Simultaneous estimation as an alternative to young eucalyptus aboveground biomass modeling in ecophysiological experiments

Thiago Wendling Gonçalves de Oliveira[1]*[iD], Rafael Rubilar[2], Carlos Roberto Sanquetta[1], Ana Paula Dalla Corte[1] and Alexandre Behling[1]

[1]Departamento de Ciências Florestais, Universidade Federal do Paraná, Av. Pref. Lothário Meissner, 900, 80210-170, Curitiba, Paraná, Brazil. [2]Forest Productivity Cooperative, Departamento de Silvicultura, Facultad de Ciencias Forestales, Universidad de Concepción, Concepción, Chile. *Author for correspondence. E-mail: thiago.wgdeoliveira@gmail.com

ABSTRACT. Accurate forest biomass estimates require the selection of appropriate models of individual trees. Thus, two properties are required in tree biomass modeling: (1) additivity of biomass components and (2) estimator efficiency. This study aimed to develop a system of equations to estimate young eucalyptus aboveground biomass and guarantee additivity and estimator efficiency. Aboveground eucalyptus biomass models were calibrated using four methods: generalized least squares (GLS), weighted least squares (WLS), seemingly unrelated regression (SUR), and weighted seemingly unrelated regression (WSUR). The approaches were compared with regard to performance, additivity, and estimator efficiency. The methods did not differ with regard to the mean biomass estimation; therefore, their performance was similar. The GLS and WLS approaches did not satisfy the additivity principle, as the sum of the biomass components was not equal to total biomass. However, this was not observed with the SUR and WSUR approaches. With regard to estimator efficiency, the WSUR approach resulted in narrow confidence intervals and an efficiency gain of over 20%. The WSUR approach should be used in forest biomass modeling as it resulted in effective estimators while ensuring equation additivity, thus providing an easy and accurate alternative to estimate the initial biomass of eucalyptus stands in ecophysiological models.

Keywords: nonlinear seemingly unrelated regression; weighting procedures; model error structure; biological consistency; forest biomass; eucalyptus management.

Introduction

Biomass is the most important variable when evaluating carbon dynamics, carbon sequestration, and forest productivity (Fu, Zeng, & Tang, 2017). Direct biomass measurements require destructive sampling, which is both complicated and expensive. However, alternative variables to biomass, such as root collar diameter, individual tree diameter at 10 cm height, diameter at breast height, total height, commercial height, live crown length, and age are compatible with individual or stand biomass models (Picard, Santi-André, & Henry, 2012; Wang, Zhao, Liu, Yang, & Teskey, 2018). Over the years, forest science has increasingly incorporated statistical models into individual and stand biomass estimations (Picard et al., 2012; Zhao, Kane, Markewitz, Teskey, & Clutter, 2015). Most studies have applied different models approaches; however, these models have considered fitted biomass components independently and have not taken into account inherent correlations among them (Parresol, 1999).

Biomass partitioning depends on the accuracy of total and component biomass estimations. For example, ecophysiological studies are required that evaluate the carbon balance and resource use (Binkley, Stape, & Ryan, 2004), nutrient allocation (Viera, Schumacher, Bonacina, Oliveira Ramos, & Rodriguez-Soalleiro, 2017), tree responses to planting density, water deficits (Hakamada, Hubbard, Ferraz, Stape, & Lemos, 2017), and forest products, such as pulpwood and bioenergy.

When developing allometric models, the sampled trees must be carefully selected in an efficiency and consistent manner to obtain accurate estimates of individual tree biomass. These models should meet expectations regarding estimation performance, biological consistency, and estimator efficiency. Estimation

performance refers to the quality of an estimate. Biological consistency refers to the additivity of components with regard to total biomass. Estimator efficiency refers to the consistency of selecting models that result in narrower confidence intervals (Parresol, 2001; Behling et al., 2018).

To meet the aforementioned expectations, models must be fit to estimate individual and total tree biomass components considering additive systems and restrictions (Parresol, 2001; Vonderach, Kändler, & Dormann, 2018). Zellner (1962) proposed seemingly unrelated regression (SUR), which considers simultaneous fitting and a correlated system of either linear or nonlinear equations.

Parresol (2001) provided a methodology for using seemingly unrelated regression, and applications of this approach have been carried out in conifer stands in China (Dong, Zhang, & Li, 2014; Wang et al., 2018), the United States (Zhao et al., 2015), and Brazil (Sanquetta et al., 2015) as well as in eucalyptus stands in Portugal (António, Tomé, Tomé, Soares, & Fontes, 2007) and Spain (Vega-Nieva, Valero, Picos, & Jiménez, 2015). These studies have taken into account stands of middle to advanced age. Even so, considering the importance of forest production in Latin America, particularly that of eucalyptus stands, environmental management for ensure high productivity (Payn et al., 2015), ecophysiological models requires that guarantee consistency and efficiency with regard to biomass estimations, especially during the first years of forest development.

For this reason, Behling et al. (2018) indicated that the best model should consider all previously identified characteristics simultaneously. In addition, both evaluation and fitting compatible biomass equations are important for predicting stand biomass at different stand ages, scaling biomass or carbon stocks, and assessing root: shoot or biomass component ratios. For example, in eucalyptus forest management plans, nutrient budgets and the potential of forests to offset anthropogenic carbon emissions are important considerations (Nizami et al., 2017). In addition, accurate individual biomass estimates may allow us to understand the difference productivity responses of different eucalyptus genotypes to management and climate effects.

This study aimed to analyze the statistical estimates, biological consistency, and efficiency of independent and additive systems models for biomass estimations in young eucalyptus stands. In particular, this study focused on an additive system for modeling individual tree biomass in young eucalyptus stands, considering that total aboveground biomass model needed to be compatible with the sum of biomass components. We hypothesized that (i) systems models would provide better biomass estimates than independent models and (ii) biological consistency would always be related to parameter efficiency in biomass models.

Material and methods

The data of this study were collected in the experimental project 'EUCAHYDRO II: Predictive model of water use, water use efficiency and drought resistance of Eucalyptus plantations and genotypes' which aimed to evaluate physiological traits and environmental factors that drive initial growth behavior in eucalyptus stands. The project was established on October 31, 2017 in the Yumbel, Bio-Bio region in southern Central Chile (37°8′0.01″ S, 72°27′34.70″ W).

The soils in the study site are volcanic black sands with low water-holding capacity. A warm-summer climate that is classified as Mediterranean (Csb) is present in the region. The annual rainfall (mean of 1299 mm) occurs predominantly during winter (~ 60%). The mean annual temperature is 14.7 °C (1979-2014), with maximum and minimum temperatures of 40°C in summer (February) and -5°C in winter (June).

The experiment was carried out with three repetitions and two borderlines. The plots consisted of sixteen trees planted with 1.5×1.0 m spacing. The data consisted of the aboveground biomass measurements of 136 sample trees divided into 34 genotypes. The selected eucalyptus genotypes were divided into three species or crosses, which are the most commonly planted species by Chilean forest companies. The species or crosses consisted of sixteen *Eucalyptus globulus* Labill. genotypes, three *Eucalyptus nitens* (H.Deane & Maiden) genotypes, and fifteen *Eucalyptus nitens* x *Eucalyptus globulus* crosses.

The soil was prepared without a subsoiler. Weed control was carried out with glyphosate (2.5 L ha⁻¹) until crown closure. Fertilization included the application of 30 g of controlled-release fertilizer (Basacote® Plus 12M) per plant at planting. The trees were irrigated at planting and until March 2018 and from November 2018 to February 2019. Both irrigation periods took place during the dry summer months and consisted of the application of 1.6 L of water m^{-2} h^{-1}. The total amount of water applied during the spring and summer months was 664.5 mm during November 2017 to March 2018 and 157.8 mm during November 2018 to February 2019.

The aboveground biomass of individual trees was measured using a destructive method (Picard et al., 2012) when the trees reached 18.5 months in age. The trees were selected according to a distribution based on the diameter 10 cm above the groundline (d), and four trees were selected per genotype. The selected trees consisted of (1) a tree with the greatest d value, (2) a tree with the d mean value plus one standard deviation, (3) a tree with the mean d value, and (4) a tree with the mean d value minus one standard deviation. For each tree that was harvested, we separated the aboveground biomass into the stem (wood + bark), branch, and leaf components. First, the individual trees were felled and removed from the experimental site. Second, d was measured with a caliper (mm), total height (h) was determined with a measuring tape (m), and the branch diameters (bd) at 2 cm from the stem insertion point were measured with a caliper (mm).

After all branches were measured, we determined the bd range per genotype. We randomly selected 15 branches per genotype, and the branches were cut 2 cm from the stem insertion point. For each cut branch, we separated the living foliage from the wood of the branch. The foliage was dried them to a constant weight at 105°C. The biomass of individual tree branches (w_b) and leaves (w_l) was estimated by fitting exponential models as a function of bd per genotype, as follow:

$$f(W_b; W_1 = \beta_0 * e^{(bd * \beta_1)}$$

After removing the branches, the stem was sectioned into three or four sections. Individual tree stem biomass was determined with a balance with a precision of 0.05 kg. A sample disc 8-cm thick was cut from each tree stem 10 cm from the groundline to determine the moisture content with an analytical balance with a precision of 0.01 g. The dry weights of the discs were measured after the samples were dried to a constant weight at 105°C. Then, individual stem biomass was calculated based on the ratio of dry biomass to fresh biomass.

The estimated biomass component models were based on a nonlinear equation with additive error terms. These are the basic model forms that have been used in other forest stand studies (António et al., 2007). The independent variables in the models were d, dh, and d^2h.

$$W_C = (\beta_b * x_1^{\beta_1} \dots x_n^{\beta_n}) + \varepsilon_1$$

where: w_c is the dry biomass component, c (stem, branches, leaves, or total) and β are model coefficients, x_n are the independent variables, and ε_t is the random error of the tth observation.

In the first step, each biomass component was independently fitted using a generalized least squares (GLS) approach with the MODEL procedure in SAS® software (Statistical Analysis System [SAS], 2009).

The model coefficients were equal to zero as the initial parameters in the interaction process. When selecting the best model for each biomass component, the significance of the parameters (α = 0.05) was considered, and the following statistics were calculated:

$$MRES = \frac{\sum_{i=1}^n (W_c - W_{c-est})}{n} \qquad RMSE = \sqrt{\frac{\sum_{i=1}^n (w_c - w_{c-est})^2}{n-p}}$$

$$R_{adj}^2 = \left(\frac{\sum_{i=1}^n (w_c - w_{c-est})}{\sum_{i=1}^n (w_c - \overline{w_c})}\right)\left(\frac{n-1}{n-p}\right) \qquad CV = \left(\frac{RMSE}{\overline{w_c}}\right) * 100$$

where; $MRES$ is the mean residuals (Kg), $RMSE$ is the root mean square error (Kg), R_{adj}^2 is the fitted coefficient of determination (%), CV is the coefficient of variation (%), w_{c-est} is the estimated dry biomass of component c, n is the number of data used in the fitting, p is the number of parameters in the model, and $\overline{w_c}$ is the mean value of the observed dependent variable for component c.

In the second step, the models were fitted using SUR to guarantee additivity (Vega-Nieva et al., 2015). The system model was composed of stem, branch, leaf, and total biomass according to the specification that total biomass was a function of the independent variables in the equations of the biomass components. SUR was fitted considering the best models selected for the independent fitting. The initial parameters in the interaction process were equal to zero. Iterative convergence to minimize the residual sum of squares was performed with the Gauss–Newton method. The analysis was performed in SAS® using the ITSUR option in the MODEL procedure.

In the third step, residual heteroscedasticity of the GLS was determined with graphical analyses. The estimated residues ($\hat{e}_i$) of the GLS model were used as the dependent variable in the error variance model, taking the natural logarithm (ln) of the square residues [$(\hat{e}_i)^2$].

$$\left(\overset{\wedge}{e_i}\right)^2 = (w_c - w_{c-est})^2 \qquad ln\left(\overset{\wedge}{e_i}\right) = \beta_{0'} + \beta_{1'} * ln(x_1) + \cdots + \beta_{n'} * ln(x_n)$$

The natural logarithm of independent variables fitted by GLS was used to define a matrix, Δ, which was an element-wise square root based on the inverse of the weight matrix for observation, i, in component c.

$$\Delta_{ic} = \frac{1}{\sqrt{x_1^{\beta_{1'}} * \dots * x_n^{\beta_{n'}}}}$$

The fourth step employed the same approach as the third step; however, the estimated SUR errors were used to model the error variance and weight calculations resulting in weighted seemingly unrelated regression (WSUR). After fixing the weighting functions in steps three and four, we refitted the equations in steps one and two using the MODEL procedure in SAS®.

According to Parresol (1999, 2001), the β estimation is obtained by minimizing the residual sum of squares:

$$S(\beta) = [y - f(X, \beta)]'[y - f(X, \beta)]$$

$$S^2(\beta) = \hat{\sigma}^2[Z(\beta)'Z(\beta)]^{-1}$$

$$\hat{\sigma}^2 = \frac{S(\beta)}{T - K}$$

$$\hat{v}(\hat{y}_i) = Z(\beta)_i' S^2(\beta) Z(\beta)_i'$$

where: $S(\beta)$ $S(\beta)$ is the residual sum squares, $S^2(\beta)$ is the variance-covariance matrix, $\hat{\sigma}^2$ is the variance, $Z(\beta)$ is the partial derivative matrix, T is the number of observations, K is the number of parameters in the model, and $\hat{v}(\hat{y}_i)$ is the variance of mean estimated observation i.

The weighted least square (WLS) estimator was defined as:

$$\beta = (X'\Psi(\hat{\theta})^{-1}X)^{-1}X'\Psi(\hat{\theta})^{-1}y$$

$$S(\beta) = [y - f(X, \beta)]'\Psi(\hat{\theta})^{-1}[\grave{y} - f(X, \beta)]$$

where: $\Psi(\hat{\theta})$ is a diagonal matrix of weights dependent on a fixed number of parameters denoted by vector θ.

For more details, see Parresol (1999). The system matrix of weights was written in block-diagonal form:

$$\Psi_i(\theta_i) = \begin{bmatrix} \Psi_1(\theta_1) & 0 & \cdots & 0 \\ 0 & \Psi_2(\theta_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \Psi_n(\theta_n) \end{bmatrix}$$

Finally, the confidence interval of the mean estimated value was defined as:

$$CI = \left[\hat{y} \pm t_{\frac{\alpha}{2}}\sqrt{\hat{v}(\hat{y})}\right]$$

where: Cl is the confidence interval, t is the value of t distribution at 95% probability.

The estimate of vector β employing SUR, which minimizes the residual sum of squares, was given by a restriction of the parameter β:

$$\beta = [y - f(X, \beta)]'(\Sigma^{-1} \otimes I)[y - f(X, \beta)]$$

$$\widehat{\Sigma}_b = \left[F(\beta)'(\widehat{\Sigma}^{-1} \otimes I)F(\beta)\right]^{-1}$$

$$\hat{\sigma}_{SUR} = \frac{\varepsilon'(\Sigma^{-1} \otimes I)\varepsilon}{MT - K}$$

$$S^2{}_{\hat{y}_j} = F_j(\beta)'(\widehat{\Sigma}_b)F_j(\beta)$$

where: Σ is the covariance matrix, $\otimes$ is the Kronecker product, I is identity matrix, $F(\beta)$ is partial derivatives for the SUR system, $\widehat{\Sigma}_b$ is the estimated covariance matrix of parameter, $\hat{\sigma}_{SUR}$ is the variance of SUR, M is the number of equations used, and $S^2{}_{\hat{y}_j}$ is estimated variance from the equation j^{th} of SUR of the ith observation ($\hat{y}_{ji}$).

The estimated covariance matrix of the parameter and variance of weighted-nonlinear seemingly unrelated regressions were calculated as follows:

$$\hat{\textstyle\sum}_{bw} = \left[F(\beta)'(\hat{\textstyle\sum}^{-1} \otimes I)(\Psi(\hat{\theta})^{-1})F(\beta)\right]^{-1}$$

$$\hat{\sigma}_{WSUR} = \frac{\varepsilon'\hat{\Delta}'(\textstyle\sum^{-1} \otimes I)\hat{\Delta}\varepsilon}{MT - K}$$

$$S^2{}_{\hat{y}_j} = F_j(\beta)'(\hat{\textstyle\sum}_{bw})F_j(\beta)$$

where: $\hat{\textstyle\sum}_{bw}$ is the covariance matrix of parameter in WSUR, $\hat{\sigma}_{WSUR}$ is the variance of WSUR equations, $\hat{\Delta}$ is the matrix of weights.

The confidence interval of the mean estimated value at 95% probability for SUR and WSUR was defined as follows:

$$CI = \left[\hat{y} \pm t_{\frac{\alpha}{2}}\sqrt{S^2{}_{y_j}}\right]$$

For more details, see Parresol (2001).

Biological consistency was assessed in relation to the additivity of the biomass components with regard to total biomass and was only tested for the equations fitted by independent models in steps one and three. If the condition $(w_{leaves} + w_{brances} + w_{stem}) - w_{total} \neq 0$ was not satisfied, then the biomass estimates were considered to lack biological consistency.

The efficiency of the estimators of the biomass equations fitted by all steps was assessed in relation to the precision of the calculated confidence intervals as follows (Parresol, 2001):

$$PIC_i = \frac{CI_i}{w_{obs_i} - w_{est_i}}$$

where: PIC_i is the precision of the confidence interval for observation i, CI_i is the confidence interval for observation i, w_{obs_i} is the observed value for observation i, and w_{est_i} is the estimated values for observation i.

The effect of the lack of additivity and efficiency on the estimates of biomass components and total biomass at the species level were evaluated with the forest inventory data. The biomass components and total biomass were estimated considering the last forest inventory assessment and extrapolated to hectare scale. The estimates were compared using a Chi-square test, in which the null hypothesis assumes that the values estimated by each step are equal at 95% probability (Behling et al., 2018).

Results

Mean bd values ranged between 7.5 and 9.0 mm among eucalyptus genotypes. The minimum and maximum bd values were 1.06 and 29.32 mm, respectively, with fewer branches present with either small or large bd. The relationship between bd and crown biomass followed similar patterns among eucalyptus genotypes, and the largest branches presented the largest biomass values (Figure 1). The R^2_{adj} and RMSE of the fitted exponential models were 0.92 ± 0.07 and 9.45 g ± 6.58, respectively, for branch biomass and 0.88 ± 0.08 and 12.96 g ± 8.67, respectively, for leaf biomass.
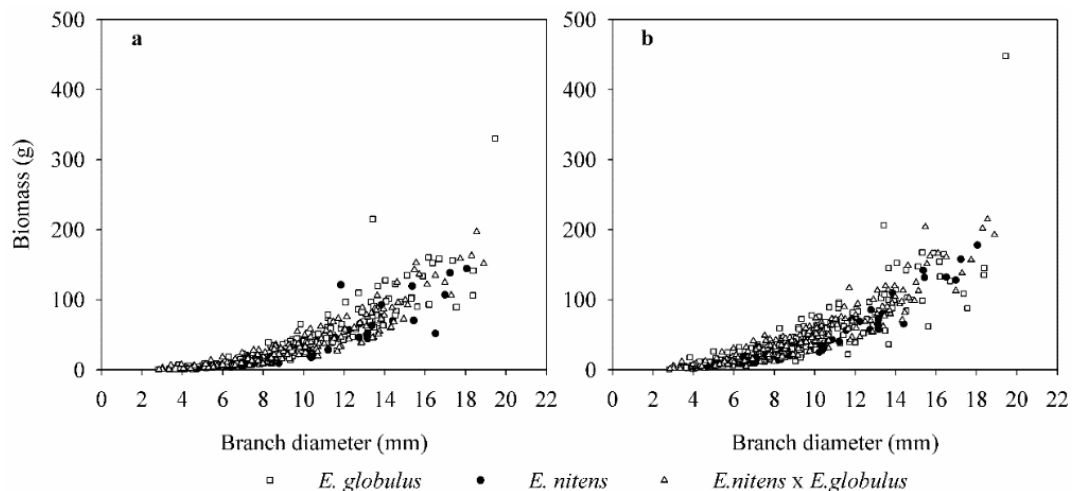

Figure 1. Relationship of branch diameter to (a) branch biomass and (b) leaf biomass for 18.5-month-old *Eucalyptus globulus*, *Eucalyptus nitens*, and *E. nitens* x *E. globulus* individuals in Yumbel, Chile.

The summary statistics for d, h, and the biomass components of all sample trees are shown in Table 1. The coefficients of variation (CV) were higher for biomass components (i.e., 34.62, 36.36, 33.32, and 31.17 for the stem, branch, leaf, and total biomass, respectively) compared those of d (17.33) or h (10.53). The Pearson correlations between the biomass components and d or h were positive (0.68 – 0.92) and significant at 1%. However, the correlation coefficients for biomass and d were higher than those for biomass and h. With regard to the relationships among biomass components, values greater than 0.90 were found, especially between stem and total biomass and branch and leaf biomass (data not shown).

Table 1. The mean, minimum (Min), maximum (Max), standard deviation (SD), and coefficient of variation (CV) values for the diameter 10 cm above the groundline (d), total height (h), and the aboveground biomass components for 18.5-month-old *Eucalyptus globulus*, *Eucalyptus nitens*, and *E. nitens* x *E. globulus* individuals in Yumbel, Chile.

Statistics	Tree variables		Biomass of aboveground components (kg)			
	d (cm)	h (m)	Stem	Branch	Leaf	Total
Mean	7.00	6.63	4.04	1.66	2.29	7.99
S.D.	1.42	0.91	1.58	0.66	0.82	2.81
CV	20.41	13.67	39.24	39.58	36.00	35.19

Biomass partitioning was similar among eucalyptus species. Individual stem biomass represented 50% of the total biomass, ranging between 31 – 68%, 38 – 60%, and 34 – 62%, for *E. globulus*, *E. nitens,* and *E. nitens* × *E. globulus,* respectively. Individual branch biomass represented 20% of total biomass and ranging between 12 – 27%, 14 – 28%, and 14 – 26% for *E. globulus*, *E. nitens,* and *E. nitens* × *E. globulus,* respectively. Individual leaf biomass represented 30% of the total biomass, and the largest variation among species ranged from 19 – 42%, 23 – 33%, and 22 – 39%, for *E. globulus*, *E. nitens,* and *E. nitens* × *E. globulus,* respectively. Leaf biomass represented approximately 50% of crown biomass.

For the equations fitted by GLS, we observed that most equations underestimated individual biomass (Table 2). Branch and leaf biomass presented the lowest worst values of R^2_{adj}, the highest RMSE, and the highest CV because these variables presented the weakest correlations with d or h when compared to those of stem and total biomass.

The selection of the best model considered the statistical analysis in Table 2 and model parsimony for SUR fitting. The best model was determined by considering the variables d and h for stem biomass, d²h for branch and total biomass and d for leaf biomass.

A pattern was detected in the increase of d and residual values, indicating that the residues fitted by SUR were not distributed in the same manner between tree sizes (Figure 2). As such, variability was present in the error model structure. To deal with this variation, the weight function was considered. The same behavior was observed when we compared GLS to WLS (data not shown). The weight functions used to correct the error variability used by WLS and WSUR are presented in Table 3 and show an improvement in the pattern of the residuals (Figure 3).

Table 2. Generalized least squares (GLS) model performance considering mean residuals (MRES), root mean square error (RMSE), adjusted coefficient of determination (R^2_{adj}), and coefficient of variation (CV) for different independent variables (I.V.) in modeling the biomass components.

| Biomass components	Performance	I.V.		
		d	d h	d²h
Stem	MRES	0.00867	-0.0005	0.00017
	RMSE	0.6724	0.6041	0.6031
	R^2_{adj}	82.00%	85.69%	85.52%
	CV	16.65%	14.95%	14.93%
Branch	MRES	0.00116	-0.0004	-0.0019
	RMSE	0.4779	0.4784	0.4788
	R^2_{adj}	47.69%	47.57%	47.49%
	CV	28.62%	28.65%	28.68%
Leaf	MRES	-0.0037	-0.0064	-0.008
	RMSE	0.5551	0.5538	0.5539
	R^2_{adj}	54.59%	54.79%	54.79%
	CV	24.26%	24.21%	24.21%
Total	MRES	0.0043	-0.0109	-0.0126
	RMSE	1.2949	1.2413	1.2384
	R^2_{adj}	78.83%	80.54%	80.64%
	CV	16.19%	15.52%	15.49%
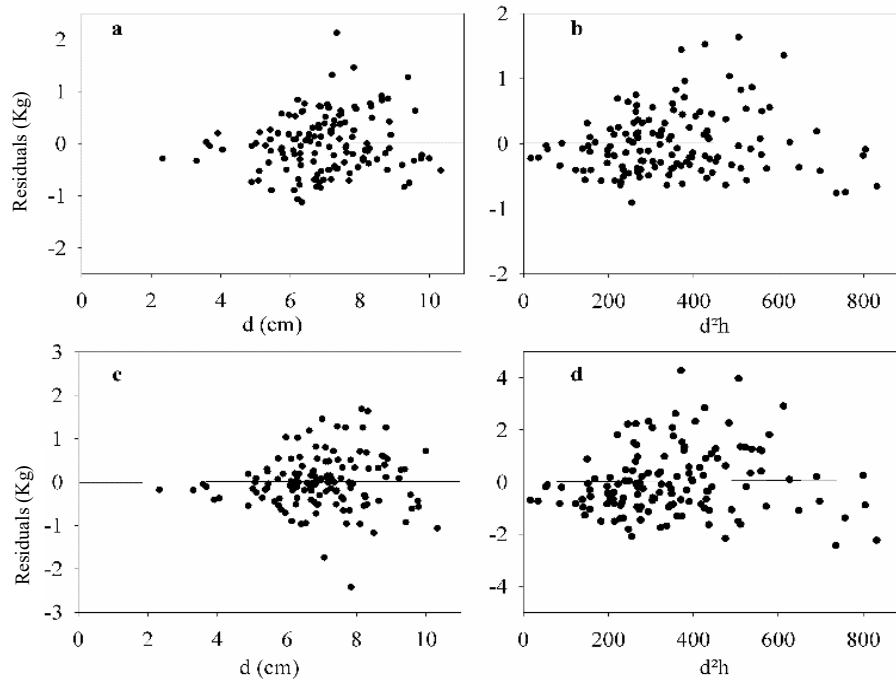
Figure 2. Residuals for the (a) leaf biomass, (b) branch biomass, (c) stem biomass, and (d) total biomass estimates for 18.5-month-old *Eucalyptus globulus*, *Eucalyptus nitens*, and *E. nitens* x *E. globulus* individuals using SUR in Yumbel, Chile.

Table 3. Weight functions used to fit biomass components considering the diameter 10 cm above the groundline (*d*) and total height (*t*) for 18.5-months-old *Eucalyptus globulus*, *Eucalyptus nitens*, and *E. nitens* x *E. globulus* individuals in Yumbel, Chile.

Biomass components	Weighted WLS	Weighted WSUR
Stem	$\dfrac{1}{\sqrt{d^{2.970717} * h^{-1.12551}}}$	$\dfrac{1}{\sqrt{d^{2.521194} * h^{-0.33379}}}$
Branch	$\dfrac{1}{\sqrt{(d^2 h)^{0.954756}}}$	$\dfrac{1}{\sqrt{(d^2 h)^{0.905923}}}$
Leaf	$\dfrac{1}{\sqrt{d^{3.365521}}}$	$\dfrac{1}{\sqrt{d^{2.689988}}}$
Total	$\dfrac{1}{\sqrt{(d^2 h)^{1.00233}}}$	$\dfrac{1}{\sqrt{(d^2 h)^{0.623724}}}$
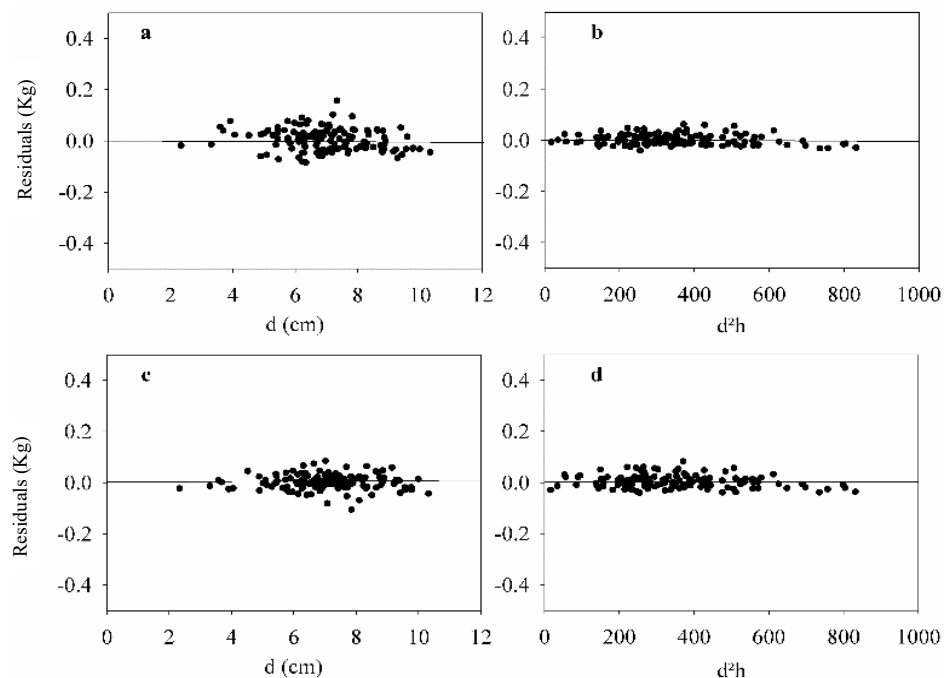
Figure 3. Residuals of the (a) leaf biomass, (b) branch biomass, (c) stem biomass, and (d) total biomass estimates for 18.5-month-old *Eucalyptus globulus*, *Eucalyptus nitens*, and *E. nitens* x *E. globulus* individuals using WSUR in Yumbel, Chile.

The first hypothesis outlined in this paper was not supported, which suggests that systems models provide better biomass estimates than those of independent models for young eucalyptus stands. The estimated coefficients and their standard deviations indicated that similar variation was present among the four methods used in this study (Table 4). Large differences were found with regard to the interceptions of the equations (β_0).

Table 4. Parameter estimation $(\hat{\beta})$ and statistical analysis results for the biomass models considering the generalized least squares (GLS), weighted least squares (WLS), seemingly unrelated regression (SUR), and weighted seemingly unrelated regression (WSUR) methods for 18.5-month-old *Eucalyptus globulus*, *Eucalyptus nitens*, and *E. nitens* x *E. globulus* individuals in Yumbel, Chile.

Biomass	Model	$\hat{\beta}_0$		$\hat{\beta}_1$		$\hat{\beta}_2$		MRES	RMSE	R^2_{adj} (%)	CV (%)
		Mean	SD	Mean	SD	Mean	SD				
Stem	GLS	0.0476	0.0109	1.4340	0.094	0.8541	0.151	-0.0005	0.604	85.47	14.95
	WLS	0.0381	0.0081	1.4759	0.093	0.9243	0.152	-0.00039	0.606	85.34	15.02
	SUR	0.0401	0.0090	1.3485	0.090	1.0313	0.143	-0.00556	0.603	85.48	14.94
	WSUR	0.0278	0.0052	1.3121	0.074	1.2542	0.126	0.03406	0.617	84.84	15.27
Branch	GLS	0.0788	0.0260	0.5264	0.055	-	-	-0.0019	0.478	47.49	28.68
	WLS	0.0463	0.0123	0.6166	0.045	-	-	-0.0022	0.484	46.30	29.00
	SUR	0.1050	0.0329	0.4775	0.052	-	-	-0.0045	0.475	47.51	28.45
	WSUR	0.0354	0.0052	0.6574	0.025	-	-	0.04203	0.488	45.46	29.22
Leaf	GLS	0.1803	0.0418	1.3017	0.114	-	-	-0.0037	0.555	54.59	24.26
	WLS	0.1257	0.0211	1.4849	0.087	-	-	-0.0104	0.561	53.56	24.53
	SUR	0.2235	0.0493	1.19320	0.108	-	-	-0.0037	0.555	54.58	24.26
	WSUR	0.1119	0.0126	1.5296	0.057	-	-	0.05298	0.562	53.39	24.58
Total	GLS	0.2013	0.0369	0.6332	0.030	-	-	-0.0126	1.238	80.64	15.49
	WLS	0.1465	0.0222	0.6868	0.025	-	-	-0.0050	1.254	80.14	15.68
	SUR	-	-	-	-	-	-	-0.0138	1.256	80.07	15.71
	WSUR	-	-	-	-	-	-	0.12908	1.281	79.26	16.02

Considering the performance of equations, we can observe that there were similar RMSE, R^2_{adj}, and CV values for the methods employed. However, there were differences in MRES using WSUR, which overestimated the biomass values for all biomass components and standard deviations of the parameters. It was possible to observe a decrease in standard deviation (SD) when a weight function was used for the variability parameter estimation.

The biomass estimates at the hectare level obtained by the GLS, WLS, SUR, and WSUR methods not significantly different when compared with a Chi-square test (p > 0.05). The mean values that were estimated according to the aforementioned methods were similar among biomass components, indicating that there was no difference with regard to the performance of the estimations (Figure 4).
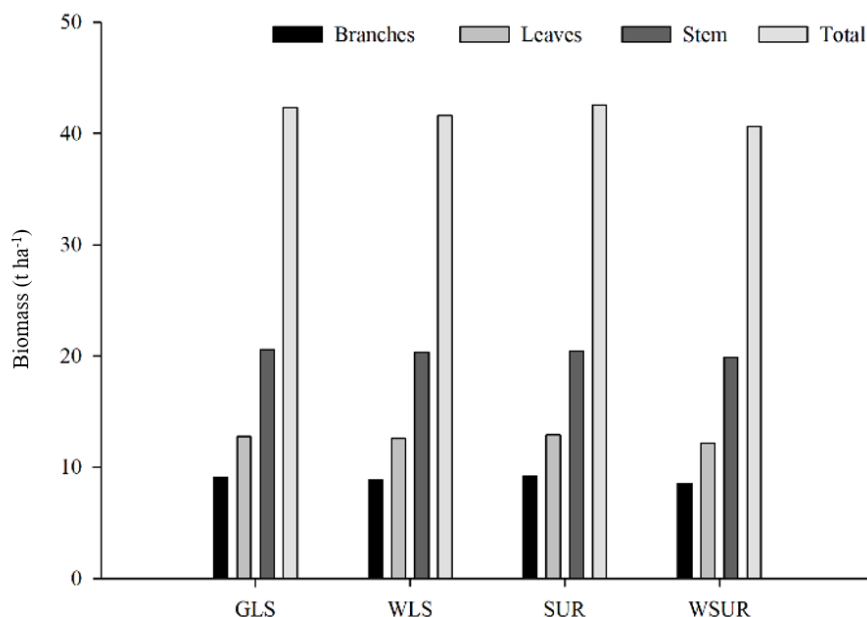

Figure 4. Biomass component estimates (*t* ha^{-1}) considering the equations fitted by the GLS, WLS, SUR, and WSUR methods for the 18.5-month-old *Eucalyptus globulus*, *Eucalyptus nitens*, and *E. nitens* x *E. globulus* individuals in Yumbel, Chile.

Biological consistency, according to the restriction that the sum of the biomass components must be equal to the total biomass, was not observed with the equations fitted by the GLS and WLS methods since ($w_{leaves} + w_{branches} + w_{stem}) - w_{total} \neq 0$. The non-additivity discrepancy of the equations fitted by GLS and WLS at the tree level ranged from -1.68 – 8.28% and -1.76 – 7.05%. Biological inconsistency was also observed at the hectare level. The non-additivity discrepancy of the equations ranged between -0.87 – 9.98% and -0.80 – 8.46% for GLS and WLS, respectively, resulting in mean differences of 0.004 t ha^{-1} and 0.005 t ha^{-1}, respectively. Regardless, these differences were not very large, the GLS and WLS equations were considered to be biological inconsistent.

Using the equations fitted by GLS, the precision of the estimates ranged from 0.56 – 24.56%, 0.38 – 34.08%, 0.02 – 26.07%, and 0.22 – 21.87% for stem, branch, leaf, and total biomass, respectively. For the equations fitted by WLS, the values for stem, branch, leaf, and total biomass ranged from 0.38 – 20.72%, 0.20 – 27.95%, 0.25 – 17.55%, and 0.13 – 17.49%, respectively. The biomass estimates of the components and total biomass using the equations fitted by WLS resulted in narrow confidence intervals and were more efficient than GLS.

The equations fitted by SUR resulted in estimate precisions ranging from 0.09 – 8.13% (stems), 0.13 – 29.99% (branches), 0.67 – 23.67% (leaves), and 0.16 - 13.51% (total biomass). As expected, SUR was more efficient when estimating total biomass; however, SUR did not perform as well when estimating branch and leaf biomass when compared to those of WLS, indicating that biological consistency did not guarantee the most efficient estimators. Therefore, the second hypothesis outlined in this paper was also not supported. Biological consistency is not always related to parameter efficiency in biomass models for young eucalyptus stands since SUR showed biological consistency; however, it does not show higher estimation efficiency for crown biomass components.

For WSUR, precision ranged from 0.02 – 9.12%, 1.53 – 11.43%, 0.11 – 13.06%, and 0.55 – 4.55% for stem, branch, leaf, and total biomass, respectively. The WUSR provided the most efficient estimators for all biomass components, resulting in a better model approaches when compared to those of GLS, WLS, and SUR. As shown in Figure 5, narrower confidence intervals were present with WSUR when compared to those of GLS, especially for branch and leaf biomass.
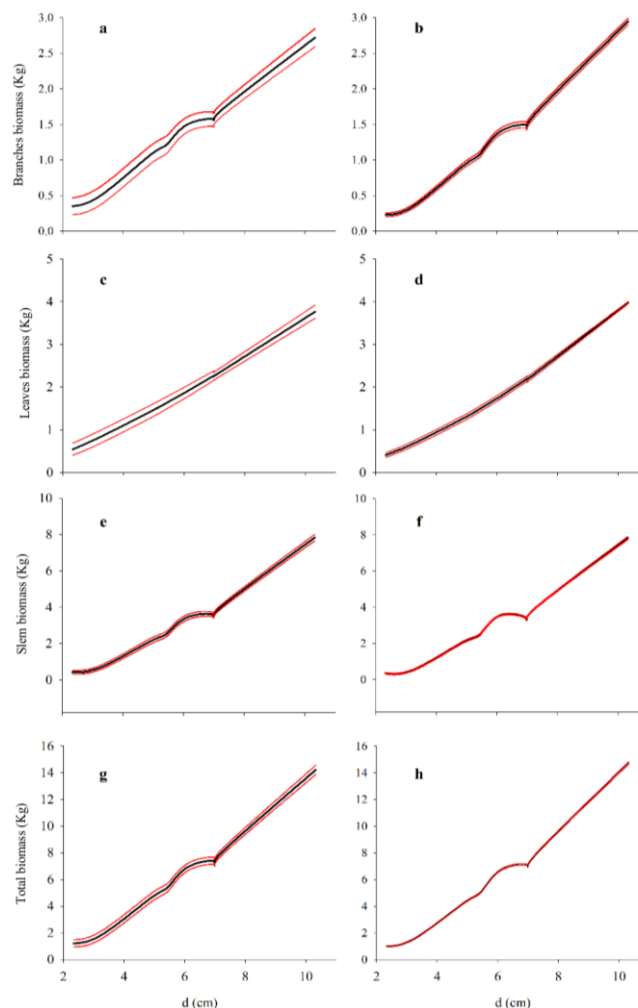

Figure 5. Differences in confidence intervals (red lines) between branch, leaf, stem, and total biomass estimates using GLS (a, c, e, g) and WSUR (b, d, f, h) for 18.5-month-old *Eucalyptus globulus*, *Eucalyptus nitens*, and *E. nitens* x *E. globulus* in Yumbel, Chile.

The mean gain efficiency in the confidence intervals of the equations fitted by WSUR was calculated by $EIC = \left(1 - \frac{PIC_{WSUR}}{PIC_{other-methods}}\right)$, and the values were 33.05, 47.34, 34.64, and 64.71% for stem, branch, leaf, and total biomass, respectively. The efficiencies using WSUR were positive and sometimes greater than 50% when compared to those of the GLS, WLS, and SUR methods (Figure 6).
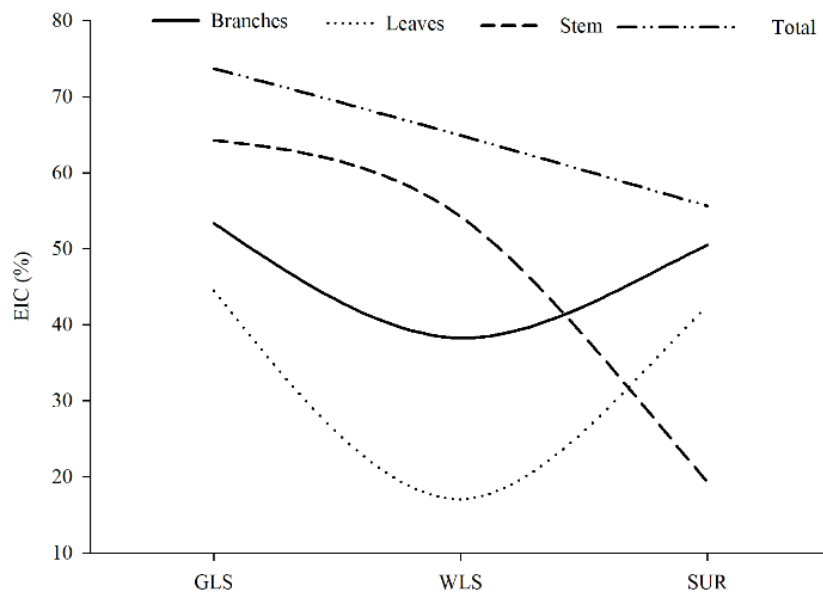

Figure 6. Efficiency of the confidence intervals for the component and total biomass estimates fitted by WUSR compared to those of GLS, WLS, and SUR.

Discussion

The worst statistical evaluation of the power functions was found for the crown biomass components, while the best was observed for stem and total biomass. Similar results have also been found in other studies, such as those by António et al. (2007) with *E. globulus* in Portugal and Viera et al. (2017) with *Eucalyptus saligna* Sm. and *Eucalyptus grandis* Hill ex Maiden × *Eucalyptus urophylla* S. T. Blake crosses in Brazil. The reasons for this are two-fold. Firstly, the reason is that the crown biomass components, leaves, and branches were correlated to a lesser degree with *d* and *h*. Secondly, leaf and branch biomass were more variable and depend on age, temperature, relative humidity, soil water content, and the species or genotype, as highlighted by António et al. (2007) and Zhang et al. (2015). The relationships between crown biomass and *d* or *h* improves as forest stands age (Behling et al., 2018).

The performance of the estimators in biomass models using the GLS, WLS, SUR, and WSUR methods were similar with regard to the quality of the statistical fitting. However, some statistics obtained by the equation fitted using SUR were better than those obtained with GLS while others were not. For WSUR, all statistics obtained by the fitted models were worse when compared to those of the other approaches. This is an effect of the flexibility of the estimator to meet the requirement of the additivity of biomass components in the SUR and WSUR methods, and it was also observed by Behling et al. (2018). Apart from this, the performance of the estimators did not affect the predictions, since the estimates of the biomass components and total biomass did not differ at either tree or hectare levels among the methods according the results of the Chi-square test. In conclusion, the GLS, WLS, SUR, and WSUR methods are equivalent with regard to performance.

Modeling biomass components and total biomass with GLS has been widely used in forest science; however, this approach does not guarantee additivity and most often results in larger errors when expanded to larger forest inventory areas (Sanquetta et al., 2015). The condition of biological consistency was not satisfied for the equations fitted through GLS and WLS, and non-additivity was observed at the tree and hectare levels, resulting in positive and negative results. The biological inconsistency percentage was higher for GLS than for WLS, mainly due to the weight function for the corrected error variance structure. In addition, biological inconsistency was also higher at tree level compared to that at hectare level given that at tree level, positive and negative results do not cancel each other out, which results in greater inconsistency when compared to that at either the plot or hectare levels, as highlighted by Behling et al. (2018).

Most of the models used in biomass estimates for planted forests were not fitted considering additivity. Given that these equations did not achieve biological consistency, it is important to take precaution when using them, especially when considering the variance of total biomass, which is a function of the variances and covariances of the estimates of the other biomass components. To deal with the lack of additivity, other studies have used system equations with restrictions on model coefficients according to total biomass (Zhao et al., 2019). In this study, the SUR method guaranteed additivity and resulted in a reduction in the differences with regard to observed and estimated total biomass due to the correlations between biomass components, coefficient restrictions, and allometric relationships, which ensured biological consistency (Wang et al., 2018). However, GLS and SUR presented residual heteroscedasticity, as has been found in other studies (Zhao et al., 2015; Vonderach et al., 2018). As the d and d^2h values increased, the absolute value of the residues increased for all biomass components (leaf, branch, stem) and total biomass. If this is not corrected, the variance among larger trees would be underestimated and that among smaller trees would be overestimated (Parresol, 2001).

A lack of consideration with regard to heteroscedasticity could lead to non-reliable parameter estimates of the biomass components, with increased variance and wide confidence intervals, which decreases estimator efficiency, as was found in this study. The correction of heteroscedasticity is accomplished by assigning a weight factor assigned to each tree (Zhang, Peng, Huang, & Zeng, 2016). Estimator efficiency results in a process with minimum variance, as was found for the WLS and WSUR methods, and results in narrow confidence intervals, especially for the crown biomass components. The WSUR improved the parameter efficiency and considered all the benefits of the simultaneous system of equations, although for some components, statistical evaluations were slightly worse than those for GLS and SUR. Nonetheless, the confidence intervals were better for WSUR compared to those of the other methods, showing that the variance in the biomass estimates was lower.

Estimator efficiency was not always related to biological consistency since WLS showed biological inconsistency; however, WLS presented greater efficiency estimators than that of SUR. Although the crown biomass components are difficult to model, it is necessary to use an estimator that results more accurate estimates, primarily because the leaves are raw material that ensure energy production and are important components for ecophysiological modeling (Binkley et al., 2004).

Zhao et al. (2015) highlighted that accurate modeling of young tree growth is necessary to understand forest development, schedule silvicultural treatments, and identify environmental variables that could influence initial biomass production. The biological consistency ensured by the additive equations is a desirable property in modeling biomass and affects the estimations of forest biomass. In addition to biomass additivity, a correction of heteroscedastic residues is established by the weight function of GLS methods and ensures that a reliable estimate of biomass with statistical efficiency is produced. Sanquetta et al. (2015) observed that the use of SUR had positive effects on confidence intervals. This was effect associated with the efficiency of the parameters obtained by applying WSUR in the present study, which made this approach a powerful tool for eucalyptus biomass models.

The precision of the confidence intervals of the equations used with the WSUR approach generated estimators that are highly efficient, constituting an important alternative to forest ecophysiological models that estimate biomass. These approaches can be easily implemented in SAS using the MODEL procedure, or in other programs, despite the mathematical complexity associated with WSUR with regard to the estimated biomass values. Future studies should focus on the implementation of these approaches with eucalyptus forest stands of different ages and in different to evaluated the validity of the ecophysiology models, considering the relationships among dendrometric variables, biomass, and environmental management strategies.

Conclusion

The fitted procedures of this study do not differ with regard to the performance evaluation since the system of equations do not improve the predictions of the biomass components. The estimated total biomass from the GLS and WLS methods are not equal to the sum of estimated biomass components, resulting in biological inconsistencies. The models in WLS and WSUR produces present smaller variance and narrow confidence intervals, resulting in a more efficient estimators when compared to GLS and SUR. It is recommended that the WSUR approach be used with independent equations in forest ecophysiology models, for accurate biomass estimates.

Acknowledgements

References

António, N., Tomé, M., Tomé, J., Soares, P., & Fontes, L. (2007). Effect of tree, stand, and site variables on the allometry of *Eucalyptus globulus* tree biomass. *Canadian Journal of Forest Research*, *37*, 895-906. DOI: 10.1139/X06-276

Behling, A., Péllico Netto, S., Sanquetta, C. R., Corte, A. P. D., 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. DOI: 10.1590/0001-3765201820170684

Binkley, D., Stape, J. L., & Ryan, M. G. (2004). Thinking about efficiency of resource use in forests. *Forest Ecology and Management, 193*, 5-16. DOI: 10.1016/j.foreco.2004.01.019

Dong, L.; Zhang, L., & Li, F. (2014). A compatible system of biomass equations for three conifer species in Northeast, China. *Forest Ecology and Management, 329*, 306-317. DOI: 10.1016/j.foreco.2014.05.050

Fu, L., Zeng, W., & Tang, S. (2017). Individual tree biomass models to estimate forest biomass for large spatial regions developed using four pine species in China. *Forest Science, 63*(3), 241-249. DOI: 10.5849/FS-2016-055

Hakamada, R., Hubbard, R. M., Ferraz, S., Stape, J. L., & Lemos, C. (2017). Biomass production and potential water stress increase with planting density in four highly productive clonal Eucalyptus genotypes. *Southern Forests: a Journal of Forest Science, 79*(3), 251-257. DOI: 10.2989/20702620.2016.1256041

Nizami, S. M., Yiping, Z., Zheng, Z., Zhiyun, L., Guoping, Y., & Liqing, S. (2017). Evaluation of forest structure, biomass and carbon sequestration in subtropical pristine forests of SW China. *Environmental Science and Pollution Research, 24*, 8137-8146. DOI: 10.1007/s11356-017-8506-7

Parresol, B. R. (1999). Assessing tree and stand biomass: critical comparisons. *Forest Science, 45*(4), 573-593.

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

Payn, T., Carnus, J. M., Freer-Smith, P., Kimberley, M., Kollert W, Liu, S., ... Wingfield, M. J. (2015). Changes in planted forests and future global implications. *Forest Ecology and Management, 352,* 57-67. DOI: 10.1016/j.foreco.2015.06.021

Picard, N., Santi-André, L., & Henry, M. (2012). *Manual for building tree volume and biomass allometric equations: from field measurement to prediction*. Rome, IT: FAO; CIRAD.

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

Statistical Analysis System [SAS]. (2009). *SAS/STAT 9.2 User's Guide* (2nd ed.). Cary, NC: SAS Institute Inc.

Vega-Nieva, D. J., Valero, E., Picos, J., & Jiménez, E. (2015). Modeling the above and belowground biomass of planted and coppiced *Eucalytpus globulus* stands in NW Spain. *Annals of Forest Science, 72*, 967-980. DOI: 10.1007/s13595-015-0493-6

Viera, M., Schumacher, M. V., Bonacina, D. M., Oliveira Ramos, L. O., & Rodríguez-Soalleiro, R. (2017). Biomass and nutrient allocation to aboveground components in fertilized *Eucalyptus saligna* and *E. urograndis* plantations. *New Forest, 48*, 445-462. DOI: 10.1007/s11056-017-9572-x

Vonderach, C., Kändler, G., & Dormann, C.F. (2018). Consistent set of additive biomass functions for eight tree species in Germany fit by nonlinear seemingly unrelated regression. *Annals of Forest Science, 75*(49). DOI: 10.1007/s13595-018-0728-4

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*(60). DOI: 10.1007/s13595-018-0738-2

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of American Statistical Association, 57*(298), 348-368. DOI: 10.1080/01621459.1962.10480664

Zhang, C., Peng, D. L., Huang, G. S., & Zeng, W. S. (2016). Developing aboveground biomass equations both compatible with tree volume equations and additive systems for single-trees in poplar plantations in Jiangsu Province, China. *Forests*, *7*(32). DOI: 10.3390/f7020032

Zhang, H., Wang, K., Xu, X., Song, T., Xu, Y., & Zeng, F. (2015). Biogeographical patterns of biomass allocation in leaves, stems, and roots in Chinas forests. *Scientific Reports, 5*, 1-12. DOI: 10.1038/srep15997

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. DOI: 10.5849/forsci.14-193

Zhao, D., Westfall, J., Coulston, J. W., Lynch, T. B., Bullock, B. P., & Montes, C. R. (2019). Additive biomass equations for slash pine trees: Comparing three modeling approaches. *Canadian Journal of Forest Research, 49,* 27-40. DOI: 10.1139/cjfr-2018-0246