Assessment of Bias in Pan-Tropical Biomass Predictions

Andrew Burt [1]*, Kim Calders [2], Aida Cuni-Sanchez [3], Jose Gómez-Dans [1,4], Philip Lewis [1,4], Simon L. Lewis [1,5], Yadvinder Malhi [6], Oliver L. Phillips [5] and Mathias Disney [1,2]

[1] Department of Geography, University College London, London, United Kingdom, [2] CAVElab - Computational and Applied Vegetation Ecology, Ghent University, Ghent, Belgium, [3] Environment Department, University of York, Heslington, United Kingdom, [4] NERC National Centre for Earth Observation (NCEO), Leicester, United Kingdom, [5] Ecology and Global Change, School of Geography, University of Leeds, Leeds, United Kingdom, [6] Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, United Kingdom

Above-ground biomass (AGB) is an essential descriptor of forests, of use in ecological and climate-related research. At tree- and stand-scale, destructive but direct measurements of AGB are replaced with predictions from allometric models characterizing the correlational relationship between AGB, and predictor variables including stem diameter, tree height and wood density. These models are constructed from harvested calibration data, usually via linear regression. Here, we assess systematic error in out-of-sample predictions of AGB introduced during measurement, compilation and modeling of in-sample calibration data. Various conventional bivariate and multivariate models are constructed from open access data of tropical forests. Metadata analysis, fit diagnostics and cross-validation results suggest several model misspecifications: chiefly, unaccounted for inconsistent measurement error in predictor variables between in- and out-of-sample data. Simulations demonstrate conservative inconsistencies can introduce significant bias into tree- and stand-scale AGB predictions. When tree height and wood density are included as predictors, models should be modified to correct for bias. Finally, we explore a fundamental assumption of conventional allometry, that model parameters are independent of tree size. That is, the same model can provide predictions of consistent trueness irrespective of size-class. Most observations in current calibration datasets are from smaller trees, meaning the existence of a size dependency would bias predictions for larger trees. We determine that detecting the absence or presence of a size dependency is currently prevented by model misspecifications and calibration data imbalances. We call for the collection of additional harvest data, specifically under-represented larger trees.

Keywords: tropical forests, above-ground biomass, allometry, prediction, error, uncertainty

1. INTRODUCTION

Above-ground biomass, AGB, is central to assessments of forest state and change because of its relationship with the carbon cycle, and ecosystem services including net primary production (Field et al., 1998; Pan et al., 2011; Costanza et al., 2014; Martin et al., 2018). The above-ground biomass of a particular tree, at a given point in time, is the result of lifetime cumulative gross primary production P_g, respiration, r, and loss, d (Roberts et al., 1993).

To measure the AGB of a tree would require: (i) its harvesting flush with the ground, (ii) the removal of water through drying, and (iii) its mass measured via weighing. These destructive measurements are necessarily limited due to their difficulty. Instead, measurements of AGB are replaced at the tree- and stand-scale with estimates predicted from allometrics (Picard et al., 2012). Allometric models exploit the correlational relationships that exist between AGB, and more readily measurable tree parameters (e.g., stem diameter, D, tree height, H, and wood density, ρ). These relationships are discovered empirically, from calibration data where AGB has been directly measured via destructive harvest, concurrent with measurement of the predictor variables.

The conventional approach for modeling such calibration data is a combination of ordinary least squares (OLS) linear regression and log-log transformation. OLS is favored because of broad coverage in the wider statistical literature and its long history in the field of allometry since introduction in the 1900s (Lapicque, 1907; Huxley, 1932). The log-log transformation is undertaken because AGB is usually observed to scale with predictor variables such as D according to a power law (Brown, 1997), and it is a convenient approach for modeling the multiplicative nature of plant growth: variance in AGB normally increases with tree size (Kerkhoff and Enquist, 2009).

Once an allometric model has been constructed from some underlying in-sample calibration data, an out-of-sample prediction of tree AGB is made by inputting the measurements of the predictor variables of that tree into the model. Stand-scale AGB is estimated by summing predictions for every tree inside a particular forest stand.

For tropical forests, these models are most often constructed at the population-scale because of diversity (Gibbs et al., 2007), implying calibration data must represent upwards of 40 000 species (Slik et al., 2015). Occasionally, more constrained models of tropical forests are available, where calibration data were acquired exclusively from either a geographic subsection of the tropics, or from a specific plant taxa (Basuki et al., 2009). However, in the interests of consistency, or because of the general unavailability of these more specific models, pan-tropical models are usually preferred.

A number of pan-tropical models exist (Henry et al., 2013), but a few have become particularly prominent (e.g., Brown et al., 1989; Chave et al., 2005, 2014; Feldpausch et al., 2012), and their subsequent predictions of stand-scale AGB are the cornerstone to multiple activities across the environmental sciences. One example usage of these predictions is to calibrate remotely sensed signals from earth observation instruments, from which regional- and global-scale AGB products are derived (Saatchi et al., 2011; Baccini et al., 2012; Avitabile et al., 2016). Another example is their provision of reference AGB stocks to intergovernmental initiatives on climate change, such as the UN-REDD program (Angelsen et al., 2012).

Such example applications of allometry will usually require error in out-of-sample predictions of stand-scale AGB be well-understood. Throughout this paper, we use ISO 5725 and BIPM definitions to describe concepts of random error, systematic error, total error, precision, trueness, accuracy, bias

and uncertainty (ISO-5725-1:1994(en), 1994; Menditto et al., 2007; JCGM-200:2012, 2012). These definitions are described below and illustrated in **Figure 1**.

Total error then, is the allometric-derived prediction of tree- or stand-scale AGB minus the true (or reference) value of AGB. Total error is the sum of two components: (i) systematic error, which describes predictable non–zero mean offsets from the true value, and (ii) random error, which describes unpredictable zero mean offsets from the true value. Trueness, precision, and accuracy are qualitative terms describing the effect on the performance of a prediction by systematic, random and total error respectively. These qualitative performance characteristics are quantitatively expressed as a bias, standard deviation and uncertainty respectively. That is, the uncertainty of a prediction should account for both systematic and random error.

Error in out-of-sample allometric predictions is potentially introduced during the selection, measurement and modeling of the in-sample calibration data, as well as in the measurement of the out-of-sample data. Possible sources of random error include: (R1) noise in the measurement of the in-sample calibration data, (R2) variance in the subsequently constructed model, which arises from the stochastic nature of plant allometry, and (R3) noise in the measurement of the out-of-sample data. Possible sources of systematic error include: (S1) biased measurement of the in-sample calibration data, (S2) bias introduced by the selected modeling methods, (S3) the possibility that the in-sample data are unrepresentative of the out-of-sample data, and (S4) biased measurement of the out-of-sample data.

The gold standard for quantifying these uncertainties in out-of-sample tree- or stand-scale predictions is direct measurement via destructive harvest. However, across tropical forests, direct measurement at the stand-scale has never been undertaken. Aside from the difficulties associated with large-scale destructive harvest, this is perhaps also due to the so-called "fallacy of misplaced concreteness" (Clark and Kellner, 2012). That is, uncertainty associated with these predictions is often ignored as a result of erroneously deeming them reference measurements, rather than the estimates they are. Indeed, only a small body of literature has considered uncertainty in out-of-sample pan-tropical predictions of AGB (Chave et al., 2004; Molto et al., 2013; Picard et al., 2015a; Réjou-Méchain et al., 2017). As outlined below, the focus of these studies has been on the precision of predictions (i.e., the effect of random error), with particular attention to sources R2 and R3.

Chave et al. (2004), using an OLS model constructed from a compilation of pan-tropical calibration data, found relative uncertainty to approximate 5–10 % at the 1 ha stand-scale (Chave et al., 2014), when accounting for source R2 using the standard error of the regression, and R3 using a Taylor series expansion. Réjou-Méchain et al. (2017), using the same model and calibration data, but perturbing the parameters of the model via a Bayesian framework to simulate further error arising from R2, found relative uncertainty to approximate 10 % at the 1 ha stand-scale (Chave et al., 2019). Picard et al. (2015a) considered R2 still further, and cognisant of the multiple, nominally suitable allometric models available, estimated their aggregate variance
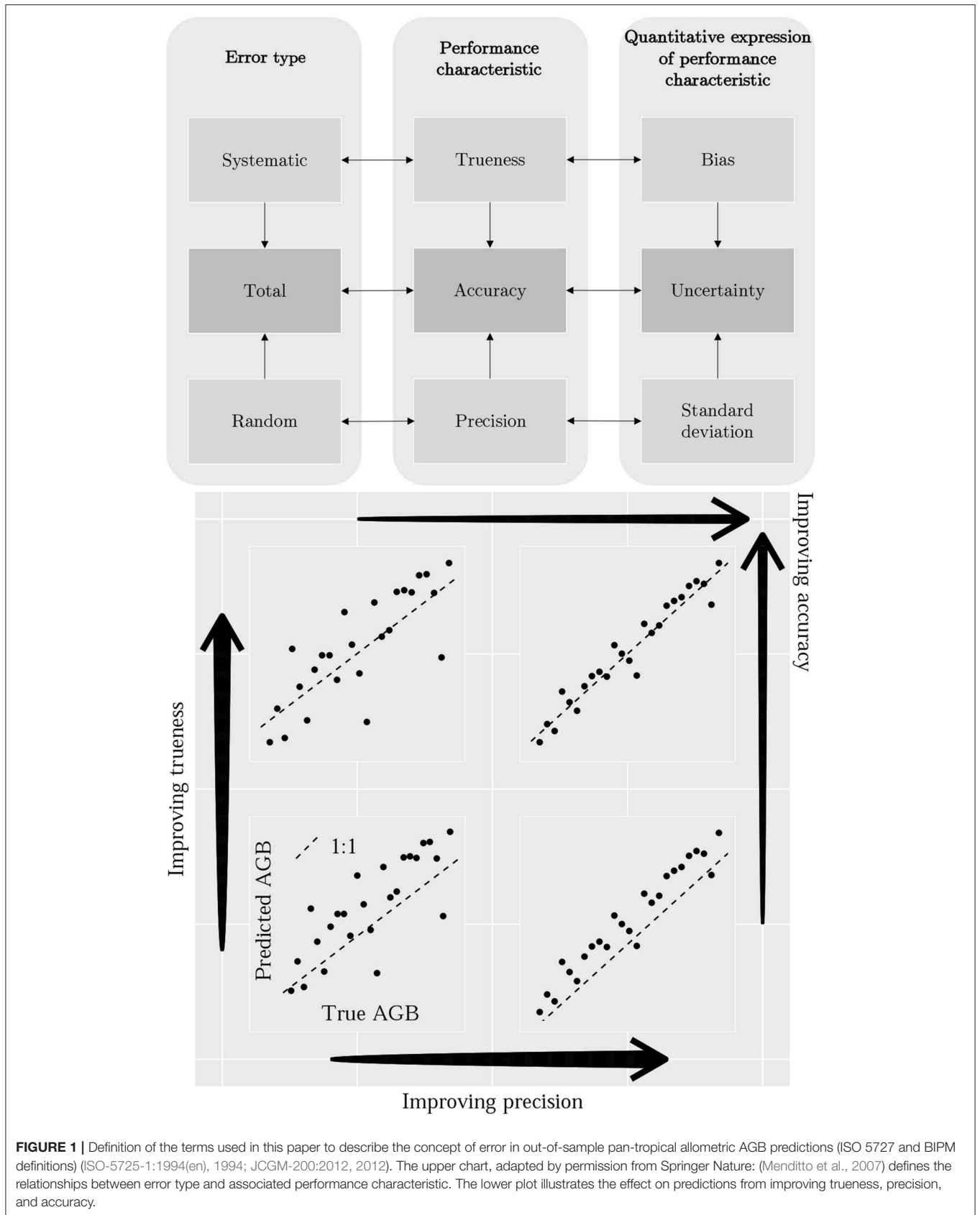
FIGURE 1 | Definition of the terms used in this paper to describe the concept of error in out-of-sample pan-tropical allometric AGB predictions (ISO 5727 and BIPM definitions) (ISO-5725-1:1994(en), 1994; JCGM-200:2012, 2012). The upper chart, adapted by permission from Springer Nature: (Menditto et al., 2007) defines the relationships between error type and associated performance characteristic. The lower plot illustrates the effect on predictions from improving trueness, precision, and accuracy.

using Bayesian model averaging, from which relative uncertainty was found to approximate 44 % at the 1 ha stand-scale.

Whilst the contribution to uncertainty by random error has received some attention, the contribution by systematic error has received considerably less. Here, we focus on systematic error in allometric-derived pan-tropical predictions of AGB, with particular attention to sources S1, S2, and S4. That is, we focus on bias introduced during the measurement and modeling of the in-sample calibration data, and measurement of the out-of-sample data.

Initially, we undertake a review of the metadata of existing pan-tropical calibration data, to note the measurement methods particular to destructive harvest experiments. We then review the underlying assumptions of OLS modeling necessary to justify unbiased predictions of out-of-sample AGB. Using open access pan-tropical calibration data, we construct several conventional models, and test whether these assumptions are met using various fit diagnostics and statistical tests. We assess and compare the precision, trueness and accuracy of predictions from these models using bootstrapping and cross-validation.

We identify several potential sources of bias, simulate their likely influence, and discuss their implication for OLS predictions of tree- and stand-scale AGB. We suggest some recommendations for quantifying and minimizing bias during the measurement and compilation of calibration data. We also discuss approaches for minimizing bias during the modeling of calibration data.

Finally, we consider one aspect of error source S3: the possibility that in-sample data are unrepresentative of out-of-sample data. We assess whether pan-tropical allometry is independent of tree size; that is, an assumption of conventional pan-tropical allometry is that the model parameters necessary for predicting the AGB of a tree are constant regardless of tree size. This is necessary to consider because calibration datasets are often imbalanced (i.e., the majority of observations will be from small trees, with relatively few large trees) (Duncanson et al., 2015; Jucker et al., 2017). If pan-tropical allometrics are dependent on tree size, these imbalances will introduce a bias into predictions of AGB for under-represented size classes.

2. METHODS

2.1. Pan-Tropical (In-sample) Calibration Data

2.1.1. Characteristics

In this paper, we consider the Chave et al. calibration dataset (Chave et al., 2014). These open access data are currently the most comprehensive available, compiled from many independent destructive harvest experiments over 7 decades, from 58 sites spanning the tropics, with measurements of AGB (kg), D (m), H (m) and ρ_b (kg m^{-3})[1] obtained from 4004 trees. **Figure 2** illustrates these data, presenting a scatter plot of AGB against D, and histograms displaying the distributions of the 4 variables.

It is worth noting two characteristics of the dataset; the first is non–constant variance in AGB. That is, for a given value of D, the

---

[1] ρ_b describes dry mass divided by wet volume.

range of values taken by AGB is not constant across scale; rather, this range increases as D increases. The second characteristic of these data is their non–uniform distribution, with the majority collected from relatively small trees. Some statistics illustrating this: AGB ranges from 1 to 76064 kg; the median and mean values are 98 and 1,134 kg; the first and third quartiles are 22 kg and 491 kg; 5.7 and 2.7 % of these data (by stem count) have AGB > 5,000 and 10,000 kg, respectively.

2.1.2. Metadata Review

In the results section, we report the protocol employed for measuring these calibration data. This is undertaken by reviewing the metadata of some of the individual studies contributing to the compiled dataset. Data considered were those published in English-language peer-reviewed journals, i.e., a total of 26 studies representing 65.3 % of the full data (by stem count) (Edwards and Grubb, 1977; Yamakura et al., 1986; Saldarriaga et al., 1988; Martinez-Yrizar et al., 1992; Brown et al., 1995; Fromard et al., 1998; Araújo et al., 1999; Nelson et al., 1999; Ketterings et al., 2001; Mackensen et al., 2000; Cairns et al., 2003; Brandeis et al., 2006; Burger and Delitti, 2008; Nogueira et al., 2008; Kenzo et al., 2009; Djomo et al., 2010; Henry et al., 2010; Niiyama et al., 2010; Ebuy et al., 2011; Ryan et al., 2011; Alvarez et al., 2012; Vieilledent et al., 2012; Colgan et al., 2013; Mugasha et al., 2013; Goodman et al., 2014; Ngomanda et al., 2014).

For the measurement of D, the measurement device, point of measurement and buttress treatment are recorded. For the measurement of H, the measurement device and whether measurement was made *in situ* or post-felling are noted. For the measurement of AGB, the methods for measuring wet mass and the subsequent conversion to dry mass are recorded. Finally, the methods used in each study for the measurement of ρ_b are also recorded.

2.2. Modeling Calibration Data

2.2.1. Ordinary Least Squares Linear Regression

The conventional approach for constructing a model to predict AGB from these calibration data is ordinary least squares (OLS) linear regression. An OLS model takes the form:

$$y = X\beta + \varepsilon \tag{1}$$

Where, y, is a nx1 vector of n observations of the dependant variable (e.g., AGB); ε, is a nx1 vector of unobserved random error in y; X, is a nxp design matrix of observations of the predictor variables (e.g., D, where p is the number of included predictor variables plus a constant term); and, β, is a px1 vector of the unknown population parameters.

The closed-form OLS solution to estimating β is minimization of the sum of the squared differences between observations and predictions of the dependant variable:

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{2}$$

Although ε are unobserved, they are estimated, and then represented by, the residuals of the model fit, e, $e_i = y_i - \hat{y}_i$ ($i = 1, ..., n$). The standard error of the regression, s, which is the OLS estimate of the standard deviation of ε, σ, which itself
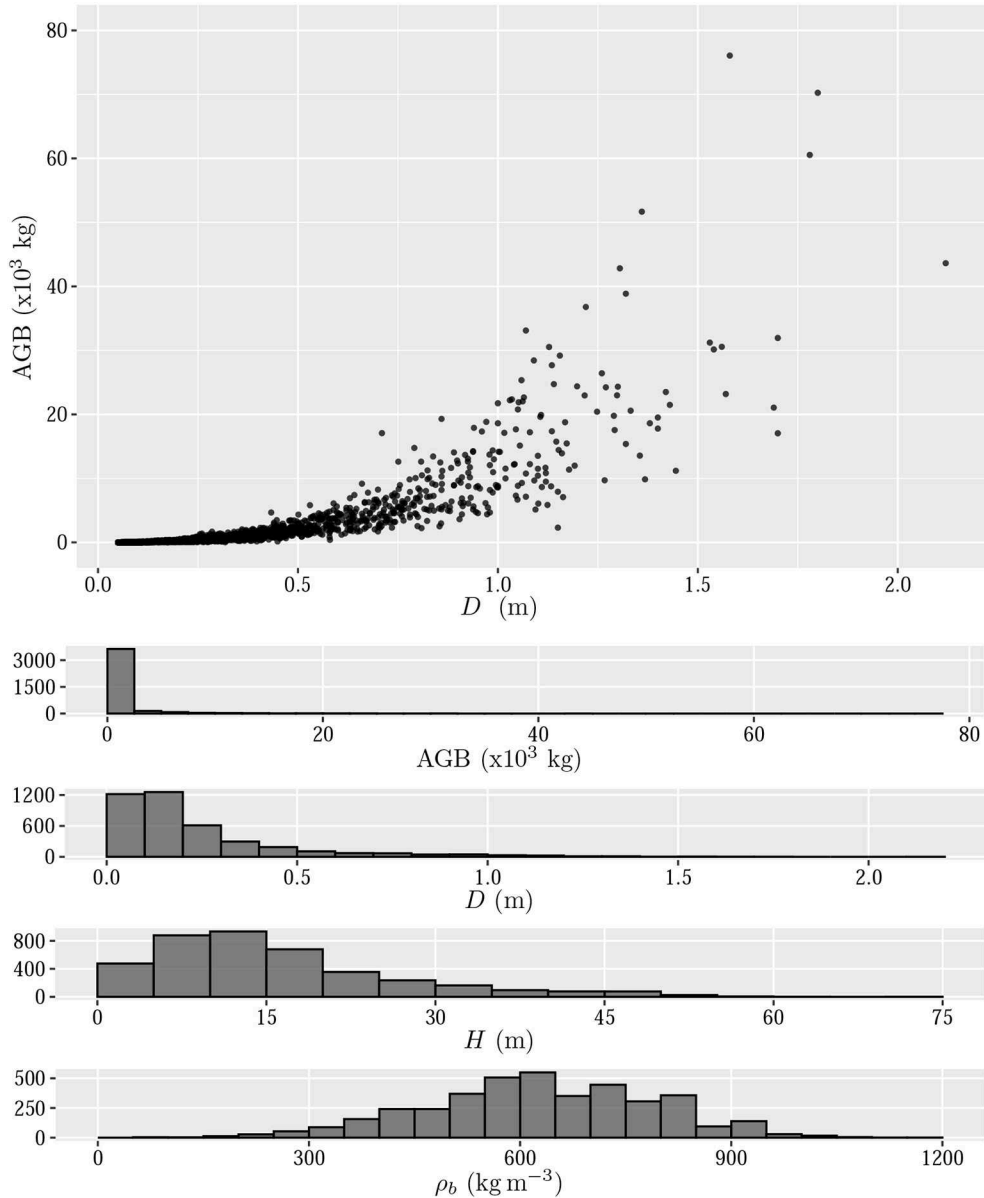
FIGURE 2 | Illustrations of the 4004 destructively harvested trees comprising the Chave et al. (2014) compilation of pan-tropical calibration data. In the upper plot, AGB is plotted against D, the dominant predictor variable in most allometric models; it can be seen that variance in AGB is non–constant. The lower histograms present the distributions of the 4 measured variables across these data. It is noted observations are non–uniformly distributed across the range of each variable, with the majority collected from relatively small trees.

is necessary for frequentist tests of statistical significance (e.g., prediction/confidence intervals), is defined as:

$$s = \sqrt{\frac{1}{(n-p)} \sum_{i=1}^{n} e_i^2} = \hat{\sigma} \qquad (3)$$

2.2.2. Assumptions and Finite-Sample Properties of Ordinary Least Squares

For $\hat{\beta}$ to be an unbiased estimate of β (i.e., the expected value of $\hat{\beta}$ is β, $\mathrm{E}[\hat{\beta}] = \beta$), the following three assumptions must be met (Hayashi, 2000):

1. Linearity: a linear relationship exists between X and y.
2. Strict exogeneity: the expected mean of ε, conditional on X, is zero; which in practice, implies ε is expected to have an unconditional zero mean, and that it is expected X is uncorrelated with ε:

$$\mathrm{E}[\varepsilon] = 0 \qquad (4)$$

$$\mathrm{E}[X^{\mathrm{T}}\varepsilon] = 0 \qquad (5)$$

3. Absence of perfect collinearity: meaning the relationship between the predictor variables is not deterministic,

which would prevent the necessary inversion of X in Equation 2.

For $\hat{\sigma}$ to have unbiased properties, $E[\hat{\sigma}] = \sigma$, and for $\hat{\beta}$ to be efficient (i.e., the variance in their estimate is the minimum), the following two assumptions must be met:

4. Absence of autocorrelation: errors, ε_i $(i = 1, ..., n)$, are uncorrelated.
5. Homoscedasticity: ε_i $(i = 1, ..., n)$, have constant variance, $\mathrm{Var}(\varepsilon) = \sigma^2$.

Finally, one further assumption is sometimes associated with OLS:

6. Normality: the error term is normally distributed:

$$\varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{6}$$

A normal error term is not required for either $\hat{\beta}$ or $\hat{\sigma}$ to retain unbiased properties, but a non–normal ε potentially invalidates t- and F-tests, or the consistency of tests of models selection, such as the Akaike information criterion, whose underlying likelihood function usually expects a normally distributed ε. However, even in such circumstances, if the model is correctly specified as above (assumptions 1–5), a non–normally distributed ε is often dismissed when n is sufficiently large by invoking central limit theorem (Pek et al., 2018).

2.2.3. Predictive Modeling

It is now necessary to consider these assumptions in the context of predicting out-of-sample AGB. These assumptions have been derived with the classical application of regression in mind: causal understanding (Shmueli, 2010). That is, $\hat{\beta}$ are interpreted as explaining the relationship of the predictor variables on the dependant variable (in this context, the term predictor variable would usually be replaced with independent variable).

Here however, our interest with $\hat{\beta}$ is solely to predict a value of AGB, for a given out-of-sample value of X, and to understand the statistical significance of this prediction. By making this fundamental distinction, it is possible to relax or discard several of the above assumptions. For simplicity, for the remainder of this subsection, X is considered to be a single predictor variable, D. For the assumption of strict exogeneity, relevant potential sources of endogeneity include:

1. Omitted variable bias. AGB is not caused by D, but is caused by the aforementioned causal variables (i.e., gross primary production, P_g, respiration, r, and losses, d):

$$AGB = \beta_0 + \beta_2 P_g + \beta_3 r + \beta_4 d + \varepsilon \tag{7}$$

Such that when these causal variables are omitted, and replaced with a non–causal predictor variable, their influence is subsumed into the error term:

$$AGB = \beta_0 + \beta_1 D + u \ (u = \varepsilon + \beta_2 P_g + \beta_3 r + \beta_4 d) \tag{8}$$

Whereby if D were correlated with the combination of omitted variables, then D and ε are correlated, which violates the assumption of strict exogeneity.

2. Systematic error in the measurement of AGB. At the most simple, if a constant bias, c, is present in the measurement of AGB, then the mean of ε is now non–zero:

$$E[\varepsilon] = c \tag{9}$$

Meaning the intercept, β_0, is biased:

$$AGB = (\beta_0 + c) + \beta_1 D + \varepsilon \tag{10}$$

3. Errors-in-variables: OLS assumes predictor variables are measured without error. In the case of a single predictor variable, the model is described as:

$$AGB = \beta_0 + \beta_1 D + \varepsilon \tag{11}$$

Where the OLS estimate of β_1 is:

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(D, AGB)}{\mathrm{Var}(D)} \tag{12}$$

However, suppose D were measured with some random error, $D' = D + \eta$ $[\eta \sim \mathcal{N}(0, \sigma_\eta^2)]$, then the estimate becomes:

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(D, AGB)}{\mathrm{Var}(D) + \mathrm{Var}(\eta)} \tag{13}$$

Meaning D and ε are now correlated via η. This manifests in a downward bias of β_1, which is often termed regression dilution.

The consequences of these various sources of endogeneity differ depending on the application of the model. If the application is explanation, then all three sources bias estimates of β, which we discuss further in the discussion section. If the application is prediction, then systematic error in the measurement of AGB will also persistently bias out-of-sample predictions of AGB. Errors-in-variables do not necessarily bias AGB predictions, although a bias will be present when the measurement error distributions between in- and out-of-sample measurements are inconsistent (Jonsson, 1994; Molto et al., 2013).

Omitted variable bias however, which potentially results in the discovery of so-called spurious relationships, can be ignored in predictive models. That is, the influence of the omitted variables on estimates of β will not bias predictions. However, as also discussed later, omitted variable bias profoundly limits the application of the model outside of prediction.

If the purpose of the model is prediction, we can also largely disregard the assumption of multicollinearity (i.e., significant correlations between the predictor variables, provided the correlation is not perfect) (Hyndman and Athanasopoulos, 2018). Also, because the calibration data are comprised of single, independent observations (with the possible exception of ρ_b, which we discuss further in the discussion section), the assumption of autocorrelation can also be disregarded.

Therefore, for OLS estimated β to have unbiased properties suitable for prediction of AGB, the following assumptions must

be met: (A1) a linear relationship exists between the predictor variables and dependant variable, (A2) the unconditional mean of ε is zero, (A3) measurement error in the predictor variables is consistent between in- and out-of-sample data. Further, $\hat{\sigma}$ has unbiased properties, and $\hat{\beta}$ become efficient, when ε is homoscedastic.

2.2.4. Log-Log Transformation

To achieve the required linearity, given the calibration data exhibit a power law relationship in real-space (**Figure 2**), log-log transformation is necessary:

$$\ln(\text{AGB}) = \beta_0 + \beta_1 \ln(D) + \varepsilon \tag{14}$$

A further beneficial trait of this transformation, in the context of calibration data where the variance in AGB is non–constant, is the increased likelihood of homoscedastic behavior from the residuals.

Once β are estimated, subsequent AGB prediction requires re-transformation of the model to real-space:

$$\text{AGB} = e^{(\hat{\beta}_0 + \varepsilon)}(D)^{\hat{\beta}_1} \tag{15}$$

That is, in real-space, ε is no longer additive (independent of the predictor variables; scale invariant), but multiplicative (dependent on the predictor variables; relative to scale).

A corollary of this re-transformation is that error is described by the log-normal distribution, which does not share the same expectation of the mean with that of the normal distribution. This mismatch introduces a bias, which is usually countered through application of a correction term (Neyman and Scott, 1960), formed using σ, as:

$$\text{AGB} = e^{(\hat{\beta}_0 + \varepsilon)} e^{\left(\frac{\hat{\sigma}^2}{2}\right)}(D)^{\hat{\beta}_1} \tag{16}$$

An implication of employing this correction term is that predictions of AGB are unbiased only when $\hat{\sigma}$ itself is unbiased ($\text{E}[\hat{\sigma}] = \sigma$).

2.2.5. Considered Model Forms

Here, we explore the fit of various model forms to the pan-tropical calibration data. The chosen selection of bivariate and multivariate models covers a range of complexities, given the predictor variables available in the calibration data. The five considered models are:

$$\ln(\text{AGB}) = \beta_0 + \beta_1 \ln(D) + \varepsilon \tag{17}$$

$$\ln(\text{AGB}) = \beta_0 + \beta_1 \ln(H) + \varepsilon \tag{18}$$

$$\ln(\text{AGB}) = \beta_0 + \beta_1 \ln(D) + \beta_2 \ln(\rho_b) + \varepsilon \tag{19}$$

$$\ln(\text{AGB}) = \beta_0 + \beta_1 \ln(D) + \beta_2 \ln(H) + \varepsilon \tag{20}$$

$$\ln(\text{AGB}) = \beta_0 + \beta_1 \ln(D) + \beta_2 \ln(H) + \beta_3 \ln(\rho_b) + \varepsilon \tag{21}$$

Once each model is fitted, we apply several diagnostics and statistical tests to the resulting residuals, e, in an effort to interpret whether the error term, ε, is homoscedastic and normally distributed. Variance of ε is assessed by visually inspecting e plotted against predicted AGB. The Breusch-Pagan and White statistical tests are applied to e to further evaluate variance of ε [null hypotheses: constant variance (homoscedasticity)] (Breusch and Pagan, 1979; White, 1980). The distribution of ε is assessed by comparing the studentised residuals, $\frac{e}{\sigma}$, with the expected normal distribution via a Quantile-Quantile plot.

The variance in $\hat{\beta}$ is quantified using confidence intervals. The classical frequentist approach for generating confidence intervals requires an unbiased estimate of σ. However, as previously identified, $\hat{\sigma}$ has unbiased properties only when ε is homoscedastic. As this assumption may not necessarily hold, confidence intervals are instead generated here using a non–parametric bootstrap.

From the calibration data, a random-with-replacement sample is drawn, from which the five OLS models are constructed. Across N draws, confidence intervals about $\hat{\beta}$, at the level α, are estimated for each model using the bias-corrected and accelerated approach (Efron, 1987).

2.3. Trueness and Accuracy of Predictions

To assess the closeness of agreement between predicted and observed AGB (accuracy) from these models, given both the random error (often referred to in a modeling context as simply variance) which affects precision, and the systematic error (similarly often referred to as bias) which affects trueness, we use k-fold cross-validation.

2.3.1. Stratified k-fold Cross-Validation

The calibration data are folded (or split) k-times, where each fold is a representative subset of the full data. Sequentially iterating through the folds, each of the five considered models are constructed from observations in the unselected folds (training data, $k - 1$). AGB is predicted by each model for each observation in the selected fold (validation data), and compared with observed AGB.

Prediction error is assessed here using the log of the accuracy ratio (Tofallis, 2015). We deliberately avoid the more widely-used mean absolute percentage error (MAPE) because of its undesirable properties including asymmetric penalty, asymmetric bounds and outlier penalty. Instead, the log of the accuracy ratio exhibits symmetric properties, and is particularly well–suited to an assortment of predictions that could reasonably be expected to span five orders of magnitude. The accuracy ratio of a prediction, Q, is defined as:

$$Q = \frac{\widehat{AGB}}{AGB} \tag{22}$$

Where the log of the accuracy ratio is defined as $\ln(Q)$.

To quantify the uncertainty and bias of predictions from each fold, we use two metrics proposed by Morley et al. (2018). First, uncertainty is assessed using the median symmetric accuracy

(MSA):

$$\text{MSA} = 100(e^{(M(|\ln(Q)|))} - 1) \tag{23}$$

Which can be readily interpreted as a percentage error. Second, bias is assessed using the symmetric signed percentage bias (SSPB):

$$\text{SSPB} = 100(sgn(M(\ln(Q)))(e^{|M(\ln(Q))|} - 1)) \tag{24}$$

Which produces a similarly interpretable percentage, whereby a positive or negative sign denotes an over- or under-estimation of the prediction respectively.

2.4. Simulating Inconsistent Measurement Error

Inconsistent measurement error in predictor variables between in- and out-of-sample data can be simulated by adding further noise to the in-sample calibration data themselves, e.g.,:

$$H' = H + \eta \ [\eta \sim \mathcal{N}(0, \sigma_\eta^2)] \tag{25}$$

From which N draws of η are made, and subsequent models constructed. The mean values of $\hat{\beta}$ across these N models are those necessary to provide unbiased predictions of AGB when the out-of-sample data are measured with η-more measurement error than that present in the calibration data.

2.5. Tree Size and Allometry

Finally, the independence of β from tree size is considered. That is, all else being equal, if pan-tropical allometry is independent of tree size, $\hat{\beta}$ should remain statistically indistinguishable between models constructed from subsets of data belonging exclusively to either small or large trees. To explore this, a series of subsets are generated from the data that contain sequentially fewer small trees, removing those below ($D \geq 0.1, 0.25, 0.5$ m, and 0.75 and 1 m). The variance in these model parameters is then estimated using the aforementioned bootstrapped BCa confidence intervals.

2.6. Methods Availability

The source code for these methods, implemented in R, is available in the *treeallom* package, which is released under the MIT license, and hosted at https://github.com/apburt/treeallom.

3. RESULTS

3.1. Review of the Calibration Data

3.1.1. Measurement of D

Across the considered studies, a measuring tape was the most commonly used measurement device, although calipers were occasionally used instead (**Table 1**). The point of measurement was often not reported, but D was referred to as "diameter-at-breast height" or "girth-at-breast-height," which is usually assumed as 1.3 m, although historically this has sometimes been considered 4.5 ft (~1.37 m). Finally, for the treatment of buttresses, two separate approaches were reported: (i) measurement directly above the buttress, and (ii) measurement 0.2 m above.

3.1.2. Measurement of H

Most often, H was measured post-felling using a tape measure (**Table 2**), although a number of studies measured H pre-harvest (i.e., with the tree *in situ*). On the resolution to which H was reported, the majority provided to the nearest 0.1 m, although this was occasionally to the nearest 1 m.

3.1.3. Measurement of AGB

For the measurement of wet mass, some studies weighed each tree in its entirety using scales (Martinez-Yrizar et al., 1992; Nelson et al., 1999; Mackensen et al., 2000; Cairns et al., 2003; Burger and Delitti, 2008; Kenzo et al., 2009; Djomo et al., 2010; Niiyama et al., 2010; Ryan et al., 2011; Vieilledent et al., 2012; Colgan et al., 2013; Mugasha et al., 2013; Ngomanda et al., 2014). Other studies mixed direct measurements with indirect measurements from volume estimates derived from diameter and length measurements. Some studies weighed the crown of each tree, but stem wet mass was derived from volume estimates for some or all trees (Edwards and Grubb, 1977; Saldarriaga et al., 1988; Araújo et al., 1999; Ketterings et al., 2001; Brandeis et al., 2006; Nogueira et al., 2008; Alvarez et al., 2012; Goodman et al., 2014). The remaining studies used volume estimates for stem and large branching for some or all trees (Yamakura et al., 1986; Brown et al., 1995; Fromard et al., 1998; Ebuy et al., 2011; Henry et al., 2010).

There was variation in the treatment of stumps, with some considering everything flush with the ground (Brandeis et al., 2006), whilst others ignored stump material (Ebuy et al., 2011). Few reported on losses from chainsaw cuts: sometimes woody swarf was weighed (Nogueira et al., 2008), and othertimes ignored (Mugasha et al., 2013). No study reported duration between felling and measurement, and on any subsequent water losses. No studies reported applying correction factors to account for either source of loss.

To estimate dry mass (AGB) from wet mass, most often subsamples were gathered from each tree, and their dry-to-wet ratio measured via oven-drying. This was usually undertaken by partitioning the wet mass into pools (e.g., stem, large branches, fine branches, twigs, leaves, and fruit), and taking subsamples from each pool. The type of subsample, the number of pools, and the number of subsamples acquired per pool varied between studies, as did the application of the dry-to-wet ratio (i.e., some derived the mean dry-to-wet ratio across the subsamples that was subsequently applied to total wet mass, whilst others applied the dry-to-wet ratio on a per-pool basis). The temperature at which the subsamples were dried and their final dry mass reported, varied from 55°C (Cairns et al., 2003) to 105°C (Ketterings et al., 2001). Some exceptions to this general approach were the selection of subsamples by height rather than pool (Vieilledent et al., 2012), taking subsamples from only a subsample of the harvested trees (Saldarriaga et al., 1988), and sourcing dry-to-wet ratios from literature (Araújo et al., 1999).

3.1.4. Measurement of ρ_b

Finally, ρ_b was often not measured. Instead, values were obtained from literature (Martinez-Yrizar et al., 1992; Araújo et al., 1999; Ebuy et al., 2011; Ngomanda et al., 2014; Mugasha et al., 2013),

TABLE 1 | The protocol employed across the 26 destructive harvest experiments for measuring the predictor variable D.

Measurement of D					
Measurement device		**Point of measurement**		**Buttress treatment**	
Diameter tape	Brown et al., 1995; Burger and Delitti, 2008; Vieilledent et al., 2012; Mugasha et al., 2013	0.10 m	Colgan et al., 2013	Directly above	Yamakura et al., 1986; Brown et al., 1995; Henry et al., 2010; Alvarez et al., 2012; Goodman et al., 2014
Calipers	Henry et al., 2010; Mugasha et al., 2013	1.30 m	Yamakura et al., 1986; Saldarriaga et al., 1988; Martinez-Yrizar et al., 1992; Brown et al., 1995; Fromard et al., 1998; Araújo et al., 1999; Nelson et al., 1999; Kenzo et al., 2009 Nogueira et al., 2008; Henry et al., 2010; Ryan et al., 2011; Alvarez et al., 2012; Vieilledent et al., 2012; Goodman et al., 2014	0.2 m above	Saldarriaga et al., 1988
DNR	Edwards and Grubb, 1977; Yamakura et al., 1986; Saldarriaga et al., 1988; Martinez-Yrizar et al., 1992; Fromard et al., 1998; Araújo et al., 1999; Nelson et al., 1999 Mackensen et al., 2000; Ketterings et al., 2001; Cairns et al., 2003; Brandeis et al., 2006; Nogueira et al., 2008; Kenzo et al., 2009; Djomo et al., 2010 Niiyama et al., 2010; Ebuy et al., 2011; Ryan et al., 2011; Alvarez et al., 2012; Colgan et al., 2013; Goodman et al., 2014; Ngomanda et al., 2014	1.36 m	Nogueira et al., 2008	DNR	Edwards and Grubb, 1977; Martinez-Yrizar et al., 1992; Fromard et al., 1998; Araújo et al., 1999; Nelson et al., 1999; Ketterings et al., 2001 Mackensen et al., 2000; Cairns et al., 2003; Brandeis et al., 2006; Burger and Delitti, 2008; Nogueira et al., 2008; Kenzo et al., 2009; Djomo et al., 2010 Niiyama et al., 2010; Ebuy et al., 2011; Ryan et al., 2011; Vieilledent et al., 2012; Colgan et al., 2013; Mugasha et al., 2013; Ngomanda et al., 2014
-	-	1.37 m	Brandeis et al., 2006	-	-
-	-	DNR	Edwards and Grubb, 1977; Mackensen et al., 2000; Ketterings et al., 2001 Cairns et al., 2003; Burger and Delitti, 2008; Djomo et al., 2010; Niiyama et al., 2010; Ebuy et al., 2011; Ngomanda et al., 2014	-	-

Reported are the measurement device, point of measurement (the height above-ground at which D is measured on the stem) and the approach for treating buttresses. DNR - did not report.

or sometimes ρ_b was not a variable under consideration, but subsequently added to these data during compilation using global databases (Edwards and Grubb, 1977; Yamakura et al., 1986; Fromard et al., 1998; Mackensen et al., 2000; Cairns et al., 2003; Burger and Delitti, 2008; Kenzo et al., 2009; Niiyama et al., 2010; Ryan et al., 2011). For those studies that did measure, the most common approach was to determine the wet volume from the subsamples (Saldarriaga et al., 1988; Brown et al., 1995; Brandeis et al., 2006; Henry et al., 2010; Vieilledent et al., 2012; Alvarez et al., 2012; Nogueira et al., 2008), although there were variations on this: e.g., only a single subsample from the stem was considered (Nelson et al., 1999), or only subsamples from the stem (Goodman et al., 2014). Other approaches involved taking cores from each tree (Djomo et al., 2010) and combining measurements with literature values (Ketterings et al., 2001).

For the measurement of wet volume, the subsamples were usually measured via Archimedes' principle (Goodman et al., 2014), but sometimes graduated cylinders (Colgan et al., 2013), estimates from geometry (Henry et al., 2010), or a combination (Brown et al., 1995). Similar to the application of the dry-to-wet ratio, ρ_b was sometimes derived from the mean across subsamples, or othertimes weighted by pool.

In summary then, measurement protocol between studies were inconsistent for each of the 4 measured variables. This is of course a largely unavoidable inevitability, given the nature of these data compiled from multiple independent studies and operators, across both a large spatial extent and time-span.

3.2. The Ordinary Least Squares Models
3.2.1. Bivariate Models
The relative strength of the correlation between the predictor variables D and H with AGB is demonstrated by the two bivariate models, with the standard error of the regression from the AGB $= f(D)$ model considerably smaller than the

	Measurement of H	
	Approach	**Measurement device**
Pre-felling	Saldarriaga et al., 1988 Djomo et al., 2010; Niiyama et al., 2010; Colgan et al., 2013; Mugasha et al., 2013	Measuring poles Sine or tangent method
Post-felling	Yamakura et al., 1986; Martinez-Yrizar et al., 1992; Brown et al., 1995; Fromard et al., 1998; Nelson et al., 1999 Mackensen et al., 2000; Ketterings et al., 2001; Cairns et al., 2003; Brandeis et al., 2006; Burger and Delitti, 2008; Nogueira et al., 2008 Henry et al., 2010; Ebuy et al., 2011; Alvarez et al., 2012; Vieilledent et al., 2012; Goodman et al., 2014; Ngomanda et al., 2014	Tape measure
DNR	Edwards and Grubb, 1977; Araújo et al., 1999; Kenzo et al., 2009; Ryan et al., 2011	-

DNR - did not report.

AGB $= f(H)$ model (**Figure 3**). Residuals from both models are heteroscedastic (**Figure 3** and **Table 3**) and non–normally distributed (**Figure 3**). It should not be expected then, that $\hat{\sigma}$ has unbiased properties (i.e., $\mathrm{E}[\hat{\sigma}] \neq \sigma$).

In the case of the AGB $= f(D)$ model, residual variance decreases with increasing predicted AGB, and a combination of light-heavy tails are observed in the distribution of studentised residuals. Multiple outliers are seen, which likely exert undesirable leverage on $\hat{\beta}$, suggesting robust regression techniques might be more appropriate. The AGB $= f(H)$ model has clear deficiencies: AGB will be underestimated for both short and tall trees.

3.2.2. Multivariate Models

Including additional predictor variables leads to a significant reduction in the standard error of the regression relative to the bivariate models (**Figure 4**). However, similar to the bivariate $AGB = f(D)$ model, residuals from the three multivariate models are heteroscedastic (**Figure 4** and **Table 4**) and non–normally distributed (**Figure 4**). Across the multivariate models, residual variance consistently decreases as predicted AGB increases. The distributions of studentised residuals exhibit various combinations of heavy/light tails and bowing.

3.3. Cross-Validation

The calibration data were folded 10 times, resulting in ~400 observations per fold. This might be similar to the stem count encountered in a 1 ha tropical forest stand, so the uncertainty and bias metrics reported here might provide something of an expectation for those at the out-of-sample stand-scale.

Prediction accuracy increased with increasing predictor variable count (**Figure 5** and **Table 5**). Uncertainty varied from a minimum of 24 % for the AGB $= f(D, H, \rho_b)$ model, to a maximum of 137 % for the AGB $= f(H)$ model. When compared with the bivariate AGB $= f(D)$ model, the AGB $= f(D, H)$ model provided 10.0 % less uncertain predictions, whereas

only a 4.6 % reduction in uncertainty was observed for the AGB $= f(D, \rho_b)$ model.

Predictions from all 5 models were persistently biased upward (**Figure 5** and **Table 5**). A small reduction in bias was observed when the multivariate models are compared with the AGB $= f(D)$ model. Overall, the minimum observed mean bias was 6 %.

3.4. Inconsistent Measurement Between In- and Out-of-Sample Data

Inconsistent measurement error in predictor variables between in- and out-of-sample data was simulated by adding further error, drawn from normal distributions, to the calibration data. Simulated error added to H had standard deviations, σ_η, of 0.25, 0.5, 1, and 2 m. Simulated error added to ρ_b had σ_η of 2, 50, 75, and 100 kg m^{-3}.

Large fluctuations are observed in the parameters of the models constructed from these various combinations of added noise (**Table 6**), which regularly fall outside the 95 % confidence intervals of the base model presented in **Figure 4**. Additional measurement error in a predictor variable manifests in a downward force on its corresponding parameter (regression dilution), and a variable upward force exerted on the remaining parameters, with a particularly pronounced effect on the intercept, β_0. That is, as measurement error inconsistency increases, the less influence that particular predictor variable has on predicted AGB.

3.5. The Effect of Tree Size on Model Parameters

The AGB $= f(D, H, \rho_b)$ model was constructed from the various considered subsets of the calibration data (these subsets contained sequentially fewer trees, removing those below diameter thresholds of $D \geq 0.1, 0.25, 0.5$ m, and 0.75 and 1 m). There is a tendency for the parameters associated with the predictor variables to increase as fewer smaller trees are considered, whilst the intercept parameter decreases (**Figure 6**). Whilst the changes in these parameters are substantial, it is noted that rarely do confidence intervals not overlap. The confidence intervals themselves rapidly inflate because of the relatively few observations in the larger size-classes.

4. DISCUSSION

The residuals of each bivariate and multivariate model were heteroscedastic and non–normally distributed. The cross-validation results found the minimum relative uncertainty in fold-scale AGB predictions achieved by these various models was ~24 %, and that predictions were also persistently upward biased by a minimum of 6 % (~400 observations per fold). Our analysis suggests that these results are likely symptoms of model misspecification. That is, the models do not account for everything they should.

4.1. Inconsistent Measurement Error

It was noted in the methods section that error in the measurement of predictor variables will not necessarily affect the trueness of AGB predictions. For example, if in-sample
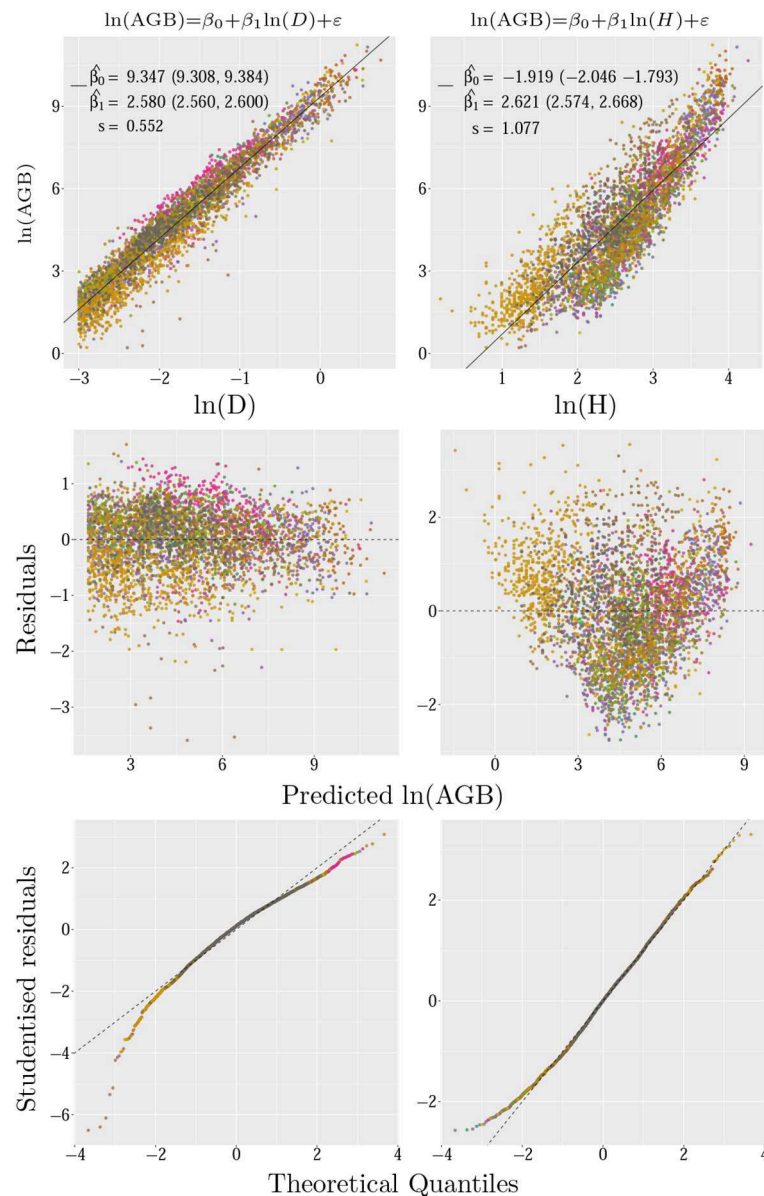
FIGURE 3 | The two bivariate pan-tropical models. Left: AGB = $f(D)$, right: AGB = $f(H)$. The upper subfigures present both models overlain on the calibration data (which are uniquely colored by each underlying study contributing to the full dataset). Provided in the upper subfigures are the OLS estimates of the population parameters, the bootstrapped BCa 95 % confidence intervals ($N = 10\,000$), and the standard error of the regression. The middle subfigures illustrate the variance of residuals across predicted B for both models, and the lower subfigures present a quantile-quantile plot comparing the distribution of studentised residuals with the expected normal distribution. It is noted that residuals from both models have non–constant variance and are non–normally distributed.

tree height, H, were measured with error, η, $H' = H + \eta$ [$\eta \sim \mathcal{N}(0, \sigma_\eta^2)$], then the subsequently constructed model characterizes the relationship AGB $= f(H')$. That is, the imprecise expectation of H is baked-in to the OLS estimate of the population parameters. Provided the out-of-sample measurement of H shares this expectation, predicting AGB using these parameters is unproblematic (Jonsson, 1994). However, if the out-of-sample measurement has a different expectation of error, then a systematic error will be introduced.

The key point then, is not the presence of measurement error itself, but the difference in its distribution between in- and out-of-sample measurements. As discussed below, we think that assuming these distributions are approximately consistent for the predictor variables H and ρ_b is unjustifiable. Crucially, if it is assumed this difference is negligible (which is the current position of all widely-used pan-tropical allometric models), when it is not, a bias of unknown direction and magnitude will be present in AGB predictions.

TABLE 3 | Results from the statistical tests applied to the residuals of both bivariate methods to assess variance.

Test	Null hypothesis	AGB = f(D)		AGB = f(H)	
		Statistic	P-Value	Statistic	P-Value
Breusch-Pagan	Var(e) = constant	13.3	< 0.001	9.2	< 0.001
White	Var(e) = constant	18.0	< 0.001	9.2	0.01

The null hypothesis of residuals having constant variance (homoscedasticity) is rejected in both models, and the alternative hypothesis of heteroscedasticity is accepted.

4.1.1. Differences Between In- and Out-of-Sample Measurement Error

In the metadata review it was noted that for the majority of in-sample data, measurement of H was made via tape measure post-felling. It would seem plausible to assume this method provides true and precise measurements, e.g., it would not be unreasonable to speculate η could take a form similar to $\eta \sim \mathcal{N}(0.0\,m, 0.5\,m)$.

However, out-of-sample measurements of H are made with the tree *in situ* using clinometers and range finders via either the tangent or sine method. Two prominent studies have explored the accuracy of these measurements in tropical forests. Larjavaara and Muller-Landau (2013) found η to take the average forms $\eta \sim \mathcal{N}(-0.8\,m, 6.8\,m)$ and $\eta \sim \mathcal{N}(-4.5\,m, 2.3\,m)$ for the tangent and sine methods respectively. Hunter et al. (2013) found η to take the average form $\eta \sim \mathcal{N}(-1.1\,m, 4.7\,m)$ for the tangent method.

This would suggest measurement error distributions between in- and out-of-sample measurement of H are significantly different. More problematic, out-of-sample H is often not measured, but replaced with predictions from models (Feldpausch et al., 2011; Sullivan et al., 2018). It is likely that these models share issues similar to those encountered here (e.g., a heteroscedastic error term means $\mathrm{E}[\hat{\sigma}] \neq \sigma$), such that the out-of-sample error structure becomes misleading.

Unlike the measurement of H, few studies have explored measurement error in ρ_b, but it would seem reasonable to suggest that obtaining a robust description of the in-sample measurement error distribution is impossible. The metadata review showed the in-sample methods include a variety of direct measurements on subsamples and cores, and acquiring values from global databases. Therefore, the mean in-sample definition of the measurement of ρ_b itself is an unknown. That is, the aggregate in-sample measurement of ρ_b, which is the expectation of the out-of-sample measurement, is some unmeasurable and unknown composite of these various methods. If the definition of the in-sample measurement is unknown, then the difference in measurement error between in- and out-of-sample measurements is unknown.

With respect to the measurement of D, it was assumed in the simulations that measurement errors were consistent. This is possibly justified as widely-used field guides for tropical forest inventorying are consistent and unambiguous in the definition of the measurement (Marthews et al., 2014). We do acknowledge however, there are reasons why in- and out-of-sample measurement error distributions might be inconsistent. For example, the metadata review identified the use of different measurement devices (e.g., diameter tape and calipers), point of

measurement and buttress treatment. Similar to the in-sample measurement of ρ_b, this would lead to the mean in-sample definition of the measurement of D being some fusion of these approaches, which cannot be mirrored by a single out-of-sample measurement. There are also possibly human factors at play: the skill and diligence of operators may vary between separate data acquisitions.

4.1.2. Implications of Inconsistent Measurement Error

The question remains then: what are the likely consequences to predictions of tree- and stand-scale AGB from inconsistent error in the measurement of in- and out-of-sample predictor variables? Given the above discussion, we think our worst-case simulations presented in **Table 6** provide a particularly conservative insight.

We assumed out-of-sample measurements of H and ρ_b were only more imprecise than in-sample measurements (i.e., measurement trueness remained consistent). We assumed these differences were characterized by normally distributed error with $2\,m$ and $100\,kg\,m^{-3}$ standard deviation respectively. Under these assumptions, our simulations of the AGB $= f(D, H, \rho_b)$ model found the parameter β_0 to change from 0.821 to 4.009, β_1 from 2.019 to 2.206, β_2 from 0.888 to 0.566, and β_3 from 0.821 to 0.508. Both absolutely and relatively, these changes in population parameters have implications to predictions of AGB.

Absolutely, these differences can be demonstrated by predicting AGB for two hypothetical trees: first a tree with $D = 0.1\,m$, $H = 20\,m$ and $\rho_b = 600\,kg\,m^{-3}$ has predicted AGB of 63.2 kg in the original model, and 52.7 kg in the simulated model, a $-16.6\,\%$ change. Second, a larger tree with $D = 1.5\,m$, $H = 40\,m$ and $\rho_b = 500\,kg\,m^{-3}$ has predicted AGB of 23,857.6 kg and 27,976.5 kg respectively, a 17.3 % change. There might therefore be some degree of cancellation when up-scaling to the stand, but this would be both a function of structural composition, and dangerous to assume.

Relatively, there are two scenarios where not accounting for inconsistent measurement error would lead to potentially spurious predictions of AGB change: inter-plot comparison and change detection. To illustrate this, we downloaded some field data from https://forestplots.net/ for 2 plots included in the Global Ecosystem Monitoring network (GEM, http://gem.tropicalforests.ox.ac.uk). These two 1 ha plots (designation: MNG-03 and MNG-04) are in close proximity to one another in l'Arboretum Raponda Walker, Estuaire, Gabon (location: $0.576°$, $9.323°$ and $0.576°$, $9.328°$). Both plots are moist, lowland, Terra Firme, secondary forests; MNG-03 has a monodominant composition whilst MNG-04 is mixed. MNG-03 has a stem count, basal area, Lorey's height and basal-area-weighted basic density of 436, $47.6\,m\,ha^{-2}$, 39.1 m and $489\,kg\,m^{-3}$ respectively; MNG-04 has 437, $34.8\,m\,ha^{-2}$, 30.8 m and $605\,kg\,m^{-3}$ respectively.

First, with respect to inter-plot comparison then, the original model predicts stand-scale AGB of 579,591 kg and 421,141 kg for MNG-03 and MNG-04 respectively, whilst the simulated model predicts 588,370 kg and 407,950 kg respectively. That is, the original model predicts a 31.7 % difference in AGB between plots, whereas the simulated models predicts 36.2 % difference.
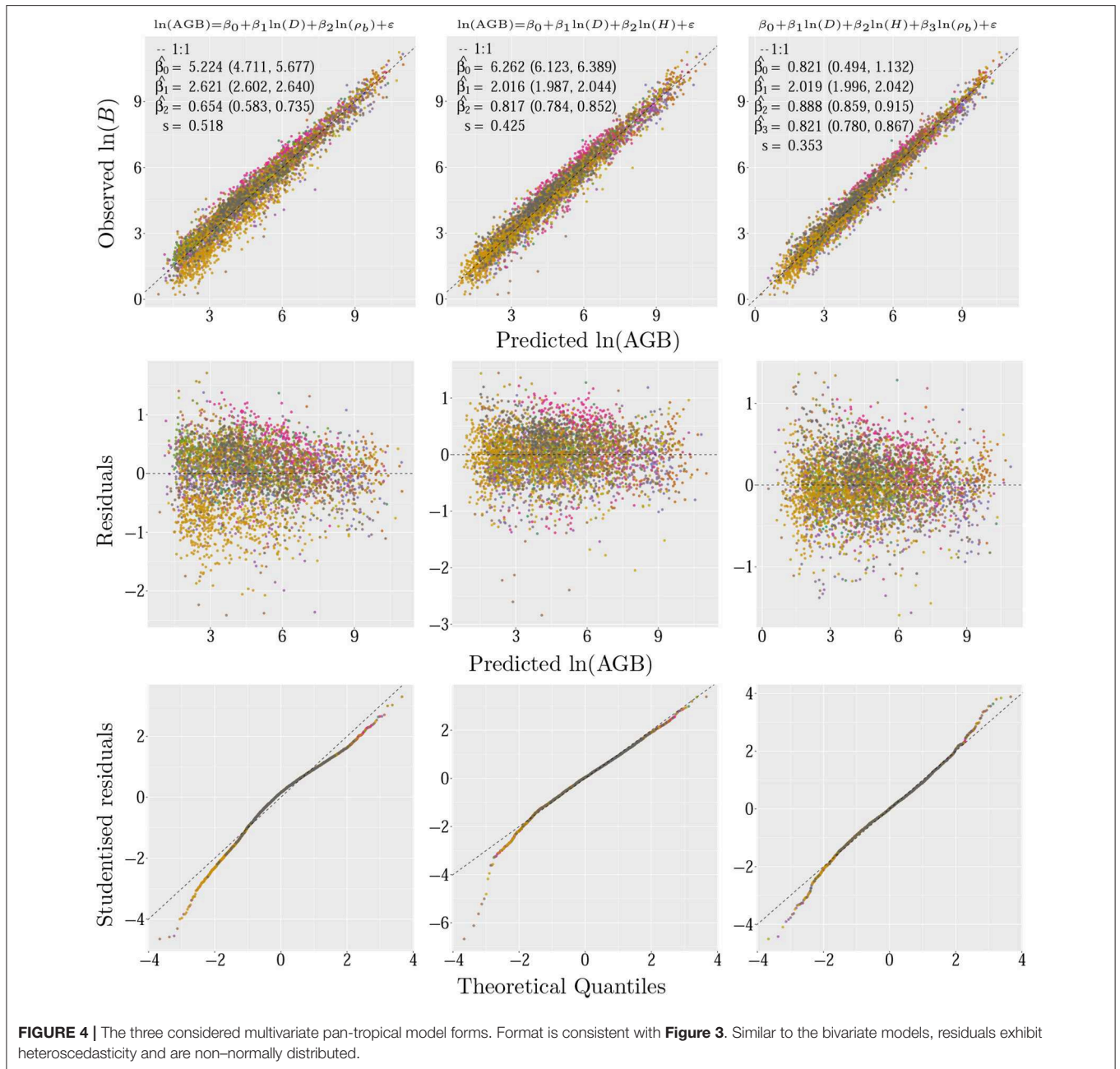
FIGURE 4 | The three considered multivariate pan-tropical model forms. Format is consistent with **Figure 3**. Similar to the bivariate models, residuals exhibit heteroscedasticity and are non–normally distributed.

TABLE 4 | Results from the statistical tests applied to the three multivariate models.

Test	Null hypothesis	AGB = f(D,ρ_b)		AGB = f(D,H)		AGB = f(D,H,ρ_b)	
		Statistic	P-Value	Statistic	P-Value	Statistic	P-Value
Breusch-Pagan	Var(e) = constant	61.6	< 0.001	51.3	< 0.001	74.6	< 0.001
White	Var(e) = constant	359.6	< 0.001	53.8	< 0.001	81.9	< 0.001

Format is consistent with Table 3. These results corroborate with those from Figure 4 in indicating the residuals of each model are heteroscedastic.

Second, to explore the implications to change detection, we hypothetically assume some changes in the composition of MNG-04 since these data were collected. We assume a uniform increase in D, H, and ρ_b of 0.01 m, 2.5 m and 25 kg m^{-3} respectively per tree. The original and simulated models now predict stand-scale AGB as 489,633 kg and 458,380 kg
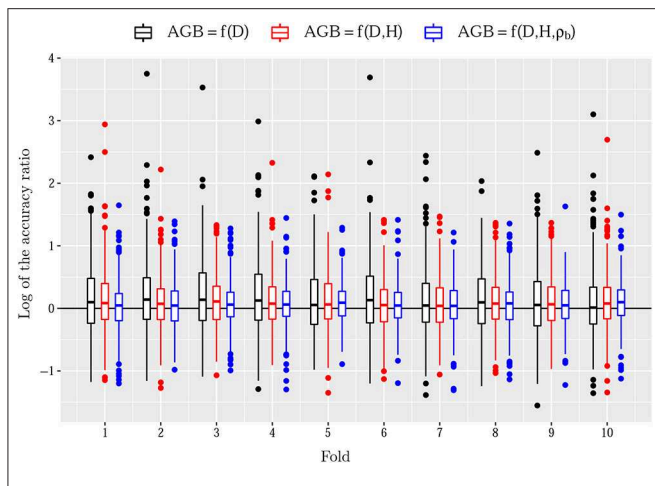
FIGURE 5 | Stratified 10-fold cross-validation results for three of the pan-tropical models. For each considered model, per validation fold, the distribution of the log of the accuracy ratio is shown (ln(A$\hat{G}$B/AGB)). Each fold contains ∼400 observations. Distributions are represented via standard format box-and-whisker. It is observed that the variance of these distributions tends to reduce as additional predictor variables are added. The median value of these distributions is consistently greater than zero, signifying predictions of AGB are generally larger than observed AGB.

TABLE 5 | Stratified 10-fold cross-validation prediction metrics for the five models.

Model	MSA (uncertainty, %)			SSPB (bias, %)		
	Mean	**Min**	**Max**	**Mean**	**Min**	**Max**
AGB = $f(D)$	40.8	36.4	46.3	9.7	0.6	15.4
AGB = $f(H)$	137.1	121.5	157.8	77.1	66.6	92.7
AGB = $f(D, H)$	30.8	27.2	35.9	7.1	3.5	10.8
AGB = $f(D, \rho_b)$	36.2	32.5	41.5	5.8	1.4	12.0
AGB = $f(D, H, \rho_b)$	24.4	22.8	27.6	6.3	0.7	11.7

Uncertainty and bias of AGB predictions from these models is quantified using the median symmetric accuracy (MSA) and the signed symmetric percent bias (SSPB) respectively. These metrics were generated per fold (∼400 observations per fold), and here the mean, minimum and maximum of the 10 values are reported for each model. It is noted minimum observed mean uncertainty is ∼24 %, and that accuracy generally improves as predictor variables are added to the model. The sign of SSPB is persistently positive, again signifying predictions of AGB are generally larger than observed AGB.
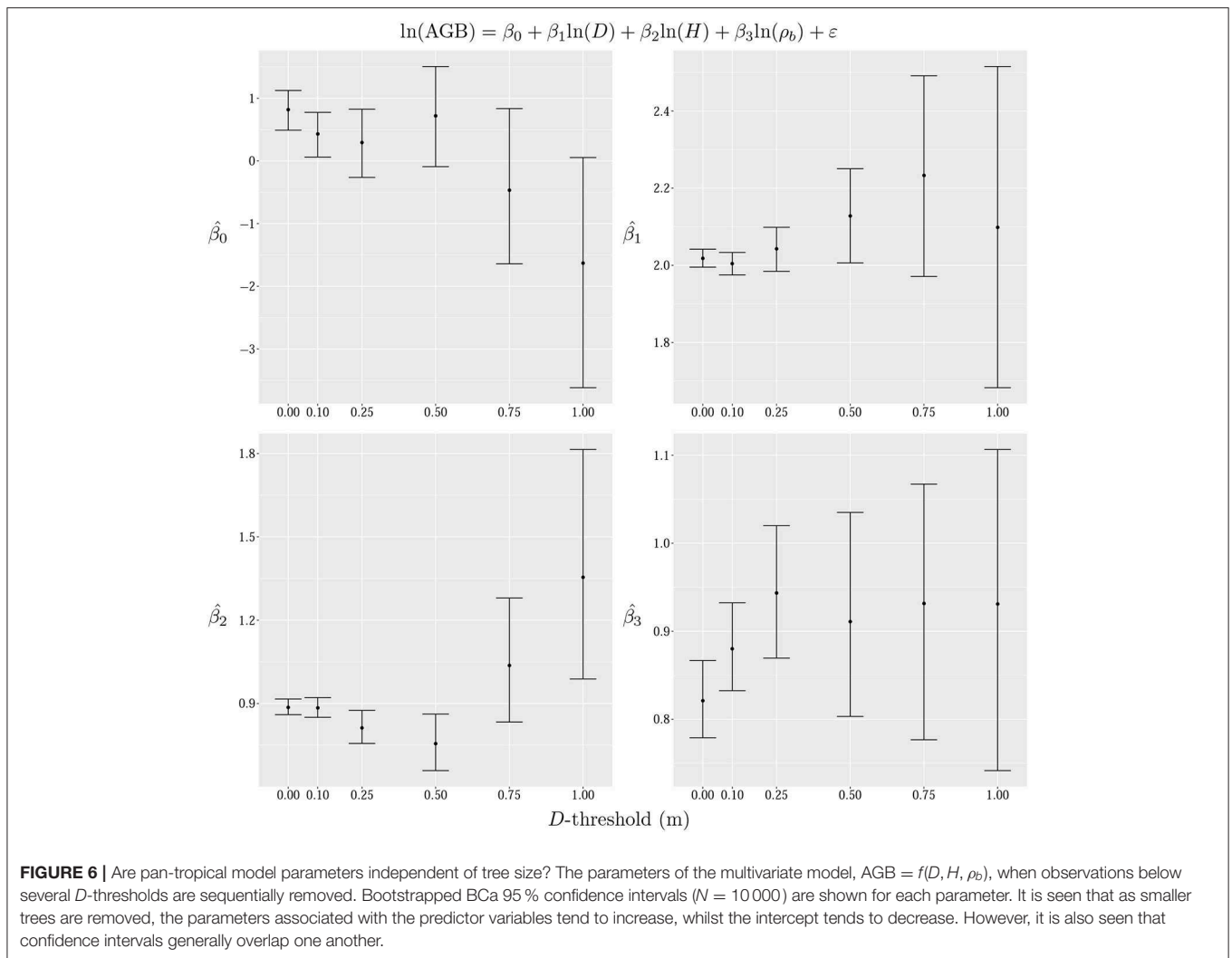
TABLE 6 | Simulating inconsistent measurement error in predictor variables between in- and out-of-sample data.

		$\ln(\text{AGB}) = \beta_0 + \beta_1\ln(D) + \beta_2\ln(H') + \beta_3\ln(\rho'_b) + \varepsilon$				
		σ_η ($H' = H + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma_\eta^2)$) (m)				
		0.00	**0.25**	**0.50**	**1.00**	**2.00**
	0	$\hat{\beta}_0 = 0.821$	0.850	0.941	1.314	2.261
		$\hat{\beta}_1 = 2.019$	2.023	2.035	2.085	2.210
		$\hat{\beta}_2 = 0.888$	0.881	0.861	0.780	0.582
		$\hat{\beta}_3 = 0.821$	0.820	0.818	0.806	0.774
		$s = 0.353$	0.354	0.358	0.372	0.409
	25	$\hat{\beta}_0 = 0.979$	1.008	1.098	1.469	2.405
		$\hat{\beta}_1 = 2.019$	2.023	2.035	2.085	2.209
		$\hat{\beta}_2 = 0.885$	0.879	0.859	0.778	0.581
		$\hat{\beta}_3 = 0.798$	0.797	0.794	0.783	0.752
		$s = 0.355$	0.356	0.360	0.374	0.410
σ_η ($\rho'_b = \rho_b + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma_\eta^2)$) (kg m^{-3})	50	$\hat{\beta}_0 = 1.427$	1.454	1.543	1.905	2.816
		$\hat{\beta}_1 = 2.019$	2.023	2.035	2.085	2.209
		$\hat{\beta}_2 = 0.879$	0.873	0.853	0.772	0.577
		$\hat{\beta}_3 = 0.731$	0.730	0.727	0.718	0.689
		$s = 0.361$	0.362	0.365	0.379	0.414
	75	$\hat{\beta}_0 = 2.044$	2.070	2.155	2.505	3.385
		$\hat{\beta}_1 = 2.020$	2.023	2.035	2.084	2.207
		$\hat{\beta}_2 = 0.871$	0.865	0.845	0.765	0.572
		$\hat{\beta}_3 = 0.638$	0.638	0.636	0.627	0.603
		$s = 0.368$	0.369	0.372	0.386	0.420
	100	$\hat{\beta}_0 = 2.708$	2.731	2.817	3.157	4.009
		$\hat{\beta}_1 = 2.020$	2.024	2.035	2.084	2.206
		$\hat{\beta}_2 = 0.862$	0.856	0.837	0.758	0.566
		$\hat{\beta}_3 = 0.539$	0.539	0.536	0.529	0.508
		$s = 0.376$	0.377	0.380	0.393	0.426

Here, the multivariate model, AGB = $f(D, H, \rho_b)$, is considered, whereby additional random error has been added to the in-sample calibration data. The additional noise is included in predictor variables H and ρ_b, which is drawn from normal distributions with increasing variance. Per distribution, 10,000 draws were made, and the mean estimate of the parameters is reported in the table. These parameters represent those required for unbiased prediction of AGB when out-of-sample data are measured with σ_η-more measurement error than the in-sample data.

respectively. That is, the original model predicts a 16.3 % increase in AGB, whilst the simulated models predicts a 12.4 % increase.

4.1.3. Including Tree Height and Wood Density in Pan-Tropical Allometry

Considering these implications, and given our assertion that these simulations of inconsistent measurement error were conservative, we think careful thought is required on how best to include H and ρ_b as predictor variables in pan-tropical allometric models. Across the literature there is a consensus that their inclusion is worthwhile: multivariate models including these variables generally exhibit a smaller standard error of the regression than bivariate D-only counterparts; H and ρ_b are therefore correlated with AGB, whilst not perfectly correlated with D. Given that in tropical forests, H and ρ_b are often observed to vary for a fixed value of D, it is therefore the expectation that their inclusion as predictors will improve tree- and stand-scale prediction accuracy. Furthermore, it has been demonstrated that at the landscape- and regional-scales, ρ_b varies systematically as a response to multiple environmental factors (Baker et al., 2004; Phillips et al., 2019). If ρ_b were excluded from pan-tropical allometry, these systematic variations would go undetected in up-scaled predictions of AGB (Mitchard et al., 2014).

These benefits of including H and ρ_b were reflected in the cross-validation results, whereby the AGB = $f(D, H)$ model yielded 10.0 % less uncertain predictions than the bivariate AGB = $f(D)$ model, and the AGB = $f(D, H, \rho_b)$ improved on this by a further 6.4 %. However, these results do not account for systematic error introduced by inconsistent errors-in-variables (e.g., in the majority of these calibration data, H was measured post-felling with a tape measure). Therefore, the decision

$$\ln(\text{AGB}) = \beta_0 + \beta_1 \ln(D) + \beta_2 \ln(H) + \beta_3 \ln(\rho_b) + \varepsilon$$

FIGURE 6 | Are pan-tropical model parameters independent of tree size? The parameters of the multivariate model, AGB = $f(D, H, \rho_b)$, when observations below several D-thresholds are sequentially removed. Bootstrapped BCa 95 % confidence intervals ($N = 10\,000$) are shown for each parameter. It is seen that as smaller trees are removed, the parameters associated with the predictor variables tend to increase, whilst the intercept tends to decrease. However, it is also seen that confidence intervals generally overlap one another.

to include these variables as predictors is balanced between reducing random error by known amount, and introducing an unknown amount of systematic error whilst inconsistent errors-in-variables remain unaccounted for.

Given the above discussion makes the case that it would be unjustifiable to assume in- and out-of-sample measurement error distributions in H and ρ_b are consistent, this would imply unknown bias is always present in AGB predictions from these conventional multivariate models. We therefore think formal steps are necessary to account for, and minimize, this bias. This action can take two forms: first, the in- and out-of-sample measurement methods become consistent, such that it is assumed the respective measurement error distributions are consistent, or second, inconsistencies are corrected for during modeling.

In the particular case of measuring H, in the above referenced studies of *in situ* error distributions, it was noted both the tangent and sine method were relatively inaccurate. Significantly, it was also seen that between the two independent studies, the resulting error distributions for the tangent method were different. This might suggest these distributions are not consistent across forest type and/or operator. It would follow then, that the in-sample measurement of H should be made from the more true and

precise measurements obtained from a tape measure post-felling. This implies that the in- and out-of-sample measurement methods will be different, and some form of modeling correction is required.

To minimize systematic error introduced by the inclusion of H in pan-tropical models then, we think the following three steps are necessary: (1) The in-sample data are measured post-felling via tape measure, where measurement error is quantified through repeated measurements, ideally by multiple operators. If calibration data are compiled from multiple individual studies, then those data where H has been measured using other methods must be excluded (e.g., *in situ* pre-harvest). (2) Out-of-sample H is measured *in situ* using the tangent or sine method, whereby measurement error is concurrently quantified, or estimated via known distributions. (3) The OLS estimators account for the inconsistencies between these two error distributions using either errors-in-variables modeling (Jonsson, 1994), or simulation approaches similar to those used here.

It would seem that the appropriate approach for including ρ_b in pan-tropical models whilst minimizing systematic error is a more open question. Firstly, the definition of the measurement of ρ_b requires standardization. Because these measurements

are currently not standardized, both across and between in- and out-of-sample data, robust quantitative descriptions of measurement errors are unavailable, meaning reliably correcting for inconsistencies and resulting bias is impossible. One approach might be to replace all measurements with values from global databases (Chave et al., 2009), but this requires careful consideration: (i) the measurement methods used to collect the underlying data are themselves likely inconsistent and (ii) errors become autocorrelated.

4.2. Is Pan-Tropical Allometry Independent of Tree Size?

An interesting question when considering systematic error in allometric-derived AGB predictions is whether model parameters are independent of tree size. That is, are the population parameters necessary for predicting the AGB of a small tree, the same as those necessary for a large tree? This question has previously been posed by others including Picard et al. (2015b), who found, using calibration data from central Africa, that bivariate power law models did not hold across all size-classes, and that some size dependency existed.

Within these specific calibration data considered here, Ploton et al. (2016) noted a break point, whereby models constructed from calibration data below and above ∼20,000 kg did not share the same population parameters. This was similarly observed in **Figure 6**: when trees belonging to specific D-classes were sequentially removed, the parameters of the AGB $= f(D, H, \rho_b)$ model changed substantially. But are these changes significant, and if so, is this a detection of size dependency?

As to the first question, it would appear these changes were not statistically significant because the bootstrapped 95 % confidence intervals for each parameter generally overlapped. Whilst the confidence intervals are compact for parameters describing the complete dataset ($n = 4004$), they quickly expand as the smaller trees are removed. This is inevitable given the non–uniform distribution of these data, where $n = 215$ and 90 for observations with $D \geq 0.75$ m and 1.0 m respectively. So the observed changes in the population parameters were not significant within these particular data, but this does not rule out the existence of a size dependency in the population.

4.2.1. Potentially Biased Measurement of Calibration Data

Even if confidence intervals were not to overlap, attributing changes (or indeed the lack of change) to a size dependency is challenging when the models are potentially misspecified. A further potential misspecification, aside from inconsistent measurement error, is that the unconditional mean error in observation of AGB is possibly non–zero ($\mathrm{E}(\varepsilon) \neq 0$).

The metadata review identified that for the measurement of wet mass (i.e., via weighing), the destructive methods introduce several sources of loss. For example, most studies did not account or correct for losses from chainsaw cuts, or water losses accrued between felling and measurement. Several studies also excluded stump material from measurement.

As shown in the methods sections, if a bias, c, were consistent across observations, then only the intercept parameter is biased, $\mathrm{E}(\hat{\beta}_0) = \beta_0 + c$. However, if bias in the observation of

AGB is correlated with tree size, the effects are more complex, and contaminate all parameters. It would again not seem unreasonable to speculate that if bias is present, that this second form is the more likely.

For example, we recently harvested 4 tropical trees in Brazil; we measured the wet mass of the stem by cutting it into manageable sections that were possible to weigh. We also estimated the losses from these cuts by estimating cut volume. Across these 4 trees, the wet masses of these four stems were 3,229, 3,636, 5,097, and 16,780 kg. The cumulative volume-derived wet mass of losses from chainsaw cuts were 28, 41, 58, and 330 kg, respectively. These losses represent ∼0.9, 1.1, 1.1, and 2.0 %, respectively. For these particular trees and measurements methods then, these losses are correlated with tree size.

Returning then to the original question, we are not trying here to suggest that observations of AGB are necessarily biased; rather that the possibility exists that AGB are biased, and it is also possible bias is correlated with tree size. In order to attribute statistically significant changes in population parameters to a dependency on tree size, it would need to be demonstrated that bias in observations of AGB is negligible.

It is also noted that the wet mass for a large section of the calibration data was not measured, but instead estimated from volume measurements (indeed the measurement method itself would appear correlated with tree size: volume-derived estimates were often used when it was logistically impracticable to weigh). Expectations of systematic error in these two measurement methods may therefore be inconsistent. Random error would likely also share a disparate expectation, which may offer a partial explanation as to why model residuals were heteroscedastic.

4.2.2. Additional Calibration Data Are Required

Answering the question of whether pan-tropical allometric models are independent of tree size would be of general scientific interest, but more specifically, it is critical to understanding the trueness of AGB predictions. Currently, the above-ground biomass of large trees is predicted from empirical relationships discovered from imbalanced calibration data (e.g., in these considered data the median value of AGB is 98 kg).

In OLS, each observation similarly influences the population parameter estimates when any leverage effects from outliers and influential points are ignored. That is, because of this imbalance, large trees currently have little influence on model parameters. If the allometric relationship is independent of tree size (implicitly, this is the assumption of current widely-used pan-tropical allometric models), then this is of little concern, but likewise, if the relationship is size dependant, predictions of AGB for the larger size-classes are biased.

To answer these questions requires the collection of more calibration data. Specifically, these new data need to be gathered from larger trees. If these data are to supplement existing data, it is more beneficial to acquire a small number of observations from larger trees, than a large number of observations from smaller trees. Indeed, adding further small trees to these calibration data will only further reduce the influence of larger trees on the OLS estimators. Additional data from larger trees will also reduce the size of confidence intervals in model parameters constructed solely from the larger trees.

Of course, in the wider context of considering whether out-of-sample data are adequately represented by in-sample data, size is only one contributing factor. Another key consideration is the geographical representation of the sample, given that allometries are geographically variable (Henry et al., 2013). These additional large trees then, would ideally be uniformly collected from across the tropics (Banin et al., 2012; Gorgens et al., 2019; Shenkin et al., 2019).

As an aside to the comment that each observation will similarly influence the OLS estimate of the population parameters, it would therefore not be sufficient to argue that a particular allometric model is suitable for predicting the AGB of a particular type of tree (e.g., a large tree or from specific geography/species), just because observations from that type are present in the calibration data, if those data are overwhelmed by observations from other types.

The form of the OLS model informs where to focus efforts in quantifying measurement error in these new data. That is, a random error term is included for observations of AGB, so the vital characteristic in the measurement of AGB is trueness, with precision a secondary concern. Whereas for observation of the predictor variables, no error term is present, meaning both characteristics of the measurement are of equal importance.

A caveat to this comment on precision in AGB, given the previous discussion on heteroscedasticity, is that we have not considered in this paper the implication of a heteroscedastic error term to predictions of AGB. It was noted in the methods section that most widely-used pan-tropical models employ a correction factor that includes $\hat{\sigma}$ when re-transforming predictions from log- to real-space. However, a heteroscedastic error term means $E(\hat{\sigma}) \neq \sigma$, which presumably biases AGB predictions.

4.3. A Note on Causality

Finally, we conclude with a comment on causality. Throughout the paper we have been careful to distinguish between prediction and explanation. In the methods sections we acknowledged that the models constructed here are endogenous: the assumption of strict exogeneity was violated by omitted variable bias.

In the introduction section it was noted the causes of above-ground biomass are lifetime cumulative gross primary production, respiration and loss. Omitting these causal variables has a fundamental implication: it would be spurious to infer from these models that D, H and ρ_b cause AGB. That is, if the D of a particular tree has changed over time, the AGB $= f(D)$ model predicts a change in AGB proportional to $D^{2.580}$, but it does not explain it.

This distinction means care must be taken with causal interpretations of allometric-derived AGB predictions. Examples of spurious causal claims might be inter-plot comparisons, where the differences in structural composition between two stands [i.e., $\sum(D, H, \rho_b)_A - \sum(D, H, \rho_b)_B$] is proposed as the explanation for their difference in predicted stand-scale AGB; or intra-plot change detection studies, where growth/death/recruitment between surveys [i.e., $\Delta \sum(D, H, \rho_b)_A$] is proposed as the explanation for change in predicted stand-scale AGB.

The models used in this paper then, are only for the purpose of prediction. For that reason, we are comfortable with the various

multivariate forms considered here that might stand accused of being a form of data dredging (Sileshi, 2014). Given that these models have no theoretical grounding, and provided they will only be used for prediction, we see no obvious reason such forms, or even more exotic forms, should not be considered, provided that the precision, trueness and accuracy of their AGB predictions are well-understood.

In conclusion, we constructed various conventional bivariate and multivariate models for predicting above-ground biomass from open access pan-tropical calibration data. We found the residuals of each model were heteroscedastic and non-normally distributed. Stratified k-fold cross-validation found the minimum uncertainty in fold-scale predictions from these models to be 24 %, and that predictions were persistently biased upward by 6 % (~400 observations per fold). These results are likely symptoms of model misspecification: in particular, that the models do not account for inconsistent measurement error in predictor variables between in- and out-of-sample measurements. Through simulation, we showed how even a conservative degree of inconsistent measurement error can potentially lead to both absolute and relative bias in tree- and stand-scale AGB predictions. We presented the case that whilst including H and ρ_b as predictor variables in pan-tropical models alongside D increased prediction precision, their inclusion introduces a bias of unknown size and direction when inconsistent measurement error remain unaccounted for. We suggested several measurement and modeling approaches to formally compensate for this bias whilst retaining the predictive benefits of these variables. Finally, we asked the question of whether pan-tropical allometric model parameters are independent of tree size. Our analysis indicates that potential model misspecifications and imbalanced calibration data currently prevent finding a definitive answer. This can only be addressed with additional calibration data, specifically from larger trees.

DATA AVAILABILITY STATEMENT

The in-sample calibration data are hosted at http://chave.ups-tlse.fr/pantropical_allometry.htm. *treeallom* is available at https://github.com/apburt/treeallom. The version described in this paper, v0.1.0 is archived at https://doi.org/10.5281/zenodo.3603434.

AUTHOR CONTRIBUTIONS

All authors conceived and designed the methods. AB wrote the manuscript and software, with contributions from all authors.

FUNDING

REFERENCES

Alvarez, E., Duque, A., Saldarriaga, J., Cabrera, K., de las Salas, G., del Valle, I., et al. (2012). Tree above-ground biomass allometries for carbon stocks estimation in the natural forests of colombia. *Forest Ecol. Manage.* 267, 297–308. doi: 10.1016/j.foreco.2011.12.013

Angelsen, A., Brockhaus, M., Sunderlin, W. D., and Verchot, L. V. (2012). *Analysing REDD+*. Bogor: Center for International Forestry Research, CIFOR.

Araújo, T. M., Higuchi, N., and de Carvalho Júnior, J. A. (1999). Comparison of formulae for biomass content determination in a tropical rain forest site in the state of pará, brazil. *Forest Ecol. Manage.* 117, 43–52.

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., et al. (2016). An integrated pan-tropical biomass map using multiple reference datasets. *Global Change Biol.* 22, 1406–1420. doi: 10.1111/gcb.13139

Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., et al. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nat. Clim. Change* 2, 182–185. doi: 10.1038/nclimate1354

Baker, T. R., Phillips, O. L., Malhi, Y., Almeida, S., Arroyo, L., Di Fiore, A., et al. (2004). Variation in wood density determines spatial patterns inamazonian forest biomass. *Global Change Biol.* 10, 545–562. doi: 10.1111/j.1365-2486.2004.00751.x

Banin, L., Feldpausch, T. R., Phillips, O. L., Baker, T. R., Lloyd, J., Affum-Baffoe, K., et al. (2012). What controls tropical forest architecture? testing environmental, structural and floristic drivers. *Global Ecol. Biogeogr.* 21, 1179–1190. doi: 10.1111/j.1466-8238.2012.00778.x

Basuki, T. M., van Laake, P. E., Skidmore, A. K., and Hussin, Y. A. (2009). Allometric equations for estimating the above-ground biomass in tropical lowland dipterocarp forests. *Forest Ecol. Manage.* 257, 1684–1694. doi: 10.1016/j.foreco.2009.01.027

Brandeis, T. J., Delaney, M., Parresol, B. R., and Royer, L. (2006). Development of equations for predicting puerto rican subtropical dry forest biomass and volume. *Forest Ecol. Manage.* 233, 133–142. doi: 10.1016/j.foreco.2006.06.012

Breusch, T. S., and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47, 1287–1294.

Brown, I. F., Martinelli, L. A., Thomas, W., Moreira, M. Z., Ferreira, C. A. C., and Victoria, R. A. (1995). Uncertainty in the biomass of amazonian forests: an example from rondônia, brazil. *Forest Ecol. Manage.* 75, 175–189.

Brown, S. (1997). *Estimating Biomass and Biomass Change of Tropical Forests*. Rome: FAO - Food and Agriculture Organization of the United Nations.

Brown, S., Gillespie, A. J. R., and Lugo, A. E. (1989). Biomass estimation methods for tropical forests with applications to forest inventory data. *Forest Sci.* 35, 881–902.

Burger, D. M., and Delitti, W. B. C. (2008). Allometric models for estimating the phytomass of a secondary atlantic forest area of southeastern brazil. *Biota Neotropica* 8, 131–136. doi: 10.1590/S1676-06032008000400012

Cairns, M. A., Olmsted, I., Granados, J., and Argaez, J. (2003). Composition and aboveground tree biomass of a dry semi-evergreen forest on mexico's yucatan peninsula. *Forest Ecol. Manage.* 186, 125–132. doi: 10.1016/S0378-1127(03)00229-9

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., et al. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* 145, 87–99. doi: 10.1007/s00442-005-0100-x

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

Chave, J., Coomes, D., Jansen, S., Lewis, S. L., Swenson, N. G., and Zanne, A. E. (2009). Towards a worldwide wood economics spectrum. *Ecol. Lett.* 12, 351–366. doi: 10.1111/j.1461-0248.2009.01285.x

Chave, J., Davies, S. J., Phillips, O. L., Lewis, S. L., Sist, P., Schepaschenko, D., et al. (2019). Ground data are essential for biomass remote sensing missions. *Surveys Geophys.* 40,863–880. doi: 10.1007/s10712-019-09528-w

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., et al. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biol.* 20, 3177–3190. doi: 10.1111/gcb.12629

Clark, D. B., and Kellner, J. R. (2012). Tropical forest biomass estimation and the fallacy of misplaced concreteness. *J. Vegetat. Sci.* 23, 1191–1196. doi: 10.1111/j.1654-1103.2012.01471.x

Colgan, M. S., Asner, G. P., and 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. doi: 10.1890/12-0922.1

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S. J., Kubiszewski, I., et al. (2014). Changes in the global value of ecosystem services. *Global Environ. Change* 26, 152–158. doi: 10.1016/j.gloenvcha.2014.04.002

Djomo, A. N., Ibrahima, A., Saborowski, J., and Gravenhorst, G. (2010). Allometric equations for biomass estimations in cameroon and pan moist tropical equations including biomass data from africa. *Forest Ecol. Manage.* 260, 1873–1885. doi: 10.1016/j.foreco.2010.08.034

Duncanson, L., Rourke, O., and Dubayah, R. (2015). Small sample sizes yield biased allometric equations in temperate forests. *Sci. Rep.* 5:17153. doi: 10.1038/srep17153

Ebuy, J., Lokombe, J. P., Ponette, Q., Sonwa, D., and Picard, N. (2011). Allometric equation for predicting aboveground biomass of three tree species. *J. Trop. Forest Sci.* 23, 125–132.

Edwards, P. J., and Grubb, P. J. (1977). Studies of mineral cycling in a montane rain forest in new guinea: I. the distribution of organic matter in the vegetation and soil. *J. Ecol.* 65, 943–969.

Efron, B. (1987). Better bootstrap confidence intervals. *J. Am. Statist. Assoc.* 82, 171–185.

Feldpausch, T. R., Banin, L., Phillips, O. L., Baker, T. R., Lewis, S. L., Quesada, C. A., et al. (2011). Height-diameter allometry of tropical forest trees. *Biogeosciences* 8, 1081–1106. doi: 10.5194/bg-8-1081-2011

Feldpausch, T. R., Lloyd, J., Lewis, S. L., Brienen, R. J. W., Gloor, M., Monteagudo Mendoza, A., et al. (2012). Tree height integrated into pantropical forest biomass estimates. *Biogeosciences* 9, 3381–3403. doi: 10.5194/bg-9-3381-2012

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240.

Fromard, F., Puig, H., Mougin, E., Marty, G., Betoulle, J. L., and Cadamuro, L. (1998). Structure, above-ground biomass and dynamics of mangrove ecosystems: new data from french guiana. *Oecologia* 115, 39–53.

Gibbs, H. K., Brown, S., Niles, J. O., and Foley, J. A. (2007). Monitoring and estimating tropical forest carbon stocks: making redd a reality. *Environ. Res. Lett.* 2:045023. doi: 10.1088/1748-9326/2/4/045023

Goodman, R. C., Phillips, O. L., and Baker, T. R. (2014). The importance of crown dimensions to improve tropical tree biomass estimates. *Ecol. Appl.* 24, 680–698. doi: 10.1890/13-0070.1

Gorgens, E. B., Motta, A. Z., Assis, M., Nunes, M. H., Jackson, T., Coomes, D., et al. (2019). The giant trees of the amazon basin. *Front. Ecol. Environ.* 17, 373–374. doi: 10.1002/fee.2085

Hayashi, F. (2000). *Econometrics*. New Jersey, NJ: Princeton University Press.

Henry, M., Besnard, A., Asante, W. A., Eshun, J., Adu-Bredu, S., Valentini, R., et al. (2010). Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of africa. *Forest Ecol. Manage.* 260, 1375–1388. doi: 10.1016/j.foreco.2010.07.040

Henry, M., Bombelli, A., Trotta, C., Alessandrini, A., Birigazzi, L., Sola, G., et al. (2013). Globallometree: international platform for tree allometric equations to support volume, biomass and carbon assessment. *iForest Biogeosci. Forestry* 6, 326–330. doi: 10.3832/ifor0901-006

Hunter, M. O., Keller, M., Victoria, D., and Morton, D. C. (2013). Tree height and tropical forest biomass estimation. *Biogeosciences* 10, 8385–8399. doi: 10.5194/bg-10-8385-2013

Huxley, J. S. (1932). *Problems of Relative Growth, 1st Edn*. London, UK: Methuen & Co. Ltd.

Hyndman, R. J., and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice, 2nd Edn*. OTexts.

ISO-5725-1:1994(en) (1994). *Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 1: General Principles and Definitions.* Geneva: Standard, International Organization for Standardization.

JCGM-200:2012 (2012). *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (vim). 2008 Version With Minor Corrections.* Saint-Cloud: Guide, BIPM.

Jonsson, B. (1994). Prediction with a linear regression model and errors in a regressor. *Int. J. Forecast.* 10, 549–555.

Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., et al. (2017). Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Global Change Biol.* 23, 177–190. doi: 10.1111/gcb.13388

Kenzo, T., Furutani, R., Hattori, D., Kendawang, J. J., Tanaka, S., Sakurai, K., et al. (2009). Allometric equations for accurate estimation of above-ground biomass in logged-over tropical rainforests in sarawak, malaysia. *J. Forest Res.* 14, 365–372. doi: 10.1007/s10310-009-0149-1

Kerkhoff, A. J., and Enquist, B. J. (2009). Multiplicative by nature: why logarithmic transformation is necessary in allometry. *J. Theoret. Biol.* 257, 519–521. doi: 10.1016/j.jtbi.2008.12.026

Ketterings, Q. M., Coe, R., van Noordwijk, M., Ambagau', Y., and Palm, C. A. (2001). Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. *Forest Ecol. Manage.* 146, 199–209. doi: 10.1016/S0378-1127(00)00460-6

Lapicque, P. L. (1907). Tableau général des poids somatique et encéphalique dans les espéces animales. *Bulletins et Mémoires de la Société d'Anthropologie de Paris* 8, 248–270.

Larjavaara, M., and Muller-Landau, H. C. (2013). Measuring tree height: a quantitative comparison of two common field methods in a moist tropical forest. *Methods Ecol. Evol.* 4, 793–801. doi: 10.1111/2041-210X.12071

Mackensen, J., Tillery-Stevens, M., Klinge, R., and Fölster, H. (2000). Site parameters, species composition, phytomass structure and element stores of a terra-firme forest in east-amazonia, brazil. *Plant Ecology* 151, 101–119. doi: 10.1023/A:1026515116944

Marthews, T. R., Riutta, T., Oliveras, I. M., Urrutia, R., Moore, S., Metcalfe, D., et al. (2014). *Measuring Tropical Forest Carbon Allocation and Cycling: A RAINFOR-GEM Field Manual for Intensive Census Plots, 3 Edn.* Global Ecosystems Monitoring network.

Martin, A. R., Doraisami, M., and Thomas, S. C. (2018). Global patterns in wood carbon concentration across the world's trees and forests. *Nat. Geoscience* 11, 915–920. doi: 10.1038/s41561-018-0246-x

Martinez-Yrizar, A., Sarukhan, J., Perez-Jimenez, A., Rincon, E., Maass, J. M., Solis-Magallanes, A., et al. (1992). Above-ground phytomass of a tropical deciduous forest on the coast of jalisco, mexico. *J. Trop. Ecol.* 8, 87–96.

Menditto, A., Patriarca, M., and Magnusson, B. (2007). Understanding the meaning of accuracy, trueness and precision. *Accred. Qual. Assurance* 12, 45–47. doi: 10.1007/s00769-006-0191-z

Mitchard, E. T. A., Feldpausch, T. R., Brienen, R. J. W., Lopez-Gonzalez, G., Monteagudo, A., Baker, T. R., et al. (2014). Markedly divergent estimates of amazon forest carbon density from ground plots and satellites. *Global Ecol. Biogeogr.* 23, 935–946. doi: 10.1111/geb.12168

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

Morley, S. K., Brito, T. V., and Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather* 16, 69–88. doi: 10.1002/2017SW001669

Mugasha, W. A., Eid, T., Bollandsås, O. M., Malimbwi, R. E., Chamshama, S. A. O. C., Zahabu, E., et al. (2013). Allometric models for prediction of above- and belowground biomass of trees in the miombo woodlands of tanzania. *Forest Ecol. Manage.* 310, 87–101. doi: 10.1016/j.foreco.2013.08.003

Nelson, B. W., Mesquita, R., Pereira, J. L. G., de Souza, S. G. A., Batista, G. T., and Couto, L. B. (1999). Allometric regressions for improved estimate of secondary forest biomass in the central amazon. *Forest Ecol. Manage.* 117, 149–167.

Neyman, J., and Scott, E. L. (1960). Correction for bias introduced by a transformation of variables. *Ann. Mathemat. Statist.* 31, 643–655.

Ngomanda, A., Obiang, N. L. E., Lebamba, J., Mavouroulou, Q. M., Gomat, H., Mankou, G. S., et al. (2014). Site-specific versus pantropical allometric

equations: which option to estimate the biomass of a moist central african forest? *Forest Ecol. Manage.* 312, 1–9. doi: 10.1016/j.foreco.2013.10.029

Niiyama, K., Kajimoto, T., Matsuura, Y., Yamashita, T., Matsuo, N., Yashiro, Y., et al. (2010). Estimation of root biomass based on excavation of individual root systems in a primary dipterocarp forest in pasoh forest reserve, peninsular malaysia. *J. Trop. Ecol.* 26, 271–284. doi: 10.1017/S0266467410000040

Nogueira, E. M., Fearnside, P. M., Nelson, B. W., Barbosa, R. I., and Keizer, E. W. H. (2008). Estimates of forest biomass in the brazilian amazon: new allometric equations and adjustments to biomass from wood-volume inventories. *Forest Ecol. Manage.* 256, 1853–1867. doi: 10.1016/j.foreco.2008.07.022

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., et al. (2011). A large and persistent carbon sink in the world's forests. *Science* 333, 988–993. doi: 10.1126/science.1201609

Pek, J., Wong, O., and Wong, A. C. M. (2018). How to address non–normality: a taxonomy of approaches, reviewed, and illustrated. *Front. Psychol.* 9:2104. doi: 10.3389/fpsyg.2018.02104

Phillips, O. L., Sullivan, M. J. P., Baker, T. R., Monteagudo Mendoza, A., Vargas, P. N., and Vásquez, R. (2019). Species matter: Wood density influences tropical forest biomass at multiple scales. *Surveys Geophys.* 40, 913–935. doi: 10.1007/s10712-019-09540-0

Picard, N., Bosela, F. B., and Vivien, R. (2015a). Reducing the error in biomass estimates strongly depends on model selection. *Ann. Forest Sci.*, 72, 811–823. doi: 10.1007/s13595-014-0434-9

Picard, N., Rutishauser, E., Ploton, P., Ngomanda, A., and Henry, M. (2015b). Should tree biomass allometry be restricted to power models? *Forest Ecol. Manage.* 353, 156–163. doi: 10.1016/j.foreco.2015.05.035

Picard, N., Saint-André, L., and Henry, M. (2012). *Manual for Building Tree Volume and Biomass Allometric Equations.* Rome; Montpellier: Food and Agricultural Organization of the United Nations and Centre de Coopération Internationale en Recherche Agronomique pour le Développement.

Ploton, P., Barbier, N., Takoudjou Momo, S., Réjou-Méchain, M., Boyemba Bosela, F., Chuyong, G., et al. (2016). Closing a gap in tropical forest biomass estimation: taking crown mass variation into account in pantropical allometries. *Biogeosciences* 13, 1571–1585. doi: 10.5194/bg-13-1571-2016

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., and Hérault, B. (2017). Biomass: an r package for estimating above-ground biomass and its uncertainty in tropical forests. *Methods Ecol. Evol.* 8, 1163–1167. doi: 10.1111/2041-210X.12753

Roberts, M. J., Long, S. P., Tieszen, L. L., and Beadle, C. L. (1993). "Chapter: Photosynthesis and production in a changing environment," in *Measurement of Plant Biomass and Net Primary Production of Herbaceous Vegetation*, eds D. O. Hall, J. M. O. Scurlock, H. R. Bolhar-Nordenkampf, R. C. Leegood, and S. P. Long (Dordrecht: Springer), 1–21.

Ryan, C. M., Williams, M., and Grace, J. (2011). Above- and belowground carbon stocks in a miombo woodland landscape of mozambique. *Biotropica* 43, 423–432. doi: 10.1111/j.1744-7429.2010.00713.x

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl. Acad. Sci. U.S.A* 108, 9899–9904. doi: 10.1073/pnas.1019576108

Saldarriaga, J. G., West, D. C., Tharp, M. L., and Uhl, C. (1988). Long-term chronosequence of forest succession in the upper rio negro of colombia and venezuela. *J. Ecol.* 76, 938–958.

Shenkin, A., Chandler, C. J., Boyd, D. S., Jackson, T., Disney, M., Majalap, N., et al. (2019). The world's tallest tropical tree in three dimensions. *Front. Forests Global Change* 2:32. doi: 10.3389/ffgc.2019.00032

Shmueli, G. (2010). To explain or to predict? *Statist. Sci.* 25, 289–310. doi: 10.1214/10-STS330

Sileshi, G. W. (2014). A critical review of forest biomass estimation models, common mistakes and corrective measures. *Forest Ecol. Manage.* 329, 237–254. doi: 10.1016/j.foreco.2014.06.026

Slik, J. W. F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L. F., Ashton, P., et al. (2015). An estimate of the number of tropical tree species. *Proc. Natl. Acad. Sci. U.S.A* 112, 7472–7477. doi: 10.1073/pnas.1423147112

Sullivan, M. J. P., Lewis, S. L., Hubau, W., Qie, L., Baker, T. R., Banin, L. F., et al. (2018). Field methods for sampling tree height for tropical forest biomass estimation. *Methods Ecol. Evol.* 9, 1179–1189. doi: 10.1111/2041-210X.12962

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *J. Operat. Res. Soc.* 66, 1352–1362. doi: 10.1057/jors.2014.103

Vieilledent, G., Vaudry, R., Andriamanohisoa, S. F. D., Rakotonarivo, O. S., Randrianasolo, H. Z., Razafindrabe, H. N., et al. (2012). A universal approach to estimate biomass and carbon stock in tropical forests using generic allometric models. *Ecol. Appl.* 22, 572–583. doi: 10.1890/11-0039.1

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.

Yamakura, T., Hagihara, A., Sukardjo, S., and Ogawa, H. (1986). Aboveground biomass of tropical rain forest stands in indonesian borneo. *Vegetatio* 68, 71–82.