

# *Reducing the error in biomass estimates strongly depends on model selection*

**Nicolas Picard, Faustin Boyemba Bosela & Vivien Rossi**

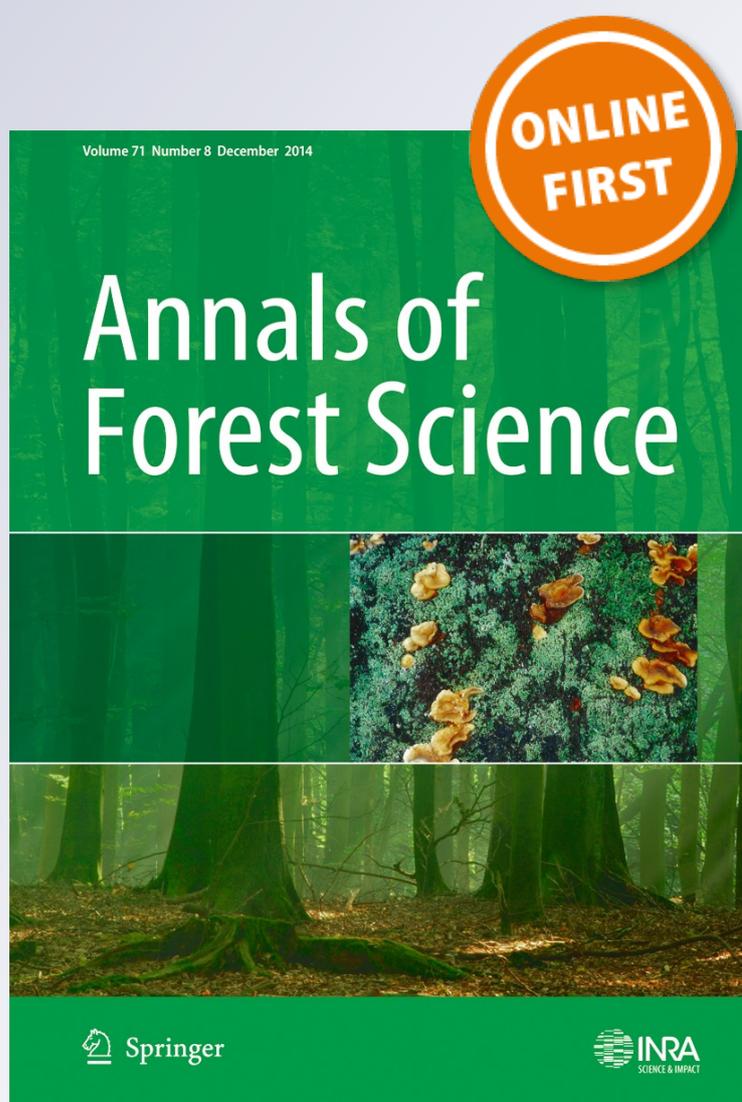
## **Annals of Forest Science**

Official journal of the Institut National de la Recherche Agronomique (INRA)

ISSN 1286-4560

Annals of Forest Science

DOI 10.1007/s13595-014-0434-9



**Your article is protected by copyright and all rights are held exclusively by INRA and Springer-Verlag France. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Reducing the error in biomass estimates strongly depends on model selection

Nicolas Picard · Faustin Boyemba Bosela ·  
Vivien Rossi

Received: 11 April 2014 / Accepted: 16 October 2014  
© INRA and Springer-Verlag France 2014

## Abstract

• **Key message** Improving the precision of forest biomass estimates requires prioritizing the different sources of errors. In a tropical moist forest in central Africa, the choice of the allometric equation was found to be the main source of error.

• **Context** When estimating the forest biomass at the landscape level using forest inventory data and allometric models, there is a chain of propagation of errors including the measurement errors, the models' prediction error, the error due to the model choice, and the sampling error.

• **Aims** This study aims at comparing the contributions of these different sources of error to the total error, to prioritize them, and improve the precision of biomass estimates.

• **Methods** Using a 9-ha permanent sample plot in a moist forest near Kisangani in the Democratic Republic of Congo and seven competing allometric models, we estimated the contributions of the different sources of error to the total error of the per hectare biomass estimate, for plot sizes ranging from 0.04 to 1 ha.

• **Results** When there was no a priori on which model being the best and for 1-ha plots, the error due to the model choice was the largest source of error (76 % of the total error). Using weights to combine the predictions of the different models into a single ensemble prediction strongly reduced this error.

• **Conclusion** Collecting training data sets on tree biomass at many sites would be needed to improve the precision of forest biomass estimates in central Africa.

**Keywords** Aboveground biomass · African moist forest · Allometric equation · Bayesian model averaging · Error propagation · Prediction error

**Handling Editor:** Laurent Saint-Andre

**Contribution of the co-authors** Nicolas Picard designed the study, undertook the analyses and drafted the manuscript.

Faustin Boyemba Bosela provided data.

Vivien Rossi was responsible for review of the manuscript and provision of comments.

N. Picard (✉)

COMIFAC, BP 20818, Yaounde, Cameroon  
e-mail: nicolas.picard@cirad.fr

F. Boyemba Bosela

University of Kisangani, BP 2012 Kisangani,  
Democratic Republic of Congo  
e-mail: faustinboyemba@yahoo.fr

V. Rossi

CIRAD, UPR Bsef, BP 2572 Yaounde, Cameroon

V. Rossi

University Yaounde 1, UMMISCO, Yaounde, Cameroon  
e-mail: vivien.rossi@cirad.fr

## 1 Introduction

Estimating forest carbon stocks in an accurate, precise and verifiable way is a current challenge for forest-based projects to attenuate greenhouse gas emissions, such as REDD+ projects (Angelsen et al. 2012). Estimating forest carbon stocks from the plot to the national level involves a combination of techniques, from field measurement to remote sensing, that are mutually dependent (Gibbs et al. 2007). At the tree level, direct biomass measurement consists in felling, drying and weighing trees. Because these measurements are costly, labour-intensive and above all destructive, they cannot be used at the forest level, thus entailing the use of biomass allometric equations to convert forest inventory data into biomass estimates. Allometric

equations are statistical models that predict the biomass of a tree from other dendrometrical characteristics (such as diameter, height, wood density) that are easier to measure and non-destructive. Several authors have highlighted that current knowledge on allometric equations in tropical rain forests needs improvement to get precise and accurate estimates of carbon stocks (Basuki et al. 2009; Djomo et al. 2011; Alvarez et al. 2012; Ngomanda et al. 2014).

When using a field inventory of plots and allometric equations, landscape-level estimates of forest carbon stocks have an error (as defined by GOF-C-GOLD 2012 and in line with the IPCC 2006 guidelines) that can be broken into two main components (Cunia 1987; Chave et al. 2004; Wagner et al. 2010; van Breugel et al. 2011): (i) sampling error, that is dependent on landscape heterogeneity, plot size and shape, and the number of plots, and (ii) model error which follows from the differences between the true plot biomass values and the model predictions. Optimizing the sampling design so as to minimize the sampling error is the goal of the forest inventory planning. The model error, in turn, can be broken into three main components (Chave et al. 2004; van Breugel et al. 2011; Molto et al. 2013): (iii) the error due to the choice of the allometric equation, (iv) the prediction error of the allometric equation, and (v) the error on the predictors of the allometric equations, i.e. measurement errors for dendrometrical variables such as diameter or height, or estimation errors for species-specific traits such as wood density. Finally, the prediction error (iv) breaks down into two components: (iv-a) the error due to the uncertainty on the model's parameter values, and (iv-b) the residual error of the model. While the former is a sampling-type error that can be squeezed down by increasing the number of observations used to fit the model, the latter represents the inter-tree variability and is not reducible at tree level unless additional predictors are included in the model.

Improving the precision and accuracy of landscape-level estimates of carbon stocks requires reducing the errors at each level of this chain of error propagation. For instance, Hunter et al. (2013) quantified the impact on the precision of biomass prediction of the measurement error of tree height. Nevertheless, not all sources of error equally contribute to the total error at the end of chain, and it is important to diagnose which sources of error contribute the most to address them in priority. Most studies so far have highlighted that the error due to model choice was prominent (Melson et al. 2011). van Breugel et al. (2011) compared the error due to the model choice, the error due to the uncertainty on the model's parameters and the sampling error to predict the forest biomass at the landscape level in Panama. Molto et al. (2013) compared the error due to the model choice and the error on the predictors of the allometric equations to predict the biomass at the tree and plot levels in French Guiana. Gonzalez et al. (2014) compared the error due to the model

choice and the error on the predictors of the allometric equations to predict the forest biomass at the landscape level in Peru. All these authors consistently concluded that the error due to the model choice was the most important source of error. However, none of these studies integrated the whole chain of propagation of errors to prioritize the different sources of error at the landscape-level.

Failing to integrate the whole chain of propagation of errors may also result in underestimated confidence intervals, which may lead to misleading results. Lewis et al. (2013) and Kearsley et al. (2013) estimated the biomass at the landscape level as if the biomass allometric equation was exact, with no associated model error. On the contrary, Cohen et al. (2013) took account of both the sampling error and of the model error, and showed that the latter significantly contributed to the variability of the per hectare biomass estimate at the landscape level. Considering only the sampling error and disregarding the model error results in undervalued confidence intervals for the per hectare biomass.

Recent development in biomass models has argued that the biomass allometry based on diameter and height was universal (i.e. valid across species and bioclimatic zones), local variations stemming from variations in the height:diameter relationship (Chave et al. 2014). Combined with the fact that forest inventory data include diameter but not height, this approach leads to height models integrated into two-entries biomass models (Banin et al. 2012; Feldpausch et al. 2012). Nevertheless, plugging height models into biomass models also means that the errors of the former propagate into the latter (Fortin and DeBlois 2010). The choice of the height model may also strongly influence the biomass estimate, as recently shown by Kearsley et al. (2013) whose reevaluation of the height:diameter relationship in central Congo basin led to a lower estimate of carbon stocks.

In this study, we decomposed the whole chain of propagation of errors when estimating the forest biomass at the landscape level to assess the relative contribution of the different sources of error. The per hectare biomass was estimated for a tropical moist forest in eastern Democratic Republic of Congo. Seven allometric equations, among the most commonly used for central African forests and including two-entries biomass equations with an integrated height model, were used to assess the error due to the model choice.

## 2 Materials and methods

### 2.1 Study area

The study was carried out in the Yoko forest reserve (0°17'44"N, 25°18'50"E), 30 km South of Kisangani in the

Oriental province of the Democratic Republic of Congo, and in the Yangambi reserve (0°46'48"N, 24°27'25"E), 100 km West of Kisangani (Boyemba Bosela 2011). The climate in this area is equatorial of the Af-type (Köppen-Geiger classification), with an average annual rainfall of 1750 mm and a mean temperature of 25 °C. The relief consists of a plateau interrupted with rivers. Soils are oxisoils that are typical of the central Congo basin. The Yoko forest is moist semideciduous forest, with *Scorodophlæus zenkeri* and *Pericopsis elata* as typical species, but also includes large patches of monodominant evergreen *Gilbertiodendron dewevrei* forest. The area of the Yangambi forest that we considered consists of former abandoned plantations (plantation date between 1937 and 1974), with the current dominant species being the planted ones. Although the Yoko and the Yangambi forests differ in their species composition and history, they belong to the same bioclimatic forest type that is, in the current state of knowledge, the only factor that guides the choice of an allometric equation in central Africa.

### 2.1.1 Forest inventory data

In April 2009, a 300 m × 300 m permanent sample plot was set up in the Yoko forest and all trees with diameter at breast height (dbh) ≥ 10 cm were inventoried. Their dbh, species and spatial coordinates were measured. This 9-ha plot was divided into  $K = 9$  1-ha plots and each 1-plot was considered as a sampling unit. Because biomass exhibits positive spatial autocorrelation (Keller et al. 2001; Chave et al. 2003; Wagner et al. 2010), these  $K$  units are not true independent

replicates. However, we considered the spatial autocorrelation at a spatial lag of 100 m to be small enough to consider these pseudo-replicates as acceptable.

### 2.1.2 Biomass data

We used tree biomass data collected at Yangambi by Ebuy Alipade et al. (2011) as a training data set to balance existing biomass allometric equations. This data set includes 12 trees from three species. We also used the dbh:height relationship fitted by Kearsley et al. (2013) at Yangambi to convert two-entries allometric equations (depending on height and dbh) into single-entry equations (depending on dbh alone). It is a Mitscherlich equation fitted to 487 trees between 10 and 125 cm dbh, with a residual standard deviation of 4.221 m:

$$H = a - b \exp(-cD) \quad (1)$$

where  $D$  is tree dbh (in cm),  $H$  is height (in m),  $a = 36.4$  m is the maximum asymptotic height,  $b = 31.7$  m is the difference between maximum and minimum height, and  $c = 0.0221 \text{ cm}^{-1}$  is the shape parameter of the curve. For all subsequent calculations, we used an average wood density of  $0.6 \text{ g cm}^{-3}$  (Chave et al. 2005; Henry et al. 2010).

### 2.2 Biomass allometric models

To estimate the error due to the model choice, seven allometric equations were considered, including two pantropical equations and five equations from central Africa (Table 1). All models except those of Dorisca et al. were fitted

**Table 1** Allometric equations to assess the error due to model choice

Reference	Country	No.	Expression	$\sigma$	$n$	$D_{\min}$	$D_{\max}$
Chave et al. (2005)	Pantropical	II.3	$\rho \exp[-1.499 + 2.148 \ln D + 0.207(\ln D)^2 - 0.0281(\ln D)^3]$	0.356 <sup>a</sup>	1504	5	156
Chave et al. (2005)	Pantropical	1.5	$0.0509 \times \rho D^2 H$	0.316 <sup>a</sup>	1349	5	156
Dorisca et al. (2011)	Cameroon	5	$\rho \times (11.1956 + 0.040326 D^2 H)$	0.374 <sup>b</sup>	51	10	134
Dorisca et al. (2011)	Cameroon	2	$\rho \times (325.5573 - 47.2984 D + 2.2942 D^2)$	0.254 <sup>c</sup>	55	10	134
Fayolle et al. (2013)	Cameroon	2	$\rho \exp[-1.183 + 1.940 \ln D + 0.239(\ln D)^2 - 0.0285(\ln D)^3]$	0.188 <sup>a</sup>	138	5	192
Ngomanda et al. (2014)	Gabon	5	$\exp[-4.060 + 4.062 \ln D - 0.228(\ln D)^2 + 1.431 \ln \rho]$	0.330 <sup>a</sup>	101	12	109
Ngomanda et al. (2014)	Gabon	11	$\exp[-2.568 + 0.952 \ln(D^2 H) + 1.189 \ln \rho]$	0.292 <sup>a</sup>	101	12	109

In all these equations, the predicted quantity is the tree aboveground biomass in kg,  $D$  is dbh in cm,  $H$  is height in m and  $\rho$  is wood density in  $\text{g cm}^{-3}$ . The number (No.) is the model number in the publication.  $\sigma$  is the residual standard error of the model.  $n$  is the number of observations used to fit the model.  $D_{\min}$  and  $D_{\max}$  are the minimum and maximum dbh (in cm) in the data set used to fit the model. The expressions of the equations include the correction factor for back-transformation when the model was fitted to log-transformed data

<sup>a</sup>Residual of log-transformed data

<sup>b</sup>The standard deviation of residuals is  $\sigma D^{2.2}$

<sup>c</sup>The standard deviation of residuals is  $\sigma D^{2.15}$

to aboveground dry biomass (as defined by the IPCC; Eggleston et al. 2006). Dorisca et al.'s models were fitted to total tree volume, which can be converted into biomass by multiplying it by wood density. No correction factor for foliage was applied in this case. All models were particular cases of the following general expression:

$$f(B/\rho^\mu) = \alpha + \beta f(D) + \gamma [f(D)]^2 + \delta [f(D)]^3 + \lambda f(D^2 H) \quad (2)$$

where  $f$  is a transformation of variables (either the identity function or the logarithm),  $B$  is tree aboveground dry biomass (in kg),  $\rho$  is wood density (in  $\text{g cm}^{-3}$ ) and  $\alpha, \beta, \gamma, \delta, \lambda, \mu$  are the model coefficients. Polynomial models (i.e. with  $f$  being the identity function) were fitted by generalized least squares regression using a power model for the residual variance, to account for the heteroscedasticity of residuals. Log-transformed models (i.e. with  $f$  being the logarithm) were fitted by ordinary least squares regression. In that case, the back-transformation induced a bias that was corrected using a correction factor  $\exp(\sigma^2/2)$ , where  $\sigma$  is the residual standard error.

Four models used only dbh and wood density as predictors (i.e.  $\lambda = 0$ ), whereas three models further included tree height as a predictor. Because tree height was not measured at Yoko, it was predicted using (1). Plugging (1) into (2) means that biomass was predicted from dbh and wood density only.

### 2.3 Error propagation

#### 2.3.1 Sampling error

At the South-East corner of each 1-ha plot, square subplots with side  $s$  (in hm) were located, with  $s = 0.2, 0.4, 0.6, 0.8$  and  $1$  hm. The per-hectare biomass  $B_k(s)$  for the  $k$ th subplot ( $k = 1, \dots, K$ ) with side  $s$  was computed from the collection of trees inventoried in this subplot. The sampling error for a subplot with side  $s$  was computed as:  $\sum_{k=1}^K [B_k(s) - \bar{B}(s)]^2 / K$ , where  $\bar{B}(s) = \sum_{k=1}^K B_k(s) / K$  is the mean per-hectare biomass of the  $K$  subplots with side  $s$ .

#### 2.3.2 Error due to the model choice

Each biomass allometric model  $m$  gave an estimate  $B_m$  of the tree biomass. Let  $\pi_m$  be the weight associated with model  $m$ , giving its probability to the best among the ensemble of allometric equations (so that  $\sum_m \pi_m = 1$ ). The estimate of tree biomass according to the ensemble of models was  $\bar{B} = \sum_m \pi_m B_m$ , and the error due to the model choice was  $\sum_m \pi_m (B_m - \bar{B})^2$ . Without further data, all models were equally likely to be the best, which meant  $\pi_m = 1/M$ , where  $M = 7$  was the number of models in the ensemble.

Given a training data set, Bayesian Model Averaging (BMA) can be used to estimate the weights  $\pi_m$  (Li et al. 2008; Picard et al. 2012). BMA assumes that there is an unknown "true" model, and that the deviations between each model  $m$  of the ensemble and this "true" model can be described by a Gaussian distribution. Thus, the distribution of tree biomass according to BMA is the mixture of  $M$  Gaussian distributions:  $\sum_m \pi_m \phi(B; B_m, \sigma_m D^2)$ , where  $\phi(\cdot; \mu, \sigma)$  is the density of the Gaussian distribution with mean  $\mu$  and standard deviation  $\sigma$ . The standard error of the deviation between model  $m$  and the true model was assumed to increase proportionally to the square dbh. The training data set at Yangambi was used to estimate the posterior probabilities  $\pi_m$  and the standard deviations  $\sigma_m$  (see Appendix A for details on the estimation method).

#### 2.3.3 Prediction error

*Error due to the uncertainty on the model's coefficients.* Computing the error that follows from the uncertainty on the values of the models' coefficients requires knowing the variance matrices of the estimator of these coefficients. Unfortunately, these variance matrices are rarely given in publications, and the original data used to fit the model are rarely available to compute them. At best, the variance (or confidence interval) of the estimate is given separately for each coefficient  $\alpha, \beta, \dots, \mu, a, b, c$  (Dorisca et al. 2011; Fayolle et al. 2013; Ngomanda et al. 2014), but this is not even always the case (e.g. Chave et al. 2005; Kearsley et al. 2013). As often with polynomial regression (on log- or non-transformed data), the estimates of the model coefficients are strongly correlated, and approximating the variance matrix of the estimator of  $\alpha, \beta, \dots$ , by the diagonal matrix with diagonal elements  $\text{Var}(\hat{\alpha}), \text{Var}(\hat{\beta}), \dots$ , may strongly inflate the prediction error.

Therefore, to compute the error due the uncertainty on the models' coefficients, we used an approximate Monte Carlo method based on the simulation of the data set used for model fitting. This Monte Carlo method is defined by pseudo-algorithms 1 and 2 for models (1) and (2), respectively.

---

**Algorithm 1** uncertainty on the coefficients of the  $H:D$  model

---

**Input:** coefficients  $a, b, c$  of model (1).

**Output:** coefficients  $\tilde{a}, \tilde{b}, \tilde{c}$  that replace  $a, b, c$ .

- 1 Draw 487 dbh values  $D_i \sim \mathcal{U}(10, 125)$ , where  $\mathcal{U}$  is the uniform distribution;
  - 2 Draw  $H_i \sim \mathcal{N}(a - b \exp(-cD_i), 4.221)$ , where  $\mathcal{N}$  is the Gaussian distribution;
  - 3 Fit model (1) to the data set of 487 observations  $(D_i, H_i)$ . Let  $\tilde{a}, \tilde{b}, \tilde{c}$  be the estimated coefficients.
-

**Algorithm 2** uncertainty on the coefficients of the allometric model

**Input:** number  $m$  of the model.

**Output:** coefficients  $\tilde{\alpha}_m, \tilde{\beta}_m, \tilde{\gamma}_m, \tilde{\delta}_m, \tilde{\lambda}_m$  of model  $m$

- 1 Draw  $n$  dbh values  $D_i \sim \mathcal{U}(D_{\min}, D_{\max})$  (values of  $n, D_{\min}, D_{\max}$  for each model are given in Table 1);
- 2 Compute  $H_i$  from  $D_i$  using (1);
- 3 Draw  $X_i \sim \mathcal{N}(\alpha_m + \beta_m f(D_i) + \gamma_m [f(D_i)]^2 + \delta_m [f(D_i)]^3 + \lambda_m f(D_i^\kappa), \sigma D_i^\kappa)$  (values of  $\sigma$  and  $\kappa$  for each model are given in Table 1, with  $\kappa = 0$  when  $f$  is the logarithm);
- 4 Compute  $B_i = \rho^\mu f^{-1}(X_i)$
- 5 Fit model  $m$  to the data set of  $n$  observations  $(D_i, H_i, B_i)$ . Let  $\tilde{\alpha}_m, \tilde{\beta}_m, \tilde{\gamma}_m, \tilde{\delta}_m, \tilde{\lambda}_m$  be the estimated coefficients.

By repeating the Monte Carlo procedure  $J$  times,  $J$  outcomes of the coefficients are obtained, whose empirical distribution approximates the distribution of the coefficients estimator. The distribution of the Monte Carlo outcomes is not exactly the same as the distribution of the estimated coefficients because the design matrix of the original data used to fit the model is not the same as the design matrix of the simulated data set. Yet, we assume that the main features of the data set were captured by the simulation method.

*Residual error* The residual errors of models (1) and (2) were also computed using a Monte Carlo method, by adding a random error to the models' predictions. This method is defined by pseudo-algorithms 3 and 4 for models (1) and (2), respectively.

**Algorithm 3** residual error of the  $H:D$  model

**Input:** error-free height prediction  $H$ .

**Output:** height prediction  $\hat{H}$ .

- 1 Draw  $\hat{H} \sim \mathcal{N}(H, 4.221)$ .

**Algorithm 4** residual error of the allometric model

**Input:** error-free prediction  $X_m$  of allometric model  $m$  and dbh  $D$  of the corresponding tree.

**Output:** prediction  $\hat{X}_m$  of allometric model  $m$ .

- 1 Draw  $\hat{X}_m \sim \mathcal{N}(X_m, \sigma D^\kappa)$  (values of  $\sigma$  and  $\kappa$  for each model are given in Table 1, with  $\kappa = 0$  when  $f$  is the logarithm).

### 2.3.4 Errors on the model predictors

Tree dbh and height are measured and thus have a measurement error. Specific wood density is estimated from the

literature and also has an associated estimation error. In this study, we disregarded the estimation error of wood density and focused on the measurement error associated with dbh and height. Dbh was assumed to be measured with a 0.5 % error (i.e. to the nearest centimetre for a tree dbh of 1 m), see pseudo-algorithm 5. Because tree height was actually not measured at Yoko but predicted from (1), measurement error on height was included as an additional residual error of model (1). Using the results of Hunter et al. (2013), measurement error on height was assumed to have a Gaussian distribution with mean zero and standard deviation 7.3 m (pseudo-algorithm 6).

**Algorithm 5** measurement error on dbh

**Input:** error-free tree dbh  $D$ .

**Output:** tree dbh  $\hat{D}$ .

- 1  $\hat{D} \sim \mathcal{U}(D \pm D/200)$ .

**Algorithm 6** measurement error on height

**Input:** error-free tree height  $H$ .

**Output:** tree height  $\hat{H}$ .

- 1 Draw  $\hat{H} \sim \mathcal{N}(H, 7.3)$ .

### 2.3.5 Total error

The total error in the estimate of the per-hectare biomass was obtained by combining all the sources of errors previously listed, and was computed as:

$$S_{\text{tot}}^2(s) = \frac{1}{KJ} \sum_{k=1}^K \sum_{m=1}^M \sum_{j=1}^J \pi_m [\mathcal{B}_{kmj}(s) - \bar{\mathcal{B}}(s)]^2 \quad (3)$$

where  $\mathcal{B}_{kmj}(s)$  was the per-hectare biomass for subplot  $k = 1, \dots, K$  with side  $s$  predicted using model  $m = 1, \dots, M$  at iteration  $j = 1, \dots, J$  of the Monte Carlo algorithm, and  $\bar{\mathcal{B}}(s) = \sum_k \sum_j \sum_m \pi_m \mathcal{B}_{kmj}(s) / (KJ)$  was the mean per-hectare biomass. The subplot biomass  $\mathcal{B}_{kmj}$  was computed using pseudo-algorithm 7.

All computations were performed using the R software (Development Core Team 2005). For the Monte Carlo algorithm,  $J = 5,000$  iterations were performed.

The different sources of error did not contribute to the total error in the same way. Whereas the sampling error and the error due to the model choice were computed as between-level variances for a modal variable (i.e. the plot and the model, respectively), the prediction error and the error on the model predictors were obtained as additional noise by a Monte Carlo method. Therefore, the total error (3) could be partitioned using the decomposition of the total sum of squares in two-way analysis of variance for a

balanced model, where the plot and the model played the role of the explanatory variables, and the prediction error and the error on the model predictors jointly corresponded to the residual error of this analysis of variance (see Appendix B for details on this error decomposition). This partition of the total error resulted in an additional error term, namely the interaction between the plot and the allometric equation. This term follows from the fact that the difference in predicted tree biomass between two allometric equations depends on tree dbh. Therefore, the difference in predicted plot biomass between two allometric equations depends on the tree size distribution within the plot.

**Algorithm 7** propagation of errors

- 1 For  $j$  of 1 to  $J$  { # Monte Carlo iteration
- 2     Pseudo-algorithm 1  $\rightarrow \tilde{a}_j, \tilde{b}_j, \tilde{c}_j$  # uncertainty on the coefficients of the  $H:D$  model
- 3     For  $m$  of 1 to  $M$  { # iteration on the allometric models
- 4         Pseudo-algorithm 2  $\rightarrow \tilde{\alpha}_{mj}, \tilde{\beta}_{mj}, \tilde{\gamma}_{mj}, \tilde{\delta}_{mj}, \tilde{\lambda}_{mj}$  # uncertainty on the coefficients of the allometric models
- 5         For  $k$  of 1 to  $K$  { # iteration on the plots
- 6             For  $i$  of 1 to  $n_k$  (where  $n_k$  is the number of trees in plot  $k$ ) { # iteration on the trees within each plot
- 7                 Let  $D_{ki}$  be the dbh of the  $i$ th tree of the  $k$ th plot. Compute  $H_{ki} = \tilde{a}_j - \tilde{b}_j \exp(-\tilde{c}_j D_{ki})$
- 8                  $D_{ki} \rightarrow$  Pseudo-algorithm 5  $\rightarrow \tilde{D}_{ki}$  # measurement error on dbh
- 9                  $H_{ki} \rightarrow$  Pseudo-algorithm 6  $\rightarrow \tilde{H}_{ki}$  # measurement error on height
- 10                  $\tilde{H}_{ki} \rightarrow$  Pseudo-algorithm 3  $\rightarrow \hat{H}_{ki}$  # residual error of the  $H:D$  model
- 11                 Compute  $X_{kmji} = \tilde{\alpha}_{mj} + \tilde{\beta}_{mj} f(\tilde{D}_{ki}) + \tilde{\gamma}_{mj} [f(\tilde{D}_{ki})]^2 + \tilde{\delta}_{mj} [f(\tilde{D}_{ki})]^3 + \tilde{\lambda}_{mj} f(\tilde{D}_{ki}^2 \hat{H}_{ki})$
- 12                  $X_{kmji}, \tilde{D}_{ki} \rightarrow$  Pseudo-algorithm 4  $\rightarrow \tilde{X}_{kmji}$  # residual error of the allometric model
- 13                 Compute  $B_{kmji} = \rho^\mu f^{-1}(\tilde{X}_{kmji})$
- 14                 Compute  $\mathcal{B}_{kmj} = (\sum_i B_{kmji})/s^2$  }

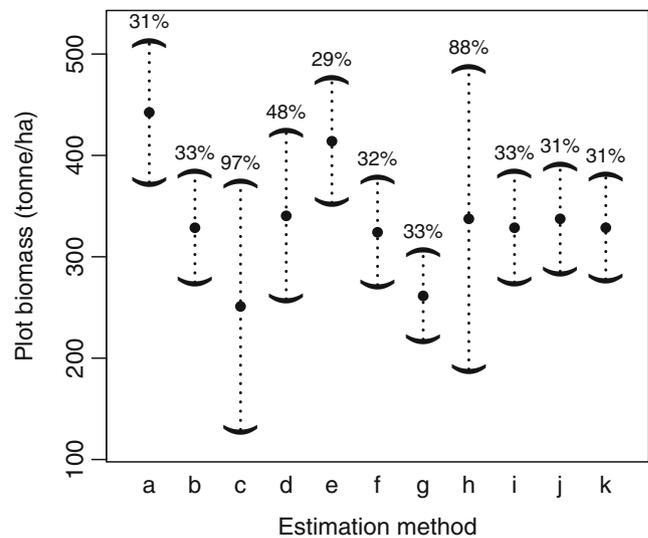
The residual error term of the analysis of variance could further be decomposed by alternatively switching on/off in the Monte Carlo algorithm each contributing source of error. By default, in Algorithm 7, all sources of error are switched on. The uncertainty on the coefficients of the  $H:D$  model could be switched off by replacing pseudo-algorithm 1 with: set  $\tilde{a} = a, \tilde{b} = b, \tilde{c} = c$ . The uncertainty on the coefficients of the allometric model could be switched off by replacing pseudo-algorithm 2 with: set  $\tilde{\alpha}_m = \alpha_m, \tilde{\beta}_m = \beta_m, \dots, \tilde{\lambda}_m = \lambda_m$ . The residual error of the  $H:D$  model or the

measurement error on height could be switched off by replacing pseudo-algorithm 3 or 6 with: set  $\hat{H} = H$ . The residual error of the allometric models could be switched off by replacing pseudo-algorithm 4 with: set  $\hat{X}_m = X_m$ . The measurement error on dbh could be switched off by replacing pseudo-algorithm 5 with:  $\hat{D} = D$ . The difference in total error when a particular source of error was switched on and when it was switched off gave the contribution of this source of error to the total error.

**3 Results**

3.1 Biomass estimate based on 1-ha sampling plots

There was strong disagreement among the allometric equations on the estimate of the aboveground biomass (Fig. 1a–g). Two models (Dorisca et al.'s model 5 and Ngomanda et al.'s model 11) brought low estimates (minimum value: 251 tonne ha<sup>-1</sup>). Two models (Chave et al.'s model II.3 and Fayolle et al.'s model 2) brought high estimates (maximum value: 442 tonne ha<sup>-1</sup>). The three remaining models brought intermediate estimates. The two models that brought low estimates used  $D^2H$  as a predictor of tree biomass, whereas



**Fig. 1** Estimate of the aboveground biomass at Yoko, DRC, according to different estimation methods based on 1-ha sampling plots (the dot indicates the estimate; the whiskers show the 95 % confidence interval; the value given on top of the upper whisker is the ratio of the amplitude of the 95 % confidence interval to the estimate). Tree biomass is computing using **a** Chave et al.'s model II.3, **b** Chave et al.'s model I.5, **c** Dorisca et al.'s model 5, **d** Dorisca et al.'s model 2, **e** Fayolle et al.'s model 2, **f** Ngomanda et al.'s model 5, **g** Ngomanda et al.'s model 11, **h** and **j** the mean of all aforementioned models with equal weights, **i** and **k** the mean of all aforementioned models with BMA weights. In **a–i**, all sources of errors are included; in **j–k**, only the sampling error is included

the two models that brought high estimates did not use tree height as a predictor, thus confirming that Kearsley et al.'s height model predicts smaller tree stature than that implied by allometric equations based on dbh alone.

There were also major differences among allometric equations in the precision of the biomass estimate. The less precise estimate, obtained with Dorisca et al.'s model 5 (Fig. 1c), had an amplitude of its 95 % confidence interval that was nearly equal to the biomass estimate. For this model, the main source of error (62 % of the square error in biomass) was the uncertainty on the model coefficients. For most equations, the amplitude of the 95 % confidence interval of the biomass estimate was approximately 30 % of the estimate (Fig. 1a, b, e–g).

Because the different allometric equations brought contrasted biomass estimates, combining these estimates into a single one using the mean resulted in a large error due to the model choice. As a consequence, the 95 % confidence interval of this mean estimate across models was also large and had an amplitude that equalled 88 % of the biomass estimate (Fig. 1h). Estimating the biomass as the mean of the predictions of the different models supposes that we have no *a priori* knowledge of the relevance of each model. Alternatively, BMA can be used to balance the prediction of each model. Using the training data set on tree biomass at Yangambi, BMA resulted in contrasted weights across models (Table 2). One single model, namely Chave et al.'s model I.5, largely outperformed all other models. As a consequence, the weighted mean estimate of biomass (using the BMA weights) was almost the same as the estimate obtained with Chave et al.'s model I.5 (Fig. 1i). Moreover, the error due to the model choice was almost null when weighting the models with the BMA weights, so that the confidence interval of the biomass mean with BMA weights was almost the same as

that of Chave et al.'s model I.5 (compare Fig. 1b and i), and was much narrower than that of the biomass mean with equal weights (compare Fig. 1h and i).

When estimating the biomass as the mean across models (with equal weights), the choice of the model was the largest source of error. As a consequence, when disregarding this error and focusing on the sampling error (i.e. the between-plot variability), the precision of the estimate greatly increased. Thus, the amplitude of the 95 % confidence interval dropped from 88 % to 31 % of the biomass estimate when considering sampling as the only source of errors (compare Fig. 1h and j). On the contrary, when estimating the biomass as a weighted mean across models (using the BMA weights), the error due to the model choice became negligible. As a consequence, the amplitude of the 95 % confidence interval was little different whether one considered all sources of error (Fig. 1i) or only the sampling error (Fig. 1k).

### 3.2 Error partition among error sources

The total square error in biomass decreased with the size of the sample plot (Fig. 2a and c). This decrease was slower than  $n$ , where  $n = (100/s)^2$  (where  $s$  is the side of the sample plot in metres) is the number of plots to sample to get a cumulated sampled area of 1 ha. In other words, the total sum of squares divided by  $n$  was smaller for plots with area 0.04 ha and increased till plots with area 1 ha (not shown here to save space). Even if the total error decreased with plot size, not all sources of error depended in the same way on plot size. The sampling error, the residual error and the plot-model interaction decreased with plot size. On the contrary, the error due to the model choice and the error due to the uncertainty on the coefficients were almost independent of the plot size (Fig. 2a and c). As a consequence, the relative contribution of each source of error to the total error changed with plot size (Fig. 2b and d).

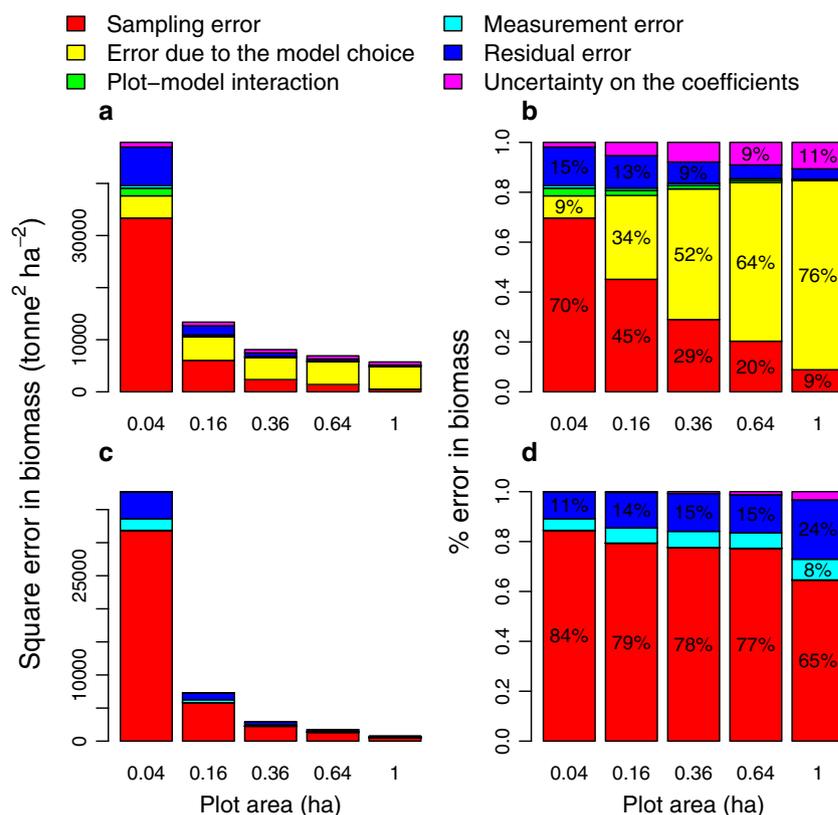
When there was no *a priori* on which model being the best, so that the model predictions were combined into a single ensemble prediction using equal weights, the relative contribution of the error due to the model choice increased from 9 % to 76 % as plot size increased from 0.04 ha to 1 ha (Fig. 2b). Meanwhile, the relative contribution of the sampling error decreased in about the same proportion. Therefore, the sampling error was the largest source of error for small plots (it accounted for 70 % of the total error for plots of 0.04 ha) whereas the error due to the model choice was the largest source of error for large plots (it accounted for 76 % of the total error for plots of 1 ha). When using equal weights for the models, the proportion of error corresponding to the residual error decreased when plot size increased (from 15 % to 4 % for a plot size from 0.04 ha to 1 ha; Fig. 2b), whereas the proportion of the error due to

**Table 2** Bayesian model averaging of seven allometric equations (see Table 1) using a training data set of tree biomass collected at Yangambi, DRC

Model	$\pi_m$	$\sigma_m$
Chave et al. (2005) II.3	< 0.01	0.359
Chave et al. (2005) I.5	> 0.99	0.322
Dorisca et al. (2011) 5	< 0.01	0.451
Dorisca et al. (2011) 2	< 0.01	0.441
Fayolle et al. (2013) 2	< 0.01	0.369
Ngomanda et al. (2014) 5	< 0.01	0.424
Ngomanda et al. (2014) 11	< 0.01	0.423

$\pi_m$  is the BMA weight of model  $m$ .  $\sigma_m$  (in  $\text{kg cm}^{-2}$ ) is the coefficient such that the standard error of the deviation in biomass between model  $m$  and the true model for a tree of dbh  $D$  equals  $\sigma_m D^2$

**Fig. 2** Partition of the error among the different sources of error when estimating the per hectare forest biomass at Yoko, DRC, for a sampling plot with side 20, 40, 60 and 80 m. The predictions of the seven allometric equations are combined into a single ensemble prediction using **a, b** equal weights, or **c, d** BMA weights. The part of each source of error is shown as **a, c** its contribution to the total sum of squares, or as **b, d** its proportion of the total error (the proportion is typed when  $>8\%$ )



the uncertainty on the coefficients increased (from 2% to 11%). For all plot sizes, the plot-model interaction and the measurement error contributed very little to the total error as compared to the other sources of error (Fig. 2a and b).

When the model predictions were weighted using the BMA weights, the pattern of error partition changed (Fig. 2c and d). Because the BMA weights were concentrated on a single model (Chave et al.'s model I.5), the error partition basically was the one obtained when using this model alone to predict biomass. As a consequence, the error due to the model choice and the plot-model interaction became negligible with respect to the other sources of error. For all plot sizes, sampling error was the largest source of error (representing 65–84% of the total error depending on plot size), followed by the residual error, then the measurement error (Fig. 2c, d).

## 4 Discussion

### 4.1 Error due to the model choice

When using 1-ha sample plots (the most common size for permanent plots in central Africa, see Lewis et al. 2013) and when all models were *a priori* equally likely, the error

due to the model choice was the greatest source of error at Yoko. It represented three-quarters of the total error. This result is consistent with the conclusions of other studies who compared the contributions of errors along part or the whole of the chain of propagation of errors (Melson et al. 2011; van Breugel et al. 2011; Molto et al. 2013; Gonzalez et al. 2014), and thus does not seem to be specific to the Yoko forest. It implies that the model choice is the source of error to address at first to improve the precision of biomass estimates at the landscape-level. This source of error can be tackled by increasing the number of sites where tree biomass measurements are performed. Few data need to be collected at each site, because data are primarily needed as training data to perform BMA and rank models, and not as calibration data to fit new allometric equations. When models' predictions can be balanced using weights that represent the probability for the model to be the best, the error due to the model choice is reduced.

Therefore, it is also important to correctly estimate the weight associated with each model. A model that is locally unrealistic must be assigned a small weight so that it does not contribute much to the ensemble prediction. BMA is a useful technique that is commonly used in climate forecast (Raftery et al. 2005; Smith et al. 2009) but probably remains underused in forestry (Li et al. 2008; Picard et al. 2012).

As an alternative, one may use model selection (Massart 2007), which requires assessing the risk of oracles. On the basis of the training data collected at Yangambi, a single model outperformed all other models. Selecting a single model does not seem to be the usual outcome of BMA (Li et al. 2008; Picard et al. 2012). More often, several models contribute to the ensemble prediction. Single model outweighting could result from a fortuitous overlap between the training data set used for BMA and the data set used to fit one the models of the BMA ensemble. However, the outperforming equation here integrated Chave et al.'s biomass equation depending on height and Kearsley et al.'s model for height, thus rendering this explanation implausible. More likely, it resulted from the very small size (12 trees) of the training data set at Yangambi.

#### 4.2 Sampling error

Sampling error remained an important source of error at Yoko, particularly when plots were small. It can be addressed using the classical techniques of sampling theory (Cochran 1977), the first driver of the error being the number of plots sampled. Sampling error expectedly decreased when the plot size increased. When optimizing forest inventories, it is commonplace to model this relationship using a power model (Zeide 1980). At Yoko, using the BMA weights to combine the models' predictions and considering the sampling error alone (without the other sources of error), the relationship between the coefficient of variation (CV, in %) of the biomass and the plot area ( $A$ , in  $m^2$ ) was:  $CV = 1395 \times A^{-0.546}$ . Therefore, contrary to the total error that decreased slower than  $1/A$ , sampling error decreased quicker than  $1/A$ . This result contrasts with what is generally found, where the exponent of  $A$  is generally greater than  $-0.5$ . For instance, Keller et al. (2001) found:  $CV = 706 \times A^{-0.350}$  in Brazilian Amazonia; Chave et al. (2003) found:  $CV = 942 \times A^{-0.45}$  in Panama; and Wagner et al. (2010) found:  $CV = 557 \times A^{-0.430}$  in French Guiana. An exponent smaller than  $-0.5$  like in Yoko implies some regularity in the spatial pattern of biomass, whereas an exponent greater than  $-0.5$  as is usual implies some clustering in the spatial pattern of biomass. At a constant effort of sampling (in terms of total area sampled), the latter means that many small plots are more efficient than few large plots, whereas the former means the opposite.

#### 4.3 Plot-model interaction

The error due to the interaction between the plot structure and the allometric equation is important to estimate emission factors. The emission factor is the difference in carbon content between two types of vegetation, and thus represents

the amount of carbon that is emitted or stored when changing from one type to the other. If the difference in predicted biomass between two allometric model is independent of the plot dbh structure (i.e. no plot-model interaction), and if the same model is used to estimate the emission factor (i.e. the biomasses of the two types of vegetation are predicted using the same model), then the estimate of the emission factor is independent of the model choice. Therefore, the model choice may have lower impact on the estimates of the emission factors than on the estimates of the biomass stocks as soon as the plot-model interaction is low. The contribution of the plot-model interaction to the total error was low at Yoko. However, because all sample plots were taken from a single 9-ha plot with an homogeneous dbh structure, it is likely that the plot-model interaction at Yoko was undervalued in comparison to a wider scale. Sampling independent plots at a landscape level with more contrasted dbh structures may result in a greater plot-model interaction than at Yoko.

#### 4.4 Residual error

Like the sampling error, the residual error decreased with plot size. The residual error in the prediction of individual tree biomass is about  $\sigma = 0.3$  on log-transformed variables (Table 1), which is actually huge considering that it corresponds to an approximate relative margin of error at level 95 % (i.e. half the width of the prediction interval at level 95 % divided by the biomass estimate) of  $\sinh(1.96 \sigma) = 62 \%$ . This residual error at the tree level is levelled off when randomly accumulating trees: trees with positive residuals compensate for trees with negative residuals, and the precision of prediction of the biomass of the average tree increases as the number of trees increases. Biomass, like volume, is likely to present differences among plots (Fortin and DeBlois 2010; Cohen et al. 2013). However, as long as this plot effect is not accounted in the allometric model, increasing the size of the plot is equivalent for the biomass prediction with accumulating more trees, hence a lower residual error at the plot level.

Although the residual error was acceptable at the scale of a 1-ha plot, it may be desirable to reduce it for individual tree biomass prediction. This may be achieved by introducing additional predictors in the allometric equations. At the species level, additional species traits (in addition to wood density) could be included, especially when a significant species effect on model residuals is found (Ngomanda et al. 2014). At the tree level, additional descriptors of tree allometry should be found, such as descriptors of crown architecture. These additional descriptors should remain simple enough to be routinely collected in large scale forest inventories.

#### 4.5 Remaining sources of error

Measurement errors at Yoko remained small as compared to the other sources of error, which confirmed previous studies (Molto et al. 2013; Gonzalez et al. 2014). If measurement techniques should be improved to reduce measurement errors, then tree height measurement should be addressed at first (Hunter et al. 2013).

The error due to the uncertainty on the models' coefficients also remained small at Yoko. This error can be reduced by increasing the sample size used for model fitting. Chave et al. (2004) showed that increasing sample size beyond 100 trees did not improve much the precision of plot biomass prediction.

#### 4.6 Implication for the confidence intervals of biomass estimates

Failing to account for the whole chain of error underestimates the biomass variability. Based on 13 1-ha plots, Kearsley et al. (2013) estimated the aboveground biomass to be  $324 \pm 40$  (95 % CI) tonne ha<sup>-1</sup> at Yangambi, significantly different from the estimate of  $396 \pm 14.3$  (95 % CI) tonne ha<sup>-1</sup> given by Lewis et al. (2013) for central African moist forests. However, neither Kearsley et al. nor Lewis et al. took account of all sources of errors. They considered the sampling error as the only source of error. Based on 9 1-ha plots and using equal weights for all models, the present study estimated the biomass at Yoko (~ 100 km from Yangambi) to be  $337 \pm 148$  tonne ha<sup>-1</sup> when accounting for all sources of errors (Fig. 1h), and  $337 \pm 52$  tonne ha<sup>-1</sup> when accounting for the sampling error alone (Fig. 1j). The former estimate is not significantly different from the African average given by Lewis et al. (2013), whereas the latter is.

Apart from the residual error and that due to the model choice, most sources of error in the current study were presumably undervalued compared to what would be obtained at a landscape level. Because spatial autocorrelation of biomass is positive, the estimate of per-hectare biomass from  $n$  contiguous plots is less variable than its estimate from  $n$  independent plots of the same size. Therefore, the sampling error obtained with the 9 contiguous 1-ha plots at Yoko is presumably smaller than what would be obtained with 9 independent 1-ha plots scattered over the landscape. We already saw that the plot-model interaction was also presumably undervalued. Because we used a constant wood density for all species, the measurement error was also undervalued. Finally, the error due to the uncertainty on the model's coefficient was presumably undervalued because we used a uniform distribution in the simulated data sets of

the Monte Carlo method, whereas original data sets used to fit (1) and (2) included more small trees than large trees. Therefore, confidence intervals of aboveground biomass estimates using independent plots at a landscape level may even be greater than those reported here.

#### 4.7 Conclusion

Quantifying uncertainty in biomass estimates is a directive of the IPCC (Eggleston et al. 2006) and, following decision 4/CP.15 of the UNFCCC (2010), will also apply to the REDD+ mechanism. Failing to integrate the whole chain of error when estimating the per-hectare biomass from inventory data and allometric equations may artificially reduce the width of the confidence interval (Fortin and DeBlois 2010). The choice of the allometric model is a major component of the total error. It is maximum when all models are equally likely. To improve the precision of regional estimates of aboveground biomass in central Africa, we recommend to collect many data sets of tree biomass at many sites, with a moderate number of trees at each site. Using BMA, such data sets would enable to assign different weights to the predictions of the different models, and thus reduce the error due to the model choice.

#### Appendix A: Estimation method for BMA

Let  $\theta = (\pi_1, \dots, \pi_M, \sigma_1, \dots, \sigma_M)$  be the vector of BMA parameters to estimate. The BMA equation corresponds to a finite mixture model (McLachlan and Peel 2000). Estimation is achieved by maximizing the likelihood. As always with finite mixture models, the likelihood cannot be maximized analytically but can be maximized numerically using the EM algorithm (McLachlan and Krishnan 2008). The EM algorithm introduces "missing data"  $z_{mi}$  that are the posterior probabilities that biomass equation  $m$  is the best model for observation  $i$ . The EM algorithm starts with an initial guess  $\theta^{(0)}$  of the parameter vector  $\theta$  and then iteratively alternates between two steps. In the E (or expectation) step, the  $z_{mi}$  are estimated given the current guess for the parameters:

$$z_{mi}^{(j)} = \frac{\pi_m^{(j-1)} \phi(B_i; B_{mi}, \sigma_m^{(j-1)} D_i^2)}{\sum_{k=1}^M \pi_k^{(j-1)} \phi(B_i; B_{ki}, \sigma_k^{(j-1)} D_i^2)}$$

where  $D_i$  and  $B_i$  are the dbh and biomass of the  $i$ th tree of the training data set,  $B_{mi}$  is the predicted biomass of the  $i$ th tree of the training data set according to model  $m$ , and superscript  $j$  refers to the  $j$ th iteration of the EM algorithm.

The M (or maximization) step consists of estimating  $\pi_m$  and  $\sigma_m$  using as weights the current estimates of  $z_{mi}$ :

$$\pi_m^{(j)} = \frac{1}{n} \sum_{i=1}^n z_{mi}^{(j)}$$

$$\sigma_m^{(j)2} = \frac{\sum_{i=1}^n z_{mi}^{(j)} [(B_i - B_{mi})/D_i^2]^2}{\sum_{i=1}^n z_{mi}^{(j)}}$$

where  $n$  is the number of observations in the training data set. The E and M steps were iterated until the  $L_1$  norm of  $\theta^{(j)}$  did not change by more than six decimal places in one iteration.

### Appendix B: Decomposition of the total error

Let  $\mathcal{B}_{kmj}(s)$  be the per-hectare biomass for the  $k$ th subplot predicted using model  $m$  at the  $j$ th iteration of the Monte Carlo algorithm. Total error (3) breaks into:

$$S_{\text{tot}}^2(s) = \frac{1}{K} \sum_{k=1}^K [\bar{\mathcal{B}}_{k\bullet\bullet}(s) - \bar{\mathcal{B}}(s)]^2 + \sum_{m=1}^M \pi_m [\bar{\mathcal{B}}_{\bullet m\bullet}(s) - \bar{\mathcal{B}}(s)]^2$$

$$+ \frac{1}{K} \sum_{k=1}^K \sum_{m=1}^M \pi_m [\bar{\mathcal{B}}_{km\bullet}(s) - \bar{\mathcal{B}}_{k\bullet\bullet}(s) - \bar{\mathcal{B}}_{\bullet m\bullet}(s) + \bar{\mathcal{B}}(s)]^2$$

$$+ \frac{1}{KJ} \sum_{k=1}^K \sum_{m=1}^M \sum_{j=1}^J \pi_m [\mathcal{B}_{kmj}(s) - \bar{\mathcal{B}}_{km\bullet}(s)]^2 \quad (4)$$

where:

$$\bar{\mathcal{B}}_{k\bullet\bullet}(s) = \frac{1}{J} \sum_{m=1}^M \sum_{j=1}^J \pi_m \mathcal{B}_{kmj}(s)$$

$$\bar{\mathcal{B}}_{\bullet m\bullet}(s) = \frac{1}{KJ} \sum_{k=1}^K \sum_{j=1}^J \mathcal{B}_{kmj}(s)$$

$$\bar{\mathcal{B}}_{km\bullet}(s) = \frac{1}{J} \sum_{j=1}^J \mathcal{B}_{kmj}(s)$$

The first term in (4) is the sampling error. The second term is the error due to the model choice. The third term represents the interaction between the plot and the allometric equation. The fourth term encompasses all remaining sources of error, i.e. the prediction error and the error on the model predictors.

### References

Alvarez E, Duque A, Saldarriaga J, Cabrera K, de las Salas G, del Valle I, Lema A, Moreno F, Orrego S, Rodríguez L (2012) Tree above-ground biomass allometries for carbon stocks estimation in the natural forests of Colombia. For Ecol Manag 267:297–308. doi:10.1016/j.foreco.2011.12.013

Angelsen A, Brockhaus M, Sunderlin WD, Verchot LV (eds.) (2012). Analysing REDD+: Challenges and choices. CIFOR Bogor, Indonesia

Banin L, Feldpausch TR, Phillips OL, Baker TR, Lloyd J, Affum-Baffoe K, Arets EJMM, Berry NJ, Bradford M, Brienen RJW, Davies S, Drescher M, Higuchi N, Hilbert DW, Hladik A, Iida Y, Abu Salim K, Kassim AR, King DA, Lopez-Gonzalez G, Metcalfe D, Nilus R, Peh KSH, Reitsma JM, Sonké B, Taedoumg H, Tan S, White L, Wöll H, Lewis SL (2012) What controls tropical forest architecture? Testing environmental, structural and floristic drivers. Global Ecol Biogeogr 21:1179–1190. doi:10.1111/j.1466-8238.2012.00778.x

Basuki TM, van Laake PE, Skidmore AK, Hussin YA (2009) Allometric equations for estimating the above-ground biomass in tropical lowland dipterocarp forests. For Ecol Manag 257:1684–1694. doi:10.1016/j.foreco.2009.01.027

Boyemba Bosela F (2011) Écologie de Pericopsis elata (Harms) Van Meeuwen (Fabaceae), arbre de forêt tropicale africaine à répartition agrégée. PhD thesis Université Libre de Bruxelles, Bruxelles

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 262:1648–1657. doi:10.1016/j.foreco.2011.07.018

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. doi: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 B 359:409–420. doi:10.1098/rstb.2003.1425

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

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

Cochran WG (1977) Sampling Techniques, 3rd edn. Wiley series in probability and mathematical statistics. Wiley, New York

Cohen R, Kaino J, Okello JA, Bosire JO, Kairo JG, Huxham M, Mencuccini M (2013) Propagating uncertainty to estimates of above-ground biomass for Kenyan mangroves: A scaling procedure from tree to landscape level. For Ecol Manag 310:968–982. doi:10.1016/j.foreco.2013.09.047

Cunia T (1987) Error of forest inventory estimates: its main components. In: Whraton EH, Cunia T (eds) Estimating tree biomass 797 regressions and their error Part E, USDA For. Serv., Northeast. 798 For. Exp. Sta., Broomall, PA, USA, Gen. Tech. Rep. NE-117, 799 pp 114

Djomo AN, Knohl A, Gravenhorst G (2011) Estimations of total ecosystem carbon pools distribution and carbon biomass current annual increment of a moist tropical forest. For Ecol Manag 261:1448–1459. doi:10.1016/j.foreco.2011.01.031

Dorjisa S, Durrieu de Madron L, Fontez B, Giraud A, Riera B (2011) Établissement d' équations entre le diamètre et le volume total

- de bois des arbres, adaptés au Cameroun. *Bois For Trop* 65:87–95
- Ebuy Alipade J, Lokombé Dimandja JP, Ponette Q, Sonwa D, Picard N (2011) Biomass equation for predicting tree aboveground biomass at Yangambi. DRC. *J Trop For Sci* 23:125–132
- Eggleston S, Buendia L, Miwa K, Ngara T, Tanabe K (eds) (2006) 2006 IPCC guidelines for national greenhouse gas inventories. Agriculture, forestry and other land use, vol 4 institute for global environmental strategies on behalf of the intergovernmental panel on climate change. Hayama, Japan
- Fayolle A, Doucet JL, Gillet JF, Bourland N, Lejeune P (2013) Tree allometry in Central Africa: Testing the validity of pantropical multi-species allometric equations for estimating biomass and carbon stocks. *For Ecol Manag* 305:29–37. doi:10.1016/j.foreco.2013.05.036
- Feldpausch TR, Lloyd J, Lewis SL, Brien RJW, Gloor M, Monteagudo Mendoza A, Lopez-Gonzalez G, Banin L, Abu Salim K, Affum-Baffoe K, Alexiades M, Almeida S, Amaral I, Andrade A, Aragão LEOC, Araujo Murakami A, Arets EJMM, Arroyo L, Aymard CGA, Baker TR, Bánki OS, Berry NJ, Cardozo N, Chave J, Comiskey JA, Alvarez E, de Oliveira A, Di Fiore A, Djagbletey G, Domingues TF, Erwin TL, Fearnside PM, França MB, Freitas MA, Higuchi N, Honorio CE, Iida Y, Jiménez E, Kassim AR, Killeen TJ, Laurance WF, Lovett JC, Malhi Y, Marimon BS, Marimon-Junior BH, Lenza E, Marshall AR, Mendoza C, Metcalfe DJ, Mitchard ETA, Neill DA, Nelson BW, Nilus R, Nogueira EM, Parada A, Peh KSH, Pena Cruz A, Peñuela MC, Pitman NCA, Prieto A, Quesada CA, Ramírez F, Ramírez-Angulo H, Reitsma JM, Rudas A, Saiz G, Salomão RP, Schwarz M, Silva N, Silva-Espejo JE, Silveira M, Sonké B, Stropp J, Taedoung HE, Tan S, ter Steege H, Terborgh J, Torello-Raventos M, van der Heijden GMF, Vásquez R, Vilanova E, Vos VA, White L, Willcock S, Woell H, Phillips OL (2012) Tree height integrated into pantropical forest biomass estimates. *Biogeosciences* 9:3381–3403. doi:10.5194/bg-9-3381-2012
- Fortin M, DeBlois J (2010) A statistical estimator to propagate height prediction errors into a general volume model. *Can J For Res* 40:1930–1939. doi:10.1139/X10-107
- Gibbs HK, Brown S, Niles JO, Foley JA (2007) Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environ Res Lett* 2:1–13. doi:10.1088/1748-9326/2/4/045023
- GOFC-GOLD (2012) A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals associated with deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation. GOFC-GOLD Report version COP18-1. GOFC-GOLD Land Cover Project Office, Wageningen University, The Netherlands
- Gonzalez P, Kroll B, Vargas CR (2014) Tropical rainforest biodiversity and aboveground carbon changes and uncertainties in the Selva Central, Peru. *For Ecol Manag* 312:78–91. doi:10.1016/j.foreco.2013.10.019
- Henry M, Besnard A, Asante WA, Eshun J, Adu-Bredu S, Valentini R, Bernoux M, Saint-André L (2010) Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of Africa. *For Ecol Manag* 260:1375–1388. doi:10.1016/j.foreco.2010.07.040
- Hunter MO, Keller M, Vitoria D, Morton DC (2013) Tree height and tropical forest biomass estimation. *Biogeosci Discuss* 10:10 529:491–10. doi:10.5194/bgd-10-10491-2013
- Kearsley E, de Haulleville T, Hufkens K, Kidimbu A, Toirambe B, Baert G, Huygens D, Kebede Y, Defourny P, Bogaert J, Beeckman H, Steppe K, Boeckx P, Verbeeck H (2013) Conventional tree height-diameter relationships significantly overestimate aboveground carbon stocks in the central Congo Basin. *Nat Commun* 4:2269. doi:10.1038/ncomms3269
- Keller M, Palace M, Hurr G (2001) Biomass estimation in the Tapajos National Forest, Brazil. Examination of sampling and allometric uncertainties. *For Ecol Manag* 154:371–382. doi:10.1016/S0378-1127(01)00509-6
- Lewis SL, Sonke B, Sunderland T, Begne SK, Lopez-Gonzalez G, van der Heijden GMF, Phillips OL, Affum-Baffoe K, Banin L, Bastin JF, Beeckman H, Boeckx P, Bogaert J, DeCanniere C, Chezeau E, Clark CJ, Collins M, Djagbletey G, Droissart V, Doucet JL, Feldpausch TR, Foli E, Gillet JF, Hamilton AC, de Haulleville T, Hladik A, Harris DJ, Hart TB, Hufkens K, Huygens D, Jeanmart P, Jeffrey K, Kamdem MN, Kearsley E, Leal ME, Llloyd J, Lovett J, Makana JR, Malhi Y, Marshall AR, Ojo L, Peh KSH, Pickavance G, Poulsen J, Reitsma JM, Sheil D, Simo M, Steppe K, Taedoung HE, Talbot J, Taplin J, Taylor D, Thomas SC, Toirambe B, Verbeeck H, Votere R, White LJT, Wilcock S, Woell H, Zemagho L (2013) Aboveground biomass and structure of 260 African tropical forests. *Philos Trans R Soc Lond B* 368:20120,295. doi:10.1098/rstb.2012.0295
- Li Y, Andersen HE, McGaughey R (2008) A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *West J Appl For* 23:223–231
- Massart P (2007) Concentration Inequalities and Model Selection École d'Été, de Probabilités de Saint-Flour XXXIII – 2003, No. 1896 in Lecture Notes in Mathematics. Springer-Verlag, Berl in Heidelberg
- McLachlan G, Peel D (2000) Finite Mixture Models. Wiley Series in Probability and Statistics. John Wiley & Sons, New York NY
- McLachlan GJ, Krishnan T (2008) The EM Algorithm and Extensions Wiley Series in Probability and Statistics, 2nd edn. John Wiley & Sons, Hoboken NJ
- Melson SL, Harmon ME, Fried JS, Domingo JB (2011) Estimates of live-tree carbon stores in the Pacific Northwest are sensitive to model selection. *Carbon Balance Manag* 6. doi:10.1186/1750-0680-6-2
- Molto Q, Rossi V, Blanc L (2013) Error propagation in biomass estimation in tropical forests. *Method Ecol Evol* 4:175–183. doi:10.1111/j.2041-210x.2012.00266.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, Botsika Bobéoré KH, Mikala Okouyi C, Nyangadouma R, Lépengué N, Mbatchi B, Picard N (2014) Site-specific versus pantropical allometric equation: Which option to estimate the biomass of a moist central african forest. *For Ecol Manag* 312:1–9. doi:10.1016/j.foreco.2013.10.029
- Picard N, Henry M, Mortier F, Trotta C, Saint-André L (2012) Using Bayesian model averaging to predict tree aboveground biomass. *For Sci* 58:15–23. doi:10.5849/forsci.10-083
- Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna Austria. <http://www.R-project.org>, ISBN3-900051-07-0
- Raftery AE, Gneiting T, Balabdaoui F, Polakowski M (2005) Using Bayesian model averaging to calibrate forecast ensembles. *Mon Weather Rev* 133:1155–1174. doi:10.1175/MWR2906.1
- Smith RL, Tebaldi C, Nychka D, Mearns LO (2009) Bayesian modeling of uncertainty in ensembles of climate models. *J Am Statist Assoc* 104:97–116. doi:10.1198/jasa.2009.0007
- UNFCCC (2010) Report of the conference of the parties on its fifteenth session, held in Copenhagen from 7 to 19 December 2009.

Addendum. Part two: Action taken by the conference of the parties at its fifteenth session. Decisions adopted by the Conference of the Parties. FCCC/CP/2009/11/Add.1, United Nations Framework Convention on Climate Change

Wagner F, Rutishauser E, Blanc L, Hérault B (2010) Effects of plot size and census interval on descriptors of forest structure and dynamics. *Biotropica* 42:664–671. doi:[10.1111/j.1744-7429.2010.00644.x](https://doi.org/10.1111/j.1744-7429.2010.00644.x)

Zeide B (1980) Plot size optimization. *For Sci* 26:251–257