

Ratio estimators and simultaneous regression estimators for biomass in mixed plantation of *Pinus elliottii* Engelm. and *Pinus taeda* L.

Ornelio P. E. Nhaduco^{1,2*}, Alexandre Behling², Sylvio Péllico-Netto², Ana P. Dalla-Corte², Carlos R. Sanquetta², Franciele A. da Silva² and Jonathan W. Trautenmüller²

¹Department of Forest Engineering, Faculty of Agronomy and Forest Engineering, Eduardo Mondlane University, 3453 Julius Nyerere Avenue, 257 Maputo, Mozambique.

²Department of Forest Science, Federal University of Paraná, 80210-170 Curitiba, Paraná, Brazil.

*Correspondence should be addressed to Ornelio P. E. Nhaduco: ornelionhaduco@gmail.com

Abstract

Aim of study: To evaluate the performance and biological consistency of ratio estimators for estimating the biomass of *Pinus elliottii* Engelm. and *Pinus taeda* L. and to compare the results obtained by this method with those generated by the nonlinear seemingly unrelated regression (*NSUR*) method in its weighted version (*WNSUR*), a procedure that was considered as a control.

Area of study: The study was carried out in the state of Paraná, in southern Brazil.

Material and methods: Data were collected on the following variables: diameter at 1.30 m height aboveground (*D*), total height (*H*), crown diameter (*dc*), crown height (*hc*), and the biomass of the following parts was quantified: roots, stem, bark, branches, needles, and total. The ratio model used for each part was obtained by the ratio between the biomass of the *m*th part (dependent variable) and the estimated stem volume (independent variable). Nonlinear allometric models were fitted simultaneously by the *NSUR* and *WNSUR* methods.

Main results: The ratio estimators did not require additional procedures to achieve additivity and, in general, became more accurate for total aboveground biomass and its parts than regression estimators (*WNSUR*).

Research highlights: Due to its simplicity of application and because it meets the premise of additivity of the biomass parts in a natural way, the ratio estimators constitute an appropriate and promising procedure for compatible estimates of tree biomass.

Keywords: additivity; biological consistency; discriminant analysis; simultaneous fitting; stem volume.

Estimadores de razón y estimadores de regresión simultánea para la biomasa en plantaciones mixtas de *Pinus elliottii* Engelm. y *Pinus taeda* L.

Resumen

Objetivo del estudio: Evaluar el desempeño y la consistencia biológica de los estimadores de razón para estimar la biomasa de *Pinus elliottii* Engelm. y *Pinus taeda* L., y comparar los resultados obtenidos mediante este método con los generados por el método de regresión no lineal aparentemente no relacionado (*NSUR*) en su versión ponderada (*WNSUR*), un procedimiento que se consideró como control.

Área de estudio: El estudio se llevó a cabo en el estado de Paraná, en el sur de Brasil.

Material y métodos: Se recopilieron datos sobre las siguientes variables: diámetro a 1,30 m de altura sobre el suelo (D), altura total (H), diámetro de la copa (dc), altura de la copa (hc), y se cuantificó la biomasa de las siguientes partes: raíces, tronco, corteza, ramas, acículas y total.

El modelo de razón utilizado para cada parte se obtuvo mediante la relación entre la biomasa de la parte *m-ésima* (variable dependiente) y el volumen estimado del tronco (variable independiente). Los modelos alométricos no lineales se ajustaron simultáneamente mediante los métodos *NSUR* y *WNSUR*.

Resultados principales: Los estimadores de razón no requirieron procedimientos adicionales para lograr la aditividad y, en general, fueron más precisos para la biomasa aérea total y sus partes que los estimadores de regresión (*WNSUR*).

Aspectos destacados de la investigación: Debido a su simplicidad de aplicación y a que cumple de manera natural con la premisa de aditividad de las partes de la biomasa, los estimadores de razón constituyen un procedimiento adecuado y prometedor para obtener estimaciones compatibles de la biomasa arbórea.

Palabras clave: aditividad; ajuste simultáneo; análisis discriminante; consistencia biológica; volumen del tronco.

Citation: Nhaduco, OPE; Behling, A; Péllico-Netto, S; Dalla-Corte, AP; Sanquetta, CR; da Silva, FA; Trautenmüller JW (2025). Ratio estimators and simultaneous regression estimators for biomass in mixed plantation of *Pinus elliottii* Engelm. and *Pinus taeda* L. Forest Systems, Volume 34, Issue 3, 20908. <https://doi.org/10.5424/fs/2025343-20908>

Received: 06/05/2024. **Accepted:** 05/06/2025. **Published:** 23/01/2026.

Copyright © 2025 CSIC. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License.

Introduction

In the current scenario of climate change, many studies have been conducted to estimate forest biomass for several purposes, highlighting the growing demand for information on stored carbon stocks (Siteo & Tchaúque, 2007; Magalhães & Seifert, 2015; Zhao et al., 2015; Shahrokhzadeh et al., 2015; Fu et al., 2017; Dai et al., 2018; Cao et al., 2019). This information, in turn, is used to estimate the amount of released CO₂, either by plant respiration, by burning, or by decomposition of biomass (Urbano et al., 2008) and to deepen the understanding of the dynamic processes that occur in forest ecosystems (Péllico Netto & Behling, 2019), which provide information for management decisions (Dai et al., 2018).

The direct (destructive) method, in which trees are felled, sectioned, and weighed for each of the parts (usually roots, stem, branches, and foliage) is the most accurate method for estimating forest biomass. However, this method is inconvenient because it is costly and time-consuming (Behling et al., 2018; Dong et al., 2018) and, consequently, it becomes practically not applicable, especially in large areas and preservation areas, where felling trees is not allowed. Therefore, separately fitted allometric models, which generally include the diameter at 1.30 m height aboveground (Dong et al., 2015), or this diameter combined with the total height of the tree (Cole & Ewel, 2006), and/or other easily measurable tree variables (Li & Zhao, 2013), are the most widely used to estimate the total biomass and its parts for different species or forest typologies (Trautenmüller et al., 2023). Nevertheless, if more than one tree part is involved, fitting the biomass of each tree part separately disregards the intrinsic correlations among the biomass of tree parts measured on the same sampled trees (Dong et al., 2014). Consequently, the biomass estimates for the individual tree parts do not add up to the total biomass estimate, and an essential and biologically consistent characteristic of the tree part regression models is that the predictions for the parts should sum to the total tree biomass prediction, a feature known as additivity (Magalhães & Seifert, 2015).

To meet the additivity assumption, different authors have reviewed and put forward various model specifications and estimation methods (Kozak, 1970; Parresol, 1999, 2001). This was highlighted chronologically as follows: Kozak (1970), recognized as a pioneer in introducing the additivity assumption in forest biomass modelling (Trautenmüller et al., 2023); Cunia & Briggs (1984), Chiyenda & Kozak (1984), Reed & Green (1985); and Parresol (1999, 2001), who introduced the two most widely known

methods: the seemingly unrelated regression (*SUR*) and the nonlinear seemingly unrelated regression (*NSUR*) for linear and nonlinear models, respectively (Dong et al., 2014). Regarding the species under study (*Pinus elliottii* Engelm. and *Pinus taeda* L.), some researchers (Parresol, 2001; Sanquetta et al., 2015; Nhaduco et al., 2021; Carvalho et al., 2022) have considered the additivity assumption for biomass parts, specifically applying the *NSUR* method. Although the seemingly unrelated regression (*SUR*) and *NSUR* methods guarantee the additivity of the tree's constituent parts, this operation is conducted in a forced manner, given the restrictions imposed on the coefficients of the total biomass equation. Furthermore, these procedures are complex and require robust statistical software, such as *SAS*, *R*, and *MATLAB*, which, in turn, necessitate a considerable level of proficiency in programming languages to be operated effectively (Trautenmüller et al., 2023).

With the objective of establishing a simple and consistent biological model that does not require additional procedures to meet the additivity assumption of the biomass tree parts but with competitive accuracy of those obtained by the regression models (*SUR* and *NSUR* procedures), Péllico Netto & Behling (2019) introduced the technique of ratio estimators for the estimation of tree biomass. Basically, this technique uses information from the auxiliary variable *X* (in this case, stem volume) that is positively correlated with the variable of interest *Y* (in this case, total biomass and its parts), as described by Cochran (1953; 1963) and Sukhatme (1954). The additivity of the parts is easily achieved through the basic principle of adding ordinary fractions. That is, if ordinary fractions are considered, with the biomass parts of the trees (roots, stem, bark, branches, and needles) as the numerators of these ratios (denoted as *Y1*, *Y2*, *Y3*, *Y4*, and *Y5*), and the stem volume (*X*) as the denominator, it becomes easy to understand that the sum of these fractions will necessarily be equal to the fraction of the total biomass (*Y*) in relation to *X*. Hence, there is perfect additivity, also referred to as natural additivity by Péllico Netto & Behling (2019).

The ratio estimators appear to represent the basic density of the respective parts by their metric unit ($\text{kg}\cdot\text{m}^{-3}$). However, these fractions are “proxy density” estimates, and the ratio obtained for the stem part is the one that most closely approximates the basic density of the wood. Therefore, Péllico Netto & Behling (2019) referred to these ratio estimators as “Scalar Coefficients Proxy of Density – *SCPD*”, rather than basic density itself. In this work, we refer to them as “Slope Coefficients” or “Ratio Estimators”, as also termed by Trautenmüller et al. (2023).

Given that it is a recently introduced methodology, its application is still in its early stages. No studies involving species of the genus *Pinus* were found. Only two studies were identified: the first, authored by the proponents of the method, Péllico Netto & Behling (2019), and the second, by Trautenmüller et al. (2023), which estimated the biomass of black wattle (*Acacia mearnsii* De Wild.) and the biomass of some species from the subtropical forests of Brazil, respectively.

In this context, under the hypothesis that ratio estimators produce reliable estimates and with biological consistency compatible with the *NSUR* method, the aims of the present study were: (i) to evaluate the reliability and biological consistency of ratio estimators in biomass estimation for a mixed stand of *P. elliottii* and *P. taeda* and (ii) to compare the results obtained by this technique with those generated by the *NSUR* method in its weighted version (*WNSUR*), a procedure that was considered as a control.

Material and methods

Database and exploratory data analysis

A total of 70 trees from a mixed stand of *P. elliottii* (33 trees) and *P. taeda* (37 trees), ranging in age from 2 to 24 years, were used for the study. These trees were subjected to measurements of dendrometric variables, including diameter at 1.30 m aboveground (*D*), total height (*H*), crown diameter (*dc*), and crown height (*hc*), as well as biomass quantification of the roots (*WR*), stem (*WS*), bark (*WBa*), branches (*WBr*), and needles (*WN*) using the direct (destructive) method to obtain their green (fresh) weights in the field and determine the dry weights in the laboratory (1), according to the methodology described by Picard et al. (2012). In proportion to the total biomass, 64.14% of the observed biomass corresponded to *WS*, 11.17% to *WBr*, 10.34% to *WR*, 8.91% to *WBa*, and 5.44% to *WN*. Aboveground biomass accounted for 89.66% of the total biomass, while below-ground biomass made up only 10.34%. The descriptive statistics, which

summarize the main characteristics of the data, are presented in Table 1, while the scatter plots, which illustrate the relationships between the variables, are shown in Fig. 1.

$$DM (\%) = \frac{PS}{PV} 100 \quad (1)$$

where: DM is dry matter (%); PS is the weight of the total dry mass of the m^{th} part; and PV is the weight of the total green mass of the m^{th} part.

Table 1. Descriptive statistics of independent and dependent variables used for fitting biomass estimation models in mixed stands of *Pinus elliottii* Engelm. and *Pinus taeda* L. in Southern Brazil.

Statistics	I	D	H	dc	hc	WR	WS	WBa	WBr	WN	WT
<i>P. elliottii</i> (n = 33)											
Min.	4.00	8.28	4.10	1.89	2.28	4.24	3.67	1.73	1.51	3.20	14.36
Max.	20.00	31.04	24.49	6.77	13.40	53.27	335.56	47.37	80.90	39.54	488.32
Mean	12.03	20.55	15.48	3.35	7.52	21.24	107.77	19.88	20.03	12.43	181.36
SD	5.17	6.30	6.01	1.27	2.60	15.72	98.20	15.71	18.60	8.06	150.51
CV (%)	42.97	30.67	38.86	37.86	34.56	74.01	91.11	79.01	92.83	64.86	82.99
<i>P. taeda</i> (n = 37)											
Min.	2.00	1.91	2.35	1.26	1.78	0.20	0.40	0.10	0.19	0.35	1.23
Max.	24.00	40.27	30.30	7.08	14.93	58.04	687.39	65.91	96.14	22.96	889.81
Mean	10.92	19.71	14.85	3.36	7.16	19.15	140.09	15.07	23.29	8.93	206.52
SD	6.59	10.59	8.94	1.57	3.41	16.44	174.73	14.91	24.86	6.00	230.30
CV (%)	60.38	53.74	60.18	46.59	47.63	85.84	124.73	98.93	106.75	67.23	111.51
<i>P. elliottii</i> and <i>P. taeda</i> (n = 70)											
Min.	2.00	1.91	2.35	1.26	1.78	0.20	0.40	0.10	0.19	0.35	1.23
Max.	24.00	40.27	30.30	7.08	14.93	58.04	687.39	65.91	96.14	39.54	889.81
Mean	11.44	20.10	15.15	3.36	7.33	20.13	124.90	17.34	21.75	10.58	194.70
SD	5.95	8.78	7.65	1.42	3.04	16.02	143.80	15.37	22.03	7.21	195.80
CV (%)	51.75	43.68	50.52	42.41	41.46	79.57	115.14	88.66	101.29	68.19	100.59

Min. is the minimum value; *Max.* is the maximum value; *SD* is the standard deviation; *CV* is the coefficient of variation; *I* is the age, in years; *D* is the diameter at 1.30 m height aboveground, in cm; *H* is the total height, in m; *dc* is the crown diameter, in m; *hc* is the crown height, in m; *WR* is the root biomass, in kg; *WS* is the stem biomass, in kg; *WB_a* is the bark biomass, in kg; *WB_r* is the branch biomass, in kg; *WN* is the needle biomass, in kg; *WT* is the total biomass, in kg.

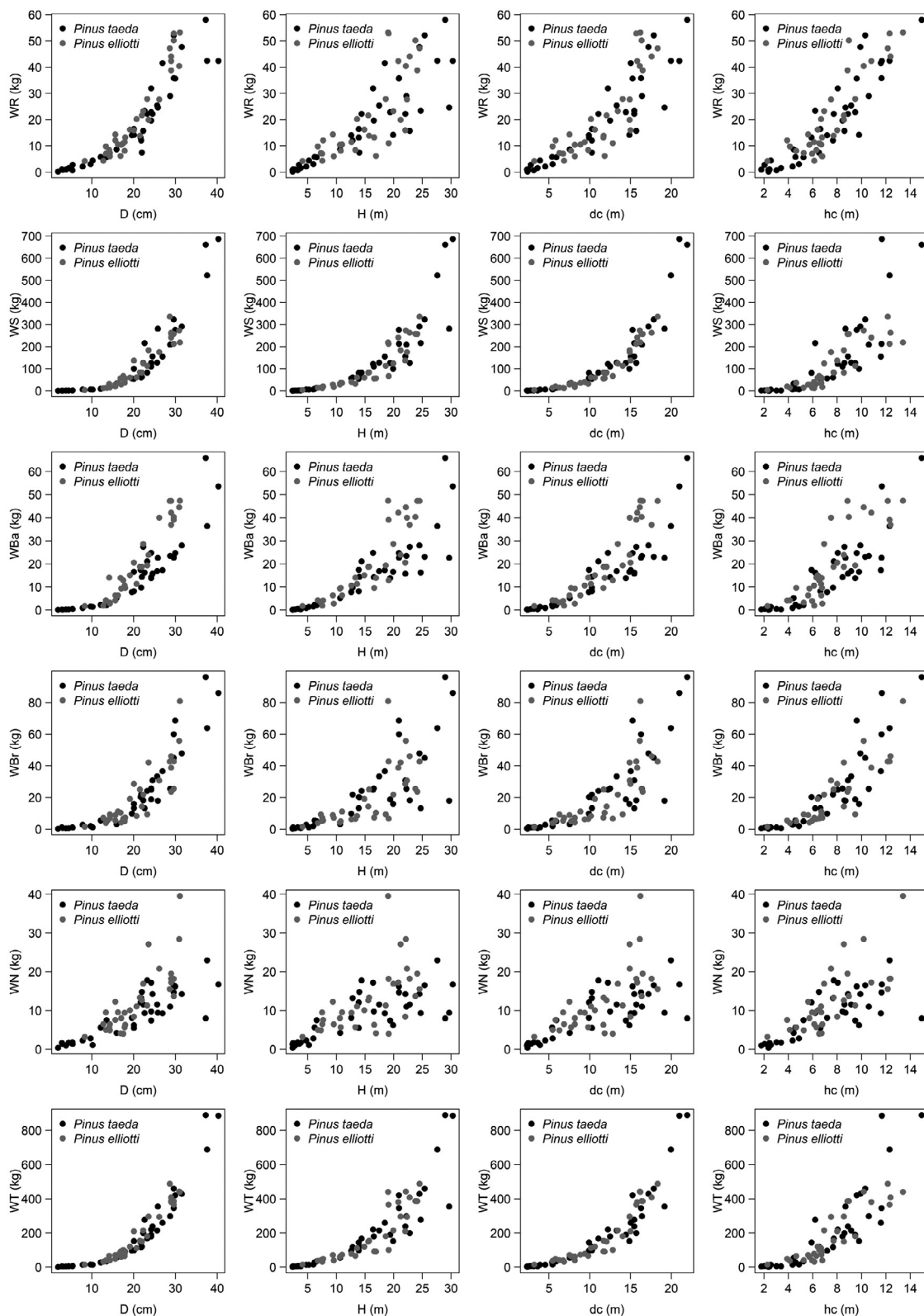


Figure 1. Scatter plots of the dependent variables as a function of the independent variables.

D is the diameter at 1.30 m height aboveground; H is the total height; dc is the crown diameter; hc is the crown height; WR is the root biomass; WS is the stem biomass; WB_a is the bark biomass; WBr is the branch biomass; WN is the needle biomass; WT is the total biomass

Model identity test

Given that the dataset consisted of two species, two non-parametric model identity tests, Chi-square test – χ^2 (2) and Kruskal-Wallis test – H (3), were performed prior to applying the procedures proposed in this study for biomass estimation. These tests aimed to evaluate whether it was necessary to separate or combine the species, that is, to determine whether biomass estimates could be derived from a generic equation or if species-specific equations were required. Although the scatter plots provided some indication, we chose to perform identity tests to avoid relying solely on visual assessment. The choice of non-parametric tests was motivated by common issues in biomass data modelling, such as non-normality and heteroscedasticity of the residuals, which were also observed in this manuscript. The test was conducted for total biomass, which represents the sum of all its individual parts. In Fig. 2, a simplified schematic representation is presented, illustrating the data stratification process and the model tested, both separately for each species and for both species together.

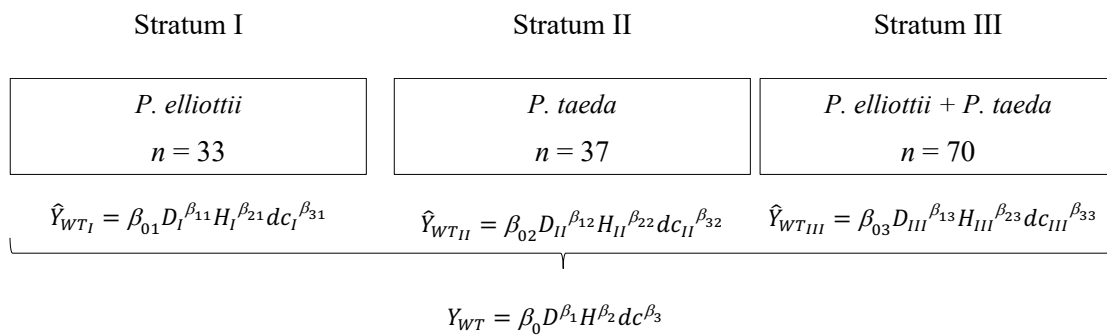


Figure 2. Representation of total biomass models ($\hat{Y}_{WT}$) fitted for each stratum (I, II, and III) based on dendrometric variables (D , H , and dc), considering stratum-specific models and the generic model for *Pinus elliottii* and *Pinus taeda*.

H_0 : There is no significant difference between the model fitted to the combined species (generic model) and the species-specific models.

H_1 : There is a significant difference between the model fitted to the combined species (generic model) and at least one of the species-specific models.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{critical} \text{ with } n - 1 \text{ df} \quad (2)$$

$$H = \frac{12}{N(N+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \dots + \frac{R_k^2}{n_k} \right) - 3(N+1) \sim \chi^2_{critical} \text{ with } k - 1 \text{ df} \quad (3)$$

where: O_i are the observed values, E_i are the expected values, n is the sample size; N is the total number of observations in all combined samples; k is the number of samples (stratum); R_k is the sum of the ranks of the k -th sample; n_k is the number of observations in the k -th sample; df is the degrees of freedom.

Ratio estimators for biomass

To estimate the biomass of *P. elliottii* and *P. taeda* by ratio estimators, the biomass of the parts (roots, stem, bark, branches, and needles) and total biomass were considered as dependent variables, and the stem volume of trees was considered as an independent variable. As the database did not contain data on the scaling of the trees, the stem volumes were estimated by integrating the [Preußner \(1974\)](#) taper curve equations fitted

by Stefanello et al. (2021) for stem volume estimation of *P. taeda* in the same region where the data for the present study were collected (Southern Brazil). The Preußner (1974) equations proposed by Stefanello et al. (2021) were adjusted for age classes of 5, 10, 15 and 20 years. It is important to highlight that in our database, three trees were more than 20 years old (with 22, 23 and 24 years old, respectively). To estimate the stem volume of these trees, the integral of the Preußner taper curve equation corresponding to the 20 years old class was used. The other alternative would be to exclude these trees from the database if they significantly compromise the precision of the estimates.

The choice of fitted equations using the methodology proposed by Preußner (1974) was due to the flexibility of the adjusted taper curves, the simplicity of volume calculations through integration and, above all, to guarantee the additivity of the volume of the parts (segments) with the total volume, which is the central theme of the present study.

The entire mathematical procedure and some examples of application of Preußner taper curve functions to estimate the stem volumes of the trees are presented in Stefanello et al. (2021) and Costa Junior et al. (2022).

Briefly, for a relative function, the integration can be performed to the relative variable “ x ” in the range $0 \leq x \leq 1$. In this case, the stem partitioning was designed to generate four functions. The first segment ranges from 0 to 25% of the stem, the second segment from 25% to 50%, the third segment from 50% to 75%, and the last one from 75% to 100%. These subdivisions coincide with the needs of volume estimates and other variables required for assortment studies (Stefanello et al., 2021). Thus, from (4) to (7), four functions were obtained to establish a continuous shape curve that connects the successive extremes of the quotients, namely:

$$y = a_{11} d_{0.5h} \frac{1}{\sqrt[w]{1+bx}} \text{ for } 0 < x \leq 0.25 \quad (4)$$

$$y = a_{21} d_{0.5h} \sqrt[w]{1-x} \text{ for } 0.25 < x \leq 0.5 \quad (5)$$

$$y = a_{31} d_{0.5h} \sqrt[w]{1-x} \text{ for } 0.5 < x \leq 0.75 \quad (6)$$

$$y = a_{41} d_{0.5h} \sqrt[w]{1-x} \text{ for } 0.75 < x \leq 1.0 \quad (7)$$

where: y is the diameter at position x ; $d_{0.5h}$ is the diameter at the middle of the stem; x is the relative length of the stem, that is, $x = \frac{h_x}{H}$, given that “ x ” exists in the real interval between 0.0 and 1.0, therefore, $\frac{0}{H} \leq x \leq \frac{h_x}{H}$; H is the tree total height; w is the power of the expression, being equal to one for obtaining a straight line and less than one for obtaining a classical parabola, that is, as it approaches zero, a strong curve is obtained, and as it approaches one, a weak curve is obtained; b is the accelerator of the parabola.

For practical application, the previous shape curve expressions were formulated into a single mathematical function or a series of successive functions along the stem. Operative transformations were applied to obtain the height at a given diameter, based on the diameter at breast height. These transformations are detailed in the work of Stefanello et al. (2021).

In the case of a relative function, therefore, the integration can be performed over the variation interval of the relative variable “ x ” or over the interval $0 \leq x \leq 1$. Since the functions in the variable “ y ” result in diameter, the integration, in order to obtain volume, will require multiplying by $\frac{\pi}{4}$ to calculate the area of the circle, then by height (H) to obtain the volume of the solid (v) (8), and taking it to “ y^2 ”, because for area calculation, the diameter is squared. Thus, we have:

$$v = \frac{\pi}{4} h \int_0^1 y^2 dx \quad (8)$$

From the obtained integrals, it was possible to estimate the total volume or the partial volume for a given desired interval. It should be considered that the integrations are performed assuming that the parts integrate with the inversion of values, as the curve has a decreasing shape from the base to the apex.

Measuring the diameter at the position corresponding to half the height ($d_{0.5h}$) is easily obtained if the tree is felled, however, this procedure is an expensive and time-consuming activity. Therefore, measuring $d_{0.5h}$ can be difficult to conduct the forest inventory (Costa Junior et al., 2022). A practical alternative is the use of taper curve as a function of diameter at breast height (D) (Stefanello et al., 2021).

The conversion of the functions to diameter at breast height will be carried out depending on the height of the tree. For trees with heights greater than 5.2 m, the conversion will take place in the first segment. In trees with heights less than 5.2 m, the conversion will occur in the second segment (Stefanello et al., 2021). For the present study, only a few trees in the 5-year class had conversion in the second segment, as most trees were over 5.2 m tall. For the remaining age classes (10, 15, and 20), the conversion occurred in the first segment.

The fitted taper curves as a function of diameter at breast height used in the present study to estimate the stem volume of the trees are presented in (9) to (28):

$$05 \text{ years} \begin{cases} 1^{st} y = a_{13.5}k \frac{1}{(1 + 100x)^{0.06}} \\ 2^{nd} y = k (1 - x)^{0.98} \\ 3^{rd} y = a_{33.5}k (1 - x)^{0.9} \\ 4^{th} y = a_{43.5}k (1 - x)^{0.874} \end{cases} \quad \text{for} \begin{cases} 0.00 < x \leq 0.25 \\ 0.25 < x \leq 0.50 \\ 0.50 < x \leq 0.75 \\ 0.75 < x \leq 1.00 \end{cases} \quad \begin{matrix} (9) \\ (10) \\ (11) \\ (12) \end{matrix}$$

$$10 \text{ years} \begin{cases} 1^{st} y = a_{13.10}k \frac{1}{(1 + 100x)^{0.06}} \\ 2^{nd} y = k (1 - x)^{0.9} \\ 3^{rd} y = a_{33.10}k (1 - x)^{0.88} \\ 4^{th} y = a_{43.10}k (1 - x)^{0.87} \end{cases} \quad \text{for} \begin{cases} 0.00 < x \leq 0.25 \\ 0.25 < x \leq 0.50 \\ 0.50 < x \leq 0.75 \\ 0.75 < x \leq 1.00 \end{cases} \quad \begin{matrix} (13) \\ (14) \\ (15) \\ (16) \end{matrix}$$

$$15 \text{ years} \begin{cases} 1^{st} y = a_{13.15}k \frac{1}{(1 + 100x)^{0.05}} \\ 2^{nd} y = k (1 - x)^{0.8} \\ 3^{rd} y = a_{33.15}k (1 - x)^{0.79} \\ 4^{th} y = a_{43.15}k (1 - x)^{0.78} \end{cases} \quad \text{for} \begin{cases} 0.00 < x \leq 0.25 \\ 0.25 < x \leq 0.50 \\ 0.50 < x \leq 0.75 \\ 0.75 < x \leq 1.00 \end{cases} \quad \begin{matrix} (17) \\ (18) \\ (19) \\ (20) \end{matrix}$$

$$20 \text{ years} \begin{cases} 1^{st} y = a_{13.20}k \frac{1}{(1 + 100x)^{0.046}} \\ 2^{nd} y = k (1 - x)^{0.78} \\ 3^{rd} y = a_{33.20}k (1 - x)^{0.77} \\ 4^{th} y = a_{43.20}k (1 - x)^{0.76} \end{cases} \quad \text{for} \begin{cases} 0.00 < x \leq 0.25 \\ 0.25 < x \leq 0.50 \\ 0.50 < x \leq 0.75 \\ 0.75 < x \leq 1.00 \end{cases} \quad \begin{matrix} (21) \\ (22) \\ (23) \\ (24) \end{matrix}$$

The resulted constants for:

5 years are:

$$a_{13.5} = 1.0530; a_{33.5} = 0.9461; a_{43.5} = 0.9126; k = D \frac{1}{(1-x)^{0.98}} \text{ for } 2^{nd} \text{ segment}; k = D(1 + bx)^{0.06} \text{ for the } 1^{st} \text{ segment} \quad (25)$$

10 years are:

$$a_{13.10} = 1.0530; a_{33.10} = 0.9862; a_{43.10} = 0.9727; k = D(1 + bx)^{0.06} \text{ for the } 1^{st} \text{ segment} \quad (26)$$

15 years are:

$$a_{13.15} = 1.0192; a_{33.15} = 0.9931; a_{43.15} = 0.9794; k = D(1 + bx)^{0.05} \text{ for the } 1^{st} \text{ segment} \quad (27)$$

20 years are:

$$a_{13.20} = 1.0061; a_{33.20} = 0.9931; a_{43.20} = 0.9794; k = D(1 + bx)^{0.046} \text{ for the } 1^{\text{st}} \text{ segment} \quad (28)$$

where: $b = 100$ is the accelerator of the parabolas; $x = \frac{1.3}{H}$ is the relative diameter at breast height.

Therefore, taking the functions for “y” (which describe the range from zero to one), for each age class, the stem volumes of the corresponding trees were obtained in (29) to (32).

For 5 years:

$$v = \frac{\pi}{4} H \left[\int_0^{0.25} a_{13.5} k^2 \frac{1}{(1+100x)^{0.12}} dx + \int_{0.25}^{0.5} k^2 (1-x)^{1.96} dx + \int_{0.5}^{0.75} a_{33.5} k^2 (1-x)^{1.8} dx + \int_{0.75}^{1.0} a_{43.5} k^2 (1-x)^{1.748} dx \right] \quad (29)$$

For 10 years:

$$v = \frac{\pi}{4} H \left[\int_0^{0.25} a_{13.10} k^2 \frac{1}{(1+100x)^{0.12}} dx + \int_{0.25}^{0.5} k^2 (1-x)^{1.8} dx + \int_{0.5}^{0.75} a_{33.10} k^2 (1-x)^{1.76} dx + \int_{0.75}^{1.0} a_{43.10} k^2 (1-x)^{1.74} dx \right] \quad (30)$$

For 15 years:

$$v = \frac{\pi}{4} H \left[\int_0^{0.25} a_{13.15} k^2 \frac{1}{(1+100x)^{0.1}} dx + \int_{0.25}^{0.5} k^2 (1-x)^{1.6} dx + \int_{0.5}^{0.75} a_{33.15} k^2 (1-x)^{1.58} dx + \int_{0.75}^{1.0} a_{43.15} k^2 (1-x)^{1.56} dx \right] \quad (31)$$

For 20 years:

$$v = \frac{\pi}{4} H \left[\int_0^{0.25} a_{13.20} k^2 \frac{1}{(1+100x)^{0.092}} dx + \int_{0.25}^{0.5} k^2 (1-x)^{1.56} dx + \int_{0.5}^{0.75} a_{33.20} k^2 (1-x)^{1.54} dx + \int_{0.75}^{1.0} a_{43.20} k^2 (1-x)^{1.52} dx \right] \quad (32)$$

For each age class, after integration, the total volume of any tree can be calculated using the resulting equations in (33) to (36):

For 5 years:

$$v = \frac{\pi}{4} H k^2 (0.34812) \quad (33)$$

For 10 years:

$$v = \frac{\pi}{4} H k^2 (0.35969) \quad (34)$$

For 15 years:

$$v = \frac{\pi}{4} H k^2 (0.38446) \quad (35)$$

For 20 years:

$$v = \frac{\pi}{4} H k^2 (0.39126) \quad (36)$$

where: 0.34812, 0.35969, 0.38446 and 0.39126 are the average Hohenald natural form factors calculated at the reference diameter $d_{0.5h}$, for the age classes 5, 10, 15 and 20 years, respectively.

The ratio is then defined as presented in (37). The entire procedures applied for its derivation are presented in P llico Netto & Behling (2019) and Trautenm ller et al. (2023).

$$\hat{R}_m = \frac{\sum_{i=1}^n Y_{mi}}{\sum_{i=1}^n X_i} \quad (37)$$

where: $\hat{R}_m$ is the estimated slope for each m^{th} part of biomass, in kg.m^{-3} ; Y_{mi} is the observed biomass of m^{th} part of i^{th} tree, in kg; and X_i is the estimated stem volume of i^{th} tree, in m^3 .

The biomass of the roots, stem, bark, branches, needles, and total biomass were estimated based on equation (38) with the slope for each m^{th} part of biomass, estimated previously in (37).

$$\hat{Y}_{mi} = \hat{R}_m \hat{v}_i \quad (38)$$

where: $\hat{Y}_{mi}$ is the estimated biomass of m^{th} part of i^{th} tree, expressed in kg; $\hat{R}_m$ is the estimated slope for each m^{th} part of biomass, expressed in kg.m^{-3} ; and $\hat{v}_i$ is the estimated stem volume of i^{th} tree, in m^3 .

Then, the variance (39), the standard deviation (40), and the standard error in percentage (41) were obtained to evaluate the performance of the ratio estimators, and the residual dispersion graphs were plotted to evaluate the occurrence or not of heteroscedasticity.

$$s_{\hat{Y}_m}^2 = \frac{N(N-n)}{n(n-1)} \sum_{m=1}^M (Y_{mi} - \hat{R}_m X_i)^2 \quad (39)$$

$$s_{\hat{Y}_m} = \sqrt{\frac{N(N-n)}{n(n-1)} \left(\sum_{m=1}^M Y_{mi}^2 - 2\hat{R}_m \sum_{m=1}^M Y_{mi} X_i + \hat{R}_m^2 \sum_{m=1}^M X_i^2 \right)} \quad (40)$$

$$s_{yx_m}(\%) = \frac{s_{\hat{Y}_m}}{\bar{y}_m} 100 \quad (41)$$

where: $s_{\hat{Y}_m}^2$ is the variance of the estimate of each biomass m^{th} , in kg^2 ; $s_{\hat{Y}_m}$ is the standard error for biomass m^{th} part, in kg; s_{yx_m} is the standard error of the estimate for the biomass m^{th} part, in %; n is the number of trees in the sampling; N is the total number of trees in the forest; $\hat{R}_m$ is the estimated slope for each m^{th} part of biomass, in kg.m^{-3} ; Y_{mi} is the biomass of m^{th} part of i^{th} tree, in kg; X_i is the estimated stem volume of i^{th} tree, in m^3 ; and $\bar{y}_m$ is the observed mean of m^{th} part of biomass, in kg.

The heteroscedasticity of the residuals was detected by graphical analysis of the residuals. Therefore, for its correction, data stratification was applied in two stages, as recommended by Trautenm ller et al. (2023), and new ratio estimates were obtained. In the first stage, stratification was based on diameter classes, to have a minimum of 23 trees per D class (Class I: 1.91–16.23 cm with 23 trees, Class II: 16.24–23.55 cm with 23 trees, and Class III: 23.56–40.27 cm with 24 trees). For this reason, it was not possible to maintain the same amplitude across the three diameter classes. This stratification based on D , established through the prior definition of the minimum number of trees per D class, is also referenced by Trautenm ller et al. (2023). In the second stage, another stratification within each diametric class was undertaken based on the estimated slope through the ratio between the observed total biomass (WT) and the estimated stem volume of the i^{th} tree ($\hat{v}_i$) (42), as shown in Fig. 3.

$$\hat{\beta}_{ih} = \frac{WT_{ih}}{\hat{v}_{ih}} \quad (42)$$

where: $\hat{\beta}_{ih}$ is the slope of i^{th} tree in the diameter class h ; WT_{ih} is the total biomass of the i^{th} tree in the diameter class h ; $\hat{v}_{ih}$ is the estimated stem volume of the i^{th} tree in the diameter class h .

Then the ratio estimators (43) should be evaluated in each stratum and specifically nominated for the total number of strata L , such that $h = 1, 2, 3 \dots L$. The entire mathematical procedure applied is presented in Trautenm ller et al. (2023).

$$\hat{R}_{ij} = \frac{\sum_{i=1}^{n_h} Y_{ij}}{\sum_{i=1}^{n_h} X_{ij}} \quad (43)$$

where: Y is the dependent variable of the model; X is the independent variable of the model; i indicates the tree in which the measurement of the variables is taken; j indicates the biomass of the tree parts in which the measurements of the variables are taken and h indicates the stratum where the trees are.

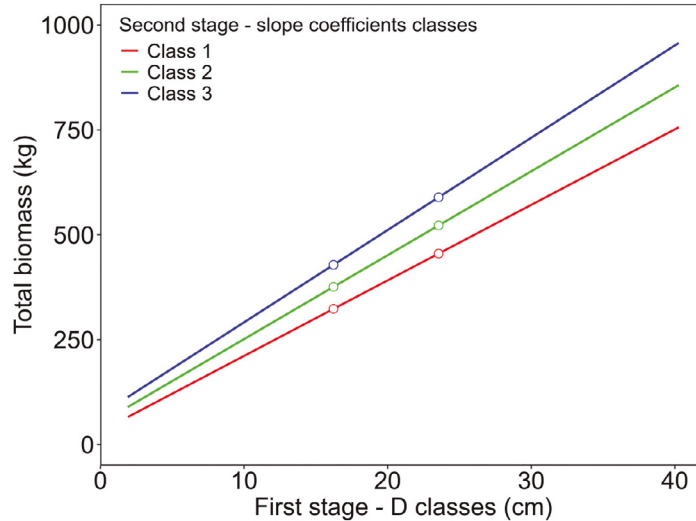


Figure 3. Schematic representation of the stratification process in two stages: stratification by diameter classes (D) (first stage of stratification) and stratification by estimated slope classes (second stage of stratification).

Source: Trautenmüller et al. (2023), with some modifications made by the authors.

A discriminant analysis was applied considering the following independent variables: diameter at 1.30 m height aboveground (D), total height (H), crown diameter (dc), crown height (hc), the ratio between crown diameter and diameter at 1.30 m aboveground (dc/D), the ratio between diameter at 1.30 m aboveground and total height (D/H), and tree age (I). The objective was to verify a posteriori classification of trees in the strata using the $\hat{\beta}_{ih}$ of each tree, obtained in (42), at the second stratification level. The discriminant function model (Hair et al., 2009) is presented in (44). The significance of the functions was assessed using the Wilks-Lambda value at a 95% probability level.

$$Z_{ik} = a + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} \quad (44)$$

where: Z_{ik} is the Z score of the discriminant function k for the i^{th} object; a is the intercept; β_i are the slope coefficients or discriminant weights for the independent j^{th} variable; and X_{ik} is the j^{th} independent variable for the i^{th} object.

In this procedure (stratification in two stages), the variance, standard deviation and standard error in percentage are presented from (45) to (47), respectively. The entire mathematical procedure applied is presented in Trautenmüller et al. (2023).

$$S^2_{\hat{Y}_{jh}} = \frac{(1-n_h f_h)}{n_h(n_h-1)} \sum_{i=1}^{n_h} (Y_{ijh} - \hat{R}_{jh} X_{ijh})^2 \quad (45)$$

$$S_{\hat{Y}_{jh}} = \sqrt{\frac{(1-n_h f_h)}{n_h(n_h-1)}} \sqrt{\frac{\sum_{i=1}^{n_h} Y_{ijh}^2 - 2\hat{R}_{jh} \sum_{i=1}^{n_h} Y_{ijh} X_{ijh} + \hat{R}_{jh}^2 \sum_{i=1}^{n_h} X_{ijh}^2}{n_h-1}} \quad (46)$$

$$S_{yx_{jh}}(\%) = \frac{t_h S_{\hat{Y}_{jh}}}{\hat{Y}_{ij}} 100 \quad (47)$$

where: $S^2_{\hat{Y}_{jh}}$ is the variance for the estimate of the j^{th} biomass parts, in stratum h for the slope coefficient, in kg^2 ; $S_{\hat{Y}_{jh}}$ is the standard error of the estimate for the j^{th} biomass part, in stratum h for the slope coefficient, in kg ; $S_{y_{x_{jh}}}$ is the standard error of the estimate for the j^{th} biomass part, in stratum h for the slope coefficient, in $\%$; n_h is the number of trees sampled in stratum h ; $\hat{R}_{jh}$ is the slope coefficient estimated for the j^{th} biomass part of each stratum h in kg.m^{-3} ; Y_{ijh} is the j^{th} biomass part, in stratum h for the slope coefficient in the i^{th} tree, in kg ; X_{ijh} is the estimated volume of the j^{th} part, in stratum h for the slope coefficient in the i^{th} tree, in m^3 ; f_h is the correction factor as a function of the number of trees in each stratum h for the slope coefficient ($f_h = n \cdot N^{-1}$); t_h is Student's t in stratum h for the slope coefficient.

Control procedure: simultaneous fitting (WNSUR)

Firstly, nine nonlinear allometric models were independently fitted for each biomass part (Table 2). The best model was selected based on the following statistical criteria: adjusted coefficient of determination (R^2_{Adj}), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and standard error of estimation (S_{yx}). The models and the quality assessment statistics considered are referenced in similar studies: Parresol (2001), Sanquetta et al. (2015), Magalhães & Seifert (2015), Dong et al. (2014, 2015, 2018), Oliveira et al. (2018). Afterwards, the nonlinear seemingly unrelated regression ($NSUR$) method was applied to force additivity based on the procedures proposed by Parresol (2001). In (48), the system of equations for simultaneous fitting is presented.

$$\begin{cases} \hat{Y}_{\text{roots}} = \beta_{11} D^{\beta_{12}} h c^{\beta_{13}} \\ \hat{Y}_{\text{stem}} = \beta_{21} D^{\beta_{22}} H^{\beta_{23}} \\ \hat{Y}_{\text{bark}} = \beta_{31} (D^2 H)^{\beta_{32}} \\ \hat{Y}_{\text{branches}} = \beta_{41} D^{\beta_{42}} h c^{\beta_{43}} \\ \hat{Y}_{\text{needles}} = \beta_{51} D^{\beta_{52}} H^{\beta_{53}} \\ \hat{Y}_{\text{total}} = \beta_{11} D^{\beta_{12}} h c^{\beta_{13}} + \beta_{21} D^{\beta_{22}} H^{\beta_{23}} + \beta_{31} (D^2 H)^{\beta_{32}} + \beta_{41} D^{\beta_{42}} h c^{\beta_{43}} + \beta_{51} D^{\beta_{52}} H^{\beta_{53}} \end{cases} \quad (48)$$

where: $\hat{Y}_{\text{roots}}$ is the estimated root biomass, in kg , $\hat{Y}_{\text{stem}}$ is the estimated stem biomass, in kg , $\hat{Y}_{\text{bark}}$ is the estimated bark biomass, in kg , $\hat{Y}_{\text{branches}}$ is the estimated branch biomass, in kg , $\hat{Y}_{\text{needles}}$ is the estimated needle biomass, in kg , $\hat{Y}_{\text{total}}$ is the estimated total biomass, in kg , D is the diameter at 1.30 m height aboveground, in cm , hc is the crown height, in m , and H is the total height, in m .

Table 2. Allometric models tested for biomass estimation in mixed stands of *Pinus elliottii* and *Pinus taeda* in Southern Brazil.

No.	Model
1	$Y = \beta_0 D^{\beta_1} + \epsilon_i$
2	$Y = \beta_0 (D^2)^{\beta_1} + \epsilon_i$
3	$Y = \beta_0 (D^2 H)^{\beta_1} + \epsilon_i$
4	$Y = \beta_0 D^{\beta_1} H^{\beta_2} + \epsilon_i$
5	$Y = \beta_0 D^{\beta_1} H^{\beta_2} d c^{\beta_3} + \epsilon_i$
6	$Y = \beta_0 D^{\beta_1} h c^{\beta_2} + \epsilon_i$
7	$Y = \beta_0 H^{\beta_1} d c^{\beta_2} + \epsilon_i$
8	$Y = \beta_0 D^{\beta_1} d c^{\beta_2} + \epsilon_i$
9	$Y = \beta_0 D^{\beta_1} h c^{\beta_2} d c^{\beta_3} + \epsilon_i$

Y is the biomass of the m^{th} part, in kg ; D is the diameter at 1.30 m height aboveground, in cm ; H is the total height, in m ; dc is the crown diameter, in m ; hc is the crown height, in m ; β_0 ; β_1 ; β_2 ; β_3 are the regression coefficients, ϵ_i = random error.

In this procedure, to formulate the null hypothesis, it was assumed that the variance of the residuals is homogeneous, tested by White's test (1980) and by graphical analysis of the residuals. If heteroscedasticity of the residuals was detected, the equations in (48) were weighted by modeling the variance structure (Parresol, 1999; Parresol, 2001; Behling et al., 2018; Trautenmüller et al., 2021) to obtain the weighted simultaneous fitting (*WNSUR*). Weighting was applied as described in (49) and (50) for independent and simultaneous fitting, respectively.

$$\hat{\beta} = (X' \Psi (\hat{\theta})^{-1} X)^{-1} X' \Psi (\hat{\theta})^{-1} y \quad (49)$$

$$\hat{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \quad (50)$$

where: $\Psi(\hat{\theta})$ is the diagonal matrix of weightings, which depends on the number of parameters (P) denoted by the vector θ of the order $(P \cdot I)$. The matrix Ω , instead of being composed of weightings, as in the traditional form of weighted regression, contains elements of a residual's matrix. X is the matrix of independent variables and y is a vector of the dependent variable.

The performance of additive equations in (48) was evaluated based on the following statistics: *Pseudo* – $R^2\%$ (51), absolute standard error of the estimate – S_{yx} (52), standard error of the estimate in percentage – $S_{yx}\%$ (53) and graphical analysis of residuals.

$$Pseudo - R^2(\%) = r_{y_m \hat{y}_m}^2 100 \quad (51)$$

$$S_{yx_m} = \sqrt{\frac{\sum_{i=1}^n (y_{im} - \hat{y}_{im})^2}{n-p}} \quad (52)$$

$$S_{yx_m}(\%) = \frac{S_{yx_m}}{\bar{y}_m} \quad (53)$$

where: $r_{y_m \hat{y}_m}^2$ is the correlation coefficient between the observed biomass matrix of the m^{th} part and the estimated biomass matrix of the m^{th} part; S_{yx_m} is the standard error of the estimate of the m^{th} biomass part; y_{im} is the m^{th} biomass observed on i^{th} tree; $\hat{y}_{im}$ is the m^{th} estimated biomass of i^{th} tree; $\bar{y}_m$ is the mean of the m^{th} biomass part; n is the number of observations; p is the number of parameters.

Comparison between ratio estimators and regression estimators (*WNSUR*)

The results of the ratio estimators and *WNSUR* estimates were compared using random distortion – *Bias*, in kg (54), mean absolute error – *MAE*, in kg (55), mean square error – *MSE*, in kg^2 (56) and root-mean-square error – *RMSE*, in kg (57).

$$Bias = \frac{1}{n} \sum_{i=1}^n (y_{ijm} - \hat{y}_{ijm}) \quad (54)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{ijm} - \hat{y}_{ijm}| \quad (55)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{ijm} - \hat{y}_{ijm})^2 \quad (56)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{ijm} - \hat{y}_{ijm})^2} \quad (57)$$

where: y_{ijm} is the m^{th} biomass observed on i^{th} tree, in j^{th} diameter class, in kg, $\hat{y}_{ijm}$ is the m^{th} estimated biomass of i^{th} tree, in j^{th} diameter class, in kg, and n is the number of observations in each j^{th} diameter class.

Finally, the performance of the biomass estimators was evaluated by the behavior of the fitting statistics defined in (54) to (57), by calculating the differences between the statistics obtained through the ratio estimators and those obtained by the *WNSUR* estimators (58).

$$D(\%) = \frac{(\hat{S}_{mh} - \hat{U}_{mh})}{\hat{S}_{mh}} 100 \quad (58)$$

where: D is the difference (in %) between the statistics obtained by ratio and regression estimators; $\hat{s}_{mh}$ are the statistics obtained by ratio estimators of each m^{th} biomass part and h stratum; $\hat{u}_{mh}$ are the statistics obtained by regression estimators of each m^{th} biomass part and h stratum.

In Fig. 4, the methodological summary for biomass estimation using ratio estimators and simultaneous regression estimators is presented.

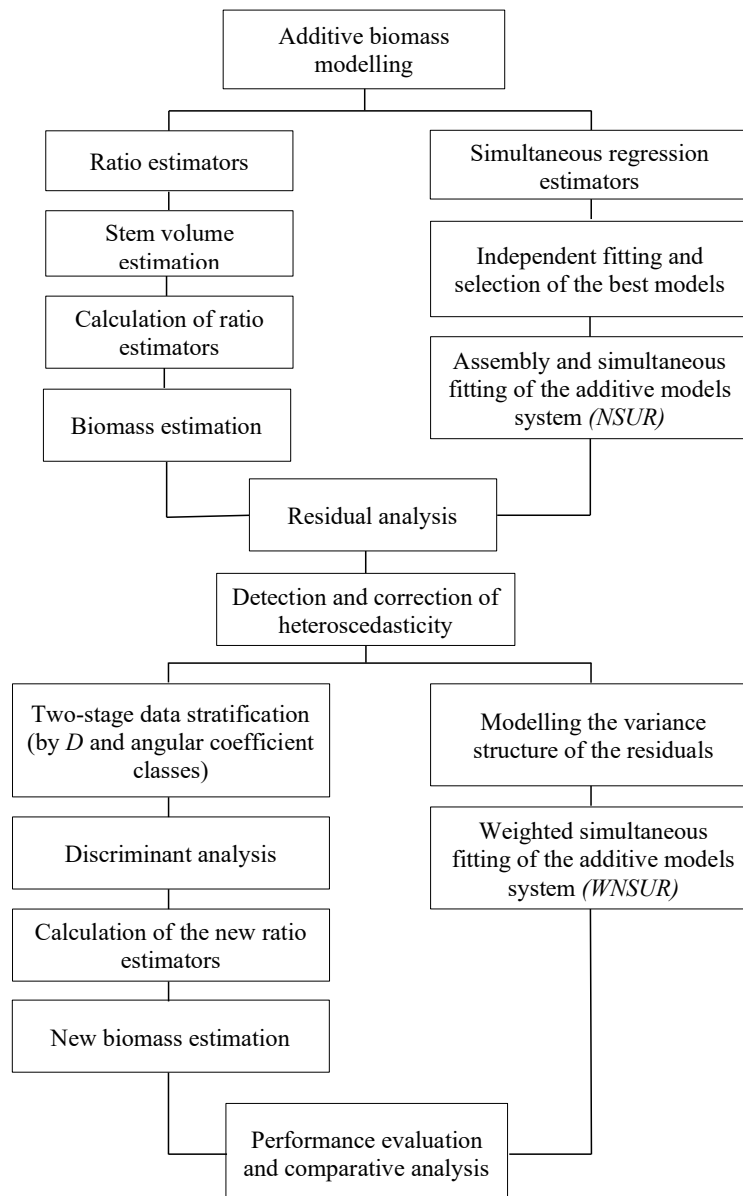


Figure 4. Flowchart of the methodology for biomass estimation of *Pinus elliottii* and *Pinus taeda*: ratio estimators and simultaneous regression estimators (*WNSUR*).

All statistical analyses involving the ratio estimators were conducted using *Microsoft Excel 2010*. Simultaneous fitting, both with and without weighting (*NSUR* and *WNSUR*), was performed using *SAS OnDemand for Academics (SAS Institute Inc., 2012)*. Discriminant analysis was carried out with the *SPSS* statistical software (IBM SPSS Statistics 21). All the graphs presented in this manuscript were generated using the *R* statistical software, except for the graphs of estimated biomass over observed biomass in different diameter strata in the ratio procedure, which were generated in *Microsoft Excel 2010*.

Results

Model identity test

The model identity tests between the specific models for *P. elliottii* and *P. taeda*, and the generic model for total biomass estimation (Table 3), showed no significant differences. The Chi-Square test results for the comparisons between the specific models and the generic model for *P. elliottii* (45.53, p -value = 0.98) and *P. taeda* (36.32, p -value = 0.99) indicated no statistical evidence to reject the null hypothesis that the models are similar. Additionally, the Kruskal-Wallis test for the comparison between all three models yielded a value of 0.45 (p -value = 0.79), further reinforcing the absence of significant differences among them. This result is visually supported by Fig. 5, which compares the three models and shows that they had the same mean rank.

Table 3. Chi-Square and Kruskal-Wallis model identity test for total biomass estimation.

Test	Model	Statistical test	p -value	Decision on H_0
Chi-Square	$\hat{Y}_{P.elliottii}$ X $\hat{Y}_{Gen.}$	45.53 ^{ns}	0.98	Not rejected
	$\hat{Y}_{P.taeda}$ X $\hat{Y}_{Gen.}$	36.32 ^{ns}	0.99	
Kruskal-Wallis	$\hat{Y}_{P.elliottii}$ X $\hat{Y}_{P.taeda}$ X $\hat{Y}_{Gen.}$	0.45 ^{ns}	0.79	

$\hat{Y}_{P.elliottii}$ is the specific model for *Pinus elliottii*; $\hat{Y}_{P.taeda}$ is the specific model for *Pinus taeda*; and $\hat{Y}_{Gen.}$ is the generic model (for both species together).

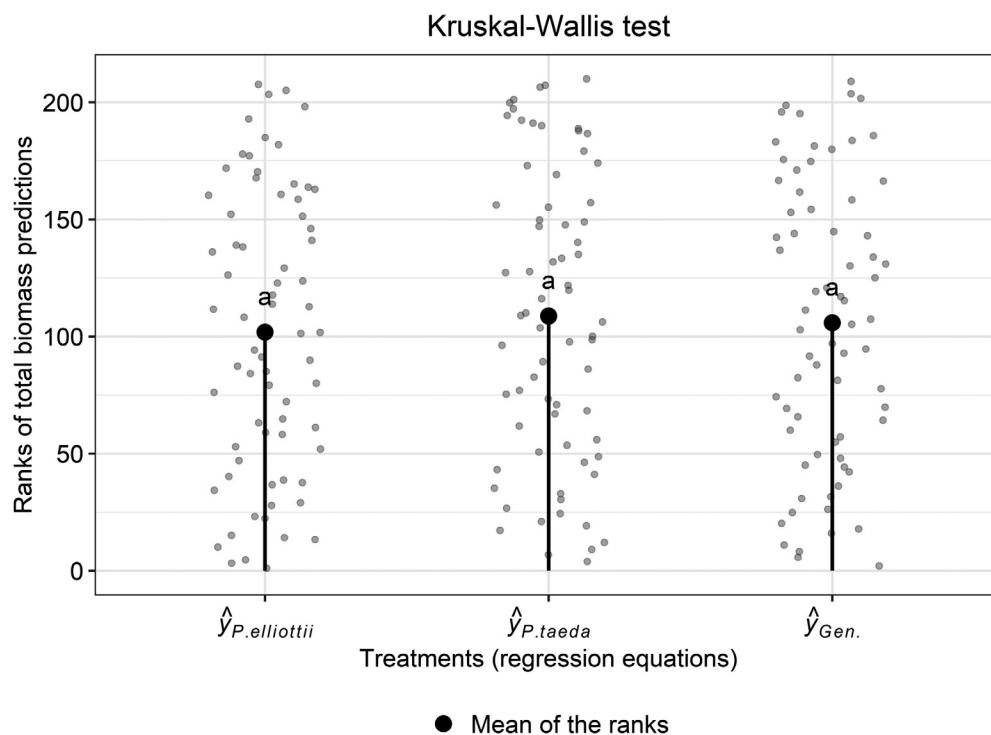


Figure 5. Comparison between the specific and generic models for total biomass estimation based on the results of the Kruskal-Wallis test.

The feasibility of unifying the data from both species was previously evidenced by the scatter plots presented in Fig. 1 (described in the Materials and Methods section), which revealed similar patterns in the nonlinear relationships between the independent variables and total biomass, as well as its parts. The overlap

in the data, with no clear distinctions in these relationships, suggested that a single model could be used to estimate the biomass of both species, simplifying the modelling process without compromising accuracy.

Ratio estimators for biomass

Table 4 presents the ratio estimators for each biomass part and their respective precision (%) obtained from fitting the entire database. The graphical analysis of the residuals' dispersion (Fig. 6) revealed the presence of heteroscedasticity, with evident variations in its dispersion according to the estimated biomass of each part and the total biomass.

Table 4. Ratio estimates for the entire database and their respective precision for roots, stem, bark, branches, needles biomass parts, and total biomass of *Pinus elliottii* and *Pinus taeda*.

Parts	Ratio estimates (kg.m^{-3})	Precision ($S_{yx}\%$)
Roots	56.87	12.31
Stem	352.71	5.28
Bark	48.98	11.37
Branches	61.45	11.07
Needles	29.89	19.14
Total	549.91	3.73

S_{yx} is the standard error of the estimate.

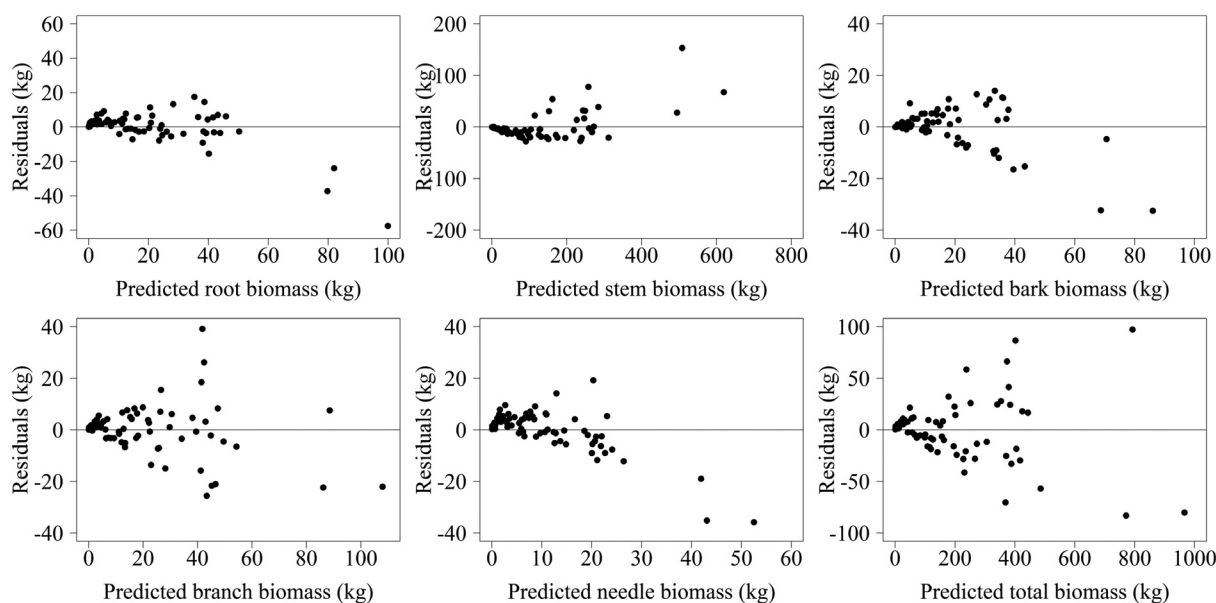


Figure 6. Graphical distribution of residuals for roots, stem, bark, branches, needles, and total biomass obtained from the entire database of *Pinus elliottii* and *Pinus taeda*.

Table 5 shows the results of the discriminant analysis. In general, this analysis provided satisfactory results, with success rates ranging from 79.20% to 91.30%, the highest for the first diameter class (1.91–16.23 cm) and the smallest for the last diameter class (23.56 – 40.27 cm). The variance explained only by the first discriminant function was also satisfactory, ranging from 68.98% to 95.13% (Fig. 7).

Table 5. Discriminant functions and their respective significance for each *D* class and hit percentage in the second stratification stage by the slope coefficient between the total biomass and the estimated stem volume of *Pinus elliottii* and *Pinus taeda*.

<i>D</i> classes (cm)	Discriminant functions	Wilks – Lambda	PC (%)
1.91 – 16.23	$Z = -17.58 - 0.02D + 0.73H - 1.99dc + 0.28hc + 9.98 + 7.44 + 0.37I$	0.149*	91.30
	$Z = -6.07 + 0.08D - 0.52H - 1.16dc - 0.09hc + 9.38 + 1.41 + 1.35I$	0.814 ^{ns}	
16.24 – 23.55	$Z = -31.10 + 0.98D - 0.32H - 10.98dc + 1.34hc + 216.66 + 4.01 + 0.21I$	0.343 ^{ns}	82.60
	$Z = -9.25 - 0.47D + 0.60H - 2.13dc - 0.25hc + 62.17 + 5.67 + 0.05I$	0.868 ^{ns}	
23.56 – 40.27	$Z = -18.48 + 1.26D + 0.05H - 5.67dc - 0.50hc + 131.51 - 2.41 - 0.32I$	0.398 ^{ns}	79.20
	$Z = -39.96 - 1.59D + 1.65H + 1.08dc - 0.31hc - 42.25 + 36.51 + 0.38I$	0.725 ^{ns}	

Z is the discriminating score; *D* is the diameter at 1.30 m height aboveground; *H* is the total height; *dc* is the crown diameter; *hc* is the crown height; *I* is the age of trees; *PC* is the hit percentage; ^{ns} is non-significant at 95% probability; * is significant at 95% probability.

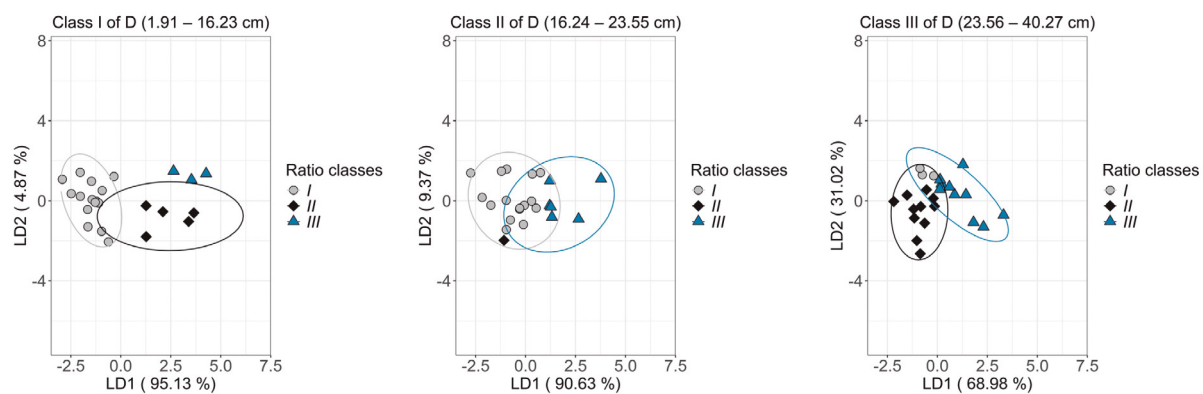


Figure 7. Graphical presentation of the discriminant analysis result: variances explained by the first (*LD1*) and second (*LD2*) discriminant function.

After the two-stage stratification, new biomass ratio estimates were obtained, as presented in Table 6. With double stratification, the performance of the ratio estimators improved significantly, with precision ranging from 1.31% (total biomass) to 5.40% (needle biomass) (Table 7). The estimated values are close to the observed values, especially for stem and total biomass, further confirming that the double stratification was indeed efficient (Suppl. Fig. S1, S2, and S3).

Table 6. Ratio estimates for the biomass of each part and total biomass of *Pinus elliottii* and *Pinus taeda* obtained after the stratification of database in diameter class at 1.30 m height aboveground (1st stratification stage) and slope coefficient between total biomass and estimated stem volume (2nd stratification stage).

Stratification stage		Ratio estimates (kg.m ⁻³)					
1 st	2 nd	Roots	Stem	Bark	Branches	Needles	Total
1.91 – 16.23	1	122.69	285.67	67.37	90.73	101.84	668.31
	2	247.95	325.95	61.32	154.33	212.30	1001.86
	3	363.30	286.57	71.99	330.73	566.63	1619.22
16.24 – 23.55	1	58.30	288.58	57.73	48.93	37.13	490.67
	2	67.02	300.20	61.29	66.48	46.09	541.09
	3	57.84	381.16	61.63	72.94	43.84	617.42
23.56 – 40.27	1	42.55	338.04	30.45	48.37	18.11	477.52
	2	47.35	349.92	42.49	52.47	19.40	511.64
	3	61.55	393.34	53.21	72.80	26.21	607.10

Table 7. Precision of the ratio estimators obtained after two-stage stratification for the biomass of roots, stem, bark, branches, needles, and total biomass of *Pinus elliottii* and *Pinus taeda*.

Method	Standard error of the estimate (%)					
	Roots	Stem	Bark	Branches	Needles	Total
Ratio	4.40	2.05	4.75	5.04	5.40	1.31

Control procedure: simultaneous fitting (WNSUR)

Table 8 presents the regression coefficients of the equations simultaneously fitted for estimating total biomass and its parts, as well as the result of White's test (1980), used to assess the homogeneity of the variance of the residuals. Based on the graphical analysis of the residuals (Fig. 8 A, B, C, D, E, and F), it was found that for the biomass of all parts and the total biomass, in the unweighted simultaneous fitting method (*NSUR*), the variability of the residuals increases as the predicted values grow, meaning that the residual variance is not constant across the fitted values, suggesting a violation of the homoscedasticity assumption. This evidence was confirmed by White's test (1980), at 95% probability, except for total biomass, where the test was not significant. The equations, after being weighted by the variance structure modelling in the weighted simultaneous fitting (*WNSUR*), showed homogeneity of the residuals, as corroborated by the results of White's test (1980) in Table 8 and the graphical analysis of the residuals (Fig. 8 G, H, I, J, K, and L).

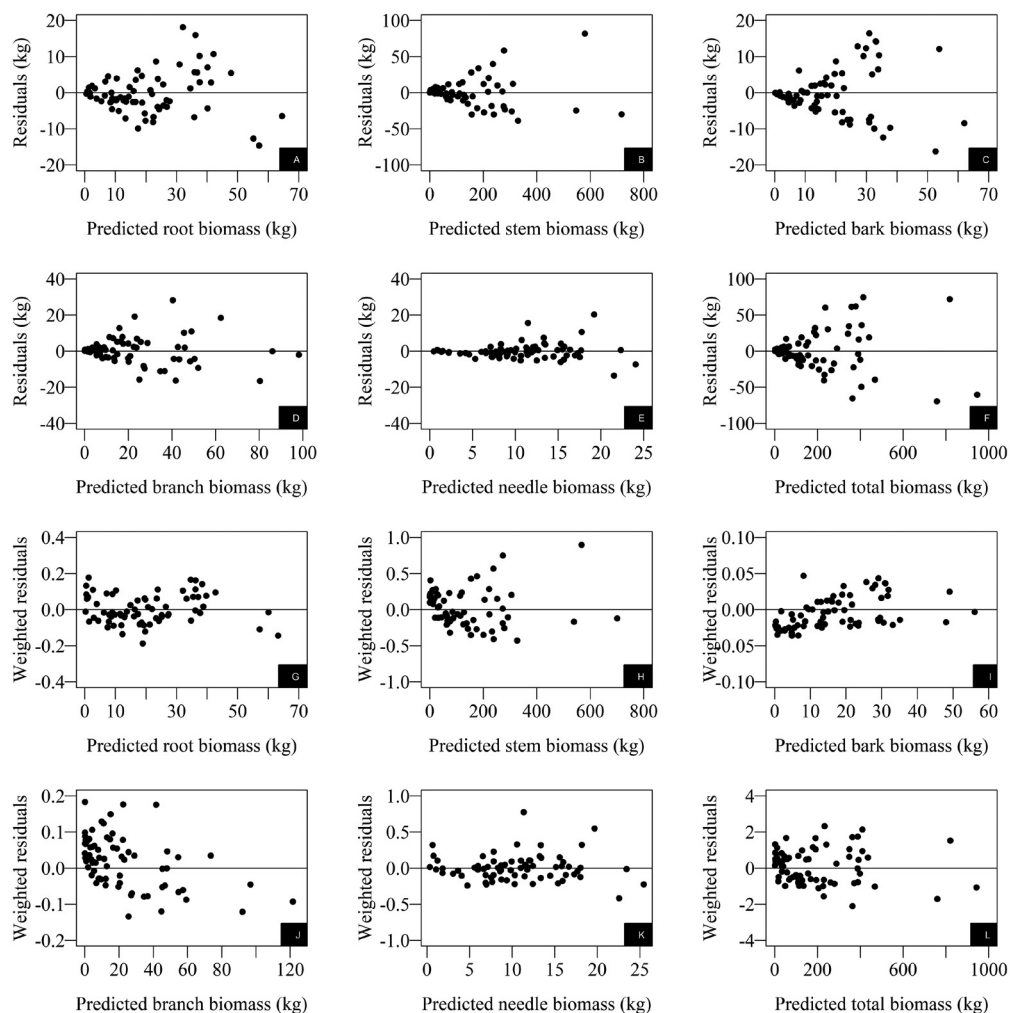
Table 8. Regression coefficients and results of the White's test for simultaneous unweighted fitting (*NSUR*) and with weighting (*WNSUR*) methods for *Pinus elliottii* and *Pinus taeda*.

Parts	b_{i1}	b_{i2}	b_{i3}	White
	Unweighted simultaneous fitting (<i>NSUR</i>)			
Roots	0.105*	1.132*	0.858*	31.14*
Stem	0.004*	1.833*	1.537*	48.63*
Bark	0.030 ^{ns}	0.706*		46.40*
Branches	0.007 ^{ns}	1.824*	1.112*	18.86*
Needles	0.226 ^{ns}	1.689*	-0.461 ^{ns}	25.44*
Total				65.01 ^{ns}

Table 8. Regression coefficients and results of the White's test for simultaneous unweighted fitting (*NSUR*) and with weighting (*WNSUR*) methods for *Pinus elliottii* and *Pinus taeda* (cont.)

Parts	b_{i1}	b_{i2}	b_{i3}	White
	Weighted simultaneous fitting (<i>WNSUR</i>)			
Roots	0.058*	1.678*	0.323 ^{ns}	10.83 ^{ns}
Stem	0.005*	1.894*	1.403*	29.32 ^{ns}
Bark	0.042*	0.666*		3.18 ^{ns}
Branches	0.002 ^{ns}	1.854*	1.511*	7.93 ^{ns}
Needles	0.151*	1.811*	-0.459*	4.02 ^{ns}
Total				62.40 ^{ns}

* significant at 95% probability; ^{ns} is not significant at 95% probability; b_{i1} , b_{i2} , and b_{i3} are the regression coefficients; $i = 1, 2, 3, 4$, and 5 for roots, stem, bark, branches and needles, respectively.

**Figure 8.** Residuals as a function of biomass estimated by simultaneous unweighted fitting (*NSUR*) method: (A) root biomass, (B) stem biomass, (C) bark biomass, (D) branch biomass, (E) needle biomass, and (F) total biomass, and by weighted simultaneous fitting (*WNSUR*) method: (G) root biomass, (H) stem biomass, (I) bark biomass, (J) branch biomass, (K) needle biomass, and (L) total biomass of *Pinus elliottii* and *Pinus taeda*.

The fitting statistics of the unweighted simultaneous fitting (*NSUR*) and the weighted simultaneous fitting (*WNSUR*) are shown in Table 9. In the same table, adequate statistics are observed for total biomass and its parts, except for bark, branches, and needles, which presented a standard error of estimate above 30%. It is worth noting that the regression coefficients and fit statistics of the nine models tested for total biomass and its parts can be found in Supplementary Table S1, and the best models selected for each biomass part to form the system of additive equations are highlighted in grey.

Table 9. Estimated statistics obtained with unweighted simultaneous fitting (*NSUR*) and with weighted simultaneous fitting (*WNSUR*) methods for *Pinus elliottii* and *Pinus taeda*.

Parts	<i>Pseudo-R</i> ² (%)	<i>S</i> _{yx} (kg)	<i>S</i> _{yx} (%)	Weight functions
Unweighted simultaneous fitting (<i>NSUR</i>)				
Roots	87.21	5.76	28.59	
Stem	98.32	18.71	14.98	
Bark	81.99	6.53	37.65	
Branches	88.65	7.49	34.42	
Needles	58.88	4.65	43.95	
Total	97.96	29.33	15.07	
Weighted simultaneous fitting (<i>WNSUR</i>)				
Roots	85.77	6.07	30.15	$\sigma^2_{WR} = D^{2,338} hc^{0,533}$
Stem	98.29	19.05	15.26	$\sigma^2_{WS} = D^{-0,076} H^{2,848}$
Bark	82.18	6.66	38.44	$\sigma^2_{WBa} = (D^2 H)^{1,231}$
Branches	87.75	8.91	40.98	$\sigma^2_{WBr} = D^{1,825} hc^{1,709}$
Needles	58.42	4.67	44.15	$\sigma^2_{WN} = D^{3,537} H^{1,687}$
Total	98.07	28.53	14.66	$\sigma^2_{WT} = D^{-1,168} H^{2,947} dc^{1,117}$

*Pseudo-R*² is the modified *R-squared*; *S*_{yx} is the standard error of the estimate; σ^2 is the residual variance for the estimation of the biomass of each component (*WR* – roots, *WS* – stem, *WBa* – bark, *WBr* – Branches, *WN* – needles) and total biomass (*WT*); *D* is the diameter at 1.30 m height aboveground; *H* is the total height; *dc* is the crown diameter; *hc* is the crown height.

Comparative analysis: ratio estimators and regression estimators (*WNSUR*)

Analysing the precisions of the estimates, the ratio estimators provided significant gains in precision for the estimation of total biomass and its parts, with precisions remaining below 6% (Table 7), compared to the *WNSUR* estimators, which had standard errors ranging from 14.66% to 44.15% (Table 9).

Figures 9, 10, and 11 present the comparative analysis between the ratio estimators and the *WNSUR* estimators through the graphical distribution of the residuals across the three diameter strata considered. In general, the ratio estimators exhibited a smaller range of residual dispersion.

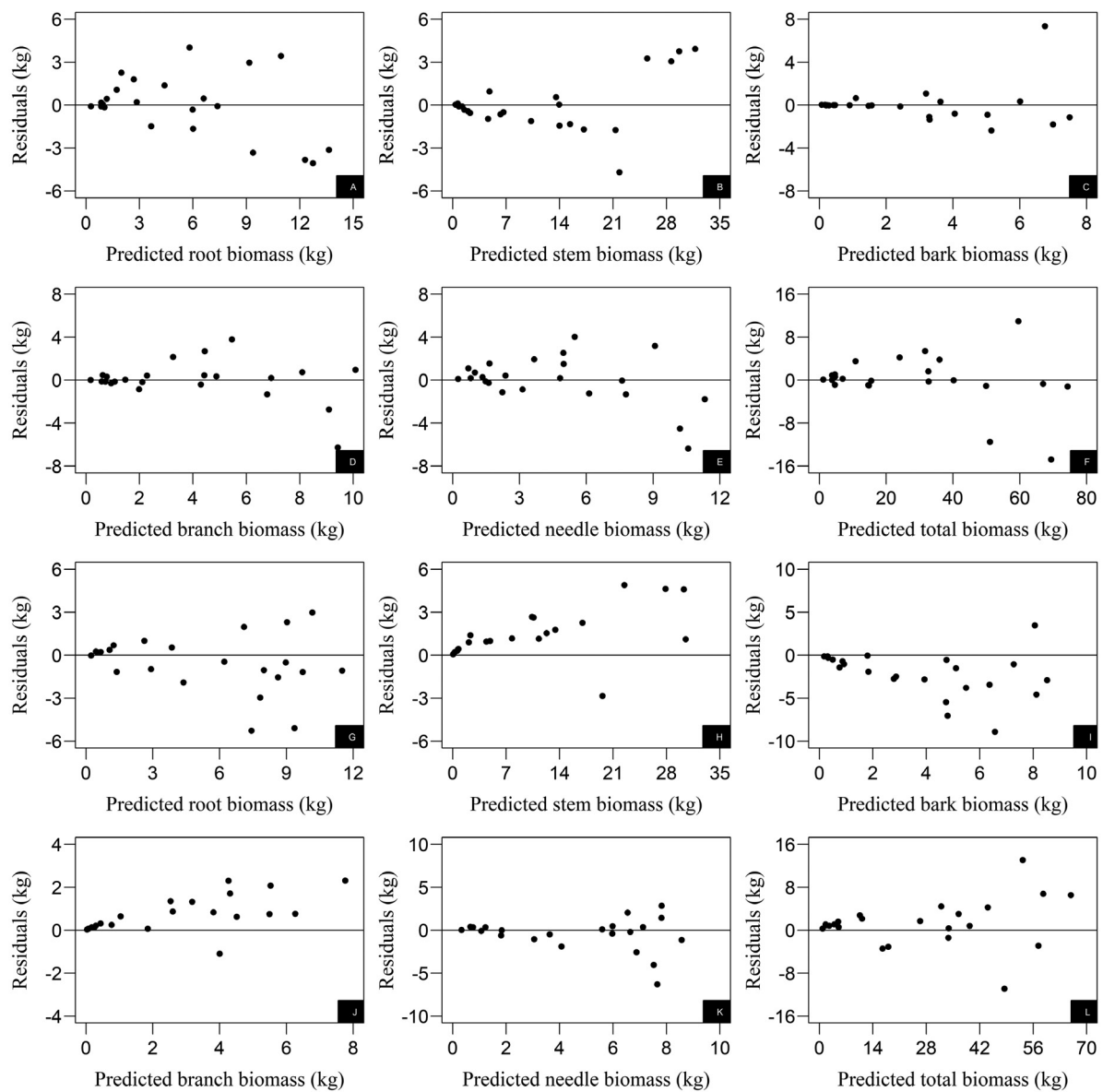


Figure 9. Comparison between ratio estimators (A, B, C, D, and F) and *WNSUR* procedure (G, H, I, J, K, and L) regarding the dispersion of residuals as a function of the estimated biomass for the parts (roots, stem, bark, branches, and needles) and total biomass in the diameter class I (1.91–16.23 cm), for *Pinus elliottii* and *Pinus taeda*.

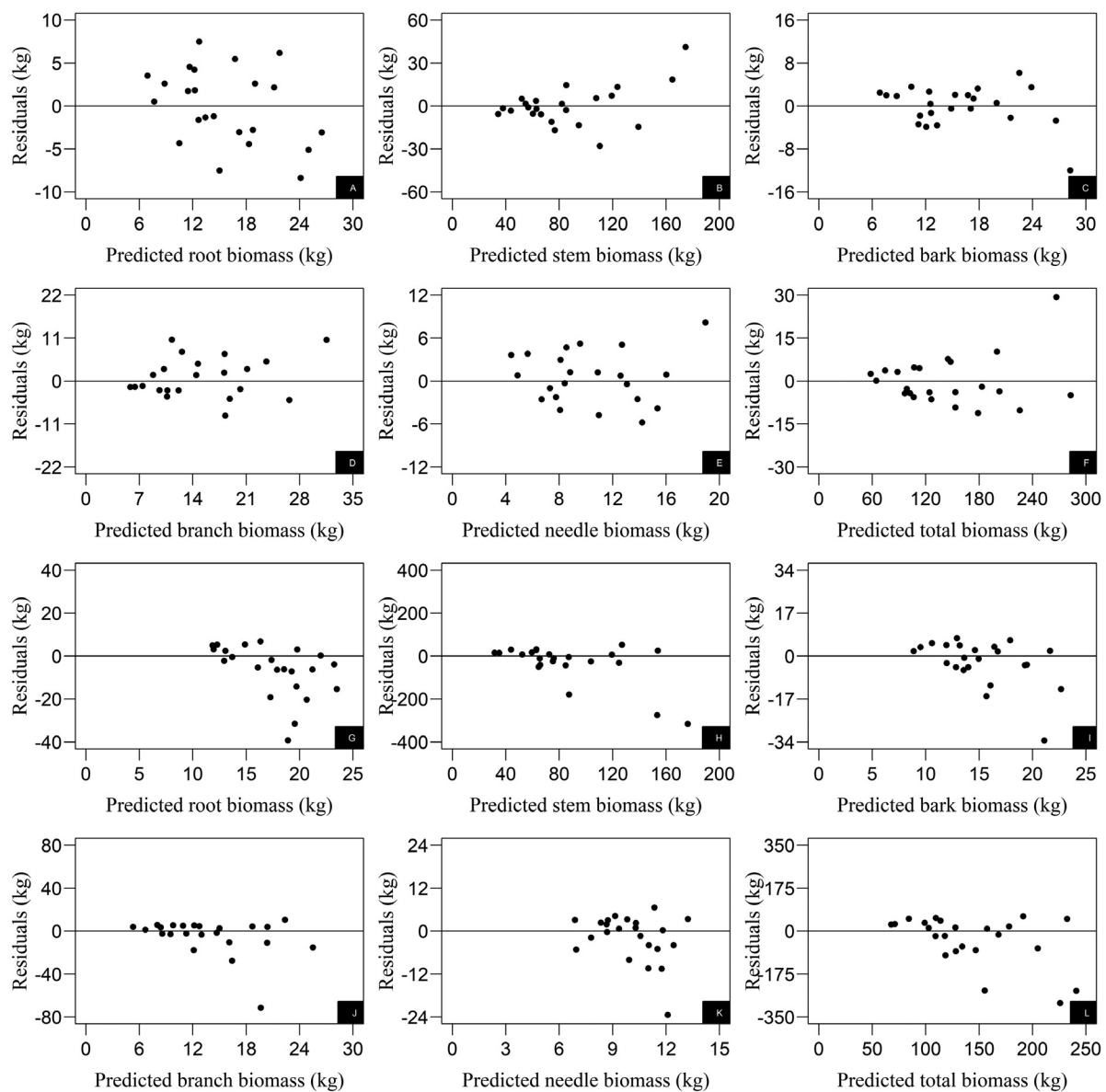


Figure 10. Comparison between ratio estimators (A, B, C, D, E, and F) and *WNSUR* procedure (G, H, I, J, K, and L) regarding the dispersion of residuals as a function of the estimated biomass for the parts (roots, stem, bark, branches, and needles) and total biomass in the diameter class II (16.24–23.55 cm), for *Pinus elliottii* and *Pinus taeda*.

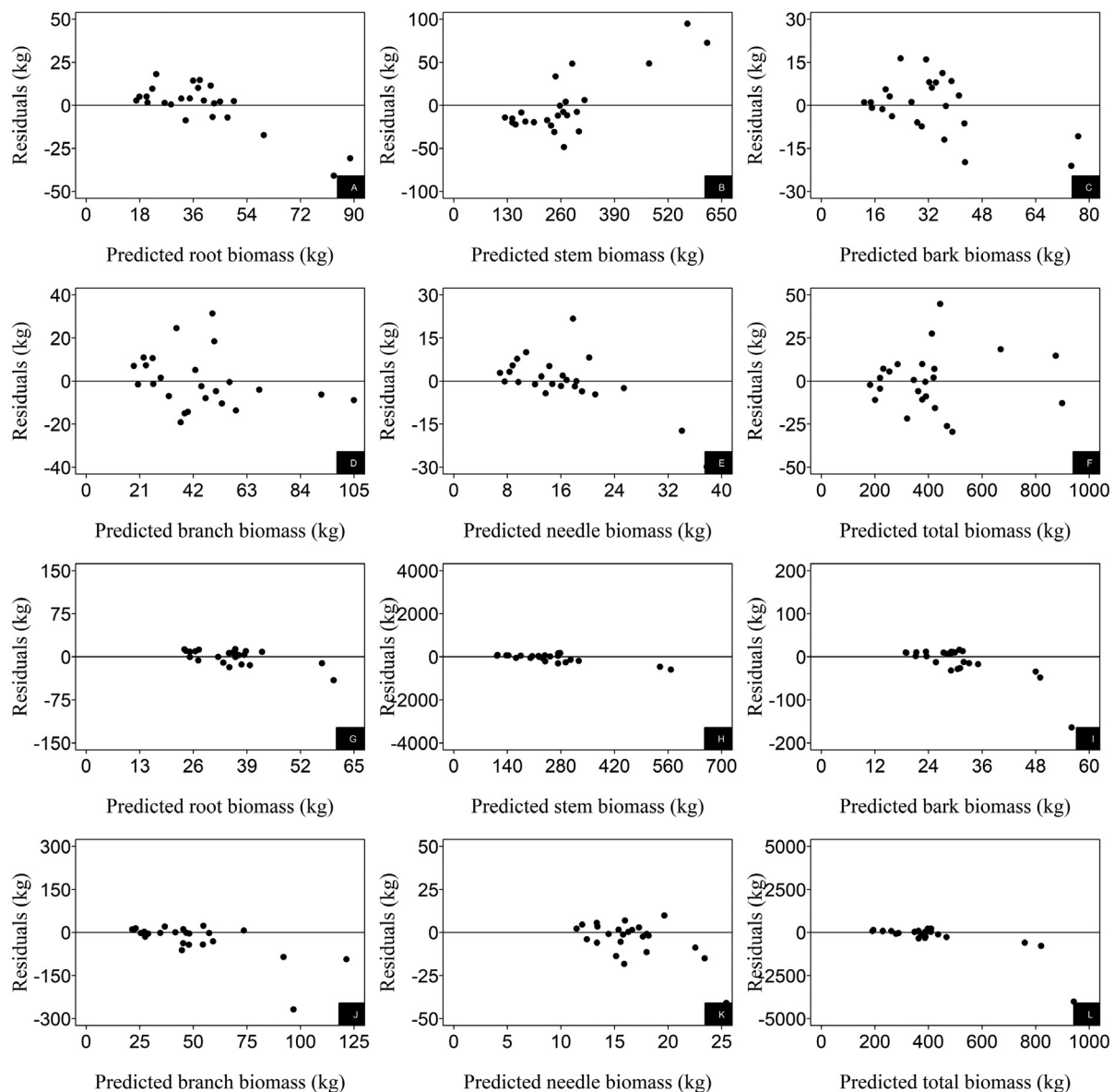


Figure 11. Comparison between ratio estimators (A, B, C, D, E, and F) and *WNSUR* procedure (G, H, I, J, K, and L) regarding the dispersion of residuals as a function of the estimated biomass for the parts (roots, stem, bark, branches, and needles) and total biomass in the diameter class III (23.56–40.27 cm), for *Pinus elliottii* and *Pinus taeda*.

The comparative analysis of the performance of ratio estimators and *WNSUR* estimators based on the behaviour of the fitting statistics is presented in Table 10. The ratio estimator performed better for most of the fitting statistics, except for the biomass of roots, stem, and needles, where the ratio estimator was better only for the *Bias* statistic.

Table 10. Comparison of the ratio estimators with *WNSUR* estimators regarding the statistics for the total biomass and its parts (roots, stem, bark, branches, and needles) for *Pinus elliottii* and *Pinus taeda*.

Parts	Statistics	Ratio estimator	<i>WNSUR</i>	Differences (%)
Roots	<i>Bias</i> (kg)	1.88E-15	0.16	-8.50E+15
	<i>MAE</i> (kg)	4.92	4.08	17.00
	<i>MSE</i> (kg ²)	68.40	36.05	47.29
	<i>RMSE</i> (kg)	8.27	6.00	27.40
Stem	<i>Bias</i> (kg)	1.83E-15	0.59	-3.21E+16
	<i>MAE</i> (kg)	12.44	11.08	10.91
	<i>MSE</i> (kg ²)	457.88	355.29	22.40
	<i>RMSE</i> (kg)	21.40	18.85	11.91
Bark	<i>Bias</i> (kg)	-4.57E-16	0.64	1.39E+17
	<i>MAE</i> (kg)	3.75	4.69	-25.23
	<i>MSE</i> (kg ²)	36.62	43.77	-19.54
	<i>RMSE</i> (kg)	6.05	6.62	-9.33
Branches	<i>Bias</i> (kg)	-1.52E-15	-0.44	-2.86E+16
	<i>MAE</i> (kg)	5.30	5.97	-12.57
	<i>MSE</i> (kg ²)	67.32	77.80	-15.56
	<i>RMSE</i> (kg)	8.20	8.82	-7.50
Needles	<i>Bias</i> (kg)	-5.08E-17	0.06	1.15E+17
	<i>MAE</i> (kg)	3.56	2.83	20.49
	<i>MSE</i> (kg ²)	36.52	21.35	41.54
	<i>RMSE</i> (kg)	6.04	4.62	23.54
Total	<i>Bias</i> (kg)	-5.08E-15	1.01	1.98E+16
	<i>MAE</i> (kg)	7.27	18.62	-156.15
	<i>MSE</i> (kg ²)	123.79	732.59	-491.81
	<i>RMSE</i> (kg)	11.13	27.07	-143.27

Bias is the random distortion; *MAE* is the mean absolute error; *MSE* is the mean square error; *RMSE* is the root-mean-square error; *WNSUR* is the weighted simultaneous fitting.

Discussion

Model identity test

The identity tests between the species-specific models and the generic model for total biomass estimation of *P. elliottii* and *P. taeda* (Table 3 and Fig. 5), along with the scatter plots (Fig. 1), indicated the feasibility of a combined modelling approach, confirming that both species exhibit comparable biomass accumulation dynamics. These results are corroborated by [Carvalho et al. \(2021\)](#), who, when analysing biomass stocks in a mixed plantation of *P. taeda* and *P. elliottii* under the same edaphoclimatic and management conditions, observed that although *P. taeda* exhibited higher mean biomass, the difference compared to *P. elliottii* was not statistically significant.

In this context, the adoption of a combined modelling approach for both species proves to be consistent with the general principle of parsimony, also known as Occam's razor ([Sober, 1981](#); [Burnham & Anderson, 2002](#)), which, when applied to the modelling process, suggests that, among competing models that adequately explain the data, the one with fewer parameters should be preferred ([Seasholtz & Kowalski, 1993](#)). This approach presents significant practical implications by reducing the need for species-specific calibrations without compromising the robustness of the estimates.

Ratio estimators for biomass

The general fit using ratio estimators yielded seemingly good estimates, with precisions below 20% (Table 4). However, the graphical analysis of the residuals' dispersion (Fig. 6) revealed heteroscedasticity for both total biomass and the biomass of individual parts, rendering these estimators inefficient, unless the non-constant variance of the residuals is addressed. For this reason, two-stage stratification was applied to correct heteroscedasticity. As a result, in addition to correcting heteroscedasticity, precision improved significantly, even for parts that, under conventional modelling (regression estimators), typically exhibit low estimation precision, such as roots, bark, branches, and needles, with the standard errors of the estimates reducing to a maximum of approximately 5% (Table 7), the lowest precision obtained. The observed improvement in the estimation precision of all biomass parts allows establishing more reliable inferences in biomass inventories. This is particularly relevant for carbon estimates, as biomass is the key variable in this process, especially when conversion factors are used to estimate carbon from biomass. Consequently, forest management becomes more efficient, as decision-making is based on more accurate information.

Applying the same procedure to estimate total aboveground biomass in subtropical forests of Brazil, Trautenmüller et al. (2023) reported precisions ranging from 51.12% to 180.13% for the entire dataset without stratification. Following the application of two-stage stratification, they observed substantial improvements, with standard errors of the estimates reduced to values between 0.58% and 3.65%. These findings underscore the effectiveness of two-stage stratification in improving estimation precision through ratio estimators, even in natural stands typically characterized by high species diversity, varying development stages, and a wide chronological range. Similarly, Pellico Netto & Behling (2019) reported precisions ranging from 1.98% to 7.05% for estimates of tree part biomass (stem and crown) and total biomass of black wattle (*A. mearnsii*). The two-stage stratification, by grouping trees based on diameter and slope coefficient classes, helps to address their structural variability, resulting in more robust ratio estimators. Consequently, this leads to improved biomass estimates, as the specific characteristics of each tree group are better assimilated, thereby generating more reliable and accurate model fit quality statistics.

The application of discriminant analysis confirmed the effectiveness of a posteriori classification of trees into the previously established groups, based on diameter and slope coefficient classes. The success rates, which ranged from 79.20% to 91.30% (Table 5), along with the substantial proportion of variance explained solely by the first discriminant function, ranging from 68.98% to 95.13% (Fig. 7), indicate that the grouping of trees was consistent. These results further strengthen the robustness of the two-stage stratification criterion, as mentioned earlier. Although the results of the discriminant analysis were satisfactory, it is important to highlight that this was the best possible result, considering the discriminant variables used. However, these results can be improved through the collection and analysis of the feasibility of including other variables that maximize the differences between groups and minimize variations within each group, such as the total observed tree volume (estimated in this study), form factor, basic wood density, additional variables related to tree architecture, among others, in addition to the variables already used in this manuscript, as suggested by MacFarlane (2015). Thus, discriminant analysis can increase the success rate of including new trees in each group (Trautenmüller et al., 2023).

Contrary to what was observed in this study, Trautenmüller et al. (2023), when applying discriminant analysis, obtained higher success rates in the larger diameter classes, associating this result with the greater variability observed in smaller diameter trees, due to the high diversity of species, ages, and growth stages. In contrast, this study, conducted with species in planted stands, which generally exhibit low heterogeneity in the factors mentioned by the referred authors, may explain the discrepancy in the obtained results.

The ratio estimators proved to be a promising technique for biomass estimation, both due to their precision and, above all, their simplicity of application. Through a simple division to estimate the coefficients, the additivity of the biomass parts is naturally ensured. With ratio estimators, the assumption that the metric unit of the independent variables must be the same as that of the dependent variable is ensured, allowing the desired property of additivity to be preserved (Pellico Netto & Behling, 2019). In other words, the slope coefficients of biomass part m (R_m), estimated by the ratio between the biomass of part m (kg) and the total stem volume (m^3), have a metric unit of $kg \cdot m^{-3}$. Since biomass is estimated by the product of the R_m coefficient ($kg \cdot m^{-3}$) and stem volume (m^3), the biomass is naturally obtained in kg.

By reducing the complexity of calculations, this approach can simplify biomass estimation for companies in the forestry sector, providing quick and consistent results, both statistically and biologically (natural

additivity), with a direct relationship between the variables. As the coefficients obtained through ratio estimation depend only on the means of the two variables involved, inventory procedures should prioritise the evaluation of individuals from classes close to the means of the selected variables, significantly reducing the required sample size and, consequently, operational costs (Péllico Netto & Behling, 2019).

Control procedure: simultaneous fitting (WNSUR)

Similarly, to the ratio estimators, the unweighted simultaneous fitting (*NSUR*) exhibited residuals' heteroscedasticity across all parts of the biomass, including the total. Although the White test (1980), presented in Table 8, indicated a different behaviour for total biomass, the visual analysis of the plots (Fig. 8 A, B, C, D, E, F) confirms the presence of heteroscedasticity for this variable as well. This suggests that, as with the other parts, the *NSUR* model is not suitable for accurately estimating total biomass. If heteroscedasticity is not corrected, as is the case with the Ordinary Least Squares method, the Generalised Least Squares method, specifically the Generalised Nonlinear Least Squares method, which involves the simultaneous fitting of nonlinear models, ceases to be efficient, which, according to Lin & Wei (2003), can lead to adverse consequences for the efficiency of the estimators.

Heteroscedasticity in biomass data is a common and virtually inevitable issue (Parresol, 2001), which explains why it has been extensively reported by several authors, notably including, among others: Parresol (1999), Parresol (2001), Nickless et al. (2011), Magalhães & Seifert (2015), Behling et al. (2018), Dong et al. (2018), Nhaduco et al. (2021), and Trautenmüller et al. (2021). According to Behling (2016), it is natural to expect that the variability of the biomass of smaller diameter trees is smaller than the variability of the biomass of larger diameter trees, causing the non-constant variance of the residuals from biomass equations. Under these circumstances, although the estimated parameters remain unbiased and consistent (Rosopa et al., 2013), they are inefficient, and the variance estimates are biased, which consequently invalidates both the significance tests (Alabi et al., 2020) and the confidence intervals (Greene, 2012). The standard errors of the estimates will be incorrect, leading to misleading inferences (Gelfand, 2015). Therefore, it is essential to assess the presence of heteroscedasticity in the data using either informal or formal methods (Nwakuya & Nwabueze, 2018) and to address it in order to obtain estimates with minimum variance and reliable prediction intervals (Parresol, 2001).

Following the correction of residual heteroscedasticity using weights derived from variance structure modelling in the weighted simultaneous fitting (*WNSUR*), it was found that this procedure provided more accurate biomass estimates for the stem and total, with precision, as measured by the standard error of the estimate ($S_{yx}\%$), of approximately 15%. Consistent with the findings of the present study (Table 9), other authors have also reported lower precision in biomass estimates for the parts roots, bark, and particularly the crown (branches and leaves) (Magalhães & Seifert, 2015; Zribi et al., 2016; Trautenmüller et al., 2021; Carvalho et al., 2022; Trautenmüller et al., 2023; Huy et al., 2023), which is due to the difficulty in establishing robust allometric relationships with the key independent variables, D and H . Results of similar magnitude were observed by Magalhães & Seifert (2015) (26.92% for roots, 48.25% for bark, and 45.23% for the crown, considered as the sum of leaves and branch biomass) in stands of *Androstachys johnsonii* Prain, and by Carvalho et al. (2022), who, similarly to the present study, conducted combined estimates of the biomass of *P. taeda* and *Pinus elliottii*, focusing, however, only on aboveground biomass without separating bark from stem, reporting 32.13% for branches and 54.70% for needles. In contrast, the stem biomass part, in general, exhibits a stronger correlation with D and H , resulting in more accurate biomass estimates for this part.

Although regression estimators (*WNSUR*) ensure additivity among the biomass parts, this assumption is not naturally achieved. The statistical dependencies among the biomass parts are forced through constraints on the regression coefficients, as described by Cunia & Briggs (1985) and Parresol (1999). Therefore, the additive estimation of total biomass results from the application of an integrated set of procedures: the selection of the best-fitting regression equations for each part, the summation of these equations, and, finally, the imposition of constraints on these equations to ensure that the sum of the partial estimates corresponds to the total biomass.

According to Péllico Netto & Behling (2019), it is a fact that researchers frequently apply volumetric models for biomass estimation, where the independent variables generally correspond to diameter at 1.30 m height aboveground (D) and total height (H), used either individually or in combination (Cole & Ewel, 2006; Dong et al., 2015) and/or other easily measurable variables (Li & Zhao, 2013). The differences in the metric

units of the dependent variables, specifically total biomass and its parts in kg, and the independent variables, namely D in cm, H , crown diameter (dc), and crown height (hc) in m, as considered in this study, constitute one of the main reasons why additivity is not naturally ensured with regression estimators. Another relevant point is that, in forest biomass modelling, the variables form factor and basic wood density are associated to the intercept and angular coefficient of the linear and nonlinear regressions, however, they exert a random influence on these equations, which compromises the natural additivity of the biomass parts, as pointed out by [Péllico Netto & Behling \(2019\)](#).

Comparative analysis: ratio estimators and regression estimators (WNSUR)

As previously discussed in detail, substantial gains in the precision of total biomass and its parts estimates were achieved through the use of ratio estimators, which ensured the assumption of additivity in a simple and natural manner, corroborating the hypothesis raised in the study. In ratio estimators, the coefficients are estimated separately, that is, in a segmented fashion for each tree group formed through two-stage stratification, which enhances the precision of the estimates and eliminates the difficulties associated with complex methodologies used to enforce the additivity of parts in regression estimators ([Kozak, 1970](#); [Chiyenda & Kozak, 1984](#); [Cunia & Briggs, 1984](#); [Cunia & Briggs, 1985](#); [Reed & Green, 1985](#); [Parresol, 1999](#); [Parresol, 2001](#)), among which the most widely used are *SUR*, for linear models, and *NSUR*, for non-linear models.

Proper data segmentation is an effective strategy for addressing variability, in contrast to approaches that attempt to model it directly, as commonly occurs with regression estimators. The application of this strategy in regression estimators is often limited by the difficulty of obtaining large samples in destructive biomass studies. Since this method involves tree felling, the number of sampled individuals tends to be reduced, thereby restricting the partitioning of data into different classes. Small sample sizes may result in biased estimates, as allometric parameters are sensitive to sample size ([Duncanson et al., 2015](#)).

In general, the ratio estimators exhibited a narrower range of residual distribution for all biomass parts as well as for total biomass (Fig. 9, 10, and 11). Ratio proved to be appropriate and reliable in all parts and total biomass, except for roots, stem, and needles biomass, where the *WNSUR* estimators performed slightly better for *MAE*, *MSE*, and *RMSE* statistics (Table 10). As observed by [Péllico Netto & Behling \(2019\)](#) in plantations of black wattle (*A. mearnsii*) and by [Trautenmüller et al. \(2023\)](#) in native forests, these being the only studies to date to apply ratio estimators for biomass estimation and to compare them with the *WNSUR* procedure, in the present study it was once again found that ratio estimators are a viable alternative to *WNSUR* regression estimators, in the present case for the genus *Pinus*, one of the most widely planted globally, including in Brazil. Therefore, if precision is a fundamental requirement for using biomass as an ecological indicator, the two-stage stratified ratio estimators are suitable for achieving satisfactory precision.

Given that the use of ratio estimators as a tool for additive biomass estimation remains in its early stages, it is recommended that future research broaden the scope of application to encompass diverse forest typologies and ecological conditions. Investigations incorporating alternative stratification approaches may further enhance the precision and applicability of ratio estimators. Furthermore, assessing the performance of these methods in contexts such as temporal monitoring, land-use changes, and varying successional stages would be valuable in consolidating their utility as a reliable tool for biomass quantification and for supporting decision-making in forest ecosystem management.

Conclusion

Based on the findings of this study, both the ratio estimators and the regression estimators (*NSUR*) exhibited residual heteroscedasticity. This issue was effectively addressed through two-stage stratification and the application of weights to the equations in the procedures of the ratio and regression estimators, respectively.

When modelling biomass using ratio estimators, the additivity of the tree biomass parts was fully satisfied in a simple and natural manner, thereby ensuring biological consistency. In contrast, simultaneous fitting (*WNSUR*), although ensuring the premise of additivity, did not naturally fulfil this requirement, necessitating additional procedures to enforce this condition.

Compared to the regression estimators (*WNSUR*), the ratio estimators proved to be appropriate and reliable for all biomass parts, except for the roots, stem, and needles biomass, where the regression (*WNSUR*) provided slightly better fits.

The ratio estimator technique yields adequate estimates and, due to its simplicity and its natural compliance with the additivity premise of biomass parts, it constitutes an effective and promising approach for generating compatible estimates of tree biomass.

Supplementary material (Table S1, Figures S1, S2 and-S3): accompanies the paper on *Forest System's* website.

Data availability: Data cannot be shared openly but are available on request from authors: Data sets generated during the current study are available from the corresponding author on reasonable request.

Acknowledgements: The authors express their sincere gratitude to the project “Dynamic levels and biological stability in the growth rates of planted forests in southern Brazil and their potential in generating carbon credits”, especially to the Centre of Excellence in Research on Biomass and Carbon – BIOFIX, Federal University of Paraná, Brazil, for kindly providing the biomass data used in this study.

Competing interests: The authors have declared that no competing interests exist.

Authors' contributions: **Ornélio P. E. Nhaduco:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Alexandre Behling:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization. **Sylvio Péllico-Netto:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ana P. Dalla-Corte:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision. **Carlos R. Sanquetta:** Investigation, Methodology, Project administration, Resources, Supervision. **Franciele A. da Silva:** Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – review & editing. **Jonathan W. Trautenmüller:** Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Funding: The authors received no specific funding for this work.

References

- Alabi OO, Ayinde K, Babalola OE, Bello HA, Okon EC, 2020. Effects of Multicollinearity on Type I Error of Some Methods of Detecting Heteroscedasticity in Linear Regression Model. *Open J Stat* 10: 664-677. <https://doi.org/10.4236/ojs.2020.104041>
- Behling A, 2016. Modelagem da biomassa de árvores para assegurar aditividade dos seus componentes. Doctoral thesis. Universidade Federal do Paraná, Brasil.
- Behling A, Péllico-Netto S, Sanquetta CR, Dalla-Corte AP, Affleck DL, Rodrigues AL, Behling M, 2018. Critical analyses when modeling tree biomass to ensure additivity of its components. *An Acad Bras Cienc* 90(2): 1759-1774. <https://doi.org/10.1590/0001-3765201820170684>
- Behling A, Péllico-Netto S, Sanquetta CR, Dalla-Corte AP, Simon AA, Rodrigues AL, Caron BO, 2019. Additive and Non-additive Biomass Equations for Black Wattle. *Floram* 26(4): e20170439. <https://doi.org/10.1590/2179-8087.043917>
- Burnham KP, Anderson DR (eds), 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York. 488 pp.
- Cao B, Domke GM, Russell MB, Walters BF, 2019. Spatial modeling of litter and soil carbon stocks on forest land in the conterminous United States. *Sci Total Environ* 654: 94-106. <https://doi.org/10.1016/j.scitotenv.2018.10.359>

- Carvalho RR, Trautenmüller JW, Carvalho SR, Júnior SC, Silva DA, Figueiredo Filho, A, 2021. Determination of the biomass stock in a mixed plantation of *Pinus taeda* L. and *Pinus elliottii* Engelm. *Adv For Sci* 8(2): 1455-1462. <https://doi.org/10.34062/afs.v8i2.10129>
- Carvalho RR, Trautenmüller JW, Kohler SV, Serpe EL, Corte AP, Silva DA, Figueiredo Filho A, 2022. Biomass modeling in a mixed plantation of *Pinus taeda* L. and *Pinus elliottii* Engelm. *Sci For* 50: e3894. <https://doi.org/10.18671/scifor.v50.27>
- Chiyenda SS, Kozak A, 1984. Additivity of component biomass regression equations when the underlying model is linear. *Can J For Res* 14: 441-446. <https://doi.org/10.1139/x84-078>
- Cole TG, Ewel JJ, 2006. Allometric equations for four valuable tropical tree species. *For Ecol Manage* 229: 351-360. <https://doi.org/10.1016/j.foreco.2006.04.017>
- Cochran, WG, 1953. *Sampling Techniques*. John Wiley & Sons, Inc., New York.
- Cochran, WG, 1963. *Sampling Technique*, 2nd ed. John Wiley & Sons, Inc., New York.
- Costa-Junior S, Péllico-Netto S, Behling A, 2022. Equações de Preussner para a estimativa do volume de árvores de acácia-negra e estoque energético para os componentes toco e ponteira. *Sci For* 50: e3869. <https://doi.org/10.18671/scifor.v50.36>
- Cunia T, Briggs RD, 1984. Forcing additivity of biomass tables: some empirical results. *Can J For Res* 14(3): 376-384. <https://doi.org/10.1139/x84-067>
- Cunia T, Briggs R, 1985. Forcing additivity of biomass tables: use of the generalized least squares method. *Can J For Res* 15(1): 23-28. <https://doi.org/10.1139/x85-006>
- Dai W, Fu W, Jiang P, Zhao K, Li Y, Tao J, 2018. Spatial pattern of carbon stocks in forest ecosystems of a typical subtropical region of southeastern China. *For Ecol Manage* 409: 288-297. <https://doi.org/10.1016/j.foreco.2017.11.036>
- Dong L, Zhang L, Li F, 2014. A compatible system of biomass equations for three conifer species in Northeast, China. *For Ecol Manage* 329: 306-317. <https://doi.org/10.1016/j.foreco.2014.05.050>
- Dong L, Zhang L, Li F, 2015. Developing additive systems of biomass equations for nine hardwood species in Northeast China. *Trees* 29: 1149-1163. <https://doi.org/10.1007/s00468-015-1196-1>
- Dong L, Zhang L, Li F, 2016. Developing two additive biomass equations for three coniferous plantation species in Northeast China. *Forests* 7(7): 136. <https://doi.org/10.3390/f7070136>
- Dong L, Zhang L, Li F, 2018. Additive biomass equations based on different dendrometric variables for two dominant species (*Larix gmelini* Rupr. and *Betula platyphylla* Suk.) in natural forests in the eastern Daxing'an Mountains, Northeast China. *Forests* 9(5): 261. <https://doi.org/10.3390/f9050261>
- Duncanson L, Rourke O, Dubayah R, 2015. Small sample sizes yield biased allometric equations in temperate forests. *Sci Rep* 5(1): 17153. <https://doi.org/10.1038/srep17153>
- 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. *For Sci* 63(3): 241-249. <https://doi.org/10.5849/FS-2016-055>
- Gelfand SJ, 2015. Understanding the impact of heteroscedasticity on the predictive ability of modern regression methods. Master's dissertation. University of Calgary, Calgary.
- Greene WH, 2012. *Econometric analysis*. 7th Edition, Pearson. 1238 pp.
- Hair JF, Black WC, Babin BJ, Anderson RE, 2009. *Multivariate data analysis*. 7th Edition, Pearson. 734 pp.
- Huy B, Khiem NQ, Truong NQ, Poudel KP, Temesgen H, 2023. Additive modeling systems to simultaneously predict aboveground biomass and carbon for *Litsea glutinosa* of agroforestry model in tropical highlands. *For Syst* 32(1): e006. <https://doi.org/10.5424/fs/2023321-19780>
- Kozak A, 1970. Methods of ensuring additivity of biomass components by regression analysis. *For Chron* 46(5): 402-405. <https://doi.org/10.5558/tfc46402-5>
- Li H, Zhao P, 2013. Improving the accuracy of tree-level above ground biomass equations with height classification at a large regional scale. *For Ecol Manage* 289: 153-163. <https://doi.org/10.1016/j.foreco.2012.10.002>
- Lin JG, Wei BC, 2003. Testing for heteroscedasticity in nonlinear regression models. *Commun Stat Theory Methods Communications in Statistics: Theory and Methods* 32(1): 171-192. <https://doi.org/10.1081/STA-120017806>
- MacFarlane DW, 2015. A generalized tree component biomass model derived from principles of variable Allometry. *For Ecol Manage* 354: 43-55. <https://doi.org/10.1016/j.foreco.2015.06.038>
- Magalhães TM, Seifert T, 2015. Biomass modelling of *Androstachys johnsonii* Prain: a comparison of three methods to enforce additivity. *Int J For Res* 17: Art ID 878402. <https://doi.org/10.1155/2015/878402>

- Nickless A, Scholes RJ, Archibald S, 2011. A method for calculating the variance and confidence intervals for tree biomass estimates obtained from allometric equations. *Afr J Sci* 107(5/6): 356. <https://doi.org/10.4102/sajs.v107i5/6.356>
- Nhaduco OPE, Dalla-Corte AP, Behling A, Sanquetta CR, 2021. Sistema de equações de biomassa de *Pinus* spp. no sul do Paraná. *Sci For* 49(129): e3457. <https://doi.org/10.18671/scifor.v49n129.24>
- Nwakuya MT, Nwabueze JC, 2018. Application of box-cox transformation as a corrective measure to heteroscedasticity using an economic data. *Am J Math Stat* 8(1): 8-12.
- Oliveira TWG, Lima EM, Sanquetta CR, Dalla-Corte AP, 2018. Ajuste simultâneo na predição de biomassa aérea em *Araucaria angustifolia* (Bertol.) Kuntze. *Biofix Sci J* 3(1): 137-144. <https://doi.org/10.5380/biofix.v3i1.58119>
- Parresol BR, 1999. Assessing tree and stand biomass: a review with examples and critical comparisons. *For Sci* 45(4): 573-593. <https://doi.org/10.1093/forestscience/45.4.573>
- Parresol BR, 2001. Additivity of nonlinear biomass equations. *Can J For Res* 31(5): 865-878. <https://doi.org/10.1139/x00-202>
- Péllico Netto S, Behling A, 2019. Additivity of tree biomass components using ratio estimate. *An Acad Bras Cienc* 91: e20180272. <https://doi.org/10.1590/0001-3765201920180272>
- Picard N, Saint-André L, Henry M, 2012. Manual for building tree volume and biomass allometric equations: from field measurement to prediction. Rome e Montpellier: Food and Agricultural Organization of the United Nations and Centre de Coopération Internationale en Recherche Agronomique pour le Développement, 215 pp. Available in: <https://www.fao.org/3/i3058e/i3058e.pdf>.
- Preußner K, 1974. Eine neue Schaftkurvengleichung und ihre Anwendung. *Wissenschaftliche Zeitschrift der Technischen Hochschule Universität Dresden* 23(11): 305-309.
- Reed D, Green EJ, 1985. A method of forcing additivity of biomass tables when using nonlinear models. *Can J For Res* 15(6): 1184-1187. <https://doi.org/10.1139/x85-193>
- Rosopa PJ, Schaffer MM, Schroeder AN, 2013. Managing heteroscedasticity in general linear models. *Psychol Methods* 18(3): 335-351. <https://doi.org/10.1037/a0032553>
- Sanquetta CR, Behling A, Dalla-Corte AP, Péllico-Netto S, Schikowski AB, Amaral MK, 2015. Simultaneous estimation as alternative to independent modeling of tree biomass. *Ann For Sci* 72: 1099-1112. <https://doi.org/10.1007/s13595-015-0497-2>
- SAS Institute Inc., 2012. SAS® OnDemand for Academics: Student User's Guide, Second Edition. Cary, NC: SAS Institute Inc
- Seasholtz MB, Kowalski B, 1993. The parsimony principle applied to multivariate calibration. *Anal Chim Acta* 277: 165-177. [https://doi.org/10.1016/0003-2670\(93\)80430-s](https://doi.org/10.1016/0003-2670(93)80430-s)
- Shahrokhzadeh U, Sohrabi H, Copenheaver CA, 2015. Aboveground biomass and leaf area equations for three common tree species of Hyrcanian temperate forests in northern Iran. *Botany* 93(10): 663-670. <https://doi.org/10.1139/cjb-2015-0078>
- Sitoe AA, Tchaúque F, 2007. Medição de Biomassa Florestal Utilizando Informação do Inventário Florestal. Ministério da Agricultura, Maputo. 52 pp.
- Sober E, 1981. The principle of parsimony. *Br J Philos Sci* 32: 145-156. <https://doi.org/10.1093/bjps/32.2.145>
- Stefanello FR, Péllico-Netto S, Behling A, Pelissari AL, Orso GA, 2021. Preußner functions for volume estimation of *Pinus taeda* L. in Southern Brazil. *N Z J For Sci* 51:4. <https://doi.org/10.33494/nzjfs512021x66x>
- Sukhatme PV, 1954. Sampling Theory Of Surveys With Application. The Indian Society of Agricultural Statistics, New Delhi, India, 491 pp.
- Trautenmüller JW, Péllico-Netto S, Balbinot R, Watzlawick LF, Dalla-Corte AP, Sanquetta CR, Behling A, 2021. Regression estimators for aboveground biomass and its constituent parts of trees in native southern Brazilian forests. *Ecol Indic* 130:108025. <https://doi.org/10.1016/j.ecolind.2021.108025>
- Trautenmüller JW, Péllico-Netto S, Balbinot R, David HC, Dalla-Corte AP, Watzlawick LF, Sanquetta CA, Mallmann AA, Engel K, Behling A, 2023. Ratio estimators for aboveground biomass and its parts in subtropical forests of Brazil. *Ecol Indic* 154:110530. <https://doi.org/10.1016/j.ecolind.2023.110530>
- Urbano E, Machado SA, Figueiredo-Filho A, Koehler HS, 2008. Modelagem da biomassa aérea em bracingais nativos da região metropolitana de Curitiba. *Floresta* 38(2): 361-372. <https://doi.org/10.5380/rf.v38i2.11631>
- Wang X, Zhao D, Liu G, Yang C, Teskey RO, 2018. Additive tree biomass equations for *Betula platyphylla* Suk. plantations in Northeast China. *Ann For Sci* 75:60. <https://doi.org/10.1007/s13595-018-0738-2>

- White H, 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4): 817-838. <https://doi.org/10.2307/1912934>
- Xing D, Bergeron JAC, Solarik KA, Tomm B, Macdonald SE, Spence JR, He F, 2019. Challenges in estimating forest biomass: use of allometric equations for three boreal tree species. *Can J For Res* 49(12): 1613-1622. <https://doi.org/10.1139/cjfr-2019-0258>
- Zeng W, Zhang L, Chen X, Cheng Z, Ma K, Li Z, 2017. Construction of compatible and additive individual-tree biomass models for *Pinus tabulaeformis* in China. *Can J For Res* 47(4): 467-475. <https://doi.org/10.1139/cjfr-2016-0342>
- Zhao D, Kane M, Markewitz D, Teskey R, Clutter M, 2015. Additive tree biomass equations for midrotation loblolly pine plantations. *For Sci* 61(4): 613-623. <https://doi.org/10.5849/forsci.14-193>
- Zribi L, Chaar H, Khaldi A, Henchi B, Mouillot F, Gharbi F, 2016. Estimate of biomass and carbon pools in disturbed and undisturbed oak forests in Tunisia. *For Syst* 25(2): e060. <https://doi.org/10.5424/fs/2016252-08062>

