ORIGINAL PAPER

Uncertainties in above ground tree biomass estimation

**Lihou Qin[1] · Shengwang Meng[2] · Guang Zhou[3] ·
Qijing Liu[1] · Zhenzhao Xu[1]**

Abstract Models of above-ground tree biomass have been widely used to estimate forest biomass using national forest inventory data. However, many sources of uncertainty affect above-ground biomass estimation and are challenging to assess. In this study, the uncertainties associated with the measurement error in independent variables (diameter at breast height, tree height), residual variability, variances of the parameter estimates, and the sampling variability of national inventory data are estimated for five above-ground biomass models. The results show sampling variability is the most significant source of uncertainty. The measurement error and residual variability have negligible effects on forests above-ground biomass estimations. Thus, a reduction in the uncertainty of the sampling variability has the greatest potential to decrease the overall uncertainty. The power model containing only the diameter at breast height has the smallest uncertainty. The findings of this study provide suggestions to achieve a trade-off between accuracy and cost for above-ground biomass estimation using field work.

Keywords Above-ground biomass · Measurement error · Residual variability · Parameter estimates · Sampling variability

✉ Qijing Liu
wdmtg666@sina.com

1   College of Forestry, Beijing Forestry University,
     Beijing 100083, People's Republic of China

2   Qianyanzhou Ecological Research Station, Key Laboratory
     of Ecosystem Network Observation and Modeling,
     Institute of Geographic Sciences and Natural Resources
     Research, Chinese Academy of Sciences, Beijing 100101,
     People's Republic of China

3   Jiangxi Academy of Forestry, Nanchang 330022,
     People's Republic of China

Introduction

Forests sequester and store large amounts of carbon and play an important role in the carbon cycle (Foley et al. 2005). Above-ground biomass (AGB), which is a crucial indicator of the carbon storage capacity of forest ecosystems (Bonan et al. 1992), is an important parameter for evaluating carbon sinks and analyzing matter and energy flow in forest ecosystems (Alongi et al. 2003). However, many sources of uncertainty can affect forest biomass prediction (Temesgen et al. 2015). One of the challenges confronting scientists is the estimation of uncertainties in forest biomass prediction (Wang et al. 2009).

The traditional method to estimate forest AGB from the individual tree scale to larger scales involves three steps (van Breugel et al. 2011): (1) models are used to predict the AGB of individual trees in forest inventory plots; (2) AGB at the plot level is estimated by summing the AGB of all trees; and (3) Forest AGB at larger spatial scales is estimated by averaging the AGB of all plots. Generally, above-ground biomass of individual trees can be obtained by harvesting trees or by predicting the AGB using models. The first method is costly, time-consuming and destructive and cannot be used at a large scale. In the second method, AGB models are established based on allometric relationships between tree AGB and other variables, such as diameter at breast height (DBH) and height (H). Many countries have established

forest inventory systems and have obtained National Forest Inventory (NFI) data at regular intervals. NFI data, in conjunction with AGB models, can be used to predict AGB of trees at the plot level and the data can be used to estimate AGB at larger scales. Therefore, this method has been widely used to estimate the forest above-ground biomass.

Traditional methods to estimate forest biomass have three types of uncertainties (Gertner 1990): uncertainty associated with measurement error, model-related uncertainty, and sampling-related uncertainty. Measurement error occur due to multiple factors, such as recording errors (in reading the tape or recording), tree form, instrument accuracy, measurement methods, measurement skills (McRoberts et al. 1994; Keller et al. 2001; Elzinga et al. 2005; Butt et al. 2013). Errors associated with NFI data and the calibration data set (CDS) can affect biomass estimation. Measurement error in CDS has a greater impact on AGB estimation (Qin et al. 2019).

Model-related uncertainty is associated with the precision of the model predictions that depend on response variables (such as volume, biomass), which are difficult to measure (Petersson et al. 2012). The accuracy of a biomass model is affected by the sample size of the calibration data, the regression methods and the type of model (Sileshi 2014). Model-related uncertainty is attributed to four primary sources: the values of the independent variables, the choice of the allometric model, residual variability, and model parameter estimates (McRoberts and Westfall 2013).

Sampling-related uncertainty occurs with large forest area biomass estimates when inventory data are used over a wide geographic area. Sampling-related uncertainty is affected by the plot size (Mauya et al. 2015), the sample size of the plots (Guo et al. 2016), and heterogeneity of the landscape (Salk et al. 2013). Studies have shown that increasing the sample size and the sample area can reduce the uncertainty caused by sampling variability (Mauya et al. 2015; Guo et al. 2016). Reducing the uncertainty in forest AGB estimation requires an understanding of uncertainty at different levels. Several researchers have predicted the effect of each different uncertainty sources on forest biomass estimation. Qin et al. (2019) quantified the effect of uncertainty due to measurement error and found that the measurement error in the CDS had a larger impact on AGB estimation. Wayson et al. (2015) used a pseudo-data approach to generate the potential distribution of the model parameters and found that this method could generate potential error structures that were used to propagate errors. McRoberts et al. (2014) used Monte Carlo simulation approaches to estimate the model uncertainty caused by residual variability and the model parameters, and found that Monte Carlo simulation approaches worked well in estimating the uncertainty in forest prediction. Fu et al. (2017) used a new method to predict sampling-related and model-related uncertainty; it was found that the model

uncertainty was larger than the sampling-related uncertainty. Molto et al. (2013) compared errors associated with model type and the independent variables at the tree and plot levels and observed that the largest source of uncertainty in AGB estimates was the biomass model. McRoberts et al (2016) estimated the uncertainty of six uncertainty sources. These studies indicate that uncertainty in forest AGB estimation has been widely investigated. However, uncertainties in AGB estimation require further in-depth study, particularly in different countries because the results from one country may not be applicable to other countries due to different measurement instruments, input variables, and methods (Berger et al. 2014). Relatively little information is available on NFI data in China. In addition, many biomass models can be used to describe the relationships between single-tree AGB and tree variables. It is necessary to determine which model type has the smallest uncertainty and which uncertainty type contributes the most to the overall error of the models.

In this study, we used Monte Carlo simulation (McRoberts et al. 2014, 2016) and the bootstrap resampling method to estimate uncertainty in AGB estimation using Chinese NFI data. The uncertainties associated with the measurement error of the variables (DBH, H), the residual variability, the variance in the model parameter estimates, and the sampling variability of the NFI data are estimated for five models. The objectives are: (1) to determine which of the four sources of uncertainty has the largest influence on the precision of AGB estimation for the different models; and, (2) to evaluate the performance of these five models for large-area AGB estimation when considering four sources of uncertainty.

Material and methods

Study area

This study was carried out in Longquan County located in eastern of China, southwestern of Zhejiang Province (Fig. 1) and covers an area of 3059 ha. It is located in the subtropical zone with a warm and humid climate. The average annual temperature is 17.6 °C, and annual precipitation is 1645.4 mm. Mountainous and hilly areas dominate the county and cover for more than 97% of the total area. Forest types include coniferous forests, evergreen broad-leaf forests, and deciduous and evergreen broad-leaf mixed forests.

National forest inventory data

NFI data collected in 2009 was used to estimate the mean AGB per hectare. The data set contained 101 permanent plots (9886 sample trees) systematically distributed using a
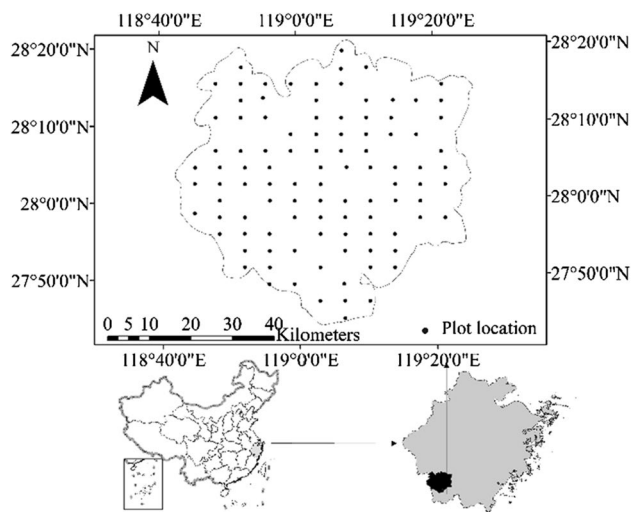
4 km × 6 km grid (Fig. 1). The plot size was 28.3 m × 28.3 m (800 m^2). In each plot, trees with a DBH ˃ 5 cm were measured. The frequency histogram of the DBH of sample trees at 2 cm intervals is shown in Fig. 2. Because measurements of H are tedious and expensive, the NFI data does not contain information on height. In this study, the height-diameter model developed by Li and Fa (2011) was used to predict the H of *Quercus*, *Cunninghamia lanceolate* (Lamb.) Hook and *Pinus massoniana* Lamb. The calibration data were obtained from the NFI data, which were collected throughout China. Height was divided into nine levels according to different sites, and used the Chapman–Richard function was used to establish the height–diameter model. This method provides more accuracy of height prediction and can be applied throughout China. For the other broad-leaved species, we use the model of height developed by Shen (2002). This
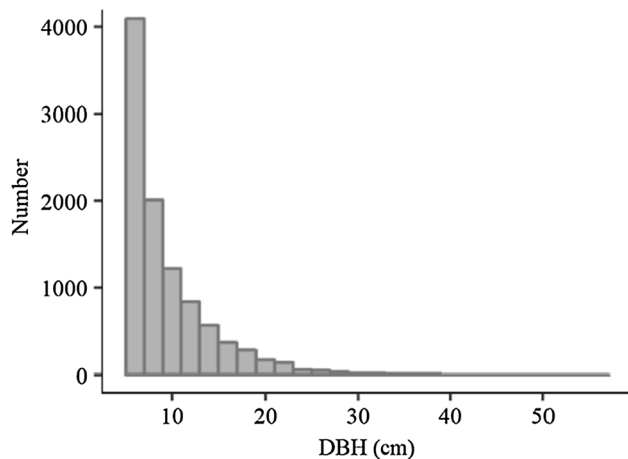
model was based on 1356 sample trees and used the Chapman–Richard function. The data set contained more than 10 common subtropical broad-leaved tree species and included a range of possible variations in average stand diameter and average height. Although, our study area was different from that of Shen (2002). But both areas are located in the mid-subtropics and are dominated by mountainous and hilly terrain. This model was used to predict the H of broad-leaved species.

Calibration data

A total of 363 harvested trees were used to develop the AGB model. This dataset was collected from 13 counties across a large geographical area in Zhejiang Province and contained 21 common tree species. In these sites, 363 NFI plots were selected (146 *Cunninghamia lanceolate* plots, 80 *Pinus massoniana* plots, 173 broad-leaved tree plots) and one average bole was selected in each plot. According to the average boles, the harvested trees were selected outside the plots. Figure 3 shows the frequency histogram of each diameter class and the relationship between the AGB of harvest trees and DBH.

After sample trees were felled at ground level, the tree species, DBH and H were recorded. The stems were cut at 2-m intervals if the H ≥ 10 m, and cut at 1-m intervals if the H < 10 m. Branches and leaves were divided into four levels: ≤ 1 cm, 1.1–2.0 cm, 2.1–3.0 cm and > 3 cm. One standard branch was selected from each level, and weight of the branch and foliage were determined. Fresh weights of bole, bark, branches, and foliage were measured in the field. The samples of each component were oven-dried until their weight stabilized. The ratio of dry mass to fresh mass was used to predict the biomass of each component, and the total AGB of each
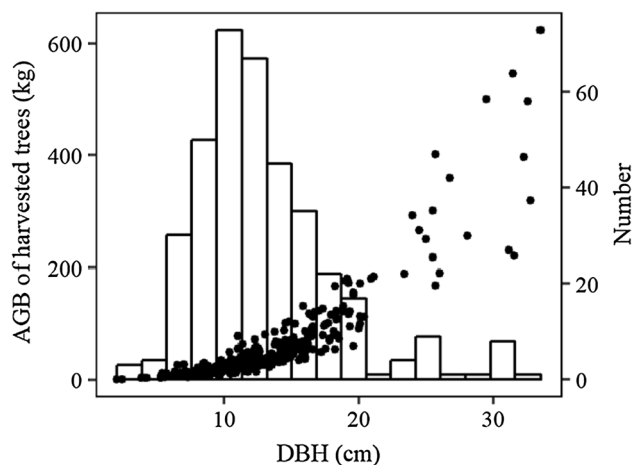

Fig. 2 Frequency histogram of each diameter class for the National Forest Inventory data

Fig. 3 Frequency histogram of each diameter class and the relationship between DBH and the AGB of harvested trees for the calibration data

harvested tree was estimated by summing all components. As shown in Figs. 2 and 3, the maximum DBH of the CDS is smaller than that of the NFI data. However, only 39 trees in the NFI data are outside of the DBH range of the CDS. Thus, the calibration data are representative the NFI data.

Measurement error data

We performed a double-blind re-measurement of 276 tree samples to estimate the measurement error. The sample size of each species group is determined according to the NFI data, and confirmed that DBH distribution was similar to the DBH distribution of the NFI data. The detailed description of this data set is available in Qin et al. (2019). According to the regulations for continuous NFI in China, the measurement error of H should not exceed 5%. In this study, it was assumed that the measurement error was followed a uniform distribution $\varepsilon_H - \mu(-0.05H, 0.05H)$; the measurement error was randomly selected from this range. Picard et al. (2015) assumed that the measurement error of DBH followed a uniform distribution due to the lack of measurement error data.

Mean AGB estimation

The calibration data was used to establish individual tree AGB allometric models and were used to predict the individual tree AGB in each plot. The tree-level AGB predictions were aggregated to produce plot-level AGB predictions. Finally, systematic sampling estimators were used with the plot-level AGB predictions to forecast the mean above-ground biomass per unit area. This procedure is referred to as hybrid inference, a term was coined by Corona et al. (2014) to describe the use of a model to predict the response variables of probability samples of auxiliary data. The population parameters (mean, total) were estimated using a probability-based estimator and the probability sample predictions (McRoberts et al. 2016; Ståhl et al. 2016).

Estimation of single-tree above-ground biomass

Many forms of allometric models can be used for developing tree AGB models. In this study, five model types were selected (Wang et al. 2015; Cecep et al. 2018; Martínez-Sánchez et al. 2020):

$$M = \alpha_1 D^{\alpha_2} + \varepsilon \tag{1}$$

$$M = \alpha_1 + \alpha_2 D + \alpha_3 D^2 + \varepsilon \tag{2}$$

$$M = \alpha_1 + \alpha_2 D + \alpha_3 D^2 + \alpha_4 D^3 + \varepsilon \tag{3}$$

$$M = \alpha_1 D^{\alpha_2} H^{\alpha_3} + \varepsilon \tag{4}$$

$$M = \alpha_1 (D^2 H)^{\alpha_2} + \varepsilon \tag{5}$$

where D is the DBH (cm), H is H (m); M is the individual tree AGB (kg), α_1, α_2, α_3 and α_4 are estimated parameters. The model performances were evaluated with the coefficient of determination (R^2).

Uncertainty caused by measurement error

The measurement error is assumed to follow a Gaussian distribution (Chave et al. 2004; Berger et al. 2014; McRoberts and Westfall 2016). We fitted a model to the SD (standard deviation) of measurement error and DBH to simulate the measurement error.

The method consisted of the following steps (Hosmer and Lemeshow 1989; Berger et al. 2014; Qin et al. 2019): (1) All trees were ranked in ascending order with respect to $\overline{D}$, where $\overline{D}$ is the mean of two DBH measurement values, and the difference (D_{Dif}) between two measurements for each tree was calculated as: $D_{Dif} = D_1 - D_2$, where D_1 and D_2 represent the first and second DBH measurement, respectively; (2) After ranking, all trees were divided into n groups according to the new order. Each group contained at least 15 trees to generate a sufficient number of groups. If the last group contained less than 15 trees, the trees in the last group were placed in the previous group; and, (3) for the ith group, the mean DBH ($\overline{D}_i$) of $\overline{D}$ and the SD (σ_{Dif}) of the differences between the two measurements were calculated. Finally, the relationship between the $\overline{D}_i$ and σ_{Dif} was described by a liner model:

$$\sigma_{Dif} = a + b\overline{D}_i \tag{6}$$

Based on this model, Monte Carlo simulations were used to predict the effect of the measurement error on AGB estimation. This procedure contained four steps:

Step 1 Using Eq. (6), the SD (σ_{Dij}) of the measurement error of DBH for the ith tree in the jth plot was predicted. For each tree in a plot, a measurement error ε_{Dij} was randomly selected from a Gaussian distribution $\varepsilon_{Dij} \sim N(0, \sigma_{Dij})$. The new DBH for each tree was then calculated by adding ε_{Dij} to the original DBH.

Step 2 If the individual tree AGB model contained H, the measurement error of H for the ith tree in the jth plot was randomly selected from the uniform distribution $\varepsilon_{Hij} - \mu(-0.05H_{i,j}, 0.05H_{i,j})$, where $H_{i,j}$ is the original H for the ith tree in the jth plot. The new H ($H'_{i,j}$) for each tree was then calculated by adding ε_{Hij} to the original H.

Step 3 For ith tree in the jth plot, the individual tree AGB was predicted using the new DBH and H simulated from step 1 and step 2. The total AGB of the jth plot was predicted by

summing all individual tree AGB: $P_j = \sum_i^n M_{i,j}$, where n is the number of trees in the jth plot and $M_{i,j}$ is the AGB of ith tree in the jth plot. The mean AGB per hectare was estimated as:

$$\overline{P} = \frac{1}{m} \sum_{j=1}^m p_j \qquad (7)$$

where, m is the number of the plot.

Step 4 Steps 1–3 were repeated 2000 times. The mean of AGB per hectare and the variance of the replications were predicted following Rubin (1987):

$$\overline{\mu}_{me} = \frac{1}{n_{rep}} \sum_{k=1}^{n_{rep}} \overline{p}^k \qquad (8)$$

$$Var(\overline{\mu}_{me}) = \left(1 + \frac{1}{n_{rep}}\right) \times W_1 + W_2 \qquad (9)$$

where, $\overline{\mu}_{me}$ is the mean AGB over replications, $W_1 = \frac{1}{n_{rep}-1} \sum_{k=1}^{n_{rep}} (\overline{\mu} - \overline{P}^k)^2$ is the between-simulation variance, $W_2 = \frac{1}{n_{rep}-1} Var(\overline{P}^k)$ is the mean within-simulation variance, $\overline{p}^k$ is the mean AGB per hectare in the kth replication and n_{rep} is the number of replications. The replications were continued until μ_{me} and $Var(\mu_{me})$ stabilized. The relative uncertainty $R_{me} = \frac{\sqrt{Var(\overline{\mu}_{me})}}{\overline{\mu}_{me}} \times 100\%$ was then used as the index to assess the uncertainty.

Uncertainty caused by residual variability

This was also estimated using Monte-Carlo simulations by adding a random error to each models' predictions (McRoberts et al. 2014). In this study, we first constructed a model from the standard deviation of the residual and the predicted AGB. The fitting method was the same as fitting models to the SD of the measurement error and DBH. McRoberts and Westfall (2013, 2016) used this method to develop models from the calibration data of tree volume and residual variability. The method had the following steps: (1) ε, $\widehat{M}$ were ranked in ascending order with respect to $\widehat{M}$, where ε is the residual of each harvested tree, $\widehat{M}$ is the model prediction of single tree AGB; (2) All trees in the CDS were divided into n groups after ranking. In this study, 20 trees were placed in a single group. If the last group contained less than 20 trees, the trees in the last group were placed in the previous group; (3) For the ith group, the mean of the $\widehat{M}$ and the SD of the ε was calculated. The relationships between the SD and the mean of the $\widehat{M}$ were then estimated using the liner model:

$$\sigma_\varepsilon = a_\varepsilon + b_\varepsilon \overline{\widehat{M}} \qquad (10)$$

where σ_ε is the SD of the ε in each group, a_ε and b_ε are the model parameters, and $\overline{\widehat{M}}$ is the mean of $\widehat{M}$ in each group. Based on this model, a four-step procedure was used to simulate the effect of residual variability on AGB estimation.

Step 1 For the ith sample tree in the jth plot, the tree AGB was predicted using the AGB model established from the CDS.

Step 2 For the ith tree in the jth plot, the SD (σ_ε) of the residual was predicted by model (10). Based on σ_ε, a random residual ($\varepsilon_{i,j}$) of each sample tree was randomly obtained from the Gaussian distribution $\varepsilon_{i,j} = N(0, \sigma_{\varepsilon ij})$, where $\sigma_{\varepsilon ij}$ is the SD of the residual of the ith tree within jth plot. Then, the new single-tree AGB value of each element in the NFI data was calculated as $M'_{i,j} = \widehat{M}_{i,j} + \varepsilon_{i,j}$, where $M'_{i,j}$ is the new AGB of ith tree in the jth plot, and $M_{i,j}$ is the original predicted value of the single tree AGB.

Step 3 The total AGB of the jth plot was estimated as $P_j = \sum_i^n M'_{i,j}$, where n is the number of trees in the plot, and $M'_{i,j}$ is the ith individual tree AGB in the jth plot. The mean of all plots was estimated as $\overline{P} = \frac{1}{m} \sum_{j=1}^m P_j$ where m is the number of the plot.

Step 4 Steps 1–3 were repeated 2000 times.

The relative uncertainty caused by the residual variability was estimated as $R_{U(rv)} = \frac{\sqrt{Var(\overline{\mu}_{rv})}}{\overline{\mu}_{rv}} \times 100\%$, where $\overline{\mu}_{rv}$ and $Var(\overline{\mu}_{rv})$ are the mean and variance of the replications, respectively; these were estimated as Eqs. (8) and (9).

Uncertainty caused by the variance of the parameter estimates

Monte-Carlo simulations can be used to simulate the variances of the parameters of the AGB model. This simulates the actual sampling process (if known) of the original data. Wayson et al. (2015) used this method to generate a large pseudo-data of variables and refitted the biomass model. McRoberts et al. (2014) also used this method to simulate the potential distribution of the parameters in the biomass model and predicted the uncertainty caused by the variance of the parameter estimates. In this study, a five-step Monte-Carlo simulation was used to estimate the uncertainty caused by model parameters:

Step 1 The original CDS was grouped into DBH classes at 1-cm intervals. If the sample size in one group was insufficient, then the sample trees were placed in the previous group. Finally, each group contained at least nine sample units.

Step 2 In each group, the AGB, DBH, and H of each harvest tree were randomly resampled until the original class size was achieved. During resampling, AGB, DBH and H were assumed to be a uniform distribution and each variable was randomly selected from the existing range.

Step 3 The new data set generated in Step 2 were used to fit a new single-tree AGB model for each model type and predict the single-tree AGB in each plot.

Step 4 The total AGB of the *j*th plot was predicted as $P_j = \sum_i^n \hat{M}_{i,j}$, where *n* is the number of trees in the plot, and $\hat{M}_{i,j}$ is the *i*th predicted individual tree biomass in the *j*th plot. The mean AGB of all plots was estimated as: $\overline{P} = \frac{1}{m}\sum_i^m P_j$, where *m* is the number of the plot.

Step 5 Steps 1–4 were repeated 2000 times. The relative uncertainty associated with the parameter estimates was estimated as $R_{par} = \frac{\sqrt{Var(\overline{\mu}_{par})}}{\overline{\mu}_{par}} \times 100\%$, where $\overline{\mu}_{par}$ and $Var(\overline{\mu}_{par})$ are the mean and variance of the replications; these were estimated using Eqs. (8)–(9).

Uncertainty caused by the sampling variability of the NFI data

In this study, bootstrap resampling was used to estimate the uncertainty associated with the sampling variability of the NFI data. This method assumed that the sample represented the population from which the sample was drawn, and each observation in the sample had an equal probability of being selected from the population (Hinkley 1988). This method can be used to estimated standard errors, confidence intervals, and variances of a survey data (Efron and Tibshirani 1986). In this study, bootstrap sampling was performed using the "bootstrap" package in the R software; the bootstrap times were set to 2000. After 2000 bootstraps times, the mean above-ground biomass and the variance of mean were estimated as:

$$\overline{\mu}_{sam} = \frac{\sum_i^n \overline{p}_{sam}^k}{n} \tag{11}$$

$$Var_{sam} = \frac{1}{n-1}\sum_{k=1}^n (\overline{\mu}_{sam} - \overline{P}_{sam}^k)^2 \tag{12}$$

where $\overline{\mu}_{sam}$ is the mean AGB after 2000 resampling steps, $\overline{p}_{sam}^k$ the mean AGB of *k*th resampling step, Var_{sam} the variance of the mean AGB after 2000 resampling steps, and *n* the number of resampling steps. The relative uncertainty associated with the sampling variability of NFI data was estimated $R_{sam} = \frac{\sqrt{Var(\overline{\mu}_{sam})}}{\overline{\mu}_{sam}} \times 100\%$.

Calculation of overall uncertainty

The overall relative uncertainty (R_T) of each model was predicted as:

$$R_T = \sqrt{\left(R_{me}\right)^2 + \left(R_{rv}\right)^2 + \left(R_{par}\right)^2 + \left(R_{sam}\right)^2} \tag{13}$$

We assumed the uncertainty between each other was independent. McRoberts and Westfall (2013) found that this

assumption is reasonable for predicting tree volume over large areas. In addition, Berger et al. (2014) used the law of error propagation and a Monte Carlo simulation to estimate the uncertainty associated with measurement error for tree volume estimations; they found that the results of the two methods were similar.

Results

Model fitting

The parameters and R^2 of the models are listed in Table 1, and the relationships between the predicted AGB and the AGB of harvested trees are shown in Fig. 4. The fits of all models to the data resulted in large R^2 values ($R^2 > 0.83$), and the model that contained DBH and H produced better fitting results than the one with only DBH.

Uncertainty calculation

The model used to predict the standard deviation of the measurement error had the form $\hat{\sigma}_D = 0.0173D + 0.0185$, with an R^2 of 0.61 (Qin et al. 2019). Figure 5 shows the mean AGB per hectare and the relative uncertainty for the five models. Both the mean AGB and the relative uncertainty tended to stabilize after 500 simulations. Table 2 shows the sources of uncertainties of the five models. The measurement error had the largest effect on the biomass estimation for model 4 and the smallest for models 1 and model 3. Although the results of the different models were affected to various degrees by the measurement error, the uncertainty related to the measurement error was negligible.

Figure 6 shows the relationships between the standard deviation of the residual and the predicted tree above-ground biomass of each group. The fit of the model describing the relationship between the estimated tree AGB and the SD of the residual was excellent and all models had very large R^2 values. Figure 7 shows the mean AGB per hectare and the relative uncertainty of all models for 2000 simulations. Both the mean AGB and the relative uncertainty tended to stabilize after 300 simulations. As shown in Table 2, the residual variability had the largest influence on the biomass

Table 1 Estimated parameters of the models

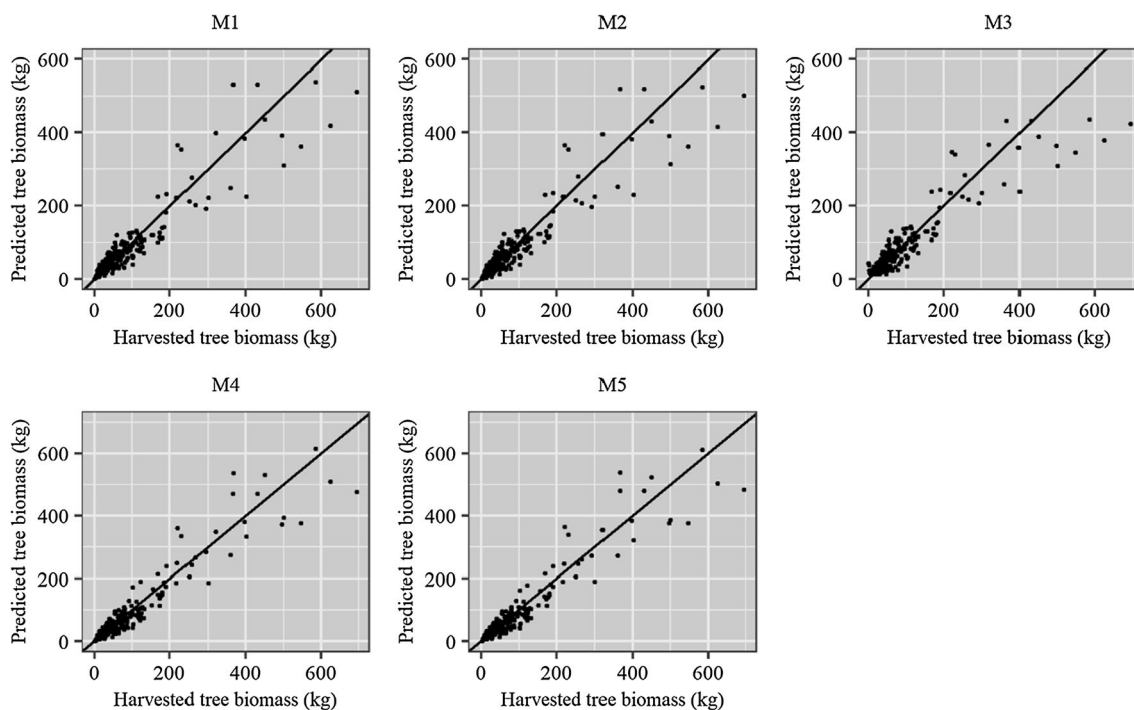| Model form	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	R^2
$M = \alpha_1 D^{\alpha_2}$	0.11	2.35	/	/	0.83
$M = \alpha_1 + \alpha_2 D + \alpha_3 D^2$	6.52	−3.01	0.45	/	0.85
$M = \alpha_1 + \alpha_2 D + \alpha_3 D^2 + \alpha_4 D^3$	71.55	−17.11	1.35	−0.02	0.84
$M = \alpha_1 D^{\alpha_2} H^{\alpha_3}$	0.05	1.70	1.00	/	0.88
$M = \alpha_1 (D^2 H)^{\alpha_2}$	0.06	0.89	/	/	0.88

Fig. 4 Relationships between the predicted AGB and the AGB of the harvested trees for different models M1–M5 are the biomass models defined in Eqs. (1)–(5)
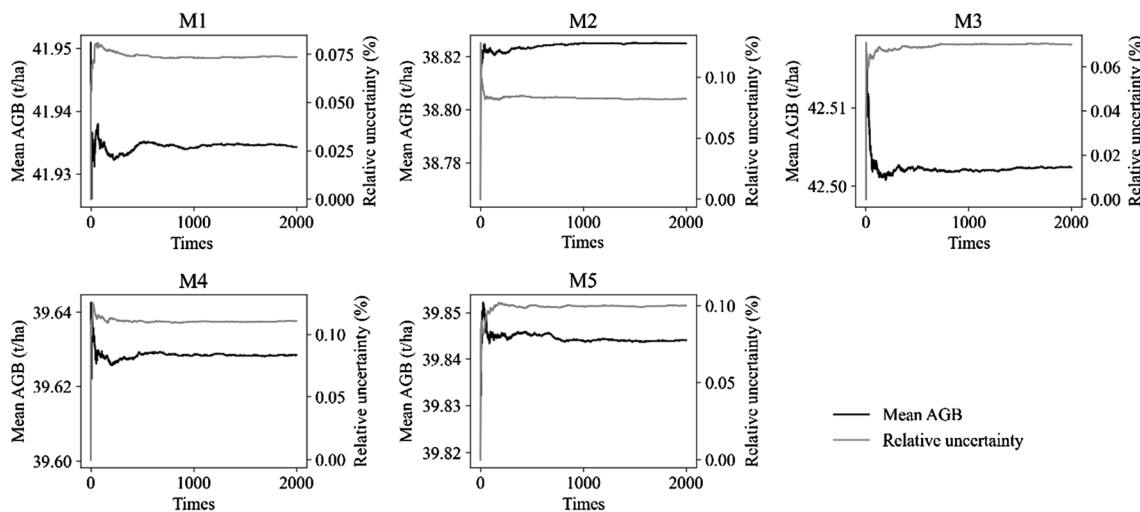

Fig. 5 Simulated mean AGB and relative uncertainty associated with the measurement error for the five models M1–M5 described by Eqs. (1)–(5)

estimation for model 1, whereas model 5 was least affected by the residual variability. The uncertainty associated with residual variability was larger than that of the measurement error but did not exceed 1.2%. Therefore, the residual variability also had a slight effect on forest biomass estimation.

The uncertainty associated with the variance of the parameter estimates (Table 2; Fig. 8) was larger than that of

the measurement error and residual variability and ranged from 3.9% (model 1) to 11.1% (model 2) for the five models.

Figure 9 shows the frequency histogram of the estimated mean above-ground biomass per unit area for the five models after 2000 bootstraps. Model 2 had the largest sampling uncertainty (10.9%), and model 3 had the smallest sampling uncertainty (9.7%). Sampling variability was the largest

Table 2 Uncertainties associated with various sources for different models

Model form	Uncertainty sources	Mean (t ha^{-1})	Relative uncertainty (%)	Total uncertainty (%)
M1	ME	41.9	0.1	11. 3
	RV	41.3	1.2	
	PU	41.9	3.9	
	SAM	41.3	10.5	
M2	ME	38.8	0.1	15. 6
	RV	38.8	1.2	
	PU	39.1	11.1	
	SAM	38.7	10.9	
M3	ME	42.5	0.1	11. 4
	RV	42.7	1.0	
	PU	42.6	5.8	
	SAM	42.9	9.7	
M4	ME	39.6	0.1	11.5
	RV	39.6	0.9	
	PU	40.7	5.2	
	SAM	39.4	10.3	
M5	ME	39.8	0.1	11. 6
	RV	39.6	0.9	
	PU	38.8	5.2	
	SAM	39.5	10.3	

ME, RV, PU and SAM represent the measurement error, residual variability, parameter estimates and sampling error respectively; M1–M5 are the 5 selected models described by Eqs. (1)–(5)

source of uncertainty for all models (Table 2). For total uncertainty, model 1 had the smallest uncertainty (11.3%), and model 2 had the largest uncertainty (15.6%) (Table 2).

Discussion

In recent years, increasing attention has focused on the assessment of uncertainty in forest biomass estimation. Several studies have quantified different types of uncertainty, such as measurement uncertainty, model uncertainty, and sampling uncertainty (Chave et al. 2004; Picard et al. 2015; Shettles et al. 2015). In this study, the uncertainty associated with the measurement error, residual variability, variance of the parameter estimates and sampling variability in NFI data for different models was estimated. The results showed that the sampling variability in the NFI data was the primary source of uncertainty. In addition, the uncertainty caused by the parameter estimates should not be overlooked.

The measurement error may be minimized by training, but is not completely avoidable (Elzinga et al. 2005). In this study, uncertainty associated with the measurement error was negligible, regardless of whether the model contained diameter or diameter and height. This is consistent with that of other studies (Berger et al. 2014; McRoberts and Westfall, 2016). In this study, we assumed that the measurement error were independent, which minimized the total error for a large number of trees, because errors will compensate each
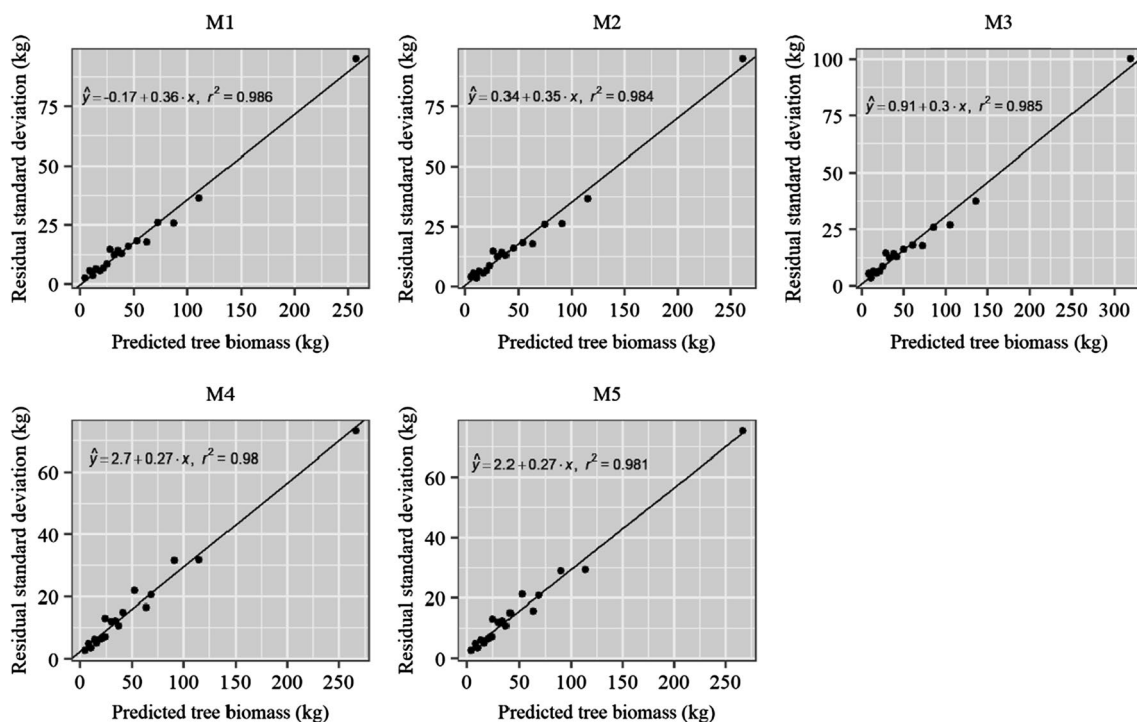

Fig. 6 Relationships between the estimated tree biomass and the SD of the residual for the 5 selected models M1–M5 described by Eqs. (1)–(5)
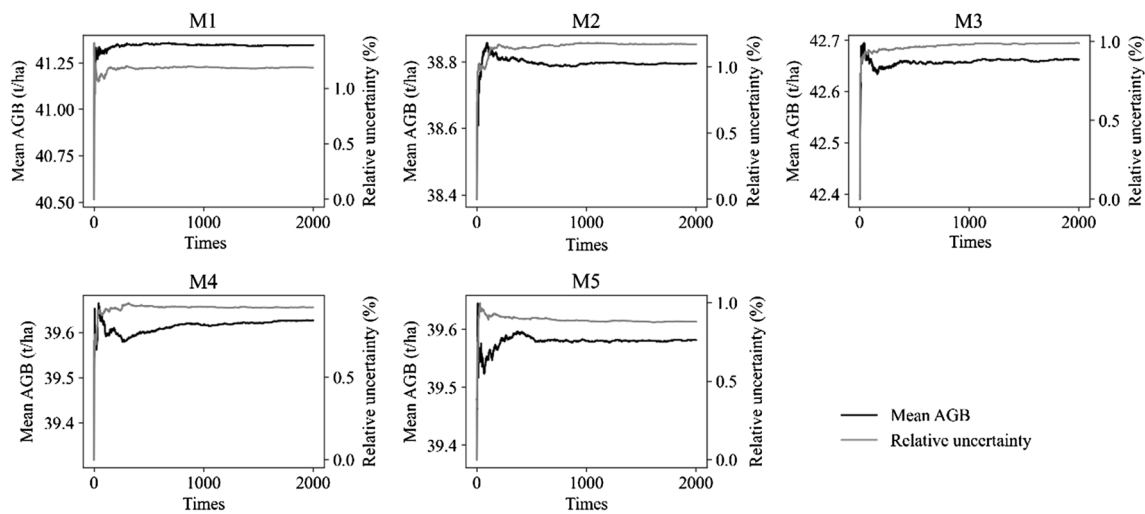
Fig. 7 Simulated mean AGB and relative uncertainty associated with the residual variability of the 5 models M1–M5 described by Eqs. (1)–(5)
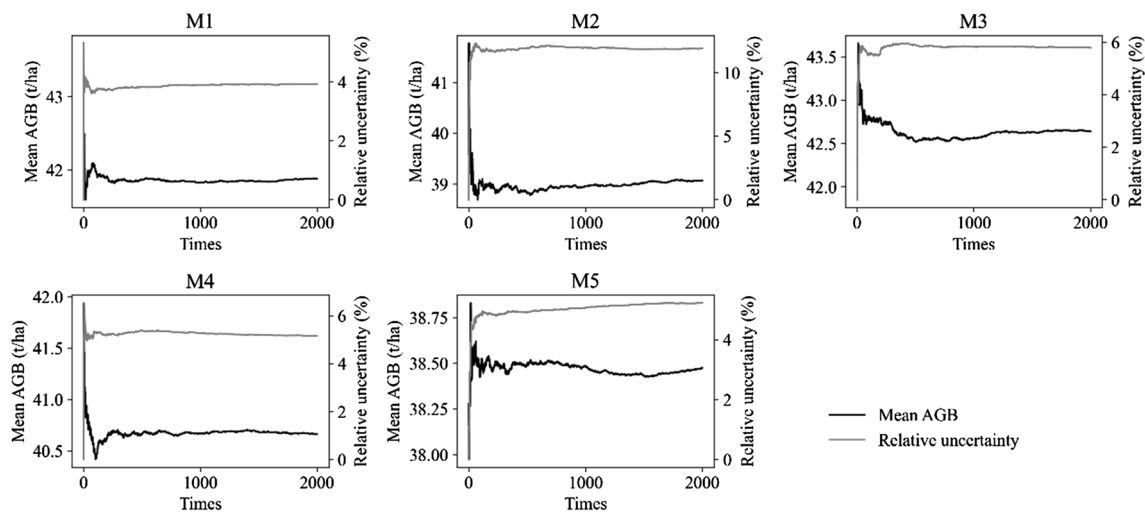

Fig. 8 Simulated mean AGB and relative uncertainty associated with the parameter estimates for 5 models M1–M5 described by Eqs. (1)–(5)

other. Dependent errors were not considered, for example, a systematic bias due to faulty material. The uncertainty caused by this source of error needs further examination. In addition, the predicted H was taken as the "true" H. This may produce other uncertaintes, but the referenced studies did not show the distribution of residual and this was not considered.

In this study, the influence of residual variabilities on the results was insignificant. This is consistent with studies by Chen et al. (2015) and Picard et al. (2015). Chave et al. (2004) and Chen et al. (2015) found that the error of single-tree above-ground biomass estimation caused by the residual varibility was more than 30%. This type of uncertainty at a regional scale was much samller than that at tree level. One reason may be that a normal distribution of the residuals was

assumed. This residual error at tree level is levelled off when residuals are randomly selested. Trees with positive residuals compensate for trees with negative residuals, resulting in a small variance between simulations. In addition, Chave et al. (2004) and Chen et al. (2015) and found that the uncertainty at the plot level was in negatively correlated with the number of trees in a plot. In this study, the average number of trees in all plots was 99, resulting in relatively low uncertainty at the plot level.

We also found that the model containing diameter at breast height and height resulted in smaller uncertainties associated with residual variability, both at the single tree level (Fig. 5) and at the regional level (Table 2). The inclusion of additional variables in biomass models typically improves model accuracy (Ketterings et al. 2001), which has
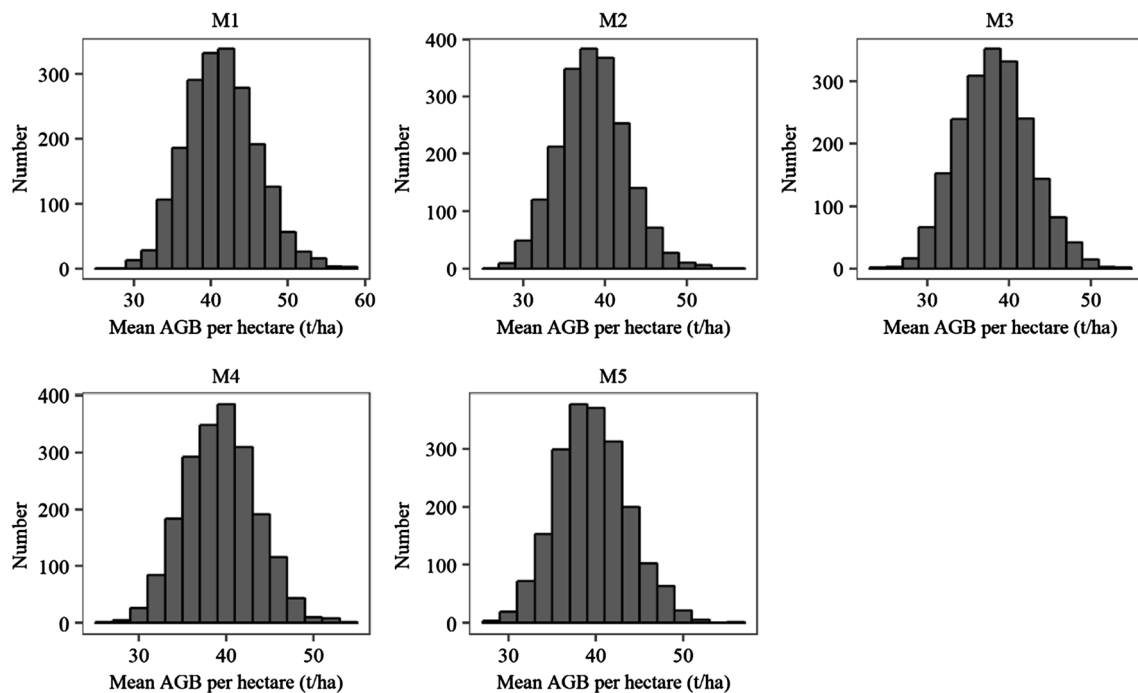
Fig. 9 Frequency histogram of the estimated mean AGB per unit area of the 5 models after 2000 bootstraps; the 5 models were described by Eqs. (1)–(5)

been confirmed in many studies. Lambert et al. (2005) found that the root mean squared error of tree biomass predictions was reduced by approximately 8% and 25% for hardwood and softwood species, respectively, if height was added to the model. Goodman et al. (2014) found that the addition of crown radius to biomass models further reduced the bias of total above-ground biomass by 11–14%. In addition, additional species traits (e.g., wood density) could be included, especially when a significant species effect on model residuals is found (Ngomanda et al. 2014).

The uncertainty caused by the variance of the model parameter was much larger than that caused by the measurement error and the residual variability. This is consistent with other studies (Breidenbach et al. 2014). Therefore, the greatest potential to improve the accuracy of biomass models is to improve the accuracy of the model parameters. These are affected by the sample size, distribution of the calibration data, and regression methods (Muller-Landau et al. 2006). Increasing the sample size can improve the accuracy of biomass models. Chen et al. (2015) reduced the sample size from 4004 to 400, and to 40, and found that the relative above-ground biomass prediction error increased from 0.7 to 2.5 and 5.7%, respectively. However, single-tree above-ground biomass measurements are difficult to obtained resulting in the limitation of the size of the calibration data set. Jenkins et al. (2004) determined the mean sample size of 2642 biomass models and found that sample sizes for nearly half of

the studies did not exceed 20 trees. Therefore, increasing the sample size appears to be a challenge in most studies. In this study, 363 sample trees were used to establish AGB models to predict subtropical forest AGB in China. In addition, different regression methods may also provide widely different estimates of the allometric parameters from the same dataset (Sileshi 2014). This aspect was not covered in this study.

As shown in Table 2, the uncertainty associated with the sampling variability of NFI data was the largest source of uncertainty. This is consistent with Berger et al. (2014) and Breidenbach et al. (2014). One approach to reducing this type of uncertainty is to increase the sample size of the plot or to increase the size of the plot (Peter and Tom 2011). Studies have found that the coefficient of variation of biomass decreased with an increase in size of plot (Chave et al. 2003). However, an increase in the number of plots or plot size may result in different spatial patterns of the biomass. A random plot distribution is more easily achieved for small plots than for a few large plots, whereas the opposite is true for a systematic distribution (Picard et al. 2015). In addition, uncertainty associated with residual variability and plot-model interaction decrease with an increase in the plot size (Picard et al. 2015). Another method to reduce sampling-related uncertainty may be the inclusion of ancillary information for stratification, such as forest characteristics or the forest structure. This method has been proved to be more accurate and efficient than systematic or random sampling

(Gharun et al. 2017), especially when remotely sensed data are used (Wallner et al. 2018).

The balance between accuracy and cost is a problem in forest above-ground biomass estimations. One method to improve the precision is to add other variables to the models, but this increases the costs of model development and field inventory. For example, accurate height measurements are time-consuming and expensive, and increase the overall cost of the forest inventory. In this study, we found that model 1 had the least uncertainties, especially when sampling variability was not considered. Our results suggest trade-off between accuracy and cost. However, different models provide different results, as shown in Table 2. One of the problems to address is how to balance the results between different models.

Conclusion

This study estimated four types of uncertainties in five forest above-ground biomass models. The results suggested that the model $M = \alpha_1 D^{\alpha_2}$ was the best to estimate above-ground biomass. This finding can be used to achieve a trade-off between accuracy and cost in biomass estimation. The results also indicate that uncertainty associated with sampling variability in national forest inventory data contributed most to overall uncertainty, followed by the uncertainty associated with the variance of the parameter estimates and the residual variability. Thus, the emphasis is on reducing the sampling-related variability if the objective is the reduction in overall uncertainty of above-ground biomass estimation. If the model-related uncertainty is to be decreased, the focus should be on reducing the uncertainty associated with the variance of the parameter estimates.

References

Alongi DM, Clough BF, Dixon P, Tirendi F (2003) Nutrient partitioning and storage in arid-zone forests of the mangroves *Rhizophora stylosa* and *Avicennia marina*. Trees 17(1):51–60. https://doi.org/10.1007/s00468-002-0206-2

Berger A, Gschwantner T, McRoberts RE, Schadauer K (2014) Effects of measurement errors on individual tree stem volume estimates for the Austrian National Forest Inventory. For Sci 60(1):14–24. https://doi.org/10.5849/forsci.12-164

Bonan GB, Pollard D, Thompson SL (1992) Effects of boreal forest vegetation on global climate. Nature 359(6397):716–718. https://doi.org/10.1038/359716a0

Breidenbach J, Antón-Fernández C, Petersson H, McRoberts RE, Astrup R (2014) Quantifying the model-related variability of biomass stock and change estimates in the Norwegian National Forest Inventory. For Sci 60(1):25–33. https://doi.org/10.5849/forsci.12-137

Butt N, Slade E, Thompson J, Malhi Y, Riutta T (2013) Quantifying the sampling error in tree census measurements by volunteers and its effect on carbon stock estimates. Ecol Appl 23(4):936–943. https://doi.org/10.1890/11-2059.1

Cecep K, Topik H, Tatang T, Omo R, Istomo. (2018) Allometric models for above- and below-ground biomass of sonneratia spp. Glob Ecol Conserv 15:e00417

Chave J, Condit R, Lao S, Caspersen JP, Foster RB, Hubbell SP (2003) Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. J Ecol 91:240–252. https://doi.org/10.1046/j.1365-2745.2003.00757.x

Chave J, Condit R, Aguilar S, Hernandez A, Lao S, Perez R (2004) Error propagation and scaling for tropical forest biomass estimates. Philos Trans R Soc Lond Ser B Biol Sci 359(1443):409–420. https://doi.org/10.1098/rstb.2003.1425

Chen Q, Vaglio Laurin G, Valentini R (2015) Uncertainty of remotely sensed aboveground biomass over an African tropical forest: propagating errors from trees to plots to pixels. Remote Sens Environ 160:134–143. https://doi.org/10.1016/j.rse.2015.01.009

Corona P, Fattorini L, Franceschi S, Scrinzi G, Torresan C (2014) Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. Can J For Res 44:1303–1311

Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1(1):54–75

Elzinga C, Shearer RC, Elzinga G (2005) Observer variation in tree diameter measurements. West J Appl For 20(2):134–137. https://doi.org/10.1093/wjaf/20.2.134

Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, Chapin FS, Coe MT, Daily GC, Gibbs HK, Helkowski JH, Holloway T, Howard EA, Kucharik CJ, Monfreda C, Patz JA, Prentice IC, Ramankutty N, Snyder PK (2005) Global consequences of land use. Science 309(5734):570–574. https://doi.org/10.1126/science.1111772

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. Can J For Res 47(8):1095–1103. https://doi.org/10.1139/cjfr-2016-0436

Gertner GZ (1990) The sensitivity of measurement error in stand volume estimation. Can J For Res 20(6):800–804. https://doi.org/10.1139/x90-105

Gharun M, Possell M, Jenkins ME, Poon LF, Bell TL, Adams MA (2017) Improving forest sampling strategies for assessment of fuel reduction burning. For Ecol Manag 392:78–89

Goodman RC, Phillips OL, Baker TR (2014) The importance of crown dimensions to improve tropical tree biomass estimates. Ecol Appl 24(4):680–698. https://doi.org/10.1890/13-0070.1

Guo H, Zhang M, Xu L, Yuan Z, Qin L, Chen T (2016) Simulation of regional forest carbon storage under different sampling densities. Acta Ecol Sin 36(14):4373–4385

Hinkley DV (1988) Bootstrap methods. J R Stat Soc 50(3):321–337

Hosmer DW, Lemeshow S (1989) Applied logistic regression. Wiley, New York, p 307

Jenkins JC, Chojancky DC, Heath LS, Birdsey RA (2004) Comprehensive database of diameter-based biomass regressions for North American tree species. U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA

Keller M, Palace M, Hurtt G (2001) Biomass estimation in the Tapajos National Forest, Brazil: examination of sampling and allometric uncertainties. For Ecol Manag 154(3):371–382. https://doi.org/10.1016/S0378-1127(01)00509-6

Ketterings QM, Coe R, van Noordwijk M, Ambagau' Y, Palm CA. (2001) Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. For Ecol Manag 146(1):199–209. https://doi.org/10.1016/S0378-1127(00)00460-6

Lambert MC, Ung CH, Raulier F (2005) Canadian national tree above-ground biomass equations. Can J For Res 35(8):1996–2018. https://doi.org/10.1139/x05-112

Li H, Fa L (2011) Height-diameter model for major tree species in China using the classified height method. Sci Silvae Sin 47(10):83–90. https://doi.org/10.11707/j.1001-7488.20111013

Martínez-Sánchez JL, Martínez-Garza C, Cámara L, Castillo O (2020) Species-specific or generic allometric equations: which option is better when estimating the biomass of Mexican tropical humid forests? Carbon Manag 11(3):241–249

Mauya EW, Hansen EH, Gobakken T, Bollandsås OM, Malimbwi RE, Næsset E (2015) Effects of field plot size on prediction accuracy of aboveground biomass in airborne laser scanning-assisted inventories in tropical rain forests of Tanzania. Carbon Balance Manag 10(1):10. https://doi.org/10.1186/s13021-015-0021-x

McRoberts RE, Westfall JA (2013) Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. For Sci 60(1):34–42. https://doi.org/10.5849/forsci.12-141

McRoberts RE, Westfall JA (2016) Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. Ann For Sci 73(3):625–633. https://doi.org/10.1007/s13595-015-0473-x

McRoberts RE, Hahn JT, Hefty GJ, Cleve JRV (1994) Variation in forest inventory field measurements. Can J For Res 24(9):1766–1770. https://doi.org/10.1139/x94-228

McRoberts RE, Moser P, Zimermann Oliveira L, Vibrans AC (2014) A general method for assessing the effects of uncertainty in individual-tree volume model predictions on large-area volume estimates with a subtropical forest illustration. Can J For Res 45(1):44–51. https://doi.org/10.1139/cjfr-2014-0266

McRoberts RE, Chen Q, Domke GM, Stahl G, Saarela S, Westfall JA (2016) Hybrid estimators for mean aboveground carbon per unit area. For Ecol Manag 378:44–56. https://doi.org/10.1016/j.foreco.2016.07.007

Molto Q, Rossi V, Blanc L (2013) Error propagation in biomass estimation in tropical forests. Methods Ecol Evol 4(2):175–183. https://doi.org/10.1111/j.2041-210x.2012.00266.x

Muller-Landau HC, Condit RS, Chave J, Thomas SC, Bohlman SA, Bunyavejchewin S, Davies S, Foster R, Gunatilleke S, Gunatilleke N, Harms KE, Hart T, Hubbell SP, Itoh A, Kassim AR, LaFrankie JV, Lee HS, Losos E, Makana JR, Ohkubo T, Sukumar R, Sun IF, Nur Supardi MN, Tan S, Thompson J, Valencia R, Muñoz GV, Wills C, Yamakura T, Chuyong G, Dattaraja HS, Esufali S, Hall P, Hernandez C, Kenfack D, Kiratiprayoon S, Suresh HS, Thomas D, Vallejo MI, Ashton P (2006) Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. Ecol Lett 9(5):575–588. https://doi.org/10.1111/j.1461-0248.2006.00904.x

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, BobéKH B, Mikala Okouyi C, Nyangadouma R, Lépengué N, Mbatchi B, Picard N (2014) Site-specific versus pantropical allometric equations: which option to estimate the biomass of a moist central African forest? For Ecol Manag 312:1–9. https://doi.org/10.1016/j.foreco.2013.10.029

Peter B, Tom N (2011) Effects of basal area factor and plot size on precision and accuracy of forest inventory estimates. North J Appl For 28(28):152–156

Petersson H, Holm S, Ståhl G, Alger D, Fridman J, Lehtonen A, Lundström A, Mäkipää R (2012) Individual tree biomass equations or biomass expansion factors for assessment of carbon stock changes in living biomass—a comparative study. For Ecol Manag 270:78–84. https://doi.org/10.1016/j.foreco.2012.01.004

Picard N, Boyemba Bosela F, Rossi V (2015) Reducing the error in biomass estimates strongly depends on model selection. Ann For Sci 72(6):811–823. https://doi.org/10.1007/s13595-014-0434-9

Qin L, Liu Q, Zhang M, Saeed S (2019) Effect of measurement errors on the estimation of tree biomass. Can J For Res 49(11):1371–1378. https://doi.org/10.1139/cjfr-2019-0034

Rubin DB (1987) Multiple imputation for nonresponse in surveys. Wiley, New York, p 258

Salk CF, Chazdon RL, Andersson KP (2013) Detecting landscape-level changes in tree biomass and biodiversity: methodological constraints and challenges of plot-based approaches. Can J For Res 43(9):799–808. https://doi.org/10.1139/cjfr-2013-0048

Shen J (2002) Establishment and application of relative H curve models for main tree species Type in Ji'an. Jiangxi For Sci Technol 1:16–19

Shettles M, Temesgen H, Gray AN, Hilker T (2015) Comparison of uncertainty in per unit area estimates of aboveground biomass for two selected model sets. For Ecol Manag 354:18–25. https://doi.org/10.1016/j.foreco.2015.07.002

Sileshi GW (2014) A critical review of forest biomass estimation models, common mistakes and corrective measures. For Ecol Manag 329:237–254. https://doi.org/10.1016/j.foreco.2014.06.026

Ståhl G, Saarela S, Schnell S, Holm S, Breidenbach J, Healey SP, Patterson PL, Magnussen S, Naesset E, McRoberts RE, Gregoire TG (2016) Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. For Ecosyst 2(2016):153–163. https://doi.org/10.1186/s40663-016-0064-9

Temesgen H, Affleck D, Poudel K, Gray A, Sessions J (2015) A review of the challenges and opportunities in estimating above ground forest biomass using tree-level models. Scand J For Res 30(4):326–335. https://doi.org/10.1080/02827581.2015.1012114

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. For Ecol Manag 262(8):1648–1657. https://doi.org/10.1016/j.foreco.2011.07.018

Wallner A, Elatawneh A, Schneider T, Kindu M, Ossig B, Knoke T (2018) Remotely sensed data controlled forest inventory concept. Eur J Remote Sens 51(1):75–87

Wang G, Oyana T, Zhang M, Adu-Prah S, Zeng S, Lin H, Se J (2009) Mapping and spatial uncertainty analysis of forest vegetation carbon by combining national forest inventory data and satellite images. For Ecol Manag 258(7):1275–1283. https://doi.org/10.1016/j.foreco.2009.06.056

Wang J, Tang J, Guo Y, Gao Y, Yao Y (2015) Aboveground biomass estimation model of two shrubs to Aohan banner northern wind-sandy area in Inner Mongolia. Agric Eng 5(6):44–47

Wayson CA, Johnson KD, Cole JA, Olguín MI, Carrillo OI, Birdsey RA (2015) Estimating uncertainty of allometric biomass equations with incomplete fit error information using a pseudo-data approach: methods. Ann For Sci 72(6):825–834. https://doi.org/10.1007/s13595-014-0436-7