iForest
Biogeosciences and Forestry

Research Article
doi: 10.3832/ifor4919-019
vol. 19, pp. 261-268

Exploring biomass modeling in the Amazon Forest: assessing the effect of sample size of model calibration datasets

Fabiano Rodrigues Pereira [(1)],
Thaís Chaves Almeida [(1)],
Hassan Camil David [(2)],
Lina Mayra Reis Galvão [(1)],
Alexandre Behling [(2)],
Angelo Augusto Ebling [(2)],
Allan Libanio Pelissari [(2)],
Xavier Simiao Chivale [(1)],
Rebecca Araujo García [(1)]

Tree-level models are commonly used to estimate the mean and total forest aboveground biomass (AGB). However, tree biomass predictions can vary significantly between models due to limited understanding of how calibration datasets influence their performance. The aim was to recommend a sample size for a dataset to calibrate tree biomass models that estimate mean AGB per unit area with the best possible precision and accuracy. The methodology consisted of (i) simulating an Amazonian forest using the Monte Carlo Method (MCM) and *m* out of *n* Bootstrap, (ii) estimating the average forest biomass using models calibrated with datasets of varying sample sizes, (iii) analyzing the precision and accuracy of the estimate of the mean AGB in Mg ha^{-1} per unit area using the models calibrated with different sample sizes, and (iv) relating characteristics of the calibration datasets with the errors produced by the models. Our findings indicate that increasing sample size improves the precision of biomass estimates, with notable gains at 200 trees. The study highlights that larger samples better represent tree diversity and support reliable biomass modeling, which is essential for effective forest management. We conclude that a minimum calibration sample size of 200 trees is sufficient to optimize biomass predictions while balancing practical sampling constraints. These results can aid in developing forest management strategies and improving carbon credit calculations in tropical ecosystems.

Keywords: Forest Simulation, Tree Biomass Modeling, Bootstrap Sampling

Introduction

Tropical forests are essential to the global carbon cycle, sequestering 40%-55% of carbon found in terrestrial plants and making a substantial contribution to primary productivity (Cuni-Sanchez et al. 2021, De Melo et al. 2024). These ecosystems play a crucial role in the context of climate change, serving as significant carbon sinks and influencing atmospheric carbon dynamics (Fearnside 2018).

To generate carbon credits under the REDD+ program (Reducing Emissions from Deforestation and Forest Degradation), accurate estimates of forest carbon stocks are essential. Efforts in carbon measurement have focused on carbon stocks in forests (Assis et al. 2020, Ekoungoulou et al. 2018), which play a significant role in mitigating climate change by removing CO_2 from the atmosphere and storing it in biomass and soil, as well as combating deforestation and forest degradation. These measures allow developing countries with verified greenhouse gas reductions and increased carbon stocks to qualify for results-based payments from international sources, such as the Green Climate Fund (GCF). Therefore, accurate estimates of carbon emissions from these forests necessitate reliable assessments of forest biomass.

The development of pantropical allometric equations has substantially improved biomass estimation across diverse tropical forest types (Chave et al. 2014). Nevertheless, multiple sources of uncertainty still affect the reliability of these estimates, making accurate biomass assessment challenging (Temesgen et al. 2015). Among these sources of uncertainty, determining the appropriate size of the calibration dataset remains a critical issue. Although larger sample sizes are generally assumed to enhance model accuracy, the relationship between sample size and predictive performance remains poorly understood.

In addition to pantropical allometric equations, forest biomass estimation has increasingly incorporated a broader range of statistical and machine-learning approaches. Traditional parametric models, including linear and nonlinear regressions, remain widely used due to their interpretability and solid theoretical foundation in forest biometrics (Garcia et al. 2026). More recently, non-parametric and artificial intelligence-based methods have gained attention for their ability to model complex and nonlinear relationships in biomass prediction (Huy et al. 2022, Antúnez et al. 2025). In particular, Random Forest and other ensemble learning algorithms have been extensively applied to remote sensing-based biomass modeling due to their strong predictive performance and robustness to multicollinearity (Campbell et al. 2021, Leite et al. 2022). However, despite their flexibility, these methods are sensitive to the quality and representativeness of calibration data, and their predictive performance

(1) Federal University of Paraná, Pref. Lothário Meissner Avenue 900, Curitiba 80210-170, PR (Brazil); (2) Department of Forest Science, Federal University of Paraná, Pref. Lothário Meissner Avenue 900, Curitiba 80210-170, PR (Brazil)

@ Hassan Camil David (hassancamil@ufpr.br)

Received: Jun 05, 2025 - Accepted: Jun 17, 2026

Citation: Pereira FR, Almeida TC, David HC, Galvão LMR, Behling A, Ebling AA, Pelissari AL, Chivale XS, García RA (2026). Exploring biomass modeling in the Amazon Forest: assessing the effect of sample size of model calibration datasets. iForest 19: 261-268. - doi: 10.3832/ifor4919-019 [online 2026-07-21]

Communicated by: Maurizio Marchi

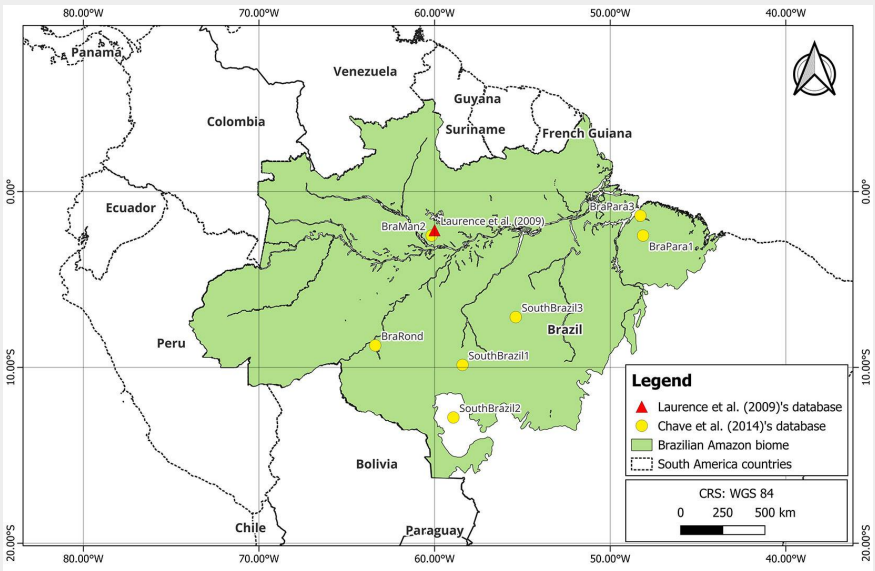

Fig. 1 - Geographical locations of the study sites. The red circle represents the study area from Laurance et al. (2009), while the black circles denote seven study sites sourced from Chave et al. (2014)'s database.

is strongly influenced by training sample size, reflecting the well-known bias-variance trade-off in statistical learning (Jenkins & Quintana-Ascencio 2020, Rajput et al. 2023). This underscores the need for a deeper understanding of how calibration dataset size influences the reliability and stability of biomass model predictions.

Although biomass can also be estimated using plot-level expansion factors (Schepaschenko et al. 2018) or remote sensing-based models (Potapov et al. 2021), tree-level calibration remains a fundamental component of forest inventory and carbon accounting. Errors originating at the individual-tree modeling stage may propagate to stand-level estimates, particularly when calibration datasets are limited or unrepresentative (Vorster et al. 2020). While previous studies have primarily focused on improving model form or predictor selection (Chave et al. 2014, Adame-Campos et al. 2019), investigations examining the role of calibration sample size have largely addressed its effects across scales, without explicitly evaluating how it influences the reliability of stand-level biomass estimates (Fassnacht et al. 2014, Hetzer et al. 2020).

Within this context, this study investigates whether variation in calibration dataset size affects the reliability of stand-level mean biomass estimates. The aim was to recommend a sample size for the dataset to calibrate tree biomass models that estimate mean aboveground biomass (AGB) per unit area with the highest possible accuracy. We hypothesize that increasing the calibration sample size enhances the accuracy and stability of stand-level mean AGB estimates in Amazonian forests, because selecting a representative sample requires multiple trees from the assessed population, and an insufficient sample size may yield a less precise model (Duncanson et al. 2015, Sullivan et al. 2018). By investigating this relationship, we hope to provide insights that could enhance forest management practices and contribute to more reliable carbon accounting in tropical forest ecosystems.

Material and methods

Study area

Although our study uses forest simulation, the study area is defined as the central Amazon, with forest stand attributes obtained from Laurance et al. (2009). The authors collected data in a forest remnant of 1,000-km² in the region of Manaus, State of Amazonas, Brazil (Fig. 1). The area consists of forest fragments, primary rainforest, and a matrix of grassland and regenerating forest at 50- to 100-m elevation (Lovejoy 1986, Laurance et al. 2002, 2009). The average precipitation is 2,600 mm per year, with a moderately strong dry season from June to August. The region also features nutritionally poor, highly acidic soils classified as Yellow Latosol (Chauvel et al. 1987).

Data

For the simulation process described in the next section, in addition to the forest stand attributes consulted in Laurance et al. (2009), we utilized trees from a sub-dataset taken from Chave et al. (2014)'s database. The database (freely available at https://zenodo.org/records/14932971) comprises 4,004 trees, with destructively collected diameter at breast height (DBH) and total height (HT) measurements, from 58 sites spanning the tropics. The sub-dataset included only trees with a DBH ≥ 10 cm, collected within or near the Brazilian Amazon Forest, encompassing the sites labeled as BraMan2, BraPara1, BraPara3, BraRond, SouthBrazil1, SouthBrazil2, and SouthBrazil3 (Fig. 1).

The sub-dataset with Amazon's trees comprises n_{AMZ} = 410 trees and will hereafter be referred to as the Amazonian sub-dataset. Descriptive statistics of the Amazonian sub-dataset are given in Tab. 1. The scatter plot of DBH and AGB for the Amazonian sub-dataset is presented in Fig. 2a.

Forest simulation and modeling

To simulate a forest, we adopted an inverted J-shaped diameter distribution typical of humid tropical forests (David et al. 2022), as well as parameters for basal area and tree density per hectare observed by Laurance et al. (2009) in a forest in central Amazonia. The algorithm for executing the forest simulation, tree biomass modeling, and forest biomass prediction was divided into two parts (see below).

Part I: Forest simulation

Part I involved simulating ten 1-ha field plots using the Monte Carlo Method (MCM), with diameter at breast height (DBH) following an inverted J-shaped distribution. The algorithm involves the steps described below.

Define *i* as a tree and *j* as a 1-ha plot.

(i) Define the parameters of the beta distribution (α; β) to reflect an inverted J-shape. To achieve this, α was set to 0.3 and β to 2.0 (eqn. 1):

$$f(x;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1|x)^{\beta-1} \quad (1)$$

(ii) For the *j*-th plot, obtain X_i = 10 + *x*100, where *x* is a random vector of size *n* drawn from the beta distribution that ranges from 0 to 1. *X* is a vector representing the DBHs, with values ranging from 10 to 100 cm, and n is a random number between 594 and 613, representing the number of trees ≥ 10 cm per ha, as in Laurance et al. (2009).

(iii) For the *j*-th plot, save the vector *X* if $G_j = \Sigma_{i=1}^{nj} (\pi X_i^2)$ is between 27 and 30 m² ha⁻¹,

Tab. 1 - Descriptive statistics of the Amazonian sub-dataset from Chave et al. (2014)'s database. (DBH): diameter at 1.30 m above ground; (HT): total height; (AGB): aboveground biomass; (SD): standard deviation; (CV): coefficient of variation.

Variable	Minimum	Mean	Maximum	Median	SD	CV (%)	n_{AMZ}
DBH (cm)	10.0	22.6	138.0	16.4	17.0	75.0	410
HT (m)	6.0	19.4	54.1	17.6	7.4	38.3	410
AGB (kg.tree^{-1})	12.3	730.5	20,416.0	142.8	1937.7	265.3	410

as in Laurance et al. (2009).

(iv) Replicate steps i-iii 10 times.

Steps i-iv define the diameter distribution of the forest, which consists of 6,014 trees distributed across ten 1-ha plots. The results of these steps are shown in Fig. 2b.

The following steps describe how the forest was simulated using trees selected from the Amazonian sub-dataset (n_{AMZ} = 410).

Denote *k* for the diameter class, where *k* = 1 corresponds to a DBH of 10-20 cm and *k* = 9 corresponds to a DBH of 90-100 cm.

(v) For the *j*-th plot created in steps i-iv, count the number of trees in the *k*-th diameter class, being this number represented by n_{jk}.

(vi) For the *j*-th plot and the *k*-th class, randomly select with replacement n_{jk} trees from the Amazonian sub-dataset (n_{AMZ} = 410).

(vii) Repeat steps v-vi for the remaining diameter classes.

(viii) Repeat steps v-vii for the remaining plots.

The simulation process allowed us to determine the actual value (parameter) of the AGB for the ten 1-ha plots, as the Amazonian sub-dataset provides the observed biomass value for each tree, along with the variables DBH and HT. Tab. 2 presents descriptive statistics for AGB, DBH, and HT of the trees in the simulated forest.

Part II: Biomass modeling under different sample sizes and uncertainty analysis

In Part II, tree biomass models were fitted using six calibration dataset sizes. The resulting equations from these fittings were used to predict the AGB in the ten 1-ha plots.

Denote *i* for trees, *j* for plots, and *m* for the sample size of the calibration dataset, where *m* = 1, $n_{calib.1}$ = 13; *m* = 2, $n_{calib.2}$ = 25; *m* = 3, $n_{calib.3}$ = 50; *m* = 4, $n_{calib.4}$ = 100; *m* = 5, $n_{calib.5}$ = 200; *m* = 6, $n_{calib.6}$ = 400 trees.

(i) For the *m*-th sample size, randomly select with replacement a sample of n_m trees from the Amazon subset, where $n_m < n_{AMZ}$, thereby implementing the *m* out of *n* Bootstrap technique.

(ii) Obtain descriptive statistics for the DBH variable from the sample selected in the previous step, including: minimum, maximum, range, mean, standard deviation, coefficient of variation (CV), Q1 (1st quartile), Q2 (2nd quartile), Q3 (3rd quartile), range between Q1 and Q2, number of trees with DBH > Q3, number of trees with DBH > 1.5 × Q3, skewness, kurtosis, and number of trees with DBH < Q1.

(iii) Fit the biomass model at the individual tree level (eqn. 2) and obtain the vector of coefficients $\hat{\beta}_p$:

$$\ln(\hat{b}_{ij}) = \hat{\beta}_0 + \hat{\beta}_1 \ln(DBH_{ij}^2) + \hat{\beta}_2 \ln(HT_{ij}) \quad (2)$$

where $\hat{\beta}_{ij}$ is the predicted tree biomass, in kg; DBH_{ij} is the diameter of the tree at 1.30 m above the ground from the selected sample, in cm; HT_{ij} is the total height of the tree from the selected sample, in m.

The application of a model that, in addition to DBH, includes tree height as an explanatory variable acknowledges that ignoring tree height in biomass estimation can lead to significant errors (Feldpausch et al. 2012, Chave et al. 2005, Goussanou et al. 2016, Djomo et al. 2010, Ekoungoulou et al. 2018).

(iv) For the *j*-th plot, predict the biomass for the *i*-th tree using the coefficient vector of $\hat{\beta}_p$ and applying the bias correction factor as described by Sprugel (1983).

(v) For the *j*-th plot, obtain the biomass at the stand level, in Mg ha^{-1}, using eqn. 3. Subsequently, calculate the average estimated and observed AGB across the 10 plots using eqn. 4 and eqn. 5, respectively:

$$\widehat{AGB}_j = \sum_{i=1}^{n_j} \hat{b}_{ij} \quad (3)$$

$$\overline{\widehat{AGB}} = \frac{1}{10} \sum_{j=1}^{10} \widehat{AGB}_j \quad (4)$$

$$\overline{AGB} = \frac{1}{10} \sum_{j=1}^{10} AGB_j \quad (5)$$

(vi) Estimate the mean error (ε_{rep}) and root mean square error ($RMSE_{rep}$) of the average AGB, in Mg ha^{-1}, using eqn. 6 and eqn. 7:

$$\varepsilon_{rep} = \overline{\widehat{AGB}} - \overline{AGB} \quad (6)$$

$$RMSE_{rep} = \sqrt{\frac{\sum_{i=1}^{n} (\widehat{AGB} | \overline{AGB})^2}{n}} \quad (7)$$

(vii) Replicate steps i-vi 5000 times.

(viii) Repeat steps i-vii for the remaining sample sizes.

Note that eqn. 4 is a model-assisted estimator of the mean AGB per unit area, eqn. 5 is a sample-based estimator of the mean AGB per unit area, and eqn. 6 and eqn. 7 are measures of accuracy of the mean AGB per unit area. Steps i-viii generate, for each sample size, two separate vectors containing 5000 mean errors and 5000 RMSE of forest biomass in Mg ha^{-1}. The last two steps of the algorithm involve measuring the precision and accuracy of the estimated mean AGB in Mg ha^{-1}, as described below.

(ix) For the *m*-th sample size, calculate the range (eqn. 8, eqn. 9) and the mean (eqn. 10, eqn. 11) of mean error and RMSE in Mg ha^{-1} across the replications:

$$A_{\varepsilon} = \varepsilon_{rep}(max) - \varepsilon_{rep}(min) \quad (8)$$

$$A_{RMSE} = RMSE_{rep}(max) - RMSE_{rep}(min) \quad (9)$$

$$\bar{\varepsilon} = \frac{1}{5000} \sum_{rep=1}^{5000} \varepsilon_{rep} \quad (10)$$

$$\overline{RMSE} = \frac{1}{5000} \sum_{rep=1}^{5000} RMSE_{rep} \quad (11)$$

(x) Repeat step ix for the remaining sample sizes.

The range of RMSE and errors was used to assess precision, as the common variance is not suitable in cases with a large number of replications. To evaluate accuracy, the mean RMSE and error across replications were considered.

To complement the evaluation of precision and accuracy, we calculated the probability of achieving an acceptable relative error threshold. For each calibration sample size, the proportion of replications in which the absolute relative error of the

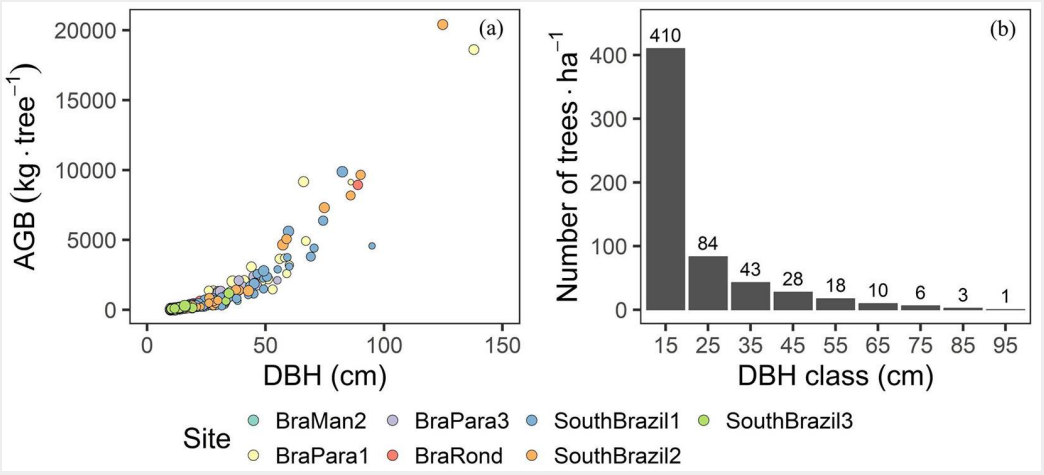

Fig. 2 - Relationship between aboveground biomass (AGB) and tree diameter at breast height (DBH) (a) and diameter distribution (b). Diameter distribution of the forest simulated (n. of trees = 6014) through the Monte Carlo method. In (a), each point represents a single tree that has been individually weighed. The size of the points corresponds to wood specific gravity.

mean AGB estimate was less than or equal to 10% was computed (eqn. 12). This probability was estimated from the distribution of 5,000 replicated simulations using an indicator function approach. Similarly, the probability of obtaining RMSE values below the same threshold was calculated (eqn. 13):

$$P_{10\%}^{\varepsilon} = \frac{1}{5000} \sum_{rep=1}^{5000} I(|\varepsilon_{rep}| \leq 10) \quad (12)$$

$$P_{10\%}^{RMSE} = \frac{1}{5000} \sum_{rep=1}^{5000} I(|RMSE_{rep}| \leq 10) \quad (13)$$

where $I()$ is an indicator function equal to 1 when the condition is satisfied and 0 otherwise. These metrics provide an intuitive measure of the reliability of the estimator across different calibration sample sizes.

To quantify the incremental improvement in precision associated with increasing calibration sample size, we calculated the percentage reduction in the range of RMSE and error between successive sample sizes. For each transition (*e.g.*, 13 to 25 trees, 25 to 50 trees, etc.), the relative decrease in the range was computed with respect to the previous sample size. This analysis allowed us to assess how much additional precision was gained when increasing the number of calibration trees and to identify potential stabilization of improvements at larger sample sizes.

Explaining error of forest biomass estimate through characteristics of model calibration dataset

In addition to analyzing precision and accuracy of the estimate of the mean AGB in Mg ha^{-1} as described above, we also related characteristics of each calibration dataset of the models to the RMSE and mean error (eqn. 6, eqn. 7) produced by the models. The descriptive statistics calculated in step ii of Part II of the algorithm were adopted as characteristics of the datasets. Note that the RMSE and error obtained from eqn. 6 and eqn. 7 and the descriptive statistics are calculated 5,000 times for each *m*-th sample size of the calibration dataset.

The first attempt to explain the error in terms of dataset characteristics involved using Pearson's correlation coefficient *r* (eqn. 14):

$$r = \frac{Cov(x, y)}{s_x \cdot s_y} \quad (14)$$

The normality of the data was assessed using the Shapiro-Wilk test. In this case, each descriptive statistic (minimum DBH, maximum DBH, etc.) for the datasets was correlated with the estimated error in forest biomass (Mg ha^{-1}). We considered a strong correlation when $r \geq 0.75$, moderate when $0.50 \leq r < 0.75$, and weak when $r < 0.50$. The calculation included only 50 data pairs, corresponding to the percentiles (0%, 2%, …, 98%, 100%) of the RMSE and mean error obtained from eqn. 6 and eqn. 7.

In summary, the analytical procedure consisted of (i) simulating an Amazonian forest using the Monte Carlo Method (MCM) and *m* out of *n* Bootstrap, (ii) estimating the average forest biomass using models calibrated with datasets of varying sample sizes, (iii) analyzing the precision and accuracy of the mean AGB estimate in Mg ha^{-1} using the models calibrated with different sample sizes, and (iv) relating characteristics of the calibration datasets with the errors produced by the models.

Tab. 2 - Descriptive statistics of the simulated forest (*n* = 6,014 trees) based on tree selection from the Amazonian sub-dataset. (DBH): diameter at 1.30 m in cm; (HT): total height in m; (AGB): aboveground biomass in Mg ha^{-1}; (SD): standard deviation; (CV): coefficient of variation.

Variable	Minimum	Mean	Median	Maximum	CV (%)
DBH (cm)	10.0	21.2	15.9	95.0	67.2
HT (m)	6.0	18.8	17.0	54.1	36.8
AGB (kg tree^{-1})	12.3	576.2	133.7	9879.0	221.2

Tab. 3 - Descriptive statistics of 5,000 mean error values for the mean AGB estimate (Mg ha^{-1}), including the percentage reduction relative to the previous sample size and the probability that |%error| ≤ 10%.

Sample size	Minimum	Mean	Median	Maximum	Range	% reduction *vs.* previous	P(|%error| ≤ 10%)
13	-244.3	17.6	5.1	1315.3	1559.6	-	0.67
25	-143.1	6.3	2.5	373.2	516.3	66.9	0.77
50	-100.6	3.1	0.7	198.3	298.8	42.1	0.86
100	-75.3	1.1	0.1	89.2	164.6	44.9	0.94
200	-50.9	0.4	-0.2	59.4	110.3	33.0	0.99
400	-37.3	-0.3	-0.7	41.0	78.3	29.0	1.00

Tab. 4 - Descriptive statistics of 5,000 RMSE values for the mean AGB estimate (Mg ha^{-1}), including the percentage reduction relative to the previous sample size and the probability that |%RMSE| ≤ 10%.

Sample size	Minimum	Mean	Median	Maximum	Range	% reduction *vs.* previous	P(|%RMSE| ≤ 10%)
13	0.0	58.7	41.7	1315.3	1315.3	-	0.43
25	0.0	36.0	28.2	373.2	373.2	71.6	0.59
50	0.0	23.7	19.2	198.3	198.3	46.9	0.77
100	0.0	16.1	13.2	89.2	89.2	55.0	0.91
200	0.0	11.3	9.5	59.4	59.3	33.5	0.99
400	0.0	8.1	7.0	41.0	41.0	30.9	1.00

Results

Our analytical procedure generated 5,000 mean errors and RMSE values using eqn. 6 and eqn. 7 for each sample size (ncalib.1 = 13, …, ncalib.6 = 400) in the calibration dataset. The histograms of the mean errors and RMSE for all sample sizes are depicted in Fig. 3. The histograms of the errors depicted an expected normal distribution, as confirmed by the Shapiro-Wilk test.

Descriptive statistics for the RMSE and errors are given in Tab. 3 and Tab. 4, where the range and mean values were obtained using eqn. 8-11, respectively. We noted that the amplitude of the RMSE and errors, which reflect the precision of the mean AGB estimate (Mg ha^{-1}), decreases as the sample size of the calibration dataset increases. The difference between the largest and smallest amplitudes indicates that the biomass estimate's precision increased by up to 32 times with the increase in sample size.

The accuracy of the biomass estimate, as reflected by the mean values in Tab. 3 and Tab. 4, showed less variation across sample sizes than the precision. We observed a reduction in the mean RMSE (from 58.7 to 8.1 Mg ha^{-1}) and a decrease in the mean error (from 17.6 to -0.3 Mg ha^{-1}) as the sample size in the calibration dataset increased. Also, the RMSE and error medians decreased with sample size, though less than the mean did. The positive error medians observed for sample sizes up to 100 indicate that biomass per unit area was frequently overestimated.

To quantify the effect of increasing calibration sample size on precision, we calculated the marginal reduction in RMSE and

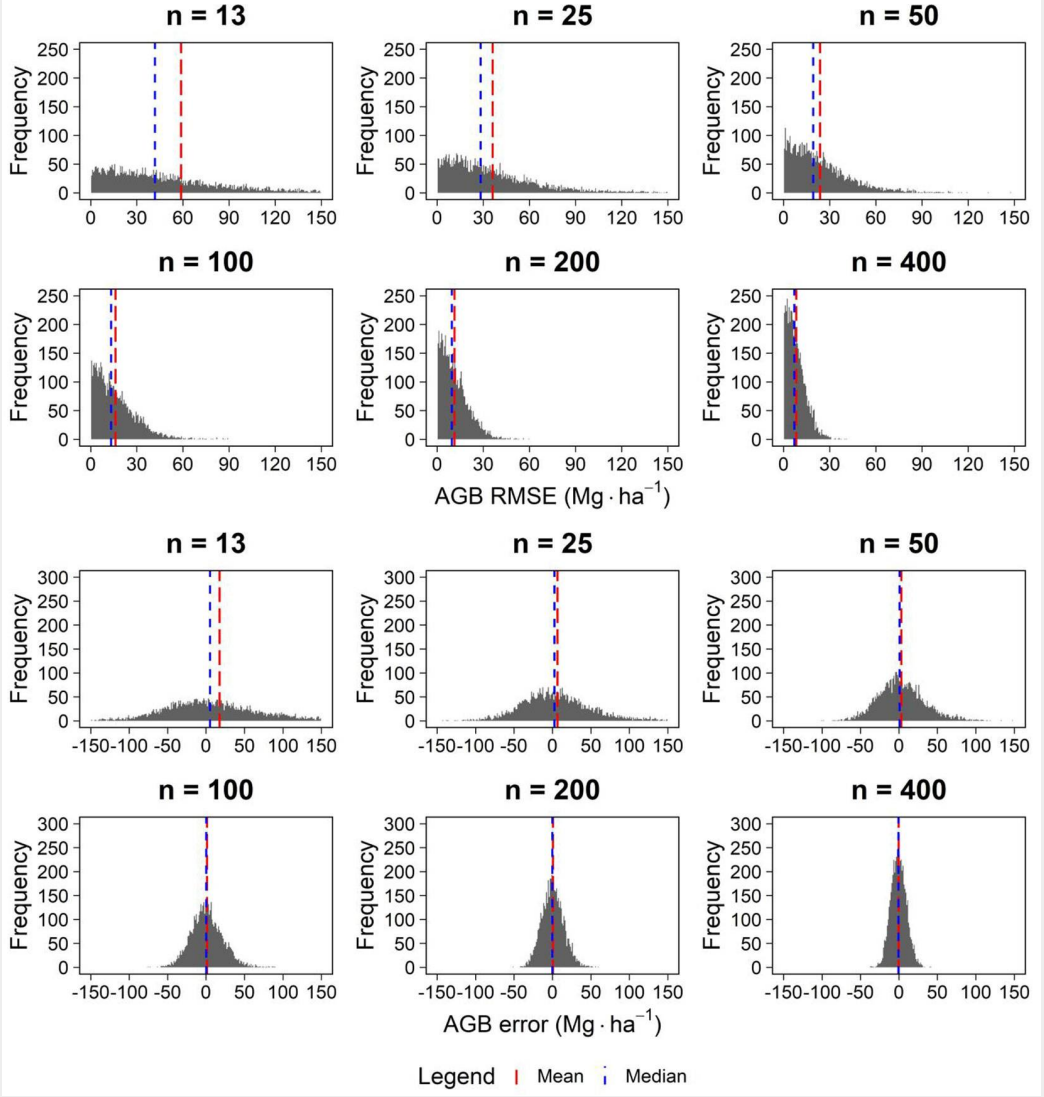

Fig. 3 - Histograms of RMSE and errors (n_{rep}=5,000) of the estimate of the mean AGB in Mg ha^{-1} (calculated using eqn. 6-7). Solid red vertical line: mean error. Dashed blue vertical line: median. Solid black vertical line: 0 (zero).

error across successive sample sizes (Tab. 3, Tab. 4). The largest improvements occurred at smaller sample sizes, with RMSE reductions of 71.6% (13-25), 46.9% (25-50), 55.0% (50-100), 33.5% (100-200), and 30.9% (200-400). A similar pattern was observed for error range (66.9%, 42.1%, 44.9%, 33.0%, and 29.0%, respectively), indicating diminishing returns and stabilization of precision gains beyond 200 trees.

The probability of achieving acceptable relative error and relative RMSE (eqn. 12-13) increased markedly with calibration sample size. For error (Tab. 3), this probability rose from 0.67 at 13 trees to 0.86 at 50 trees and reached ~1.00 from 200 trees onward, indicating the stabilization of the average bias at relatively large sample sizes. For RMSE (Tab. 4), the probability increased from 0.43 (13 trees) to 1.00 (400 trees). Notably, with 100 trees, the probability of obtaining RMSE ≤ 10% already exceeded 0.90, demonstrating high reliability at this calibration size. However, with 200 trees, the probability reached approximately 1.0 for both error and RMSE. These results indicate that improvements diminish beyond 200 trees, supporting this value as a practical threshold for stable and precise mean AGB estimation.

The correlation between the estimated RMSE and error of forest biomass and each descriptive statistic was considered weak for all tested sample sizes, as the values were less than 0.50, except between the error and the coefficient of variation of DBH when the sample size was 200 (r = 0.51, indicating a moderate correlation). However, none of the metrics explained the behavior of the estimation error, as evidenced by the low correlation values.

The percentage error of the mean AGB

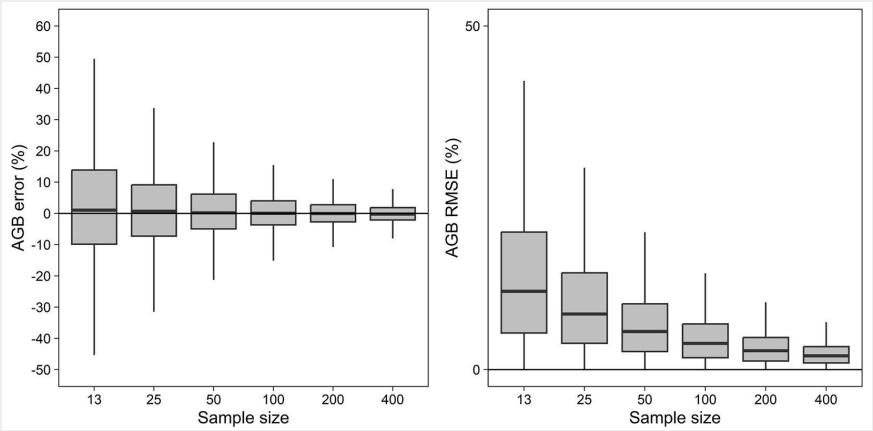

Fig. 4 - Boxplots of relative errors of the mean AGB estimate (Mg ha^{-1}) for different sample sizes (number of trees) of model calibration datasets.

estimate for the tested sample sizes is shown in Fig. 4. The sample sizes of 200 and 400 exhibited the lowest percentage error ranges, specifically -14.70% to 17.13% and -10.76% to 11.84%, respectively. The smaller the number of samples, the greater the error range, with amplitudes of 450.04%, 148.98%, 86.23%, 47.49%, 31.83%, and 22.61% for 13, 25, 50, 100, 200, and 400 samples, respectively. The same patterns were observed for RMSE values. Notably, when the sample size was doubled from 200 to 400, the error and RMSE amplitudes decreased by 9.22% and 11.84%, respectively. Almost all mean values of percentage error were positive, ranging from 5.08% for the smallest sample size to -0.11% for 200.

Discussion

The high variability of tree attributes (Tab. 1) is common in tropical forests, as they are characterized by structural and floristic diversity, along with the presence of large trees (*e.g.*, Max. AGB = 20,416 kg tree^{-1}, Tab. 1 – Akindele & LeMay 2006, Ter Steege et al. 2013, Worbes & Junk 1999). This high population heterogeneity increases the challenge of biomass estimation in natural forests (Cysneiros et al. 2017). However, large trees must be included in the sample, as these individuals contribute a substantial proportion of the biomass and accurately reflect the population's composition (Feldpausch et al. 2012, Aabeyir et al. 2020). Using 5,000 repetitions across the different sample sizes in this study effectively reduced the effect of sample composition on the regression model fit, allowing for a more reliable analysis of model adjustments, consistent with the findings of Leão et al. (2021).

The results underscore the reliability of the equations employed to estimate stand-level biomass, as they consistently approached the true mean despite the high variability observed in smaller samples. Furthermore, the normal distribution of the errors suggests a solid theoretical foundation.

The underestimation of biomass increased with sample size, while RMSE values decreased (Tab. 3, Tab. 4), suggesting a reduction in the influence of larger trees in the modeling process. Given the inverted-J distribution (*e.g.*, most observations are of small trees, with only a small proportion of large trees), the larger the sample, the greater the tendency for more small trees to be selected. In line with this pattern, evidence in the literature indicates that allometric models focusing on large trees tend to overestimate biomass for smaller trees (Deans et al. 1996, Van Breugel et al. 2011).

In contrast, Dutca et al. (2020) found that although sample size strongly influenced accuracy, it showed limited impact on the precision of biomass predictions. Conversely, Duncanson et al. (2015), studying temperate forests across multiple sites, reported that the accuracy of allometric models is highly affected by sample size, with small samples yielding biased coefficients. Similarly, Aabeyir et al. (2020) indicated that calibration with larger sample sizes yields more reliable coefficients. This behavior aligns with the gradual increase in mean AGB estimate (Mg ha^{-1}) precision as more trees were used for model calibration (Tab. 3, Tab. 4, Fig. 4), indicating that small samples may not adequately represent the sampled population; thus, the larger the sample size, the greater the likelihood that additional observations (trees) yield consistent results (Brooks & Barcikowski 2012). Although model diagnostics and parameter significance were not evaluated individually for each fitted equation, the observed variability among simulations reflects not only differences in sample composition but also the instability of parameter estimates associated with limited calibration datasets. Additionally, Chave et al. (2004), in their study on biomass estimation in tropical forests, recommended that allometric equations be constructed from large samples, ideally comprising at least 100 trees spanning a broad diameter range. Such findings underscore the importance of evaluating the optimal sample size for forest management (Maxwell et al. 2008), as the number of sampled trees directly impacts fieldwork labor, survey costs, and the accuracy of models applied in forest inventory.

Beyond comparisons with previous sample-size studies, these findings have implications for broader biomass estimation methodologies. While many studies emphasize improvements in model form, predictor selection, or spatial scaling strategies (Chave et al. 2014, Adame-Campos et al. 2019), comparatively less attention has been given to the structural role of calibration sample size in stand-level biomass inference (Fassnacht et al. 2014, Hetzer et al. 2020). Plot-based expansion approaches and remote sensing frameworks typically prioritize plot representativeness or landscape-level predictive performance (Schepaschenko et al. 2018, Potapov et al. 2021, Campbell et al. 2021), often assuming that tree-level allometric relationships are sufficiently stable. However, because large-scale biomass products ultimately depend on these underlying relationships (Leite et al. 2022), the present results demonstrate that calibration sample size alone can substantially influence the variability of stand-level mean AGB estimates, even when model structure remains unchanged. These findings reinforce the importance of treating calibration design as a structural component of model reliability rather than merely an operational constraint.

Although the literature indicates that sample sizes can exceed 200 trees (Romero et al. 2022), they often do not (Woortmann et al. 2018, Moura et al. 2022, De Melo et al. 2024). Hetzer et al. (2020) found that around 100 plots are necessary for biomass estimations; however, in the present study, a significant reduction in error was observed when the sample size was increased from 100 to 200 trees (*e.g.*, reduction of 33.0% for error and 33.5% for RMSE – Tab. 3, Tab. 4). While sampling 200 trees provided good precision, the improvement in precision from doubling the sample size (from 200 to 400) was not substantial enough to justify the additional effort. Furthermore, a sample size of 200 is preferable, as it increases the chances of adequately representing a diverse range of tree sizes, as recommended by Chave et al. (2004). Additionally, recognizing that it is not always possible to obtain such a sample size in practice, one should consider that when working with a limited sample size, it is crucial to be aware of the trade-off between sample size and the precision of the mean AGB estimate (Mg ha^{-1}) at the stand level.

The ability to replicate findings is essential across all scientific disciplines to ensure the stability and generalizability of results. A reliable approach to guarantee this generalization is to apply these models to independent samples separate from the data used for model fitting. In this context, a limitation of the present study is that the same dataset was used to simulate forest structure and calibrate biomass models, which may influence the independence of model validation and should be considered when interpreting the generalizability of the results. Thus, developing a model with high accuracy should be regarded as only the first step in the validation process, as common validation methods do not eliminate the need to validate models on other samples (Brooks & Barcikowski 2012). As observed in this study, an insufficient calibration sample size may yield satisfactory results for a specific dataset but present limitations when transferred to other forest conditions, leading to wasted time, field effort, and costs for inconclusive or less applicable outcomes. Consequently, simulation emerges as an essential tool to support the validation process, allowing assessment of model performance at the stand level, a practice still rarely applied in the development of biomass models (Dutca et al. 2020, David et al. 2022). Finally, although the forest structure here was simulated according to a typical inverted-J distribution representative of central Amazon forests, the variability level of the simulated stands remained within the ecological range reported by Laurance et al. (2009). Therefore, our recommendation applies to forests with comparable structural heterogeneity, and future research should extend this framework using independent datasets and forests with contrasting diameter distributions and lower structural variability.

Conclusions

The sample size affects the precision of the mean AGB estimate (Mg ha^{-1}) per unit area. Larger sample sizes yield better preci-

sion, but it does not affect accuracy. To compose a model calibration dataset, one should consider that precision can improve as the sample size increases. To estimate the mean and total forest biomass in the Amazon using tree-level modeling, we recommend a calibration dataset comprising 200 randomly sampled trees, with DBHs spanning the forest range.

Acknowledgements

David HC: conceptualization, methodology, supervision and writing - original draft, review & editing; Pereira FR: formal analysis, investigation, methodology, validation, and writing - original draft; Almeida TC: formal analysis, investigation, methodology, validation, and writing - original draft; García RA: methodology and writing; Galvão LMR: methodology and writing; Chivale XS: methodology and writing; Behling A: methodology and writing - review; Pelissari AL: methodology and writing - review; Ebling AA: methodology and writing - review.

References

Aabeyir R, Adu-Bredu S, Agyare WA, Weir MJ (2020). Allometric models for estimating aboveground biomass in the tropical woodlands of Ghana, West Africa. Forest Ecosystems 7: 1-23. - doi: 10.1186/s40663-020-00250-3

Adame-Campos RL, Ghilardi A, Gao Y, Paneque-Gálvez J, Mas JF (2019). Variables selection for aboveground biomass estimations using satellite data: a comparison between relative importance approach and stepwise Akaike's information criterion. ISPRS International Journal of Geo-Information 8: 245. - doi: 10.3390/ijgi8060245

Akindele SO, LeMay VM (2006). Development of tree volume equations for common timber species in the tropical rain forest area of Nigeria. Forest Ecology and Management 226 (1-3): 41-48. - doi: 10.1016/j.foreco.2006.01.022

Antúnez P, Wehenkel C, Basave-Villalobos E, Calixto-Valencia CG, Valenzuela-Encinas C, Ruiz-Aquino F, Sarmiento-Bustos D (2025). Predictive modeling of volume and biomass in *Pinus pseudostrobus* using machine learning and allometric approaches. Forest Science and Technology 21: 110-122. - doi: 10.1080/21580103.2025.2456295

Assis TO, De Aguiar APD, Von Randow C, De Paula Gomes DM, Kury JN, Ometto JPH, Nobre CA (2020). CO_2 emissions from forest degradation in Brazilian Amazon. Environmental Research Letters 15 (10): 104035. - doi: 10.1088/1748-9326/ab9cfc

Brooks GP, Barcikowski RS (2012). The PEAR method for sample sizes in multiple linear regression. Multiple Linear Regression Viewpoints 38 (2): 1-16. [online] URL: http://ojs.lib.ua.edu/glmj/article/view/262

Campbell MJ, Dennison PE, Kerr KL, Brewer SC, Anderegg WR (2021). Scaled biomass estimation in woodland ecosystems: testing the individual and combined capacities of satellite multispectral and lidar data. Remote Sensing of Environment 262: 112511. - doi: 10.1016/j.rse.2021.112511

Chauvel A, Lucas Y, Boulet R (1987). On the genesis of the soil mantle of the region of Manaus, Central Amazonia, Brazil. Experientia 43: 234-241. - doi: 10.1007/BF01945546

Chave J, Condit R, Aguilar S, Hernandez A, Lao S, Perez R (2004). Error propagation and scaling for tropical forest biomass estimates. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359 (1443): 409-420. - doi: 10.1098/rstb.2003.1425

Chave J, Andalo C, Brown S, Cairns MA, Chambers JQ, Eamus D, Folster H, Fromard F, Higuchi N, Kira T, Lescure JP, Nelson BW, Ogawa H, Puig H, Riera B, Yamakura T (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145: 87-99. - doi: 10.1007/s00442-005-0100-x

Chave J, Réjou Méchain M, Búrquez A, Chidumayo E, Colgan MS, Delitti WBC, Duque A, Eid T, Fearnside PM, Goodman RC, Henry M, Martínez-Yrízar A, Mugasha WA, Muller-Landau HC, Mencuccini M, Nelson BW, Ngomanda A, Nogueira EM, Ortiz-Malavassi E, Pélissier R, Ploton P, Ryan CM, Saldarriaga JG, Vieilledent G (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. Global Change Biology 20 (10): 3177-3190. - doi: 10.1111/gcb.12629

Cuni-Sanchez A, Sullivan MJP, Platts PJ, Lewis SL, Marchant R, Imani G, Hubau W, Abiem I, Adhikari H, Albrecht T, Altman J, Amani C, Aneseyee AB, Avitabile V, Banin L, Batumike R, Bauters M, Beeckman H, Begne SK, Bennett AC, Bitariho R, Boeckx P, Bogaert J, Bräuning A, Bulonvu F, Burgess ND, Calders K, Chapman C, Chapman H, Comiskey J, De Haulleville T, Decuyper M, DeVries B, Dolezal J, Droissart V, Ewango C, Feyera S, Gebrekirstos A, Gereau R, Gilpin M, Hakizimana D, Hall J, Hamilton A, Hardy O, Hart T, Heiskanen J, Hemp A, Herold M, Hiltner U, Horak D, Kamdem MN, Kayijamahe C, Kenfack D, Kinyanjui MJ, Klein J, Lisingo J, Lovett J, Lung M, Makana JR, Malhi Y, Marshall A, Martin EH, Mitchard ETA, Morel A, Mukendi JT, Muller T, Nchu F, Nyirambangutse B, Okello J, Peh KSH, Pellikka P, Phillips OL, Plumptre A, Qie L, Rovero F, Sainge MN, Schmitt CB, Sedlacek O, Ngute ASK, Sheil D, Sheleme D, Simegn TY, Simo-Droissart M, Sonké B, Soromessa T, Sunderland T, Svoboda M, Taedoumg H, Taplin J, Taylor D, Thomas SC, Timberlake J, Tuagben D, Umunay P, Uzabaho E, Verbeeck H, Vleminckx J, Wallin G, Wheeler C, Willcock S, Woods JT, Zibera E (2021). High aboveground carbon stock of African tropical montane forests. Nature 596: 536-542. - doi: 10.1038/s41586-021-03728-4

Cysneiros VC, Pelissari AL, Machado SA, Figueiredo-Filho A, David HC (2017). Cluster and discriminant analyses for stem volume modelling of tree species groups in an Amazon rainforest. Journal of Tropical Forest Science 29 (3): 325-333. [online] URL: http://www.jstor.org/stable/44272910

David HC, Barbosa RI, Vibrans AC, Watzlawick LF, Trautenmuller JW, Balbinot R, Ribeiro SC, Jacovine LAG, Corte APD, Sanquetta CR, Silva AC, Freitas JV, MacFarlane DW (2022). The tropical biomass and carbon project - An application for forest biomass and carbon estimates. Ecological Modelling 472: 110067. - doi: 10.1016/j.ecolmodel.2022.110067

De Melo AWF, Lima AJN, De Oliveira MVN, Dos Santos J, Brown IF, Do Amaral EF, Da Silva SS, Oliveira I, Camargo PB, Higuchi N (2024). To improve estimates of neotropical forest carbon stocks more direct measurements are needed: an example from the Southwestern Amazon. Forest Ecology and Management 570: 122195. - doi: 10.1016/j.foreco.2024.122195

Deans JD, Moran J, Grace J (1996). Biomass relationships for tree species in regenerating semideciduous tropical moist forest in Cameroon. Forest Ecology and Management 88 (3): 215-225. - doi: 10.1016/S0378-1127(96)03843-1

Djomo AN, Ibrahima A, Saborowski J, Gravenhorst G (2010). Allometric equations for biomass estimations in Cameroon and pan moist tropical equations including biomass data from Africa. Forest Ecology and Management 260 (10): 1873-1885. - doi: 10.1016/j.foreco.2010.08.034

Duncanson L, Rourke O, Dubayah R (2015). Small sample sizes yield biased allometric equations in temperate forests. Scientific Reports 5 (1): 89. - doi: 10.1038/srep17153

Dutca I, Mather R, Ioras F (2020). Sampling trees to develop allometric biomass models: how does tree selection affect model prediction accuracy and precision? Ecological Indicators 117: 106553. - doi: 10.1016/j.ecolind.2020.106553

Ekoungoulou R, Nzala D, Liu X, Niu S (2018). Tree biomass estimation in central African forests using allometric models. Open Journal of Ecology 8 (3): 209-237. - doi: 10.4236/oje.2018.83014

Fassnacht FE, Hartig F, Latifi H, Berger C, Hernández J, Corvalán P, Koch B (2014). Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. Remote Sensing of Environment 154: 102-114. - doi: 10.1016/j.rse.2014.07.028

Fearnside PM (2018). Brazil's Amazonian forest carbon: the key to Southern Amazonia's significance for global climate. Regional Environmental Change 18: 47-61. - doi: 10.1007/s10113-016-1007-2

Feldpausch TR, Lloyd J, Lewis SL, Brienen RJW, Gloor M, Monteagudo Mendoza A, Lopez-Gonzalez G, Banin L, Abu Salim K, Affum-Baffoe K, Alexiades M, Almeida S, Amaral I, Andrade A, Aragão LEO, Araujo Murakami A, Arets EJM, Arroyo LA, Aymard CGA, Baker TR, Bánki OS, Berry NJ, Cardozo N, Chave J, Comiskey JA, Alvarez E, De Oliveira A, Di Fiore A, Djagbletey G, Domingues TF, Erwin TL, Fearnside PM, França MB, Freitas MA, Higuchi N, Honorio CE, Iida Y, Jiménez E, Kassim AR, Killeen TJ, Laurance WF, Lovett JC, Malhi Y, Marimon BS, Marimon-Junior BH, Lenza E, Marshall AR, Mendoza C, Metcalfe DJ, Mitchard ETA, Neill DA, Nelson BW, Nilus R, Nogueira EM, Parada A, Peh KSH, Pena Cruz A, Peñuela MC, Pitman NCA, Prieto A, Quesada CA, Ramírez F, Ramírez-Angulo H, Reitsma JM, Rudas A, Saiz G, Salomão RP, Schwarz M, Silva N, Silva-Espejo JE, Silveira M, Sonké B, Stropp J, Taedoumg HE, Tan S, Steege H, Terborgh J, Torello-Raventos M, Van Der Heijden GMF, Vásquez R, Vilanova E, Vos VA, White L, Willcock S, Woell H, Phillips OL (2012). Tree height integrated into pantropical forest biomass estimates. Biogeosciences 9: 3381-3403. - doi: 10.5194/bg-9-3381-2012

García RA, Galvão LMR, Chivale XS, Almeida TC, Pereira FR, Martins-Neto RP, Sanquetta CR, David HC (2026). Comparing 30 tree biomass models to estimate forest biomass in the Amazon. Forests 17 (2): 213. - doi: 10.3390/f17020213
Goussanou CA, Guendehou S, Assogbadjo AE, Kaire M, Sinsin B, Cuni-Sanchez A (2016). Specific and generic stem biomass and volume models of tree species in a West African tropical semi-deciduous forest. Silva Fennica 50 (2): 1474. - doi: 10.14214/sf.1474
Hetzer J, Hut A, Wiegand T, Dobner HJ, Fischer R (2020). An analysis of forest biomass sampling strategies across scales. Biogeosciences 17 (6): 1673-1683. - doi: 10.5194/bg-17-1673-2020
Huy B, Truong NQ, Khiem NQ, Poudel KP, Temesgen H (2022). Deep learning models for improved reliability of tree aboveground biomass prediction in the tropical evergreen broadleaf forests. Forest Ecology and Management 508: 120031. - doi: 10.1016/j.foreco.2022.120031
Jenkins DG, Quintana-Ascencio PF (2020). A solution to minimum sample size for regressions. PLoS One 15: e0229345. - doi: 10.1371/journal.pone.0229345
Laurance SG, Laurance WF, Nascimento HE, Andrade A, Fearnside PM, Rebello ER, Condit R (2009). Long-term variation in Amazon forest dynamics. Journal of Vegetation Science 20 (2): 323-333. - doi: 10.1111/j.1654-1103.2009.01044.x
Laurance WF, Lovejoy TE, Vasconcelos HL, Bruna EM, Didham RK, Stouffer PC, Gascon C, Bierregaard RO, Laurence SG, Sampaio E (2002). Ecosystem decay of Amazonian forest fragments: a 22-year investigation. Conservation Biology 16 (3): 605-618. - doi: 10.1046/j.1523-1739.2002.01025.x
Leão FM, Nascimento RGM, Emmert F, Santos GGA, Caldeira NAM, Miranda IS (2021). How many trees are necessary to fit an accurate volume model for the Amazon forest? A site-dependent analysis. Forest Ecology and Management 480: 118652. - doi: 10.1016/j.foreco.2020.118652
Leite RV, Silva CA, Broadbent EN, Amaral CH, Liesenberg V, Almeida DRA, Mohan M, Godinho S, Cardil A, Hamamura C, Faria BL, Brancalion PHS, Hirsch A, Marcatti GE, Dalla Corte AP, Zambrano AMA, Costa MBT, Matricardi EAT, Silva AL, Goya LRR, Valbuena R, Mendonça BAF, Silva Junior CHL, Aragão LEOC, García M, Liang J, Merrick T, Hudak AT, Xiao J, Hancock S, Duncanson L, Ferreira MP, Valle D, Saatchi S, Klauberg C (2022). Large-scale multi-layer fuel load characterization in tropical savanna using GEDI spaceborne lidar data. Remote Sensing of Environment 268 (1-2): 112764. - doi: 10.1016/j.rse.2021.112764
Lovejoy TE (1986). Edge and other effects of isolation on Amazon forest fragments. In: "Conservation Biology: The Science of Scarcity and Diversity". Sinauer Associates, Sunderland, MA, USA, pp. 257-285.
Maxwell SE, Kelley K, Rausch JR (2008). Sample size planning for statistical power and accuracy in parameter estimation. Annual Review of Psychology 59: 537-563. - doi: 10.1146/annurev.psych.59.103006.093735
Moura M, Sanquetta M, Sanquetta C, Bastos A, Corte A (2022). Dynamics of biomass stocks in areas of secondary forest located in western Amazon between 2016 and 2020. Agriculture and Forestry/Poljoprivreda i Šumarstv 68 (1): 7-23.
Potapov P, Li X, Hernandez-Serna A, Tyukavina A, Hansen MC, Kommareddy A, Pickens A, Turubanova S, Tang H, Silva CE, Armston J, Dubayah R, Blair JB, Hofton M (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment 253 (4): 112165. - doi: 10.1016/j.rse.2020.112165
Rajput D, Wang WJ, Chen CC (2023). Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics 24 (1): 207. - doi: 10.1186/s12859-023-05156-9
Romero FMB, Jacovine LAG, Torres CMME, Ribeiro SC, Da Rocha SJSS, Novais TDNO, Gaspar RO, Da Silva LF, Vidal E, Leite HG, Staudhammer CL, Fearnside PM (2022). Aboveground biomass allometric models for large trees in southwestern Amazonia. Trees, Forests and People 9: 100317. - doi: 10.1016/j.tfp.2022.100317
Schepaschenko D, Moltchanova E, Shvidenko A, Blyshchyk V, Dmitriev E, Martynenko O, See L, Kraxner F (2018). Improved estimates of biomass expansion factors for Russian forests. Forests 9 (6): 312. - doi: 10.3390/f9060312
Sprugel DG (1983). Correcting for bias in log-transformed allometric equations. Ecology 64: 209-210. - doi: 10.2307/1937343
Sullivan MJ, Lewis SL, Hubau W, Qie L, Baker TR, Banin LF, Phillips OL (2018). Field methods for sampling tree height for tropical forest biomass estimation. Methods in Ecology and Evolution 9 (5): 1179-1189. - doi: 10.1111/2041-210X.12962
Temesgen H, Affleck D, Poudel K, Gray A, Sessions J (2015). A review of the challenges and opportunities in estimating aboveground forest biomass using tree-level models. Scandinavian Journal of Forest Research 30 (4): 326-335. - doi: 10.1080/02827581.2015.1012114
Ter Steege H, Pitman NC, Sabatier D, Baraloto C, Salomão RP, Guevara JE, Phillips OL, Castilho CV, Magnusson WE, Molino JF, Monteagudo A, Núñez Vargas P, Montero JC, Feldpausch TR, Coronado EN, Killeen TJ, Mostacedo B, Vasquez R, Assis RL, Terborgh J, Wittmann F, Andrade A, Laurance WF, Laurance SG, Marimon BS, Marimon BH, Guimarães Vieira IC, Amaral IL, Brienen R, Castellanos H, Cárdenas López D, Duivenvoorden JF, Mogollón HF, Matos FD, Dávila N, García-Villacorta R, Stevenson Diaz PR, Costa F, Emilio T, Levis C, Schietti J, Souza P, Alonso A, Dallmeier F, Montoya AJ, Fernandez Piedade MT, Araujo-Murakami A, Arroyo L, Gribel R, Fine PV, Peres CA, Toledo M, Aymard C. GA, Baker TR, Cerón C, Engel J, Henkel TW, Maas P, Petronelli P, Stropp J, Zartman CE, Daly D, Neill D, Silveira M, Paredes MR, Chave J, Lima Filho Dd, Jørgensen PM, Fuentes A, Schöngart J, Cornejo Valverde F, Di Fiore A, Jimenez EM, Peñuela Mora MC, Phillips JF, Rivas G, van Andel TR, von Hildebrand P, Hoffman B, Zent EL, Malhi Y, Prieto A, Rudas A, Ruschell AR, Silva N, Vos V, Zent S, Oliveira AA, Schutz AC, Gonzales T, Trindade Nascimento M, Ramirez-Angulo H, Sierra R, Tirado M, Umaña Medina MN, van der Heijden G, Vela CI, Vilanova Torre E, Vriesendorp C, Wang O, Young KR, Baider C, Balslev H, Ferreira C, Mesones I, Torres-Lezama A, Urrego Giraldo LE, Zagt R, Alexiades MN, Hernandez L, Huamantupa-Chuquimaco I, Milliken W, Palacios Cuenca W, Pauletto D, Valderrama Sandoval E, Valenzuela Gamarra L, Dexter KG, Feeley K, Lopez-Gonzalez G, Silman MR (2013). Hyperdominance in the Amazonian tree flora. Science 342 (6156): 592. - doi: 10.1126/science.1243092
Van Breugel M, Ransijn J, Craven D, Bongers F, Hall JS (2011). Estimating carbon stock in secondary forests: decisions and uncertainties associated with allometric biomass models. Forest Ecology and Management 262 (8): 1648-1657. - doi: 10.1016/j.foreco.2011.07.018
Vorster AG, Evangelista PH, Stovall AE, Ex S (2020). Variability and uncertainty in forest biomass estimates from the tree to landscape scale: the role of allometric equations. Carbon Balance and Management 15 (1): 283. - doi: 10.1186/s13021-020-00143-6
Woortmann CP, Higuchi N, Santos Jd, Silva RP (2018). Allometric equations for total, above- and below-ground biomass and carbon of the Amazonian forest type known as campinarana. Acta Amazonica 48 (2): 85-92. - doi: 10.1590/1809-4392201700673
Worbes M, Junk WJ (1999). How old are tropical trees? The persistence of a myth. IAWA Journal 20 (3): 255-260. - doi: 10.1163/22941932-90000689