[\beta = (X'X)^{-1} X'Y]$ for multiple regression models were fitted by the method of ordinary least squares, as follows:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} 1 & \dots & r_{1n} \\ \vdots & 1 & r_{2n} \\ r_{n1} & r_{n2} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{12} \\ r_{13} \\ \vdots \\ r_{1n} \end{bmatrix}$$

where: r = simple correlations; n = 1, 2, ..., n variables; $\beta =$ parameters to be estimated.

By solving this system of equations, it was possible to predict the direct and indirect effects of variable X_1 over variable Y, in the following way:

$$r_{YX_1} = p_{YX_1} + p_{YX_2}r_{12} + \dots + p_{YX_n}r_{1n}$$

where:

 r_{YX_1} is the linear correlation between the variables Y and X_1 ;

 p_{YX_1} is the direct effect of variable X_I on Y;

 p_{YX_2} is the indirect effect of variable X_1 through X_2 on Y;

 $p_{YX_n}r_{1n}$ is the indirect effect of variable X_I through X_n on Y;

 r_{12} is the simple correlation of variables X_1 and X_2 ;

 r_{1n} is the simple correlation of the variables X_l and X_n .

When multicollinearity was detected, an alternative model to the ordinary least squares was used as a method of ridge regression in which the estimations are obtained with the increase of the main diagonal elements of the matrix X'X, which included a constant (k) chosen arbitrarily ($0 \le k \le 1$). The system of normal equations was obtained in the following manner, $\beta = (X'X + k)^{-1} - 1X'Y$, and is detailed in the sequence:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} 1+k & \dots & r_{1n} \\ \vdots & 1+k & r_{2n} \\ r_{n1} & r_{n2} & 1+k \end{bmatrix}^{-1} \begin{bmatrix} r_{12} \\ r_{13} \\ \vdots \\ r_{1n} \end{bmatrix}$$

where r = simple correlations; n = 1, 2, ..., n variables; $\beta = \text{parameters to be estimated}$; k = constant chosen arbitrarily $(0 \le k \le 1)$.

All tests, diagnoses, and analyses were performed using the software R (R Development Core Team, 2017) and Genes (CRUZ, 2016).

RESULTS

In the present study, all correlations between the allometric and dendrometric variables were significant (p-value <0.01) and positive (Table 1). The tree total aboveground biomass was strongly correlated with *dbh* (0.93) and biomass of the component stem (0.94) and branches (0.90). Such high correlations between the total tree biomass and its components were expected, because the greater the biomass of the stem and branches components are, greater is the tree total aboveground biomass of each tree measured. Despite the low correlation coefficient between total height and biomass of the leaves (0.49), this value was significant (p-value <0.01).

Table 1. Spearman correlation among dendrometric variables, biomass of components (stem, branches, and leaves), and tree total aboveground biomass in the subtropical forests of Brazil.

Tabela 1. Correlação de Spearman entre as variáveis dendrométricas, componentes de biomassa (fuste, galhos e folhas) e biomassa total acima do solo em floresta subtropical no Brasil.

	/			1		
	BT	dbh	h	BS	BB	BL
BT	1	0.93**	0.78**	0.94**	0.90**	0.71**
dbh		1	0.76**	0.90**	0.81**	0.63**
H			1	0.82**	0.60**	0.49**
BS				1	0.76**	0.60**
BB					1	0.71**
BL						1

**Significance at 99% probability calculated with a *t*-test, *BT* is tree total aboveground biomass, *dbh* is diameter 1.3 m aboveground, *h* is total height, *BS* is biomass of the stem, *BB* is biomass of the branches, and *BL* is biomass of leaves.

In the present study the descriptive statistics showed the great amplitude and variance of the data, such as the tree total aboveground biomass that varied within a wide range from 3.4 to 5,471.1 kg (Table 2). Figure 2 shows the relationship between all variables, and some of them are not linear.

Table 2.	Descriptive statistics of dendrometric variables, biomass of components (stem, branches and leaves) and
	tree total aboveground biomass in the subtropical forests of Brazil.

Tabela 2.	Estatística descritiva das variáveis dendrométricas, componentes de biomassa (fuste, galhos e folhas) e
	biomassa total acima do solo em florestas subtropicais no Brasil.

biomassa totar actina do solo cin norestas succiopicais no Brasil.									
Statistics	dbh	h	BS	BB	BL	BT			
Minimum	5.3	3.9	2.4	0.1	0.1	3.4			
Average	16.8	11.6	92.6	99.4	3.8	203.7			
Maximum	82.0	27.4	2,153.4	3,907.7	59.1	5,471.1			
Variance	154.9	14.29	42,754.4	123,416.9	44.2	304,523.2			
Standard Deviation	12.45	3.78	206.77	351.31	6.65	551.8			
CV (%)	79.0	33.9	269.9	350.4	209.9	299.6			
KS (statistics)	0.18	0.10	0.33	0.39	0.28	0.36			
KS (p-value)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001			

KS is the Kolmogorov-Smirnov test for normality, CV (%) is the Coefficient of Variation in %, *dbh* is the diameter at 1.3 m aboveground, *h* is the total height, *BS* is the biomass of the stem, *BB* is the biomass of the branches, *BL* is the biomass of the leaves and *BT* is the tree total aboveground biomass.



Figure 2. Relationship between variables diameter at breast height (*dbh*), total height (*h*) and biomass of the stem (BS), branches (BB), leaves (BL) and tree total aboveground biomass (BT).



The Kolmogorov-Smirnov test was applied to allometric and dendrometric data and revealed no normality for all proposed variables (Figure 3). Therefore, the correlations should be evaluated by a non-parametric method, such as the Spearman correlation, to test the degrees of significance among variables.



Figure 3. Evaluation of the normal distribution fitting of the variables diameter at breast height (A), total height (B) and biomass of the stem (C), branches (D), leaves (E) and tree total aboveground biomass (F).

Figura 3. Avaliação do ajuste de distribuição normal das variáveis diâmetro à altura do peito (A), altura total (B) e biomassa do fuste (C), galhos (D) e folhas (E) e biomassa aérea total acima do solo (F).

Table 1 shows the Spearman simple correlation values. However, these do not express the real causes and effects among the variables, as can be revealed through the deconstruction of direct and indirect effects of the independent variables on the dependent ones, in all proposed combinations presented in Table 3. With the use of tree total aboveground biomass as a dependent variable, the high coefficient of determination ($R^2 = 0.97$) and the small residual effect (0.16) indicated the effective contribution of independent variables in the model (Table 3). The variables showed weak multicollinearity (NC = 45.34) based on the method described by Montgomery *et al.* (2012).

 Tabela 3. Efeitos diretos e indiretos das variáveis dendrométricas, componentes de biomassa (fuste, galhos e folhas) e biomassa acima do solo em florestas subtropicais no Brasil.

Variable		Direct	Indirect effect					ŕ	D2	0	ŀ	
Dependent	Independent	effect	dbh	h	BS	BB	BL	BT	I	K-	e	ĸ
	dbh	0.27	-	0.16	-	0.07	-0.01	0.55	0.90	0.88		
	h	0.22	0.19	-	-	-0.05	-0.01	0.47	0.82		0.34	0.066
BC	BS	-	-	-	-	-	-	-	-			
<i>D</i> 5	BB	-0.08	0.20	0.13	-	-	-0.02	0.53	0.76			
	BL	-0.02	0.16	0.10	-	-0.06	-	0.42	0.60			
	BT	0.59	0.25	0.18	-	0.07	-0.01	-	0.94			
	dbh	0.17	-	0.07	-0.09	-	0.11	0.69	0.81			
	h	-0.10	0.13	-	-0.08	-	0.09	0.56	0.60	0.79	0.46	0.073
תת	BS	-0.10	0.15	-0.08	-	-	0.11	0.68	0.76			
DD	BB	-	-	-	-	-	-	-	-			
	BL	0.18	0.11	-0.05	-0.06	-	-	0.53	0.71			
	BT	0.74	0.16	-0.06	-0.08	-	0.14	-	0.90			
	dbh	0.05	-	-0.01	-0.03	0.29	-	0.33	0.63	0.51	0.70	0.093
	h	-0.01	0.04	-	-0.03	0.21	-	0.28	0.49			
DI	BS	-0.03	0.04	-0.01	-	0.26	-	0.34	0.60			
BL	BB	0.36	0.05	-0.01	-0.02	-	-	0.33	0.71			
	BL	-	-	-	-	-	-	-	-			
	BT	0.36	0.06	-0.01	-0.03	0.33	-	-	0.71			
	dbh	0.17	-	0.03	0.41	0.28	0.04	-	0.93			
	h	0.04	0.13	-	0.37	0.21	0.03	-	0.78			
$\mathbf{D}T$	BS	0.45	0.15	0.03	-	0.27	0.04	-	0.94	0.97 (0.16	-
В1	BB	0.36	0.14	0.02	0.33	-	0.05	-	0.90			
	BL	0.06	0.11	0.02	0.27	0.25	-	-	0.71			
	DT											

BT = tree total aboveground biomass; dbh = diameter at 1.3 m above ground level; h = total height; BS = biomass of the stem; BB = biomass of the branches; BL = biomass of the leaves; \hat{r} = Spearman correlation coefficient; R^2 = coefficient of determination; e = residual effect of the variable; k = constant included in the main diagonal of the matrix X'X; - no information for this variable.

The use of stem biomass as a dependent variable required correction of multicollinearity. The inclusion of the constant k (0.066) in the matrix X'X resulted in NC = 38.1, $R^2 = 0.88$, and residual effect = 0.34 (Table 3). The tree total aboveground biomass (0.59), dbh (0.27), and total height (0.22) variables presented the greatest direct effects on stem biomass. The direct effects of dbh and h on BS are greater than the effects of BB and BL components on BS, because dbh is taken on the stem and the stem height is part of the total height, what was already expected, and the components BB and BL present a small cause and effect relationship on BS.

When considering branches biomass as dependent variable, path analysis showed $R^2 = 0.79$, residual effect = 0.46, and NC = 36.0 with a constant k = 0.073, therefore the tree total aboveground biomass (0.74), *dbh* (0.17) and leaves component (0.18) variables presented the greater direct effects (Table 3) than the others (*h* and stem biomass) on the biomass of branches.

The biomass equation for the leaf component presented an $R^2 = 0.51$, residual effect = 0.70, and with the use of the constant *k* (0.093), the *NC* was 38.8. These results indicated that the tree total aboveground biomass

FLORESTA, Curitiba, PR, v. 49, n. 3, p. 587 - 596, jul/set 2019. Trautenmüller, J. W. *et.al.* ISSN eletrônico 1982-4688 DOI: 10.5380/rf.v49 i3.60782

Table 3.
 Direct and indirect effects of dendrometric variables on the biomass of components (stem, branches and leaves) and tree total aboveground biomass in the subtropical forests of Brazil.

(0.36) and stem component (0.36) variables provide greater efficiency in explaining the biomass of the component leaves.

DISCUSSION

Linear correlation studies are important for modeling biometric variables, especially the biomass of trees with *dbh* and *h*. When these linear correlations between independent and dependent variables are significant, they are appropriate to compose allometric models, usually fitted by means of linear or non-linear regression. Several studies found in the literature use regression to express biomass allometric relationship with *dbh* and *h* (ZHAO; KANE, 2017, ARAÚJO *et al.*, 2018), but few studies show the magnitude of these correlations between these variables (BURGER; DELITTI, 2008).

The correlations of the variables dbh and h with the total biomass of the trees can result in values equal or greater than 0.90, as found by Burger and Delitti (2008). Návar (2014) mentioned error estimates between 50 and 60% for the best adjusted equations for aboveground biomass modeling in tropical forests in Mexico, using as independent variables dbh, h, wood density and basal area. Araújo *et al.* (2018), evaluating biomass allometric equations in a restoration area of the Atlantic Forest, observed error estimates between 40 and 50% for the best equations when the data set was stratified by diameter classes and ecological groups. In this context, the present work intends to elucidate the reasons why the variables dbh and h, despite having a high correlation with the biomass, when included in the regression models do not concomitantly result in low errors for the biomass estimates.

The errors in regression models are expressed by the degree of variability among the variables that compose them. In native forests, for example, biomass variability for small trees do not behave in the same way as for large trees, which can be observed in Figure 2A. Consequently, when the regression models are fitted for the entire data set, these variations affect the magnitude of the errors.

An alternative to reduce these errors is the stratification of the data by the size of the trees. The relationship between the total biomass and *dbh* allows to identify that the variance between diameter classes increases with the *dbh* growth (Figure 2A), therefore stratification in diameter classes is appropriated to reduce errors. A similar relationship (BT and base diameter) can be seen in the article by Lima et al. (2017), in which they deal with the modeling of total biomass of trees in dry tropical forest in the state of Pernambuco, Brazil, and the maximum *dbh* is 35 cm due to the forest typology. In a study with the species *Betula platyphylla* Suk., Wang et al. (2018) found a similar relationship between wood biomass and *dbh*, considering a maximum *dbh* of 15 cm. These relationships evidenced the increase of the variance with the diameter growth, independent of the diameter dispersion.

In Path Analysis the variables showed weak multicollinearity (NC = 45.34) based on the method described by Montgomery *et al.* (2012), which states that if NC is less than 100, multicollinearity does not interfere on the estimation of the parameters (i.e. weak multicollinearity). In the other hand, if NC varies from 100 to 1,000 (multicollinearity is moderate to strong), or finally, if NC is greater than 1,000 (severe multicollinearity) the estimated parameters are affected to the extension that, when the multicollinearity tends to be perfect, the coefficients becomes undetermined and standard deviations tend to infinity (GREENE, 2012).

The Spearman correlation of tree total biomass aboveground with dbh is 0.93, but its direct effect is only 0.17, which is a small contribution to the total calculated correlation. The direct effects of stem and branch biomass components are more expressive than the direct effects of dbh and h on the tree total aboveground biomass. This evidence is biologically consistent because the sum of stem and branch biomass compose more than 90% of the tree total aboveground biomass. This result was expected because the stem and branch components are part of tree total aboveground biomass. The path analysis allowed us to detect the real associations between the studied variables and select the best ones for modeling biomass, because it because clear that we could evaluate the direct and indirect effects between all variables. In general, for the prediction of the components of a tree, the tree total aboveground biomass presented the highest direct effect, this result was also expected, because the biomass components of a tree are part of the tree total biomass.

For all models presented in Table 3, the dependent variables *BS*, *BB* and *BL* presented the highest direct affects with total biomass, and for this one, as dependent variable, the highest direct effects occurred with *BS* and *BB*. These evidences show that better correlation effects are not obtained when using the conventional concepts of direct correlation between the dependent and the independent variables. The path analysis provides the background for the deconstruction of direct and indirect effects of the correlations to achieve the best participation of the variables in the proposed models.

For modeling tree total aboveground biomass from the information of the cause and effect of all explanatory variables on the dependent variable, the stem and branch biomass components should be included along with dbh. It was observed that the biomass of leaves and the h presented the least direct influence on tree

total aboveground biomass than the other variables evaluated. For stem and branch biomass components, including tree total aboveground biomass and *dbh* could produce better estimates of these variable biomass components. In this way, estimations of biomass should include biomass components in the regression model. In this case, the modeling can be developed using two or three-stage least square regressions (GREENE, 2012).

CONCLUSIONS

- The use of simple correlation indicates that any variable can be used in modeling tree total aboveground biomass and its components, therefore, this decision can lead to an incorrect choice of independent variables without cause and effect in the biomass modeling. Biomass components of stems and branches have a direct effect on the tree total aboveground biomass that is greater in magnitude than tree diameter at breast height.
- The independent traditional variables *dbh* and *h* frequently presents high correlations with tree total aboveground biomass and with its components, but the direct effects of *dbh* and *h* on tree biomass are inexpressive, reason why path analysis is important when modeling biomass aiming to obtain better biological results on allometric equations. Variables that present small direct effect on the biomass shall not be included in the model.
- The inclusion of at least one biomass component is necessary so that the modeler can obtain satisfactory results while estimating tree total aboveground biomass. For all components, the tree total aboveground biomass should be included in the regression model.

ACKNOWLEDGEMENTS

The authors thank the financial support of the Research Foundation of Rio Grande do Sul State (FAPERGS), processes 10/0164-5 and 10/1818-5, and Capes for the Doctorate fellowship of the first author.

REFERENCES

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. M.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v.22, n.6, p.711-728, 2013.

ARAÚJO, E. J. G.; LOUREIRO, G. H.; SANQUETTA, C. R.; SANQUETTA, M. N. I.; DALLA CORTE, A. P.; PÉLLICO NETTO, S.; BEHLING, A. Allometric models to biomass in restoration areas in the Atlantic rain forest. **Floresta e Ambiente**, v.25, n.1, e20160193. 2018.

AZEVEDO, A. M.; SEUS, R.; GOMES, C. L.; FREITAS, E. M.; CANDIDO, D. M.; SILVA, D. J. H.; CARNEIRO, P. C. S. Correlações genotípicas e análise de trilha em famílias de meios-irmãos de couve de folhas. **Pesquisa Agropecuária Brasileira**, v.51, n.1, p.35-44, 2016.

BALBINOT, R.; TRAUTENMÜLLER, J. W.; CARON, B. O.; BORELLA, J.; COSTA JÚNIOR, S.; BREUNIG, F. M. Vertical distribution of aboveground biomass in a seasonal deciduous forest. **Revista Brasileira de Ciências Agrárias**, v.12, n.3, p.361-365, 2017.

BURGER, D. M.; DELITTI, W. B. C. Allometric models for estimating the phytomass of a secondary Atlantic Forest area of southeastern Brazil. **Biota Neotropica**, v.8, n.4, p.131-136, 2008.

CRUZ, C.D. Genes Software – extended and integrated with the R, Matlab and Selegen. Acta Scientiarum. Agronomy, v.38, n.4, p.547-552, 2016.

DEWEY, D. R.; LU, K. H. A Correlation and path-coefficient analysis of components of crested wheatgrass seed production. **Agronomy journal**, v.51, n.9, p.515-518, 1959.

GREENE, W. H. Econometric analysis. Seventh Edition, Pearson, 2012. 1238p.

LIMA, R. B.; ALVES JÚNIOR, F. T.; OLIVEIRA, C. P.; SILVA, J. A. A.; FERREIRA, R. L. C. Predicting of biomass in Brazilian tropical dry forest: a statistical evaluation of generic equations. Anais da Academia Brasileira de Ciências, v.89, n.3, p.1815-1828, 2017.

LÚCIO, A. D.; STORCK, L.; KRAUSE, W.; GONÇALVES, R. Q.; NIED, A. H. Relações entre os caracteres de maracujazeiro-azedo. **Ciência Rural**, v.43, n.2, p.225-232, 2013.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. Introduction to linear regression analyses. (5th edition). Wiley, 2012, 661p.

NÁVAR, J. Pan tropical biomass equations for Mexico's dry forests. Agronomía Colombiana, v.32, n.3, p.367-376. 2014.

OLIVEIRA, N.; RODRÍGUEZ-SOALLEIRO, R.; HERNÁNDEZ, M. J.; CAÑELLAS, I.; SIXTO, H.; PÉREZ-CRUZADO, C. Improving biomass estimation in a *Populus* short rotation coppice plantation. **Forest Ecology and Management**, v.391, p.194–206, 2017.

OLIVOTO, T.; SOUZA, V.Q.S.; NARDINO, M.; CARVALHO, I.R.; FERRARI, M.; PELEGRIN, A, J.; SZARESKI, V.J.; SCHIMIDT, D. Multicollinearity in Path Analysis: A Simple Method to Reduce Its Effects. Agronomy Journal, v.109, n.1, p.131-142, 2017

R Development Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing. Vienna, Austria. 2017.

RESENDE, R. T.; MARCATTI, G. E.; PINTO, D. S.; TAKAHASHI, E. K.; CRUZ, C. D.; RESENDE, M. D. V. Intra-genotypic competition of Eucalyptus clones generated by environmental heterogeneity can optimize productivity in forest stands. **Forest Ecology and Management**, v.380, p.50–58, 2016.

SALLA, V. P.; DANNER, M. A.; CITADIN, I.; SASSO, S. A. Z.; DONAZZOLO, J.; GIL, B. V. Análise de trilha em caracteres de frutos de jabuticabeira. **Pesquisa Agropecuária Brasileira**, v.50, n.3, p.218-223, 2015.

SIEGEL, S.; CASTELLAN JR, N. J. Estatística não-paramétrica: para ciências do comportamento. Porto Alegre: Artmed, 2ª Ed. 2006. 448 p.

TENA, E.; MEKBIB, F.; AYANA, A. Correlation and path coefficient analyses in sugarcane genotypes of Ethiopia. American Journal of Plant Sciences, v.7, p.1490-1497, 2016.

WANG, X.; ZHAO, D.; LIU, G.; YANG, C.; TESKEY, R. O. Additive tree biomass equations for *Betula platyphylla* Suk. plantations in Northeast China. **Annals of Forest Science**, v.75, p.60, 2018.

WRIGHT, S. Correlation and causation. Journal of Agricultural Research, v.20, p.557-585, 1921.

ZANETTE, V. H.; KURCHAIDT, S. M.; CAMARGO, L. P.; WATZLAWICK, L. F.; KOEHLER, H. S. Ajuste de modelos de regressão para a estimativa da biomassa aérea para seis regiões do estado do Paraná. **Enciclopédia Biosfera**, v.14, n.26, 29–43, 2017.

ZHAO, D.; KANE, M. New variable-top merchantable volume and weight equations derived directly from cumulative relative profiles for loblolly Pine. **Forest Science**, v.63, n.3, p.261–269, 2017.