Sampling trees to develop allometric biomass models: How does tree selection affect model prediction accuracy and precision?

Ioan Dutcă [a,b,*], Richard Mather [b] and Florin Ioraș [b]

[a] Department of Silviculture, Transilvania University of Brasov, 1 Șirul Beethoven, Brașov 500123, Romania

[b] Buckinghamshire New University, Queen Alexandra Rd, High Wycombe HP11 2JZ, UK

[*] Corresponding author E-mail: idutca@unitbv.ro

Abstract

Developing allometric biomass models is an important process because reliability of forest biomass and carbon estimations largely depend on the accuracy and precision of such models. The effects of tree sampling on tree aboveground biomass (AGB) prediction accuracy and precision are complex and can, therefore, be difficult to quantify. In this paper we use a Monte Carlo simulation to investigate how model prediction accuracy and precision are affected by tree sampling approaches. Because diameter at breast height (D, in cm) is the most common predictor of tree AGB (in kg dry weight), we focused our analysis on the AGB-D relationship. The following sample characteristics were investigated: (i) sample size; (ii) extent of the D-range (difference between the largest and the smallest D value); (iii) position of D-range (characterized by the starting point of D-range); and (iv) the size-distribution (distribution of D) of sample trees. We found that, although the natural variability of AGB-D relationship was a key driver for both prediction accuracy and precision, the above sample characteristics were important for improving prediction accuracy. Although having a negligible effect on precision, both sample size and size-distribution of sample trees, greatly influenced prediction accuracy. We demonstrate that selecting a constant number of trees for each D class (i.e. uniform distribution of the sample trees over the D-range) generally produced models that were more accurate predictors of AGB. The extent and position of D-range, although considerably affecting the goodness

of fit and the standard errors of allometric model parameters, had only a marginal effect on AGB prediction accuracy and precision. Furthermore, we showed that R^2 was a poor indicator of model prediction accuracy and precision, due to its sensitivity to changes in D-range. These findings inform certain practical recommendations we report for improving the accuracy and precision of biomass prediction.

Keywords: allometric biomass models, tree sample size, aboveground biomass, diameter at breast height, diameter distribution, sampling characteristics

Abbreviations

D	tree diameter at breast height (in cm);
AGB	aboveground biomass of a tree (in kg dry weight);
D-range	an interval of simulated D observations used to develop an allometric model, and characterized by the starting and ending points of the interval;
S_3	a D-range between 0.1 and 60 cm;
S_2	a D-range between 10 and 60 cm;
S_1	a D-range between 20 and 60 cm;
I_{min}	a D-range between 30 and 60 cm;
B_1	a D-range between 30 and 70 cm;
B_2	a D-range between 30 and 80 cm;
B_3	a D-range between 30 and 90 cm;
I_{max}	a D-range between 0.1 and 90 cm;
RSE	residual standard error;
n	sample size;
β_0	the intercept of a linear allometric model in logarithmic scale;
β_1	the slope of linear allometric model in logarithmic scale;
$SE(\beta_0)$	standard error of the intercept;
$SE(\beta_1)$	standard error of the slope;
R^2	coefficient of determination;
P_A	standard deviation of relative bias, reported as a measure of prediction accuracy;
P_P	mean coefficient of variation of predicted biomass, reported as a measure of prediction precision.

1. Introduction

It is widely accepted that forests play a critical role in the fight against climate change (Grassi et al., 2017), and that the accumulation of carbon in tree biomass is regarded as an important service

provided to society. However, the development of sustainable mitigation measures and programmes such as REDD+ (Reducing Emissions from Deforestation and Forest Degradation) requires that accumulation of carbon in forests is accurately and precisely estimated. Estimating carbon accumulation in forests is typically achieved using forest inventory records, to which allometric models are applied (Brown, 2002; Chave et al., 2004; Clark et al., 2001; Stephenson et al., 2014). To determine carbon sequestration forest biomass is first estimated, then, using a constant proportionality ratio, e.g. 0.47 (IPCC, 2006), the equivalent carbon content may then be calculated, which can be further converted to express CO_2. Therefore, since the ratio between biomass and carbon is a constant, the terms 'carbon accumulation' and 'biomass accumulation' have approximately the same meaning.

Producing accurate and precise predictions of biomass is challenging for several reasons. First, it needs an unbiased forest inventory design with accurate measurements of tree attributes. Second, it requires that allometric biomass models are representative for the forest inventory data to which the model is applied. Selection of the allometric model has been shown to be an important step for reducing biomass prediction uncertainty (Picard et al., 2015). Allometric biomass models are nonlinear regression models that typically use tree diameter at breast height (D, in cm) and/or tree height (H, in m) to predict tree aboveground biomass (AGB, in kg dry weight). Models are based on a sample of trees for which biomass was measured. Representativeness of the model to the forest inventory data requires that sample trees are selected from the inventoried population. Allometric biomass models were shown to be greatly influenced by site conditions (Dutcă et al., 2018a). This in turn may increase the complexity of tree sampling and reduce their transferability of the models to other sites (Dutcă, 2019).

The range of tree sizes and their distribution across the range are important prerequisites for determining sample strata. The range represents the difference between largest and the smallest value of predictor (e.g. D) for the sample trees used to build the model. The distribution of sample trees (on D-range) is often referred to as 'D class distribution' (Chave et al., 2004; Roxburgh et al., 2015) because D is usually measured in forest inventories in scales of increment categories (e.g. intervals of 2 cm). However, when developing allometric biomass models, diameter at breast height (D) is measured as accurately as possible and represented as a continuous variable.

Because allometric models are site-specific (Dutcă, 2019; Dutcă et al., 2018a), there are numerous examples of published allometric models based on trees sampled from one or few forest stands (Chojnacky et al., 2014; Jia et al., 2015; Marziliano et al., 2015; Morhart et al., 2016, 2013; Mosseler et al., 2014; Zianis et al., 2005), which therefore have limited and less than optimal D-range. Alternatively, allometric models may be deliberately developed to represent biometrics of small trees only (e.g. Pajtík et al. 2008; Dutcă et al. 2010; Blujdea et al. 2012; Ciuvat et al. 2013). Nevertheless, tree size is subject to natural limitations; maximum tree height is influenced by physiological stress and resource abundancy as well as hydraulic constraints (Koch et al., 2004). Although maximum tree height is physically limited, trees continue to accumulate biomass by increasing their diameter (Stephenson et al., 2014). Generic allometric models and biomass databases often include very large trees, for example, D of up to 212 cm (Chave et al., 2014), up to 293 cm (Jucker et al., 2017) or even as much as 648 cm (Falster et al., 2015).

The process of biomass measurement is very resource intensive. It is, therefore, important that sampling is optimized to ensure that the resulting allometric model predicts biomass as accurately and precisely as possible. In this paper, using a Monte Carlo analysis, we investigate which approaches of tree selection affect biomass prediction accuracy and precision and how these factors exert their influence. The sample characteristics that were investigated are: (i) sample size; (ii) the extent of D-range (i.e. difference between largest and the smallest sample tree); (iii) position of D-range (i.e. the starting or ending point of the range); and (iv) the distribution of sample trees (i.e. the frequency distribution of selected trees across the D-range).

To demonstrate the effects of sample characteristics on biomass prediction accuracy and precision we performed a simulation study. This involved the following steps: (1) bivariate sets of AGB-D data were simulated to capture key characteristics of the sample trees (e.g. AGB-D variability, sample size, D-range, size-distribution of the sample trees); (2) allometric biomass models were fitted to simulated data; (3) the allometric biomass models were then applied to predict the biomass in a plot and the errors from model parameters and residual variability were propagated to determine their effects on plot AGB prediction; (4) the AGB prediction accuracy and precision (at plot level) were assessed; (5) an examination was made to identify which characteristics of the sample

121 trees considered in the first step (i.e. AGB-D variability, sample size, D-range, size distribution of the
122 sample trees) affected the model's prediction accuracy and precision, and to determine the nature and
123 extent of these affects. Our study aims to inform improvements in the overall accuracy and precision
124 of biomass prediction for forests, and to suggest measures for developing robust allometric biomass
125 models.

2. Material and methods

2.1. Some rationale on the simulation design

Although logarithmic transformation (Huxley, 1932; Snell, 1892) is widely regarded as a standard procedure in the development of allometric biomass models, its use is the subject of some debate (Kerkhoff and Enquist, 2009; Packard, 2012; Packard and Boardman, 2008; Xiao et al., 2011). The standard assumptions of this type of transformation are: (i) heteroscedasticity, which is common in allometric models, is entirely removed by transformation; and (ii) because errors are lognormally distributed when back-transformed (original scale), they will be normally distributed in log-log scale. If these two assumptions hold true, then the back-transformed errors can be assumed to be multiplicative (Cole and Altman, 2017). In other words, the back-transformed residuals may be expressed as a ratio between observed and predicted biomass and therefore indicate the percent variation of observed biomass relative to predicted biomass. However, if the two assumptions do not hold true, then the logarithmic transformation is not recommended, as the general assumptions of a linear model (e.g. normality of residuals, homogeneity of variance) would not be met. Xiao et al. (2011) showed that although both the multiplicative and the additive error-type relationships occur in nature, multiplicative errors were much more frequent. Also, because diameter at breast height (D) is the most common predictor of individual tree aboveground biomass (AGB), we have focused our simulation on AGB-D relationship, starting with a log-log linear model:

$$\ln(\mathrm{AGB}) = \beta_0 + \beta_1 \cdot \ln(\mathrm{D}) + \varepsilon \quad \text{(Eq. 1)}$$

Where: AGB is the aboveground biomass (in kg dry weight); D is the diameter at breast height (in cm); 'ln' is the natural logarithm; β_0 and β_1 are the model parameters in logarithmic scale; and ε is the additive error term (additive for the log-log scale), normally distributed with a mean of zero. We then defined some true parameters for a hypothetical population. Because the population is hypothetical, to make the values of parameters credible, we derived the parameters from a real biomass dataset reported by Schepaschenko et al. (2017). The true model parameters for our hypothetical population were:

$$\ln(\mathrm{AGB}) = -2.11 + 2.33 \cdot \ln(\mathrm{D}) + \varepsilon \quad \text{(Eq. 2)}$$

Starting with these true parameters, we generated random sets of ln(AGB) – ln(D) data which were further fitted. The error term (ε in Eq. 2) is normally distributed with the mean zero and standard error of residuals, RSE. The resulting model was then applied to a plot dataset to estimate the biomass. Each generated dataset had specified characteristics, such as RSE (residual standard error) of log-log model, number of observations, D-range extent, position and distribution. A Monte Carlo approach (described below) was used.

2.2. Natural variability of AGB-D relationship

Sampling design should capture the natural variability of AGB-D relationship that is intrinsic to the population. Because we assumed that heteroscedasticity is removed by logarithmic transformation and that errors are lognormally distributed in original scale, the natural (or intrinsic) variability of AGB-D relationship can be expressed as the residual standard error (RSE) of the log-log linear model (see Eq. 2). Since the residuals of a back-transformed log-log linear model show relative variation of AGB (relative to predicted AGB), the RSE can be interpreted, for original scale, as a form of coefficient of variation (Cole and Altman, 2017). We tested two values of RSE in this study, 0.2 and 0.3, which can be interpreted as 20% and 30% coefficient of variation. These two values lie within the expected range for allometric biomass models (Roxburgh et al., 2015).

2.3. Sample characteristics

2.3.1. Number of observations (sample size)

The number of sample trees necessary to develop an allometric model depends on the precision required, the level of intrinsic variability in the AGB-D relationship and other factors. Roxburgh et al. (2015) performed a simulation study to find the number of sampled trees necessary to develop allometric models. They concluded that, given the intrinsic variability of trees and the differences between distribution of tree diameters used to construct the model and the distribution of tree diameters of the inventory data, a number of anywhere between 17 to 166 trees were required to obtain prediction with a standard deviation within 5% from the mean. However, Picard et al. (2012)

suggested that approximately a minimum number of 100 trees was needed to construct reliable volume models. In our simulation design we tested three values of sample size, n = 100, n = 150 and n = 1000 trees. The first two values (n = 100 and n = 150) were intended to determine the effect of a 50% increase in sample size, as to compare it to a 50% increase in RSE (from RSE = 0.2 to RSE = 0.3). The third value (n = 1000) was intended to see how increasing the sample size influences model prediction performance.

2.3.2. The extent of D-range

The range of diameter at breast height (D) used in allometric biomass models varies greatly. In a review of allometric models, Zianis et al. (2005) most models were based on a relatively narrow D-range with no consistent starting point (minimum D) for the range. For example, the largest tree of 90 cm was recorded in an allometric model for *Quercus ilex* in Italy and the minimum recorded diameter was 20 cm. Comparable maximum limits of D-range are reported in recent biomass datasets for boreal and temperate forests (Schepaschenko et al., 2017; Ung et al., 2017), and larger D-range are reported for trees sampled in tropical regions (Chave et al., 2014; Falster et al., 2015; Jucker et al., 2017). For our simulation study, we assumed a maximum D-range in allometric biomass models between 0.1 and 90 cm (after the D-range reported by Zianis et al., 2005), and divided the range into three equal diameter intervals of 30 cm. Starting from the second interval (i.e. I_{min} = [30, 60]), we gradually expanded I_{min} in two directions (i.e. towards small diameter and towards large diameters) until reaching the limits of the maximum D-range. This resulted in seven D-ranges. We examined the entire D-range (i.e. I_{max} = [0.1, 90]), thereby testing a total of eight D-ranges (as summarised in Table 1).

Table 1

D-ranges used for simulation (D is the diameter at breast height)

Code	D-range (cm)	Description
S_3	[0.1, 60]	I_{min} + 30 cm towards small diameters
S_2	[10, 60]	I_{min} + 20 cm towards small diameters
S_1	[20, 60]	I_{min} + 10 cm towards small diameters

I_{min}	[30, 60]	The minimum D-range
B_1	[30, 70]	I_{min} + 10 cm towards large diameters
B_2	[30, 80]	I_{min} + 20 cm towards large diameters
B_3	[30, 90]	I_{min} + 30 cm towards large diameters
I_{max}	[0.1, 90]	The maximum D-range

207

208 **2.3.3. The position of D-range**

209 The position of D-range is characterized by the starting point of D-range. Each member of each pair
210 of identical D-range extent began at a different position (Table 1). For example, the ranges S_1 and B_1
211 have the same 40 cm range but their starting positions differ by 10 cm. This difference increases to 20
212 cm for S_2 vs. B_2 and to 30 cm for S_3 vs. B_3 (Table 1).

213

214 **2.3.4. Distribution of sample trees**

215 The frequency distribution required for sampling trees and for developing robust models is an
216 important consideration because it determines the level of resources and logistics required for
217 measuring biomass. If trees were entirely randomly sampled, the sample size-distribution would
218 approach that of the population. However, trees are not entirely randomly sampled because the sample
219 is first stratified for each D-class, before random sampling is conducted within D-classes (McRoberts
220 et al., 2015). A 'D class' groups trees within a specified D-range. Thus, for a 2 cm D class the entire
221 D-range is divided into intervals (classes) of 2 cm (e.g. D = 10 to 12 cm). Workers therefore are able
222 to determine how they represent frequency distributions through their selection of the range
223 represented and the bins for each D-class. Nevertheless, the distribution of sample trees will influence
224 how well the model is informed across the range of D, with consequences for confidence in model
225 prediction. In our simulation, we explored four types of distribution (Fig. 1):

226 (a) Uniform distribution on D-range (Fig. 1, a) of the sample frequency, where a constant number
227 of sample trees is selected for each D class.

228 (b) Normal distribution on D-range of the sample frequency (Fig. 1, b), where the sample
229 frequency reflects a normal distribution of D. In other words, the largest number of sample
230 trees is from the middle of D-range and decreases towards the margins of the range;

(c) Uniform distribution on ln(D)-range (Fig. 1, c1), which, for the original scale is equivalent to inverse of uniform distribution (Fig. 1, c2, the result of exponentiation of observations sampled from a uniform distribution on ln(D)-range).

(d) Normal distribution on ln(D)-range (Fig. 1, d1). This is equivalent to lognormal distribution on D-range (Fig. 1, d2). For both, the uniform and normal distribution on ln(D)-range, a larger number of small trees is sampled compared to large trees (Fig. 1, c2 and d2).

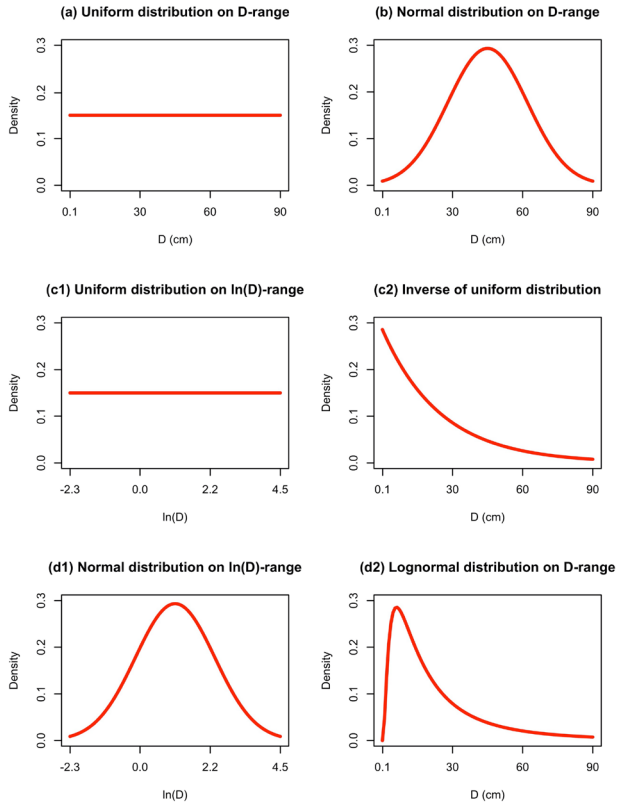

Fig. 1. Distributions of sample trees used for simulations: (a) Uniform distribution on D-range (D is the tree diameter at breast height); (b) Normal distribution on D-range; (c1) Uniform distribution on ln(D)-range, which is equivalent to the inverse of uniform distribution (c2); (d1) Normal distribution on ln(D)-range, for which, the equivalent of original scale is the lognormal distribution (d2).

It is relatively straightforward to define D-limit ranges for uniform distributions. However, the normal distribution for D theoretically extends to infinity. For our simulation we therefore sampled from a truncated normal distribution, for which the lower and upper bounds of D-range were established using the 'truncnorm' package in R (Mersmann et al., 2018). We set the D-range to correspond to ±

two standard deviations, equal to an interval expected to include 95% of observations from a normal distribution. The mean of the normal distribution (μ_d) was the mean of D of the corresponding sample:

$$\mu_d = D_{min} + \frac{(D_{max} - D_{min})}{2} \quad \text{(Eq. 3)}$$

and the standard deviation (σ_d) was calculated as:

$$\sigma_d = \frac{\mu_d - D_{min}}{2} \quad \text{(Eq. 4)}$$

where D_{min} and D_{max} are the minimum and maximum limits of the D-range of interest (Table 1). For example, the normal distribution for I_{min} = [30, 60] cm was defined by the mean, μ_d = 45, with standard deviation, σ_d = 7.5.

2.4. Plot data

We compared the accuracy and precision of model simulations for estimating the biomass in a plot. Each allometric model developed on simulated data was applied to estimate the biomass in a 500 m^2 plot. The plot contained 21 trees for which biomass was predicted as a function of D using all simulated models. Because I_{min} = [30, 60] was the largest interval common to all the D-ranges tested, we selected a plot that contained only tree diameters that fell within this interval (Fig. 2). The purpose of this plot was therefore to provide a reference for prediction for all the simulated models in this study. In total, 0.96 million allometric models (5000 simulations × 2 RSE values × 3 sample sizes × 4 types of distribution × 8 D-ranges) were simulated. Therefore, the value of AGB predicted from this plot is that it provides a baseline for comparing AGB results predicted by other model that use different sample characteristics.

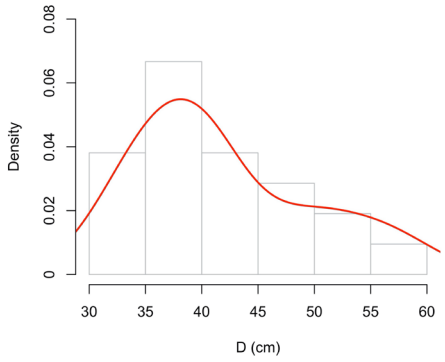

Fig. 2. The size distribution of the 21 sample trees in the plot. Note: D is the diameter at breast height; the red curve represents the kernel density; the grey bars represent the density of each D-class (width of 5 cm).

It is known that models have a poorer prediction performance at the extremes of the covariate range. For example, a biomass model developed on sample trees with a D-range of 0.1 to 90 cm would normally perform best when predicting biomass for trees at the centre of D-range (D = 45 cm) and progressively worse approaching the sample extremes of D = 0.1 cm or D = 90 cm. Therefore, one study objective was to investigate how models perform across the D-range. Consequently, another reason for working with a single plot with D-range restricted to I_{min} was to investigate the performance of models when only part of the D-range was used for prediction. A third reason for working with only one plot was to exclude other potentially confounding sources of uncertainty. In this study we aimed to describe only that uncertainty arising from model parameters and residuals, and intentionally avoided introducing potentially confounding effects of between site variations.

2.5. Monte Carlo simulation

A Monte Carlo analysis was used to assess the effects of sampling approaches on biomass prediction. We followed the next steps:

1. For the *k*th simulation (K = 5000, is the total number of simulations), an allometric model was developed and then applied to predict biomass in the plot. The allometric model was developed based on simulated ln(AGB)-ln(D) data selected from the hypothetical population:
    1.1. defined a vector representing the errors of log-log linear model. The length of this vector was equal to the sample size (i.e. three values of sample size were used in this analysis, n =100, n = 150 and n = 1000, see section 2.3.1). The elements of the vector were randomly selected from a normal distribution with the mean zero and standard deviation either 0.2 or 0.3. Later in the simulation design, the standard deviation of this distribution will become the residual

296 standard error (RSE) of the allometric model. Two values of RSE were used, RSE = 0.2 and
297 RSE = 0.3, see section 2.2.
298 1.2. defined a vector containing sample ln(D) values, which were randomly selected from a
299 specific distribution type (i.e. four types of distribution were used, see section 2.3.4) and a
300 specific D-range (i.e. a total of eight ranges were used, Table 1). Because models were fitted
301 in log-log scale for uniform and normal distributions of D-range (Fig. 1, a and b), we
302 randomly selected the sample D values from a uniform and normal distribution on D-range
303 and then log-transformed the sampled values (to obtain ln(D) values). For uniform and
304 normal distributions on ln(D)-range, we sampled the ln(D) values directly in log-log scale,
305 from a uniform and respectively normal distribution on ln(D)-range (Fig. 1, see c1 and d1).
306 For each of the k^{th} simulation, a distinct set of ln(D) values was generated, $\ln(D)_{(k)}$.
307 1.3. defined a vector (the length of the vector equals the sample size, see section 2.3.1) containing
308 the sample ln(AGB) values. Using the $\ln(D)_{(k)}$ values (obtained at step 1.2) and the error term
309 (obtained at step 1.1) in Eq. 2, we generated the set of ln(AGB) values, which is also distinct
310 for each of the k^{th} simulation, $\ln(AGB)_{(k)}$.
311 1.4. fitted a linear model on the bivariate set of $\ln(AGB)_{(k)}$ (obtained at step 1.3) and $\ln(D)_{(k)}$
312 values (obtained from step 1.2):

313 $$\ln(AGB)_{(k)} = \beta_{0(k)} + \beta_{1(k)} \cdot \ln(D)_{(k)} + \varepsilon_{(k)} \quad \text{(Eq. 5)}$$

314 1.5. We retained the standard errors of model parameters, $SE(\beta_{0(k)})$ and $SE(\beta_{1(k)})$, and the
315 coefficient of determination for the k^{th} simulation ($R^2_{(k)}$):

316 $$R^2_{(k)} = 1 - \frac{\sum(\ln(AGB)_{i(k)} - \widehat{\ln(AGB)}_{i(k)})^2}{\sum(\ln(AGB)_{i(k)} - \overline{\ln(AGB)}_{(k)})^2} \quad \text{(Eq. 6)}$$

317 Where $\ln(AGB)_{i(k)}$ is the *i*th observed ln(AGB) in the k^{th} simulation; $\widehat{\ln(AGB)}_{i(k)}$ is the *i*th
318 predicted ln(AGB) in the k^{th} simulation and $\overline{\ln(AGB)}_{(k)}$ is the mean of all ln(AGB) values in
319 the k^{th} simulation.
320 1.6. defined the variance-covariance matrix to account for the covariance between $\beta_{0(k)}$ and $\beta_{1(k)}$
321 in the following steps.

2. The allometric model developed within steps #1.1 to #1.6 (one model for each k^{th} simulation) was used to estimate the plot biomass. To propagate the uncertainty from each allometric model (i.e. from model parameters and residual variance) to the plot level estimates, a loop of J = 5000 repetitions was used, adapted from McRoberts et al. (2015, 2016). For the j^{th} repetition:

2.1. defined a vector containing two values ($\beta_{0(j)}$ and $\beta_{1(j)}$) sampled at a time from a bivariate normal distribution (based on variance-covariance matrix of the allometric model developed at step 1.6, and on model parameters, $\beta_{0(k)}$ and $\beta_{1(k)}$, from step 1.4);

2.2. defined a vector containing one error term (ε_j) sampled at a time (one for each j^{th} repetition) from a normal distribution with the standard deviation equal to the residual standard error of the k^{th} allometric model (Eq. 5).

2.3. calculate the predicted biomass for each tree ($\widehat{AGB_i}$) in the plot based on the sampled parameters (from step 2.1) and error (from step 2.2):

$$\widehat{AGB_i} = \exp(\beta_{0(j)} + \beta_{1(j)} \cdot D_i + \varepsilon_j) \quad \text{(Eq. 7)}$$

2.4. calculate the predicted plot biomass ($\widehat{AGB_j}$) as the sum of individual tree predictions:

$$\widehat{AGB_j} = \sum_{i=1}^{m} \widehat{AGB_i} \quad \text{(Eq. 8)}$$

Where m = 21, and m is the total number of trees in the plot.

3. The mean plot biomass, standard error of the mean and the relative bias were calculated over all J repetitions:

3.1. the mean predicted plot AGB over J repetitions:

$$\overline{\widehat{AGB_k}} = \frac{1}{J}\sum_{j=1}^{J} \widehat{AGB_j} \quad \text{(Eq. 9)}$$

3.2. standard error of the mean:

$$\hat{\sigma}_k = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}(\widehat{AGB_j} - \overline{\widehat{AGB}_k})^2} \quad \text{(Eq. 10)}$$

3.3. relative bias:

$$Bias_k(\%) = \frac{\left(\overline{\widehat{AGB}_k} - \mu\right)}{\mu} \cdot 100 \quad \text{(Eq. 11)}$$

where μ is the plot AGB, based on true population parameters (plot true AGB) and was calculated by applying the model based on true parameters (see Eq. 2) with a correction factor (Baskerville,

1972; Goldberger, 1968). The model was applied to all m = 21 trees in the plot and then the sum of individual tree biomasses was calculated. RSE is the residual standard error and can take one of two possible values, 0.2 and 0.3 (see section 2.2):

$$\mu = \sum_{i=1}^{m}(\exp(2.11 + \frac{RSE^2}{2}) \cdot D_i^{2.33}) \quad \text{(Eq. 12)}$$

4. Measures of prediction accuracy and precision were calculated over all simulations (K = 5000 simulations):

4.1. The standard deviation of relative bias, reported as a measure of prediction accuracy (P_A):

$$P_A = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(Bias_k - \overline{Bias})^2} \quad \text{(Eq. 13)}$$

Where $\overline{Bias} = \frac{1}{K}\sum_{k=1}^{K}(Bias_k)$

4.2. The mean coefficient of variation of predicted biomass, reported as a measure of prediction precision (P_P):

$$P_P = \frac{1}{K}\sum_{k=1}^{K}\frac{\hat{\sigma}_k}{\overline{\widehat{AGB}_k}} \cdot 100 \quad \text{(Eq. 14)}$$

Where $\hat{\sigma}_k$ is the standard error of predicted biomass (Eq. 10); $\overline{\widehat{AGB}_k}$ is the mean predicted plot biomass (Eq. 9).

2.6. Prediction accuracy and precision

Prediction accuracy and precision are used to describe the performance of an estimator (Walther and Moore, 2005). This study adopts the definition that prediction accuracy is the difference between a predicted value and the true value (Walther and Moore, 2005). Because our simulation design calculated 5000 values (therefore 5000 'differences' between predicted and true plot AGB, which are normally distributed with a mean of zero), accuracy was reported as the standard deviation for these 5000 values (Standard deviation of relative bias, P_A, Eq. 13). Furthermore, prediction precision is a measure of 'the statistical variance of an estimation procedure' (Walther and Moore, 2005) which is a form of uncertainty arising from random variation. In this study, the precision was reported as the mean coefficient of variation of predicted biomass at plot level (P_P) in Eq. 14.

374 **2.7. Data processing**

375 Simulation analyses were performed in R (R Core Team, 2017) with the RStudio interface (RStudio
376 Team, 2016) and using the packages “MASS” (Venables and Ripley, 2002) and “rtruncnorm”
377 (Mersmann et al., 2018).

378

3. Results

3.1. The effects on standard errors of model parameters and on goodness of fit

The simulation results demonstrate that with increasing D-range, the standard errors of model parameters ($SE(\beta_0)$ and $SE(\beta_1)$ in Eq. 5) decreased while the R^2 values (Eq. 6) increased (Fig. 3 and Appendix 1). Greater standard errors denote a less precise estimation of model parameters, whereas larger R^2 values indicate a better fit of the model to the data. The effects were stronger when the D-range increased towards including small trees (Fig. 3, $S_1 – S_3$) compared to large diameter trees (Fig. 3, $B_1 – B_3$). When increasing the extent of D-range, the largest reduction of $SE(\beta_0)$ and $SE(\beta_1)$ and the largest increase of R^2 occurred for normal distribution on ln(D)-range (Fig. 3, d1-d3). Although in Fig. 3 only presents results for n = 100 and RSE = 0.3, similar patterns were obtained for other values of sample size and RSE (Appendix 1).

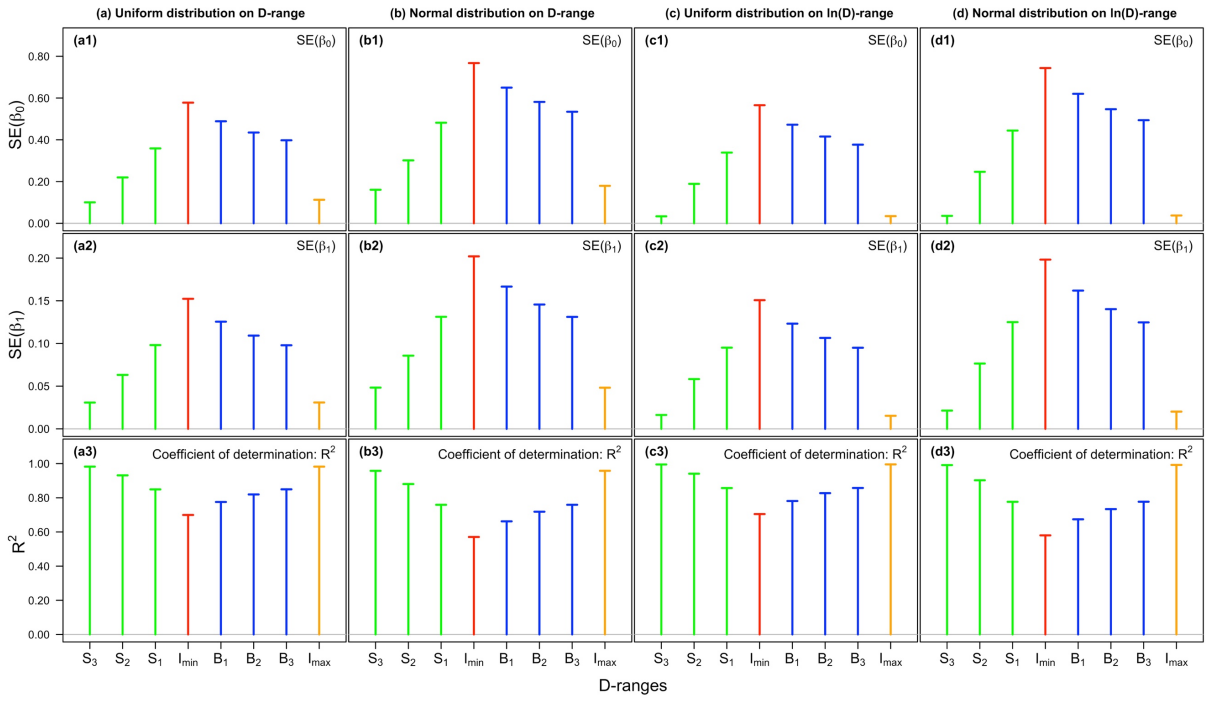

Fig. 3. The standard errors of model parameters $SE(\beta_0)$ and $SE(\beta_1)$, and the model goodness of fit (R^2) for a log-log transformed allometric biomass model (Eq. 5), different types of sample tree distribution and different D-ranges. For D-ranges S_3 to I_{max} (x-axis), see Table 1. Note: Each column of graphs, referred to as (a) to (d), represents a different type of sample tree distribution (for more information

see section 2.3.4); SE(β_0) is the standard error of the intercept in Eq. 5 and was calculated as the mean over all K=5000 simulations: $SE(\beta_0) = \frac{1}{K}\sum_{k=1}^{K}[SE(\beta_{0(k)})]$, where SE($\beta_{0(k)}$) is from step 1.5 in section 2.5; SE(β_1) is the standard error of the slope in Eq. 5, calculated as $SE(\beta_1) = \frac{1}{K}\sum_{k=1}^{K}[SE(\beta_{1(k)})]$, where SE($\beta_{1(k)}$) is from step 1.5 in section 2.5; R^2 is the coefficient of determination, calculated as $R^2 = \frac{1}{K}\sum_{k=1}^{K}(R^2{}_{(k)})$, where $R^2{}_{(k)}$ is from Eq. 6. This figure only presents data for models based on one value of sample size (n = 100) and one value of residual standard error (RSE = 0.3); the data for all values of sample size tested in this study (i.e. n = 100, n = 150 and n = 1000) and all values of RSE (i.e. RSE = 0.2 and RSE = 0.3) are presented in Appendix 1.

The standard errors of model parameters were affected by both RSE and sample size. However, the model goodness of fit (R^2) was affected mainly by the RSE with sample size only having a slight influence.

When RSE was increased by 50% (from 0.2 to 0.3) the standard errors of model parameters (intercept and slope) increased by the same 50% rate (SD = 0.31%; calculated based on values presented in Table A1, and Table A2 in Appendix1) whereas the effect on R^2 was dependent on the extent of the D-range and on the type of distribution (Fig. 3). For models based on smaller D-ranges and on trees sampled over a normal distribution (on either D or ln(D)), the effects of increasing RSE on R^2 were stronger.

When sample size was increased by 50% (from 100 to 150 trees), the standard errors of model parameters reduced, on average, by 18.7% (SD = 0.36%). When sample size was increased by 1000% (from 100 to 1000) the standard errors decreased by 68.7% (SD = 0.33%). Nevertheless, increasing the sample size by 50% (from 100 to 150) and tenfold (from 100 to 1000) led to relatively small changes in mean values for R^2 of only 0.07% and 0.18% respectively (see Appendix 1).

3.2. The effects on biomass prediction accuracy

As expected, residual standard error (RSE) was an important driver for prediction accuracy (expressed as standard deviation of relative bias, P_A, Eq. 13). A low P_A value means that the difference between predicted AGB and true AGB is small, and therefore the model is more accurate. When RSE was increased from 0.2 to 0.3 (therefore, by 50%), P_A increased by approximately the same ratio (i.e. by an average of 51.4%, SD = 2.3%; mean and SD were calculated from 96 P_A values presented in Table A4, Appendix1, using all possible permutations for 8 D-ranges, 3 values of sample size and 4 types of distribution). The effect was stronger for models based on shorter D-ranges (Fig. 4 and Table A4 in Appendix 1). Sample size was also an important factor affecting biomass prediction accuracy, although its effect was weaker when compared to that of RSE. When sample size was increased by 50% (from 100 to 150), P_A decreased by an average of 18.4% (SD = 1.2%; calculated on 96 values in Table A4). Increasing the sample size by tenfold (from 100 to 1000) resulted in an average decrease of P_A of 67% (SD = 0.8%; calculated on 96 values in Table A4). These effects were very similar to those found for standard errors of model parameters (when sample size increased by 50%, the standard errors decreased by 18.7%; when sample size increased tenfold, the standard errors decreased by 68.7%).

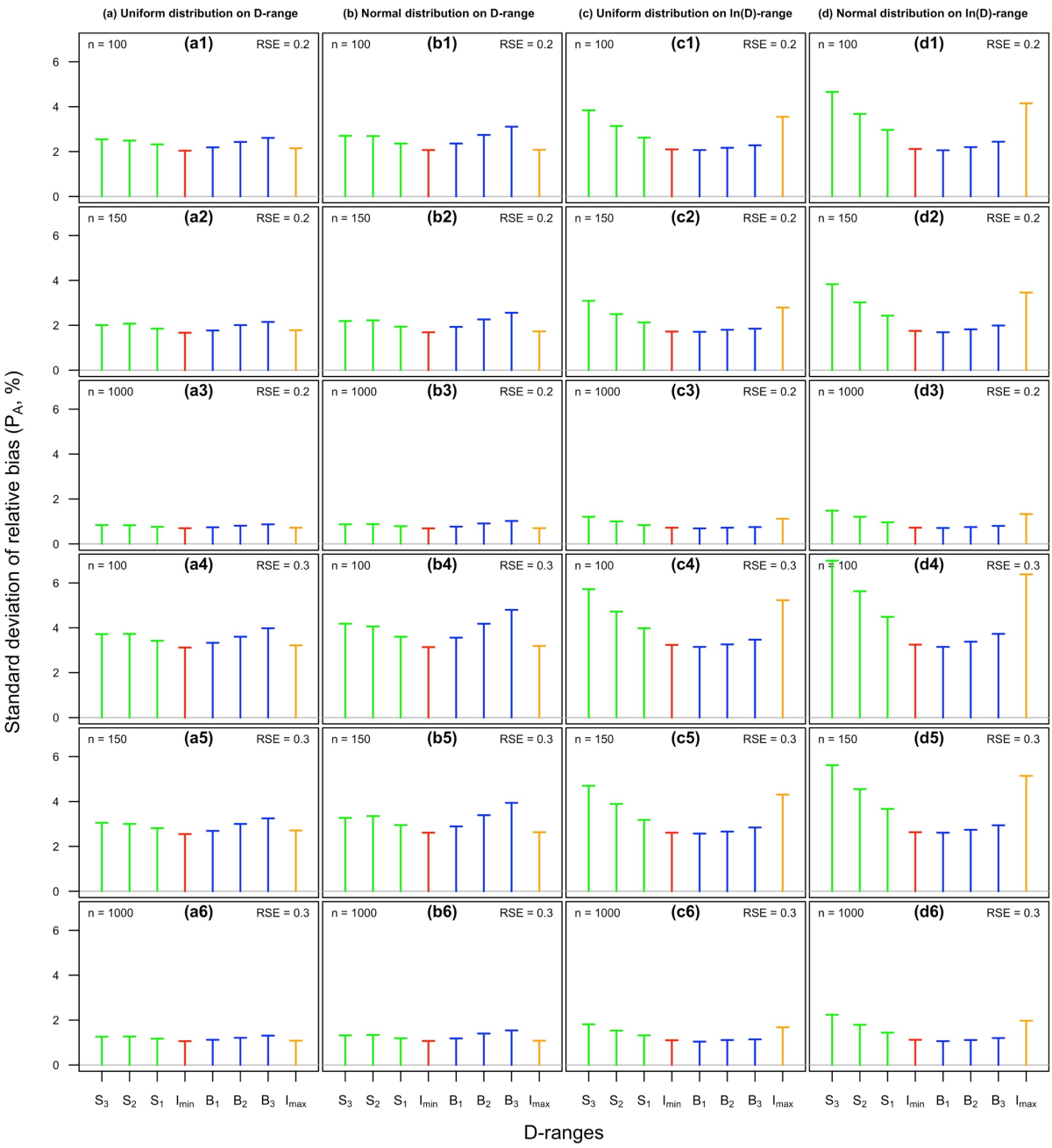

Fig. 4. The standard deviation of relative bias, describing prediction accuracy (P_A, Eq. 13, see section 2.5) for different characteristics of the sample. For D-ranges S_3 to I_{max} (x-axis), see Table 1. Note: Each column of graphs, referred to as (a) to (d), represents a different type of sample tree distribution (for more information see section 2.3.4); The rows 1-3 are for sample sizes (n) of 100, 150 and 1000 trees respectively and RSE = 0.2. Rows 4-6 repeat the same sample sizes for RSE = 0.3.

444 The variation in P_A values was lowest for uniform distribution on D-range (Fig. 4, a1-a6). This means
445 that models constructed with trees selected along a uniform distribution of D-range produced more
446 stable prediction accuracies across the D-range represented by models. In other words, sampling a
447 constant number of trees for each D-class mitigates losses in allometric model accuracy when only
448 limited D-range is available for prediction.

449 However, models that were based on trees selected over uniform or normal distributions over
450 transformed ln(D) range (Fig. 4, c1-c6 and d1-d6), produced larger P_A values for $S_1 – S_3$ ranges
451 compared to $B_1 – B_3$. The cause of these differences lies in how well the model was informed over the
452 range of D = 30 to 60 cm. We mentioned above (section 2.3.4) that the uniform or normal distribution
453 on ln(D) range (see Fig. 1, c1, c2, d1 and d2) assume that a greater number of smaller trees are
454 selected than larger ones. Therefore, the models based on uniform and normal distribution on ln(D)-
455 range (Fig. 4, c1-c6 and d1-d6) are better informed towards the left (small tree) side of D-range
456 distribution. However, the models based on $S_1 – S_3$ (in Fig. 4, c1-c6 and d1-d6) emphasise the right
457 (larger tree) side of D-range for prediction (e.g. models based on S_3 were developed for D = 0.1 to 60
458 cm and were used to predict biomass of trees with D = 30 to 60 cm), which is less well informed.
459 Therefore, the models based on $B_1 – B_3$ ranges produced more accurate predictions of AGB compared
460 to models based on $S_1 – S_3$ ranges.

461 Because the models based on $S_1 – S_3$ and $B_1 – B_3$ ranges used only part of the entire available
462 D-range for prediction (e.g. the model based on S_3 although being developed for D = 0.1 to 60 cm,
463 was used to predict the biomass of trees with D = 30 to 60 cm), these were preferentially tuned to
464 predict I_{min} with $S_1 – S_3$ or $B_1 – B_3$. Since prediction accuracy is poorer at the margins of D-range (for
465 any given model) it is to be expected that P_A values increase slightly (for models based on $S_1 – S_3$ and
466 $B_1 – B_3$ in comparison to models based on I_{min}). However, both I_{min} and I_{max} based models used the
467 central portion of D-range for prediction and therefore these two can be compared to assess how
468 increasing the extent of D-range affects prediction accuracy. Increasing the range from I_{min} to I_{max} did
469 not improve the prediction accuracy and had the opposite effect. This was especially notable for
470 distributions on ln(D)-range (Fig. 4, c1-c6 and d1-d6) for which the P_A value increased by up to 98%.

For models based on uniform and normal distribution on D-range (Fig. 4, a1-a6 and b1-b6) a much smaller increase, of up to 6.6%, was observed.

We demonstrated the effects of increasing D-range from I_{min} to I_{max} when the number of observations remained constant. Therefore, although the models based on I_{max} exhibit greater R^2 and smaller standard errors for model parameters (Fig. 3), their prediction accuracy was poorer compared to models based on I_{min} (Fig. 4, see I_{min} vs. I_{max}). This suggests that the absolute number or density of observations for each part of D-range (or for each diameter class) is important. For the specific D-range of the plot data (i.e. D = 30 to 60 cm), the models based on I_{max} had a lower density of observations, compared to models based on I_{min}, since the same number of observations had to be distributed over a wider D-range (in the case of I_{max} based models). These results are important, because they demonstrate in comparison to model fitting and the standard errors of model parameters, that RSE (in log-log scale) and the absolute number of trees across the D-range are more important determinants of prediction accuracy.

3.3. The effects on biomass prediction precision

Although increasing the D-range the standard errors of model parameters decrease and the R^2 increases (Fig. 3), producing therefore improved models, this improvement was not reflected in the precision of biomass prediction (here, expressed as the mean coefficient of variation of predicted biomass, P_P, in Eq. 14). The P_P did not decrease with the increasing D-range and in some cases even increased slightly (Table 2).

Table 2

The mean coefficient of variation of predicted biomass (P_P, Eq. 14), for uniform and normal distribution on D-range and ln(D) range, for sample sizes of n = 100, n = 150 and n = 1000, for residual standard error RSE = 0.2 and RSE = 0.3 and for D-ranges S_3, S_2, S_1, I_{min}, B_1, B_2, B_3 and I_{max} (for more information on D-ranges, see Table 1).

D-range	Uniform distribution on D-range			Normal distribution on D-range			Uniform distribution on ln(D)-range			Normal distribution on ln(D)-range		
	n=100	*n*=150	*n*=1000	*n*=100	*n*=150	*n*=1000	*n*=100	*n*=150	*n*=1000	*n*=100	*n*=150	*n*=1000
RSE = 0.2												
S_3	20.32	20.29	20.22	20.35	20.30	20.22	20.53	20.41	20.23	20.78	20.54	20.26
S_2	20.31	20.24	20.21	20.32	20.32	20.20	20.44	20.33	20.22	20.54	20.38	20.22
S_1	20.30	20.26	20.20	20.30	20.30	20.22	20.39	20.29	20.20	20.44	20.34	20.22
I_{min}	20.27	20.25	20.20	20.30	20.25	20.21	20.28	20.28	20.20	20.26	20.23	20.20
B_1	20.28	20.27	20.21	20.28	20.25	20.21	20.25	20.26	20.21	20.26	20.27	20.21
B_2	20.29	20.31	20.21	20.39	20.31	20.22	20.29	20.26	20.20	20.25	20.25	20.21
B_3	20.35	20.32	20.21	20.45	20.35	20.23	20.27	20.27	20.21	20.30	20.27	20.21
I_{max}	20.28	20.25	20.21	20.31	20.27	20.21	20.49	20.36	20.20	20.64	20.49	20.25
RSE = 0.3												
S_3	30.94	30.82	30.72	31.07	30.82	30.73	31.30	31.02	30.74	31.68	31.36	30.78
S_2	30.91	30.82	30.71	30.96	30.88	30.70	31.18	30.98	30.73	31.28	31.01	30.75
S_1	30.85	30.81	30.72	30.90	30.88	30.71	30.98	30.89	30.72	31.08	30.94	30.72
I_{min}	30.80	30.78	30.71	30.81	30.78	30.69	30.85	30.80	30.71	30.89	30.83	30.70
B_1	30.86	30.75	30.70	30.90	30.85	30.70	30.89	30.81	30.72	30.85	30.80	30.68
B_2	30.95	30.83	30.69	31.04	30.83	30.73	30.94	30.77	30.69	30.93	30.83	30.70
B_3	30.93	30.88	30.72	31.09	30.98	30.73	30.92	30.81	30.70	30.96	30.88	30.72
I_{max}	30.82	30.78	30.71	30.79	30.76	30.70	31.22	30.99	30.73	31.44	31.14	30.76

From Table 2 it can be seen that P_P is highly related to residual standard error (RSE). Earlier it was mentioned (section 2.2) that RSE in log-log scale can be interpreted as a form of coefficient of variation for the original D-range scale. The slight increases in P_P values over and above base levels of 20% and 30% (for RSE values of 0.2 and 0.3 respectively) are due to uncertainty in model parameters, since P_P values contain errors propagated from both model parameters and residual variance. Therefore, RSE was the main driver of model prediction precision, with a very small proportion produced by uncertainty in model parameters (up to 5.3%). Increasing RSE by 50% (from 0.2 to 0.3) resulted in an average increase in P_P of 52.1% (SD = 0.2%; the mean and SD were calculated on the 96 P_P values presented in Table 2, for each value of RSE), regardless of sample size, D-range and distribution type. However, sample size, although greatly influencing prediction accuracy, had little effect on prediction precision. Since increasing the sample size directly affected the standard errors of model parameters (producing a decrease in standard errors) and since the propagated errors from model parameters represent only a very small proportion of P_P (up to 5.3%), it is to be expected that sample size will have little effect on prediction precision. Increasing the number of observations by 50% (from 100 to 150), had the effect of reducing P_P by 0.33% (SD = 0.29%), and increasing observations tenfold (from 100 to 1000) led to a reduction in P_P by 0.81% (SD = 0.56%).

However, both these effects were found not to be significantly different from zero change ($p = 0.26$ and $p = 0.16$ respectively).

518 **4. Discussion**

519

520 **4.1. Factors influencing biomass prediction accuracy and precision**

521 The effects of tree sampling and data treatment approaches on biomass prediction accuracy and
522 precision are subtle and can sometimes be counterintuitive. Findings here reveal certain
523 characteristics of sampling strategies that are important for improving model prediction accuracy and
524 precision. Of these it is the natural variability of the AGB-D relationship (expressed by RSE) that is
525 the main driver for prediction accuracy and precision, thus an increase in RSE of 50% resulted in
526 proportionally similar improvement in accuracy and precision. Increasing sample size was also found
527 to be important for improving model accuracy but less so for improving precision. The finding that
528 the effect of sample size on prediction accuracy depended on RSE and D-range, and was a function of
529 $1/\sqrt{n}$, where n is the sample size, was consistent with results published from earlier studies (Chave et
530 al., 2004; Picard et al., 2012).

531 Analyses demonstrate how a wider D-range improves model fit and the standard errors of
532 model parameters (Fig. 3). This may also help to ensure that results from statistical tests are properly
533 representative of allometric model performance, because the reduction of standard errors will increase
534 the likelihood that null hypotheses (for no difference) are correctly rejected in analyses such as t- and
535 F- tests (Dutcă et al., 2018b). However, we also showed that, although the model based on a wider D-
536 range had a better fit, the prediction accuracy was poorer (Fig. 4, see I_{min} vs. I_{max}). This result, which
537 may be surprising, can be explained by the frequency of the observations across the D-range. If the
538 number of observations remain constant, increasing the D-range inevitably reduces the density of
539 observations with negative consequences on AGB (aboveground biomass) prediction accuracy. Often,
540 increasing the range of D is achieved by merging datasets for different D-ranges. In this event, the
541 density of observations across the D-range is not reduced and the resulting increase of sample size
542 increases prediction accuracy.

543 Furthermore, Roxburgh et al. (2015) suggested that the optimal size distribution of sample
544 trees to develop allometric models is the one that most closely matches the distribution of trees to
545 which the model is applied. Although our plot data appears to be lognormally distributed (Fig. 2), the

546 greatest accuracy (lowest P_A value) was obtained for models based on a uniform distribution of D-
547 range. This finding is in contradiction with results reported by Roxburgh et al. (2015). Because our
548 plot D data only appeared to be lognormal, we further investigated this phenomenon by generating a
549 new D dataset of 1000 observations lognormally distributed on I_{max} range. We investigated whether
550 the model based on uniform distribution (developed for the same I_{max} range) produced lower P_A and
551 P_P values (when predicting AGB of this new D dataset of 1000 observations) than the model based on
552 lognormal distribution. The results confirmed that uniform distribution on D-range produced lower P_A
553 and P_P values (model based on uniform distribution: P_A = 3.2% and P_P = 30.8%; model based on
554 lognormal distribution: P_A = 6.3% and P_P = 31.4%). We repeated the comparison, for models based on
555 uniform vs. normal distribution on D-range, when predicting AGB of 1000 trees normally distributed.
556 Again, the model based on uniform distribution produced lower P_A and P_P values compared to model
557 based on normally distributed sample trees (model based on uniform distribution: P_A = 3.5% and P_P
558 =30.8%; model based on normal distribution: P_A = 3.6% and P_P = 30.9%). Therefore, our results
559 indicate that models based on uniform distribution of the sample trees on D-range perform better
560 (produce more accurate and precise predictions) regardless of distribution of the trees to which the
561 model is applied.
562
563 **4.2. Small trees are more informative in allometric models**
564 We demonstrate that, for models based on similar number of observations and similar extent of D-
565 range (and similar residual standard errors in logarithmic scale), if models include smaller diameter
566 trees, the standard errors of model parameters were reduced and R^2 values were greater (e.g. see S_3 vs.
567 B_3 in Fig. 3). Therefore, it is suggested that small trees are generally more informative in allometric
568 models, compared to large trees. However, this seemingly anomalous finding can be explained by (or
569 represents the indirect effect of) the heteroscedastic nature of the relationship between biomass and
570 tree diameter. The variance in allometric models is not constant and increases with D (Zianis, 2008).
571 As a result, to fit a nonlinear model the observations are usually weighted inversely to residual
572 variance (the lower the residual variance, the larger the weight and vice-versa) (Dutcă et al., 2019).
573 Logarithmic transformation on the other hand, performs a similar function: it re-scales data so that

574 units are stretched for small values of variables (D and AGB) and compressed for large ones.
575 Therefore, log-log transformation more heavily weights the influence of small trees over large ones,
576 to ensure that residuals are comparable residuals across predictor range (i.e. homoscedasticity).
577 As the lowest residual variance usually occurs for the smallest D values (Zianis and
578 Mencuccini, 2004), small trees are more heavily weighted and have a greater influence on regression
579 models than larger trees. Therefore, small trees impart more information to models, and exert greater
580 overall influence over the standard errors of model parameters and goodness of fit. Given the fact that
581 small trees require less effort for biomass measurement, they are highly cost-effective to sample.
582 Nevertheless, we have demonstrated that, although the models that included small trees produced
583 smaller standard errors of model parameters and larger R^2 values, they did not necessarily produce
584 more accurate or precise predictions of AGB (Fig. 4 and Table 2).
585
586 **4.3. Selection criteria of allometric models**
587 Goodness of fit (R^2 of linear model in log-log scale) is often reported with allometric biomass models,
588 and is widely accepted as a criterion for model selection (Sanquetta et al., 2018). The assumption is
589 that a model with the best fit will reasonably predict the biomass of other trees. Our results confirm
590 that R^2 was not affected by sample size (Sanquetta et al., 2018). However, we showed that R^2 was a
591 poor indicator of model prediction performance with respect to both accuracy and precision. Plotting
592 the R^2 against P_A (Fig. 5, a) and P_P (Fig. 5, b) we observed no clear relationship between R^2 and
593 model prediction accuracy or precision.
594 Although not sensitive to changes in sample size, R^2 was sensitive to variations in D-range
595 (Fig. 3 and Appendix 1). Models yielded greater values of R^2 for the maximum extents of D-range
596 (i.e. I_{max}, see Fig. 3) and when distribution of sampled trees was uniform on ln(D)-range ($R^2 = 0.998$,
597 Fig. 3 and Table A3, Appendix 1). However, we showed that the extent of D-range did not affect
598 prediction accuracy nor precision, and that actually the models based on trees sampled along a ln(D)-
599 range produced poorer prediction accuracies. These findings suggest that R^2 may not be a reliable
600 indicator of model prediction performance.

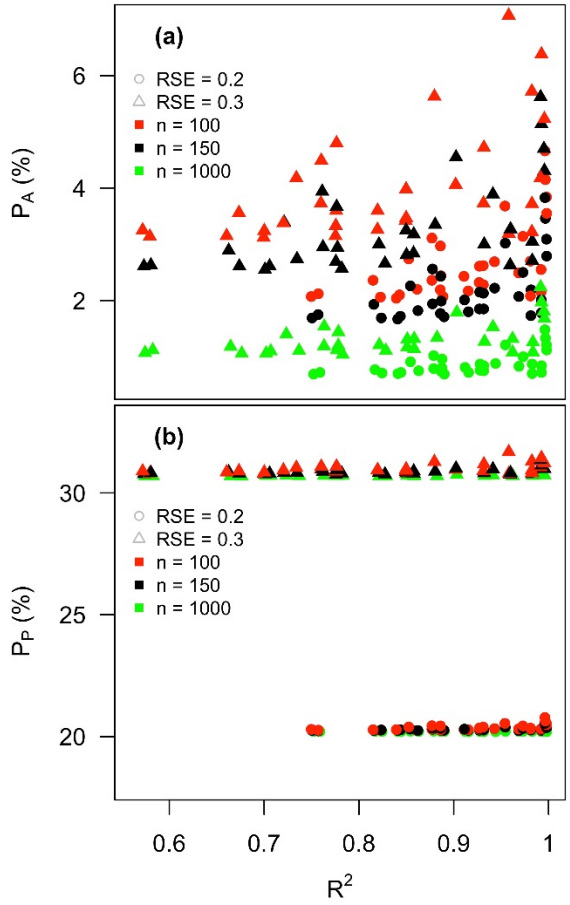

Fig. 5. The relationship between model goodness of fit (R^2, Eq. 6) and prediction accuracy (P_A, standard deviation of relative bias in %, Eq. 13) (a) and between R^2 and prediction precision (P_P, mean coefficient of variation of predicted biomass in %, Eq. 14) (b). Note: The plotted P_A values are from Table A4 (Appendix 1); the P_P values are from Table 2; the model R^2 values are from Table A3 (Appendix 1); larger P_A values show lower prediction accuracy; larger P_P values show lower prediction precision.

4.4. Limitations of the study

Our study has the following limitations. Firstly, the conclusions are only valid if the assumptions hold that heteroscedasticity is removed by logarithmic transformation and that errors are normally distributed in log-log scale. Secondly, because the study was limited to the relationship between AGB and D, the conclusions should not be extrapolated to other types of relationships. Thirdly, this study did not consider the uncertainty arising from between site variation. Fourthly and finally, we have

assumed that the diameters of trees in the inventory (plot) dataset were always within the D-range used to construct the model. We did not investigate the consequences of predicting AGB of trees outside the range of diameters used to construct the models.

4.5. Recommendations

Study findings suggest that the following guidelines will be useful in the preparation of reliable allometric models:

(1) **Select a constant number of trees for each D class (use a uniform distribution of sample trees).** Results demonstrate that the models based on uniformly distributed sample trees over the D-range (D is the diameter at breast height) produced more accurate AGB predictions (AGB is the aboveground tree biomass), regardless of D-distribution of the inventory dataset. Also, variations in prediction accuracy across D-range were minimal.

(2) **Using R^2 as criterion for model selection should be done with caution**. Findings suggest that R^2 (coefficient of determination) alone is not a strong indicator of model prediction performance.

(3) **Use strategies to avoid unnecessary large levels of RSE in allometric models**. Because RSE (Residual Standard Error of the model in log-scale) is indicative of the intrinsic AGB variability for any given D, it cannot be naturally reduced. However, because RSE was a key driver of both prediction accuracy and precision, it is recommended that strategies are adopted to help reduce unnecessary AGB variability, such as: (i) avoiding using generic allometric models, where species effect is ignored and, therefore, to use species-specific allometric models wherever possible; (ii) test and include additional predictors in the models that may explain part of the residual variance, such as tree height, crown diameter and wood density.

5. Conclusions

The key conclusions drawn from this study are as follow: (i) residual variance was the most important driver of model's prediction accuracy and precision; (ii) increasing the sample size improved prediction accuracy (although its effect was weaker than that of residual standard error), but had negligible effect on prediction precision; (iii) increasing the extent of D-range, although improving both the goodness of fit and standard errors of model parameters, did not affect prediction accuracy nor precision; (iv) the size distribution of sample trees was important for prediction accuracy; we found that uniform distribution of D-range was optimal, regardless of the distribution of the inventory dataset; (v) small trees were more informative in allometric models, due to the effects of inherently heteroscedastic variance; (vi) R^2 was not a good indicator of prediction performance of allometric models.

Acknowledgements

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CCCDI – UEFISCDI, project number ERANET-FACCE ERAGAS – FORCLIMIT (82/2017), within PNCDI II. We thank the anonymous reviewers whose comments have greatly improved this manuscript.

655 **References**

656 Baskerville, G.L., 1972. Use of Logarithmic Regression in the Estimation of Plant Biomass. Can. J.
657 For. Res. 2, 49–53. doi:10.1139/x72-009

658 Blujdea, V.N.B., Pilli, R., Dutca, I., Ciuvat, L., Abrudan, I.V., 2012. Allometric biomass equations for
659 young broadleaved trees in plantations in Romania. For. Ecol. Manage. 264, 172–184.
660 doi:10.1016/j.foreco.2011.09.042

661 Brown, S., 2002. Measuring carbon in forests: current status and future challenges. Environ. Pollut.
662 116, 363–372. doi:10.1016/S0269-7491(01)00212-3

663 Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., Perez, R., 2004. Error propagation and
664 scaling for tropical forest biomass estimates. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359, 409–
665 20. doi:10.1098/rstb.2003.1425

666 Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque,
667 A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A.,
668 Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-
669 Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014.
670 Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Chang.
671 Biol. 20, 3177–3190. doi:10.1111/gcb.12629

672 Chojnacky, D.C., Heath, L.S., Jenkins, J.C., 2014. Updated generalized biomass equations for North
673 American tree species. Forestry 87, 129–151. doi:10.1093/forestry/cpt053

674 Ciuvat, A.L., Abrudan, I.V., Blujdea, V., Dutca, I., Nuta, I.S., Edu, E., 2013. Biomass Equations and
675 Carbon Content of Young Black Locust (Robinia pseudoacacia L.) Trees from Plantations and
676 Coppices on Sandy Soils in South-Western Romanian Plain. Not. Bot. Horti Agrobot. Cluj-
677 Napoca 41, 590–592. doi:10.15835/NBHA4129355

678 Clark, D.A., Brown, S., Kicklighter, D.W., Chambers, J.Q., Thomlinson, J.R., Ni, J., 2001. Measuring
679 Net Primary Production in Forests: Concepts and Field Methods. Ecol. Appl. 11, 356–370.
680 doi:10.1890/1051-0761(2001)011[0356:MNPPIF]2.0.CO;2

681 Cole, T.J., Altman, D.G., 2017. Statistics Notes: Percentage differences, symmetry, and natural
682 logarithms. BMJ 358, j3683. doi:10.1136/bmj.j3683

683 Dutcă, 2019. The Variation Driven by Differences between Species and between Sites in Allometric
684 Biomass Models. Forests 10, 976. doi:10.3390/f10110976
685 Dutcă, I., Abrudan, I.V., Stăncioiu, P.T., Blujdea, V.N.B., 2010. Biomass conversion and expansion
686 factors for young Norway Spruce (Picea abies (L.) Karst.) trees planted on non-forest lands in
687 Eastern Carpathians. Not. Bot. Horti Agrobot. Cluj-Napoca 38.
688 doi:http://dx.doi.org/10.15835/nbha3835450
689 Dutcă, I., Mather, R., Blujdea, V.N.B., Ioraș, F., Olari, M., Abrudan, I.V., 2018a. Site-effects on
690 biomass allometric models for early growth plantations of Norway spruce (Picea abies (L.)
691 Karst.). Biomass and Bioenergy 116, 8–17. doi:10.1016/j.biombioe.2018.05.013
692 Dutcă, I., McRoberts, R.E., Næsset, E., Blujdea, V.N.B., 2019. A practical measure for determining if
693 diameter (D) and height (H) should be combined into D2H in allometric biomass models. For.
694 An Int. J. For. Res. 92, 627–634. doi:10.1093/forestry/cpz044
695 Dutcă, I., Stăncioiu, P.T., Abrudan, I.V., Ioraș, F., 2018b. Using clustered data to develop biomass
696 allometric models: The consequences of ignoring the clustered data structure. PLoS One 13,
697 e0200123. doi:10.1371/journal.pone.0200123
698 Falster, D.S., Duursma, R.A., Ishihara, M.I., Barneche, D.R., FitzJohn, R.G., Vårhammar, A., Aiba,
699 M., Ando, M., Anten, N., Aspinwall, M.J., Baltzer, J.L., Baraloto, C., Battaglia, M., Battles, J.J.,
700 Bond-Lamberty, B., van Breugel, M., Camac, J., Claveau, Y., Coll, L., Dannoura, M.,
701 Delagrange, S., Domec, J.-C., Fatemi, F., Feng, W., Gargaglione, V., Goto, Y., Hagihara, A.,
702 Hall, J.S., Hamilton, S., Harja, D., Hiura, T., Holdaway, R., Hutley, L.S., Ichie, T., Jokela, E.J.,
703 Kantola, A., Kelly, J.W.G., Kenzo, T., King, D., Kloeppel, B.D., Kohyama, T., Komiyama, A.,
704 Laclau, J.-P., Lusk, C.H., Maguire, D.A., le Maire, G., Mäkelä, A., Markesteijn, L., Marshall, J.,
705 McCulloh, K., Miyata, I., Mokany, K., Mori, S., Myster, R.W., Nagano, M., Naidu, S.L.,
706 Nouvellon, Y., O'Grady, A.P., O'Hara, K.L., Ohtsuka, T., Osada, N., Osunkoya, O.O., Peri,
707 P.L., Petritan, A.M., Poorter, L., Portsmuth, A., Potvin, C., Ransijn, J., Reid, D., Ribeiro, S.C.,
708 Roberts, S.D., Rodríguez, R., Saldaña-Acosta, A., Santa-Regina, I., Sasa, K., Selaya, N.G.,
709 Sillett, S.C., Sterck, F., Takagi, K., Tange, T., Tanouchi, H., Tissue, D., Umehara, T., Utsugi,
710 H., Vadeboncoeur, M.A., Valladares, F., Vanninen, P., Wang, J.R., Wenk, E., Williams, R., de

711 Aquino Ximenes, F., Yamaba, A., Yamada, T., Yamakura, T., Yanai, R.D., York, R.A., 2015.
712 BAAD: a Biomass And Allometry Database for woody plants. Ecology 96, 1445–1445.
713 doi:10.1890/14-1889.1
714 Goldberger, A.S., 1968. The Interpretation and Estimation of Cobb-Douglas Functions. Econometrica
715 36, 464–472. doi:10.2307/1909517
716 Grassi, G., House, J., Dentener, F., Federici, S., den Elzen, M., Penman, J., 2017. The key role of
717 forests in meeting climate targets requires science for credible mitigation. Nat. Clim. Chang. 7,
718 220–226. doi:10.1038/nclimate3227
719 Huxley, S.J., 1932. Problems of Relative Growth, 1st ed. The Dial Press, New York.
720 doi:10.1038/129775a0
721 IPCC, 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Institute for Global
722 Environmental Strategies.
723 Jia, Q., Liu, Q., Li, J., 2015. Individual-based fine root biomass and its functional relationship with
724 leaf for Pinus tabuliformis in northern China. Eur. J. For. Res. 134, 705–714.
725 doi:10.1007/s10342-015-0884-0
726 Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., Dalponte, M., van Ewijk,
727 K.Y., Forrester, D.I., Haeni, M., Higgins, S.I., Holdaway, R.J., Iida, Y., Lorimer, C., Marshall,
728 P.L., Momo, S., Moncrieff, G.R., Ploton, P., Poorter, L., Rahman, K.A., Schlund, M., Sonké, B.,
729 Sterck, F.J., Trugman, A.T., Usoltsev, V.A., Vanderwel, M.C., Waldner, P., Wedeux, B.M.M.,
730 Wirth, C., Wöll, H., Woods, M., Xiang, W., Zimmermann, N.E., Coomes, D.A., 2017.
731 Allometric equations for integrating remote sensing imagery into forest monitoring programmes.
732 Glob. Chang. Biol. 23, 177–190. doi:10.1111/gcb.13388
733 Kerkhoff, A.J., Enquist, B.J., 2009. Multiplicative by nature: Why logarithmic transformation is
734 necessary in allometry. J. Theor. Biol. 257, 519–521. doi:10.1016/j.jtbi.2008.12.026
735 Koch, G.W., Sillett, S.C., Jennings, G.M., Davis, S.D., 2004. The limits to tree height. Nature 428,
736 851–854. doi:10.1038/nature02417
737 Marziliano, P.A., Lafortezza, R., Medicamento, U., Lorusso, L., Giannico, V., Colangelo, G., Sanesi,
738 G., 2015. Estimating belowground biomass and root/shoot ratio of Phillyrea latifolia L. in the

739 Mediterranean forest landscapes. Ann. For. Sci. 72, 585–593. doi:10.1007/s13595-015-0486-5
740 McRoberts, R.E., Chen, Q., Domke, G.M., Ståhl, G., Saarela, S., Westfall, J.A., 2016. Hybrid
741 estimators for mean aboveground carbon per unit area. For. Ecol. Manage. 378, 44–56.
742 doi:10.1016/J.FORECO.2016.07.007
743 McRoberts, R.E., Moser, P., Zimermann Oliveira, L., Vibrans, A.C., 2015. A general method for
744 assessing the effects of uncertainty in individual-tree volume model predictions on large-area
745 volume estimates with a subtropical forest illustration. Can. J. For. Res. 45, 44–51.
746 doi:10.1139/cjfr-2014-0266
747 Mersmann, O., Trautmann, H., Steuer, D., Bornkamp, B., 2018. truncnorm: Truncated Normal
748 Distribution.
749 Morhart, C., Sheppard, J., Spiecker, H., 2013. Above Ground Leafless Woody Biomass and Nutrient
750 Content within Different Compartments of a P. maximowicii × P. trichocarpa Poplar Clone.
751 Forests 4, 471–487. doi:10.3390/f4020471
752 Morhart, C., Sheppard, J.P., Schuler, J.K., Spiecker, H., 2016. Above-ground woody biomass
753 allocation and within tree carbon and nutrient distribution of wild cherry (Prunus avium L.) – a
754 case study. For. Ecosyst. 3. doi:10.1186/s40663-016-0063-x
755 Mosseler, A., Major, J.E., Labrecque, M., Larocque, G.R., 2014. Allometric relationships in coppice
756 biomass production for two North American willows (Salix spp.) across three different sites.
757 For. Ecol. Manage. 320, 190–196. doi:10.1016/j.foreco.2014.02.027
758 Packard, G.C., 2012. Is non-loglinear allometry a statistical artifact? Biol. J. Linn. Soc. 107, 764–773.
759 doi:10.1111/j.1095-8312.2012.01995.x
760 Packard, G.C., Boardman, T.J., 2008. Model selection and logarithmic transformation in allometric
761 analysis. Physiol. Biochem. Zool. 81, 496–507. doi:10.1086/589110
762 Pajtík, J., Konôpka, B., Lukac, M., 2008. Biomass functions and expansion factors in young Norway
763 spruce (Picea abies [L.] Karst) trees. For. Ecol. Manage. 256, 1096–1103.
764 doi:10.1016/j.foreco.2008.06.013
765 Picard, N., Boyemba Bosela, F., Rossi, V., 2015. Reducing the error in biomass estimates strongly
766 depends on model selection. Ann. For. Sci. 72, 811–823. doi:10.1007/s13595-014-0434-9

767 Picard, N., Saint-André, L., Henry, M., 2012. Manual for building tree volume and biomass allometric
768 equations: from field measurement to prediction. FAO and CIRAD, Rome, Italy, and
769 Montpellier, France.
770 R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for
771 Statistical Computing, Vienna, Austria.
772 Roxburgh, S.H., Paul, K.I., Clifford, D., England, J.R., Raison, R.J., 2015. Guidelines for constructing
773 allometric models for the prediction of woody biomass: How many individuals to harvest?
774 Ecosphere 6, 1–27. doi:10.1890/ES14-00251.1
775 RStudio Team, 2016. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.
776 Sanquetta, C.R., Dalla Corte, A.P., Behling, A., de Oliveira Piva, L.R., Péllico Netto, S., Rodrigues,
777 A.L., Sanquetta, M.N.I., 2018. Selection criteria for linear regression models to estimate
778 individual tree biomasses in the Atlantic Rain Forest, Brazil. Carbon Balance Manag. 13, 25.
779 doi:10.1186/s13021-018-0112-6
780 Schepaschenko, D., Shvidenko, A., Usoltsev, V., Lakyda, P., Luo, Y., Vasylyshyn, R., Lakyda, I.,
781 Myklush, Y., See, L., McCallum, I., Fritz, S., Kraxner, F., Obersteiner, M., 2017. A dataset of
782 forest biomass structure for Eurasia. Sci. Data 4, 170070. doi:10.1038/sdata.2017.70
783 Snell, O., 1892. Die Abhängigkeit des Hirngewichtes von dem Körpergewicht und den geistigen
784 Fähigkeiten. Arch. Psychiatr. Nervenkr. 23, 436–446. doi:10.1007/BF01843462
785 Stephenson, N.L., Das, A.J., Condit, R., Russo, S.E., Baker, P.J., Beckman, N.G., Coomes, D.A.,
786 Lines, E.R., Morris, W.K., Rüger, N., Álvarez, E., Blundo, C., Bunyavejchewin, S., Chuyong,
787 G., Davies, S.J., Duque, Á., Ewango, C.N., Flores, O., Franklin, J.F., Grau, H.R., Hao, Z.,
788 Harmon, M.E., Hubbell, S.P., Kenfack, D., Lin, Y., Makana, J.-R., Malizia, A., Malizia, L.R.,
789 Pabst, R.J., Pongpattananurak, N., Su, S.-H., Sun, I.-F., Tan, S., Thomas, D., van Mantgem, P.J.,
790 Wang, X., Wiser, S.K., Zavala, M.A., 2014. Rate of tree carbon accumulation increases
791 continuously with tree size. Nature 507, 90–93. doi:10.1038/nature12914
792 Ung, C.H., Lambert, M.C., Raulier, F., Guo, X.J., Bernier, P.Y., 2017. Biomass of trees sampled
793 across Canada as part of the Energy from the Forest Biomass (ENFOR) Program.
794 doi:https://doi.org/10.23687/fbad665e-8ac9-4635-9f84-e4fd53a6253c

795 Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S, 4th ed. Springer, New York.
796 Walther, B.A., Moore, J.L., 2005. The concepts of bias, precision and accuracy, and their use in
797 testing the performance of species richness estimators, with a literature review of estimator
798 performance. Ecography (Cop.). 28, 815–829. doi:10.1111/j.2005.0906-7590.04112.x
799 Xiao, X., White, E.P., Hooten, M.B., Durham, S.L., 2011. On the use of log-transformation vs.
800 nonlinear regression for analyzing biological power laws. Ecology 92, 1887–1894.
801 doi:10.1890/11-0538.1
802 Zianis, D., 2008. Predicting mean aboveground forest biomass and its associated variance. For. Ecol.
803 Manage. 256, 1400–1407. doi:10.1016/J.FORECO.2008.07.002
804 Zianis, D., Mencuccini, M., 2004. On simplifying allometric analyses of forest biomass. For. Ecol.
805 Manage. doi:10.1016/j.foreco.2003.07.007
806 Zianis, D., Muukkonen, P., Mäkipää, R., Mencuccini, M., 2005. Biomass and stem volume equations
807 for tree species in Europe. Finnish Society of Forest Science, Finnish Forest Research Institute.
808