forests

MDPI

Article

Allometric Models for Estimation of Forest Biomass in North East India

Arun Jyoti Nath [1,*], Brajesh Kumar Tiwari [2], Gudeta W Sileshi [3], Uttam Kumar Sahoo [4], Biplab Brahma [1], Sourabh Deb [5], Ningthoujam Bijayalaxmi Devi [6], Ashesh Kumar Das [1], Demsai Reang [1], Shiva Shankar Chaturvedi [2], Om Prakash Tripathi [7], Dhruba Jyoti Das [8] and Asha Gupta [9]

1 Department of Ecology and Environmental Science, Assam University, Silchar 788011, Assam, India; biplab.brahma@yahoo.com (B.B.); asheshkd@gmail.com (A.K.D.); demn000@gmail.com (D.R.)
2 Department of Environmental Studies, North Eastern Hill University, Shillong 793022, India; bktiwarinehu@gmail.com (B.K.T.); sschaturvedinehu@gmail.com (S.S.C.)
3 School of Agricultural, Earth & Environmental Sciences, University of Kwazulu-Natal, Pietermaritzburg 4041, South Africa; sileshigw@gmail.com
4 Department of Forestry, Mizoram University, Aizawl 796004, India; uksahoo_2003@rediffmail.com
5 Department of Forestry and Biodiversity, Tripura University, Suryamaninagar 799022, India; drsourabhdeb@gmail.com
6 Department of Botany, Sikkim University, Gangtok 737102, India; bijayalaxn@yahoo.co.in
7 Department of Forestry, North Eastern Regional Institute of Science and Technology, Itanagar 791109, India; tripathiom7@gmail.com
8 Rain Forest Research Institute, Jorhat 785010, India; dhrubajyoti.india@gmail.com
9 Department of Life Sciences, Manipur University, Imphal 795003, India; ashaguptamu@gmail.com
* Correspondence: arunjyotinath@gmail.com; Tel.: +91-384-227-0824

Received: 18 October 2018; Accepted: 23 January 2019; Published: 28 January 2019

Abstract: In tropical and sub-tropical regions, biomass carbon (C) losses through forest degradation are recognized as central to global terrestrial carbon cycles. Accurate estimation of forest biomass C is needed to provide information on C fluxes and balances in such systems. The objective of this study was to develop generalized biomass models using harvest data covering tropical semi-evergreen, tropical wet evergreen, sub-tropical broad leaved, and sub-tropical pine forest in North East India (NEI). Among the four biomass estimation models (BEMs) tested $AGB_{est} = 0.32(D^2H\delta)^{0.75} \times 1.34$ and $AGB_{est} = 0.18D^{2.16} \times 1.32$ were found to be the first and second best models for the different forest types in NEI. The study also revealed that four commonly used generic models developed by Chambers (2001), Brown (1989), Chave (2005) and Chave (2014) overestimated biomass stocks by 300–591 kg tree^{-1}, while our highest rated model overestimated biomass by 197 kg tree^{-1}. We believe the BEMs we developed will be useful for practitioners involved in remote sensing, biomass estimation and in projects on climate change mitigation, and payment for ecosystem services. We recommend future studies to address country scale estimation of forest biomass covering different forest types.

Keywords: Biomass estimation models; forest ecosystems; remote sensing; winners curse

1. Introduction

Land based climate change mitigation strategies have received much global attention in the recent past, due to the large sink capacity and economic viability [1]. Among the terrestrial ecosystems, tropical and sub-tropical forests are considered central to global terrestrial carbon (C) stocks [1,2]. Tropical and sub-tropical regions are well-recognized for losing forests due to agricultural expansion

and bio-energy production [3]. Information on tropical forest biomass and C fluxes is gaining both economic and political currency in renewable energy development, C credit markets and research in global environmental change. Since the approval of the REDD$^+$ (reducing emissions from deforestation and forest degradation) during the Conference of the Parties (COP 19) in November 2013, the Warsaw Framework has become a formal mechanism for creating performance-based C financing. The Warsaw Framework requires consistency in methods, definitions, and comprehensiveness in measurement, reporting and verifying emissions by sources and removals by sinks, and changes in C storages [4].

In spite of the significant progress in biomass estimation methods, quantifying C stocks in tropical and sub-tropical forests is still challenging. A large degree of uncertainty exists in measured C stocks and fluxes in tropical and sub-tropical forests [5–9]. Some of the uncertainty results from the lack of consistencies in methods, thus leading to widely varying results even among similar studies. Based on remotely sensed data, two independent studies, namely Harris et al. [5] and Baccini et al. [8], published maps of tropical forest C storages, which are widely used in REDD$^+$ projects [6]. For 2000–2005, Harris et al. [5] reported gross emission of 0.81 Pg (Petagram, 1 Pg = 10^{15} g) C year^{-1} compared to 2.22 Pg C year^{-1} by Baccini et al. [8]. Although the maps developed in both the studies used the same LiDAR data from the Geoscience Laser Altimeter System (GLAS) across the tropics, the maps reveal substantial differences in total biomass stocks, with little consistency in the direction of differences [6,9]. Differences of this magnitude are reason for concern, not only in policy formulation, but also global climate change science [9]. According to Mitchard [6] one of the causes of the differences between the two maps is the difference in the biomass estimation models (BEMs) used to estimate biomass from the ground plots. The choice of BEMs can significantly influence local, regional and global biomass estimates. In addition, the choice of BEMs poses a practical limit to the accuracy with which remote sensing methods can predict regional biomass [6]. This highlights the need for the development of region-specific BEMs.

The introduction of C credits under REDD has financial implications to C stock in tropical and sub-tropical forest ecosystems [10]. Therefore, accurate estimation of C loss and sequestration is fundamental to the initiatives of managing forested ecosystems for reducing CO_2 emissions [11] and model applications to climate change studies [12]. Accuracy in estimates of forest C stocks are limited by the challenge of developing robust models to estimate tree biomass [13]. Different direct and indirect methods are used for biomass estimation. In the direct methods, a sample of trees in a given area is harvested and measured for estimation of dry weight in different tree components (e.g., trunk, branches, and leaves). Direct methods can be expensive, especially when dealing with large sample areas and several species [14]. In the indirect method, biomass is usually estimated using BEMs, which relate measurable variables such as total tree height, diameter at breast height, and woody density to total tree biomass [15–17]. Therefore, the use of indirect methods is often preferred over direct methods [18].

Multi-species pan tropical models have been developed for estimation of above ground biomass (AGB) for major forest types [16,19]. However, such models may not accurately predict biomass of forests in different ecological regions of the world [20,21]. Species-specific models [16] and LiDAR technology can also be unreliable for application to mixed species stands.

In the tropical and sub-tropical parts of North East India (NEI), forests cover 66% of the total geographical area [22]. These forests have significant influence on regional and national C balance. In NEI, different species-specific models have been developed for *Pinus kesiya* [23], *Hevea brasiliensis* [24], and *Barringtonia acutangula* [25]. Given the uncertainty in biomass estimation, these species-specific models may not have wider application in mixed species tropical forests. Biomass and C stock for diverse forest ecosystems in NEI have also been estimated [26–29] using various generic models including those developed by Brown et al. [30], Chambers et al. [31] and Chave et al. [16,19]. However, the accuracy of biomass estimates using these models has rarely been tested. Therefore, accurate estimation of biomass and C stock in NEI will form the baseline for regional/national C balance datasets.

The objective of this study was to develop generalized biomass models using harvest data covering tropical semi-evergreen, tropical wet evergreen, sub-tropical broad leaved, and sub-tropical pine forest in NEI. The aim of this study is to present generalized BEMs for NEI for use by practitioners involved in projects on climate change mitigation, and payment for ecosystem services.

2. Methods

2.1. Descriptions of the Study Region

This study covered NEI, which consists of eight different states: Assam, Arunachal Pradesh, Mizoram, Meghalaya, Manipur, Nagaland, Sikkim, and Tripura. NEI covers 26.3 million hectares (M ha) equivalent to 8% of total geographical area of India [32]. NEI is situated at the confluence of the Indo-Chinese, Indo-Malayan, and Indian bio-geographical realms. Due to this unique geographical location, NEI represents numerous forest types falling within one of the biodiversity hotspots of the world, the Indo-Burma biodiversity hotspot [33]. In NEI, 17.2 M ha of land is covered with forests, which constitutes ~25% of India's total forest area [32] and is represented by five broad forest types based on the elevation, forest structure and composition: (i) tropical semi-evergreen (up to 600 m), (ii) tropical wet evergreen (up to 900 m), (iii) sub-tropical broad leaved (900–1900 m), (iv) sub-tropical pine (1000–3500 m), and (v) alpine temperate (above 3500 m) [34,35].

2.2. Sampling Strategies

Sample tree data (a total of 303 trees) were collected from four major forest types (Table 1, Figure 1) accounting for over 90% of the forest cover in NEI. These four forest types span over tropical to sub-tropical climates. Alpine zone that represents temperate forests of NEI accounts for only 5%–7% of total forest cover of NEI. Therefore, this forest type was not included in this study.

To cover diverse tree sizes for harvest, seven diameter classes were formed: 10.1–20 cm, 20.1–30 cm, 30.1–40 cm, 40.1–50 cm, 50.1–60 cm, 60.1–70 cm, and 70.1–90 cm. Tree sizes $\leq$10 diameter at breast height (D) were not considered for harvest. Sample trees were harvested from a minimum of four different diameter classes for each of the four different forest types. This sampling strategy was adopted because for certain forest types, trees were distributed up to maximum of four diameter classes. Then selected trees were cut at ground level and total height (m) was measured. After felling, tree components were separated in to leaf, branch, and bole, and fresh weight of each component was measured in the field with a digital balance. Sub-samples (1 kg) of each component were collected, taken to the laboratory and dried at 65 °C until a constant weight was reached. Then, the fresh weight to dry weight ratio was used to compute the dry mass of the total tree.

Table 1. Dominant and co-dominant tree species sampled for the present study.

Forest Types	Altitudinal Range (Meters)	Species
Alpine Temperate	>3500	Data not available for this study
Sub-Tropical Pine	1000–3500	*Pinus kesiya* Royle ex Gordon, *Pinus roxburghii* Sarg.
Sub-Tropical Broad Leaved	900–1900	*Schima wallichii* Reinw. ex Blume, *Quercus oblongata* D. Don, *Ficus benghalensis* L., *Machilus gamblei* King ex Hook.f., *Mallotus philippensis* (Lam.) Muller.-Arg., *Myrica sapinda* Wall., *Terminalia myriocarpa* Van Heurck & Mull., *Terminalia chebula* (Gaertn) Retz, *Toona ciliata* M.J. Roem, *Juglans regia* L., *Alnus nepalensis* D. Don
Tropical Wet Evergreen	Up to 900	*Tectona grandis* L.f., *Macaranga denticulata* (Blume) Muller. -Arg., *Mesua ferrea* L., *Dipterocarpus turbinatus* C.F.Gaertn
Tropical Semi-Evergreen	Up to 600	*Albizia procera* (Roxb.) Benth., *Syzygium cumini* (L.) Skeels, *Macaranga peltata* (Roxb.) Muller. -Arg., *Bauhinia variegata* L., *Artocarpus chama* Buch. –Ham.

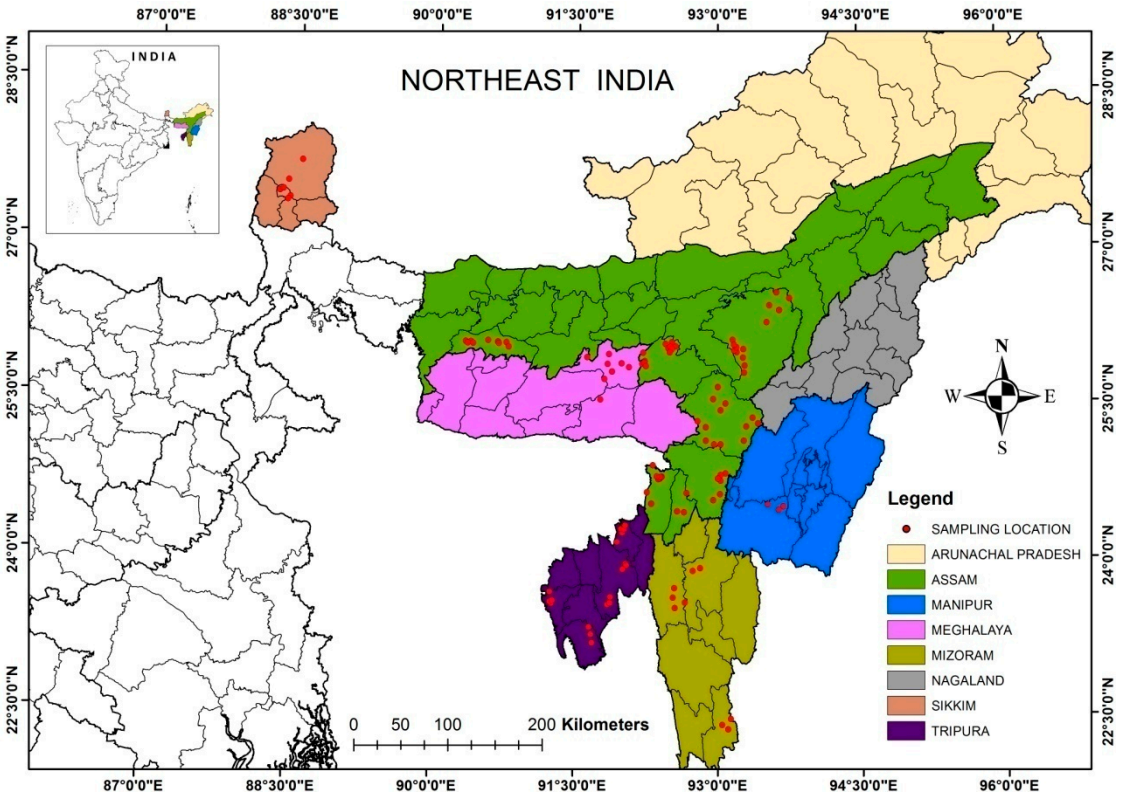

Figure 1. Geographical location of sampling points.

2.3. Model Development

Development of an empirical biomass model is subject to the use of appropriate independent variables and covariates that are likely to influence biomass and the chosen variables. Although the selection of independent variables still remains a matter of debate [17], D and total tree height (H) are the most commonly used variables when developing BEMs [16,17]. The compound forms of D and H with or without wood density (δ) are also widely used in BEMs [13,19]. The conventional approach in the development of BEMs has been to use models with only fixed effects, i.e., without considering covariates. As biomass estimates may vary not only with the fixed effects, but also

covariates such as forest type, stand density, site quality, competition, etc., it is important to include such covariates as sources of additional variation in the model. In this study, we used both conventional and a linear mixed modeling (LMM) framework where covariates are considered in model development. The general form of the LMM model is given as:

$$Y_i = \beta X_i + u_i Z_j + \varepsilon_i. \quad (1)$$

where Y_i is the n-dimensional response vector, β is the p-dimensional parameter vector for fixed-effects, X_i is the nxp design matrix for fixed-effects, u_i is the q-dimensional vector of parameters for random effects; Z_i is the nxq design matrix for random-effects, ε_i is the n-dimensional error vector [36]. It is assumed that the expectation (E) is $E(\varepsilon_i) = E(u_i) = 0$, and the variances (V) and co-variances (Cov) are: $Var(u_i) = R_i$, $Var(u_i) = D_i$, $Cov(\varepsilon_i, u_i) = 0$, respectively. Furthermore, ε_i and u_i are assumed to be normally distributed as: $\varepsilon_i \sim N(0,R_i)$ and $u_i \sim N(0,D_i)$ [36].The rationale for the LMM is that both the fixed and random parameters can be entered in the model simultaneously, thus providing consistent estimates of parameters and their standard errors than the conventional method [37].

In this study, we had only information on forest type, but information was lacking on stand density, site quality and competition. Initially, we entered a random intercept and forest type as random effects in the model, but this created problem with convergence of algorithms and testing the significance of the random intercept. Although convergence criteria were met, the final hessian was not positive definite when the intercept was entered as a random effect. Therefore, we used only forest type as the random effect. Here, four forest types, namely, sub-tropical pine, sub-tropical broad leaved, tropical wet evergreen and tropical semi-evergreen forest were considered in the model.

Both the conventional and LMM versions of four commonly used BEMs [17] were compared here. In the conventional framework, the four models are formulated as follows:

Model 1: $\ln(AGB) = \ln(a) + bln(D) + \varepsilon;$
Model 2: $\ln(AGB) = \ln(a) + bln(D^2H) + \varepsilon;$
Model 3: $\ln(AGB) = \ln(a) + bln(D^2H\delta) + \varepsilon;$
Model 4: $\ln(AGB) = \ln(a) + bln(D) + cln(H) + dln(\delta) + \varepsilon$

where AGB is total above-ground biomass in kg dry matter tree^{-1}, a is the intercept, b, c and d are slope parameters, D is diameter at breast height in centimeters, H is total height in meters, δ is wood density (or specific gravity) in kg m^{-3} and ε is the random error. In all models we had H measurements, but δ was obtained from Global Wood Density Database [38]. Upon back transformation, each model was multiplied with a correction factor (CF) estimated from the residual variance or mean square error (MSE) using the formula $CF = exp(MSE/2)$ [39]. The CF accounts for the back transformed of the error and a step in log-transformed data in BEMs [17].

In the LMM framework, the models were formulated as follows:

Model 1: $\ln(AGB) = \ln(a) + bln(D) + uF + \varepsilon;$
Model 2: $\ln(AGB) = \ln(a) + bln(D^2H) + uF + \varepsilon;$
Model 3: $\ln(AGB) = \ln(a) + bln(D^2H\delta) + uF + \varepsilon;$
Model 4: $\ln(AGB) = \ln(a) + bln(D) + cln(H) + dln(\delta) + uF + \varepsilon$

where AGB, D, H, δ, a, b, c, d and ε are defined as in the conventional. The added quantity is the parameter u, which is the random effect of forest type (F). In all LMM models, it is assumed that u and ε are uncorrelated, random and normally distributed variates with mean 0 and variance of 1. In these models, there are two variance components; the variance associated with the random effect (σ_u^2) and the residual variance (σ_ε^2).

To ensure comparability of parameters, we estimated parameters of both the conventional and LMM models using the PROC MIXED procedure in the SAS system. The SAS codes used for estimation of parameters of the conventional and LMM models are given in Tables A1 and A2, respectively. PROC

MIXED uses the restricted maximum likelihood (REML) estimation method. The REML method is known to produce consistent estimates of the variance-covariance matrix [37] and asymptotic standard errors of the covariance parameter estimates.

The performance of the different models was assessed using the bias-corrected Akaike information criterion (AICc) and the coefficient of determination (R^2). In the LMM framework, the R^2 is not routinely estimated due to theoretical problems in its definition and practical difficulties in implementation [40]. As such, PROC MIXED does not normally report it. Therefore, R^2 was estimated as: $R^2 = 1 - RV_m/RV_i$, where RV_m is the residual variance of the full model, and RV_i is the residual variance of empty (intercept-only) model [40].

LMMs are based on the theory of empirical best linear unbiased predictors (EBLUPs) of the random effects and the best linear unbiased estimates (BLUEs) of the fixed effects [37]. The advantage of EBLUPs estimation is that it does not require normality of the random effects [41]. As such, LMMs provide an efficient approach to small area (or domain) estimation by incorporating random effects that account for dissimilarities between domains (forest type in our analysis). Taking advantage of this property, localized AGB predictions that are specific to each forest type were generated using the intercepts of the different forest types for the two highly ranked LMM model. Finally, the standardized residuals generated using the linear mixed effects model were plotted against the independent values to check for heteroscedasticity in residuals. Outliers were detected by checking the studentized residuals, and values below −2 or +2 were considered as outliers [17]. Heteroscedasticity will typically be manifested as residuals whose magnitude is correlated with that of the response variable and a plot of the residuals against the predicted values will reveal a megaphone pattern if the errors are not homogenous [17].

2.4. Model Validation

Cross validation is usually recommended to determine how accurately BEMs will perform when applied to an independent dataset. Usually 5-fold or 10-fold cross validation provides a good balance between bias and variance [17]. Therefore, a 10-fold cross-validation was employed to evaluate the predictive performance of the four selected models. The goodness of fit criteria was calculated for the validation dataset using the lava and forecast packages of the R package (Appendix A Table A3). Specifically, the adjusted coefficient of determination (Adj R^2), root mean square of error (RMSE), the AICc and Bayesian information criterion (BIC) were estimated from the 10-fold cross validation using the forecast packages of R (see Appendix A).

2.5. Comparison with Generic Models

We also compared our highest ranked LMM model with the following commonly used generic models developed by Brown et al. (1989) [30], Chamber et al. (2001) [31], Chave et al. (2005) [16] and Chave et al. (2014) [19] for broadly defined forest types. We chose these models for comparison with our own models because they are commonly used for biomass estimation in NEI by previous researchers [26–29].

Chamber's model is given as

$$AGB = exp\left[-0.370 + 0.333\ln D + 0.9333 lnD^2 - 0.122 lnD^3\right] \quad (2)$$

The model proposed by Brown et al. (1989) for tropical regions is:

$$AGB = 13.2579 - (4.8945D) + 0.6713D^2 \quad (3)$$

Chave's model 1 was based on his equation for tropical wet forests (Chave et al. 2005) given as:

$$AGB = 0.0776\left(\rho D^2 H\right)^{0.940} \quad (4)$$

Chave's model 2 is Equation (4) of Chave et al. (2014) said to be the best-fit pan tropical model. It is given as:

$$AGB = 0.0673\left(\rho D^2 H\right)^{0.976} \quad (5)$$

We determined the appropriateness of these generic models by comparing with our highest ranked model using the R^2, RMSE, relative prediction error (Error), mean absolute percentage error (MAPE) and AICc. We also compared the 95% CI (confidence interval) of the slopes (b) of the regression of measured against fitted values to assess whether or not significant prediction errors exist. If significant prediction errors exist $b \neq 1$, and the 95% CI of b will not cover 1 [17].

3. Results

Parameter estimates of models 1–4 generated using the conventional method and the linear mixed effects modeling (LMM) framework are given in Table 2. The AICc and adjusted R^2 show that models 1–4 fitted using LMM are superior to those fitted using the conventional method (Table 2). Therefore, all inferences hereafter will be based on the models fitted using the mixed effects framework. Estimates of the covariance parameters for the different LMM models are given in Table 3. In all models the variance of random effect was not significantly different from zero, while the residual variance was significantly larger than zero (Table 3). The largest residual variance was recorded in Model 2.

Table 2. Comparison of models fitted using the conventional (CONV) method with models fitted within a linear mixed effects (LME) framework where forest type was used as a random effect. The Akaike information criterion (AIC), adjusted R^2 (Adj R^2) and RMSE were used for comparing CONV with LMM.

Model Parameters							Adj	
Model		***ln(a) (SE)*** †	***b (SE)***	***c (SE)***	***d (SE)***	**AIC** ‡	**R^2**	**RMSE**
1	CONV	−2.12 (0.34)	2.32 (0.10)			769	0.62	0.851
	LME	−1.73 (0.45)	2.16 (0.10)	-		**705**	**0.705**	**0.749**
2	CONV	−2.30 (0.35)	0.82 (0.04)			771.7	0.619	0.852
	LME	−1.64 (0.44)	0.74 (0.04)	-		**734.5**	**0.675**	**0.785**
3	CONV	−1.83 (0.32)	0.82 (0.04)			770.3	0.621	0.850
	LME	−1.25 (0.43)	0.75 (0.04)	-		**725.7**	**0.685**	**0.773**
4	CONV	−2.12 (0.39)	2.00 (0.16)	0.43 (0.16)	0.37 (0.34)	763.2	0.627	0.842
	LME	−1.21 (0.49) ^{ns}	2.22 (0.16)	−0.08 (0.16) ^{ns}	0.87 (0.30)	**698.8**	**0.710**	**0.740**

† Figures in parentheses are asymptotic standard errors (SE) of parameter estimates; ‡ AIC values in bold face indicate the better model when the conventional was compared with LMM; ns: not significant: Note: Model 1: $ln(AGB) = ln(a) + b^*ln(D)$; Model 2: $ln(AGB) = ln(a) + b^*ln(D^2H)$; Model 3: $ln(AGB) = ln(a) + b^*ln(D^2H\delta)$; Model 4: $ln(AGB) = ln(a) + b^*ln(D) + c^*ln(H) + d^*ln(\delta)$.

Table 3. Estimates of covariance parameters and their significance for the different models.

Model §	**Covariance Parameter**	**Variance Estimate**	**Z Value** †	***p* Value** ‡
1	Forest	0.38 (0.32)	1.18	0.1187
	Residual	0.56 (0.05)	12.21	<0.0001
2	Forest	0.30 (0.26)	1.16	0.1236
	Residual	0.62 (0.05)	12.20	<0.0001
3	Forest	0.33 (0.28)	1.17	0.1218
	Residual	0.60 (0.05)	12.20	<0.0001
4	Forest	0.45 (0.38)	1.18	0.1182
	Residual	0.55 (0.05)	12.16	<0.0001

† The Z value is the Wald Z-test for covariance parameter estimates; ‡ p value represents the significance of the Z test of the hypothesis that the variance of the effect is 0; § Specification of models is as in Table 2.

Table 4 provides goodness-of-fit statistics of the cross validation of models 1–4. Models 1–3 were indistinguishable in terms of RMSE, AICc, BIC and R^2 (Table 4). Although the cross-validation goodness of fit criteria indicates that model 4 is the best, the mixed model analysis revealed that parameters a and d in Model 4 are not significantly different from zero. Therefore, Model 3, the one

with the next smaller RMSE, AICc and BIC, was chosen as the best model. Model 3 is given as $AGB_{est} = 0.32(D^2H\delta)^{0.75} \times 1.34$ in the arithmetic domain where 1.34 is the correction factor. The second best option would be the power law model given as $AGB_{est} = 0.18D^{2.16} \times 1.32$ in the arithmetic domain. The AGB values predicted using the different models are shown in Figure A1. In all models, the residuals did not reveal heterogeneity of variance (Appendix A Figure A1).

The LMM estimates of the intercept (and SE) of Model 3 were −1.01 (0.34) for sub-tropical pine, −0.83 (0.36) for sub-tropical broad leaved, −1.10 (0.31) for tropical wet evergreen, and −2.08 (0.31) for tropical semi evergreen forest. The predictions produced using these estimates for each forest type are presented in Figure 2. For Model 1, the forest type-specific intercept were −1.34 (0.34) for sub-tropical pine, −1.46 (0.37) for sub-tropical broad leaved, −1.49 (0.31) for tropical wet evergreen, and −2.64 (0.32) for tropical semi-evergreen forest, and corresponding the predictions for each forest type are presented in Figure 3. The discrepancies in predictions between the conventional and the LMM models were largest in tropical wet evergreen forest (Figure 2d), and the discrepancies increased with tree diameter (Figure 3a). On the other hand, the smallest discrepancy was observed in tropical semi-evergreen forest subtropical pine forest (Figures 2d and 3d).

Table 4. Comparison models using cross validation goodness of fit criteria including the adjusted coefficient of determination (Adj R^2), root mean square of error (RMSE), the bias-corrected Akaike information criterion (AICc) and Bayesian information criterion (BIC).

Model	Adj R^2	RMSE	AICc	BIC
1	0.62	0.728	−93.7	−82.7
2	0.62	0.729	−93.1	−82.0
3	0.62	0.726	−94.4	−83.4
4	0.63	0.713	−98.5	−83.8

Note: Model 1: ln(AGB) = ln(a) + b*ln(D); Model 2: ln(AGB) = ln(a) + b*ln(D^2H); Model 3: ln(AGB) = ln(a) + b*ln(D^2Hδ); Model 4: ln(AGB) = ln(a) + b*ln(D) + c*ln(H) + d*ln(δ).

Table 5. Comparison of the predictions of our highest rated model with the generic models using the various goodness of fit criteria and the slope (*b*) of the measured above ground biomass (AGB) with the fitted values.

Model	R^2	RMSE	Error	MAPE	AICc	*b* (95% CL)
Our highest rated model	0.869	141	197.4	235.7	3001.8	1.06 (1.01–1.11)
Chamber's model	0.870	140	591.4	595.4	2997.9	0.33 (0.32–0.35)
Brown's model	0.848	151	303.6	314.8	3045.7	0.59 (0.56–0.61)
Chave's model 1	0.829	161	299.6	313.2	3081.8	0.42 (0.40–0.45)
Chave's model 2	0.821	165	372.8	382.7	3096.5	0.32 (0.31–0.34)

With high R^2 and low AICc, RMSE and MAPE values, our highest rated model was also superior to the generic model developed by Brown and the two pan-tropical models developed by Chave (Table 5). Although our highest rated model was comparable with Chamber's model in terms of R^2, AICc, and RMSE, it was superior in terms of MAPE and prediction error (Table 5). With $b < 1$ all the generic models also have significant prediction errors (Table 5). Compared to our highest rated model, Chamber's model and Chaves model 2 severely overestimated AGB especially for trees with DBH exceeding 40 cm (Figure 4a). The largest average deviation from our model prediction (595.5 kg tree^{-1}) was recorded with Chamber's model, while the lowest (191.4 kg tree^{-1}) was recorded with Brown's model (Figure 4b). However, the deviations increased with increasing tree diameter (Figure 3b). Irrespective of tree diameter, when errors in estimation were evaluated, Chamber's, Brown's, Chave Model 1 and Chave Model 2 overestimated biomass stock by 591.4, 303.6, 300 and 372.8 kg tree^{-1}, respectively, while our highest rated model overestimated biomass stock by 197.4 kg tree^{-1}.

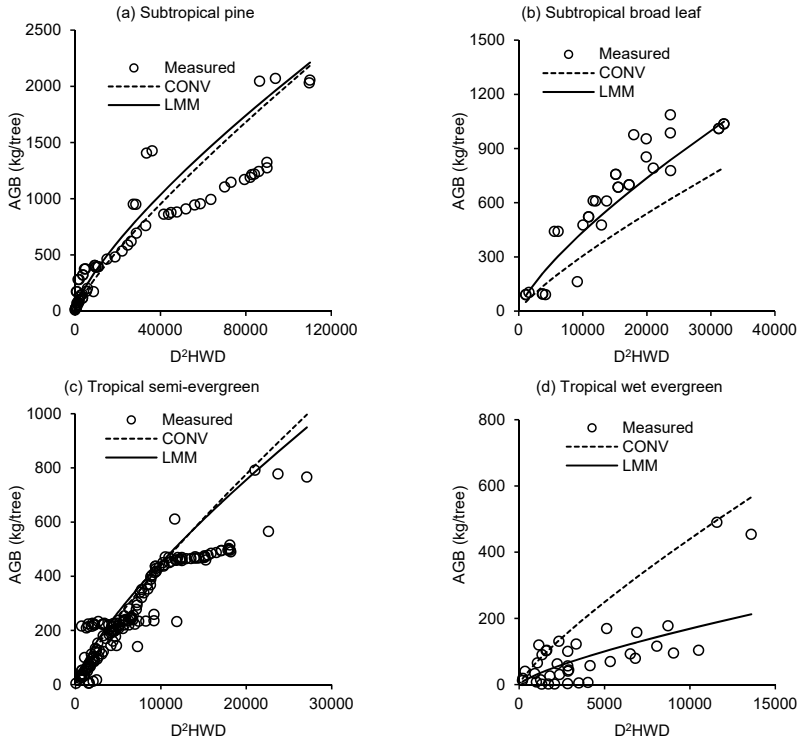

Figure 2. The observed AGB (open circles) and fitted lines for the four forest types produced using the highest rated model (Model 3) in the arithmetic domain. The dashed fitted lines represent predictions produced using the conventional (CONV) method, while the smooth lines represent predictions produced using the LMM procedure.

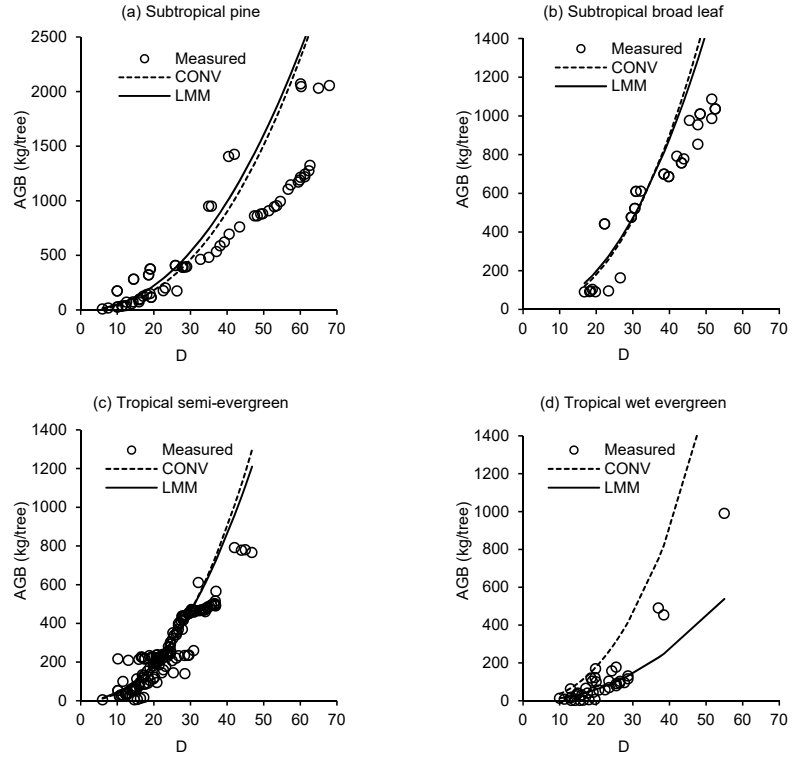

Figure 3. The observed AGB (open circles) and fitted lines for the four forest types produced using the second highest rated model (Model 1) in the arithmetic domain. The dashed fitted lines represent predictions produced using the conventional (CONV) method, while the smooth lines represent predictions produced using the LMM procedure.

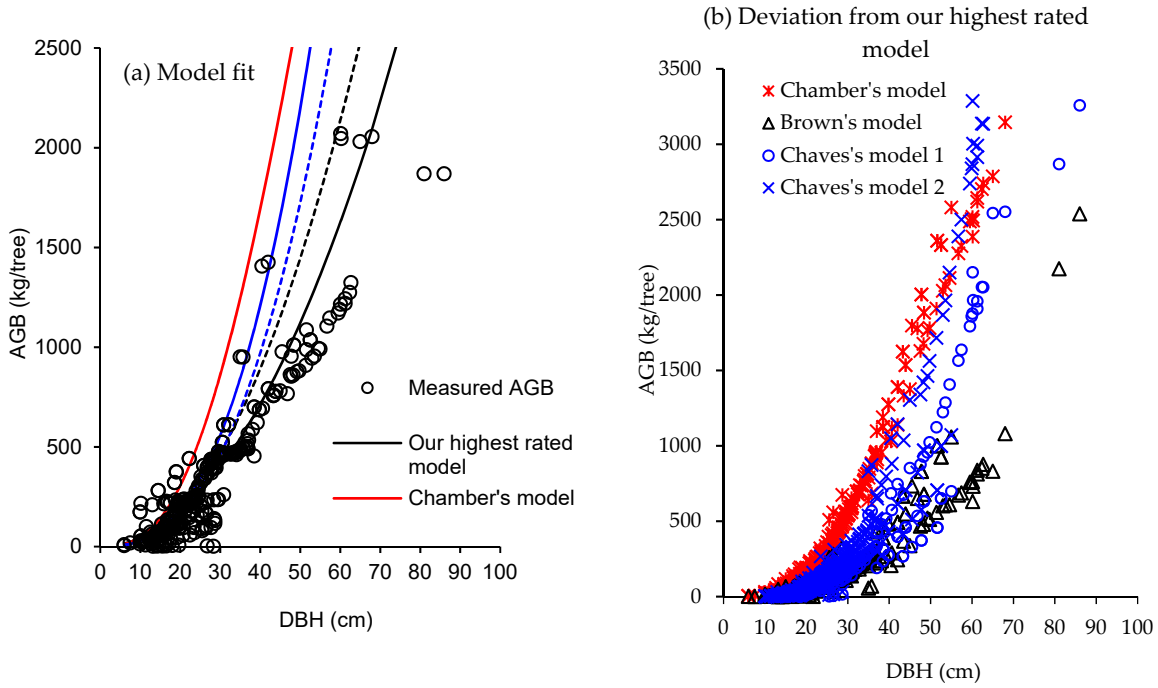

Figure 4. Comparison of models using (**a**) fitted values and (**b**) departures from our highest rated model across values of tree diameter at breast height (DBH).

4. Discussion

Our analysis indicated that models 1–4 fitted using the LMM approach is superior to those fitted using the conventional method. This is probably because mixed effects models can account for the clustered or nested structure of data in forestry. Interestingly, the discrepancies in predictions between the conventional and the LMM models were large in tropical wet evergreen forest.

Our analysis also indicated that the models with wood specific gravity (Model 3 and 4) fitted the data better than models without it although wood specific gravity was obtained from the literature. This is in disagreement with analyses by Stegen [42] that shows inconsistent relationship between forest biomass and wood specific gravity. Although recent analysis has also shown that wood specific gravity has a very weak contribution to AGB [43,44], we believe its inclusion can improve prediction. The use of fewer explanatory variables has been recommended for ease in model application and validation [17]. In the present study D and H were directly measured and could be included in BEMs. In the study area model 3 was found to be more appropriate than Model 1, 2 and 4. Although the cross validation shows that Model 4 is slightly better than Model 3, parameters a and c of Model 4 were not significantly different from zero. Therefore, Model 4 cannot be reliably used for predictive purposes. The limitation of our highest rated model is that it requires H and δ data, which are often not available in many situations. In situations where measured height (H) and wood density (δ) are lacking, second highest rated model, i.e., $AGB_{est} = 0.18D^{2.16} \times 1.32$, may be used for biomass estimation in NEI.

Although species-specific BEMs have often been applied across multiple sites, they are not necessarily applicable to other species, especially those with differing wood densities [16] and plant functional types [45]. The generic models [16,19,30,31] currently in use in NEI severely overestimated the biomass stock when applied to our dataset. This suggests that our model is better suited for AGB estimation for diverse forest types in NEI. The differences in biomass stock estimation between our highest rated model and the generic models may be attributed to differences in the forest types. The generic models were widely used because of the high goodness of fit to the sample data used to develop them. Models usually have a grossly inflated performance in-sample compared to their performance in follow-up studies. This is called the winner's curse [46] in statistics literature. It must be noted that a good model fit does not necessarily translate into good predictions of AGB at the

landscape level or outside the study area [47]. The ability of a model to describe the data at hand (in sample fit) is sometimes confused with predictive power (out-of-sample fit). The value of cross validation is to avoid bias in predictions of biomass [45]. Cross validation also provides a better method for model assessment as it estimates how accurately a model will perform when applied to an independent dataset. It will also curtail problems such as over-fitting [17]. Therefore, we caution against overdependence on generic models without validating their suitability in new areas.

5. Conclusions

Availability of appropriate biomass models for species-diverse forest ecosystems was a major constraint in the study of C balance in NEI. The models developed in the present study can be applied for multi-site and multi-species evaluation of stand biomass and C assessment at local and regional scale. Such models can also find application in studies of plant community and forest structure, and remote sensing methods. Future studies may incorporate harvest data from Alpine zone with larger data set for country scale estimation and validation. We also recommend future studies to address country scale estimation of forest biomass covering different forest types.

Author Contributions: A.J.N., B.K.T. formulated the research work. G.W.S., A.J.N., analyzed the data and wrote the first draft. B.B., A.K.D., U.K.S., S.D., N.B.D., D.R., S.S.C., O.P.T., D.J.D., A.G. shared their data set for model development. All authors jointly discussed the results, drew conclusions and finalized the manuscript. All authors read and approved the final manuscript.

Funding: Authors acknowledges Department of Science and Technology (DST), GOI (DST/IS-STAC/CO_2-SR-224/14(c)-AICP-AFOLU-1), for funding this work.

Acknowledgments: We acknowledge the contribution made by the Junior Research Fellows engaged under the DST AICP-NEH Project towards generating the field data that were used in this paper.

Conflicts of Interest: The authors declare that they have no competing interests.

Availability of Data and Materials: Data will be made available on completion of this on-going research project (completion date March 2019) from the corresponding author on reasonable request.

Abbreviations

Adj R^2	adjusted coefficient of determination
AGB	above ground biomass
AIC	akaike information criterion
BIC	bayesian information criterion
BEMs	biomass estimation models
CF	correction factor
COP	conference of the Parties
C	carbon
EBLUPs	empirical best linear unbiased predictors
CONV	conventional
H-D	height-diameter
PRESS	prediction residual error sum of square
LMM	linear mixed modeling
R^2	coefficient of determination
REDD	Reducing Emissions from Deforestation and Forest Degradation
RMSE	root mean square of error
MSE	mean square error
NEI	North East India

Appendix A

Table A1. PROC MIXED codes for estimating parameters of the conventional models.

```
/*Code for fitting Model 1*/
Proc mixed data = Biomass method = REMLcovtest ratio ic;
Class Forest;
Model lnAGB = lnD/solution outp = check cl;
Run;
```

```
/*Code for fitting Model 2*/
Proc mixed data = Biomass method = REMLcovtest ratio ic;
Class Forest;
Model lnAGB = lnDDH/solution outp = check cl;
Run;
```

```
/*Code for fitting Model 3*/
Proc mixed data = Biomass method = REMLcovtest ratio ic;
Class Forest;
Model lnAGB = lnWDDH/solution outp = check cl;
Run;
```

```
/*Code for fitting Model 4*/
Proc mixed data = Biomass method = REMLcovtest ratio ic;
Class Forest;
Model lnAGB = lnDlnHlnW/solution outp = check cl;
Run;
```

Table A2. PROC MIXED codes for estimating parameters of the LMM models.

```
/*Code for fitting Model 1*/
Proc mixed data = Biomass method = REMLcovtest ratio ic;
Class Forest;
Model lnAGB = lnD/solution outp = check cl;
Random Forest;
ESTIMATE "1" intercept 1 | Forest 1;
ESTIMATE "2" intercept 1 | Forest 0 1;
ESTIMATE "3" intercept 1 | Forest 0 0 1;
ESTIMATE "4" intercept 1 | Forest 0 0 0 1;
Run;
```

```
/*Code for fitting Model 2*/
Proc mixed data = Biomass method = REMLcovtest ratio ic;
Class Forest;
Model lnAGB = lnDDH/solution outp = check cl;
Random Forest;
ESTIMATE "1" intercept 1 | Forest 1;
ESTIMATE "2" intercept 1 | Forest 0 1;
ESTIMATE "3" intercept 1 | Forest 0 0 1;
ESTIMATE "4" intercept 1 | Forest 0 0 0 1;
Run;
```

```
/*Code for fitting Model 3*/
Proc mixed data = Biomass method = REMLcovtest ratio ic;
Class Forest;
Model lnAGB = lnWDDH/solution outp = check cl;
Random Forest;
ESTIMATE "1" intercept 1 | Forest 1;
ESTIMATE "2" intercept 1 | Forest 0 1;
ESTIMATE "3" intercept 1 | Forest 0 0 1;
ESTIMATE "4" intercept 1 | Forest 0 0 0 1;
Run;
```

```
/*Code for fitting Model 4*/
Proc mixed data = Biomass method = REMLcovtest ratio ic;
Class Forest;
Model lnAGB = lnDlnHlnW/solution outp = check cl;
Random Forest; ESTIMATE "1" intercept 1 | Forest 1;
ESTIMATE "2" intercept 1 | Forest 0 1;
ESTIMATE "3" intercept 1 | Forest 0 0 1;
ESTIMATE "4" intercept 1 | Forest 0 0 0 1;
Run;
```

Table A3. R scripts for the four models.

Model1<-lm(lnAGB ~ lnD, data = Biomass_data) Model2<-lm(lnAGB ~ lnDDH, data = Biomass_data) Model3<-lm(lnAGB ~ lnWDDH, data = Biomass_data) Model4<-lm(lnAGB ~ lnD + lnH + lnW, data = Biomass_data)
The R script for cross validation using 10-fold using the lava packages: cv(list(Model1, Model2, Model3, Model4), k = 10, data = Biomass_data)
Various goodness of fit indices were calculated using the following codes using the "forecast" package of R: Model1_results < -t(data.frame(CV(Model1))) Model2_results < -t(data.frame(CV(Model2))) Model3_results < -t(data.frame(CV(Model3))) Model4_results < -t(data.frame(CV(Model4))) Model_results < -rbind(Model1_results, Model2_results, Model3_results, Model4_results)

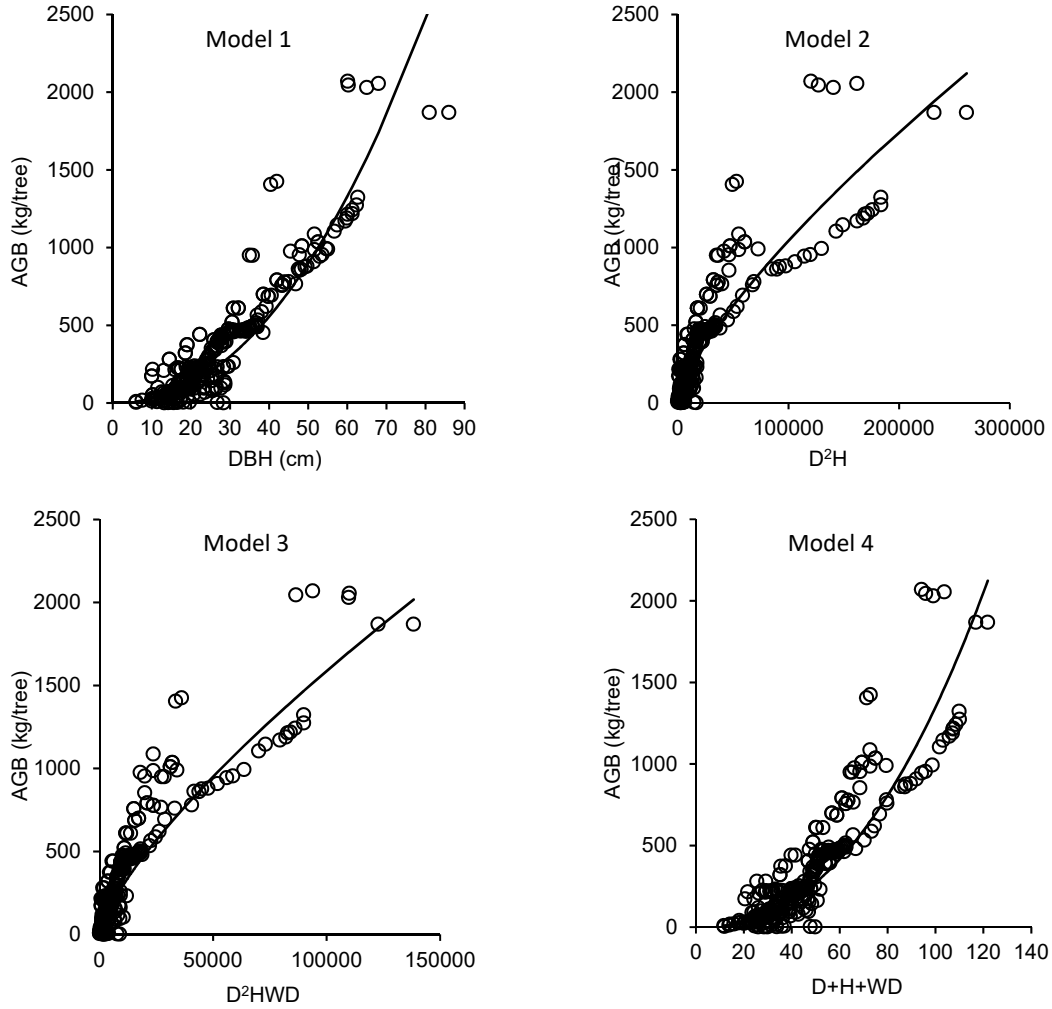

Figure A1. The observed AGB (open circles) and fitted values (smooth lines) using models 1–4 in the arithmetic domain.

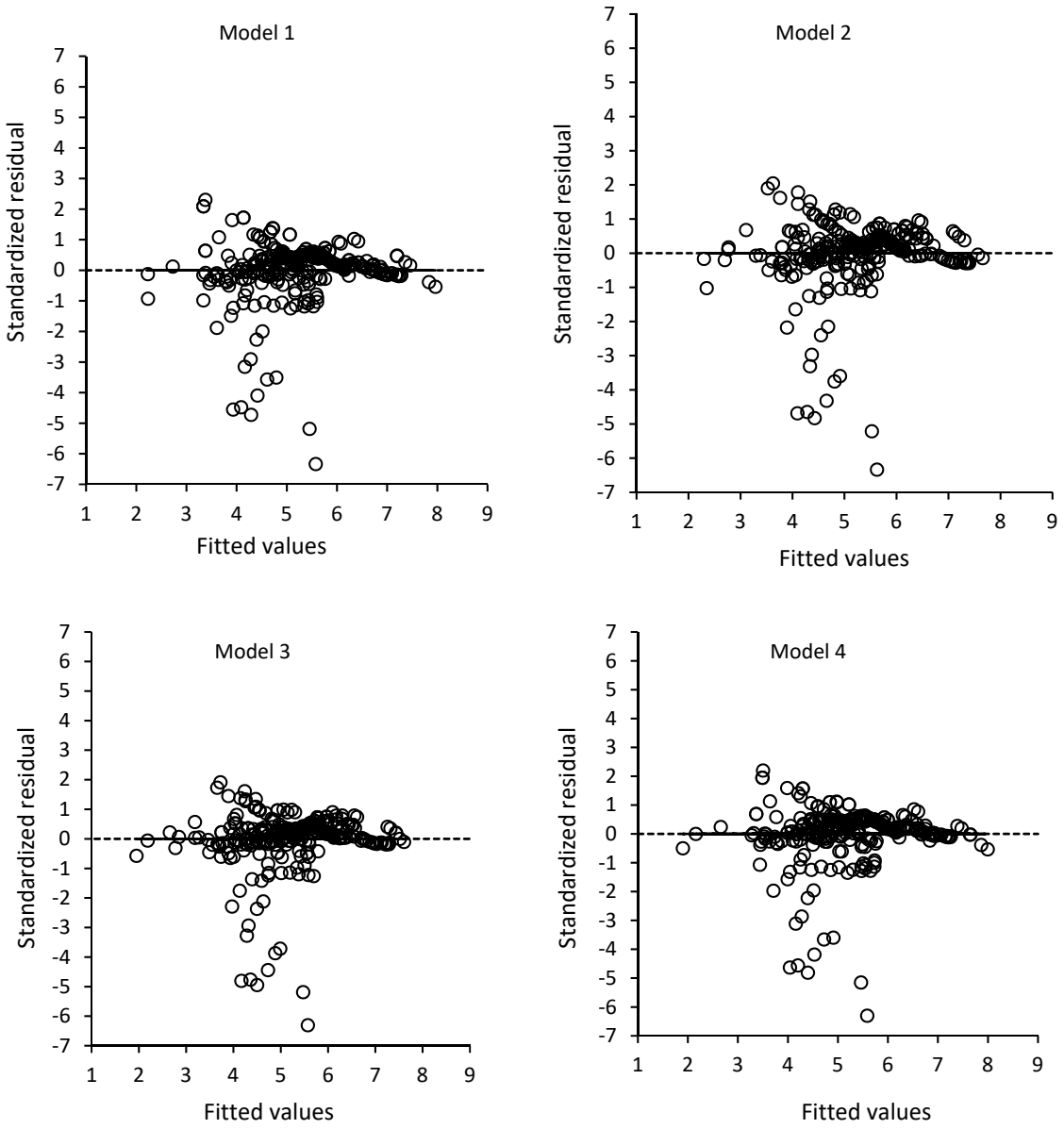

Figure A2. Plots of standardized residuals against the fitted values on the logarithmic scale

References

1. Bloom, A.; Exbrayat, J.; van der Velde, I.; Feng, L.; Williams, M. The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1285–1290. [CrossRef]
2. Sullivan, M.J.P.; Talbot, J.; Lewis, S.L.; Phillips, O.L.; Qie, L.; Begne, S.K.; Chave, J.; Cuni-Sanchez, A.; Hubau, W.; Lopez-Gonzalez, G.; et al. Diversity and carbon storage across the tropical forest biome. *Sci. Rep.* **2017**, *7*, 39102. [CrossRef]
3. Hosonuma, N.; Herold, M.; De Sy, V.; De Fries, R.S.; Brockhaus, M.; Verchot, L.; Angelsen, A.; Romijn, E. An assessment of deforestation and forest degradation drivers in developing countries. *Environ. Res. Lett.* **2012**, *7*, 044009. [CrossRef]
4. UNFCCC. Modalities for National Forest Monitoring Systems. 2013. Available online: http://unfccc.int/files/meetings/warsaw_nov_2013/decisions/application/pdf/cop19_fms.pdf (accessed on 10 October 2018).
5. Harris, N.L.; Brown, S.; Hagen, S.C.; Saatchi, S.S.; Petrova, S.; Salas, W.; Hansen, M.C.; Potapov, P.V.; Lotsch, A. Baseline map of carbon emissions from deforestation in tropical regions. *Science* **2012**, *336*, 1573. [CrossRef]

6. Mitchard, E.T.A.; Saatchi, S.S.; Baccini, A.; Asner, G.P.; Goetz, S.J.; Harris, N.L.; Brown, S. Uncertainty in the spatial distribution of tropical forest biomass: A comparison of pan-tropical maps. *Carbon Balance Manag.* **2013**, *8*, 10. [CrossRef] [PubMed]
7. Woodhouse, I.H.; Mitchard, E.T.A.; Brolly, M.; Maniatis, D.; Ryan, C.M. Radar backscatter is not a 'direct measure' of forest biomass. *Nat. Clim. Chang.* **2012**, *2*, 556–557. [CrossRef]
8. Baccini, A.; Goetz, S.J.; Walker, W.S.; Laporte, N.T.; Sun, M.; Sulla-Menashe, D.; Hackler, J.; Beck, P.S.A.; Dubayah, R.; Fried, M.A.; et al. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nat. Clim. Chang.* **2012**, *2*, 182–185. [CrossRef]
9. Zarin, D.J. Carbon from tropical deforestation. *Science* **2012**, *336*, 1518–1519. [CrossRef]
10. Canadell, J.; Raupach, M.R. Managing forests for climate change mitigation. *Science* **2008**, *320*, 1456. [CrossRef]
11. Pelletier, J.; Ramankutty, N.; Potvin, C. Diagnosing the uncertainty and detectability of emission reduction for REDD+ under current capabilities: An example for Panama. *Environ. Res. Lett.* **2011**, *6*, 024005. [CrossRef]
12. Poulter, B.; Hattermann, F.; Hawkins, E.; Zaehle, S.; Sitch, S.; Restrepo-Coupe, N.; Heyder, U.; Cramer, W. Robust dynamics of Amazon dieback to climate change with perturbed ecosystem model parameters. *Glob. Chang. Biol.* **2010**, *16*, 2476–2495. [CrossRef]
13. Goodman, R.C.; Phillips, O.L.; Baker, T.R. The importance of crown dimensions to improve tropical tree biomass estimates. *Ecol. Appl.* **2014**, *24*, 680–698. [CrossRef]
14. Xiao, C.W.; Ceulemans, R. Allometric relationships for below and aboveground biomass of young Scots pines. *Fore. Ecol. Manag.* **2004**, *203*, 177–186. [CrossRef]
15. Montes, N.; Gauquelin, T.; Badri, W.; Bertaudiere, V.; Zaoui, E.H. A non-destructive method for estimating above-ground forest biomass in threatened woodlands. *Fore. Ecol. Manag.* **2000**, *130*, 37–46. [CrossRef]
16. Chave, J.; Andalo, C.; Brown, S.; Cairns, M.A.; Chambers, J.Q.; Eamus, D.; Fölster, H.; Fromard, F.; Higuchi, N.; Kira, T.; et al. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* **2005**, *145*, 87–99. [CrossRef] [PubMed]
17. Sileshi, G.W. A critical review of forest biomass estimation models, common mistakes and corrective measures. *Fore. Ecol. Manag.* **2014**, *329*, 237–254. [CrossRef]
18. Vashum, K.T.; Jayakumar, S. Methods to estimate aboveground biomass and carbon stock in natural forests—A review. *J. Ecosyst. Ecogr.* **2012**, *2*, 1000116. [CrossRef]
19. Chave, J.; Réjou-Méchain, M.E.; Búrquez, A.; Chidumayo, E.; Colgan, M.S.; Delitti, W.B.; Duque, A.; Eid, T.; Fearnside, P.M.; Goodman, R.C.; et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Chang. Biol.* **2014**, *20*, 3177–3190. [CrossRef]
20. Litton, C.M.; Kauffman, J.B. Allometric models for predicting above-ground biome two widespread woody plants in Hawaii. *Biotropica* **2008**, *40*, 313–320. [CrossRef]
21. Vahedi, A.A.; Mataji, A.; Hodjati, S.M.; Djomo, A. Allometric equations for predicting aboveground biomass of beech hornbeam stands in the Hyrcanian forests of Iran. *J. For. Sci.* **2014**, *60*, 236–247. [CrossRef]
22. *India State of Forest Report 2017*; Ministry of Environment, Forests and Climate Change, Govt of India: New Delhi, India, 2017.
23. Baishya, R.; Barik, S.K. Estimation of tree biomass, carbon pool and net primary production of an old-growth *Pinus kesiya* Royle ex. Gordon forest in north-eastern India. *Ann. For. Sci.* **2011**, *68*, 727–736. [CrossRef]
24. Brahma, B.; Sileshi, G.W.; Nath, A.J.; Das, A.K. Development and evaluation of robust tree biomass equations for rubber tree (*Hevea brasiliensis*) plantations in India. *For. Ecosyst.* **2017**, *4*, 14. [CrossRef]
25. Nath, S.; Nath, A.J.; Sileshi, G.W.; Das, A.K. Biomass stocks and carbon storage in *Barringtonia acutangula* floodplain forests in North East India. *Biomass Bioenergy* **2017**, *98*, 37–42. [CrossRef]
26. Baishya, R.; Barik, S.K.; Upadhaya, K. Distribution pattern of aboveground biomass in natural and plantation forests of humid tropics in northeast India. *Trop. Ecol.* **2009**, *50*, 295–304.
27. Borah, N.; Nath, A.J.; Das, A.K. Aboveground biomass and carbon stocks of tree species in tropical forests of Cachar district, Assam North East India. *Int. J. Ecol. Environ. Sci.* **2013**, *39*, 97–106.
28. Borah, M.; Das, D.; Kalita, J.; Prasanna, H.; Borua, D.; Phukan, B.; Neog, B. Tree species composition, biomass and carbon stocks in two tropical forest of Assam. *Biomass Bioenergy* **2015**, *78*, 25–35. [CrossRef]
29. Waikhom, A.C.; Nath, A.J.; Yadava, P.S. Aboveground biomass and carbon stock in the largest sacred grove of Manipur, North East India. *J. For. Res.* **2018**, *29*, 425–428. [CrossRef]

30. Brown, S.; Gillespie, A.J.R.; Lugo, A.E. Biomass estimation methods for tropical forests with application to forestry inventory data. *For Sci.* **1989**, *35*, 881–902.
31. Chambers, J.Q.; Santos, J.D.; Ribeiro, R.J.; Higuchi, N. Tree damage, allometric relationships, and above-ground net primary production in central Amazon forest. *Fore. Ecol. Manag.* **2001**, *152*, 73–84. [CrossRef]
32. Roy, P.S.; Kushwaha, S.P.S.; Murthy, M.S.R.; Roy, A.; Kushwaha, D.; Reddy, C.S.; Behera, M.D.; Mathur, V.B.; Padalia, H.; Saran, S. *Biodiversity Characterisation at Landscape Level: National Assessment*; Indian Institute of Remote Sensing: Dehradun, India, 2012.
33. Mittermeier, R.A.; Robles-Gil, P.; Hoffmann, M.; Pilgrim, J.; Brooks, T.; Mittermeier, C.G.; Lamoreux, J.; Da Fonseca, G.A.B. *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*; CEMEX: Mexico City, Mexico, 2004.
34. Champion, H.G.; Seth, S.K. *A Revised Survey of the Forest Types of India*; Natraj Publishers: Dehradun, India, 1968; p. 404, (Reprinted 2005).
35. Poffenberger, M.; Barik, S.K.; Choudhury, D.; Darlong, V.; Gupta, V.; Palit, S.; Upadhyay, S. *Forest Sector Review of Northeast India*; Community Forestry International: Santa Barbara, CA, USA, 2006.
36. Chen, D.; Huang, X.; Zhang, S.; Sun, X. Biomass modeling of larch (*Larix* spp.) plantations in China based on the mixed model, dummy variable model, and Bayesian hierarchical model. *Forests* **2017**, *8*, 268. [CrossRef]
37. Jiang, J. Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. *Stat. Sin.* **1998**, *8*, 861–885.
38. Chave, J.; Coomes, D.A.; Jansen, S.; Lewis, S.L.; Swenson, N.G.; Zanne, A.E. Towards a worldwide wood economics spectrum. *Ecol. Lett.* **2009**, *12*, 351–366. [CrossRef]
39. Sileshi, G.W. The fallacy of retification and misinterpretation of the allometry exponent. 2015. Available online: https://www.researchgate.net/profile/Gudeta_Sileshi2/publication/281204953_The_fallacy_of_reification_and_misinterpretation_of_the_allometry_exponent/links/55dc322808aec156b9b008a2/The-fallacy-of-reification-and-misinterpretation-of-the-allometry-exponent.pdf (accessed on 10 October 2018). [CrossRef]
40. Nakagawa, S.; Schielzeth, H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **2013**, *4*, 133–142. [CrossRef]
41. Das, K.; Jiang, J.; Rao, J.N.K. Mean squared error of empirical predictor. *Ann. Stat.* **2004**, *32*, 818–840.
42. Stegen, J.C.; Swenson, N.G.; Valencia, R.; Enquist, B.J.; Thompson, J. Above-ground forest biomass is not consistently related to wood density in tropical forests. *Glob. Ecol. Biogeogr.* **2009**, *18*, 617–625. [CrossRef]
43. Molto, Q.; Rossi, V.; Blanc, L. Error propagation in biomass estimation in tropical forests. *Methods Ecol. Evol.* **2013**, *4*, 175–183. [CrossRef]
44. Lima, A.J.N.; Suwa, R.; de Mello Ribeiro, G.H.P.; Kajimoto, T.; dos Santos, J.; da Silva, R.P.; Souza, C.A.S.; Barros, P.C.; Noguchi, H.; Ishizuka, M.; et al. Allometric models for estimating above- and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil. *For. Ecol. Manag.* **2012**, *277*, 163–172. [CrossRef]
45. Paul, K.I.; Radtke, P.J.; Roxburgh, S.H.; Larmour, J.; Waterworth, R.; Butler, D.; Brooksbank, K.; Ximenes, F.; et al. Validation of allometric biomass models: How to have confidence in the application of existing models. *For. Ecol. Manag.* **2016**, *412*, 70–79. [CrossRef]
46. Thaler, R. *The Winner's Curse*; Free Press: New York, NY, USA, 1992.
47. Marra, D.M.; Higuchi, N.; Trumbore, S.E.; Ribeiro, G.H.P.M.; dos Santos, J.; Carneiro, V.M.C.; Lima, A.J.N.; Chambers, J.Q.; Negrón-Juárez, R.I.; Holzwarth, F.; et al. Predicting biomass of hyperdiverse and structurally complex Central Amazon forests—A virtual approach using extensive field data. *Biogeosci. Discuss.* **2015**, *12*, 15537–15581. [CrossRef]