Journal of Tropical Forest Science 29(3): 282–296 (2017)
https://doi.org/10.26525/jtfs2017.29.3.282296

CONSTRUCTION OF MULTISPECIES ALLOMETRIC EQUATIONS: IS THERE A STATISTICAL PALLIATIVE FOR DESTRUCTIVE TREE SAMPLING?

TB Mayaka[1, *], R Eba'a-Atyi[2] & ST Momo[3, 4]

[1]*Faculty of Science, University of Dschang, PO Box 67 Dschang, Cameroon*
[2]*Center for International Forestry Research, PO Box 2008, Messa, 17 Yaoundé, Cameroon*
[3]*Laboratory of Plant Systematics and Ecology, Department of Biology, Higher Teachers' Training College, University of Yaoundé I, PO Box 047, Yaoundé, Cameroon*
[4]*Institut de Recherche pour le Développement, Unité Mixte de Recherche Boulevard de la Lironde, TA A51/PS2, 34398 Montpellier CEDEX 5, France*

*th.mayaka@gmail.com

Submitted May 2016; accepted November 2016

To avoid the problem of destructive tree sampling, we tested an indirect estimation procedure whereby aboveground tree volumes are estimated using affine-transforms of traditional bole volume equations. The study compared the predictive performance of the (1) proposed procedure, (2) allometric aboveground volume equations, and (3) simple upscaling of empirical tree bole volumes (serving as benchmark procedure). The study assessed species- and tree-individual random deviations from mean procedure effect on prediction errors. Six volume equations were each fitted to aboveground volume (α- version) and bole volume (β- version); predictions of latter volumes were affine-transformed to estimate aboveground tree volumes. Bole height, total height and diameter at breast height (dbh) were measured for 59 trees from 10 species in eastern Cameroon. The Schumacher and Hall equation ranked first in quality-of-fit. The direct and indirect approaches applied with this volume equation predicted the aboveground tree volumes equally well across all species (groupings), with bias (± RMSE of 0.153 ± 2.512 and 0.178 ± 2.56 m^3 respectively. Finally, trees within species accounted for 49.3% of the total variability in volume prediction error vs only 3.5% for species. Recommendations have been made for improvement relatively to data requirements and model building.

Keywords: Aboveground volume, affine transform, belowground volume, conjugate models, jackknife estimator, predictive performance, volume equation

INTRODUCTION

Tropical rainforests are the world's most important carbon sink for mitigating the impact of carbon emission on global climate change (Houghton et al. 2009, Marshall et al. 2012). These forests are the main focus of the international policy to 'reduce emissions from deforestation and degradation' in developing countries (REDD+) (Kuyah et al. 2012, Marshall et al. 2012, Angelsen et al. 2013) under which forest carbon stocks must be estimated with as much accuracy and precision as possible. This requirement has rekindled interest in the construction of allometric equations for estimating aboveground tree biomass (Zianis & Mencuccini 2004, Picard et al. 2012a, Kuyah et al. 2012, Ploton et al. 2015).

However, ecological and methodological constraints have impeded the development of allometric equations globally, with a particular acuteness in the Congo Basin. Ecologically, the contribution of the Congo Basin forests to the global carbon cycle is difficult to assess due to several sources of uncertainty. These include (1) high plant diversity, often exceeding 300 different species ha^{-1} (Lewis et al. 2004), (2) high variability in forest types, wood densities, tree heights, biomass and volume (Feldpausch et al. 2011, Lewis et al. 2013), (3) poor knowledge of natural stand dynamics such as growth, mortality and recruitment (Picard et al. 2012b, Mayaka et al. 2014) and (4) human disturbances (logging and deforestation) responsible for releasing greenhouse effect gas (IPCC 2007, Marshall et al. 2012).

Methodologically, while the development of tree biomass equations is embedded in the theoretical frame of regression analysis, it suffers from sampling process errors. In effect the data used in constructing biomass equations typically include small tree diameters (Ketterings et al. 2001, Ebuy et al. 2011). This is because the process is not only destructive as trees must be felled, but it is also lengthy and cumbersome as trees must be cut into compartments that are weighed and sampled for determining their wood specific density. However, these comprehensive tree measurements have not eliminated the main difficulties inherent to regression analysis. Such difficulties include the choice and inclusion of independent variables in sufficient number and appropriate metrics and lack of standard models together with the difficulty in model selection (Schabenberger & Pierce 2002). Inadequate tackling of these issues is likely to result in prediction errors, further compounding the bias due to high proportion of small trees in the samples. The multifactor variability in tree size, volume and biomass observed between locations is readily handled using linear mixed effects models (Pinheiro & Bates 2000, Venables & Ripley 2002, Bolker et al. 2008). The predictor variables used in multispecies allometric equations are, in decreasing order of importance, tree stem diameter , woody density, tree height and forest type, all of which, wood density aside, have traditionally been used to construct volume equations (Avery & Burkhart 2002).

How could tree volume (hence biomass and carbon stocks) be extrapolated in situations where destructive sampling of trees is prohibited? Five such situations can be envisaged as follows: firstly, protected areas (parks, reserves and sanctuaries) do not allow any use forms that may conflict with the preservation of the biological diversity of tropical rainforest. Secondly, vast swathes of forested lands remain inaccessible for lack of economic interest, inadequate logistics and road infrastructure, or difficult terrain such as marshes and mountains. Thirdly, only traditional dendrometric measurements (diameter and height) may be obtained from some individual trees. Fourthly, destructive sampling of (mostly small) trees may cause too much disruption in the sustainable management of forest concessions, especially under reduced-impact logging schemes (Mayaka et al. 2014). Fifthly, the need may occasionally arise to estimate the volume or biomass of already logged trees for policing or seeking compensation from illegal harvesting.

The objectives of this study were to propose a statistical approach for extrapolating stem volume to total aboveground volume (up to group 1 branches with basal diameter ≥ 20 cm) when destructive sampling is prohibited and to assess which proportion of the total variability in volume prediction error is due to species and individual trees (within species).

MATERIALS AND METHODS

Notation and terminology

Throughout this paper, volume will be considered in relation to five tree compartments, namely, stump, butt swell, buttresses (if present), log and tree top. Tree top consists of the stem part above the main tree bole plus branches which have basal diameters of 20 cm or larger. The term aboveground volume will refer to total volume of the five tree compartments. A scaling function refers to any transformation that maps bole volume to aboveground volume. A particular class of scaling function is the affine function expressed as $\varphi(x) = \varphi_0 + \varphi_1 \times x$, where x = arbitrary argument and (φ_0, φ_1) is a pair of real numbers. In this work, extrapolation is defined as a simple procedure whereby an independent tree measurement (D*, H*) together with a previously fitted volume equation are used to predict the tree bole volume $\hat{V}_b^*$ which is then mapped onto tree aboveground volume $\hat{V}_a^E$ using a scaling function, as illustrated in the bottom part of Figure 1.

The notations used are as follows:

(1) D and H are tree diameter (at breast height or aboveground buttresses if present) and height respectively

(2) V_a and V_b are respectively the aboveground and bole volumes

(3) $G = \{g_u, u = 1, 2, \ldots 6\}$ is a set of six contending models for predicting both volume types

(4) $V_{au} = g_u\ (D, H; \alpha_{pu})$ and $V_{bu} = g_u\ (D, H; \beta_{pu})$ are conjugate models used for predicting aboveground volume and bole volume respectively; also referred to as α- and β-versions of model, g_u, where $\alpha_{pu} = (\alpha_1, \ldots, \alpha_{pu})$ and $\beta_{pu} = (\beta_1, \ldots, \beta_{pu})$ are p_u-dimensional vectors of unknown regression parameters and p_u = the number of dimensions for parameter in model u.

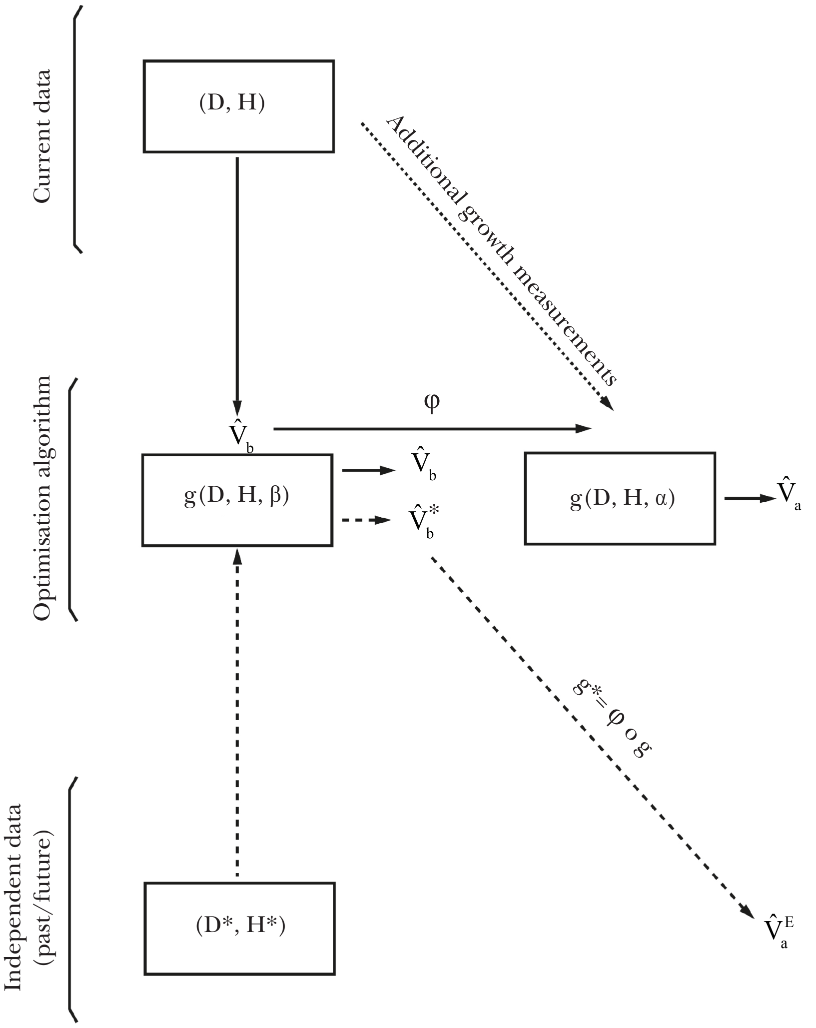

Figure 1 The proposed indirect approach for predicting aboveground tree volumes ($\hat{V}_a^E$) in relation to the direct approach for predicting aboveground and bole volumes ($\hat{V}_a$ and $\hat{V}_b$ respectively)

(5) $\hat{\varphi}_{-ts}$, $\hat{\alpha}_{-ts}$ and $\hat{\beta}_{-ts}$ are the jackknife estimators of model parameters, leaving out tree t (t = 1, 2,..., n_s) of species s (s = 1, 2,..., S).

Study site

The study was conducted in the Kadeï Division, east region of Cameroon, within a forest concession (3° 53'–3° 56' N, 14° 48'–14° 51' E) of 86,096 ha at an elevation range of 600–850 m above sea level. The vegetation forms part of the semi-deciduous Guinea-Congolese dense tropical rainforest with high species richness. The wet equatorial climate exhibits an annual average temperature of 25 °C, an average annual precipitation of 1500 mm (with range ± 100 mm), and four seasons (resulting from the influence of the monsoon and harmattan), which are a long dry season (November–March), a short dry season (June–August), a light wet season (March–June) and a heavy wet season (August–November). Ferralsols (ferrallitic red soils) are common in the concession, except in flood plains and swamps where alluvial and colluvial sediments overlay hydromorphic soils.

Study species

The study used 59 sample trees earmarked for felling and belonging to 10 different species (see details in Table 1). Five species provided at least five trees each—ayous (*Triplochiton scleroxylon*), sapelli (*Entandrophragma cylindricum*), iroko (*Milicia excelsa*), tali (*Erythrophleum ivorense*), and okan (*Cylicodiscus gabunensis*) and the other five species—abam (*Chrysophyllum beguei*), frake (*Terminalia superba*), ilomba (*Pycnanthus angolensis*), padouk (*Pterocarpus soyauxii*), and lati (*Amphimas pterocarpoides*) were pooled in a remainder group.

Table 1 Composition of the study sample

Species	Sample size
Ayous (*Triplochiton scleroxylon*, Malvaceae)	22
Sapelli (*Entandrophragma cylindricum*, Meliaceae)	9
Iroko (*Milicia excelsa*, Moraceae)	9
Tali (*Erythrophleum ivorense,* Caesalpiniaceae)	7
Okan (*Cylicodiscus gabunensis*, Fabaceae)	5
Remainder[1]	7
Total[2]	59

[1]The remainder group comprised seven trees including three lati (*Amphimas ferrugineus*, Fabaceae) and one tree of each of the following four species: abam (*Chrysophyllum beguei*, Sapotaceae), frake (*Terminalia superba*, Combretaceae), ilomba (*Pycnanthus angolensis*, Myristicaceae) and padouk (*Pterocarpus soyauxii*, Fabaceae); [2]the actual model fitting used one observation less (i.e. ayous tree identified as outlier in Figure 2)

Tree measurements

Tree measurements included diameter at breast height (dbh, D) measured at 1.30 m above the ground or 30 cm above buttresses using diameter tape and total height was obtained by adding the lengths of the different tree sections: stump, butt swell (if present), bole and crown. The following approaches were used in computing volumes of tree components.

(1) Length (L) together with the average of two perpendicular diameters were recorded at lower ($\bar{d}_l$) and upper ($\bar{d}_u$) ends of the stump and butt swell. Assuming the stump to be cylindrical, its volume was obtained as $\pi \times L \times (\bar{d}_l + \bar{d}_u)^2/4$.

(2) Bole diameters were measured at 2-m intervals and the resulting volume was computed using the Smalian formula. Total bole volume was obtained as the sum of volumes of individual bole sections.

(3) Volume of a buttress was obtained as $(4 - \pi) L_b H_b W_b)/12$ (Henry et al. 2010), where L_b, H_b, and W_b, are the length, height and width of the buttress respectively.

(4) Branches in the crown were subdivided into two groups (Henry et al. 2010) depending on whether basal diameter was 20 cm and above (group 1) or less than 20 cm (group 2). The stem portion above the bole and the branches in group 1 were measured at 1-m interval (whenever possible) and their volume obtained using the same formula as for the stump.

Statistical techniques

Statistical techniques used were modelling assumptions, model fitting and selection and predictive performance assessment. All computations were performed using version 3.0.2 of software R (2013).

Modelling assumptions

We make the usual assumptions of linear and nonlinear regression models:

(1) Covariates (dbh and bole height) are free of measurement errors.

(2) Error terms, ε, are a Gaussian variables identically and independently distributed (iid) with mean zero and variance σ_ε^2.

(3) Regression parameters form a vector of unknown constant values to be estimated by an optimisation method such as the ordinary least squares for linear regression or, alternatively, the Gauss–Newton and Newton–Raphson approaches to least squares in case of nonlinear regression (Bates & Watts 1988, Schabenberger & Pierce 2002, Venables & Ripley 2002).

For the purpose at hand, we further assumed the following:

(1) Aboveground volume is an affine function of the bole volume, i.e. $V_a = \varphi_0 + \varphi_1 \times V_b + \epsilon$, where $\epsilon \sim$ iid $N(0, \sigma^2)$.

(2) $G = \{g_u, u = 1, 2, \ldots 6\}$ is a set of nested models that is closed under the affine transformation φ, i.e. for all $u = 1, 2, \ldots 6$, g_u is nested in some

g_{u+k} with $k \geq 0$ and $\varphi(g_u) = g_u^*$ is also a member of the set G, given some reparameterisation.

(3) Fnctions g_u ($u = 1, 2, \ldots 6$) are each twice differentiable with respect to regression parameters.

Model fitting and selection

We compared six of the most widely used volume equations in the forestry literature (Clutter et al. 1983, Avery & Burkart 2002, Fonweban et al. 1995, 2012). These equations are functions of stem diameter D and height H of the generic form g_u (D, H; θ_{pu}), $u = 1, 2, \ldots, 6$ which are displayed in Table 2 together with model nesting pattern, typology and image under the affine transformation. Three approaches lend themselves in predicting the aboveground volumes using stem diameter and height.

Approach 1 (benchmark): Using the scaling function, *in casu* the affine regression model (AR), the following aboveground volume predictions and prediction errors were derived:

$$\hat{V}_{ats}^{AR} = \hat{\varphi}\left(V_{bts}\right) = \hat{\varphi}_0 + \hat{\varphi}_1 \times V_{bts} \tag{1}$$

$$d_{ts}^{AR} = V_{ats} - \left(\hat{\varphi}_0 + \hat{\varphi}_1 \times V_{bts}\right) \tag{2}$$

where subscripts a and b refer aboveground and belowground volumes respectively, and subscripts t and s index trees and species respectively.

Alternatively, a mechanistic regression model is fitted to aboveground (V_a) and bole (V_b) volumes, giving a pair of conjugate models, henceforth referred to as α- and β-versions of the model. More precisely,

$$\begin{cases} V_{auts} = g_u\left(D_{ts},\ H_{ts};\ \alpha_{p_u}\right) + \varepsilon_{uts},\ \alpha\text{−version} \\ V_{buts} = g_u\left(D_{ts},\ H_{ts};\ \beta_{p_u}\right) + \varepsilon_{uts},\ \beta\text{−version} \\ \qquad \left(1 \leq t \leq n_s, 1 \leq s \leq S\right) \end{cases} \tag{3}$$

where all terms are as previously defined with reference to tree t of species s and S = 6, total number of species in the study.

Approach 2: This was a direct prediction of the aboveground volume using the α-version of model in equation 3, from which the prediction errors were derived as:

$$\begin{aligned} d_{ts}^{Mu} &= V_{ats} - \hat{V}_{ats} \\ &= V_{ats} - g_u\left(D_{ts}, H_{ts}; \hat{\alpha}_{p_u}\right) \end{aligned} \tag{4}$$

where superscript Mu refers to estimation error obtained with model u.

Approach 3: This was an indirect prediction of the aboveground volume in two steps, using the β-version of model in equation 3 to predict the bole volume which was then expanded into the aboveground volume using the scaling function in equation 1. The aboveground volumes predicted were:

$$\begin{aligned} \hat{V}_{auts}^{E} &= \hat{\varphi}\left(\hat{V}_{buts}\right) \\ &= \hat{\varphi}_0 + \hat{\varphi}_1 \times g_u\left(D, H; \hat{\beta}_{p_u}\right) \\ &= g_u^*\left(D_{ts}, H_{ts}; \hat{\beta}_{p_u^*}^*\right) \end{aligned} \tag{5}$$

Table 2 Form, number of parameters, nesting pattern, regression type and affine transform of the compared multispecies volume equations

u	Form of equation	p_u	Nesting model	Regression type	Affine transform	
					$\varphi_0 = 0$	$\varphi_0 \neq 0$
1	$V_1 = \theta_0 + \theta_1 D^2$	2	V_5	Linear	V_1	V_1
2	$V_1 = \theta_0 + \theta_1 D^2 H$	2	V_6	Linear	V_2	V_2
3	$V_3 = \theta_1 D^{\theta_2}$	2	V_4	Intrinsically linear	V_3	V_5
4	$V_4 = \theta_1 D^{\theta_2} H^{\theta_3}$	3	V_6	Intrinsically linear	V_4	V_6
5	$V_5 = \theta_0 + \theta_1 D^{\theta_2}$	3	V_6	Nonlinear	V_5	V_5
6	$V_6 = \theta_0 + \theta_1 D^{\theta_2} H^{\theta_3}$	4	V_6	Nonlinear	V_6	V_6

The compared multispecies volume equations are of the generic form $V_u = g_u$ (D, H; θ_{pu}), where D = dbh (cm), H = bole height (m), u = indexing integer ($1 \leq u \leq 6$) and θ_{pu} = vector-valued parameter of dimension p_u; also shown for each model are the nesting model, regression type and image under the affine transformation $\varphi(x) = \varphi_0 + \varphi_1(x)$; model 3 is nested in both models 4 (under H_0: $\theta_3 = 0$) and 5 (under H_0: $\theta_0 = 0$)

for which the extrapolation errors were as follows:

$$\begin{aligned} d_{ts}^{Eu} &= V_{ats} - \hat{V}_{auts}^{E} \\ &= V_{ats} - g_u^{*}\left(D_{ts}, H_{ts}; \hat{\beta}_{p_u^*}^{*}\right) \end{aligned} \qquad (6)$$

where $\hat{\beta}_{p_u^*}^{*} = \psi\left(\hat{\beta}_{p_u}\right)$ is some reparameterisation induced by the scaling function $\varphi(\cdot)$.

The proposed indirect approach is equivalent to a first-order Taylor's approximation (Dudewicz & Mishra 1988) for small differences $\alpha - \hat{\beta}^{*}$. We recommend the systematic upscaling of the bole volume predictions when the distance between the estimated α- and β- parameter vectors is larger than 5% of the length of the estimated α-parameter vector, i.e. $\left\|\hat{\alpha}_{p_u} - \hat{\beta}_{p_u}\right\| / \left\|\hat{\alpha}_{p_u}\right\| \geq 0.05$ as further expounded in Appendix 1.

We used information criteria (ICs) for model selection (Burnham & Anderson 2002), including the most widespread Akaike's information criterion, $AIC = -2l + 2P$, its finite- size correction $AIC_c = AIC + \frac{2P(P+1)}{n-P-1}$, and the second most common but more conservative Schwarz or Bayesian information criterion (BIC) = $-2l + P \times \log(n)$, where l = log-likelihood, $n = \sum_s n_s = 59$ is the sample size, and P = number of model parameters (the variance of error terms included, i.e. $P = p_u + 1$). Approaches 2 and 3 were applied with each of the k (i.e. the number of models that pass the selection procedure by AIC criterion) selected models ($k < 6$) giving a total of $2 \times k$ factorial combinations which, when combined with the benchmark (approach 1), formed an augmented factorial design with $(2 \times k + 1)$ compared procedures.

Predictive performance assessment

The main purpose of this study was to assess whether the aboveground volume extrapolation (approach 3) is nearly as good as the allometric tree volume equation (approach 2) or even a simple upscaling of the empirical bole volume (approach 1 or benchmark). Given the small sample size of 59 trees, the jackknife or leave-one-out cross-validation (LOOCV) technique was used to assess the predictive performances of the $(2 \times k + 1)$ contending procedures. The LOOCV consists in fitting repeatedly a model by leaving out each observation in turn, thus providing for the omitted observation a prediction that is independent from the training data (Stone 1974, Arlot & Celisse 2010). Using jackknife prediction errors in which the jackknife estimators of model parameters $\hat{\varphi}_{-ts}$, $\hat{\alpha}_{-ts}$ and $\hat{\beta}_{-ts}$ are substituted respectively into equations 2, 4, and 6, the prediction bias or mean errors (in m^3) were calculated as follows (all superscripts and unnecessary subscripts have been dropped for simplicity): $B_s = \sum_{t=1}^{n_s} d_{ts}/n_s$ (for species s) and $B = \sum_{s=1}^{S} w_s B_s$ (over all species, where $w_s = n_s / \sum_1^S n_s$ is the weight for species s). The precision of prediction (also in m^3) was measured by the root mean square errors $RMSE = \sqrt{B_s^2 + V_s}$ for species s and $\sqrt{\sum_s w_s\left(B_s^2 + V_s\right)}$ over all species.

In keeping with our second purpose (which was to assess which proportion of the total variability in volume prediction error was due to species and individual trees within species) stated in the introduction, we estimated the variance components of species and trees within species on the prediction errors. The prediction procedures enter the following ANOVA model as levels of a fixed-effect factor (Pinheiro & Bates 2000):

$$d_{pts} = \mu + \tau_p + r_s + r_{ts} + \varepsilon_{pts} \qquad (7)$$

where the response variable d_{pts} = prediction error obtained with procedure p ($p = 1, 2, \ldots, 2 \cdot k + 1$) on tree t ($t = 1, 2, \ldots, n_s$) of species s ($s = 1, 2, \ldots, S$), μ = overall average prediction error, τ_p = fixed effect of procedure on prediction error, $r_s \sim \text{iid}\, N\left(0, \sigma_S^2\right)$ = species–individual random deviations from the mean prediction error μ, $r_{ts} \sim \text{iid}\, N\left(0, \sigma_S^2\right)$ = random deviations of individual trees (nested in species) from mean effect of procedure p and $\varepsilon_{pts} \sim \text{iid}\, N\left(0, \sigma_\varepsilon^2\right)$ = random residual errors. Again, for sake of simplicity, the augmented factorial design structure is not shown in the above ANOVA model equation (see Schaarschmidt and Vaas (2009) for detailed statistical analysis guidelines). Finally, the algorithm of the proposed prediction approach is detailed in Appendix 2.

RESULTS

The summary statistics of tree diameter and heights (Appendix 3) and the bole-to-aboveground volume ratios (Appendix 4) are shown for contextualisation. Ayous was the largest species in the data set and it exhibited the widest range of variation in terms of dbh ($\bar{D}$ = 1.27 ± 0.30 m) and total height ($\bar{H}_t$ = 49.3 ± 3.0 m). By contrast, tali had the smallest tree sizes ($\bar{D}$ = 0.95 ± 0.16 m and $\bar{H}_t$ 43.6 ± 6.2 m). The species in the remainder group had dbh of 1.05 ± 0.08 m and height 47.1 ± 3.8 m. The estimated ratios of bole volume-to-aboveground volume (i.e. up to group branches of basal diameter ≥ 20 cm) ranged from 0.81 (ayous) to 0.90 (iroko) and did not differ significantly from 0.84 for all species (groupings) according to the 95% percentile confidence interval.

Model fits and assessment

The bivariate and marginal distributions of both volume types are shown in Figure 2, which clearly indicates an outlier: an ayous tree. By discarding the outlier, the relation between aboveground and bole volumes became linear and strong. The fitted line overlaid on the scatter plot had coefficients (with standard errors) $\hat{\varphi}_0 = -2.255 \pm 0.894$ ($p < 0.05$) and $\hat{\varphi}_1 = 1.336 \pm 0.052$ ($p < 0.001$, and a coefficient of determination $r^2 = 0.92$. Table 3 shows the fit of α- and β-versions for all six contending models, including parameter estimates and information criteria. Models 4 and 3 ranked respectively first and second. The distance between estimates of α- and β-parameter vectors relative to the length of α-parameter vector for both selected models was twice the 5% threshold (10.7 and 10.1% for models 4 and 3 respectively), thus requiring an adjustment of bole volume predictions.

Predictive performance assessment

The mean prediction errors (with root mean square errors in parentheses) for the compared procedures are presented in Table 4. The benchmark procedure underestimated the aboveground volume for all species, except for ayous and tali. The predictions based on model 3 whether adjusted (β-version) or not (α-version) were the least precise and least

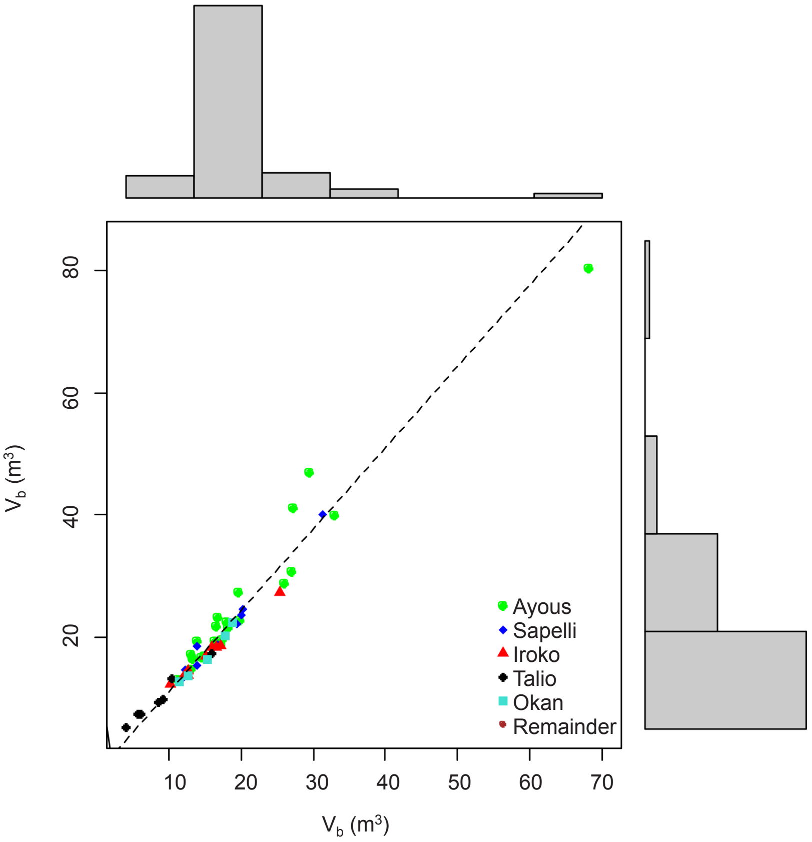

Figure 2 Bivariate and marginal distributions of bole volume (V_b) and aboveground volume (V_a) the obvious outlier among ayous trees was omitted from model fitting

Table 3 Model fits (including estimated parameters with standard errors in parentheses) and assessment based on quality-of-fit (using information criteria AIC, AIC_c and BIC for initial ranking) together with predictive performance (using mean errors and root mean square errors of prediction for final ranking)

Model		Parameter estimate[1]				Information criteria[2]			Rank
Version	Index(u)	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_3$	$\hat{\theta}_4$	AIC	AIC_c	BIC	
α	1	0.26 (1.34)	14.24*** (0.94)	-	-	318.2	312.7	324.4	5
	2	0.94 (0.80)	0.52*** (0.02)	-	-	267.9	262.3	274.0	3
	3	2.64*** (0.03)	2.07*** (0.16)	-	-	-14.5	-20.0	-8.3	2
	4	0.09 (0.26)	1.92*** (0.10)	0.79*** (0.08)	-	-71.3	-78.6	-63.1	1
	5	-0.83 (6.83)	15.3* (6.64)	1.90** (0.56)	-	320.2	312.9	328.4	6
	6	1.41 (2.80)	0.73 (0.56)	2.09*** (0.28)	0.88*** (0.18)	269.2	260.4	279.5	4
β	1	2.99** (1.11)	9.84*** (0.77)	-	-	295.8	290.2	301.9	6
	2	3.37*** (0.75)	0.36*** (0.02)	-	-	260.8	255.2	266.9	3
	3	2.50*** (0.03)	1.87*** (0.16)	-	-	-12.1	-17.6	-5.9	2
	4	0.10 (0.30)	1.73*** (0.11)	0.74*** (0.09)		-55.1	-62.2	-46.9	1
	5	-14.16 (19.95)	26.86 (19.90)	0.94 (0.58)	-	294.9	287.6	303.1	5
	6	-3.85 (5.37)	2.79 (2.50)	1.35*** (0.33)	0.55** (0.18)	260.2	251.4	270.5	4

[1]The parameter estimates for models 3 and 4 are those of the fitted log-linear regression; [2]the model selection used preferably the finite-size correction of the Akaike's information criterion (AIC_c) and the Bayesian information criterion (BIC) as the sample size of the training set (n = 58, after one outlier was omitted) relative to the number of parameters ($3 \leq P \leq 5$) is less than 40 (see text for details); *, **, *** = significant at probability levels of 5, 1 and 0.1% respectively

accurate but differed only slightly between them. By comparison, the predictions obtained with model 4, also known as Schumacher and Hall volume equation (Schumacher & Hall 1933), were more accurate and precise compared with the other procedures (Figure 3). Direct and indirect approaches based on this model predicted the aboveground tree volume equally well across all species (groupings), giving 0.153 ± 2.512 and 0.178 ± 2.560 m^3 (bias ± RMSE) respectively. However, both procedures poorly predicted the aboveground volumes of ayous trees (RMSE = 3.093 and 3.223 m^3 respectively for direct and indirect procedures).

Larger variability of prediction errors was observed among trees within species than between species (Figure 3). Variance component estimates were as follows $\hat{\sigma}_S^2 = 0.3316$, $\hat{\sigma}_T^2 = 4.6412$ and $\hat{\sigma}_\varepsilon^2 = 4.4480$ adding up to 9.4208. Hence, trees within species accounted for 49.3% of the total variation as against 3.5% for species.

DISCUSSION

Model output, selection and predictive performance

This study included aboveground volumes up to group 1 branches, i.e. branches with basal diameter ≥ 20 cm. The proposed approach relied on the assumed linear relationship between aboveground and bole volumes, i.e. $V_a = \varphi_0 +$

Table 4 Mean bias (in m^3, with root mean square errors in parentheses) of three prediction approaches of tree aboveground volumes

Approach[1]	Procedure		Species[2]						Overall
	Model	Adjustment	Ayous	Sapelli	Iroko	Tali	Okan	Rem[3]	
Benchmark	AR	No	0.706 (3.592)	-0.214 (1.231)	-1.232 (2.003)	0.668 (1.587)	-1.023 (1.212)	-0.057 (1.638)	-0.192 (2.507)
Direct	M4, α	No	0.156 (3.093)	0.066 (2.413)	-0.192 (1.803)	0.044 (2.525)	-0.508 (0.886)	1.334 (2.231)	0.150 (2.512)
	M3, α	No	1.536 (4.00)	-0.600 (4.954)	0.895 (2.356)	-3.315 (4.228)	-2.410 (3.510)	2.372 (3.656)	-0.253 (3.911)
Indirect	M4, β	Yes	0.044 (3.223)	0.050 (2.342)	-0.301 (1.777)	0.540 (2.652)	-0.50 (0.855)	1.237 (2.111)	0.178 (2.560)
	M3, β	Yes	1.524 (4.075)	-0.592 (4.808)	0.861 (2.312)	-3.193 (4.181)	-2.502 (3.594)	2.361 (3.650)	-0.257 (3.906)

[1] The compared approaches form an augmented factorial design of five procedures including α- and β-versions of models 3 and 4 (M3 and M4) plus the benchmark affine regression (AR)-based approach serving to upscale bole volume predictions from β-versions of models, the bottom two lines correspond to the proposed indirect approach; [2] see Table 1 for scientific names and sample sizes; [3] Rem = group of remainder species (see Table 1 for composition)

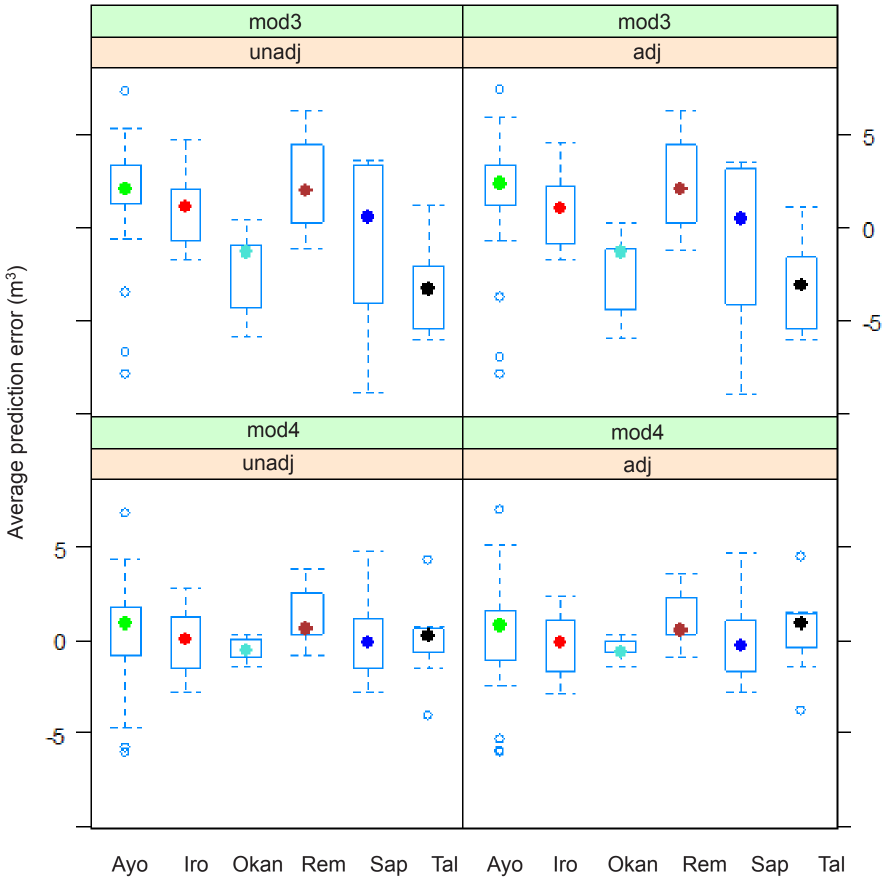

Figure 3 Box plot of the distribution of prediction errors by species (groupings) conditional on prediction procedures; prediction procedures are factorial combinations of adjusted (adj) and unadjusted (unadj) models (mod) 3 and 4; species names are abbreviated as follows: Ayo = ayous, Iro = iroko, Rem = remainder group (including lati, abam, frake, ilomba and padouk), Sap = sapelli, Tal = tali

$\varphi_1 \times V_b$ as verified with the dataset in this study (Figure 2). Empirically, the linear relationship would seem reasonable when the bole-to-aboveground volume was around 0.84 as in this study (but see Ploton et al. 2015 and further below). Theoretically the linearity assumption is not only useful but also realistic as it probably holds independently of the structural and morphological characteristics of trees (Bentley et al. 2013). Indeed, this seemed to be the case since from a simple dimension analysis, the volume of a tree is $V \propto D^2 H$ of which the bole volume is an approximation. On the other hand, $V \propto D^d H^h$ $(2 < d + h < 3)$ from fractal geometry analysis since it is acknowledged that a positive number between 2 and 3 estimated tree crown dimension better and it is also assumed that the "overall shape of a tree (stem and crown) may posses [sic] similar fractal dimension" (Zianis & Mencuccini 2004).

Linear relationship above provided a benchmark to assess other approaches for predicting aboveground volume. For null intercept , the prediction approach is equivalent to ratio estimation, namely $\hat{V}_a = \hat{\varphi}_1 \times V_b$ (Mayaka et al. 2014). This simpler more intuitive rule consists of using slope of regression through the origin for expanding bole volume into aboveground volume. Alternatively, the scaling function may be a power function, namely, $V_a = \varphi_0 V_b^{\varphi_1}$ with the value of φ_1 $(1 \leq \varphi_1 < 2)$ depending on the sum d + h of the exponents of tree diameter and height referred to above. The proposed approach still applies when scaling function and allometric equations are power functions, provided the logarithm metrics is used. Indeed, taking model 4 (which reduced to model 3 for $\theta_3 = 0$), bole volumes $V_b = \beta_1 D^{\beta_2} H^{\beta_3}$ would be mapped onto aboveground volumes $V_a = \beta_1^* D^{\beta_2^*} H^{\beta_3^*}$ or, equivalently, $\log V_a = \log\left(\beta_1^*\right) + \beta_2^* \log D + \beta_3^* \log H$ with the following adjusted β-parameters $\log\left(\beta_1^*\right) = \log\varphi_0 + \varphi_1 \log\beta_1$, $\beta_2^* = \varphi_1 \beta_2$ and $\beta_3^* = \varphi_1 \beta_3$.

The outcome of model selection and ranking was clear-cut, with models 4 and 3 ranking respectively first and second, well above the others, by the three information criteria AIC, AIC_c, and BIC. The final model ranking obtained with cross-validation did not differ from that obtained initially with information criteria. The advantage of cross-validation over information criteria is its quasi-universality: cross-validation generally yields good model selection performances provided data are identically and independently distributed (Arlot & Celisse 2010). However, this comparative advantage comes at higher computational cost and lesser accuracy in selection frameworks where penalised criteria (of which information criteria form a special class) has been designed to be optimal, such as least-squares regression (Arlot & Celisse 2010).

Further reduction in the number of selected models prescribed in step 5 of the algorithm (Appendix 2) was neither necessary nor possible on practical and procedural grounds. From a practical standpoint, only two models were selected on the first selection round out of six (which was a small number to begin with). From a procedural perspective, the images of models 3 and 4 under the affine transformation (i.e. models 5 and 6 respectively) did not pass the first selection round.

The performance of model 4 corroborates previous results obtained with *Eucalyptus saligna* stands in Cameroon and Kenya (Shiver & Brister 1990, Fonweban et al. 1995) and further accounts for the widespread and successful application that model, i.e. Schumacher and Hall equation, has enjoyed in forestry (Bailey 1994, Fonweban et al. 1995, 2012). In contrast, despite its good reputation, equation 2, usually referred to as the combined variable equation (Spurr 1952), performed poorly on both fit quality and predictive performance. This result confirmed earlier arguments that inclusion of tree height and crown area improved accuracy of allometric equations that were based on tree stem diameter alone (Ketterings et al. 2001).

Species contributed less to the variability in prediction errors than did trees within species, implying that mixed-effects modelling techniques should be an integral part of the proposed prediction procedure. Direct and indirect predictions of aboveground volumes obtained across species using the Schumacher and Hall's equation were equally precise. The indirect approach was slightly more biased by 0.025 m^3 or 16.3% compared with the direct approach, a discrepancy that was inconsequential in light of potential ecological gains. Finally, all three approaches (benchmark, α-model and β-model) readily apply to estimation of carbon stock, since tree biomass equations are identical to volume equations up to approximately a constant which is either multiplicative or additive depending on the metrics used. In the case of logarithm metrics, various aboveground biomass equations analogous of our models 3 and 4 can be seen elsewhere (Kuyah et al. 2012). The results of this study cannot, however, be extended to the whole crown (i.e. beyond group 2 branches with basal diameter < 20 cm) nor generally to larger trees in which crown mass represent 50% or more of the total tree biomass (Ploton et al. 2015).

Perspectives and recommendations

We examined ways to improve our proposed prediction procedure with respect to data requirements and model fitting as well as validation.

Data requirements

The poor predictive performance we recorded on species with smallest sample sizes (i.e. species in the remainder group) suggested that sufficient number of trees should be sampled on a small number of species rather than the reverse. Ayous was the study species with the largest sample size (n = 22) and also the largest trees. Thus, we recommend sampling two dozen trees per reference species over the variation range of the two key dendrometric variables (dbh and total height). Furthermore, allocation of the sample size should be representative of the typical frequency distribution of tree size as measured by diameter, or perhaps even better, the height-to-diameter ratio for the species concerned. In other words the shape of the histogram should reflect that of the theoretical frequency distribution of tree size which, in the case of diameter, is typically either bell-shaped (ayous, sapelli, and tali) or reversed-J (padouk and ilomba) (Durrieu de Madron & Forni 1997, Picard et al. 2012b). Alternatively, when grouping is done by architecture or morphology types, sampling should be representative of the frequency distribution of tree size for the grouping factor. The predictive performance of pantropical allometric equations can be further improved through the use of habitat (location) covariables. The latter may comprise soil characteristics (depth, texture, C:N ratio, pH and sum of base cations; Marshall et al. 2012, Lewis et al. 2013) and climate (rainfall and temperature; Lewis et al. 2013).

The suggestion to increase sample sizes (and number of sites as indicated below) may conflict with our stated objective to limit the destructive sampling of trees. However, the contradiction is only apparent, as an optimal estimation procedure will contribute in keeping such destruction within acceptable bounds while meeting statistical thresholds (power, bias and precision) acceptable to all parties.

Modelling techniques

We examined three specific issues in connection to small sample sizes and multifactor variability. Firstly, since allometric equations will continue to use a relatively limited number of trees per species and other classificatory factors, bootstrapping can be useful in calculating the precision of estimated model parameters (Efron & Tibshiriani 1993). Secondly, variability of tree volume and biomass is due to various ecological, environmental and human factors (Marshall et al. 2012, Lewis et al. 2013). Accordingly, the construction of allometic equations should include sites and trees nested in species as random effect factors. Alternatively, trees may be grouped by any other meaningful factors such as architectural, morphological or functional types. On the other hand, using models with appropriate error covariance structure improves statistical inference through deflated residual variance, better precision and increased power of tests, provided the survey design and underlying biological mechanisms are heeded (Davidian & Giltinan 1995, Pinheiro & Bates 2000). Thirdly, the main effects and interaction of habitat (i.e. location) covariables mentioned earlier and noted hereafter ($X_1,\ldots X_K$) are readily included in the mean response of allometric equations. This inclusion is, however, contingent on the form of the volume equation considered:

(1) For models 1,2, 5 and 6, by simply replacing θ_0, the intercept or linear parameter as the case may be, with $\gamma\ (X_1,\ldots X_K) = \gamma_0 + \sum_k \gamma_k X_k + \sum_{k \geq h} \sum_h \gamma_{kh} X_h X_k$, where γ_0 = intercept (overall mean effect), γ_k = main effect of X_K and γ_{kh} = interaction between X_K and X_h.

(2) For models 3 and 4, replace the linear parameter θ_1 with $e^{\gamma\ (X1,\ldots XK)}$, where $\gamma\ (X_1,\ldots X_K)$ is as just described.

Formally, including habitat covariates in model 4 leads to the volume equation

$$\ln V_{lstk} = (\gamma_0 + r_l + r_s + r_{ts}) + \sum_k \gamma_k X_k + \sum_{k \geq h} \sum_h \gamma_{kh} X_h X_k + \theta_2 \ln D + \theta_3 \ln H \qquad (8)$$

where all terms are as previously defined, r_l, r_s and r_{ts} are random deviations from intercept γ_0 respectively for individual location l, individual species s and individual trees t nested within species s (alternatively trees may be grouped by morphology types). The bole volume predictions obtained with the β-version of the model in equation 8 would then be adjusted to obtain aboveground volume by (1) using a scaling function as already indicated and (2) incorporating the best linear unbiased predictors of the random variables r_l and r_s which are significantly different from zero. In case tree biomass is the focus, the proposed prediction procedure readily applies as indicated earlier.

The cross-validation performed in this study is but one of the validation procedures available to modellers for testing the predictive performance of their models (Stone 1974, Arlot & Celisse 2010). The preliminary model ranking and selection based on information criteria is optimal and should be used preferably over cross-validation for the purpose of weeding up and cutting down computational costs, especially when comparing a large number of models. However, the predictive performance of models with subsequent independent data should use bias and RMSE as criteria. In cases where destructive sampling is either banned or impractical as premised in this study, actual volume should be obtained by alternative means such as LiDAR-t (Dassot et al. 2011, Kaasalainen et al. 2014, Hosoi et al. 2013a, b).

CONCLUSIONS

This study proposed an indirect approach for extrapolating aboveground tree volume using affine transformation of bole volume predictions, in an attempt to avoid the destructive sampling of trees when building multispecies allometric equations. The proposed approach was equivalent to the direct allometric equations in estimating aboveground volume across species (groupings), in terms of prediction bias and accuracy obtained with the Schumacher and Hall equation. The use of linear mixed-effects models revealed that trees within species contributed the greatest variability to volume prediction errors. The proposed approach readily applies to the prediction of aboveground tree biomass (and hence carbon stock) through the inclusion of wood density as constant, whether in a multiplicative or additive manner. It is suggested that multi-species and multi-location models incorporate habitat covariates to account for the variability of tree volume that is related to ecological and environmental factors. Together with other non-destructive methods, the proposed approach complements current efforts in building multispecies allometric equations, especially in situations where trees cannot be felled due to multiple constraints (imposed by conservation and management options as well as limited logistics and field inaccessibility).

ACKNOWLEDGEMENTS

Destructive data from Cameroon were collected with the financial support from the IRD project PPR FTH-AC 'Changements Globaux, Biodiversité et Santé en Zone Forestière d'Afrique Centrale' and the support from the Alpicam Company. The insightful comments of JN Fonweban and GN Shu on the earlier drafts of the manuscript are highly acknowledged.

REFERENCES

Angelsen A, Brockhaus M, Sunderlin WD & Verchot L. 2013. Introduction. Pp. 1–12 in Angelsen A et al. (eds) *Analysing REDD+: Challenges and Choices*. CIFOR, Bogor.

Arlot S & Celisse A. 2010. A survey of cross-validation procedures for model selection. *Statistics Surveys* 4: 40–79.

Avery TE & Burkhart HE. 2002. *Forest Measurements. Fifth edition.* McGraw Hill, New York.

Bailey R. 1994 A compatible volume-taper model based on the Schumacher and Hall generalized constant form factor equation. *Forestry Science* 40: 303–313.

Bates DM & Watts DG. 1988. *Nonlinear Regression Analysis and Its Application.* Wiley & Sons, New York.

Bentley LP, Stegen JC, Savage VM et al. 2013. An empirical assessment of tree branching networks and implications for plant allometric scaling models. *Ecology Letters* 16: 1069–1078.

Bolker BM, Brooks ME, Clark CJ et al. 2008. Generalized linear mixed-models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution* 24: 127–135.

Burnham KP & Anderson DR. 2002. Information and likelihood theory: a basis for model selection and inference. Pp. 49–97 in Burnham KP & Anderson DR (eds) *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach. Second Edition.* Springer Science, New York.

Clutter JL, Fortson JC, Pienaar LV, Brister GH & Bailey RL. 1983. *Timber Management: A Quantitative Approach.* John Wiley & Sons, New York.

Davidian M & Giltinan DM. 1995. *Nonlinear Models for Repeated Measurements.* Chapman and Hall, New York.

Dassot M, Constant T & Fournier M. 2011. The use of terrestrial LiDAR technology in forest science: application fields, benefits and challenges. *Annals of Forest* Science 68: 959–974.

Dudewicz EJ & Mishra SN. 1988. *Modern Mathematical Statistics*. Wiley & Sons, New York.

Durrieu de Madron L & Forni E. 1997. Aménagement forestier dans l'est du Cameroun: structure du peuplement et périodicité d'exploitation. *Bois et Forêts des Tropiques* 254: 39–50.

Ebuy J, Lokombe Dimandja JP, Ponette Q, Sonwa DJ & Picard N. 2011. Allometric equation for predicting aboveground biomass of three tree species. *Journal of Tropical Forest Science* 23: 125–132.

Efron B & Tibshiriani RJ. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, London.

Feldpausch TR, Babnin L, Phillips OL et al. 2011. Height-diameter allometry of tropical forest trees *Biogeosciences* 8: 1081–1106.

Fonweban JN, Mayaka TB & Seukep JB. 1995. Construction and validation of tree volume equations for *Eucalyptus saligna* in Cameroon. *Commonwealth Forestry Review* 74: 355–360.

Fonweban J, Gardiner B & Auty D. 2012. Variable-top merchantable volume equations for Scots pine (*Pinus sylvestris*) and Sitka spruce (*Picea sitchensis* (Bong.) Carr.) in Northern Britain. *Forestry*. doi:10.1093/forestry/cpr069.

Henry M et al. 2010. Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of Africa. *Forest Ecology and Management* 260: 1375–1388.

Hosoi F, Nakai Y & Omasa K. 2013a. 3-D voxel-based solid modeling of a broad-leaved tree for accurate volume estimation using portable scanning LIDAR. *ISPRS Journal of Photogrammetry and Remote Sensing* 82: 41–48.

Hosoi F, Nakai Y & Omasa K. 2013b. Voxel tree modeling for estimating leaf area density and woody material volume using 3-D LIDAR data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science* II-5/W2: 115–120.

Houghton RA, Hall F & Goetz SJ. 2009. Importance of biomass in the global carbon cycle. *Journal of Geophysical Research* 114: G00E03. doi:10.1029/2009JG000935.

IPCC (Intergovernmental Panel on Climate Change). 2007. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Parry ML et al. (eds). Cambridge University Press, Cambridge.

Kaasalainen S, Krooks A, Liski J et al. 2014. Change detection of tree biomass with terrestrial laser scanning and quantitative structure modelling. *Remote Sensing* 6: 3906–3922.

Ketterings QM, Coe R, Noordwijk VM, Ambagau Y & Palm CA. 2001. Reducing uncertainty in the use of allometric biomass equations for predicting aboveground tree biomass in mixed secondary forests. *Forest Ecology and Management* 146: 199–209.

Kuyah S, Dietz J, Muthuri C et al. 2012. Allometric equations for estimating biomass in agricultural landscapes. I: Aboveground biomass. *Agriculture, Ecosystems and Environment* 158: 216–224.

Lewis SL, Malhi Y & Phillips OL. 2004. Fingerprinting the impacts of global change on tropical forests. *Philosophical Transactions of the Royal Society B* 359: 437–462.

Lewis SL, Sonké B, Sunderland T et al. 2013. Aboveground biomass and structure of 260 African tropical forests. *Philosophical Transactions of the Royal Society B* 368: 20120295. https://dx.doi.org/10.1098/rstb.2012.0295.

Marshall AR, Willcock S, Platts PJ et al. 2012. Measuring and modelling above-ground carbon and tree allometry along a tropical elevation gradient. *Biological Conservation* 154: 20–33.

Mayaka TB, Eba'a-Atyi R & Nkie MC. 2014. On volume recovery index and implications for sustainable logging in Congo Basin. *Forest Ecology and Management* 313: 292–299.

Picard N, Saint-Andre L & Henry M. 2012a. *Manuel de Construction d'Équations Allométriques Pour l'Estimation du Volume et la Biomasse des Arbres: De la Mesure de Terrain à la Prédiction*. Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Rome and Montpellier. http://www.fao.org/docrep/018/i3058f/i3058f.pdf

Picard N, Gourlet-Fleury S & Forni E. 2012b. Stock recovery rates are not the panacea to assess timber yield sustainability: evidence from managed Central African forests. *Forest Ecology and Management* 281: 12–22.

Pinheiro JC & Bates DM. 2000. *Mixed Effects Models in S and S-Plus*. Springer, New York.

Ploton P, Barbier N, Momo ST et al. 2015. Closing a gap in tropical forest biomass estimation: accounting for crown mass variation in pantopical allometries. *Biogeosciences Discussions* 12: 19711–19750. doi:10.5194/bgd-12-19711-2015.

Schaarschmidt F & Vaas L. 2009. Analysis of trials with complex treatment structure using multiple contrast tests. *HortScience* 44: 188–195.

Schabenberger O & Pierce FJ. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press, New York.

Schumacher FX & Hall FS. 1933. Logarithmic expression of timber tree volume. *Journal of Agricultural Research* 47: 719–734.

Shiver BD & Brister GH. 1990. Tree and stand volume functions for *Eucalyptus saligna* in Western Kenya. Pp 200–210 in Burkhart HE (ed) *IUFRO Proceedings on Research in Forest Mensuration, Growth, and Yield*. 5–11 August 1990, Montreal.

Spurr SH. 1952. *Forest Inventory*. Ronald Press, New York.

Stone M. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society (Series B)* 36: 111–147.

Venables WN & Ripley BD. 2002. *Modern Applied Statistics with S*. Fourth edition. Springer, New York.

Zianis D. & Mencuccini M. 2004. On simplifying allometric analyses of forest biomass. *Forest Ecology and Management* 187: 311–332.

Appendix 1

Analytical and geometrical features of the proposed approach

This appendix further examines a number of analytical and geometrical issues inherent to the proposed approach for extrapolating tree aboveground volume. To begin with, since G the set of compared models is closed under the affine transformation φ, both g_u and its image g_u^* belong to G. Next, the jackknifed prediction errors $d_{ts}^{Eu} = V_{ats} - \hat{\varphi}\left(\hat{V}_{buts}\right) = V_{ats} - g_u^*\left(D, H, \hat{\beta}_{-ts}^*\right)$ are approximated by the first-order Taylor's series expansion of $g_u^*(\cdot) = \varphi(g_u(\cdot))$ namely, $\sum_{j=1}^{p_u^*} \frac{\partial g_u^*}{\partial \alpha_j}\left(\hat{\beta}_j^*\right)\left(\alpha_j - \hat{\beta}_j^*\right)$ to the extent that the differences $\left(\alpha_j - \hat{\beta}_j^*\right)$, $j = 1, 2, \ldots, p_u^*$ are sufficiently small, or equivalently, if in the Euclidian space of dimension p_u^*, the vectors $\hat{\beta}^*$ and α are close. This condition implies that $\hat{\beta}^*$ is also close to $\hat{\alpha}$, by virtue of the least squares estimation method. Now, the image of an arbitrary argument x under the affine transformation φ is either $\varphi_1 \cdot x$ or $\varphi_0 + \varphi_1 \cdot x$, according as the null hypothesis H_0: $\varphi_0 = 0$ is is rejected at 5% probability level or less. In either case, the affine transformation induces an adjustment of the estimated parameter vector such that $\hat{\beta}^* = \psi\left(\hat{\beta}\right)$. By way of illustration, the adjusted bole volume prediction under growth model 4 is $\hat{\varphi}\left(\hat{V}_4\right) = \hat{\varphi}_0 + \hat{\varphi}_1 \times \hat{V}_4$, which upon expansion yields $\hat{\varphi}_0 + \left(\hat{\varphi}_1 \hat{\beta}_1\right) D^{\hat{\beta}_2} H^{\hat{\beta}_3} \equiv g_6^*\left(D, H; \hat{\beta}^*\right)$. An interesting special case is the identity function $\varphi \equiv I$, where the image of an argument is the argument itself, i.e. $\varphi(x) = x$, meaning that the estimated parameter need not be scaled up. This may be the case only if, for all practical purpose, the distance $\hat{\Delta}(\alpha, \beta)$ between $\hat{\beta}$ and $\hat{\alpha}$ is very negligible compared to the distance $\hat{\Delta}(\alpha, 0)$ between 0 and $\hat{\alpha}$, say $\hat{\Delta}(\alpha, \beta) / \hat{\Delta}(\alpha, 0) < 5\%$. The empirical distances are obtained for tree $t(t=1,2,\ldots,n_s)$ of species $s(s=1,2,\ldots,S)$ respectively as $\hat{\Delta}_{-ts}(\alpha, \beta) = \left\|\hat{\alpha}_{-ts} - \hat{\beta}_{-ts}\right\|$ and $\hat{\Delta}_{-ts}(\alpha, 0) = \hat{\alpha}_{-ts} - 0$ using the jackknife technique. The ratio of average distances must therefore be less than the threshold value, in casu $\bar{\Delta}\left(\alpha, \beta\right) / \bar{\Delta}\left(\alpha, 0\right) < 5\%$, where $\bar{\Delta}\left(\alpha, \beta\right) = \sum_{s=1}^{S} \sum_{t=1}^{n_s} \hat{\Delta}_{-ts}\left(\alpha, \beta\right) / \sum_{s=1}^{S} n_s$, and $\bar{\Delta}(\alpha, 0)$ is computed analogously.

Appendix 2

Algorithm for implementing the proposed approach

The algorithm for implementing the proposed prediction approach is as follows.

Step 1: Fit α- and β-models to the complete data set, omitting possible outliers and influential observations;

Step 2: Rank and select the α- and β-models separately, based on information criteria, AIC, AIC_c and BIC;

Step 3: Obtain the jackknife prediction errors for the $2 \times k + 1$ prediction procedures and compute the precision and accuracy for each procedure;

Step 4: Use the prediction accuracy (or bias, B) and precision (RMSE) obtained in step 3 to update the model ranking in step 2;

Step 5: Further downsize the set of previously selected models using the following rule: from the initial subset of selected β-models obtained in step 2, take model g_u such that $g_u^* = \varphi(g_u)$, so far as g_u^* had passed at least the first round of selection for α-models (based on information criteria, ICs); alternatively, from the initial subset of selected α-models take model g_u^* such that $g_u = \varphi^{-1}(g_u^*)$, so far as g_u has passed at least the first round of selection for β-models (based on information criteria, ICs).

Appendix 3

Summary of tree measurement data

Species		Dbh		H_b		H_t	
	n	Range	$\bar{X} \pm SD$	Range	$\bar{X} \pm SD$	Range	$\bar{X} \pm SD$
Ayous	22	0.88–2.12	1.27 ± 0.30	20.6–37.1	29.3 ± 3.8	45.3–57.3	49.3 ± 3.0
Sapelli	9	1.00–1.78	1.21 ± 0.23	18.3–32.3	26.2 ± 4.4	42.5–52.7	48.5 ± 3.8
Iroko	9	0.94–1.26	1.07 ± 0.10	25.6–29.6	27.9 ± 1.2	45.6–55.6	49.6 ± 2.8
Tali	7	0.68–1.13	0.95 ± 0.16	13.8–20.8	17.5 ± 2.7	32.6–51.6	43.6 ± 6.2
Okan	5	1.03–1.36	1.16 ± 0.13	17.6–27.5	23.4 ± 3.9	45.9–49.6	48.4 ± 1.5
Remainder[1]	7	0.95–1.15	1.05 ± 0.08	24.7–33.1	28.0 ± 3.2	42.9–53.2	47.1 ± 3.8

The following are included for each species: sample size (n), in addition to variation range and $\bar{X} \pm SD$ of diameter at breast height (Dbh, m), bole height (H_b, m), bole volume (V_b, m^3) and total volume (V_t, m^3); [1]the remainder group comprises seven trees (three lati and one tree of each of abam, frake, ilomba and padouk)

Appendix 4

Bootstrap estimates of bole to aboveground volume ratio

Ratio of bole volume to aggregate volume for the whole data are given by species (groupings) and for all species combined

Species	n	(bole to aboveground volume ratio)[1]				
		Estimate	Bias[2]	Var	MSE[3]	95% CI[4]
Ayous	22	0.81	3.25	4.71	4.71	(0.76, 0.85)
Sapelli	9	0.84	9.50	3.19	3.20	(0.80, 0.87)
Iroko	9	0.90	-4.42	0.90	0.91	(0.88, 0.92)
Tali	7	0.87	-33.93	6.61	6.72	(0.82, 0.92)
Okan	5	0.89	3.82	3.43	3.43	(0.86, 0.93)
Remainder	7	0.85	5.84	7.92	7.92	(0.79, 0.90)

[1]Overall ratio estimate was obtained using a thousand bootstrap resamples stratified by species; [2]: average of the ratio statistic of 1,000 bootstrap replicates minus the estimate of that ratio statistic in the data; [3]; [4]limits of the 95% confidence interval (CI) are respectively the 25th and 975th ordered values of the 1000 resample estimates of the ratio statistic of interest (Efron & Tibshiriani 1993).