Available online at http://www.ifgdg.org

Int. J. Biol. Chem. Sci. 18(6): 2205-2220, December 2024

International Journal of Biological and Chemical Sciences

ISSN 1997-342X (Online), ISSN 1991-8631 (Print)

Original Paper http://ajol.info/index.php/ijbcs http://indexmedicus.afro.who.int

Optimal tree sampling for ecosystem-specific biomass allometry modeling in Congo Basin forests

Gabin Finagnon LALY[1], Gilbert ATINDOGBE[1], Afouda Hospice AKPO[1], Titilayo Oyélèyè Fafounkè DOTCHAMOU[1], Barthélemy KASSA[2] and Noël Houédougbé FONTON[1*]

[1]*Laboratoire d'étude et de Recherche en Statistique Appliquée et Biométrie, Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, Bénin.*

[2] *Laboratoire d'Ecologie Appliquée, Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, Bénin.*

[*]*Corresponding author; E-mail: hnfonton@gmail.com; Tel. +00299 97 27 45 09.*

Received: 04-04-2024 Accepted: 23-07-2024 Published: 31-12-2024

ABSTRACT

Allometric equations are fundamental for estimating biomass in forests and their accuracy depends heavily on the quality and representativeness of the data used to construct them. This study aimed to benchmark tree sampling techniques and determine the optimal number of sample trees for constructing allometric equations. Ten sampling strategies consisting of the combination of two allometric models and five sampling techniques were evaluated. Random sampling techniques and four sampling techniques with eight diameter size-classes based on cumulative frequency distribution were compared. A wide range of sample data was simulated using a parametric resampling method to ensure unbiased sampling and a representative spread of observations. Data were derived from 15 inventory plots in three Congo Basin forest reserves. Results showed that uncertainty due to differences in size class distribution was minimized by a sampling technique, which effectively represents large trees. High sample sizes were required for precision in the absence of large trees. Sample sizes uncertainty was influenced by stand characteristics, mainly the shape of the inventory plot and data distribution. This study reveals that the biomass prediction uncertainty depends on the population's specific characteristics, the type of allometric model used, and the representativeness of large trees in the sample.

Keywords: Allometric equation, Biomass estimation, Sampling techniques, Parametric resampling, Diameter size-classes, Congo Basin.

INTRODUCTION

The successful implementation of climate mitigation and carbon sequestration depends on the quality of information on the carbon balance and the capacity of forest ecosystems to mitigate the effects of changing climate (Lung and Espira, 2015). This information depends on the robust estimate of carbon stocks (Saatchi et al., 2011). Accurate estimates for woody biomass were of scientific concern for the biomass markets and carbon trading. The most biomass pool directly affected by deforestation and degradation was the aboveground biomass (Saatchi et al., 2011).

Uncertainty on the amount of forest biomass has been pointed out by Lewis et al.

DOI : https://dx.doi.org/10.4314/ijbcs.v18i6.12

(2009), Chave et al. (2014). Modern approaches such as remote sensing or LiDAR were of interest for the estimation of the aboveground biomass (AGB). Allometric equations constitute the common tool for estimating biomass and forest carbon stock (van Breugel et al., 2011; 2016; Ganamé et al., 2021). They were also used to quantify land-based sequestration activities that constituted a key component so that more attention must be given for the accuracy of the methods used for their development (Roxburh et al., 2015).

Appropriate allometric equation becomes a major concern for an accurate estimation of forest biomass as pointed out by Rutishauser et al. (2013) and for the emission factor (Picard et al., 2016). It is also the preference between available models (from other areas or pan-tropical) and the local models as well as the choice of predictors (van Breugel et al., 2011). Allometric equations are established with harvested trees, operations that are manual, expensive and time-consuming. Despite the difficulties, allometric equations were reported more accurate and flexible (Jalkanen et al., 2005) for biomass estimation. Therefore, pantropical allometric equations for AGB estimation were established with large dataset from many continents (Chave et al., 2005, 2014) including data from several study sites in Africa (Djomo et al., 2016).

Some studies have reported the use of site-specific or ecosystem specific allometric equations for better estimate of biomass (Manzo et al., 2015; Dembele et al., 2023; Ngomanda et al., 2014; Djomo et al. 2016). In the Congo basin, multi-species allometric equations were established but the sampling strategies were variable. Ebuyi et al. (2011) harvested 12 tree-samples including three species: *Gilbertiodendron dewevrei* (Limbali), *Autranella congolensis* (Mukulungu), and *Drypetes likwa* with tree diameters at breast height (dbh) ranging from 24.4 to 52.2 cm. The 71 trees belonging to 31 species sampled by Djomo et al. (2010) were harvested in three 10 m × 10 m (100 m²) squared plots but only 4.2% of the trees had dbh greater than 10.0 cm. Fayolle et al. (2013) harvested 138 tree-samples belonging to 42 species and distributed over a wide range of dbh from 5.3 to 192.5 cm. A particular sampling had been observed with Ngomanda et al. (2014) who sampled 101 trees, belonging to 10 species and distributed approximately 10 trees per species and per diameter size-class; tree dbh ranged from 11.8 cm to 109.4 cm. Fayolle et al. (2018) compiled dataset from six sites in the Congo basin, comprising 845 tropical trees belonging to 55 African species and covering a large range of diameters 10–208 cm. One can reasonably admit that there is no standard strategy for trees sampling for the establishment of ecosystem-specific allometric equations for biomass estimation.

Trees sampling strategy must follow a suitable sampling methodology. The number of sample trees appeared correlated with the available budget and the recommendation of Chave et al. (2004) and van Breugel et al. (2011). Roxburgh et al. (2015) did an excellent work in this area but for plantation forests. The variability of the biomass of large trees was put in evidence by Lewis et al. (2009) in particular for the Congo basin forests and Romero et al. (2022) in Amazon rainforest. Logistically, biomass data-gathering of large trees requires important equipment. Lung and Espira (2015) reported that diameter, stem densities of large trees, and wood density were the most important variables influencing biomass, accounting for over 75% of the variation in the estimates. Hence, the aim of this study was to answer which sampling strategies for reducing uncertainty in allometric biomass estimation by comparing sampling techniques, determining a sufficient number of sampled trees according to forest stand structures, and assessing the importance of including largest trees.

MATERIALS AND METHODS

Data set used

For a given ecosystem, the stand data easily available was from temporary and permanent plots of the management and the logging inventories. Therefore, the inventory plot dataset from 15 permanent plots of the Central African Regional Program for the

Environment (CARPE) was used. The plots were established in three forest reserves: Dzanga Sangha (Balinga et al., 2006), Monts de Cristal (Sunderland et al., 2004) and Waka (Balinga 2006). In each plot, trees with diameter at 1.30 m above ground level (dbh) greater than 10 cm were measured and their species were identified. The number of trees by plot ranged from 314 to 623. The two-parameter Weibull distribution function (Equation 1) was applied to each inventory plot data to estimate its shape and scale.

$$f(D) = \frac{\alpha}{\beta}\left(\frac{D}{\beta}\right)^{\alpha-1} e^{-\left(\frac{D}{\beta}\right)^{\alpha}} \qquad (1)$$

The shape values of the plots ranged from 1.35 to 1.76 (Table 1), indicating that their diameter distributions were left-skewed and consistent with the structure of natural forests. The maximum diameter was 188 cm with an average of 24.8 cm and a median of 17.8 cm. The number of large trees per plot ranged from 5 to 36 for diameters greater than 70 cm, from 1 to 14 for diameters greater than 90 cm, and from 0 to 11 for diameters greater than 100 cm. Tree species richness ranged from 66 to 119, confirming that these plots were located in the ecoregion of tropical moist forest.

Destructive biomass data with 283 sample trees (dbh≥10 cm) from published works (Djomo et al., 2010; Henry et al., 2010; Ebuyi, et al., 2011; Fayolle, et al., 2013; Ngomanda et al., 2014) were compiled. These data were drawn from the dense evergreen forest and the transition forest between the dense evergreen forest and semi-deciduous forest. The mean diameter was of 44.9 cm, with a median of 37.6 cm and 25% of trees had a diameter greater than or equal to 70 cm. The number of trees with diameter greater than 70 cm and 90 cm was respectively 93 and 47. The shape value was equal to 1.79 and the greatest diameter was 192.5 cm. The total number of species was 75.

For tree species where total height and wood specific gravity data were not available in both the destructive and inventory datasets, total height was estimated using the height allometric equation from Djomo et al. (2016)

$$H = Exp(1.190 + 0.406 \times \ln(D) + 0.036 \times (\ln(D))^2$$

Wood specific gravity data were sourced from the international wood density table (Zanne et al., 2009). For tree species absent from this table, the average wood density of the plot was assigned. The two allometric equations used for parametric resampling were established using the biomass data and are presented as follow

$$\ln(AGB_i) = -1.272 + 2.421\ln(D_i) + \ln(\rho_i)$$

with residual standard error (RSE) of 0.305 and

$$\ln(AGB_i) = -1.581 + 2.273 \times \ln(D_i) + 0.259 \times \ln(H) + \ln(\rho_i)$$

with RSE=0.299 (Fonton et al., 2017).

Components of uncertainty in biomass allometric model

The overall uncertainty of a given allometric model used to estimate biomass is determined by three key components as presented by Roxburgh et al. (2015). The first is the inherent variability of biomass on the natural scale, which increases with diameter (Figure 1a) and remains constant on the logarithmic scale (Xiao et al., 2011). This variability is reflected in the standard deviation of residuals around the line of best fit (Figure 1 b).

The second source of variability was the number of sampled trees used to establish the allometric equation, known as sampling uncertainty. The accuracy of biomass prediction improves with the number of trees used in the equation development (Roxburgh et al., 2015). The third source of uncertainty relates to the homogeneity of sampled trees to the broader population to which the model is applied. As pointed out by Snowdon et al. (2002) and Picard et al. (2012), the reduction of the precision occurs when the size-class distribution of allometric models differs from that of the population, mainly in term of diameter size-class distribution. This study addresses these three sources of uncertainty by implementing controlled sampling strategies.

Sampling strategies

Ten sampling strategies were used and characterized by the combination of two allometric models and five sampling techniques. The choice of this methodological approach was to control the three components of uncertainty: the variability of biomass with increasing diameter, the number of sampled trees (sampling uncertainty) and their representativeness according to the forest stand structure.

The two allometric models were natural logarithm transformed to reduce the inherent variability at natural scale. They differed with the number of predictor variables. The first was with two predictor variables called Model 1 (Equation 2),

$$\ln(AGB) = \ln(\beta_0) + \beta_1 \ln(D) + \beta_3 \ln(\rho) \quad (2)$$

and the second with three predictor variables called, Model 2 (Equation 3),

$$\ln(AGB) = \ln(\beta_0) + \beta_1 \ln(D) + \beta_2 \ln(H) + \beta_3 \ln(\rho) \quad (3)$$

where (D) was the diameter at breast height, (H) the tree height and (ρ) the wood specific gravity. According to Fonton et al. (2017), these two models were the best one for two and three predictor variables with available destructive *AGB* published data with 283 sampled trees ($D \geq 10$ cm) in Congo basin. Many studies have highlighted the importance of tree height as third predictor variable in the *AGB* equation (Chave et al., 2014; Djomo et al., 2016; Fonton, et al., 2017).

To control the second and third components of uncertainty, the algorithm of the sampling techniques was guided by a wide range of number of sampled trees and their representativeness according to the diameter size-class distribution of the landscape to which the allometric model was applied. The first sampling technique was the random sample on natural scale (S1): each tree had the same probability to be selected. The number of the sampled trees was ranged from 10 to 300 by step of 5 and the sampled trees were selected with a uniform distribution in the range D_{min} (the minimum diameter) to D_{max} (the maximum). Based on the relatively smaller number of large trees in natural forests (Chave et al. 2005; Roxburgh et al. 2015), three other sampling techniques were considered, the S2, S3 and S4. For S2, random samples were selected from cumulative frequency distribution of diameter and split into diameter size-classes. Size-class boundaries were based on the cumulative sums of diameter listed in order of increasing tree diameter. The cumulative sum of the largest tree was divided into eight that corresponded to the upper limit of first size-class. The upper limit of the remaining size-class was successively determined by multiplying this value by 2 to 7. The number of eight diameter size-classes was reasonably in accordance with Snowdon et al. (2002) and Dietz and Kuyah (2011). From each diameter size-class thus delimited, equal numbers of sampled trees, ranging from 2 to 30 were randomly selected. The other sampling techniques (S3 and S4) were similar to S2 concerning the eight diameter size-class boundaries, but the cumulative sums were made with the diameter squared (S3) and with the logarithm of diameter (S4). The later sampling technique (S5) was a stratified random sample on the natural scale where the range from D_{min} to D_{max} was divided into eight diameter size-classes distribution. The first 6 size-classes were from $D = 10\,\text{cm}$ to $D < 70\,\text{cm}$ with equal interval of 10 cm, while the two others were $70\,\text{cm} \leq D < 100\,\text{cm}$ and $D \geq 100\,\text{cm}$ respectively. The numbers of sampled trees ranging from 2 to 30, were randomly selected by size-class.

Parametric re-sampling simulation

The simulation design used in this study was the one proposed by Roxburgh et al. (2015) which differs from bootstrapped resampling procedure used by Chave et al. (2004) and van Breughel et al. (2011) that led to biased biomass prediction due to insufficient observed data. The simulation data were generated across a wide range of tree numbers to ensure their representativeness around the fitted line using parametric resampling. This parametric resampling was applied to ensure unbiased sampling across the entire diameter size range and representative spread of observations around the fitted line. In the theoretical framework of linear regression, the dependent variable was considered as random and the observed values of the response variable were denoted as a realization of a random variable. In equations 2 and 3, the variables $ln(D)$, $ln(H)$, and $ln(\rho)$ were nonrandom but that was not the case of the response variable $\ln(AGB)$ considered as random variable with normal distribution, so that:

$$\ln(AGB) \overset{iid}{\approx} N(\hat{Y}, \sigma^2) \quad (4)$$

where $\hat{Y}$ = Equation 1 or 2 and σ^2 the residual variance.

For each tree *i*, the simulated value of response variable was computed as:

$$\ln(AGB_i) = \hat{Y} + \varepsilon_i = \ln(\beta_0) + \beta_1 \ln(D_i) + \beta_3 \ln(\rho_i) + \varepsilon_i \quad \text{or} \quad (5)$$

$$= \ln(\beta_0) + \beta_1 \ln(D_i) + \beta_2 \ln(H) + \beta_3 \ln(\rho_i) + \varepsilon_i \quad (6)$$

where $\varepsilon_i \overset{iid}{\approx} N(0, \sigma^2)$.

For each number of sample trees n_s, a vector of normal random values ε_i of size n_s were generated and the equation 5 or 6 was applied to compute a vector of $\ln(\text{AGB})$. A new allometric model was established and used for biomass prediction at the plot level. For each sampling strategy and each sample tree numbers n_s, the coefficient of variation of the predicted biomass (CV_{n_s}) was computed with 1000 randomly constructed models as:

$$CV_{n_s} = 100 \times \hat{\sigma}_{AGB} / \bar{x}_{AGB} \quad (7)$$

where $\hat{\sigma}_{AGB}$ and $\bar{x}_{AGB}$ were the standard deviation and the mean of the predicted AGB according to each size n_s.

Variability of large trees

The accuracy of allometric equation without large trees was evaluated with truncated sample dataset. The truncated sample dataset was the sample dataset without large trees. According to Lung and Espera (2015), 44% of *AGB* and 25% of total *AGB* by plot were explained by large trees with dbh >50 cm while for Silk et al. (2010, 2013), large trees corresponded to diameter more than 70 cm. Thus, four truncated sample datasets were considered: those containing trees with diameter lower than 50 cm, 70 cm, 90 cm and 100 cm. The coefficient of variation of the predicted biomass was computed with 1000 randomly constructed models for each sampling strategy and each sample size (except for sampling technique S1 and S5).

Statistical analysis

The required number of sampled trees to achieve a given level of precision was estimated. Then for each sampling strategy, an empirical power function was established to determine the relationship between CV_{n_s} and the sample size n_s (Equation 9) as:

$$CV_{n_s} = a \times n_s^b \quad (9)$$

where a was the intercept and b the slope. To estimate these two unknown parameters Equation 9 was log-transformed (Equation 10).

$$\ln(CV_{n_s}) = \ln(a) + b \times \ln(n_s) \quad (10)$$

For a better appreciation of the relationship between CV_{n_s} and n_s, the quality of this adjustment was evaluated with two statistical parameters, the residual standard error (RSE) and the adjusted coefficient of determination (R^2aj). The analysis of variance was carried out to compare the ten sampling strategies mainly the interactions with forest type. The Snedecor-Newton-Keuls comparison of mean test (SNK) was used to set up the homogeneity group of sampling strategies. For each sampling strategies, the sample size n_α required according to each precision level CV_α of 2.5%, 5%, 7.5% or 10% was computed as (Equation 12):

$$n_\alpha = \exp\left[\frac{\ln(CV_\alpha) - \ln(a)}{b}\right] \quad (11)$$

The number of sampled trees were calculated for CV_α equal to 2.5%, 5%, 7.5% and 10% corresponding respectively to 97.5%, 95%, 92.5% and 90% confidence level with the corresponding error on biomass estimation of 5%, 10%, 15% and 20% respectively.

The importance of including large trees in the sample for the establishment of AGB equation was highlighted comparing the coefficients of variation of predicted biomass of truncated data set ((CV_{tns}) to those of the data set (CV_{ns}) with the *t* student paired wise test. The precision gap of including largest trees ($P_{gap}\%$) was also quantified as the percentage of the difference between CV_{tns} and CV_{ns} so that:

$$P_{gap}\% = 100 \times (CV_{tns} - CV_{ns}) / CV_{ns} \qquad (12)$$

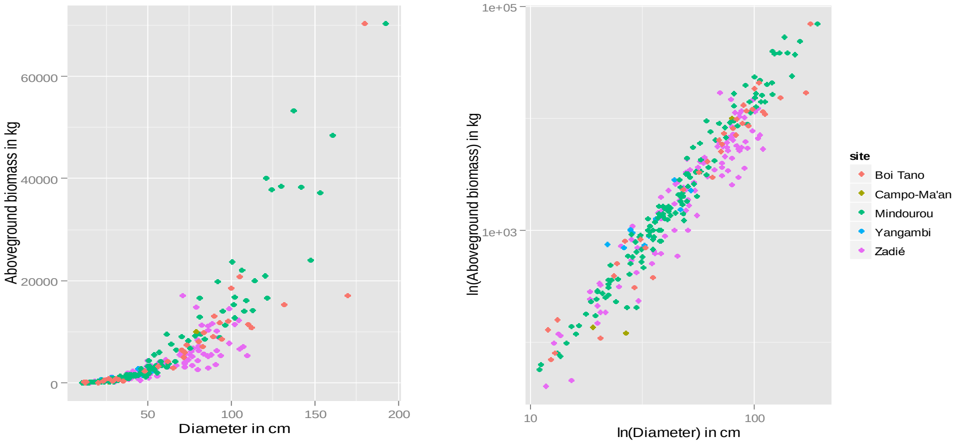
If $P_{gap}\%$ is positive, the allometric equation established with sampling from dataset is more accurate than sampling from truncated dataset (without large trees).

The statistical software R, version 3.6.0 (R core team, 2019) was used for data processing and statistical analysis.

Table 1: Dendrometrical characteristics of the inventory plots and biomass dataset.

Forest reserve	Plot	Density	N_Esp	BA	dbh≥70 cm	dbh≥90 cm	dbh≥100 cm	Shape	Scale
Dzanga Sangha	1	424	107	23,96	9	3	3	1,59	24,55
Dzanga Sangha	2	435	119	29,64	11	5	4	1,54	26,60
Dzanga Sangha	3	314	66	29,65	16	14	11	1,35	28,43
Dzanga Sangha	4	429	96	22,42	8	1	0	1,76	24,75
Dzanga Sangha	5	489	104	35,31	16	9	7	1,55	27,34
Monts de Cristal	1	546	87	28,84	5	2	1	1,72	24,64
Monts de Cristal	2	587	87	44,91	20	10	6	1,56	28,30
Monts de Cristal	3	557	96	40,01	21	10	5	1,51	26,91
Monts de Cristal	4	547	86	39,31	23	13	9	1,49	26,57
Monts de Cristal	5	569	103	44,52	36	12	9	1,41	26,72
Waka	1	393	83	33,35	22	4	2	1,54	29,83
Waka	2	422	100	32,34	17	4	4	1,59	28,59
Waka	3	501	109	39,70	14	2	1	1,67	29,99
Waka	4	554	105	47,72	24	13	7	1,51	29,67
Waka	5	623	106	45,15	28	12	5	1,52	27,18
Biomass data		283	75		93	47	32	1.79	63.74

N_Esp= Tree species richness; *BA*=basal area (m^2ha^{-1});



(a) (b)

Figure 1: Site distribution of above ground biomass vs. diameter at breast height (a) and the natural logarithm transformation distribution (b).

RESULTS

Analyzing parameters of the power function

The adjustment quality of the power function relating the coefficient of variation of predicted biomass (CV_{n_S}) to the sample size (n_S) was analyzed using residual standard error and coefficient of determination values (Table 2). Independently of the allometric models, R^2aj and RSE were 0.99 and less than 0.03 respectively, except for sampling technique S1. The statistical significance of the estimated intercept and slope of the power function relationship was then confirmed, allowing for comparison of sampling strategies based on these parameters. The analysis of variance considering all the sampling factors (sampling techniques, model types and forest reserves) showed significant interactions among them. Subsequent analyses were performed for each model type and, if necessary, by forest reserve.

For allometric Model 1, there was a significant difference among sampling techniques, with high intercept parameter values of 4.44 and 4.06 for S1 and S4 respectively, while their slope parameters were the lowest (-0.56 and -0.51, Table 3). For allometric Model 2, a significant interaction occurred between sampling techniques and forest reserves.

The forest reserve MC was characterized by the highest value of intercept and the lowest value for slope, followed by DZ. With multiple comparison of mean, S2, S3 and S5 form a homogeneous group that was different from S1 and S4 for the intercept while for the slope, S1 was different from the four others which are homogeneous group (Table 3). Taking into account the results, it appeared that there was a relation between the parameters of the plot stand, the power function parameters and the sample size.

Number of sampled trees to achieve levels of precision based on the result

To illustrate the underlying power function relationship between number of sampled trees and precision, differences among sampling strategies were evident with the computation of the number of trees to achieve different levels of precision. In Table 4, the number of trees (and their ranges) were presented by allometric model for each sampling technique. Sample sizes were smallest for S3 and S5, and largest for S2, followed by S4 and S1. At a precision level of CV=5% for Model 1, significant differences in sample size were observed among sampling techniques, with S1 requiring the highest sample size (167) and S3 (42) and S5 (44) the lowest. Similar trends were observed for Model 2, where S1 required the largest sample size (204), and S3 (43) and S5 (47) the smallest.

The sample sizes computed in Table 4 were subject to uncertainties, evaluated with the coefficient of variation of 10% to 25% for Model 1 and 14% to 42% for Model 2 at a precision level of 5%. These uncertainties could be attributed to stand parameters of inventory plots. This variability was analyzed to determine the appropriate number of sampled trees relative to stand parameters. Across each sampling strategy, the intercepts on logarithm scale between the biomass precision CV% and the number of sampled trees varied, as illustrated in Figure 2.

The variation between intercept of the power function (Equation 9) was explained by the shape of the Weibull distribution of inventory plots data (Figure 2.a). This variation was computed with the second-order polynomial relationship between the intercept values y and the shape x of the Weibull distribution applied to the plots inventory data for each sampling strategy as

$$y = z_0 + z_1 \times x + z_3 \times x^2 \qquad 13$$

with z_i the regression coefficients. The estimated regression of Equation 13, showed that the intercept was inversely proportional to the shape. This trend reached a saturation with S3 and S5 at shape value higher than 1.7 (Figure 2.b). A correction of the intercept was used to generalize equation 11 (Equation 14) with CV_α the biomass precision level and n_α the adjusted number of sampled trees.

$$n_\alpha = \left(\frac{\exp(z_0 + z_1 \times x + z_3 \times x^2)}{CV_\alpha} \right)^{\frac{1}{|b|}} \qquad 14$$

The adjusted sample sizes as function of the shape values ranging from 1.30 to 1.80 were computed by step of 0.05 (Table 5).

Prediction gap with dataset of large trees

The comparison of the observed coefficient of variation of predicted biomass across the 15 inventory plots using the pairwise t-test revealed significant differences, except for the truncated diameter dataset. The magnitude of this difference varied with sampling strategies, with no significant interactions observed among forest types. As presented in Table 6, the gap $(P_{gap}\%)$ ranged from 8.0 to 10.4 and 23.9 to 31.8 respectively for Model 1 and Model 2 considering trees with dbh<100 cm. Sampling technique S4 was characterized by greater precision regardless the Dmax.

Table 2: Summary ranges of the residual standard errors (RSE) and coefficient of determination (R^2aj) of the relationship between CV_{n_s} and n_s for the ten sampling strategies.

Strategy technique	Model 1		Model 2	
	RSE	R^2aj	RSE	R^2aj
S1	**(0,032 0,091)**	**(0,964 0,995)**	**(0,534 0,807)**	**(0,674 0,797)**
S2	(0,019 0,031)	(0,994 0,998)	(0,021 0,027)	(0,995 0,997)
S3	(0,018 0,027)	(0,995 0,998)	(0,017 0,025)	(0,996 0,998)
S4	(0,019 0,028)	(0,995 0,998)	(0,022 **0,064**)	(**0,982** 0,997)
S5	(0,019 0,026)	(0,995 0,998)	(0,021 0,028)	(0,995 0,997)

Model 1=allometric model with D and ρ , Model 2= allometric model with D, H and ρ , S1=random sample on natural scale;S2= random sample with cumulative frequency distribution of diameter, S3= the random sample with cumulative frequency distribution on diameter squared, S4= random sample with cumulative frequency distribution on logarithm scale and S5= stratified random sample with diameter size-classes distribution on natural scale; S2 to S5 are split into eight diameter size-classes distribution.

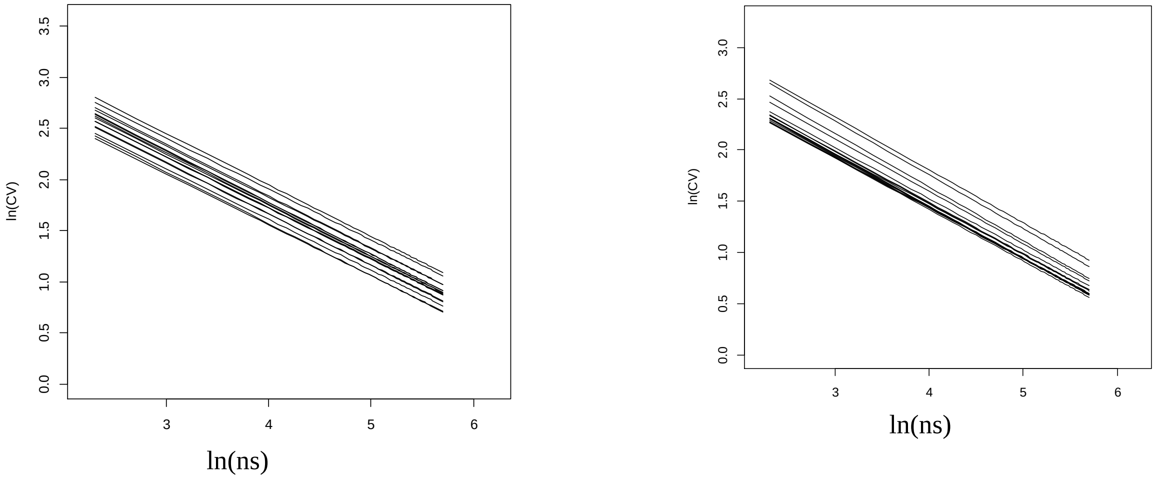
Table 3: Sampling technique comparison of mean by Student-Newmn-Keuls method for Intercept and slope parameters of the power function relationship between coefficient of variation of predicted biomass (CV_{n_s}) and the sample size used to establish the allometric models (n_s); in bracket the standard error of mean.

Sampling techniques	Model 1	Model 2	DZ	MC	WA
	Intercept				
S1	4.437 (0.137) **a**	8.578 (0.917) **a**	8.608 (0.830) **a**	9.335 (0.766) **a**	7.789 (0.418) **a**
S2	3.747 (0.120) **c**	3.906 (0.246) **c**	3.963 (0.299) **c**	3.976 (0.270) **c**	3.780 (0.138) **c**
S3	3.475 (0.060) **d**	3.495 (0.066) **d**	3.485 (0.088) **c**	3.529 (0.071) **c**	3.472 (0.019) **d**
S4	4.063 (0.128) **b**	4.610 (0.407) **b**	4.714 (0.450) **b**	4.810 (0.391) **b**	4.305 (0.206) **b**
S5	3.500 (0.067) **d**	3.545 (0.158) **d**	3.548 (0.183) **c**	3.625 (0.186) **c**	3.461 (0.053) **d**
	Slope				
S1	-0.555 (0.012) **c**	-1.313 (0.146) **c**	-1.312 (0.115) **b**	-1.437 (0.128) **b**	-1.188 (0.076) **b**
S2	-0.502 (0.004) **a**	-0.519 (0.013) **a**	-0.527 (0.014) **a**	-0.519 (0.013) **a**	-0.511 (0.008) **a**
S3	-0.501 (0.006) **a**	-0.503 (0.005) **a**	-0.499 (0.004) **a**	-0.507 (0.005) **a**	-0.503 (0.003) **a**
S4	-0.508 (0.004) **b**	-0.576 (0.034) **b**	-0.590 (0.032) **a**	-0.596 (0.026) **a**	-0.542 (0.013) **a**
S5	-0.502 (0.004) **a**	-0.505 (0.011) **a**	-0.506 (0.008) **a**	-0.512 (0.014) **a**	-0.498 (0.006) **a**

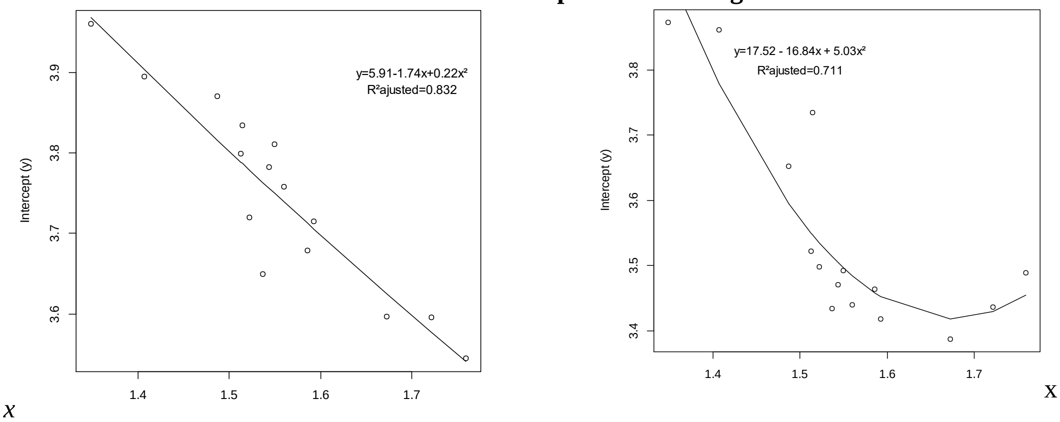
M1= Allometric model with two predictor variables M2= Allometric model with two predictor variables; DZ= Dzanga Sangha forest, MC=Monts de Cristal forest, WA= Waka forest.

Table 4: Number of sample trees to achieve biomass prediction with a given precision (CV_{ns}) by sampling techniques and by allometric model with data from 15 1-ha plots data distributed across 3 forest reserves in the Congo Basin.

Model type	Sampling technique	Predicted AGB CV 2.5%	5%	7.5%	10%
Model 1	S1	583 (139)	**167 (39)**	81 (19)	48 (11)
	S2	287 (65)	**72 (17)**	32 (7)	18 (4)
	S3	165 (16)	**42 (4)**	18 (2)	10 (1)
	S4	504 (124)	**129 (32)**	58 (15)	33 (8)
	S5	173 (24)	**44 (6)**	20 (3)	11 (1)
Model 2	S1	346 (57)	**204 (38)**	149 (31)	120 (27)
	S2	333 (125)	**88 (36)**	41 (17)	24 (10)
	S3	170 (24)	**43 (6)**	19 (3)	11 (2)
	S4	633 (230)	**194 (82)**	97 (45)	59 (29)
	S5	185 (46)	**47 (13)**	21 (6)	12 (4)



A. Precision vs. number of sampled trees on logarithm scale



B. Intercept vs. shape and the adjusted model

Figure 2: Relationships between predicted biomass precision and the sample size on logarithm scale for S2 - model 1 and S5 - model 2 applied to inventory data and the variation of the Intercept explained by the shape of 15 plots.

Table 5: Number of sampled trees adjusted to stand shape for precision of CV=5% by model type and sampling technique.

Model type	Sampling technique	Stands shape value									
		1.30	**1.35**	**1.40**	**1.50**	**1.55**	**1.60**	**1.65**	**1.70**	**1.75**	**1.80**
Model 1	S1	271	240	215	176	162	149	139	131	124	118
	S2	123	110	98	79	71	64	58	53	48	43
	S3	60	54	49	42	40	39	38	38	38	39
	S4	248	210	180	138	123	111	102	94	89	84
	S5	65	58	53	45	42	40	39	38	38	38
Model 2	S1	715	505	371	226	188	162	146	136	133	135
	S2	331	234	170	99	80	66	56	49	45	42
	S3	70	60	52	43	41	39	38	39	40	42
	S4	1102	716	480	237	175	133	105	85	71	62
	S5	145	103	76	49	42	38	36	36	38	42

Table 6: Gap percentages ($P_{gap}\%$) of predicted biomass with diameter lower than 50, 70, 90 and 100 cm by sampling technique and allometric model.

Sampling technique	**Model 1**				**Model 2**			
	50	**70**	**90**	**100**	**50**	**70**	**90**	**100**
S2	67,3	28,8	13,3	***9,4***	345,0	112,9	47,7	***31,8***
S3	88,7	34,3	15,0	***10,4***	425,2	126,1	45,5	***27,8***
S4	48,0	21,2	11,3	***8,0***	246,6	80,9	35,7	***23,9***

S2= random sample with cumulative frequency of size-class diameter on natural scale, S3= the random with cumulative frequency of size-class diameter squared, S4= random sample with cumulative frequency of size-class diameter on logarithm scale, all were split into eight diameter size-classes distribution.

DISCUSSION

The sample size (n) for modeling biological phenomenon has retained more attention in biometrics with an increasing precision as a function of $\sqrt{n}$ (Picard et al., 2012). Despite the development of biomass allometric equation for prediction use, it is surprising that little attention has been given to the optimum sample size versus the performance of biomass prediction, except for Roxburgh et al. (2015) in planted forests. This situation can be explained by the challenges associated with the cost and time-consuming nature of harvest data collection, and the variability in tree numbers and diameter classes that result from it. Nonetheless, it remains crucial to estimate the precision achievable with the resources invested. Some guides have been provided by Chave et al. (2004) and van Breughel et al. (2011) with investigation of error propagation. Therefore, many sources of uncertainty have been revealed (Ketterings al., 2001; Chave et al., 2004; Molto et al., 2012) and the most important one is the error due to the choice of allometric equations or model misspecification (Chave et al., 2004; van Breugel et al., 2011; Melson et al., 2011; Molto et al., 2012; Picard et al., 2016). This major source of variation of the predicted biomass is correlated with the number of trees used to calibrate allometric models (Chave et al. 2004) where fewer trees lead to increased uncertainty in aboveground biomass estimates at the landscape level (van Breugel et al., 2011). The central concern of this study was to determine the optimal number of sampled trees required to establish accurate aboveground biomass allometric equations.

The present results revealed that the uncertainty in predicted biomass was function of the characteristics of the population to which the model is applied, the allometric model type and the representativeness of large trees relative to the sample size. The random sample on natural scale (S1) is characterized by the highest sample sizes compared to other techniques. In contrast, the random sample on the square of diameter with an eight size-class distribution (S3) yields the lowest sample sizes. While for S1, all the trees had the same probability of being part of the sample, S3 was characterized by one equal number of trees per diameter size-class and privileged diameter classes of large trees based on the cumulative sum of the squared diameters. This privilege decreased with S2 where the algorithm were based on the cumulative sum of the diameters and even more so with S4, which used the logarithm of diameters. This difference between the sampling techniques is attributed to the size-class distribution of trees used to establish the allometric model, which aligns with the population characteristics, as advocated by Fu et al. (2017). This size-class distribution is not solely the unique source of uncertainty in predicted biomass. The uncertainty on the sample sizes for 5% level of precision (Table 4) ranged from 4 to 39 and 6 to 82 respectively for Model 1 and Model 2. This source of error was analyzed with the shape of the inventory data. The intercept was a function of the shape in second-polynomial relationship and a new relationship between the sample size and the power function was adjusted. This confirmed the approach of Dietz and Kuyah (2011) that stipulated selecting trees to span more evenly the potential diameter range to which the model is expected to be applied.

Integrating height as a third predictor variable improved the quality of predicted biomass. The study showed that, with this model type, more sample size data were required to achieve the same level of precision. Uncertainty occurred with the number of predictor variables. This source of uncertainty can be explained by the model's coefficient of variation as a component of prediction error which was inversely proportional to the number of observations used for model fitting. The simulated study of Knofczynski and Mundfrom, (2008) demonstrated that the sample size increased with the number of predictor variables. Thus, the selected sample size must take into account the number of predictor variables in the model.

Overall, precise AGB estimates are presented for the standard precision level of 5%. For all sampling techniques, the number of trees required ranged from 176 to 42 and 237 to 43 respectively for Model 1 and Model 2 with a stand shape value of 1.50. This result is similar to the minimum sample trees recommended by Chave et al. (2004). However, care must be taken regarding the diameter structure of the forest to which the allometric model is applied. While the slopes of the relation between predicted biomass precision and the number of sample trees on a logarithm scale appear identical, the intercepts were correlated with the shape of the diameter structure of the tree population.

With allometric Model 1, the sampling technique with the fewest sample trees ranging from 60 to 40 were S3; it was followed by S5 (65 to 38) and S2 (123 to 43). For Model 2, S3 with 70 to 42 was characterized by lowest number of sample trees, followed by S5 (145 to 42) and S2 (331 to 42). For these three sampling techniques, uncertainty due to differences in size class distribution was minimized according to the algorithm. S3 and S5 can be reasonably considered the best-performing sampling techniques, while S4 and S1 were the poorest and should be avoided.

Conclusion

The common tool for estimating biomass and forest carbon stock is the allometric equation despite modern approaches. Ecosystem-specific allometric equations have been identified by many studies as more accurate than regional or pantropical equation.

For the biomass allometric equation establishment, sampled trees were harvested. Taking into account the components of uncertainty in biomass equation estimation, the strategies for tree sampling were based on the inherent variability of biomass with increasing diameter, the number of sampled trees and the representativeness of the sampled trees according to the forest.

The results of this study provide necessary information for management and scientific decisions on sampling strategy and sample trees for the establishment of forest biomass allometric equations. While many questions are asked about the number and diametric structure of sampled trees for data collection, this study fills that gap. It shows that the accuracy of carbon stock estimation in forest reserves should rely on these findings to reduce uncertainty and to get more reliable results. Therefore, the definition of the sampling techniques and number of sample trees must consider the stand shape value and level of precision. Hence, the relationship between the intercept of the power function (predicted aboveground biomass coefficient of variation versus sample size) and the shape of stand was investigated. The accuracy of predicted biomass taking into account the large trees in sampled biomass data remained very important based on the results obtained.

COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHORS' CONTRIBUTIONS

The authors confirm contribution to the paper as follows: NHF conceived the idea, compiled the data. GA and FGL analyzed and interprated the results with NHF. AHA, FGL and TOFD wrote the manuscript with editorial support and reviewed from NHFand BK. All authors have read and agreed to the published the manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank the PREREED of COMIFAC for the assistance in accessing the data from previous biomass data collection.

REFERENCES

Balinga M. 2006. A vegetation assessment of Waka national park, Gabon. Central African Regional, Program for the Environment, Smithsonian Institution, 124 p.

Balinga M, Moses S, Fombod E. 2006. A preliminary assessment of the vegetation of the Dzanga Sangha protected area complex, Central African Republic. Central African Regional, Program for the Environment, Smithsonian Institution, 152 p.

Chave J, Réjou-Méchain M, Búrquez A, Chidumayo E, Colgan MS, Delitti WBC, Duque A, Eid T, Fearnside PM, Goodman RC, Henry M, Martínez-YrízarA, Mugasha WA, Muller-Landau HC, Mencuccini M, Nelson BW, Ngomanda A, Nogueira EL, Ortiz-Malavassi EO, 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**: 3177–3190. DOI: https://doi.org/10.1111/gcb.12629

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: https://doi.org/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: https://doi.org/10.1007/s00442-005-0100-x

Dembele B, Gomgnimbou AP, Yameogo JT, Ouedraogo OW, Hien M. (2023). Détermination du carbone stocké par les ligneux dans la bande de servitude de la rivière Kou dans la Province du Houet (Burkina Faso). *International Journal of Biological and Chemical Sciences*, **17**(1): 281-292. DOI : https://doi.org/10.4314/ijbcs.v17i1.20

Dietz J, Kuyah S. 2011. *Guidelines for Establishing Regional Allometric Equations for Biomass Estimation through Destructive Sampling*. World Agroforestry Centre (ICRAF): Nairobi.

Djomo AN, Picard N, Fayolle A, Henry M, Ngomanda A, Ploton P, McLellan J, Saborowski J, Adamou I, Lejeune P. 2016. Tree allometry for estimation of carbon stocks in African tropical forests. *Forestry*, **89**(4): 446-455. DOI: https://doi.org/10.1093/forestry/cpw025

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**: 1873–1885. DOI: https://doi.org/10.1016/j.foreco.2010.08.034

Ebuy J, Lokombe DJ, Ponette Q, Sonwa D, Picard N. 2011. Allometric equation for predicting aboveground biomass of three tree species. *Journal of Tropical Forest Science,* **23**(2): 125–132. DOI: http://www.jstor.org/stable/23616912

Fayolle A, Doucet JL, Gillet J-F, Bourland N, Lejeune P. 2013. Tree allometry in Central Africa: Testing the validity of pantropical multi-species allometric equations for estimating biomass and carbon stocks. *Forest Ecology Management,* **305**: 29–37. DOI: https://doi.org/10.1016/j.foreco.2013.05.036

Fayolle A, Ngomanda A, Mbasi M, Barbier N, Bocko Y, Boyemba F, Medjibe V. 2018. A regional allometry for the Congo basin forests based on the largest ever destructive sampling. *Forest Ecology and Management,* **430**: 228-240. DOI: https://doi.org/10.1016/j.foreco.2018.07.030

Fonton HN, Medjibé V, Djomo A, Kondaoulé J, Rossi V, Ngomanda A, Maïdou H. 2017. Analyzing accuracy of the power functions for modeling aboveground biomass prediction in Congo Basin tropical forests. *Open Journal of Forestry,* **7**: 388-402. DOI: https://doi.org/10.4236/ojf.2017.74023

Fu Y, Lei Y, Zeng W, Hao R, Zhang G, Zhong Q, Xu M. 2017. Uncertainty assessment in aboveground biomass estimation at the regional scale using a new method considering both sampling error and model error. *Canadian Journal of Forest Research*, **47**(8): 1095-1103. DOI: https://doi.org/10.1139/cjfr-2016-0436

Ganamé M, Bayen P, Ouédraogo I, Balima LH, Thiombiano A. 2021. Allometric models for improving aboveground biomass estimates in West African savanna ecosystems. *Trees, Forests and People,* **4**: 100077. DOI: https://doi.org/10.1016/j.tfp.2021.100077

Henry M, Besnar A, Asante WA, Eshun J, Adu-Bredu S, Valentin R, Bernoux M, Saint-Andre L. 2010. Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of Africa. *Forest Ecology and Management*, **260**: 1375–1388. DOI: https://doi.org/10.1016/j.foreco.2010.07.040

Jalkanen A, Mäkipää R, Ståhl G, Lehtonen A, Petersson H. 2005. Estimation of the biomass stock of trees in Sweden. *Annals of Forest Sciences,* **62**: 845–851. DOI: https://doi.org/10.1051/forest:2005075

Knofczynski TG, Mundfrom D. 2008. Sample sizes when using multiple linear regression for prediction. *Educational and Psychological Measurement,* **68**: 431-442. DOI: https://doi.org/10.1177/0013164407310131

Lewis SL, Lopez-Gonzalez G, Sonké B, Affum-Baffoe K, Baker TR, Ojo LO, Wöll H. 2009. Increasing carbon storage in intact African tropical forests. *Nature,* **457**(7232): 1003-1006. DOI: https://doi.org/10.1038/nature07771

Lung M, Espira A. 2015: The influence of stand variables and human use on biomass and carbon stocks of a transitional African forest: Implications for forest carbon projects. *Forest Ecology and Management*, **351**: 36–46. DOI: https://doi.org/10.1016/j.foreco.2015.04.032

Manzo OL, Moussa M, Issoufou HBA, Abdoulaye D, Morou B, Youssifi S, Mahamane A, Paul R. 2015. Equations allométriques pour l'estimation de la biomasse aérienne de Faidherbia albida (Del.) Achev dans les agrosystèmes d'Aguié, Niger. *International Journal of Biological and Chemical Sciences*, **9**(4): 1863-1874. DOI: https://doi.org/10.4314/ijbcs.v9i4.12

Ngomanda A, Engone Obiang NL, Lebamba J, Moundounga Mavouroulou Q, Gomat H, Mankou GS, Loumeto J, Midoko Iponga D, Kossi Ditsouga F, Zinga Koumba R. 2014. Site-specific *versus* pantropical allometric equations: Which option to estimate the biomass of a moist central African forest? *Forest Ecology and Management*, **312**: 1–9. DOI: https://doi.org/10.1016/j.foreco.2013.10.029

Picard N, Henry M, Fonton HN, Kondaoulé J, Fayolle A, Birigazzi L, Sola G, Poultouchidou A, Trotta C, Maïdou H. 2016. Error in the estimation of emission factors for forest degradation in central Africa. *Journal of Forestry Research*, **21**(1): 23-30. DOI: https://doi.org/10.1007/s10310-015-0510-5

Picard N, Laurent S-A, Henry M. 2012. *Manual for Building Tree Volume and Biomass Allometric Equations: from field Measurement to Prediction*. Food and Agricultural Organization of the United Nations, Rome and Centre de Coopération Internationale en recherche Agronomique pour le Développement, Montpellier, France.

Romero FMB, Jacovine LAG, Torres CMME, Ribeiro SC, da Rocha SJSS, Novais TDNO, Fearnside PM. 2022. Aboveground biomass allometric models for large trees in southwestern Amazonia. *Trees, Forests and People,* **9**: 100317. DOI: https://doi.org/10.1016/j.tfp.2022.100317

Roxburgh SHKI, Paul D, Clifford JR, England R, Raison RJ. 2015. Guidelines for constructing allometric models for the prediction of woody biomass: how many individuals to harvest? *Ecosphere*, **6**(3): 38. DOI: http://dx.doi.org/10.1890/ES14-00251.1

Rutishauser E, Noor'an F, Laumonier Y, Halperin J, Rufi'ie, Hergoualc'h K, Verchot L. 2013. Generic allometric models including height best estimate forest biomass and carbon stocks in Indonesia. *Forest Ecology and Management*, **305**: 219–225. DOI: https://doi.org/10.1016/j.foreco.2013.07.013

Saatchi SS, Harris NL, Brown S, Lefsky M, Mitchard ETA, Salas W, Zutta BR., Buermann W, Lewis SL, Hagen S, Petrova S, White L, Silman M, Morel A. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *PNAS*, **108**(24): 9899-9904. DOI: https://doi.org/10.1073/pnas.1019576108

Slik JWF, Aiba S, Brearley FQ, Cannon CH, Forshed O, Kitayama K, Nagamasu H, Nilus R, Payne J, Paoli G, Poulsen AD, Raes N, Sheil D, Sidiyasa K, Suzuki E, van Valkenburg JLCH. 2010. Environmental correlates of tree biomass, basal area, wood specific gravity and stem density gradients in Borneo's tropical forests. *Global Ecology and Biogeography*, **19**: 50-60. DOI: https://doi.org/10.1111/j.1466-8238.2009.00489.x

Slik JW, Paoli G, McGuire K, Amaral I, Barroso J, Bastian M, Blanc L, Bongers F, Boundja P, Clark C, Collins M, Dauby G, Ding Y, Doucet J-L, Eler E, Ferreira L, Forshed O, Fredriksson G, Gillet J-F, Harris D, Leal M, Laumonier Y, Malhi Y, Mansor A, Martin E, Miyamoto K, Araujo-Murakami A, Nagamasu H, Nilus R, Nurtjahya E, Oliveira A, Onrizal O, Parada-Gutierrez A, Permana A, Poorter L, Poulsen J, Ramirez-Angulo H, Reitsma J, Rovero F, Rozak A, Sheil D, Silva-

Espejo J, Silveira M, Spironelo W, ter Steege H, Stevart T, Navarro-Aguilar GE, Sunderland T, Suzuki E, Tang J, Theilade I, van der Heijden G, van Valkenburg J, Van Do T, Vilanova E, Vos V, Wich S, Wöll H, Yoneda T, Zang R, Zhang M-G, Zweifel N. 2013. Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics. *Global Ecology and Biogeography*, **22**: 1261-1271. DOI: https://doi.org/10.1111/geb.12092

Snowdon P, Raison RJ, Keith H, Ritson P, Grierson P, Adams M, Montagu K, Bi H, Burrows W, Eamus D. 2002. *Protocol for sampling tree and stand biomass*. National Carbon Accounting System Technical Report No. 31. The Australian Greenhouse Office, Canberra, Australia.

Sunderland T, Walters G, Issembe Y. 2004. Preliminary study of the vegetation of the national park of Mbe, Crystal Mounts, Gabon. Central African Regional Programme for the Environment, Smithsonian Institution, 47 p.

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**: 1648-1657. DOI: https://doi.org/10.1016/j.foreco.2011.07.018

Xiao X, White EP, Hooten MB, Durham SL. 2011. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. *Ecology*, **92**(10): 1887–1894. DOI: https://doi.org/10.1890/11-0538.1

Zanne AE, Lopez-Gonzalez G, Coomes DA, Ilic J, Jansen S, Lewis SL, Miller RB, Swenson NG, Wiemann MC, Chave J. 2009. Data from: Towards a Worldwide Wood Economics Spectrum. Dryad Digital Repository. DOI: http://dx.doi.org/10.5061/dryad.234