Re-evaluation of individual diameter:height allometric models to improve biomass estimation of tropical trees

Alicia Ledo,1,6 Thomas Cornulier,1 Janine B. Illian,2,3 Yoshiko Iida,4 Abdul Rahman Kassim,5 and David F. R. P. Burslem1

1School of Biological Sciences, University of Aberdeen, Aberdeen, AB24 3UU United Kingdom
2University of St. Andrews, Centre for Research into Ecological and Environmental Modelling, CREEM, The Observatory, Buchanan Gardens, St Andrews, KY16 9LZ United Kingdom
3Department of Mathematical Sciences, Norwegian University of Science and Technology, NTNU, NO-7491, Trondheim, Norway
4Forestry and Forest Product Research Institute, Kyushu Research Center, 4-11-16 Kurokami, Chuo-ku, Kumamoto, 860-0862 Japan
5Forest Research Institute of Malaysia, 68100, Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia

Abstract. Accurate estimation of tree biomass is necessary to provide realistic values of the carbon stored in the terrestrial biosphere. A recognized source of errors in tree aboveground biomass (AGB) estimation is introduced when individual tree height values ($H$) are not directly measured but estimated from diameter at breast height (DBH) using allometric equations. In this paper, we evaluate the performance of 12 alternative DBH:$H$ equations and compare their effects on AGB estimation for three tropical forests that occur in contrasting climatic and altitudinal zones. We found that fitting a three-parameter Weibull function using data collected locally generated the lowest errors and bias in $H$ estimation, and that equations fitted to these data were more accurate than equations with parameters derived from the literature. For computing AGB, the introduced error values differed notably among DBH:$H$ allometric equations, and in most cases showed a clear bias that resulted in either over- or under-estimation of AGB. Fitting the three-parameter Weibull function minimized errors in AGB estimates in our study and we recommend its widespread adoption for carbon stock estimation. We conclude that many previous studies are likely to present biased estimates of AGB due to the method of $H$ estimation.

Key words: bias; biomass; error; inaccuracy; modelling; tropical forest.

Introduction

Estimating tree height ($H$) as a function of diameter at breast height (DBH) using an allometric equation is a widespread practice in ecology and forestry, and the accuracy of these estimates have important implications for international environmental policies. These functions are used to represent how individuals and species allocate resources, to evaluate site quality, and to estimate wood stocks and above ground biomass (AGB) in forest vegetation. Accurate quantification and assessment of the AGB and carbon stocks in forest vegetation underpins policies to mitigate carbon dioxide emissions such as the UN-REDD+ program and the recommendations of the Intergovernmental Panel on Climate Change (Gibbs et al. 2007).

The only way to measure AGB directly is via destructive sampling. This approach has been used to develop empirical equations that estimate the individual AGB for tropical trees as a function of DBH and, in some cases, $H$ and wood density ($\rho$) (Feldpausch et al. 2012, Chave et al. 2014). Equations that include $H$ as a dependent variable are more accurate and errors arise in AGB estimation when $H$ is not available (Feldpausch et al. 2011, Chave et al. 2014). However, in most tropical forest inventories, only DBH is recorded, therefore $H$ is not available or, at best, measured on a subsample of trees. In this case, an allometric equation that relates specific tree DBH to $H$ can be fitted by using empirical data. The classic equations for relating DBH to $H$ employed in forestry are of the linear regression type, including a logarithmic transformation of the variables (Packard 2014). Probability distribution models, such as the Weibull or the Gompertz distributions, have also been used successfully (Molto et al. 2014). Although these models can be fitted and parametrized using actual data for local studies where suitable data exist, pantropical or regional allometric equations have also been proposed in the literature that allow estimation of tree $H$ from DBH (and sometimes climate) where local data are unavailable (Feldpausch et al. 2011). Since, for living trees, AGB has to
be estimated, there is always some uncertainty in the AGB values that has to be taken into account. Random unsystematic errors are not as concerning as systematic errors (Muller-Landau et al. 2014), because positive and negative errors are expected to follow a normal distribution and to be compensated for large samples. On the other hand, systematic errors introduce either positive or negative bias, and result in an overestimation or underestimation of AGB at the stand level that scales with the number of individuals included in the samples. A lack of $H$ data has been identified as a major sources of error in AGB estimation (Chave et al. 2004).

No study that compares a wide range of the proposed equations to estimate $H$ as a function of DBH has been published. In addition, the magnitude of errors and bias arising from the use of allometry estimated tree height ($\hat{H}$) instead of field measured individual $H$ in AGB calculation have not been assessed. In this paper, we present a comprehensive study comparing the performance, in terms of estimated values, mean square error, and bias, of the main proposed DBH:$H$ allometric models based on local data sets for three tropical forests: a lowland semi-deciduous forest in Panama, a lowland evergreen rain forest in Malaysia, and a montane cloud forest in Peru. We then evaluate precision and bias in AGB estimation caused by the use of $\hat{H}$ derived from the different DBH:$H$ equations, and determine whether the introduced compound errors due to $\hat{H}$ estimation were DBH dependent. Third, we determine the compound error in AGB estimation due to $\hat{H}$ estimation. The method we present here can be used for evaluating the performance of allometric models for any forest vegetation that has been sampled using equivalent techniques.

**METHODS**

**Study sites**

The first data set was collected from two seasonal lowland tropical moist forests on Barro Colorado Island (BCI) and the Gigante Peninsula in the central zone of Panama (Condit et al. 1998). Four different sources were used to compile data for a total of 8407 trees from 178 species, with tree diameters ranging from 1 to 295 cm (Bohlman and O’Brien 2006, Larjavaara and Muller-Landau 2013, Wright et al. 2010; and 2746 trees from Helene Muller-Landau, unpublished data). The second data set was from a lowland rainforest in the Pasoh National forest reserve in Peninsula Malaysia. Two different sources were used to compile data for a total of 6415 trees from 488 species, with tree diameters ranging from 1 to 585 cm (Iida et al. 2012; and 1112 trees [see data availability] measured in 2015 for this study). The third data set came from Bosque de Neblina de Cuyas, a montane cloud forest situated in the Andean range in north Peru (Ledo et al. 2012). This data set contains 13414 trees from 73 species, with DBH ranging from 0.1 to 257.8 cm. Unlike the first two data sets, the inventory in Cuyas was not selective, and all free-standing trees on three plots totaling 3 ha were included. Broken or damaged trees, palms, and lianas were excluded from the analyses in all cases.

**Comparison of allometric equations to estimate $H$ as a function of DBH**

We tested two kinds of allometric equations for estimating $H$ as a function of DBH (DBH:$H$). First, we fitted local empirical DBH:$H$ equations using a range of probability distribution models that have been employed both in ecology and forestry. We will refer to this first group of equations as fitted equations. Second, we used parametrized regional and pantropical DBH:$H$ equations that have been proposed in the literature, and this second group of equations will be referred to as literature equations. The seven fitted equations and the five literature equations we tested are presented in Table 1. We used non-linear least-squares for parameter estimation, via the nls function in R v3.1.1 (R Core Team 2014), which assumes the errors to be independent and normally distributed. The R code adopted for all the analyses can be found in Data S1. Measurement errors have a variance that depends on tree size, but this paper is focused exclusively on the performance of the DBH:$H$ equations rather than errors associated with data acquisition. We randomly selected one-half of the available data to fit the equations and used the second half to validate them in every case. We evaluated the performance of all these allometric models on the basis of their bias and root-mean-squared error (RMSE), where $\text{RMSE}_H = \sqrt{\left(\frac{\sum (\hat{H}_i - H_i)^2}{n-p}\right)}$ and $\text{bias}_H = \left(\frac{\sum (\hat{H}_i - H_i)}{n}\right)$. Here $n$ is the number of trees, and $p$ is the number of parameters in the model. $\hat{H}_i$ is the estimated value of tree height for tree $i$ derived from the fitted equation and $H_i$ is the value of tree height measured by the original studies. That value can be obtained using either field techniques or remote sensing. We acknowledge that height measurements obtained using these techniques are not equivalent, but a formal comparison of measurement techniques is beyond the scope of this paper (but see Hunter et al. [2013] and Larjavaara and Muller-Landau [2013]). If errors are not systematic, but depend on tree size, larger trees may contribute more to the total error. Hence, we also examined the individual error in $\hat{H}$ estimation ($\text{IE}_{H}$) as $\text{IE}_{H} = \hat{H} - H$ and observed how this quantity varies among trees of different size and among equations (Table 1). Finally, we re-fitted the best model using the entire data set for each site so that our estimates are based on as much information as possible and uncertainty in the estimates is reduced. Those models can be used in future studies at these locations.

**Error compounding in AGB estimation when H is estimated from a DBH:$H$ allometric equation**

We used the recently developed model proposed in Chave et al. (2014) to estimate AGB for individual trees in...
the tropics. To obtain a ρ value for our target species, we used the Global Wood Density Database (Zanne et al. 2010). We calculated AGB using the measured DBH and H data, and then with the estimated $\hat{H}$ values from the different DBH:H equations (Table 1). We assumed that AGB values calculated using Chave et al. (2014) represented actual values and calculated the individual errors in AGB estimation associated with estimating $H$ as $\hat{AGB}_H - \bar{AGB}_H$ and then the RMSE and bias

$$\text{RMSE}_{\text{AGB}} = \sqrt{\frac{\sum (\hat{AGB}_H - \bar{AGB}_H)^2}{n - p}}$$

as and

$$\text{bias}_{\text{AGB}} = \frac{\sum (\hat{AGB}_H - \bar{AGB}_H)}{n}.$$  

We then studied in more detail whether the compound errors in AGB due to $H$ estimation were size dependent and how this varied among DBH:H equations. To do so, we split the trees in each data set into 10-cm diameter classes from class 1 (0–10 cm) to class 10 (90–100 cm) plus class 11 (100–150 cm) and class 12 (>150 cm). The cumulative error in AGB in each class was calculated as the sum of individual errors:

$$\text{Error AGB}_\text{class} = \sum_{\text{tree}} (\hat{AGB}_H - \bar{AGB}_H)_{\text{class}}.$$  

We repeated these analyses but weighting the error in each DBH class by the total AGB in the class calculated using actual $H$

$$\text{Weight Error AGB}_\text{class} = \frac{\sum_{\text{tree}} (\hat{AGB}_H - \bar{AGB}_H)}{\sum \bar{AGB}_H_{\text{class}} \times 100}.$$

### Compounding of errors in estimated AGB due to $H$ estimation at plot level

Tropical forest inventories are often based on plots of 20 × 20 m or multiples of this basic unit (Burslem and Ledo 2015). Hence, we assembled simulated communities of trees from our data sets to mimic the communities on plots of this size to evaluate the compound error in AGB at the plot scale; which were 195, 216, and 170 trees in Central Panama, Pasoh, and Bosque de Cuyas, respectively. We randomly selected this number of trees to represent each plot with a probability weight per size class set to reproduce the observed DBH frequency distribution in each community. This weighted sampling was not required for the Bosque de Cuyas plot because the DBH and $H$ data were derived from complete inventories on actual plots. We defined the plot-level AGB error, PlotErrorAGB, as the sum of individual errors of all trees selected in the simulated or actual plots:

$$\text{Plot Error AGB} = \sum_{\text{tree}} (\hat{AGB}_H - \bar{AGB}_H)_{\text{plot}}.$$  

The $H$ values used to derive AGB$H$ were calculated from the DBH:H equations presented in Table 1. We repeated this sampling 999 times per site in order to obtain 95% confidence limits for the mean of PlotErrorAGB for the random plot subsamples.

### Results

#### Comparison of DBH:H allometric models

The mean $\hat{H}$ predicted values, RMSE$_H$ and bias$_H$ differed among DBH:H models, as did the shape of the DBH:H curve for each model (Fig. 1, with numeric results in Appendix S1). Fitted DBH:H equations based on local data (Eqs. 1–8 in Table 1) showed smaller RMSE$_H$ suggesting a better performance than their literature counterparts (Eqs. 9–12 in Table 1). The three-parameter Weibull function (Eq. 7) was the most accurate model in all three studied forests, in terms of lowest values of RMSE$_H$ and
bias \( \hat{H} \), followed by the Canham-Michaelis-Menten equation (Eq. 5) and the power law (Eq. 4). Most fitted equations underestimated \( H \), whereas equations derived from the literature always overestimated \( H \) (Appendix S1). Individual error distributions were notably different among models (Fig. 2). Linear-type equations (Eqs. 1–3) and literature-parameterized equations (Eqs. 9–12) severely overestimated tree height for trees with larger DBH values (Fig. 2). Once again, the three-parameter Weibull model (Eq. 7) generated fits with error distributions that were reasonable and centered in all three cases (Fig. 2, Appendix S1). The fitted equations using all individuals measured per site are presented in Appendix S2.

**Error compounding in AGB estimation when \( H \) is estimated from a DBH: \( H \) allometric equation**

The values of RMSE\(_{\text{AGB}}\) and bias\(_{\text{AGB}}\) varied strongly among equations (results in Appendix S3), which suggests that the selection of DBH: \( H \) model affects the accuracy of AGB estimation. In general, fitted equations yielded results with less error than equations derived from the literature. Literature equations (Eqs. 9–12), the linear equation (Eq. 1) and the power law (Eq. 4), overestimated AGB in every case, and the log-log equation (Eq. 3) in two of the three cases. The other distributions underestimated AGB in most cases, but to an extent that was different among forests (Appendix S3). The three-parameter Weibull model (Eq. 7) introduced the lowest values of RMSE\(_{\text{AGB}}\) and bias\(_{\text{AGB}}\) both for Central Panama and Bosque de Cuyas. For Pasoh, the Canham-Michaelis-Menten equation (Eq. 6) and the three-parameter Weibull (Eq. 7) were the most accurate equations. Owing to the use of estimated \( \hat{H} \) values, the error in estimated AGB increased strongly with DBH (Fig. 3). The big trees showed the largest errors in AGB both in absolute and proportional terms, which was expected due to the bias found in most DBH: \( H \) equations. The three-parameter
Fig. 2. Individual tree height estimation error in meters, $\text{IE}_H$, against DBH (cm), for each individual using the different DBH-$H$ equations presented in Table 1. Each orange dot corresponds to a tree. The mean value of the error is the navy line, with the 95% confidence interval of the mean in blue. The gray dotted line represents the 0 error. The figure is presented for the central Panama case only. Plots for the rest of the equations were similar and are presented in Appendix S4. P, pantropical equation; E, environmental equation; and R, regional equation. See Table 1 for equation details. (Color figure can be viewed at wileyonlinelibrary.com.)

Fig. 3. (Left) Values of error$A\text{GB}_{\text{class}}$, the error in $A\text{GB}$ (estimated aboveground biomass) contributed by each diameter class to the total error using different DBH-$H$ models (Table 1). (Right) Values of Weight Error$A\text{GB}_{\text{class}}$, the Error$A\text{GB}_{\text{class}}/A\text{GB}_{\text{class}}$. The number of trees in each diameter class were as follows: 0–10 cm, 4711 trees; 10–20 cm, 1045 trees; 20–30 cm, 709 trees; 30–40 cm, 426 trees; 40–50 cm, 447 trees; 50–60 cm, 283 trees; 60–70 cm, 194 trees; 70–80 cm, 110 trees; 80–90 cm, 155 trees; 90–100 cm, 108 trees; and DBH > 100 cm, 228 trees. Class bounds are indicated by the gray marks on the horizontal axis. The figure represents the data for the central Panama case only. Graphs for the other two study sites are similar and are presented in Appendix S5.
Weibull equation performed better for AGB estimation than the other equations because it was the least biased for the height of the larger trees (Fig. 3).

Compounding of errors in estimated AGB due to H estimation at plot level

The compounding of errors in estimated AGB at plot level due to inaccurate H estimation differed notably among DBH:H equations, as did the variability of errors across simulated plots (Fig. 4). The literature equations (Eqs. 9–12) always yielded overestimated AGB values, which were more than twice the AGB value. Fitted equations Eqs. 4–7 yielded the most accurate AGB estimations in every case, with mean error values around 0 (Fig. 4). Results from the three-parameter Weibull equation (Eq. 7) outperformed the other equations in terms of RMSE and biomass and resulted in more realistic AGB values. However, the range of the mean values obtained from the simulations was still wide, as was the variance (Fig. 4). Differences among forests were observed for some equations, which reflect the different characteristics of the original data sets. The Pasoh data set contains a small number of very big trees with irregular shapes, and these individuals are represented by large errors in estimated AGB for some equations (Fig. 3).

**Discussion**

**Describing DBH:H allometries**

Some DBH:H equations generated severely biased height predictions and should not be used (Fig. 2). The bias is mainly caused by inaccurate estimation of H for large trees (Figs. 2 and 3). The most biased equations are those that use a linear expression or assume an increasing curve that never reaches an asymptote, and this result was consistent whether the equations were fitted to local data or derived from regional or pantropical data sets available in the literature (Appendix S1). In some cases the estimated height of large trees is double the measured value (Fig. 1). These non-saturating functions are not ecologically realistic because tree height increment decreases with time, particularly once a tree has reached the forest canopy (Koch et al. 2004). On the other hand, small individuals invest comparatively more resources in stem height elongation, in order to overcome low light availability and reach the canopy (Taiz and Zeiger 2002). In contrast to linear regression models, nonlinear distribution models approaching a horizontal asymptote for trees with large DBH values are more realistic (Eqs. 5–8, 11) and the predicted H values differed from the actual values by only a few meters for large trees, and by a few centimeters for smaller trees (Appendix S1; Fig. 1).

Among those models, the three-parameter Weibull outperformed the other models (Appendix S1; Figs. 1–3) and emerged as the least biased function among those we tested (Appendix S1). Although measurement errors due to height estimation from the ground are not accounted for in our analyses (Hunter et al. 2013, Larjavaara and Muller-Landau 2013), they are unlikely to differentially affect the performance of different allometric models.

**Implications for AGB estimation**

Previous studies have found that including height measurements in allometric models improves the precision of AGB estimates (Feldpausch et al. 2012, Chave et al. 2014), and our study clearly supports this conclusion. If H is measured on a subsample of trees, the allometric equation used to predict H for unmeasured trees not only has to be accurate, but also, and more importantly, unbiased. Unsuitable equations introduce directional errors that either overestimate or underestimate...
AGB, and this error will increase as the DBH of the trees increases, resulting in less realistic values of AGB (Fig. 4), in which the estimated AGB can have a multiplicative error. A special case of interest is the estimated AGB of large trees. Big trees store disproportionately large amounts of carbon, so they are the main agents responsible for variation in carbon storage at stand level (Körner 2009). However, the amount of error in \( \bar{H} \) estimation is greatest for very large trees (Fig. 3), resulting in an increased potential bias in AGB estimation if \( \bar{H} \) is not estimated accurately. In addition, the precision of the AGB equation also decreases for very small or large trees (Chave et al. 2014), so errors due to inaccurate \( \bar{H} \) values may compound and the resulting carbon stock values may be highly unrealistic. The selection of inaccurate models in prior studies may have yielded inaccurate estimates of AGB and carbon stock values. Imprecise estimation of AGB values may lead to imprecise estimates of emissions and existing quantities of carbon stored in tropical forests. This could have harmful consequences for climate change evaluation and hence mitigation policies.

**Acknowledgments**

The first author was supported by the European Union under a IEF Marie-Curie Action. We thank Helene Muller-Landau and S. Joseph Wright for providing data on tree heights for some trees in Panama and for comments that helped to develop this study.

**Literature Cited**


**Supporting Information**

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/eap.1450/full

**Data Availability**

Data associated with this paper are available in figshare: https://dx.doi.org/10.6084/m9.figshare.3892506.v1