Forest Ecology and Management 329 (2014) 237–254

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Review

A critical review of forest biomass estimation models, common mistakes and corrective measures

Gudeta W. Sileshi *

World Agroforestry Centre, Southern Africa Programme, Chitedze Research Station, P.O. Box 30798, Lilongwe, Malawi

ARTICLE INFO

Article history:
Received 1 April 2014
Received in revised form 21 June 2014
Accepted 24 June 2014
Available online 18 July 2014

Keywords:
Allometry
Confirmation bias
Freedman paradox
Isometry
Hypothetico-deductive
Uncertainty

ABSTRACT

The choice of biomass estimation models (BEMs) is one of the most important sources of uncertainty in quantifying forest biomass and carbon fluxes. This review was motivated by many mistakes and pitfalls I encountered in the recent literature regarding BEMs. The most common mistakes were the arbitrary choice of analytical methods, model dredging and inadequate model diagnosis, ignoring collinearity, uncritical use of model selection criteria and uninformative reporting of results. Sometimes, errors in parameter estimates were not checked and model uncertainty was ignored when interpreting and reporting results. Consequently, biologically implausible and statistically dubious equations such as $ln(M) = ln(a) + b(lnD) + c(lnD)^2 + d(lnD)^3 + e(ln\rho)$ have been published as allometric models. These are perpetuated in the literature, databases and field manuals and will pose a serious threat to the integrity of future forest biomass estimates. Through worked examples, I also illustrate that (1) allometric coefficients can be biased by the choice of analytical procedures and methodological artefacts; (2) collinearity of predictors can result in coefficients with unacceptable levels of error; (3) the R^2 and Akaike information criterion (AIC) have been misused and have resulted in the selection of implausible BEMs; and (4) differences in the definition of model "bias" has sometimes led to contradictory reports. I propose corrective measures for most of these problems and provide suggestions for prospective authors on how to avoid pitfalls in interpretation and reporting of results.

Contents

* Current address: 5600 Lukanga R, Lusaka, Zambia. Tel.: +26097835097.
E-mail addresses: sileshigw@gmail.com, sgwelde@yahoo.com

http://dx.doi.org/10.1016/j.foreco.2014.06.026
0378-1127/

1. Introduction

Estimation of forest biomass usually involves three steps: (1) application of biomass estimation equations (BEMs) to measurements of individual trees in forest inventory plots, (2) the summation of individual tree biomass to yield plot level estimates, and (3) the calculation of mean values across plots to yield a forest-level estimate (van Breugel et al., 2011). Since all these steps involve indirect methods, a large degree of uncertainty exits in estimated forest biomass, carbon stocks and fluxes (Ahmed et al., 2013; Clark and Kellner, 2012; Molto et al., 2013a; Somogyi et al., 2007; van Breugel et al., 2011). Lack of consensus on definitions, methodological inconsistencies and assumptions also lead to widely differing results even among similar studies (Clark and Kellner, 2012; Somogyi et al., 2007).

The selection of BEMs can influence local, regional and global biomass estimates. Recent studies demonstrates that the uncertainly in the selection of BEMs poses a practical limit to the accuracy with which remote sensing methods can predict regional biomass (Ahmed et al., 2013; Zhao et al., 2012). Since the choice of BEMs is the first critical step, there has been a rapid increase in efforts to develop locally appropriate equations. The IPCC guidelines (e.g. IPCC, 2003) and current standard operating procedures for carbon measurement (e.g. Walker et al., 2012) recommend the use of existing (previously published) equations where available. However, they do not provide guidance or standards for selecting model from the existing pool or developing new ones. Consequently, researchers involved in the development and application of BEMs are faced with many challenges: (1) the choice between existing equations and developing new ones; (2) the choice between species-specific and multi-species models; (3) the choice between simple bivariate power-law (typical allometric) functions and models with multiple predictors; and (4) the many predictors with ill-understood relationships when developing new models with multiple predictors.

The choice from existing models is challenging because there is a bewildering array of BEMs with varying degrees of complexity in the literature. For example, the GlobeAllomeTree database contains over 706 equations from Europe, 2843 from North America and 1058 from Africa (Henry et al., 2013). Sometimes several equations are available for one species, and having different equations for a single species would be a problem. The sheer number of equations also makes the choice of BEMs from databases a daunting task. The second problem arises because there are not enough species-specific equations for the species-rich tropical forests (Chave et al., 2005). Therefore, it is argued that multi-species pan-tropical models are more appropriate. The third challenge arises because the bivariate power-law models are believed to be too simple to deal with the idiosyncrasy of tropical forest biomass. Nevertheless, parameter estimation is challenging when several predictors are considered in a model. In addition, increasing mode complexity necessitates the use of large sample sizes and concurrent resource (time, labor) requirements. Even if sample sizes is not limiting, uncertainty is inevitable due to the incomplete knowledge about the data generating process, natural variability, measurement errors and modeling errors when several predictors are to be modeled. Lastly, when constructing new models the choice of predictors to be included in the final model is often challenging because the relationships between the many variables is not well understood.

As result of all these challenges, many mistakes and pitfalls in data analyses and interpretation occur when constructing BEMs. Sometimes, these mistakes pass unchecked or may even be difficult to check because the reporting is usually incomplete. When a BEM is published, usually it finds its way into databases such as GlobeAllomeTree (Henry et al., 2013) without information regarding the uncertainty about its parameters. This hinders informed choice of BEMs from databases. This review was motivated by these and many other concerns I have in the application of BEMs. The objective of this review is therefore to (1) point out the common mistakes and their consequences, and (2) provide suggestions for prospective authors to make better informed choice of models and for reviewers to objectively assess the merit of "new" models before they are published. For brevity, I will focus only on biomass equations, and I have excluded volume expansion equations.

2. Methods

This analysis started with a critical review of BEMs published in the last 10 years (2004–2014). In the course of the review, I identified several instances of errors in the estimation of power-law parameters and debate regarding discrepancies between empirical and theoretical values of the exponents of the mass-diameter and height-diameter allometric scaling. Therefore, I conducted a meta-analysis in order to highlight some of the difficulties in interpretation of the parameters. The meta-analysis was also intended to test which theoretical model has more support from empirical studies. I compiled a large dataset of allometric power-law parameters from existing databases in Henry et al. (2010), Jenkins et al. (2004), Návar (2009), Ter-Mikaelian and Korzukhin (1997), Zianis and Mencuccini (2004) and the more recent literature. In total there were 639 entries of the normalization constant (hereafter called intercept for brevity) and 765 of the exponent from equations across the globe that have estimated total aboveground biomass (M in kg) from stem diameter (D at 1.3 in cm). I extracted the power-law parameters either from what Návar (2009) calls nonlinear regression ($Y = aX^b$) or intrinsically linear regression ($ln(Y) = ln(a) + b(lnX)$. I then compared the empirical distribution of the exponent (b) with theoretical values with the aim of identifying the most likely theoretical model for the range of values published so far. I applied both normal and kernel density estimation to make inferences about the probability density distribution of the power-law intercept and exponent.

In order to examine potential influences of methodological artefacts on parameters, I explored the heterogeneity in the power-law coefficients with sample size and estimation methods. For this purpose, I re-analyzed datasets from Table 1 of Ebuy et al. (2011), Appendix A of Djomo et al. (2010), Table 1 of Henry et al. (2010), Tables S2–S7 of Mascaro et al. (2011), Appendix A of Nogueira et al. (2008) and Kuyah et al. (2012). For brevity, hereafter I will refer to these datasets as Ebuy, Djomo, Henry, Mascaro, Nogueira and Kuyah datasets. I selected these datasets because the requisite information was available online or directly from authors. In order to illustrate heterogeneity in allometric coefficients due to estimation methods, I fitted models to these datasets using nonlinear regression, ordinary least square (OLS) regression, robust regression, maximum likelihood and Bayesian analyses with Markov chain Monte Carlo (MCMC) simulation. For consistency of parameter estimation across datasets, I converted the biomass in Henry's dataset into kg per tree as it was originally reported in Mg per tree. This was necessary because changes in the measurement unit of M or D will change the estimates of the intercept.

In addition, I compiled parameters of BEMs with multiple predictors in order to examine how far published values depart from each other. In order to illustrate uncertainty around estimation of coefficients and the pervasive nature of influence statistics and collinearity, I fitted the commonly used BEMs to the data from Ebuy, Djomo, Henry and Kuyah. Using Henry's dataset, I also made more detailed illustrations of other problems encountered in

fitting BEMs. I chose Henry's dataset for this purpose as it is typical of the range of sample sizes used in the literature. Uncertainty around a model parameter (θ) is usually indicated by the standard error (SE). Better still is the percent relative standard errors (PRSE) defined as $PRSE = 100\left(\frac{SE}{|\theta|}\right)$ (Dumont et al., 2013; McCune and Grace, 2002). For ecological studies point estimates of θ are considered unreliable when PRSE is greater than 20% (McCune and Grace, 2002). In other statistical applications including demographic surveys and health statistics, point estimates of θ are considered unreliable if PRSE > 25%. This cut-off point has been used by the Office of Economic and Statistical Research in Australia and the U.S. National Center for Health Statistics and the Centre for Disease Control (CDC, 2010).

3. Synthesis

In this section I will discuss common mistakes that occur in the selection or development of BEMs. In order to provide background and ensure clarity in Section 3.2, I will first describe the classes of commonly used BEMs.

3.1. Classes of commonly used BEMs

The literature review revealed two main classes of BEMs in common use, namely (1) simple power-law BEMs and (2) models with multiple predictors.

3.1.1. Simple power-law BEMs

The power-law functions are the typical allometric equations. Allometry designates the relative change of one biological trait ($\Delta Y/Y$) in relation to the relative change of a second one ($\Delta X/X$). In order to avoid confusion in the field of relative growth, Huxley and Teissier (1936) agreed to consistently use the term allometry and the conventional power-law equation: $Y = aX^b$ where a is called normalization (proportionality) constant and b is the exponent. Generally, allometric scaling assumes a power function because it is supported by the notion of growth as a multiplicative process (Gayon, 2000; Marquet et al., 2005; Packard, 2014; Shea, 1985; West et al., 1999). Power-law relationships are characterized by scale invariance (self-similarity) and universality (Marquet et al., 2005). This allows predictions over a wide range of scales and levels of organizations, revealing the existence of universal principles within the seemingly idiosyncratic nature of biological and physical systems (Brown et al., 2002; Marquet et al., 2005). Allometric scaling has been established in many animal and plant traits and some appear to be universal, occurring in virtually all taxa of organisms and types of environments (Brown et al., 2002; Hendriks, 2007; Price et al., 2007).

In the rest of this review, I will refer to the simple power-law BEM $Y = aX^b$ or its logarithmic form

$$ln(Y) = ln(a) + b(lnX) \text{ or } log(Y) = log(a) + b(logX) \quad \text{(BEM1)}$$

A common mistake among the studies that used BEM 1 (e.g. Basuki et al., 2009; Djomo et al., 2010; Ebuy et al., 2011; Fayolle et al., 2013; Lima et al., 2012; Moore, 2010; Návar, 2009; Ngomanda et al., 2014; Picard et al., 2012) is that the log-linear form of this equation has been erroneously presented as $ln(Y) = a + b(lnX)$, i.e. instead of $ln(a)$ the intercept term was written as "a".

The intercept is hypothesized to be determined by several physiological and allocation traits of trees and thus can vary widely. The exponent (b) can be perceived as a distribution coefficient for the growth resources between Y and X, i.e. when X increases by 1%, Y increases by b% (Pretzsch and Dieler, 2012). BEM 1 can be either isometric ($b = 1$) or allometric ($b \neq 1$) depending on the exponent. Isometric scaling represents a simple 1:1 rule between two biological traits Y and X. For example, the scaling of above-ground biomass with below-ground biomass has been widely reported to be isometric across forest types (Cheng and Niklas, 2007; Hui et al., 2014; Yang and Luo, 2011).

Table 1
Allometric scaling relationships between total above-ground biomass (M) and tree diameter (D), M and stem height (H), H and D, M and tree density (N), crown radius (CR) and D, crown area (CA) and D assuming the constant stress (STRESS), elastic similarity (ELASTIC) models, metabolic theory of ecology (MTE) and geometric similarity (GEOM) models.

Co-variation	GEOM	STRESS	ELASTIC	MTE
M–D	$M = aD^3$	$M = aD^{5/2}$	$M = aD^{8/3}$	$M = aD^{8/3}$
M–H	$M = aH^3$	$M = aH^5$	$M = aH^4$	$M = aH^4$
H–D	$H = aD^1$	$H = aD^{1/2}$	$H = aD^{2/3}$	$H = aD^{2/3}$
M–N	–	–	–	$M = aN^{-4/3}$
CR–D	$CR = aD^1$	–	$CR = aD^{2/3}$	$CR = aD^{2/3}$
CA–D	$CA = aD^2$	–	$CA = aD^{4/3}$	$CA = aD^{4/3}$

– No explicit predictions are available.

There are different theoretical allometric models derived from physical and biological first principles that predict universal scaling relationships between Y and X. These are the geometric similarity, stress similarity, elastic similarity and metabolic theory of ecology (MTE) (Niklas, 1995; West et al., 1999). In each model, some physical optimization principle is invoked to explain the origin of allometric relationships. The geometric similarity model (hereafter GEOM) predicts a scaling of M with D or H and other predictors as in Table 1. The constant stress similarity model (hereafter STRESS) also called the concept of adaptive growth and uniform stress theory posits that the shape of a tree's stem is influenced by mechanical loading. It assumes that the trunk tapers such that stress produced by wind pressure along the stem is equalized resulting in H scaling with diameter D as $H = aD^{1/2}$ (or $D = aH^2$). The elastic similarity model (hereafter ELASTIC) considers a tree trunk as a self-supporting tapering column and thus predicts that H will scale with D as $\alpha D^{2/3}$ (or $D = \alpha H^{3/2}$) to resist buckling of a tree under its own weight. Although STRESS and ELASTIC are derived from biophysical principles, they make different predictions for the exponents relating Y with X (Table 1). MTE uses scaling relations based on metabolism and biomechanics to quantify how trees use resources, fill space, and grow (Enquist et al., 1999, 2009; West et al., 1999, 2009). Recent theoretical work (Enquist, 2002; Enquist et al., 2009; West et al., 2009) and empirical studies (Anfodillo et al., 2012) indicate that such relationships at an individual level shape the structure of the whole forest community. As such the entire forest behaves as if it were a hierarchically branching resource supply network that mimics the branching network of a single tree (e.g. Enquist et al., 2009; West et al., 2009).

Each of these models establishes universal exponents for the allometric relationships among several predictors such as M, D, H, tree density (N), crown radius (CR) and crown area (CA) as summarized in Table 1. For the sake of brevity in this review, I will focus only on the scaling of M–D and H–D because of their direct relation to forest biomass estimation. Under each, I will illustrate instances of poor analyses and misinterpretation of allometric coefficients in the published literature. Through the meta-analysis, I will also attempt to test which of the theoretical models have overwhelming support from empirical studies published so far.

3.1.1.1. Allometric scaling of tree biomass with diameter (M–D). The different allometric scaling theories predict different exponents for the M–D relationship. The STRESS predicts a scaling with an exponent of 5/2, while ELASTIC and MTE predict a common scaling exponent of 8/3 (Table 1). The difference is that ELASTIC assumes biomechanical while MTE assumes metabolic scaling under optimized tree architecture (allometric ideal plants) and resource

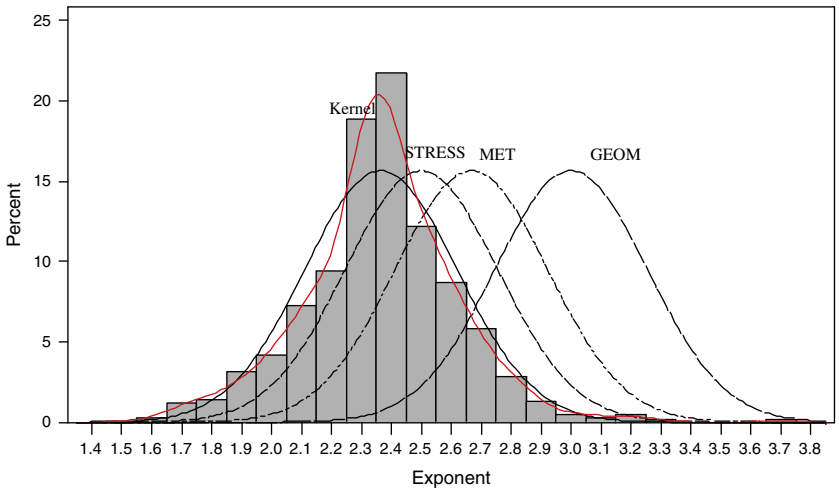

Fig. 1. The statistical distributions of the exponent (b) of the M–*D* scaling based on data from 684 equations. Bars represent the distribution of observed values, while the red line represents the kernel density distribution. The smooth black line represents the fitted normal distribution of the empirical exponents with a mean = 2.37, mode = 2.36, median = 2.36, standard deviation = 0.26, skewness = 0.17 and kurtosis = 1.57. The estimated quantiles for the normal distribution are: 10% Q = 2.04, 25% Q = 2.19, 50% Q = 2.37, 75% Q = 2.54 and 90% Q = 2.69. The distribution of theoretical expectations are represented by broken lines, i.e. STRESS (b = 5/2), MTE (b = 8/3) and GEOM (b = 3.0) given the data. All tests (Shapiro–Wilk, Cramer-von Mises and Anderson–Darling) indicate significant departure (P < 0.01) of empirical values of the exponents from a normal distribution. Note that the STRESS and MET have stronger support by the empirical data than the GEOM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

transport (Niklas and Spatz, 2004; West et al., 1999). There is a growing body of evidence suggesting that *M–D* scaling in multi-species tree data behave the same way as single species data (Tredennick et al., 2013; Kuyah et al., 2012, 2014). However, discrepancies between empirical and theoretical values of the exponent have triggered fierce debate. In an effort to explain the sources variation, some researchers (e.g. Návar, 2009; Pilli et al., 2006; Zapata-Cuartas et al., 2012; Zianis, 2008) have tried to synthesize metadata. However, some of the analyses have resulted in erroneous conclusions as will be illustrated in the following sections.

The meta-analysis in this study reveals that the empirical distribution of the exponent overlaps with the distribution expected under the STRESS and ELASTIC/MTE models (Fig. 1a). This highlights that all the theoretical models except GEOM have some support from empirical studies. Contrary to the conclusion by Návar (2009), this analysis does not reject the theoretical values of the exponent (b = 5/2 or 8/3). The normal and kernel density distributions of the exponent have a single mode (Fig. 1). However, statistical tests indicate that the empirical distribution of the exponent significantly departs from normal (see Fig. 1 and Appendix A Figs. 1 and 2 for explanation). This is indicative of the fact that the studies available so far do not adequately represent the sampling distribution of the allometric exponent. Therefore, the confidence intervals (CI) computed from such datasets are expected to be unreliable. This analysis also suggests that the scaling exponent is stochastic in nature and better represented by a probability distribution. The stochastic nature of the exponent is consistent with Niklas and Spatz (2004) who state that scaling relationships vary across species due to species-specific differences in biomass partitioning patterns and ecological responses to different environmental conditions. MET also predicts a log–log nonlinear relationship across species with small body sizes ultimately converging on the 8/3 scaling as body size increases (Niklas and Spatz, 2004).

As in the exponent, the empirical distribution of the intercept has one mode but it significantly departs from normality (see Appendix A for details). The intercept is inversely related to the exponent (Fig. 2a). This has also been observed in earlier studies with smaller sample sizes (e.g. Fehrmann and Kleinn, 2006; Pilli et al., 2006; Zapata-Cuartas et al., 2012; Zianis, 2008). As illustrated in Fig. 3A, large uncertainty in values of the intercept occurs where the exponent is <2.40.

The review indicated that heterogeneity in empirical values of the allometric parameters can result from methodological artefacts including small sample size, choice of regression methods and fitting of models to parts of a dataset. For example, extremely small or large values of the exponent (Fig. 2b) and intercept (Fig. 2c) were associated with small sample sizes of less than 50 trees. As sample size increased variability decreased and the exponent appeared to converge around the theoretical values of 2.5 for STRESS (Fig. 2b). When sample sizes are less than <50 trees, inferences about discrepancies between empirical and theoretical values are known to be problematic (Coomes and Allen, 2009; see also Table 2). Over 60.4% of values of the exponent in the literature were estimated from a sample size of less than 30 trees, while medium to large sample sizes ($\geqslant$50 trees) accounted only for 19.3% of the equations. Therefore, a great deal of uncertainty exists in the >60% of values of the exponent.

Different regression methods can also give widely differing estimates of the allometric coefficients (Table 2) from the same dataset. Irrespective of regression method, the exponent of Ebuy's dataset (sample size of 12 trees) was very small relative to theoretical predictions. The intercept in Ebuy's dataset was also estimated with very large errors across regression methods. This re-enforces the finding that large uncertainty may be associated with estimates of the exponent in small samples (Fig. 2b). The estimates of the exponent from the nonlinear method were outside the OLS, ML and Bayesian 95% credible intervals (CI) for the data from Djomo, Kuya and Noguiera (Table 2). The exponent estimated using the nonlinear regression of the dataset of Mascaro was outside the 95% CI of the OLS, ML and Bayesian although the CIs themselves overlapped. The robust estimate of the exponent was outside the Bayesian 95% CIs for the data from Djomo (Table 2). This highlights the danger of taking 95% CI at face value and using them to reject/accept a hypothesis as in Návar (2009).

The arbitrary classification of trees as juvenile and mature (Pilli et al., 2006) or small and large (Muller-Landau et al., 2006) and fitting equations to a restricted part of the tree size range is also a potential source of bias in allometric coefficients (see Gould, 1966 for more). By dividing a dataset of 1504 trees into small

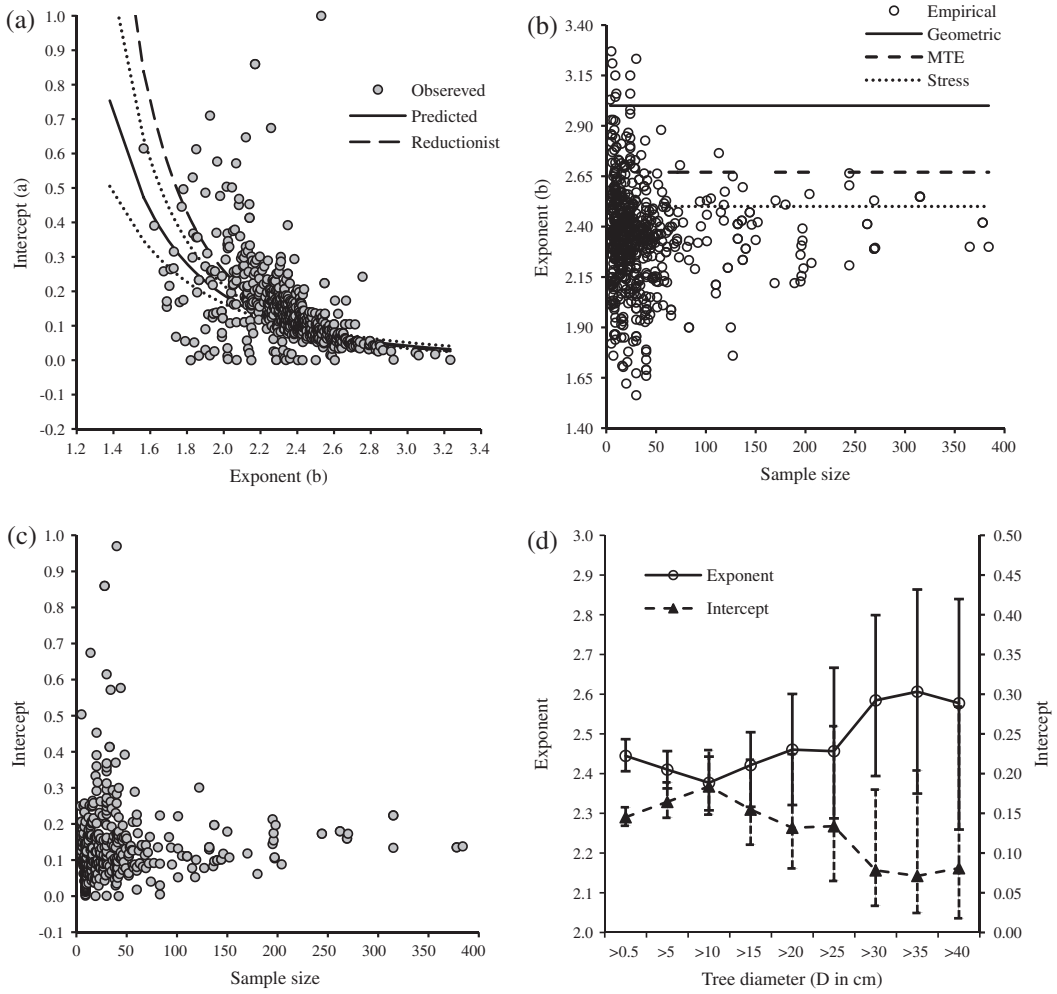

Fig. 2. The relationship between the intercept and the exponent of the M–*D* scaling (a), their distribution in relation to sample size (b and c) and tree size (d). Dotted lines in A represent the 95% confidence intervals of predictions generated using maximum likelihood method. The estimated power law equation was $2.52b^{-3.73}$ compared to the reductionist method of Zianis (2008) $a = 7.028b^{-4.756}$, which predicted values (dashed line) outside the 95% CIs. Figure 2d shows the increase in the exponent (bars are 95% CI) and corresponding decrease in the intercept with successive removal of smaller diameter trees (i.e. >1, >5, >10, >15, >20, >25 and 30 cm) and refitting the models. Model fit (R^2) with the remaining sample sizes (i.e. 596, 468, 335, 228, 170, 138 and 109 trees) only changed slightly (R^2 = 0.979, 0.958, 0.945, 0.928, 0.902, 0.881 and 0.882, respectively).

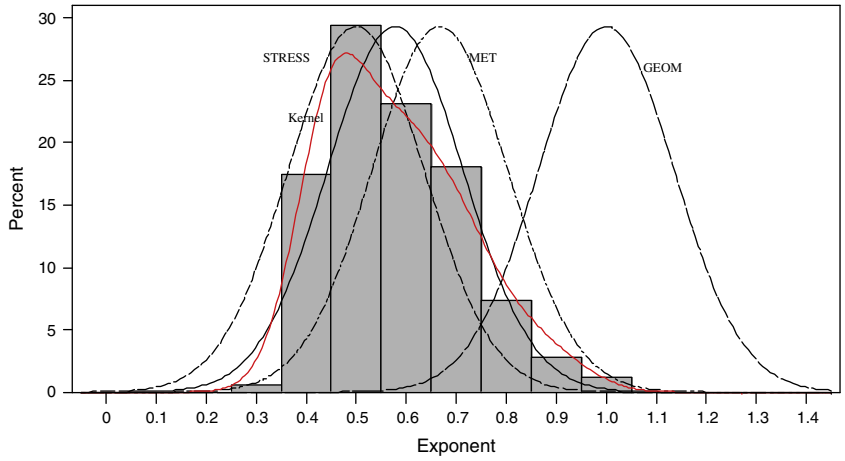

Fig. 3. The statistical distributions of the exponent of the *H*–*D* scaling. Bars represent the distribution of observed values, while the red line represents the kernel density distribution. The smooth line represents the approximation to normal distribution (b = 0.58; mode = 0.45; median = 0.56; standard deviation σ = 0.14; skewness = 0.66 and kurtosis = −0.27). The estimated quantiles for normal distribution are: 10% Q = 0.40, 25% Q = 0.49, 50% Q = 0.58, 75% Q = 0.67 and 90% Q = 0.75. The broken lines represent the distributions expected under the theoretical models, i.e. STRESS (b = 0.50), MTE (b = 0.667) and GEOM (μ = 1.0) given the data (i.e. e standard deviation σ = 0.136). All tests indicate significant departure (P < 0.01) of the exponents from normality. Note that the STRESS and MET have stronger support by the empirical data than the GEOM.

Table 2
Heterogeneity in estimates of the allometric exponent (b) and intercept (a) of $M = aD^b$ due to choice of regression methods, i.e. nonlinear, robust regression, log-linear ordinary least square (OLS), log-linear maximum likelihood (ML) and Bayesian analysis. In the robust regression, the MM estimator was used. Figures in parenthesis are 95% confidence intervals and Bayesian credible intervals (CIs).

Data set	*N*	Nonlinear[a]	Robust[b,c]	OLS[b]	ML[b]	Bayesian[b,c]
Exponent						
Ebuy	12	1.64 (0.82–2.47)	1.36 (0.79–1.93)	1.53 (0.94–2.12)	1.53 (0.99–2.07)	1.54 (0.89–2.19)
Henry	42	2.41 (1.87–2.95)	2.43 (2.29–2.58)	2.40 (2.27–2.54)	2.40 (2.27–2.53)	2.40 (2.27–2.54)
Djomo	71	3.83 (3.68–3.96)	2.07 (1.94–2.20)	2.33 (2.21–2.45)	2.33 (2.22–2.44)	2.33 (2.21–2.44)
Kuyah	72	3.09 (2.80–3.38)	2.49 (1.41–2.56)	2.48 (2.41–2.55)	2.48 (2.40–2.55)	2.48 (2.40–2.55)
Mascaro	150	2.30 (2.12–2.47)	2.33 (2.28–2.38)	2.22 (2.16–2.28)	2.22 (2.17–2.28)	2.22 (2.17–2.28)
Nogueira	264	2.30 (2.21–2.40)	2.44 (2.39–2.49)	2.39 (2.33–2.45)	2.39 (2.37–2.45)	2.39 (2.34–2.44)
Intercept						
Ebuy	12	**3.62 (-7.46–14.7)**	**9.30 (1.27–68.0)**	**5.19 (0.65–41.6)**	**5.21 (0.77–34.8)**	**5.05 (0.75–37.34)**[d]
Henry	42	**0.18 (-0.31–0.67)**	0.17 (0.10–0.30)	0.18 (0.11–0.31)	0.18 (0.11–0.30)	0.18 (0.11–0.31)
Djomo	71	**0.0006 (0.0002–0.0009)**	0.16 (0.13–0.18)	0.12 (0.10–0.14)	0.12 (0.10–0.14)	0.12 (0.10–0.14)
Kuyah	72	**0.007 (-0.002–0.015)**	0.08 (0.06–0.11)	0.09 (0.07–0.11)	0.09 (0.07–0.11)	0.09 (0.07–0.11)
Mascaro	150	0.24 (0.10–0.37)	0.18 (0.17–0.20)	0.23 (0.20–0.25)	0.23 (0.20–0.25)	0.23 (0.20–0.25)
Nogueira	264	0.29 (0.17–0.40)	0.17 (0.17–0.20)	0.19 (0.16–0.22)	0.19 (0.16–0.22)	0.19 (0.16–0.22)

N is the sample size.
Bold figures are unreliable estimates according to the RSE (>30%). According to Hougard's measure of skewness (>0.25) the intercept values estimated using non-linear regression were unreliability in the data from Ebuy, Henry, Djomo and Kuyah.
[a] In the non-linear regression, untransformed data were analyzed using Gauss–Newton algorithm.
[b] In OLS, ML regression and Bayesian analyses, *M* and *D* data were transformed into log*e* (i.e. ln) before analyses.
[c] Bayesian analysis was implemented using the GENMOD procedures of SAS. Here, Jeffreys' prior was used because it was the more appropriate one for the data based on the deviance information criterion (DIC).
[d] Note that even the Bayesian estimate has such an unrealistically wide CI indicating unreliability of estimates from the small sample used in Ebuy's data.

and large, Muller-Landau et al. (2006) estimated the values of the exponent (b) and its 95% CIs as $b = 2.65$ (2.58–2.72) for 1018 small trees, $b = 2.42$ (2.38–2.47) for 481 large trees and $b = 2.61$ (2.58–2.64) for the whole dataset. The authors interpreted the non-overlapping 95% CI for small and large trees as a reflection of heterogeneity in the MTE exponents. In order to illustrate how this can also emerges as an artefact of model fitting, I combined datasets from Ebuy, Djomo, Henry, Mascaro, Nogueira and Kuyah and successively removed plants with $D < 1$, 5, 10, 15, 20, 25, 30 and 35 cm and fitted the *M*–*D* scaling to the remaining data. This resulted in increases in the exponent (and widening of its 95% CIs) with successive removal of smaller diameter trees and a corresponding decrease in the intercept (Fig. 2d). This highlights the fact that with current methods it is not possible to confidently tell whether heterogeneity in the exponent is an artefact or a reflection of its size-dependence. Fitting models to a part of the dataset is only appropriate where biological thresholds (break points) exist (Friedman et al., 2013). In such instances a more reliable approach is to fit piece-wise regression (Friedman et al., 2013; Moncrieff et al., 2011).

Errors in measurement of either *M* or *D* can also influence estimates of the allometric coefficients. For example, the ambiguity with which tree diameter (D) at breast height has been defined in the past could be one source (Brokaw and Thompson, 2000). The mean difference between *D* measured at 130 and 140 cm was 3.5 mm but this resulted in 4% higher biomass when calculated using *D* measured at 130 cm than at 140 cm (Brokaw and Thompson, 2000). According to Molto et al. (2013a) a bias of ±10% in *D* estimates can result in ±20% or more bias in *M* estimates. Just by rescaling *D* to $D_{0.1}$ Fehrmann and Kleinn (2006) found a shift in the values of the exponent from 2.40 to 2.56, a value closer to the theoretical (2.50 for STRESS and 2.67 of MTE). The way above-ground biomass is reported is also variable; i.e. some reported woody biomass alone while others reported woody + foliage biomass. When such studies are combined in meta-analyses, they are likely to produce distorted mean estimates of the exponent.

Given all these sources of errors and the non-normality of the exponent (Fig. 1), statistical comparison of point estimates and their 95% CI from meta-analyses with theoretical values can be misleading. Therefore, the conclusion by Návar (2009) and his outright rejection of the null hypothesis are premature. Picking equations arbitrarily from existing literature and databases (e.g. GlobAllomeTree) can also result in unreliable biomass estimates. The use of average values of the intercept (e.g. $a = 0.10$ or 0.142) as proposed in the 'global' model of Zianis (2008) is likely to provide estimates biased to an unknown degree. The reductionist approach of Zianis (2008) is also likely to be inadequate because the values of the intercept widely fluctuate especially when the exponent is <2.4 (Fig. 3c). Using this method, Zianis and Mencuccini (2004) found $a = 7.028b^{-4.756}$ from a meta-analysis of 279 empirical equations. My analysis of 639 equations gave $a = 2.581b^{-3.78}$ leaving predictions from the reductionist model outside the 95% CI of the empirical values (Fig. 2a). This highlights the need for rigorous validation of existing models before applying them to local conditions.

3.1.1.2. Allometric scaling of tree height with diameter (H–D). As in the *M*–*D* scaling, theories of allometric scaling provide competing predictions for *H*–*D* relationships (Table 1). The STRESS predicts an exponent of 1/2 while ELASTIC and MTE predict 2/3. According to Niklas and Spatz (2004) metabolism and hydraulic transport rather than mechanical constraints drive *H*–*D* scaling. This leads to a log–log nonlinear relationship across species for small *D*, but ultimately converges on the 2/3 scaling rules as body size increases (Niklas and Spatz, 2004). Although the ELASTIC/MTE predictions have been supported by empirical results (King et al., 2009; Niklas, 1995), in recent reports the 2/3 scaling has been challenged (e.g. Ducey, 2012; Pretzsch and Dieler, 2012; Russo et al., 2008; Watt and Kirschbaum, 2011). For example, Russo et al. (2008) statistically compared empirical values with the exponent of 2/3 and concluded that there was virtually no support for MTE predictions. Watt and Kirschbaum (2011) similarly found exponents ranging from 0.73 to 1.43 and concluded that this clearly violates the assumptions underlying allometric relationships. Contrary to these conclusions, results of the meta-analysis indicate that empirical values of the exponent widely overlap with theoretical values predicted by the STRESS and ELASTIC/MTE (Fig. 3). This suggests that there is not enough ground for rejecting predictions of any of these theories. The departures from theory have been shown

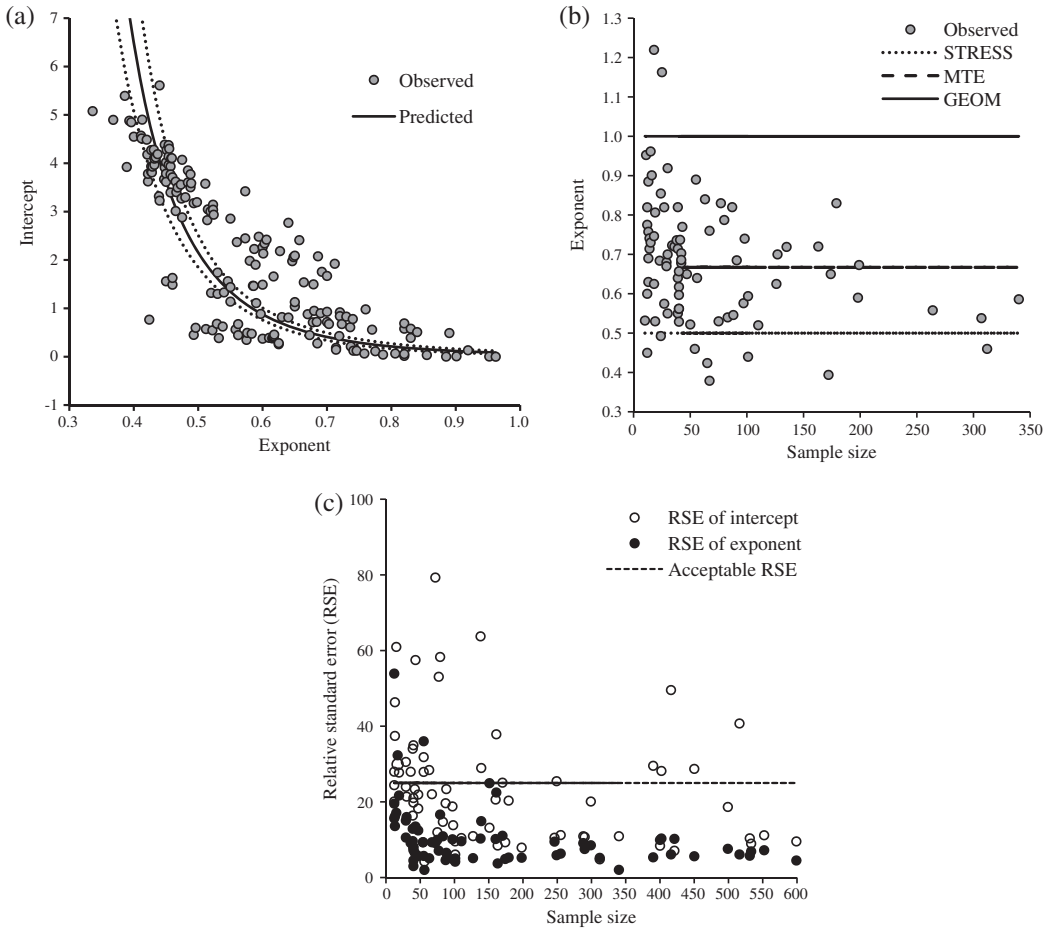

Fig. 4. The relationship between the intercept and the exponent of the *H–D* scaling (a), the distribution of the exponent in relation to sample size and theoretical values (b) and the error (RSE%) with which the coefficients have been estimated (c). Dotted lines in (a) represent the 95% confidence intervals estimated using maximum likelihood methods assuming a power-law relationship between the intercept and exponent.

to corresponds with trade-offs related to disturbance (e.g. herbivory, fire, etc.) and competitive interactions (Moncrieff et al., 2011; Tredennick et al., 2013).

As in the *M–D* scaling, the allometric coefficients are inversely related (Fig. 4a). The *H–D* exponent has been estimated relatively accurately (PRSE < 30) especially with sample sizes of 30–50 trees (Fig. 4b). From the same dataset the intercept has often been estimated with larger errors (Fig. 4c). This is probably because methods for *H* measurement are not standardized and *H* is also very difficult to measure accurately in closed-canopy tropical forests. For example, Hunter et al. (2013) found that the precision of *H* measurements ranged from 3% to 20% of total *H*, leading to a mean error of 16% in the estimate of individual tree biomass. Larjavaara and Muller-Landau (2013) reported overestimation of *H* by more than 100% or underestimation by 20% depending on the method used.

Heterogeneity in the *H–D* scaling parameters can also result from the arbitrary classification of trees, inappropriate analytical methods and small sample sizes. For example, spurious values of the exponent may be obtained due to the classification of trees according to size, and separately fitting models for each size classes. Muller-Landau et al. (2006) fitted the *H–D* scaling to 7430 small trees (below canopy) and found $b = 0.649$ (0.642–0.655). The corresponding value for 1613 large canopy trees was $b = 0.460$ (0.443–0.477) compared to $b = 0.593$ (0.590–0.597) for the whole dataset of 9043 trees. Similarly, Watt and Kirschbaum (2011) fitted models separately and found $b = 0.53$ for trees with $D < 10$ cm and $b = 1.06$ for $D \geqslant 10$ cm. Using reduced major axis regression, the exponent was even higher at $b = 1.19$ Watt and Kirschbaum (2011). As depicted in Fig. 2d, entirely different values of the exponent could have been found if a different classification was applied. This highlights that such findings can arise either as artefacts of the model fitting or the data generating process. MET acknowledges that fitting data for shrubs and saplings (with relatively few branches may not have an area filling architecture) can results in a different exponents from trees (Enquist, 2002). It must be noted that the whole idea behind scaling theories is that they are consistent across size classes and fitting models to sub-sets of data is contrary to that principle (see Gould, 1966). However, if well-established biological thresholds exist fitting piece-wise regression may be more appropriate (Friedman et al., 2013; Moncrieff et al., 2011).

Since the exponent is inversely related to the intercept, any other factor that influences the correct estimation of the intercept can also result in errors in estimation of the exponent. Onthogenetic factors and biophysical constraints on plant growth and function can be a source of heterogeneity in the intercept, and thus their influence is likely to be reflected in the exponent. For example, Osunkoya et al. (2007) demonstrated the role of vertical stratification of light and shifts in the *H–D* scaling with ontogeny, especially for the understory species (Osunkoya et al., 2007). Among other factors cited as sources of heterogeneity in the *H–D* scaling are evergreenness, wood density (Ducey, 2012), stand density, topographical exposure, soil carbon to nitrogen ratio and air temperature (Watt and Kirschbaum, 2011).

The conclusion from the review of *M–D* and *H–D* above is that allometric coefficients are estimated with a large degree of error and some of the figures in the published literature may be unreliable. It must be noted that in reality the true value of allometric coefficients are unknowable because they are parameters. We can only obtain their estimates from samples and sometimes estimates may be biased to an unknown degree. Since the scaling relationships governing tree height, diameter, and mass are adaptive and functionally flexible (Niklas and Spatz, 2004), allometric coefficients should be plastic. Therefore, heterogeneity in their values should be the rule rather than the exception. Contrary to these facts, some authors have interpreted deviations of empirical values from theoretical values as proof of violation of theoretical predictions (e.g. Coomes and Allen, 2009; Muller-Landau et al., 2006; Návar, 2009; Russo et al., 2008). There is also a widespread misconception about the concept of "invariance" in allometric scaling. This does not mean a single fixed value of the exponent across all empirical studies or a mean from meta-analyses. Rather it means that the exponent is a parameter that possesses the same statistical properties at a wide range of scales and levels of biological organization (Marquet et al., 2005). Allometric scaling theories only set limits within which organisms can live and function. However, there may be stronger limits determined by onthogeneny and/or ecological interactions such as competition and disturbance (Moncrieff et al., 2011; Tredennick et al., 2013). Thus theoretical values should be interpreted as "expectations" for allometrically ideal plants. The meta-analysis highlighted that the STRESS, ELASTIC and MTE models have some support by published studies, and their better use would help in furthering our understanding of mechanisms. Unfortunately, in most recent analyses the focus has only been on debating about the MTE. Since the alternative theories of STRESS, ELASTIC and MTE do not make mutually exclusive predictions, formal quantitative testing will be challenging. This highlights the need for adopting a pluralistic approach to explain a richer variety of processes. Assertions such as "we found virtually no support for any prediction" or "the hypothesis that the exponent is equal to 8/3 was rejected" are uninformative or outright misleading. Researchers should also be cautious of misreading into such statements because some of the conclusions could easily emerge from confirmation bias (*sensu* MacCoun, 1998) rather than the data generating process.

The review and analyses above also highlight that some of the heterogeneity may be attributed to methodological artefacts leading to larger errors in the intercept and the exponent. Therefore, future research should focus more on identifying the drivers of the variation in the parameters (especially the intercept) at the relevant scale and subjecting them to more extensive empirical and theoretical scrutiny.

3.1.2. BEMs with multiple predictors

The BEMs based on power-law functions are believed to be too simple to deal with the idiosyncrasy of tropical forest biomass. Therefore, models with multiple predictors are often claimed to be more useful to quantify biomass at a given scale. Across landscapes and forest types, biomass can vary with species, population density, growth stage, management, topography, site characteristics, etc. At the plot level, variation in biomass can be quantified using numerical models that describe the relationships between biomass and measurements such as *D*, *H*, wood density (ρ), diameter at base (Db), basal area (BA), crown area (CA) and/or site characteristics. BEMs with multiple predictors refer to this kind of models. An infinite number of models can be derived using combinations of these predictors, and their polynomial or interaction terms (see Appendix C in Henry et al., 2011; Sawadogo et al., 2010). In this review, I will only focus on the following broad class of models which appear widely in the recent literature. For brevity, I will divide the most common models into three categories as (1) models with linear and polynomial terms (BEM 2), (2) models with linear and/or interaction terms (BEM 3); and (3) models with compound derivatives (BEM 4).

Models with linear and polynomial terms are usually framed as follows:

$$ln(M) = ln(a) + b(lnD) + c(lnD)^2 + d(lnD)^3 + e(ln\rho) \quad \text{(BEM2)}$$

Chave et al. (2005) constructed BEM 2 with the assumption (citing Niklas, 1995) that a polynomial relationship between *H* and *D* provides a reasonable generalization of the power-law model. Contrary to this assumption, Niklas (1995 on page 220) states that a quadratic equation is a poor model ..... since this indicates tree height is predicted to decrease for very large stem diameters. Several sub-models of BEM2 have been derived assuming different values of *c*, *d* and e (see Appendix C in Henry et al., 2011; Ngomanda et al., 2014; Appendix A in Zianis and Mencuccini, 2004). The disadvantage of BEM 2 is that its parameters vary widely across studies (see Table 3) meaning that it generalizes poorly. In addition, it has several conceptual and statistical problems. First, the coefficients of the quadratic and cubic terms are not amenable to meaningful interpretation because of the confounding of the various terms. Secondly, as a polynomials fitted to logarithms cannot be re-expressed in the arithmetic domain and regression coefficients may not be interpreted in the context of the original measurements. Thirdly, the motivation for including polynomials (D^2 and D^3) is to test for curvature in addition to linear effects in a model. However, a 3rd degree polynomial assumes "increase–decrease-increase" response with increase in the predictor. Therefore, the *M–D* relationship implied in BEM 2 is biologically implausible.

Among the models with linear and/or interaction terms, the following is the most common one:

$$ln(M) = ln(a) + b(lnD) + c(lnH) + d(ln\rho) \quad \text{(BEM3)}$$

A number of sub-models may be derived from this general model assuming different values of *c* and *d*. For example when $c = d = 0$, BEM 3 is the same as BEM 1. Like BEM 2, the parameters of this model vary widely across published studies (Table 3). Sometimes, interaction terms may also be included and the model can be very complicated. Typical examples of this are found in Sawadogo et al. (2010) and many others in Appendix C of Henry et al. (2011). Hybrids between BEM 2 and BEM 3 such as $ln(M) = ln(a) + b(lnD) + c(lnD)^2 + d(lnH) + e(ln\rho)$ have also been used (e.g. Colgan et al., 2013; Moore, 2010).

A typical example of models with compound derivatives is the following

$$ln(M) = ln(a) + bln(\rho D^2 H) \quad \text{(BEM4)}$$

Depending on availability of ρ or *H* data, BEM 4 could be reduced to $ln(M) = ln(a) + bln(D^2H)$ or $ln(M) = ln(a) + bln(\rho D^2)$. Parameters of BEM 4 are less variable compared to those of BEM 2 and 3 (Table 2). However, this does not mean that it has the power-law properties of scale-invariance (self-similarity) and universality.

As in BEM 1, the intercept term in BEMs 2–4 has been erroneously presented as "*a*" instead of "$ln(a)$" in a number of publications (see Basuki et al., 2009; Chave et al., 2005; Colgan et al., 2013; Djomo et al., 2010; Fayolle et al., 2013; Henry et al., 2010; Lima et al., 2012; Moore, 2010; Ngomanda et al., 2014; Picard et al., 2012; Vieilledent et al., 2012). This is a non-trivial error because it can mislead researchers who select equations from the literature or databases such as the GlobeAllomeTree. For example, unwary readers could use the value of "*a*" without back-transforming (as e^a or 10^a) in the geometric domain. Therefore, prospective

Table 3
Variability in parameter estimates of recently published multi-species equations.

Model	Forest	Model specification and application	*a*	*b*	*c*	*d*	*e*
BEM 2		$ln(M) = ln(a) + b(lnD) + c(lnD)^2 + d(lnD)^3 + e(ln\rho)$					
	Dry	Chave et al. (2005)	−1.02	1.82	0.20	−0.03	0.39
		Alvarez et al. (2012)	3.65	−1.70	1.17	−0.12	1.29
	Moist	Alvarez et al. (2012)	1.96	−1.10	1.17	−0.12	1.06
		Alvarez et al. (2012)	2.41	−1.29	1.17	−0.12	0.45
		Djomo et al. (2010) data[a]	−1.75	1.98	−0.01	0.03	0.11
		Ebuy et al. (2011)	234.31	−191.1	53.01	−4.85	0.19
		Fayolle et al. (2013)	−1.17	1.91	0.25	−0.03	0.98
		Henry et al. (2010) data[a]	−3.96	5.34	−0.90	0.09	1.04
		Kuyah (data)	−0.29	1.19	0.33	−0.03	0.74
		Ngomanda et al. (2014)	6.93	−5.41	2.42	−0.24	1.42
	Wet	Alvarez et al. (2012)	3.13	−1.54	1.17	−0.12	1.77
		Alvarez et al. (2012)	1.84	−1.26	1.17	−0.12	−0.22
		Alvarez et al. (2012)	1.66	−1.11	1.17	−0.12	0.33
BEM 3		$ln(M)=ln(a) + b(lnD) + c(lnH) + d(ln\rho)$					
	Dry	Chave et al. (2005)	−2.68	1.81	1.04	0.38	
		Alvarez et al. (2012)	−2.22	2.08	0.59	1.09	
	Moist	Alvarez et al. (2012)	−2.22	2.08	0.59	1.09	
		Alvarez et al. (2012)	−2.92	2.08	0.59	0.39	
		Chave et al. (2005)	−2.99	2.14	0.82	0.81	
		Djomo et al. (2010)	−6.28	0.05	4.03	0.33	
		Ebuy et al. (2011)	−0.02	1.31	0.67	−0.68	
		Henry et al. (2010) data[a]	−1.23	2.36	0.08	0.90	
		Kuyah et al. (2012)	−1.51	2.30	0.06	0.76	
		Ngomanda et al. (2014)	−2.78	1.84	1.06	1.16	
	Wet	Alvarez et al. (2012)	−2.29	2.08	0.59	1.02	
		Alvarez et al. (2012)	−3.67	2.08	0.59	−0.36	
		Alvarez et al. (2012)	−2.86	2.08	0.59	0.45	
		Chave I.1 Wet	−2.41	2.04	0.66	0.75	
BEM 4		$ln(M)=ln(a) + bln(\rho D^2H)$					
	Dry	Chave et al. (2005)	−2.19	0.92			
		Alvarez et al. (2012)	−2.33	0.94			
	Moist	Alvarez et al. (2012)	−2.33	0.94			
		Alvarez et al. (2012)	−2.26	0.94			
		Chave et al. (2005)	−2.56	0.94			
		Djomo et al. (2010) data[a]	−2.69	0.91			
		Ebuy et al. (2011) data[a]	0.45	0.66			
		Kuyah et al. (2012) data[a]	−1.85	0.88			
		Henry et al. (2010) data[a]	−1.68	0.88			
	Wet	Alvarez et al. (2012)	−2.49	0.94			
		Alvarez et al. (2012)	−2.03	0.94			
		Alvarez et al. (2012)	−2.29	0.94			

[a] Parameters were estimated from data in the paper or provided by author.

authors and reviewers should take note of this in the future. Developers of the GlobeAllomeTree should also endeavour to correct this error or explicitly state that "*a*" stands for $ln(a)$ or $log(a)$ in the relevant equations in the database.

In other publications (see Henry et al., 2010; Mate et al., 2014; Mbow et al., 2013; Sawadogo et al., 2010; Singh et al., 2011; Segura and Kanninen, 2005) untransformed values have been analyzed. Where measured H and ρ are lacking, "estimated" values or those in databases have been used as input variables (Chave et al., 2005; Molto et al., 2013a,b). H has often been estimated from D using various models (e.g. power, Chapman–Richards, exponential, logistic, Gompertz, Michaelis–Menten and Weibull, etc.) and "estimated H" (hereafter called H_{est}) has been included in BEM 2–4 (e.g. Ngomanda et al., 2014; Djomo et al., 2010; Lima et al., 2012; Rutishauser et al., 2013; Vieilledent et al., 2012). Some studies (e.g. Molto et al., 2013b; Rutishauser et al., 2013; Vieilledent et al., 2012) have reported slight improvement in model fit statistics by including Hest. However, the inclusion of H_{est} in BEM 2–4 is riddled with the following interrelated problems: (1) inclusion of H_{est} as an "independent" predictor is conceptually flawed because in reality H_{est} and D are highly correlated; (2) this will introduce statistical problems in parameters of BEM 2 and 3; (3) where compound derivatives of D and H are included (e.g. BEM 4) there is no unique way to partition the variance in the response; (4) there is a great deal of uncertainty about the appropriate model to be used to derive H_{est} from D. There are over 70 H–D models (See Feldpausch et al., 2011; Li and Zhao, 2013; Huang et al., 2000; Lima et al., 2012; Molto et al., 2013b) each producing different values of H_{est}. Applying a model from one ecoregion to different ecoregions can also result in up to 29% overestimations or up to 22% underestimations of H (Huang et al., 2000). Taken together these problems increase uncertainty about parameters of BEM 3 and 4 when H_{est} is used.

Even when measured ρ and H are available, their inclusion in BEM 3 and 4 will introduce two sources of errors over and above those expected in BEM 1: (1) errors due to estimation of model parameters and (2) errors due to measurement of ρ and H. For example, recent studies (Mate et al., 2014; Wiemann and Williamson, 2014) have demonstrated high variability of ρ with tree height. This highlights that errors would occur if ρ at breast height were assumed to represent whole stem ρ. Although inclusion of H for example in BEM 3 may result in improvement of goodness-of-fit statistics (e.g. increase in R^2), the significance of such improvement differs greatly (Li and Zhao, 2013).

It must be noted that BEMs 2–4 are neither based on any allometric theory nor do they follow the power-law function that typical allometric models assume. Nevertheless, in many publications on biomass estimation including general guidelines (e.g. IPCC, 2003; Picard et al., 2012; Walker et al., 2012) and databases

(GlobeAllomeTree), the rubric "allometric model" has been used to mean all kinds of volume and biomass equations. The dressing up of all kinds of equations with the title "allometric" has partly contributed to publication of biologically implausible and statistically dubious equations such as BEM 2 without scrutiny. BEMs can take any functional form (linear or nonlinear) and may involve multiple predictors. On the other hand typical allometric models are bivariate power-law function and thus are a subset of the broader class of biomass estimation equations. An undesirable outcome of the liberal use of the term "allometric model" to refer to BEMs 2–4 will be the emergence of terminology, approaches and standards in biomass estimation that are inconsistent with other disciplines of science. In order to facilitate effective communication and to avoid such inconsistencies, there is an urgent need for circumscription of the term "allometric models".

BEMs 2–4 are essentially based on data dredging (see 3.2.2), which has become easier to implement with the advent of increased computing power. Data dredging involves trawling through data in the hope of identifying patterns and possible associations (Burnham and Anderson, 2002). Based on the hunches developed during data exploration, researchers fit several different models and reject some before publishing selected models. The danger is that if one searches long enough one will always find some model to fit a dataset, an extreme case of which is the Freedman's paradox (Lukacs et al., 2010). When computerized search is employed, biologically implausible combinations with high descriptive ability are likely to turn up. Testing hypotheses suggested by the data can then identify false leads resulting in a biased selection of a single seemingly "best" model.

Since statistical testing of hypotheses regarding underlying theories is not the motivations in this process, there is often no reference point against which parameters of such models can be judged. For example, in over 70% of the papers (in the reference list below) with "allometry" in their title specifically referring to forest biomass estimation, there is no indication of any hypothesis. The temptation to focus only on BEMs with multiple predictors emerges from the desire to improve model fit statistics such as R^2 and RMSE. Unfortunately, focusing on the best relationship for the current dataset without paying attention to the data generating process and testing hypotheses about the functional form of the model will not provide ecological insight. The consequence of all these will be discussed under Section 3.2.2.

3.2. Common mistakes and corrective measures

Although the review identified several mistakes, in the following sections I will focus on the most common ones. These are summarized under six broad categories: (1) the arbitrary choice of analytical methods; (2) model dredging; (3) inadequate model diagnosis; (4) ignoring collinearity; (5) uncritical use of model selection criteria; and (6) uninformative reporting of results. These mistakes usually occur unintentionally because the researcher either misunderstood some aspects of the analysis or did not have adequate knowledge of better ways of doing. I identify a problem as "common" when it has been repeated in more than two publications in the last 10 years and cite those as illustrative examples. However few they are, such mistakes can have serious implications to global forest biomass and carbon estimation.

3.2.1. Arbitrary choice of analytical methods

Fitting bivariate allometric models to dataset is not a simple task and much effort has been put into refining and standardizing analytical methods (Packard, 2014; Packard et al., 2011; Taskinen and Warton, 2013; Warton et al., 2006; Xiao et al., 2011). The task becomes more complicated when fitting BEMs with multiple predictors because such refinement is lacking. Oftentimes researchers give little thought to the choice of methods and it is usually unclear why they choose one method over another. Some used ordinary least square (OLS) regression while others applied nonlinear regression (see Návar, 2009; Pilli et al., 2006; Ter-Mikaelian and Korzukhin, 1997 for reviews). From Appendix A of Ter-Mikaelian and Korzukhin (1997) I estimated that 66% of the *M–D* equations were analyzed using OLS log-linear regression and 14% using nonlinear regression. In the remaining 20%, allometric (power-law) coefficients were calculated from two or more equations. Note that different regression methods may produce different values of the allometric parameters.

The present analysis (Table 2) illustrates that estimates of the allometric parameters could differ depending on the egression method and the sample size used. For example, the exponent estimated using nonlinear regression is slightly higher than that produced using OLS, ML or Bayesian analysis. Similarly, according to Hui et al. (2014) log-linear regression and nonlinear regression generate larger scaling constants than reduced major axis (RMA) regression. I did not include RMA in this comparison because it is not applicable to the *M–D* scaling according to the criteria by Friedman et al. (2013). With small samples (e.g. Ebuy), parameter estimation can be inaccurate. A relatively large sample size (>50) is required to estimate parameters accurately using OLS (Figs. 2a and 4b; Coomes and Allen, 2009). In the case of BEMs with multiple predictors, the sample size has to be doubled or tripled depending on the number of parameters to be estimated. OLS regression can be useful only when the predictors are few, are not collinear, and have a well-understood relationship to the response variables. However, if any of these conditions break down, OLS is inefficient and inappropriate. The distribution of errors is often asymmetrical because a few large trees in the sample can exert undue influence on regression coefficients estimated using OLS.

Most often the response variable and predictor are transformed into logarithms to ensure symmetry and reduce heterogeneity of the residuals. However, this has its own limitations (Packard, 2014; Packard et al., 2011). In some cases the decision to transform or not appears to be arbitrary and in many cases no reason is given (e.g. Henry et al., 2010; Mbow et al., 2013; Singh et al., 2011; Segura and Kanninen, 2005). It must be noted that the error distribution determines which method performs better (Packard, 2014; Packard et al., 2011; Xiao et al., 2011). Non-linear regression is better for data with multiplicative, heteroscedastic, log-normal error while nonlinear regression better characterizes data with additive, homoscedastic, normal errors (Xiao et al., 2011). In nonlinear regression, it is assumed that the error term is normally distributed and additive on the arithmetic scale. In contrast, log-linear regression assumes that errors are normally distributed and additive on the logarithmic scale (Packard et al., 2011). Violation of statistical assumptions of error can lead to biased point estimates if OLS or MLE are applied especially with small samples. Bayesian methods need a much smaller sample size to estimate parameters accurately (Zapata-Cuartas et al., 2012). The hierarchical Bayesian framework has been shown to provide a robust approach (see Price et al., 2009; Tredennick et al., 2013) for estimating allometric coefficients. Prospective researchers should explore this approach in preference to the OLS and MLE methods.

3.2.2. Model dredging

To some degree model dredging (*sensu* Guthery et al., 2005) has become common place among those who develop BEMs with multiple predictors (see for example in Alvarez et al., 2012; Chave et al., 2005; Djomo et al., 2010; Fayolle et al., 2013; Henry et al., 2010; Ngomanda et al., 2014; Sawadogo et al., 2010; Segura and Kanninen, 2005). In what could be interpreted as a full-fledged model dredging, Henry et al. (2010) tested 960 models and published only three as the best based on the AIC. Equations in Table 2

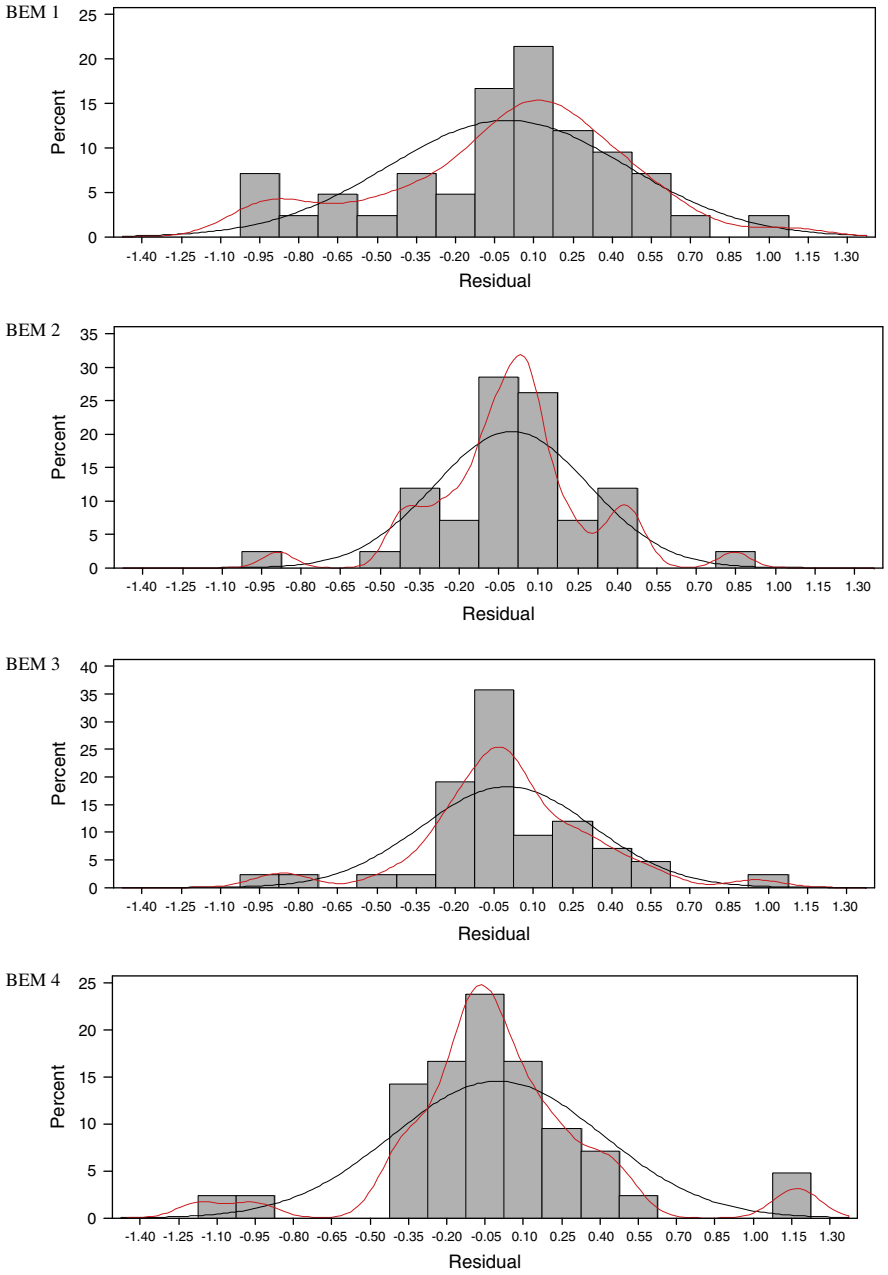

Fig. 5. Test of normality and homoscedasticity of residuals for different models fitted to the dataset from Henry. The bars are the observed compared with black smooth line representing a normal distribution ($\mu = 0$, δ) and red smooth line representing a kernel density function. Except in BEM 1 the distribution of residuals is significantly different from normal according to the Shapiro–Wilk and Cramer von Mises tests. White's test indicated slight departure from homogeneity of residuals in BEM 1 ($\chi^2 = 5.0$; $P = 0.082$) and BEM 4 ($\chi^2 = 5.1$; $P = 0.079$) but note in BEM 2 and BM 3 ($P > 0.10$).

of Sawadogo et al. (2010) and equation 15 of Segura and Kanninen (2005) are the true products of this mistake. Model dredging is unacceptable because exploratory analysis, hypothesis testing and prediction using the same dataset can lead to confirmation bias. If this trend is not corrected, it is likely to result in a situation akin to Fortsche's (1963) story in "Chaos in the brickyeard" where no effort is made even to maintain the distinction between a pile of bricks and a true edifice. Therefore, I strongly recommend that BEMs are constructed in the framework of hypothetico-deductive process. Hypotheses should be specified *a priori* based on a body of theory regarding underlying ecological processes. This is critical because theory provides structure, mechanistic insight and predictive power. BEMs formulated without reference to theory will result merely in the accumulation of situation-bound descriptions that are of limited predictive value about forest biomass trajectories.

3.2.3. Inadequate model diagnosis

Estimating regression coefficients entails more than simply fitting a line to a set of data. It is important to determine if the underlying assumptions are met or not and the equation accurately model the data generating process. Authors rarely check even for the most basic of diagnostic statistics such as normality and homoscedasticity of errors and influence statistics. Heteroscedasticity will typically be manifested as residuals whose magnitude is correlated with that of the response variable. A plot of the residuals

against the predicted values will reveal a megaphone pattern if the errors are not homogenous as in Fig. 2a of Huang et al. (2000) and Fig. 1 of Návar (2009). When comparing models, simple graphical exploration using normal and kernel density distributions can reveal problems which cannot be revealed by model fit statistics. In order to illustrate this I fitted BEM 1–4 to Henry's dataset and present the test of normality and homogeneity of residuals in Fig. 5. The figures highlight important differences among the models that cannot be ignored. For example, the kernel density reveals multimodal distribution of residuals in BEM 2–4 (Fig. 5). White's test indicated homogeneity of residuals in BEM 2 and 3 but some heterogeneity in BEM 1 and 4.

Test of normality and homogeneity alone are not adequate. Therefore, I conducted additional diagnostics to check whether certain observations have undue "influence" on the coefficients of one model compared to another. In the following discussion, I will focus on the diagnosis of influence, outliers and leverage points. I have also illustrated a complete set of other diagnostics in Appendix B. Taskinen and Warton (2013) demonstrate that classical tests that are widely used in allometry are sensitive to outliers, which can either be wrong entries or products of the regression fitting. There are two types of the later, i.e. outliers in the response variable (Y) and outliers with respect to the predictors (Xi) called leverage points. Outliers in both X and Y directions are also possible, and these were dubbed bivariate contaminations by Taskinen and Warton (2013). When the outliers are leverage points, residual plots may not reveal them. Outlier may go undetected in transformed data because logarithmic transformation usually draws them toward the center of the distribution (Packard et al., 2011). An objective way to detect outliers is to check the studentized residuals. Values exceeding −2.0 or +2 represent outliers that can cause serious heteroscedasticity. Influential points (Cook's D) are predictor combinations with unusually large weights in determining regression coefficients. Presence of outliers and influential points does not mean that the observations are unreliable but indicate an inaccurate model fit. Heterogeneity of residuals and influence statistics are inevitable because of the preponderance of small trees and a few large trees resulting in skewed M and D data. Although large individuals may constitute <5% of the trees in the landscape, they hold >48% of the biomass (e.g. Kuyah et al., 2012, 2014). These few extremes can exert large influence on coefficients especially where the model involves polynomials or compound derivatives. For example, a single large tree in Henry's dataset had such influence on BEM 2. This is because the squaring and cubing of D magnifies its influence (see Appendix B Fig. 1 and Appendix C Fig. 1 for details).

In order to illustrate this point, I made detailed comparison of models with respect to outliers and leverage points in datasets from Ebuy, Henry, Djomo and Kuya (Table 4). The distribution of residuals significantly departed from normality in BEM 3. Inclusion of polynomial also increased the magnitude of influence statistics in BEM 2 compared to BEM 1 (Table 4; see also Appendix B Fig. 1). Across the different datasets, BEM 1 produced no outliers while BEM 2 and 3 produced large outliers (1–8% of the observations). BEM 1 also had fewer leverage points (2–8% of the observations) compared to BEM 2, where 16–29% of the observations in each dataset were leverage points (Table 4). The proportion of outliers and leverage points was much higher when untransformed data were analyzed (see Appendix D). Influence statistics appear to be magnified in BEM 4 by the multiplication of D^2 by H and ρ especially for largest trees. In such case the residual can be large and distort the fitted values (see Appendix C Fig. 1). The solution to this problem is not exclusion of large trees from the analysis, but to fit models without polynomials or compound derivatives. Exclusion of large trees will reduce generalizability of the model as such trees often drive global biomass and carbon trajectories (Stephenson et al., 2014).

The impact of outliers and leverage points is usually reflected in the inflation of the standard errors (SE) of regression coefficients. A model becomes unreliable when SEs are inflated and if PRSE is greater than 25% for one or more parameters (for details see Appendix D). In order to reduce the influence of outliers on allometric coefficients, robust regression analysis has been recommended (Taskinen and Warton, 2013). Since the combined effect of outliers can destabilize coefficients of models with multiple predictors, I recommend the use of robust regression in preference to OLS regression. While there are many estimators, currently the most preferred is the MM-estimator (Andersen, 2008). The MM-estimator combines the high asymptotic efficiency of M-estimators with the high breakdown of S-estimators and addresses outliers in

Table 4
Illustrative examples of outliers (% of data), and leverage points (% of data) multicollinearity (VIF) in parameter estimates in the datasets from Ebuy, Henry, Djomo and Kuyah. VIF > 5 indicates significant collinearity.

Model	Model specification		Parameter	Ebuy (N = 12)	Henry (N = 42)	Djomo (N = 71)	Kuyah (N = 72)
BEM 1	$ln(M) = ln(a) + b(lnD)$	Outliers[a]		0	0	1.4	0
		Leverage[a]		8.3	2.4	7.0	5.6
		VIF	b	1.0	1.0	1.0	1.0
BEM 2	$ln(M) = ln(a) + b(lnD) + c(lnD)^2 + d(lnD)^3 + e(ln\rho)$	Outliers[a]		8.3	4.8	1.4	0
		Leverage[a]		**16.7**	**23.8**	**28.4**	**29.2**
		VIF	b	**5492.8**	**483.0**	**52.5**	**975.7**
			c	**24405.0**	**2350.6**	**302.8**	**4474.4**
			d	**7471.9**	**757.5**	**126.4**	**1340.9**
			e	1.3	1.1	1.1	1.6
BEM 3	$ln(M) = ln(a) + b(lnD) + c(lnH) + d(ln\rho)$	Outliers		8.3	4.8	1.4	0
		Leverage		**16.7**	**21.4**	**16.9**	2.8
		VIF	b	1.4	**9.9**	**3300.2**	3.8
			c	1.4	**10.3**	**3298.6**	2.8
			d	1.1	1.5	1.0	1.7
BEM 4	$ln(M) = ln(a) + bln(\rho D^2 H)$	Outliers		0	**11.9**	1.4	1.4
		Leverage		0	7.1	7.0	2.8
		VIF	b	1.0	1.0	1.0	1.0

Outliers and leverage points were computed using robust regression. If outliers and/or leverage points constitute >10% of the entries, model parameters are likely to be unreliable.
Bold figures indicate unreliable parameter estimates.
[a] Parameters were estimated from data in the paper or provided by author.

both the dependent and independent variables. MM refers to multiple *M*-estimation procedures (where "*M*" is for maximum likelihood type). In S-estimation, the "*S*" refers to scale.

3.2.4. Ignoring collinearity

Many measurements of tree dimensions are correlated with each other either allometrically or isometrically (Table 1). Therefore, collinearity will be inevitable when two or more of these are entered together with either their polynomial (e.g. BEM 2) or interaction terms (e.g. Table 2 of Sawadogo et al., 2010). The common diagnostic test for this is the variance inflation factor (VIF). Although many researchers suggest that multicollinearity is only severe at VIF > 5, values as low as 2 can have significant impacts on parameter estimates (Graham, 2003). I estimated VIF for BEM 1–4 in the dataset from Ebuy, Henry, Djomo and Kuyah. The results show that the VIF is several magnitudes larger than the cut-off point of 5 in BEM 2 and 3 in all datasets (Table 4). The problem was more pronounced in Djomo's dataset where *H* estimated from *D* (H_{est}) was used in BEM 3 (Table 4). This means that significant collinearity can be expected if H_{est} is included in a model together with *D* and/or where sample sizes are small. In the presence of collinearity, the main effects will have unstable regression coefficients (inflated standard errors) with erratic signs (i.e. negative signs where positive sign is expected or vice versa) as in Table 3. This hinders interpretation of the coefficients in a biologically meaningful way. Unfortunately, many authors have ignored this problem due to misplaced assumptions. For example, Sawadogo et al. (2010) states that multicolinearity was not violated since none of the explanatory variable showed a correlation coefficient of <0.90 with another variable. This is wrong because correlation coefficients are inadequate when more than two explanatory variables are included. The correlation coefficient cannot account for the moderation of one explanatory variable by another in models with multiple predictors. The correct metric is thus the variance inflation factor (VIF) or its inverse (i.e. 1/VIF called tolerance).

Collinear variables can give the appearance of a highly significant model (i.e. high R^2, low RMSE, and low AIC). If collinearity is large, (a) variables which were previously shown to be significant will have inflated standard errors (see Appendix D); (b) regression coefficients can have values inconsistent with previous reports, e.g. a negative coefficient when the true effect is positive or values of extreme magnitudes; (c) dramatic changes in the estimates of the coefficients due to addition or deletion of one or a few observations. These problems are evident in BEM 2 and 3 (see Table 3). For example, in BEM 2 not only the magnitude but also the sign of estimates differs across studies for the same regression coefficients (see Table 3). This highlights the fact that if any of the parameter values of BEM 2 in Table 3 were to be applied elsewhere, prediction errors will be large. This was already evident from Table 5 of Chave et al. (2005) where their models II.1–II.4 resulted in wide swings in bias. Therefore, it is not surprising that this models over-estimate biomass by up to 400% in studies in Brazil (Lima et al., 2012), Columbia (Alvarez et al., 2012), Gabon (Ngomanda et al., 2014), Ghana (Henry et al., 2010), Indonesia (Basuki et al., 2009), Madagascar (Vieilledent et al., 2012), Peruvian Andes (Girardin et al., 2010) and South Africa (Colgan et al., 2013). Although many authors tried to explain the discrepancy between their findings and Chave's model II in terms of differences in wood density, a more plausible explanation is the instability of the parameters in BEM 2 (Table 3 and 4). Therefore, such models should be used with a great deal of caution. When absolutely necessary, BEM 2 and 3 should be simplified through an objective method. I strongly recommend the use of partial least square regression (PLS) for this purpose. When the predictors are many, highly collinear and sample sizes are small, PLS is very reliable for identifying relevant predictors and their magnitude of influence (Carrascal et al., 2009). In order to illustrate the utility of this method I fitted BEMs 2 and 3 using PLS to the data from Djomo, Henry and Kuyah. The results in Table 5 suggest that the first factor explains 92–97% of the variance in *M* and 64–85% of the variance in the predictors. This means that the inclusion of *H* and ρ makes only a small contribution in BEM 3. This conclusion has also emerged from a recent analysis by Molto et al. (2013a). Indeed, *H* accounts for such a small variation beyond that accounted for by *D* that the chance of committing an error is high by adding *H* as a significant predictor (Ter-Mikaelian and Korzukhin, 1997). It is not surprising then that recent studies (Kearsley et al., 2013) found significant overestimation of aboveground carbon stocks using conventional BEMs employing *H*–*D* relationships. Recent analysis by Molto et al. (2013a) has also shown that ρ has a very weak contribution to *M* (i.e. coefficient not different from zero). These findings highlight the need for caution in applying BEM 3 to datasets outside sites of their origin.

3.2.5. Uncritical use of criteria for model selection

The need for model selection is particularly great in BEMs where (1) the models fitted to the data constitute only an approximation to the data generating process; (2) the number of potentially important predictors is large; and (3) the sample size is relatively small. Researchers often compare BEMs using criteria and indices selected in an ad hoc manner. The notations and definition of some of the indices also vary from one publication to another. Therefore, in this section I will review the criteria and indices used in assessing model fit and adequacy, and circumstances that produce problematic interpretations.

3.2.5.1. Coefficient of determination. The coefficient of determination (R^2) and its adjusted version are the most commonly used criteria in the BEM literature. However, in many instances it has been used uncritically. Whether the estimated coefficients are meaningful or not the model with the largest R^2 was claimed to be the best (see Mbow et al., 2013; Sawadogo et al., 2010; Segura and Kanninen, 2005). The R^2 is deceptive because firstly it increases

Table 5
Percent explained variance by factors from the partial least square regression of data from Henry, Djomo and Kuya.

Dataset	Model	Dependent				Independent			
		Factor 1	Factor 2	Factor 3	Total	Factor 1	Factor 2	Factor 3	Total
Henry	BEM 2	95.0	1.0	2.5	98.4	74.0	25.0	1.0	100
	BEM 3	96.2	0.4	1.8	98.4	64.4	32.8	2.8	100
Djomo	BEM 2	93.5	1.8	1.6	96.9	71.0	17.9	11.0	99.9
	BEM 3	95.5			95.5	67.1			67.1
Kuyah	BEM 2	97.4	0.6	1.0	99.0	85.1	13.9	1.0	100
	BEM 3	92.9	5.2	0.9	99.0	73.7	9.6	16.8	100

BEM 2: $Ln(M) = ln(a) + b(lnD) + c(lnD)^2 + d(lnD)^3 + e(ln\rho)$.
BEM 3: $Ln(M) = ln(a) + b(lnD) + c(lnH) + d(ln\rho)$.

with the addition of polynomial terms. In order to illustrate this, I sequentially added polynomial terms of D and fitted the following models to the untransformed data from Henry. Note the increase in R^2 and the coincident increase in VIF:

$M = -4.93 + 0.20D$	$R^2 = 0.607$; VIF = 1.0
$M = 0.73 - 0.06D + 0.002D^2$	$R^2 = 0.745$; VIF = 8.5
$M = -2.21 + 0.20D - 0.002D^2 + 1.47 \times 10^{-5}D^3$	$R^2 = 0.753$; VIF = 74 − 383
$M = 5.10 - 0.66D + 0.02D^2 - 1.11\text{x}10^{-4}D^3 + 7.3 \times 10^{-12}D^4$	$R^2 = 0.856$; VIF = 163 − 1898

Similarly, the R^2 will keep increasing with the inclusion of new predictors. For example, R^2 increased from 0.607 to 0.644 with the addition of H and ρ in the following models.

$M = -2.04 + 0.28D - 0.24H$	$R^2 = 0.622$; VIF = 5.2
$M = -9.60 + 0.31D - 0.32H + 14.45\rho$	$R^2 = 0.644$; VIF = 5.8

The increase in VIF means that the predictive ability is eroded. What many unwary researchers forget is that the R^2 gives the false impression that model fit to the data is improving. This is called the Feedman's paradox, i.e. you will always get a large R^2 value ($P < 0.05$) by chance alone if you do variable selection by randomly taking variables that have no correlation with your dependent variable. Unaware of this problem, some researchers conduct "best subset" model selection using methods such as maximum R^2 or adjusted R^2. However, the model thus selected can be a very bad representation of the data generating process (see Appendix E for explanations). Finally, using the same data even a higher R^2 value can be obtained with a smaller sample size. Therefore, one cannot argue that simply because a high R^2 has been obtained, a "correct" model has been found. In that sense, the use of R^2 for comparison of BEMs with multiple predictors should be discouraged.

3.2.5.2. Root mean square of error (RMSE). RMSE is referred to as the residual standard error (RSE) in some publications. In order to avoid confusion with the relative standard error (which is also abbreviated RSE), I will only refer to the RMSE. While RMSE serves to aggregate the residuals into a single measure of predictive power, it can be just as unreliable as the R^2. The RMSE decreases with increases in R^2, and therefore both represent the same thing but in opposite directions. The smallest value of RMSE is usually found in over-fit models (see Appendix D Table 1). Therefore, RMSE is practically useless for comparing models where collinear variables are included. It also gives a relatively high weight to large errors due to the squaring of errors before they are averaged. This means that RMSE is most useful when large errors are particularly undesirable. Another disadvantage is that there is no criterion for a good value of RMSE as it depends on the units in which the variable is measured and the transformation used (see Appendix Table 1).

3.2.5.3. Cross-validation. Although cross-validation is a necessary step in developing predictive models (James et al., 2013), this has been done only in very few publications (e.g. Colgan et al., 2013; Kuyah et al., 2012; Ngomanda et al., 2014). Cross validation is a better method of model assessment than residuals as it estimates how accurately BEMs will perform when applied to an independent dataset. It will also curtail problems such as over-fitting. The simplest type of cross-validation is the twofold (or hold-out) technique which consists of randomly dividing the original samples into two equal sets of training and test data. However, in some cases, the data may not have sufficient observations to create a sizable training set and a validation set that represent the population well. In these cases, a K-fold cross validation is an attractive alternative. This consists in dividing the original dataset into K subsets of approximately equal size (called folds) and using each subset, one after the other, as validation datasets, with the model being fitted on the K-1 subsets remaining. Cross validation tends to have decreasing bias but increasing variance as the number of folds increases. Therefore, a very important consideration is the optimum value of K to be used. According to James et al. (2013) 5-fold or 10-fold cross validation provide a good balance between bias and variance.

3.2.5.4. Regression of observed against predicted values. Sometimes linear regression of predicted values (Mp) against observed values (Mo) is used to judge how well different models fit empirical data (e.g. Alvarez et al., 2012; Kaitaniemi, 2004; MacFarlane et al., 2014). According to Piñeiro et al. (2008) this is incorrect because it leads to an erroneous estimate of the slope and intercept of the relationship between Mp and Mo. The correct approach is regression of Mo (in the y-axis) against Mp (in the x-axis) and testing the significance of slope ($b = 1$) and intercept ($a = 0$) (Piñeiro et al., 2008). If significant prediction errors are present, $a \neq 0$ and $b \neq 1$. This way, underestimation or overestimation of each predicted values from the 1:1 line (ideal model) can be explored graphically. In addition, I recommend testing departures from this criterion through examination of the 95% CI of a and b. This is a more transparent way for comparing new models with existing models. Using Henry's dataset, I illustrated this approach in Appendix C.

3.2.5.5. Information criteria. Akaike's information criterion (AIC) and Bayesian information criterion (BIC) were proposed to overcome the inherent weaknesses of traditional variable selection methods such as stepwise selection, forward and backward elimination (Lukacs et al., 2010). The "AIC-best" model is based on an optimal trade-off between bias and variance, which involves just the right number of predictors in a model. However, optimal bias-variance trade-off does not necessarily imply ecological reality (Guthery et al., 2005). AIC is based on the assumption that the candidate models are all good reflections of reality (i.e., statistically sound and biologically meaningful). AIC also does not assume the true model is in the set of models being compared. Therefore, it can identify a "best" model even from a set of outrageously implausible ones. For example, some of the models in Appendix D (e.g. BEM 2, 5, 10–11) are either biologically implausible or have statistical problems, yet some (e.g. BEM 2 and 15) have been identified as AIC-best. Therefore, it is up to the author to make sure that the candidate models included in the comparison are good reflections of the data generating process.

AIC can also be used to conduct "best subset" model selection. However, the AIC-best can be useless for practical application outside conditions where the current dataset was collected (see Appendix E for explanations). Nevertheless, in almost all papers I reviewed researchers have chosen the model favored by AIC as the "best". This is contrary to the central messages of Burnham and Anderson (2002) that interpretation should not be based on a single "best" model, but should explicitly acknowledge uncertainty among models that are similarly consistent with the data. AIC values *per se* are inconsequential; rather it is the Akaike weights that are more informative when comparing models. In order to demonstrate this, I calculated the AIC weights from Table 3 of Ngomanda et al. (2014) and summarized the results in Table 6. The AIC weights indicate that none of the first six models was overwhelmingly supported by the data in Ngomanda et al. (2014). Model 5, which was claimed to be the best by Ngomanda et al. (2014), has only 45% chance of being the best for the data. Among

Table 6
Comparison of models by Ngomanda et al. (2014) using the AIC, AIC weights and the error (PRSE) with which each parameters was estimated. PRSE was calculated as 100 * (SE/θ) where SE and θ are the standard error and the parameter value, respectively, in Table 3 of Ngomanda.

Model	Model description	PRSE								
		R^2	AIC	AICw	*a*	*b*	*c*	*d*	*e*	*f*
1	$\ln M = \ln(a) + b(\ln D)$	0.896	128.0	0.00	21.0	3.4				
2	$\ln M = \ln(a) + b(\ln D) + c(\ln D)^2$	0.899	127.1	0.00	**39.9**	24.6	29.6			
3	$\ln M = \ln(a) + b(\ln D) + c(\ln D)^2 + d(\ln D)^3$	0.901	127.1	0.00	**102.2**	**51.5**	**75.4**			
4	$\ln M = \ln(a) + b(\ln D) + f(\ln \rho)$	0.941	72.1	0.08	22.7	2.5				11.5
5	$\ln M = \ln(a) + b(\ln D) + c(\ln D)^2 + f(\ln \rho)$	0.944	68.6	0.45	**32.4**	17.9	**43.0**			11.2
6	$\ln M = \ln(a) + b(\ln D) + c(\ln D)^2 + d(\ln D)^3 + f(\ln \rho)$	0.946	68.5	0.47	**115.4**	**62.9**	**78.2**	**35.6**		11.3
7	$\ln M = \ln(a) + b(\ln D) + c(\ln D)^2 + d(\ln D)^3 + \ln(\rho)$	0.941	73.5	0.00	**103.0**	**54.4**	**72.1**	**33.2**		
8	$\ln M = \ln(a) + b(\ln D) + e(\ln H)$	0.929	91.1	0.00	11.2	8.1			14.7	
9	$\ln M = \ln(a) + e\ln(D^2H)$	0.924	95.7	0.00	10.2				2.9	
10	$\ln M = \ln(a) + b(\ln D) + e(\ln H) + f(\ln \rho)$	0.956	44.5	0.31	13.1	5.7			17.4	12.9
11	$\ln M = \ln(a) + e\ln(D^2H) + + f(\ln \rho)$	0.956	42.9	0.69	9.2				2.2	11.9

AIC weights were calculated as described on page 77 of Burnham and Anderson (2002).
Bold figures indicate unreliable parameter estimates. Any estimate with PRSE > 30 was deemed to be unreliable (see explanation under materials and methods).

the models with *H*, only model 11 has some support (69%). In fact models 3, 5, 6 and 7 should not have been included in the comparison because their parameters have unacceptably large errors (PRSE > 30%). I also conducted bet subset selection using Henry's dataset and demonstrate in Appendix E how this method can be misleading. While AIC identified the model with *D*, *H*, ρ, *V* as predictors as the best for Henry's dataset, the AIC weight shows that model 1 has only 25% chance of being the best (Appendix E Table 1). In Appendix E Table 2, I demonstrate that the AIC-best model has unstable parameters. Partial least square regression (Appendix E Table 3) indicated that only three factors explain >97% of the variance. An important lesson from Table 6 and Appendix E is that models with the highest R^2 and lowest AIC are not necessarily the best because they have very low likelihood (small AIC weight) or their parameters could be unstable (PRSE > 30). Therefore, the current ritualistic use of AIC to compare a plethora of models should be discouraged. Instead, AIC weights should be used for comparing a set of biological plausible and statistically sound models, i.e. those that have passed rigorous diagnosis.

3.2.5.6. Prediction error and associated metrics. Regardless of how the prediction is produced, the prediction error (e_i) is simply $e_i = \frac{Mo-Mp}{n}$ where *Mo* and *Mp* are the observed and predicted biomass of the *i*th tree, respectively, and *n* is the number of trees. While bias is one of the most common statistics used for comparing BEMs, subtle differences were found in its definition among reports published even in the same journal. In Basuki et al. (2009), van Breugel et al. (2011), Chave et al. (2005), Colgan et al. (2013), Djomo et al. (2010), Henry et al. (2010) and Macfarlane et al. (2014), bias of a given model was calculated as the mean of the residual although reported under different names (see Table 6). Fayolle et al. (2013) used the relative error calculated as the ratio of the absolute error and the observed value for each tree. Although the formula is not given, it sounds more like the formula in Pilli et al. (2006) and Zianis and Mencuccini (2004). Ngomanda et al. (2014) used the "relative bias" although it is not clear from their formula how this was calculated. Rutishauser et al. (2013) and Vieilledent et al. (2012) used more or less the same formula; the difference being the median in the former and the mean in the later. This is only a small sample of papers to illustrate how confusingly bias is defined in the biomass estimation literature. Obviously, the means and standard deviations of "bias" calculated using the various methods will be different and this will give a misleading picture about performance of the same model across different studies. In order to allow direct comparison of the different definitions, I have transcribed the various formulae using common notations (Table 7).

There are several disadvantages of the first 5 formulae in Table 7: (1) values being infinite or undefined when *Mo* approaches zero; (2) values being extremely skewed when *Mo* values are close to zero; (3) heavier penalty on positive errors than on negative errors; (4) unrealistic mean values when negative and positive

Table 7
Different definitions of bias and error as published in the literature and transcribed to a common notation using Pi to represent predicted and *Mo* to represent observed mass (i.e. ABG).

Author	Name of metric	Original formulae	Common notation for "bias"
1. Chave et al. (2005)	Error (%)	$100\left(\frac{AGBp-AGBm}{AGBm}\right)$	$\frac{100}{n}\sum_{i=1}^{n}\frac{Mp-Mo}{Mo}$
2. MacFarlane et al. (2014)	Error of prediction	$\left(\frac{AGBp-AGBm}{AGBm}\right)$	$\frac{100}{n}\sum_{i=1}^{n}\frac{Mp-Mo}{Mo}$
3. Basuki et al. (2009)	Average deviation	$\frac{100}{n}\sum_{i=1}^{n}\frac{Yi-Yi}{Yi}$	$\frac{100}{n}\sum_{i=1}^{n}\frac{Mp-Mo}{Mo}$
4. Djomo et al. (2010)	Relative error (%)	$100 \times \frac{Pi-Mi}{Mi}$	$\frac{100}{n}\sum_{i=1}^{n}\frac{Mp-Mo}{Mo}$
5. Ngomanda et al. (2014)	Relative bias	$\frac{1}{n}\sum_{i=1}^{n}\frac{Bi-Bi}{Bi}$	$\frac{1}{n}\sum_{i=1}^{n}\frac{M?-M?}{M?}$
6. Rutishauser et al. (2013)	Bias	$median\left(100 \times \left(\frac{1}{exp(\varepsilon_i)}\right)\right)$	$median\left(100 \times \left(\frac{1}{exp(Mp-Mo)}\right)\right)$
7. Vieilledent et al. (2012)	Bias (%)	$100 \times \left(\frac{1}{exp(\varepsilon_i)-1}\right)$	$100 \times \left(\frac{1}{exp(Mp-Mo_i)}\right)$
8. Paul et al. (2013)	Mean square error of prediction (MSEP)	$\frac{1}{n}\sum_{i=1}^{n}(Ei-Oi)^2$	$\frac{1}{n}\sum_{i=1}^{n}(Mp-Mo)^2$
9. Pilli et al. (2006)	Relative difference	$100 \times \frac{\lvert Mp-Mr\rvert}{Mr}$	$\frac{100}{n}\sum_{i=1}^{n}\frac{\lvert Mp-Mo\rvert}{Mo}$
10. Zianis and Mencuccini (2004)	Relative difference	$\frac{\lvert M-M\rvert}{M}$	$\frac{100}{n}\sum_{i=1}^{n}\frac{\lvert Mo-Mp\rvert}{Mo}$
11. Proposed	MAPE		$\frac{100}{n}\sum_{i=1}^{n}\frac{\lvert Mo-Mp\rvert}{Mo}$

RE is relative error; ε_i are individual errors in Rutishauser et al. (2013). Since Bi are not known in Ngomanda et al. (2014), it was not possible to transcribe.
In Zianis (2008) V_r denotes the observed and V_p predicted mass.

errors offset each other. For example, on average a very small bias is indicated when under-estimates and over-estimates cancel out. This gives the impression of a good model while hiding a huge variability. For example, Chave's model I.3 had the highest variability (−27% to +26%) and error in estimation (PRSE > 502) but the mean bias was the lowest (0.57). The use of absolute values (as in Zianis and Mencuccini, 2004) or squared values (as in Paul et al., 2013) prevents negative and positive errors from offsetting each other.

In order to avoid confusion among the different formula, I propose the use of the mean absolute percentage error (MAPE), whose statistical properties are well known and widely used in forecasting and model comparison in ecology and environmental assessment (Yao et al., 2013). Note that values calculated using formulae 1–5 in Table 7 will be different from that of MAPE. In order to illustrate this, I calculated MAPE and the corresponding PRSEs from the sample size and the standard deviation in Table 3 and 5 of Chave and summarized in Appendix F. MAPE was on average 2–5 times that produced using Chave's formula but the PRSEs of MAPE are smaller compared to those from Chave's formula (see Appendix F).

The model with the lowest bias or MAPE is not necessarily the best as it may be associated with large errors of estimation. At present there is also no cut-off point for what constitutes an acceptable value of either bias or MAP. This gives individual author liberty to declare models based on the lowest values from their dataset. This limits the utility of bias or MAPE for comparison of models across different studies. Although we cannot define a universal cut-off point for what constitutes a large bias (or MAPE), for objective model selection and consistent reporting it is critical to adopt a cut of point or a range of values. For example, models with MAPE > 10% (+PRSE > 30%) may be considered unreliable. This needs to be further investigated and an agreement needs to be reached.

3.2.6. Uninformative reporting

The main pitfalls in this area include (but not only limited to) incomplete reporting of (1) the process and criteria used for selecting existing models; (2) procedures and criteria for selecting new models locally; (3) model diagnosis and validation; and (4) measures of uncertainty in model parameters. Regarding the first problem, authors rarely provide information on how and why they selected a particular model(s) from the existing pool of models for comparison with their locally developed model(s). For example, in Table 3 there are over 20 different equations for moist forests. The question is what criteria will researchers use to pick a model appropriate for a particular condition? Sometimes, a single model is arbitrarily picked from the literature and compared with local models even where the evidence for its validity is not overwhelming. Unless objective criteria are available, confirmation bias can result in cherry-picking of models.

Secondly, when developing new models a great deal of data dredging occurs prior to publication of a seemingly "best" model. However, nothing is reported regarding the steps taken to arrive at the so-called "best" model. Simply reporting a "best" model is not only uninformative but it can be outright misleading. Therefore, it is important to report whether the model was selected with a hunch following data dredging or using some formal procedure such as best subset selection. If AIC was used as a criterion, log-likelihoods, number of parameters estimated and the AIC weights should be reported. However, it is not acceptable to report AIC and *P* values (statistical significance) together although this is common in some reports (e.g. Ebuy et al., 2011; Ngomanda et al., 2014; Sawadogo et al., 2010). If one reports P values, one is implicitly claiming that hypotheses have been tested. Unless one has formulated a hypothesis before running the analyses, variable selection and *post hoc* testing of hypothesis is unacceptable because the *P* values will be overstated.

In almost all publications, information on model diagnostics (e.g. residual plots, influence statistics, VIF, etc.) and cross-validation tests were lacking. This makes the objective selection of existing models from databases very difficult. In future, it is important that authors include model diagnostics and results of cross-validation tests as Supplementary materials.

Measures of error and uncertainty (e.g. SE, 95% CIs) in the estimated coefficients are also rarely reported. Where different models were compared, measures of bias are often reported without an estimate of the error associated with them. This can be misleading because the model with the smallest bias could in reality have large errors if large under-estimates and over-estimates off-set each other. Therefore, I recommend reporting of parameter estimates and measures of bias together with their PRSEs. Models that have one or more coefficients with PRSE > 25% should also not be published.

In addition, I propose the following questions as a checklist for prospective authors: (1) Have the model selection/fitting steps been adequately described? (2) Are parameter estimates reported together with measures of error? (3) Are model diagnostic and cross-validation tests included? (4) Is the reporting of other statistical tests complete and informative? These questions will also help reviewers to judge the merit of the paper objectively. I would also like to reiterate the call by Mascaro et al. (2011) that published models should be accompanied by the raw data as an appendix. This will allow individual investigators to judge whether the model published is adequate for their needs.

4. Conclusions

This review identified some confusion in the use of the term "allometry" and a number of mistakes in the forest biomass estimation literature. I have illustrated how the allometric exponent is prone to estimation and interpretation errors. I have also illustrated how the use of BEMs with a large number of predictors can lead to tailoring the model too much to the current data at the expense of generalizability, interpretability and ecological insight. Therefore, the importance of choosing a reasonably simple functional expression involving the minimum of non-interpretable parameters cannot be overemphasized. Simpler models are those with fewer parameters (2–4) and "independent" predictors (1–3) and without polynomial terms. For example, a model with D and ρ will have 3 parameters to estimate and more tractable than one with D, ρ and H. Simple predictive models are preferred for a number of practical reasons: (1) simpler models are easier to put to test in replication and cross-validation tests; (2) fewer trees need to be cut to estimate their parameters with certainty; (3) simpler models suffer less from influence statistics and collinearity; (4) parameters of simpler models are easier to interpret in a biologically meaningful way. However, this does not mean models of more complex mathematical form should not be used if the objective is description rather than prediction. If predictive models are to be developed, clear hypotheses should be formulated before data collection. The review has also revealed a number of non-trivial mistakes and inconsistencies in the formulation of equations. If not corrected, these can have far-reaching consequences on the integrity of the BEMs and their practical application.

Acknowledgements

I am grateful to Shem Kuyah for the insightful discussions regarding problems in forest biomass estimation and sharing his data I used as one of the examples. I am also indebted to the researchers who made their original data available online as supplementary online materials or within their papers, thereby

enabling me (and others) to re-examine their findings from a different perspective.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foreco.2014.06.026.

References

Ahmed, R., Siqueira, P., Hensley, H., Bergen, K., 2013. Uncertainty of forest biomass estimates in north temperate forests due to allometry: implications for remote sensing. Remote Sens. 5, 3007–3036.

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. Manage. 267, 297–308.

Andersen, R., 2008. Modern Methods for Robust Regression. Sage Publications Inc., Thousand Oaks, USA, pp 128.

Anfodillo, T., Carrer, M., Simini, F., Popa, I., Banavar, J.R., Maritan, A., 2012. An allometry-based approach for understanding forest structure, predicting tree-size distribution and assessing the degree of disturbance. Proc. R. Soc. B 280, 20122375.

Basuki, T.M., van Laake, P.E., Skidmore, A.K., Hussin, Y.A., 2009. Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests. For. Ecol. Manage. 257, 1684–1694.

Brokaw, N., Thompson, J., 2000. The H for DBH. For. Ecol. Manage. 129, 89–91.

Brown, J.H., Gupta, V.K., Li, B.L., Milne, B.T., Restrepo, C., West, G.B., 2002. The fractal nature of nature: power laws, ecological complexity and biodiversity. Phil. Trans. R. Soc. Lond. B 357, 619–626.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practice Information-Theoretic Approach. Springer Verlag, New York.

Carrascal, L.M., Galván, I., Gordo, O., 2009. Partial least squares regression as an alternative to current regression methods used in ecology. Oikos 118, 681–690.

CDC/National Center for Health Statistics, 2010. Reliability of estimates. <http://www.cdc.gov/nchs/ahcd/ahcd_estimation_reliability.htm> (access 15.06.14).

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.-P., Nelson, B.W., Ogawa, H., Puig, H., Riéra, B., Yamakura, T., 2005. Tree allometry and improved estimation of carbonstocks and balance in tropical forests. Oecologia 145, 87–99.

Cheng, D., Niklas, K., 2007. Above- and below-ground biomass relationships across 1543 forested communities. Ann. Bot. 99, 95–102.

Clark, D.B., Kellner, J.R., 2012. Tropical forest biomass estimation and the fallacy of misplaced concreteness. J. Veg. Sci. 23, 1191–1196.

Colgan, M.S., Asner, G.P., Swemmer, T., 2013. Harvesting tree biomass at the stand level to assess the accuracy of field and airborne biomass estimation in savannas. Ecol. Appl. 23, 1170–1184.

Coomes, D.A., Allen, R.B., 2009. Testing the metabolic scaling theory of tree growth. J. Ecol. 97, 1369ced co.

Djomo, A.N., Ibrahima, A., Saborowski, J., Gravenhorst, G., 2010. Allometric equations for biomass estimations in Cameroon and pan moist tropical equations including biomass data from Africa. For. Ecol. Manage. 260, 1873–1885.

Ducey, M.J., 2012. Evergreenness and wood density predict height–diameter scaling in trees of the northeastern United States. For. Ecol. Manage. 279, 21–26.

Dumont, C., Mentre, F., Gaynor, C., Brendel, K., Gesson, C., Chenel, M., 2013. Optimal sampling times for a drug and its metabolite using SIMCYP-simulations as prior information. Clin. Pharmacokinet. 52, 43–57.

Ebuy, J., Lokombé, J.P., Ponette, Q., Sonwa, D., Picard, N., 2011. Allometric equation for predicting aboveground biomass of three tree species. J. Trop. For. Sci. 23, 125–132.

Enquist, B.J., 2002. Universal scaling in tree and vascular plant allometry: toward a general quantitative theory linking plant form and function from cells to ecosystems. Tree Physiol. 22, 1045–1064.

Enquist, B.J., West, G.B., Charnov, E.L., Brown, J.H., 1999. Allometric scaling of production and life-history variation in vascular plants. Nature 401, 907–911.

Enquist, B.J., West, G.B., Brown, J.H., 2009. Extensions and evaluations of a general quantitative theory of forest structure and dynamics. Proc. Natl. Acad. Sci. USA 106, 7046–7051.

Fayolle, A., Doucet, J.L., Gillet, J.F., Bourland, N., Lejeune, P., 2013. Tree allometry in Central Africa: testing the validity of pantropical multi-species allometric equations for estimating biomass and carbon stocks. For. Ecol. Manage. 305, 29–37.

Fehrmann, L., Kleinn, C., 2006. General considerations about the use of allometric equations for biomass estimation on the example of Norway spruce in central Europe. For. Ecol. Manage. 236, 412–421.

Feldpausch, T.R., Banin, L., Phillips, O.L., Baker, T.R., Lewis, S.L., et al., 2011. Height-diameter allometry of tropical forest trees. Biogeosciences 8, 1081–1106.

Fortsche, B.K., 1963. Chaos in the brickyard. Science 142, 339.

Friedman, J., Bohonak, A.J., Levine, R.A., 2013. When are two pieces better than one: fitting and testing OLS and RMA regressions. Environmetrics 24, 306–316.

Gayon, J., 2000. History of the concept of allometry. Amer. Zool. 40, 748–758.

Girardin, C.A.J., Malhi, Y., Aragao, L.E.O.C., Mamani, M., et al., 2010. Net primary productivity allocation and cycling of carbon along a tropical forest elevational transect in the Peruvian Andes. Glob. Change Biol. 16, 3176–3192.

Gould, S.J., 1966. Allometry and size in ontogeny and phylogeny. Biol. Rev. 41, 587–640.

Graham, M.H., 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84, 2809–2815.

Guthery, F.S., Brennan, L.A., Peterson, M.J., 2005. Information theory in wildlife science: critique and viewpoint. J. Wildlife Manage. 69, 457–465.

Hendriks, J.A., 2007. The power of size: a meta-analysis reveals consistency of allometric regressions. Ecol. Model. 205, 196–208.

Henry, M., Besnard, A., Asante, W.A., 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. Manage. 260, 1375–1388.

Henry, M., Picard, N., Trotta, C., Manlay, R., Valentini, R., Bernoux, M., Saint-André, L., 2011. Estimating tree biomass of sub-Saharan African forests: a review of available allometric equations. Silva Fennica 45, 477–569.

Henry, M., Bombelli, A., Trotta, C., Alessandrini, A., Birigazzi, L., et al., 2013. GlobAllomeTree: international platform for tree allometric equations to support volume, biomass and carbon assessment. iForest 6, 326–330.

Huang, S., Price, D., Titus, S.J., 2000. Development of ecoregion-based height–diameter models for white spruce in boreal forests. For. Ecol. Manage. 129, 125–141.

Hui, D., Wang, J., Shen, W., Le, X., Ganter, P., et al., 2014. Near isometric biomass partitioning in forest ecosystems of China. PLoS ONE 9 (1), e86550.

Hunter, M.O., Keller, M., Vitoria, D., Morton, D.C., 2013. Tree height and tropical forest biomass estimation. Biogeosciences 10, 8385–8399.

Huxley, J.S., Teissier, G., 1936. Terminology of relative growth. Nature 137, 780–781.

IPCC, 2003. In: Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Wagner, F., (Eds.). Good Practice Guidance for Land Use, Land-use Change and Forestry. Institute for Global Environmental Strategies (IGES), Hayama, Japan, 600 pp.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning with Applications in R. Springer Science+Business Media, New York, 426 pp.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2004. Comprehensive Database of Diameter-Based Biomass Regressions for North American Tree Species. USDA General Technical Report NE-319.

Kaitaniemi, P., 2004. Testing the allometric scaling laws. J. Theor. Biol. 228, 149–153.

Kearsley, E., de Haulleville, T., Hufkens, K., et al., 2013. Conventional tree height–diameter relationships significantly overestimate aboveground carbon stocks in the Central Congo Basin. Nat. Commun. 4. http://dx.doi.org/10.1038/ncomms3269.

King, D.A., Davies, S.J., Tan, S., Noor, N.S.M., 2009. Trees approach gravitational limits to height in tall lowland forests of Malaysia. Funct. Ecol. 23, 284–291.

Kuyah, S., Dietz, J., Muthuria, C., Jamnadassa, R., Mwangi, P., Coe, R., et al., 2012. Allometric equations for estimating biomass in agricultural landscapes: I. Aboveground biomass. Agric Ecosyst. Environ. 158, 216–224.

Kuyah, S., Sileshi, G.W., Njoloma, J., Mng'omba, S., Neufeldt, H., 2014. Estimating aboveground tree biomass in three different miombo woodlands and associated land use systems in Malawi. Biomass Bioenergy 66, 214–222.

Larjavaara, M., Muller-Landau, H.C., 2013. Measuring tree height: a quantitative comparison of two common field methods in a moist tropical forest. Method. Ecol. Evol. 4, 793–801.

Li, H., Zhao, P., 2013. Improving the accuracy of tree-level aboveground biomass equations with height classification at a large regional scale. For. Ecol. Manage. 289, 153–163.

Lima, A.J.N., Suwa, R., De Mello Ribeiro, G.H. Pires, Kajimoto, T., dos Santos, J., et al., 2012. Allometric models for estimating above- and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro. Brazil. For. Ecol. Manage. 277, 163–172.

Lukacs, P.M., Burnham, K.P., Anderson, D.R., 2010. Model selection bias and Freedman's paradox. Ann. Inst. Stat. Math. 62, 117–125.

MacCoun, R.J., 1998. Biases in the interpretation and use of research results. Annu. Rev. Psychol. 49, 259–287.

MacFarlane, D.W., Kuyah, S., Mulia, R., Dietz, J., Muthuri, C., Van Noordwijk, M., 2014. Evaluating a non-destructive method for calibrating tree biomass equations derived from tree branching architecture. Trees 28, 807–817.

Marquet, P.A., Quiñones, R.A., Abades, S., Labra, F., Tognelli, M., Arim, M., Rivadeneira, M., 2005. Scaling and power-laws in ecological systems. J. Expl. Biol. 208, 1749–1769.

Mascaro, J., Litton, C.M., Hughes, F., Uowolo, A., Schnitzer, S.A., 2011. Minimizing Bias in biomass allometry: model selection and log-transformation of data. Biotropica 43, 649–653.

Mate, R., Johansson, T., Sitoe, A., 2014. Biomass equations for tropical forest tree species in Mozambique. Forests 5, 535–556.

Mbow, C., Verstraete, M.M., Sambou, B., Diaw, A.T., Neufeldt, H., 2013. Allometric models for aboveground biomass in dry savanna trees of the Sudan and Sudan-Guinean ecosystems of Southern Senegal. J. For. Res. 19, 340–347.

McCune, B., Grace, J.B., 2002. Analysis of Ecological Communities. Gleneden Beach, OR, USA, 304 pp.

Molto, Q., Rossi, V., Blanc, L., 2013a. Error propagation in biomass estimation in tropical forests. Method. Ecol. Evol. 4, 175–183.

Molto, Q., Hérault, B., Boreux, J.-J., Daullet, M., Rousteau, A., Rossi, V., 2013b. Predicting tree heights for biomass estimates in tropical forests. Biogeosci. Discuss. 10, 8611–8635.
Moncrieff, G.R., Chamaillié-Jammes, S., Higgins, S.I., O'Hara, R.B., Bond, W.J., 2011. Tree allometries reflect a lifetime of herbivory in an African savanna. Ecology 92, 2310–2315.
Moore, J.R., 2010. Allometric equations to predict the total above-ground biomass of radiata pine trees. Ann. For. Sci. 67, 806.
Muller-Landau, H.C., Condit, R.S., Chave, J., Thomas, S.C., et al., 2006. Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. Ecol. Lett. 9, 575–588.
Návar, J., 2009. Biomass component equations for Latin American species and groups of species. Ann. For. Sci. 66, 208–216.
Ngomanda, A., Obiang, N.L.E., Lebamba, J., Mavouroulou, Q.M., Gomat, H., Mankou, G.S., Loumeto, J., Iponga, D.M., Ditsouga, F.K., Koumba, R.Z., Bobé, K.H.B., Okouyi, C.M., Nyangadouma, R., Lépengué, N., Mbatchi, B., Picard, N., 2014. Site-specific versus pantropical allometric equations: Which option to estimate the biomass of a moist central African forest? For. Ecol. Manag. 312, 1–9.
Niklas, K.J., 1995. Size-dependent allometry of tree height, diameter and trunk-taper. Ann. Bot. 75, 217–227.
Niklas, K.J., Spatz, H.-C., 2004. Growth and hydraulic (not mechanical) constraints govern the scaling of tree height and mass. Proc. Natl. Acad. Sci. USA 101, 15661–15663.
Nogueira, E.M., Fearnside, P.M., Nelson, B.W., Barbosa, R.I., Keizer, E.W.H., 2008. Estimates of forest biomass in the Brazilian Amazon: new allometric equations and adjustments to biomass from wood-volume inventories. For. Ecol. Manage. 256, 1853–1867.
Osunkoya, O.O., Omar-Ali, K., Amit, N., Dayan, J., Daud, D.S., Sheng, T.K., 2007. Comparative height–crown allometry and mechanical design in 22 tree species of Kuala Belalong rainforest, Brunei. Borneo. Am. J. Bot. 94, 1951–1962.
Packard, G.C., 2014. Multiplicative by nature: logarithmic transformation in allometry. J. Exp. Zool. 322, 202–207.
Packard, G.C., Birchard, G.F., Boardman, T.J., 2011. Fitting statistical models in bivariate allometry. Biol. Rev. 86, 549–563.
Paul, K., Roxburgh, S.H., England, J.R., Ritson, P., et al., 2013. Development and testing of allometric equations for estimating above-ground biomass of mixed-species environmental plantings. For. Ecol. Manage. 310, 483–494.
Picard, N., Saint-André, L., Henry, M., 2012. Manual for Building Tree Volume and Biomass Allometric Equations: From Field Measurement to Prediction. FAO, Rome, and CIRAD, Montpellier, 215 pp.
Pilli, P., Anfodillo, T., Carrer, M., 2006. Towards a functional and simplified allometry for estimating forest biomass. For. Ecol. Manage. 237, 583–593.
Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: observed vs predicted or predicted vs observed. Ecol. Model. 216, 316–322.
Pretzsch, H., Dieler, J., 2012. Evidence of variant intra- and interspecific scaling of tree crown structure and relevance for allometric theory. Oecologia 169, 637–649.
Price, C.A., Enquist, B.J., Savage, V.M., 2007. A general model for allometric covariation in botanical form and function. Proc. Natl. Acad. Sci. 104, 13204–13209.
Price, C.A., Ogle, K., White, E.P., Weitz, J.S., 2009. Evaluating scaling models in biology using hierarchical Bayesian approaches. Ecol. Lett. 12, 641–651.
Russo, S.E., Wiser, S.K., Coomes, D.A., 2008. A re-analysis of growth–size scaling relationships of woody plant species. Ecol. Lett. 11, 311–312.
Rutishauser, E., Noor'an, F., Laumonier, Y., Halperin, J., Hergoualc'h Rufi'ie, K., Verchot, L., 2013. Generic allometric models including height best estimate forest biomass and carbon stocks in Indonesia. For. Ecol. Manage. 307, 219–225.
Sawadogo, L., Savadogo, P., Tiveau, D., Dayamba, S.D., Zida, D., Nouvellet, Y., Oden, P.C., Guinko, S., 2010. Allometric prediction of above-ground biomass of eleven woody tree species in the Sudanian savanna-woodland of West Africa. J. Forest. Res. 21, 475–481.
Segura, M., Kanninen, M., 2005. Allometric models for tree volume and total aboveground biomass in a tropical humid forest in Costa Rica. Biotropica 37, 2–8.
Shea, B.T., 1985. Bivariate and multivariate growth allometry: statistical and biological considerations. J. Zool. 206, 367–390.
Singh, V., Tewari, A., Kushwaha, S.P.S., Dadhwal, V.K., 2011. Formulating allometric equations for estimating biomass and carbon stock in small diameter trees. For. Ecol. Manage. 261, 1945–1949.
Somogyi, Z., Cienciala, E., Mäkipä, P., Muukkonen, P., Lehtonen, A., Weiss, P., 2007. Indirect methods of large-scale forest biomass estimation. Eur. J. Forest. Res. 126, 197–207.
Stephenson, N.L., Das, A.J., Condit, R., Russo, S.E., 2014. Rate of tree carbon accumulation increases continuously with tree size. Nature 507, 90–93.
Taskinen, S., Warton, D.I., 2013. Robust tests for one or more allometric lines. J. Theor., Biol. 333, 38–46.
Ter-Mikaelian, M.T., Korzukhin, M.D., 1997. Biomass equations for sixty-five North American tree species. For. Ecol. Manage. 97, 1–24.
Tredennick, A.T., Bentley, L.P., Hanan, N.P., 2013. Allometric convergence in savanna trees and implications for the use of plant scaling models in variable ecosystems. PLoS ONE 8 (3), e58241.
van Breugel, M., Ransijn, J., Craven, D., Bongers, F., Hall, J.S., 2011. Estimating carbon stock in secondary forests: Decisions and uncertainties associated with allometric biomass models. For. Ecol. Manage. 262, 1648–1657.
Vieilledent, G., Vaudry, R., Andriamanohisoa, S.F.D., Rakotonarivo, S.O., Randrianasolo, Z.H., et al., 2012. A universal approach to estimate biomass and carbon stock in tropical forests using generic allometric models. Ecol. Appl. 22, 572–583.
Walker, S.M., Pearson, T., Casarim, F.M., Harris, H., Petrova, S., Grais, A., Swails, E., Netzer, M., Goslee, K.M., Brown, S., 2012. Standard Operating Procedures for Terrestrial Carbon Measurement, Version 2012. Winrock International.
Warton, D.I., Wright, I.J., Falster, D.S., Westoby, M., 2006. Bivariate line fitting methods for allometry. Biol. Rev. 85, 259–291.
Watt, M.S., Kirschbaum, M.U.F., 2011. Moving beyond simple linear allometric relationships between tree height and diameter. Ecol. Model. 222, 3910–3916.
West, G.B., Brown, J.H., Enquist, B.J., 1999. The fourth dimension of life: fractal geometry and allometric scaling of organisms. Science 284, 1677–1679.
West, G.B., Enquista, B.J., Browna, J.H., 2009. A general quantitative theory of forest structure and dynamics. Proc. Natl. Acad. Sci. USA 106, 7040–7045.
Wiemann, M.C., Williamson, G.B., 2014. Wood Specific Gravity Variation with Height and its Implications for Biomass Estimation. Forest Products Laboratory Research Paper FPL-RP-677. USD, Forest Service, Madison, WI, 9 p.
Xiao, X., White, E.P., Hooten, M.B., Durham, S.L., 2011. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. Ecology 92, 1887–1894.
Yang, Y., Luo, Y., 2011. Isometric biomass partitioning pattern in forest ecosystems: evidence from temporal observations during stand development. J. Ecol. 99, 431–437.
Yao, X., Fu, B., Lu, Y., Sun, F., Wang, S., et al., 2013. Comparison of four spatial interpolation methods for estimating soil moisture in a complex terrain catchment. PLoS ONE 8 (1), e54660.
Zapata-Cuartas, M., Sierra, C., Alleman, L., 2012. Probability distribution of allometric coefficients and Bayesian estimation of aboveground tree biomass. For. Ecol. Manage. 277, 173–179.
Zhao, F., Guo, Q., Kelly, M., 2012. Allometric equation choice impacts lidar-based forest biomass estimates: a case study from the Sierra National Forest, CA. Agric. Forest. Meteorol. 165, 64–72.
Zianis, D., 2008. Predicting mean aboveground forest biomass and its associated variance. For. Ecol. Manage. 256, 1400–1407.
Zianis, D., Mencuccini, M., 2004. On simplifying allometric analyses of forest biomass. For. Ecol. Manage. 187, 311–332.