UN-REDD
PROGRAMME

PART B-5:
Tree allometric equations in Evergreen broadleaf, Deciduous, and Bamboo forests in the South East region, Viet Nam

UN-REDD PROGRAMME
Viet Nam

October 2012

Hanoi, Viet Nam

Tree allometric equation development for estimation of forest above-ground biomass in Viet Nam -- Evergreen broadleaf, Deciduous, and Bamboo forests in the South East region

By

Dr. Nguyen Dinh Hung, Division of Sciences, Techniques and Technologies

Msc. Nguyen Van Bay, The Center for Forest Information and Consultancy

Eng. Nguyen Dinh Binh, The Center for Forest Information and Consultancy

Msc. Nguyen Cao Tung, The Center for Forest Information and Consultancy

Forest Inventory and Planning Institute

Edited by:

Akiko Inoguchi, Gael Sola, Matieu Henry and Luca Birigazzi, FAO

Recommended citation: Hung, N.D., Bay, N.V., Binh, N.D., Tung, N.C. (2012) Tree allometric equations in Evergreen broadleaf, Deciduous, and Bamboo forests in the South East region, Vietnam, in (Eds) Inoguchi, A., Henry, M. Birigazzi, L. Sola, G. Tree allometric equation development for estimation of forest above-ground biomass in Viet Nam, UN-REDD Programme, Hanoi, Viet Nam.

ACKNOWLEDGEMENT

We deeply acknowledge the UN-REDD Vietnam Programme who has provided budget and technical assistant for this Study. Special thanks are given to Dr. L. Saint-André (CIRAD-INRA) and Dr. M. Henry (FAO) for their useful lectures and teaching materials from the workshop on development of allometric equation for biomass estimation. We are very grateful to Mrs. Akiko Inoguchi (FAO-VN) and Dr. Pham Manh Cuong (VN-FOREST and VRO) for their frequent advice and regular supervision.

The authors would like to thank and all staff of Department of Agriculture and Rural Development of Binh Thuan province, especially Mr. Mai Kieu - the Deputy Director and other staff of Department of Forestry, Forest Protection Department, and Phan Dung Forest Protection Management Board for their help and cooperation in the implementation process of this project.

We also want to thank colleagues from Vietnam REDD Office, Vietnam Forestry University, Research Centre for Forest Ecology and Environment, Tay Nguyen University and FAO who have contributed valuable comments for improving this report.

EXECUTIVE SUMMARY

This report describes the process of developing biomass allometric equations and biomass conversion and expansion factors for biomass estimation of the evergreen broadleaf, deciduous and bamboo forests in the South East Region of Vietnam. Destructive sampling was done to collect biomass data of sample trees and use these data as dependent variables in the multiple regression analyses. Equations were developed using various different statistical models and regression approaches and then compared. For equations developed using the least squares approach, the adjusted R^2 was used for comparison. For equations developed using the maximum likelihood approach, the Akaike Information Criterion with correction (AICc) was used as for comparison. Cross validation tests were conducted to assess the errors of prediction and compare the equations across different regression approaches. For woody forests, the best chosen equations were compared with previouly published equations, including those of Basuki *et al.* (2009), Brown (1997) and Chave *et al.* (2005).

For evergreen broadleaf forest, analyzed results of nine statistical models using three regression approaches have lead to the recommendation of using the following four equations, which are the best for each group of input variables:

Equation[1]	Expected value of error[2] (%)	Range of error[3] (95% CL)
$AGB = 0.1277 \times D^{2.3943}$	0.909	-12.78% ÷ 16.76%
$AGB = 0.0530 \times (D^2H^{0.7})^{1.0072}$	-0.467	-13.54% ÷ 14.36%
$AGB = 0.2328 \times (D^{2.4}\rho)^{0.9933}$	0.679	-10.05% ÷ 12.38%
$AGB = 0.0968 \times (D^2H^{0.7}\rho)^{10037}$	-0.666	-10.81% ÷ 10.28%

[1] AGB is the above-ground biomass in kg; D is the diameter at breast height in cm; H is the height in m; and ρ is the wood density in g/cm^3 of the tree.

[2] The error here means the error (in percentage) of the predicted total AGB as compared to the measured total AGB of a set of trees.

[3] These ranges of error apply when predicting the total AGB for datasets of 37 or more trees. For datasets with smaller number of trees, the ranges of error may be larger.

The results also indicated that the inclusion of height and wood density as additional input variables contributes to the improvement of prediction. Therefore, whenever these variables are available, equations that use them should be applied. Moreover, the inclusion of wood density improves the robustness of prediction much more than the inclusion of height so wood density should be given the first priority when considering additional variables. The comparison with previously published equations has shown that all three previously published equations tend to overestimate the AGB of trees in the dataset in this study. The total AGB errors of the equations of Basuki et al. (2009), Brown (1997) and Chave et al. (2005) for the current dataset are 11.2%, 57.2% and 35.8%, respectively. This indicates that countries need to develop their own specific allometric equations in order to improve the certainty of biomass prediction and carbon stock assessment.

An attempt was also made to estimate BCEF and BEF for evergreen broadleaf forests. The results show that BCEF and BEF do not depend on DBH but vary around a constant, which is 0.715 for BCEF and 1.256 for BEF.

For deciduous forest, analyzed results of nine statistical models using three regression approaches have lead to the recommendation of using the four following equations, which are the best for each group of input variables:

Equation	Expected value of error (%)	Range of error[1] (95% CL)
$AGB = 0.0670 \times D^{2.5915}$	-1.082	-12.92% ÷ 13.33%
$AGB = 0.0154 \times (D^2H^{0.7})^{1.1682}$	-0.913	-11.68% ÷ 11.82%
$AGB = 0.0560 \times (D^{2.4}\rho)^{1.1655}$	-0.681	-11.46% ÷ 11.84%
$AGB = 0.0159 \times (D^2H^{0.7}\rho)^{1.2275}$	-1.850	-10.79% ÷ 7.34%

[1]These ranges of error apply when predicting the total AGB for datasets of 19 or more trees. For datasets with smaller number of trees, the ranges of error may be larger.

Both the inclusions of height and wood density improve the biomass prediction. Unlike evergreen broadleaf forests, the role of wood density in biomass prediction of deciduous forests is less important. This is because the number of species in deciduous forests is quite small and therefore the variation in wood density is smaller than in evergreen broadleaf forests. Comparison with published equations has shown that while the equations of Basuki and Brown overestimate the total AGB of the current dataset by, respectively, 6.7% and 39.7%, the equation of Chave *et al.* adapts very well (the total AGB error is 1.7%) with the current dataset.

Similar to evergreen broadleaf forests, the calculation of BCEF and BEF for deciduous forest shows that they do not depend on DBH but vary around a constant, which is 0.834 for BCEF and 1.396 for BEF.

For bamboo forests, analyzed results of four statistical models using three regression approaches have lead to the recommendation of using the following two equations:

Equation	Expected value of error (%)	Range of error[1] (95% CL)
$AGB = 0.1006 \times D^{2.2220}$	0.327	-6.76% ÷ 7.84%
$AGB = 0.0644 \times D^{1.9696}H^{0.3426}$	0.265	-6.66% ÷ 7.61%

[1]These ranges of error apply when predicting the total AGB for datasets of 40 or more trees. For datasets with smaller number of trees, the ranges of error may be larger.

The results also show that the inclusion of H only slightly improves the accuracy as well as the robustness of the prediction. Because heights of standing bamboos are quite difficult to measure accurately, it is recommended that for bamboo forests, it is not necessary to include the variable H in biomass prediction.

Age-class specific equations were also developed for bamboos. The analyzed results show that although age-class specific equations help to improve the robustness of biomass prediction, the accuracy is degraded. Therefore, it is suggested that the general equations developed for all age classes should be used for bamboo forests with a balanced proportion of bamboos in each age class (i.e., each age class accounts for about one third of the bamboos) and age-class specific equations should be used otherwise.

In order to improve the certainty of biomass prediction in the studied region, the next studies should concentrate on the development of equations and BCEFs specified to each tree family or wood density class. Since the ranges of error of the best models for evergreen broadleaf and deciduous forests are still quite large (±10%), destructive sampling of more sample trees is also recommended.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION AND ACRONYMS

AD	Activity Data
AFOLU	Agriculture, Forestry and Other Land Use
AGB	Above-Ground Biomass
AIC	Akaike Information Criterion
AICc	Akaike Information Criterion with correction
BCEF	Biomass Conversion and Expansion Factor
BEF	Biomass Expansion Factor
BGB	Below-Ground Biomass
CFIC	Centre for Forest Information and Consultancy
CL	Confidence Limits
COP	Conference of the Parties
CV	Coefficient of Variation
DBH	Diameter at Breast Height
EF	Emission Factors
Est.	Estimate
FAO	Food and Agriculture Organization of the United Nations
FIPI	Forest Inventory and Planning Institute
FSIV	Forest Science Institute of Vietnam
GHG-I	Green House Gas -Inventory
IPCC	Inter-governmental Panel on Climate Change
LULUCF	Land Use, Land Use Change and Forestry
MRV	Measurement, Reporting and Verification
QA/QC	Quality Assessment/Quality Control
SSE	Sum of Squared Errors
TNU	Tay Nguyen University
UNFCCC	United Nations Framework Convention on Climate Change
UN-REDD	United Nations Collaborative Program on Reducing Emissions from Deforestation and Forest Degradation in Developing Countries
VFU	Vietnam Forestry University
VRO	Vietnam REDD Office
WD	Wood Density
IV%	Importance Value (%)

1 INTRODUCTION

The objective of this study is to:

- conduct forest biomass field measurements for two plots in evergreen broadleaf forests, one plot in deciduous forest plot and one bamboo forest plot in Binh Thuan province of the South East Region;
- carry out development of allometric equations for biomass estimation in the South East Region through application of collected data in regression analysis, using different variables, to develop allometric equations for estimation of forest biomass; and
- carry out error assessment for their developed allometric equations using the independent data from sample trees collected for error assessment.

This report describes the implementation process and results of the second phase of the Study conducted by CFIC. The report is organized as follows: Section 2 describes materials and methods used in the Study. Section 3 gives a description of the surveyed areas. Sections 4, 5 and 6 present the results together with discussion for each espective forest type of evergreen broadleaf, deciduous, and bamboo forests. Finally, conclusions and recommendations are provided in Section 7

2 MATERIALS AND METHODS

2.1 Sampling strategy

2.1.1 Woody Forests (Evergreen and Deciduous Forests)

Location and design of the plot

Criteria for sample plot establishment in woody forests

The establishment of sample plots were conducted to meet the following criteria: i) representativeness (based on assessment of experts) of the forest types being studied; ii) representativeness for topographic conditions; and iii) covering a number of different tree sizes; iv) the sample plots should be set up on less disturbed forests where large sized trees are available (preferably in rich forests, and as a minimum in medium (quality) forests[1]).

The area of each sample plot is 1 ha. The plot is a square of 100 m x 100 m. In steep areas (slope gradient larger than 20°), four sub-sample plots of 0.25 ha (50 m x 50 m) each were used instead.

Description of sample plot BT-PD-01

The information of the sample plot BT-PD-01 is given in Table 1 and its position on satellite image is given in Figure 1 below.

Table 1: Description of sample plot BT-PD-01

Plot name:	**BT-PD-01**
Administrative location:	Compartment no. 7 - Phan Dũng commune - Tuy Phong district - Bình Thuận province
Owner/manager:	Tuy Phong Protection Forest Management Board
Coordinate (VN2000 projection):	Long = 108° 38' 46" E; Lat = 11° 30' 16" N
Altitude:	340 m
Slope:	15°
Plot area:	1 ha
Plot size:	100 x 100 (m)
Forest type:	Evergreen broadleaf forest (forest with ≥ 75% of broadleaf tree species and green all around the year)
Forest status:	IIIB (forest that has been affected at medium-level; the structure of trees with DBH ≥ 40 cm has been changed)
Volume (estimated):	300 m^3/ha

[1] According to Circular 34/TT-BNN issued by MARD, a rich forest is a forest with a standing wood volume of 201 – 300 m^3/ha and that of medium forest is 101 – 200m^3/ha.

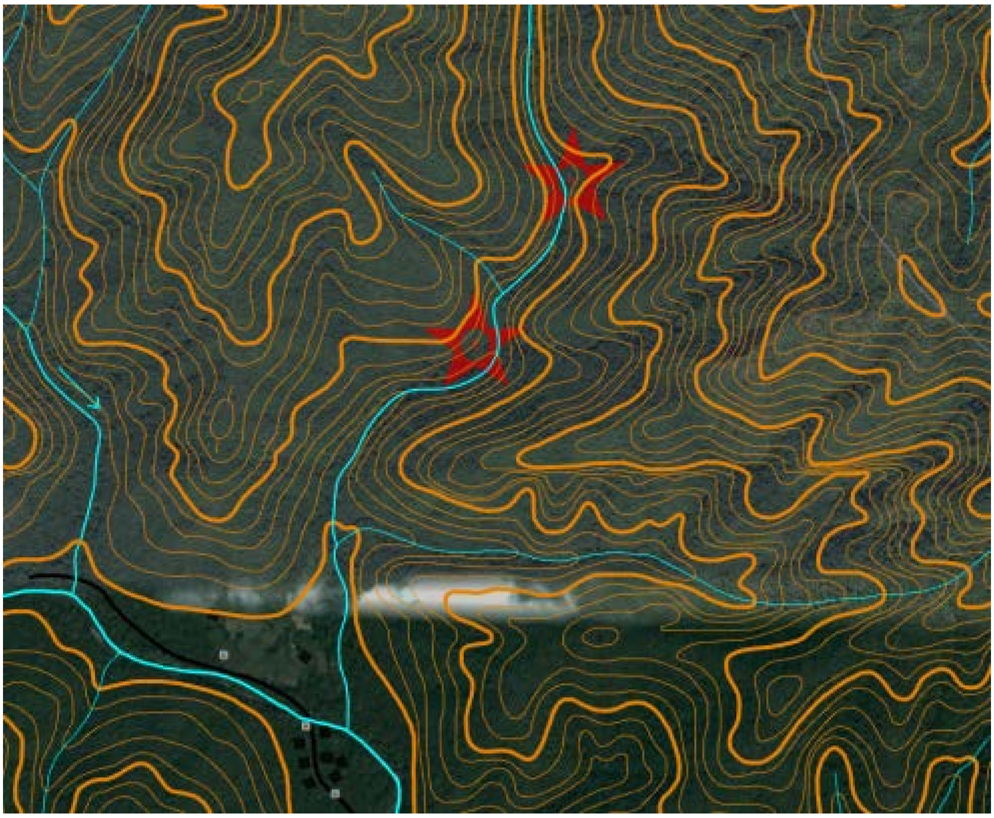


Figure 1: Position of sample plots BT-PD-01 and BT-PD-04 on satellite image

Description of sample plot BT-PD-02

Table 2: Description of sample plot BT-PD-02

Plot name:	**BT-PD-02**
Administrative location:	Compartment no. 9 - Phan Dũng commune - Tuy Phong district - Bình Thuận province
Owner/manager:	Tuy Phong Protection Forest Management Board
Coordinate (VN2000 projection):	Long = 108° 37' 46" E; Lat = 11° 29' 45" N
Altitude:	320 m
Slope:	15°
Plot area:	1 ha
Plot size:	100 x 100 (m)
Forest type:	Evergreen broadleaf forest
Forest status:	IIIA3 (forest that has been affected at medium-level; the structure of trees with DBH ≥ 40 cm has been changed)
Volume (estimated):	260 m^3/ha

Figure 2: Positions of sample plot BT-PD-02 on satellite image

Description of sample plot BT-PD-03

Table 3: Description of sample plot BT-PD-03

Plot name:	**BT-PD-03**
Administrative location:	Compartment no. 16 - Phan Dũng commune - Tuy Phong district - Bình Thuận province
Owner/manager:	Tuy Phong Protection Forest Management Board
Coordinate (VN2000 projection):	Long = 108° 38' 43" E; Lat = 11° 25' 40" N
Altitude:	230 m
Slope:	3°
Plot area:	1 ha
Plot size:	100 x 100 (m)
Forest type:	Deciduous forest (forest with ≥ 75% of seasonal deciduous tree species)
Forest status:	RIIIA3 (deciduous forest that has been affected at medium-level; the structure of trees with DBH ≥ 40 cm has been changed)
Volume (estimated):	160 m^3

Figure 3: Position of sample plot BT-PD-03 on satellite image

Selection of the sampling trees

The selection of the tree is the result of diameter measurement of all the trees within each plot

All the trees in the sample plots were grouped into DBH classes. The interval of DBH classes is 10 cm, and the DBH classes are: 5 – 14.9 cm; 15 – 24.9 cm; 25 – 34.9 cm; 35 – 44.9 cm; 45 – 54.9 cm; 55 – 64.9 cm; 65 – 74.9 cm. The total number of sample trees for harvesting is 55 trees for each plot in each forest type (50 trees for development of allometric equations and 5 trees for validation). The number of sample trees for each DBH class is chosen proportionally with the number of trees in the class, with at least three sample trees harvested for each DBH class. Then the sample trees in each DBH class in the sample plots were randomly selected. Due to time and budget limitation, trees with DBH ≥ 75 cm were not sampled in this study. The numbers of felled trees for each species and each DBH class are given in Table 4 and Table 5, respectively.

Table 4: Number of felled trees divided by species (evergreen broadleaf sample plots)

No	Local Name	Scientific Name	Number of felled trees		
			BT-PD-01	BT-PD-02	Total
1	Bằng lăng	Lagerstroemia calyculata	7	6	13
2	Bằng lăng ổi	Lagerstroemia crispa		3	3
3	Bình linh nghệ	Vitex ajugaeflora	2	2	4
4	Căm xe	Xylia xylocarpa	4	5	9
5	Cẩm lai	Dalbergia oliveri		3	3
6	Cẩm liên	Shorea siamensis		1	1

No	Local Name	Scientific Name	Number of felled trees		
			BT-PD-01	BT-PD-02	Total
7	Cẩm thị	Diospyros maritima	3		3
8	Chiêu liêu	Terminalia chebula	2	4	6
9	Chiêu liêu nghệ	Terminalia triptera	1		1
10	Cò ke	Microcos paniculata	5	1	6
11	Dầu rái	Dipterocarpus alatus	12	4	16
12	Dầu trà beng	Dipterocarpus obtusifolius		3	3
13	Gáo	Adina polycephala	4	1	5
14	Gáo vàng	Adina pilulifra		2	2
15	Giáng hương	Pterocarpus macrocarpus		3	3
16	Ké	Nephelium sp.		1	1
17	Kơ nia	Irvingia malayana	1	2	3
18	Lim đá	Sp1	1	1	2
19	Lim xẹt	Peltophorum pterocarpum		1	1
20	Móng bò	Bauhinia sp.	1		1
21	Săng mây	Antheroporum pierrei	1	1	2
22	Sao đen	Hopea recopei	2		2
23	Sến	Madhuca sp.		1	1
24	Sến mủ	Shorea roxburghii		3	3
25	Sổ	Dillenia scabrella	1		1
26	Tà quang	Sp2	1		1
27	Thành ngạnh	Cratoxylon pruniflorum		1	1
28	Thị nhong	Diospyros sp.		1	1
29	Trâm	Syzygium sp.	1		1
30	Trâm đỏ	Syzygium oblatum	3		3
31	Vậy nước	Cephalanthus tetrandra	1	1	2
32	Vên vên	Anisoptera costata		1	1
33	Vừng	Careya arborea		2	2
34	Xoài rừng	Mangifera minitifolia	2	1	3
Total			55	55	110

Table 5: Number of felled trees divided by DBH class (evergreen broadleaf sample plots)

No	DBH class (cm)	Number of felled trees		
		BT-PD-01	BT-PD-02	Total
1	5 – 14.9	14	15	29

2	15 – 24.9	15	12	27
3	25 – 34.9	9	11	20
4	35 – 44.9	7	6	13
5	45 – 54.9	4	5	9
6	55 – 64.9	3	3	6
7	65 – 74.9	3	3	6
Total		55	55	110

For deciduous forest, a total of 55 sample trees were felled for destructive biomass measurement in the deciduous sample plot. The numbers of felled trees divided by species and DBH classes are given in Table 6 and Table 7, respectively.

Table 6: Numbers of sample trees by species (deciduous sample plot)

No.	Local Name	Scientific Name	N
1	Bồ hòn	Sapindus saponaria	1
2	Căm xe	Xylia xylocarpa	6
3	Cà chắc	Shorea obtusa	3
4	Cà đuối	Cryptocarya petelotii	2
5	Chiêu liêu khế	Terminalia alata	3
6	Dầu lông	Dipterocarpus intricatus	21
7	Gáo vàng	Adina pilulifra	2
8	Kơ nia	Irvingia malayana	6
9	Nhàu	Morinda citrifolia	1
10	Thành ngạnh	Cratoxylon pruniflorum	1
11	Trắc	Dalbergia cochinchinensis	3
12	Trâm	Syzygium sp.	4
13	Vừng	Careya arborea	2
		Total	55

Table 7: Numbers of sample trees by DBH class (deciduous sample plot)

DBH Class	N trees	N sampled
5 – 14.9 cm	398	16
15 – 24.9 cm	220	13
25 – 34.9 cm	74	11
35 – 44.9 cm	32	9
45 – 54.9 cm	8	6
≥ 55 cm	2	0
Total	734	55

2.1.2 Bamboo Forests

Location and design of the plots

Criteria for sample plot establishment

The criteria for bamboo sample plot establishment are: i) representativeness (based on assessment of experts) of the forest types being studied; ii) representativeness for topographic conditions; and iii) covering a number of different bamboo sizes; iv) the sample plots should be set up on less disturbed area. The area for one bamboo sample plot is 0.5 ha, which is half of that for woody forest (because the variation in bamboo forests is often smaller than woody forests). The shape of the plot is rectangular (100m x 50m).

Description of sample plot BT-PD-04

Table 8: Description of sample plot BT-PD-04

Plot name	BT-PD-04
Administrative location	Compartment no. 16 - Phan Dũng commune - Tuy Phong district - Bình Thuận province
Owner/manager	Tuy Phong Protection Forest Management Board
Coordinate (VN2000 projection)	Long = 108° 38' 56" E; Lat = 11° 30' 32" N
Altitude	365 m
Slope	24°
Plot area	1 ha
Plot size	100 x 100 (m)
Forest type	Bamboo forest (forest predominated by bamboo species)
Forest status	Medium
Density (estimated)	4,100 trees/ha

The position of the plot BT-PD-04 is shown in Figure 1 above.

Selection of the sampling trees

Firstly, all bamboos were grouped into DBH classes. The interval of DBH class is 1 cm, and DBH classes are: 2 – 2.9 cm; 3 – 3.9 cm; 4 – 4.9 cm; 5 – 5.9 cm; 6 – 6.9 cm; 7 – 7.9 cm; 8 – 8.9 cm; etc. Next, the sample bamboos from each DBH class were randomly selected following the next crtiteria: i) Samples should be allocated in proportion with the number of bamboos in each DBH class; ii) Samples should be representative of the age class; iii) The number of samples would be determined based on the number of DBH classes identified, and the bamboo age. No minimal number of sample have been setted up but there is at least 4 samples per dbh class. The total number of samples for harvesting is 120 (100 bamboos for development of allometric equations and 20 bamboos for validation).

Biomass data of 120 sample bamboos were collected. The numbers of bamboos divided by DBH classes and age classes are given in Table 9. Data on species name, DBH, height, age class, and dry weight of each component of these sample bamboos are given in Annex 8.

Table 9: Numbers of sample bamboos by DBH classes and age classes

DBH Class	Age class			Total
	Old	Medium	Young	
2.0 – 2.9 cm	1	3	2	6
3.0 – 3.9 cm	4	5	4	13
4.0 – 4.9 cm	4	5	6	15
5.0 – 5.9 cm	8	8	6	22
6.0 – 6.9 cm	7	9	8	24
7.0 – 7.9 cm	8	6	6	20
8.0 – 8.9 cm	6	6	4	16
9.0 – 9.9 cm	1	1	2	4
Total	39	43	38	120

2.2 Variables measurement and calculation for volume and biomass

2.2.1 Field measurements

Woody Forests (evergreen and deciduous broad leaved forests)

Measurement of tree DBH and identification of tree species

All live trees with DBH from 5 cm and above in the sample plots were measured. The information collected include: i) tree species (Vietnamese and scientific names); and ii) DBH of trees.

Destructive measurement of fresh biomass of sample trees

Firstly, the measurement point for DBH was marked, then the tree was felled at its base following logging procedures. Following this, using measuring tapes measurements were taken for :

a) Diameter and height of the stump;
b) DBH at 1.3 m;
c) Total tree height (from the stump to the top of the crown).
d) Length of tree bole - from the stump to the first main branch;
e) Length of tree bole - from the stump to the point where diameter becomes 10 cm;

Next, the tree was separated into different components (stem, branches and leaves) and the weights of these components were weighed immediately in the field using a digital scale with the maximum capacity of 500 kg and the precision of 0.1 kg.

Collecting samples for analysis of dry oven mass and wood density

Sampling for dry mass analysis was done immediately after completion of measurement of fresh weight of each tree component. The following steps were conducted:

Samples for dry mass analysis: collect three samples per tree of stem, branches and leaves. The samples are taken from different positions of the stem, and different parts of branches and leaves so that they are representative for the parts being sampled. Following ICRAF (2011), the samples of the stem and branch were about 0.5 to 1.0 kg in weight. The samples of the leaves were about 0.3 – 0.5 kg in weight. The samples for dry mass analysis are weighed immediately and carefully using two digital scales (one is Ohaus BC15 with the maximum weighing capacity of 15 kg and the precision of 0,5 g and the other is Ohaus SPS2001F with the maximum weighing capacity of 2 kg and the precision of 0,1 g) to accurately determine the fresh weight of each sample.

Samples for wood density analysis: five wood discs were taken from the stem. The sampling positions were at stump level (0.0 m), at 1/5; 2/5; 3/5; and 4/5 of stem length. The wood discs were 5 – 10 cm thick. For small discs (diameter ≤ 20 cm), the whole disc is taken as is. For large discs (diameter > 20 cm), only a radial section of the disc is taken.

The field measurement of forest biomass were conducted through sample plots following draft version of the Guidelines on Destructive Measurement for Forest Biomass Estimation developed (UN-REDD Vietnam 2012).

Bamboo Forests

Measurement of bamboos

In the bamboo sample plots, four sub-plots with an area of 400 m^2 (20m x 20m) each were established.

DBH was measured using diameter tapes and age class (old, medium and young) was determined for each bamboo with DBH over 2cm in the sub-plots. After a bamboo is measured, the bamboo was marked with white paint to avoid missed or repeated measurement.

Destructive measurement of fresh biomass of sample bamboo

First, the bamboo was felled using a hand saw. Next, measurement of height of the felled bamboo was taken. Finally, the bamboo was separated into different components: stem, branches and leaves and measured immediately for weight of each component using a scale.

Collecting samples for dry mass analysis

For bamboo forests, only samples for dry mass analysis were collected. Samples were collected immediately after measurement of fresh weights of each bamboo component. Out of 120 sample bamboos for fresh biomass measurement, 70 samples were selected for sampling of dry mass analysis. The selected bamboos for sampling of dry mass analysis should be representative of each age group and DBH class. For each sample bamboo selected for dry mass analysis, six sub-samples are collected. Four sub-samples are taken from the stem (at the stump level; ¼; ½; and ¾ of stem length positions), one sub-sample from branches and one sub-sample from leaves. The sub-samples are collected such that their weights are 0.5 – 1.0 kg for stem and branch sub-samples, and 0.3 – 0.5 kg for leave sub-samples. The sub-samples are then weighed immediately in the field using a high precision scale (Ohaus SPS2001F with the maximum weighing capacity of 2 kg and the precision of 0,1 g) to accurately determine their weights.

2.2.2 Laboratory measurements

Analysis of oven dry mass and wood density

After the completion of the destructive measurement in the field, the collected samples were sent immediately to laboratories in the Forest Science Institute of Vietnam (FSIV) for oven dry mass and wood density analyses. Dry mass of samples were determined using oven drier at a temperature of 105°C until the samples reached constant weights. Basic wood densities of all wood discs are determined at the moisture content of 0%. Wood density measurements methodology followed the National standard TCVN

8048-2: 2009. The wood volume was determined using the water displacement method with prism shaped and minimum sized: 20 x 20 x 25 mm subsamples. Wood densities was then calculated with the following formula:

$$SWD = \frac{SDW}{SV} \quad \text{Formula (1)}$$

Where: *SWD* is the wood density of the sample in g/cm^3; *SDW* is the dry weight of sample cube and *SV* is the volume of sample cube.

Calculation of total dry biomass

The total dry weights (TDW) for each component of the sample trees are calculated based on the total fresh weights (TFW) of each component measured in the field and the ratios of dry weight to fresh weight calculated for each component in the laboratory. The formula for TDW calculation is as follows:

$$TDW_c = TFW_c \frac{SDW_c}{SFW_c} \quad \text{Formula (2)}$$

Where: TDW_c is the total dry weight of a component *c* (stem, branches, or leaves); TFW_c is the total fresh weight of this component measured in the field; SDW_c and SFW_c are the dry weight and fresh weight and the samples for this component.

The total above-ground biomass of a tree is the sum of its total dry weights of three components: stem, branches and leaves. The formula is:

$$TDW_{tree} = TDW_{stem} + TDW_{branch} + TDW_{leave}. \quad \text{Formula (3)}$$

2.2.3 Other variables

According to IPCC 2003, BEF is – when used to calculate aboveground biomass of forests – the ratio of aboveground oven-dry biomass of trees to oven-dry biomass of the commecial volume, dimensionless. The biomass of commercial volume can be calculated as commercial volume times wood density or directly measured as the biomass of tree bole. In this study the formula used is:

$$BEF = \frac{AGB_{total}}{AGB_{stem}}$$

2.3 Model fitting and selection

2.3.1 Regression Analysis

Regression analyses are conducted using the SAS software. For evergreen broadleaf and deciduous forests, the variables include DBH (D, cm), height (H, m) and wood density (ρ, g/cm^3). The following nine models are used:

$AGB = aD^b$ Model (1)

$AGB = a(D^2H)^b$ Model (2)

$AGB = a(D^2H^{0.7})^b$ Model (3)

$AGB = aD^bH^c$ Model (4)

AGB = a$(D^{2.4}\rho)^b$ Model (5)

AGB = a$D^b\rho^c$ Model (6)

AGB = a$(D^2H\rho)^b$ Model (7)

AGB = a$(D^2H^{0.7}\rho)^b$ Model (8)

AGB = a$D^bH^c\rho^d$ Model (9)

Where *a*, *b*, and *c* are the coefficients needed to be found. Form (5), Form (7) and Form (10) are based on the results of previous analyses using other datasets from the regression analysis for the stem volume equation using the form V = aD^2H^b and V = cD^d, the optimal values for *b* and *d* are, respectively, approximately 0.7 and 2.4 (unpublished data).

For bamboo forest, the variables include DBH (D) and height (H). The Forms (3) to (6) above are used for regression analysis.

Three approaches of regression analysis were used to find the coefficients;

- The first approach is to apply the least squares optimization to the original equations.
- The second approach is to transform the above equations to the logarithmical form and then apply the least squares optimization to the transformed equations.
- The third approach is to use the maximum likelihood optimization to the original equations.

2.3.2 Comparison of models

In order to evaluate the in-sample performance of the models, three indicators were employed: the adjusted R^2 ($\bar{R}^2$) value, the sum of squared error (SSE) and Akaike information criterion with correction (AICc, Burnham and Anderson 2002). The AICc is calculated by the following formula:

$$AIC_c = -2\ln(L) + \frac{2kn}{n-k-1}$$ Formula (4)

Where *L* is the maximum likelihood of the model, *k* is the number of parameters needed to be estimated, and *n* is the size of the sample dataset.

The $\bar{R}^2$ and SSE are used to measure the goodness of fit for equations that are developed using the least squares method. The AICc is used to measure the goodness of fit for equations that are developed using the maximum likelihood method.

2.3.3 Cross validation and error assessment

To avoid overfitting of the models, cross validation tests were conducted. The sample dataset is randomly divided into two sub-sets: a training subset and a testing sub-set. The sizes of the training and testing sub-set are, respectively, 2/3 and 1/3 the size of the original dataset. For each division, the training sub-set is used to fit the models and then the fitted models are used to predict the total dry weights of the testing sub-set. These predicted total dry weights are then used to calculate the errors (in percentage) as compared to the measured total dry weight of the testing sub-set. The above procedure was repeated one million times to generate probability density functions of the total AGB error of each equation. Each probability density function is then approximated by a log-normal distribution (Formula (4)) in order to compare the performance of the equations in practice and estimate the confidence intervals of the error for each equation.

$$f_x(x;\alpha,\sigma,\mu) = \frac{\alpha}{(\alpha x+1)\sigma\sqrt{2\pi}} \times e^{-\frac{(\ln(\alpha x+1)-\mu)^2}{2\sigma^2}} \qquad \text{Formula (5)}$$

To facilitate the cross validation tests, a program was written in the C language. The program was validated by comparing its results with the SAS software. With the same dataset, the program generated the same results with the SAS software for every combination of the statistical models and regression approaches.

3 RESULTS FOR EVEGREEN BROAD LEAVED FORESTS

3.1 Result 1: forest and trees characteristics

3.1.1 Species composition

A total of 79 species were identified among the 1,715 trees of the two evergreen broadleaf sample plots. For 85 of these trees (N% = 5%) species names could not be identified. Table 10 provides a list the ten most dominant species. The full list of species is given in Annexes.

Table 10: Summary of tree species composition (evergreen broadleaf sample plots)

No.	Local Name	Scientific Name	N	G	N%	G%	IV%
1	Bằng lăng	Lagerstroemia calyculata	318	15.593	18.54	19.57	19.06
2	Căm xe	Xylia xylocarpa	223	8.543	13.00	10.72	11.86
3	Dầu rái	Dipterocarpus alatus	71	9.760	4.14	12.25	8.20
4	Cò ke	Microcos paniculata	169	2.588	9.85	3.25	6.55
5	Bình linh nghệ	Vitex ajugaeflora	66	3.041	3.85	3.82	3.83
6	Thẩu tấu	Aporosa sphaerosperma	83	1.690	4.84	2.12	3.48
7	Kơ nia	Irvingia malayana	53	2.675	3.09	3.36	3.22
8	Gáo	Adina polycephala	29	3.487	1.69	4.38	3.03
9	Gạo	Bombax malabarica	2	3.714	0.12	4.66	2.39
10	Sến	Madhuca sp.	42	1.500	2.45	1.88	2.17
		Total			61.57	66.01	63.79

Based on the importance value (IV%) index, which is calculated by taking the average of N% and G%, the most dominant species is *Lagerstroemia calyculata*, with an IV% value of 19.06 and accounts for 18.5% of the total number of trees and 19.6% of the basal area. The second most dominant species is *Xylia xylocarpa*, with an IV% value of 11.86%. Other species with IV% ≥ 5% include *Dipterocarpus alatus* (IV% = 8.20) and *Microcos paniculata* (IV% = 6.55%). The ten most dominant species account for 61.57% of the total number of trees and 66.01% of the total basal area.

3.1.2 Forest structure

There are a total of 1,658 trees in the two studied sample plots. The average density is 829 trees/ha. The N-D distribution of all the trees in the two studied sample plots is given in Figure 4. The number of trees decreases with increase in DBH.

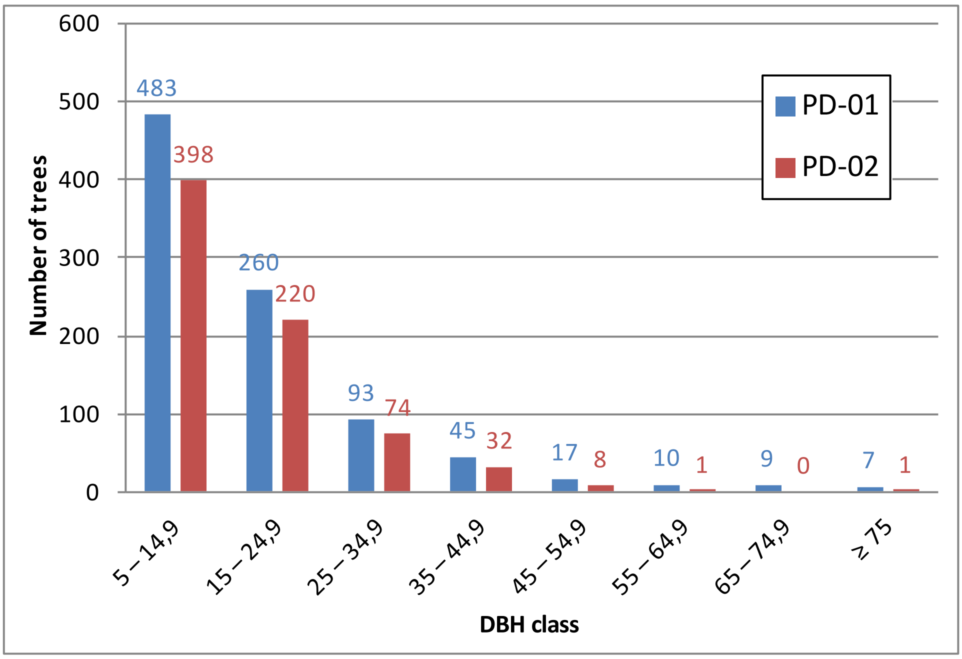


Figure 4: N-D distribution of trees (evergreen broadleaf sample plots)

3.1.3 Biomass of sample trees

Biomass data of a total of 110 sample trees are collected.

The results of dry mass analysis of 110 sample trees are given in Table 11. As an average, the stem has the highest ratio followed by branches. The coefficient of variation (CV, %) of the ratios is smallest in branches and highest in leaves.

Table 11: Ratio of dry biomass to fresh biomass (evergreen broadleaf forests)

Statistical values	Dry to fresh mass ratio		
	Stem	Branch	Leaf
Min	0.455	0.431	0.229
Max	0.617	0.580	0.451
Avg	0.539	0.507	0.343
Stdev	0.040	0.035	0.052
CV(%)	7.416	6.852	15.298

From the data of fresh biomass and the dry-to-fresh mass ratio data, the dry biomass of each component of the trees were calculated using the Formula (2).

Data on species name, DBH, height, volume, fresh biomass of each component, dry-fresh mass ratios and the converted dry biomass of each tree component of these 110 sample trees are given in Annex 5.

3.1.4 Relation between H and diameter

The SAS software was employed to do a regression analysis of the logarithm function correlating H(m) and DBH (cm). The resulting equation is H = 10.190 × ln(DBH) – 12.026 (R^2 = 0.834; F value = 544.2; $p < 0.001$; Figure 5). The graphic exploration using the dataset of the Study, indicates a correlation between H and DBH.

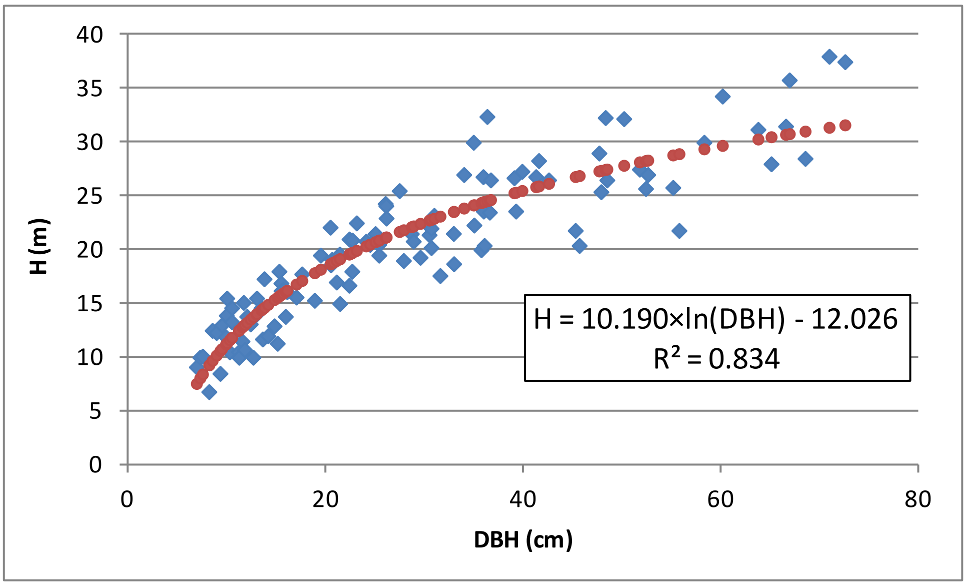


Figure 5: Correlation function between H (m) and DBH (cm) (evergreen broadleaf sample plots)

3.1.5 Wood density analysis

Table 12 below provides a summary of wood density analysis results for the species in the sample dataset. The average wood density is 0.565 ± 0.078 g/cm³. For those species that have previous known values, the wood densities analyzed in this Study seem to be smaller. One of the reasons may be that some of the wood densities collected in earlier studies[2] are dried wood densities (i.e., calculated by dividing oven-dried mass to dried volume) while in this Study the wood densities are calculated by dividing the oven-dried mass to the green volume over bark of the samples. Another reason may be due to differences in the method of analyzing the wood densities in different laboratories such as using different volume measurement methods. The wood densities of all 110 sample trees are provided in Annex 5.

Table 12: Wood density analysis for species (evergreen broadleaf sample plots)[3]

No	Local Name	Scientific Name	N	Wood density (g/cm³)			
				Min	Max	Avg	Known Value*
1	Bằng lăng	Lagerstroemia calyculata	13	0.518	0.559	0.536	0.610
2	Bằng lăng ổi	Lagerstroemia crispa	3	0.623	0.641	0.629	0.670
3	Bình linh nghệ	Vitex ajugaeflora	4	0.490	0.567	0.529	

[2] By the Research Centre for Forest Ecology and Environment (RCFEE).

[3] Source: from the wood density database collected by RCFEE in the initial phase of this Study.

No	Local Name	Scientific Name	N	Wood density (g/cm^3)			
				Min	Max	Avg	Known Value*
4	Căm xe	Xylia xylocarpa	9	0.620	0.714	0.657	1.150
5	Cẩm lai	Dalbergia oliveri	3	0.629	0.660	0.640	1.055
6	Cẩm liên	Shorea siamensis	1	0.586	0.586	0.586	0.910
7	Cẩm thị	Diospyros maritima	3	0.558	0.629	0.588	
8	Chiêu liêu	Terminalia chebula	6	0.594	0.672	0.639	
9	Chiêu liêu nghệ	Terminalia triptera	1	0.588	0.588	0.588	
10	Cò ke	Microcos paniculata	6	0.535	0.569	0.549	
11	Dầu rái	Dipterocarpus alatus	16	0.471	0.595	0.534	0.780
12	Dầu trà beng	Dipterocarpus obtusifolius	3	0.650	0.667	0.657	0.850
13	Gáo	Adina polycephala	5	0.320	0.390	0.360	
14	Gáo vàng	Adina pilulifra	2	0.571	0.591	0.581	0.650
15	Giáng hương	Pterocarpus macrocarpus	3	0.538	0.604	0.574	0.820
16	Ké	Nephelium sp.	1	0.512	0.512	0.512	
17	Kơ nia	Irvingia malayana	3	0.674	0.700	0.683	0.980
18	Lim đá	Sp1	2	0.575	0.620	0.597	
19	Lim xẹt	Peltophorum pterocarpum	1	0.395	0.395	0.395	0.600
20	Móng bò	Bauhinia sp.	1	0.532	0.532	0.532	
21	Săng mây	Antheroporum pierrei	2	0.648	0.681	0.664	
22	Sao đen	Hopea recopei	2	0.540	0.554	0.547	0.740
24	Sến mủ	Shorea roxburghii	4	0.570	0.599	0.585	0.890
25	Sổ	Dillenia scabrella	1	0.564	0.564	0.564	
26	Tà quang	Sp2	1	0.666	0.666	0.666	
27	Thành ngạnh	Cratoxylon pruniflorum	1	0.599	0.599	0.599	
28	Thị nhong	Diospyros sp.	1	0.693	0.693	0.693	
29	Trâm	Syzygium sp.	1	0.491	0.491	0.491	
30	Trâm đỏ	Syzygium oblatum	3	0.530	0.590	0.561	
31	Vậy nước	Cephalanthus tetrandra	2	0.357	0.395	0.376	
32	Vên vên	Anisoptera costata	1	0.590	0.590	0.590	0.650
33	Vừng	Careya arborea	2	0.491	0.508	0.500	
34	Xoài rừng	Mangifera minitifolia	3	0.464	0.495	0.476	
All trees			110	0.320	0.714	0.565	

3.2 RESULT 2: Modeling of the stem volume

The stem volume has not been measured during the field work so no model has been developed.

3.3 RESULT 3: Modeling of Aboveground biomass

3.3.1 Modeling per tree compartments

With DBH only

Allometric equations for each component (stem, branches and leaves) of the tree are also developed. Only the power model which uses the input variable D (Model (1)) is used here. The regression analyses are done using the SAS software. The results are given in Table 13, Figure 6, Figure 7 and Figure 8 below.

Table 13: Results of the second regression approach relating dry biomass (kg) of each part of the tree with DBH (cm) for Model (1) (evergreen broadleaf)

Part of tree	Parameter *a*				Parameter *b*				R^2	Pr > F
	Est.	Std. err.	95% CL		Est.	Std. err.	95% CL			
Stem	0.1138	0.0142	0.0856	0.1420	2.3513	0.0387	2.2746	2.4280	0.9162	<.0001
Branch	0.0070	0.0022	0.0026	0.0115	2.7063	0.0986	2.5109	2.9016	0.7167	<.0001
Leaf	0.0085	0.0029	0.0028	0.0143	1.9217	0.1060	1.7116	2.1319	0.6396	<.0001

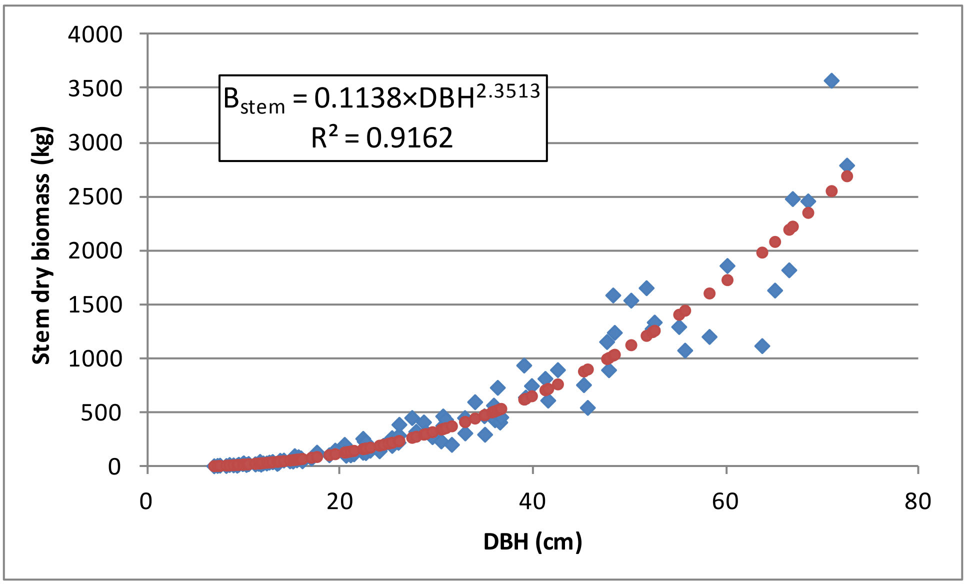


Figure 6: Equation for estimating dry stem biomass from DBH (cm) (evergreen broadleaf)

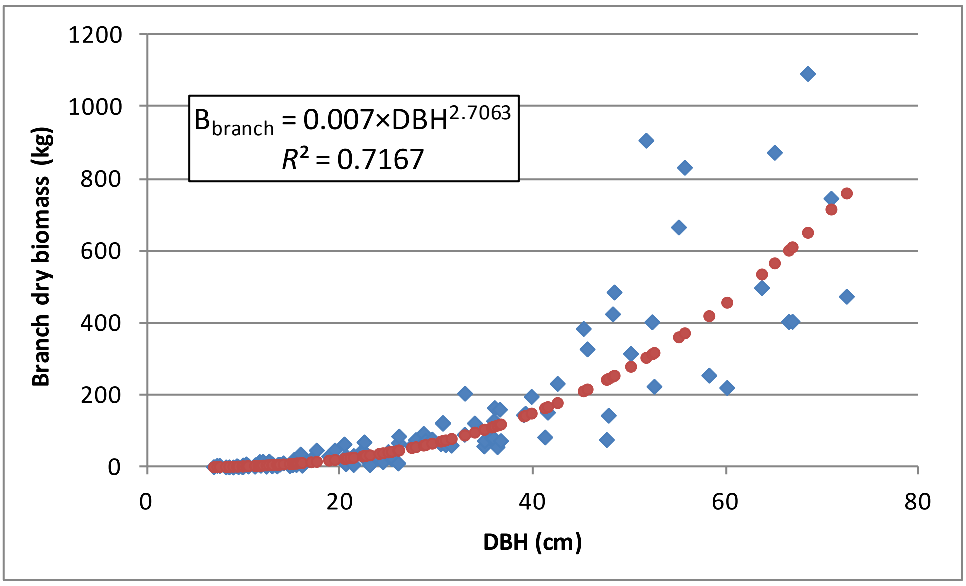


Figure 7: Equation for estimating dry branch biomass from DBH (cm) (evergreen broadleaf)

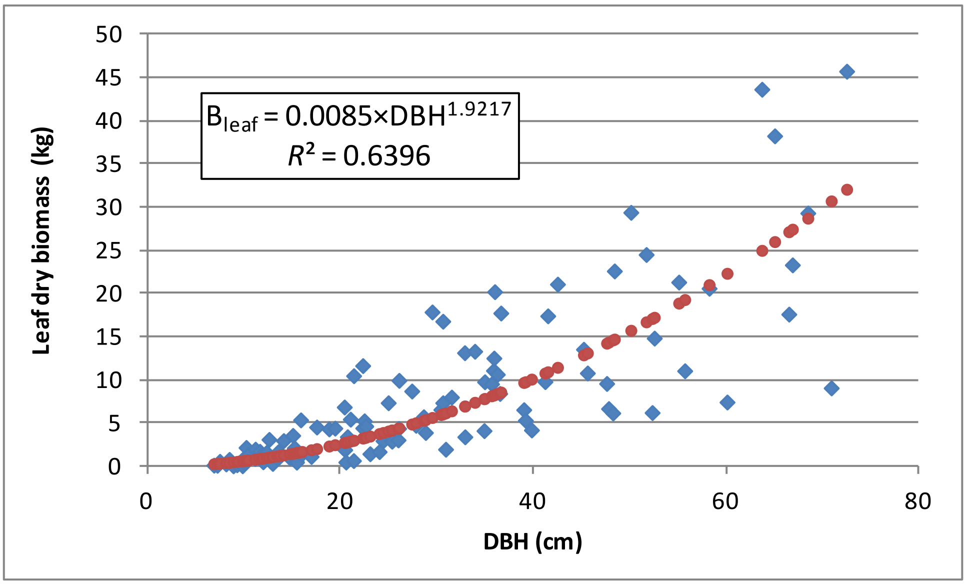


Figure 8: Equation for estimating dry leaf biomass from DBH (cm) (evergreen broadleaf)

It is observed from the above figures that stem biomass correlates strongest to the DBH (R^2 = 0.92), followed by branch biomass (R^2 = 0.72). Leaf biomass has weak correlation with DBH (R^2 = 0.64).

With all explanatory variables

The modeling of tree components has not been done with all the explanatory variables.

3.3.2 Modeling of total aboveground biomass

Model fitting

First, regression analyses using the first approach (least squares optimization of the original equations) for statistical Models (1) through (9) are applied using the procedure NLIN of the SAS software. The analyzed results are given in Table 14.

Table 14: Regression analyses using the first approach (evergreen broadleaf)

Model No		a*	b*	c*	d*	$\bar{R}^2$	SSE	Pr > F
Model (1)	$B = aD^b$	0.1575	2.3389			0.9287	5,506,201	<.0001
Model (2)	$B = a(D^2H)^b$	0.0637	0.9014			0.9260	5,731,524	<.0001
Model (3)	$B = a(D^2H^{0.7})^b$	0.0771	0.9716			0.9296	5,447,732	<.0001
Model (4)	$B = aD^bH^c$	0.1041	2.1335	0.3661		0.9308	5,298,038	<.0001
Model (5)	$B = a(D^{2.4}\rho)^b$	0.2503	0.9842			0.9593	3,143,630	<.0001
Model (6)	$B = aD^b\rho^c$	0.2549	2.3603	1.0047		0.9589	3,142,787	<.0001
Model (7)	$B = a(D^2H\rho)^b$	0.1083	0.8998			0.9558	3,434,671	<.0001
Model (8)	$B = a(D^2H^{0.7}\rho)^b$	0.1353	0.9703			0.9599	3,106,945	<.0001
Model (9)	$B = aD^bH^c\rho^d$	0.1776	2.1506	0.3528	0.9839	0.9615	2,922,987	<.0001

* All parameters are significant at $p < 0.001$.

All equations have quite high $\bar{R}^2$ value, indicating that they can all be used to estimate forest biomass. The equation derived from Model (1) is among the equations with the lowest $\bar{R}^2$, as it uses only the predictor D. Nevertheless, Model (1) is considered a useful model as H and ρ are difficult and costly to measure and the $\bar{R}^2$ value is only 3.5% off compared to the most optimal model. Among the three models that use only D and H as the input variables, only Model (3) and Model (4) have slightly higher $\bar{R}^2$ values as compared to Model (1) (Model (2) has lower $\bar{R}^2$), suggesting that the inclusion of H does not significantly improve the prediction accuracy for this dataset. Models that use ρ as an input variable (Models (5) to (10)) have significantly higher $\bar{R}^2$ values as compared to models that do not use ρ (Models (1) to (4)), indicating that the inclusion of ρ can significantly improve the accuracy of the prediction. Among the three models that use only input variables D and H, Model (4) has the highest $\bar{R}^2$. Between the two models that use only D and ρ as input variables, Model (5) performs better. Models (7) to (9) all using three input variables, have the highest $\bar{R}^2$. Among these three, Model (9) has the highest $\bar{R}^2$value.

Next, regression analyses using the second approach (least squares optimization of the logarithmically transformed forms) are performed using the procedure NLIN in the SAS software. The analyzed results are provided in Table 15.

Table 15: Regression analyses using the second approach (evergreen broadleaf)

Model No	a*	b*	c*	d*	$\bar{R}^2$	SSE	Pr > F
Model (1)	0.1230	2.3965			0.9727	6.6613	<.0001
Model (2)	0.0387	0.9424			0.9756	5.9470	<.0001

Model (3)	0.0511	1.0085			0.9764	5.7492	<.0001
Model (4)	0.0553	2.0534	0.6400		0.9763	5.7394	<.0001
Model (5)	0.2243	0.9957			0.9811	4.6063	<.0001
Model (6)	0.2128	2.3919	0.9175		0.9810	4.5911	<.0001
Model (7)	0.0680	0.9407			0.9845	3.7915	<.0001
Model (8)	0.0939	1.0057			0.9851	3.6306	<.0001
Model (9)	0.0941	2.0382	0.6596	0.9302	0.9849	3.6121	<.0001

* All parameters are significant at $p < 0.001$.

It is observed that with the same model, the coefficients estimated using the second approach are quite different from those estimated using the first approach. Performance of models ranked by $\bar{R}^2$ using the second approach is similar to that using the first approach. There are some small differences. Model (2) now has the higher $\bar{R}^2$ than Model (1) and Model (3) is now the most optimal among three models that use only D and H. Among the three models that use all three input variables, Model (8) now has the highest $\bar{R}^2$.

Finally, regression analyses using the third approach (maximum likelihood optimization) are done by the procedure NLP in the SAS software. The analyzed results are given in Table 16.

Table 16: Regression analyses using the third approach (evergreen broadleaf)

Model No	**a***	**b***	**c***	**d***	**Log*L***	**AICc**
Model (1)	0.1277	2.3943			-610.95	1226.01
Model (2)	0.0396	0.9417			-602.76	1209.63
Model (3)	0.0530	1.0072			-601.77	1207.65
Model (4)	0.0525	2.0104	0.7123		-601.77	1209.76
Model (5)	0.2328	0.9933			-588.74	1181.59
Model (6)	0.2458	2.3811	1.0768		-588.60	1183.42
Model (7)	0.0694	0.9393			-575.87	1155.85
Model (8)	0.0968	1.0037			-573.70	1151.52
Model (9)	0.1012	2.0074	0.6987	1.0650	-573.60	1155.58

* All parameters are significant at $p < 0.001$.

It is observed from the table that the values of the coefficients estimated using the third approach are very close to those estimated using the second approach. Model (1), which uses only D as the input variable, has the highest (least optimal) AICc value. Models that use two variables D and ρ (i.e. Model (5) and Model (6)) perform better than models that use two variables D and H (i.e. Models (2), (3) and (4)) in terms of the AICc, suggesting that the inclusion of ρ is more important than the inclusion of H in improving the prediction accuracy. Among the three models that use only D and H, Model (3) has the lowest (most optimal) AICc value. Between two models that use only D and ρ, Model (5) performs better in terms of AICc and should be chosen. Finally, Models (7) to (9), which use all three input variables, have the lowest AICc values. Among them, Model (8) has the best AICc.

Cross validation and error assessment

To avoid over-fitting of the models, cross validation tests were carried out. Table 17 shows the properties of the approximated probability density functions of the total AGB error for every equations developed using the first approach.

Table 17: Properties of probability density functions of total AGB error (%) for equations developed using the first regression approach (evergreen broadleaf)

Model No	α	Σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0104	0.0833	0.0025	0.571	0.236	-0.429	0.0499	0.9994	-14.26	17.30	31.56
Model (2)	0.0098	0.0812	0.0166	2.047	1.705	1.024	0.0476	0.9994	-13.54	19.58	33.12
Model (3)	0.0101	0.0808	0.0129	1.621	1.292	0.638	0.0493	0.9994	-13.43	18.54	31.97
Model (4)	0.0099	0.0823	0.0019	0.534	0.191	-0.491	0.0482	0.9994	-14.86	17.87	32.73
Model (5)	0.0091	0.0562	-0.0009	0.073	-0.102	-0.450	0.0644	0.9997	-11.62	12.75	24.37
Model (6)	0.0105	0.0709	-0.0051	-0.243	-0.482	-0.957	0.0595	0.9994	-12.77	13.64	26.41
Model (7)	0.0077	0.0509	0.0170	2.394	2.223	1.883	0.0597	0.9995	-10.27	16.02	26.29
Model (8)	0.0083	0.0513	0.0137	1.830	1.669	1.347	0.0635	0.9995	-10.07	14.65	24.72
Model (9)	0.0112	0.0756	0.0014	0.381	0.126	-0.383	0.0593	0.9995	-12.17	14.38	26.55

It is observed that all models have very high R^2, indicating that Formula(5) is a good form to approximate the probability density functions of the total AGB error. In this table, the means (or expected values) of error indicate the accuracy while the ranges of error show the robustness of the models. Model (1), which uses only D as the input variable, can be considered accurate (mean = 0.571) but among the least robust models (the range of error is from -14.26% to 17.30%). Models that use H as an additional input variable (i.e. Models (2), (3) and (4)) do not have better accuracy or robustness. This is probably caused by the utilization of an inappropriate regression approach. Models that use ρ as an additional variable (i.e. Model (5) and Model (6)) are more accurate (i.e., their means are closer to zero) and have smaller ranges of error as compared to Models (1) to (4), confirming the importance of using ρ as an variable for biomass estimation. Finally, Models (7) to (9), though using all three predictors, do not improve the accuracy nor the robustness of biomass prediction as compared to Models (5) and (6). The probability density functions of error for some selected models are shown in Figure 9.

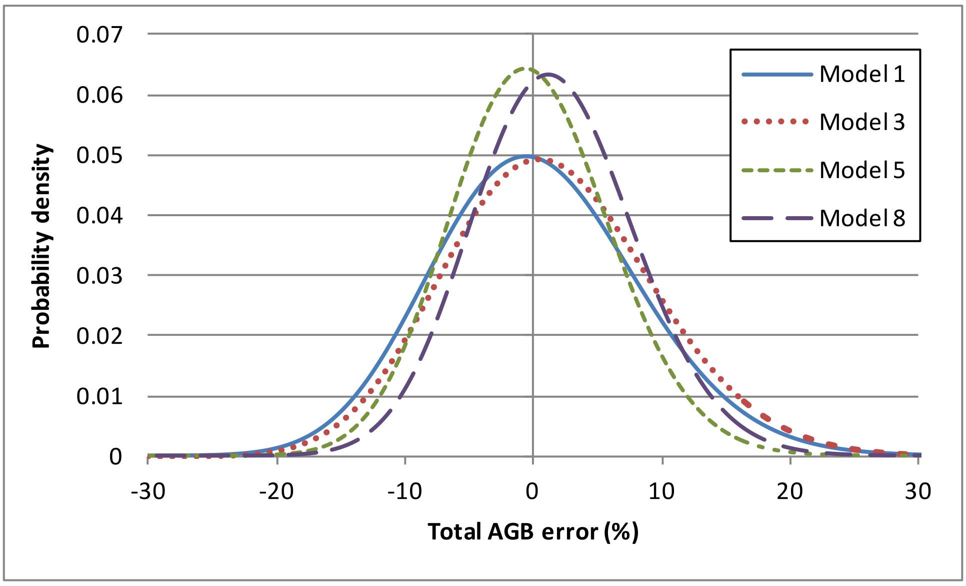


Figure 9: Probability density functions of total AGB error (%) for selected equations developed by the first regression approach (evergreen broadleaf)

Next, the cross validation test is performed for equations derived using the second regression approach. The results are provided in Table 18.

Table 18: Properties of probability density functions of total AGB error (%) for equations developed using the second regression approach (evergreen broadleaf)

Model No.	α	σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0128	0.0937	-0.0294	-2.423	-2.842	-3.675	0.0565	0.9995	-14.95	12.99	27.94
Model (2)	0.0128	0.0953	-0.0449	-3.073	-3.411	-4.084	0.0563	0.9995	-16.14	11.91	28.05
Model (3)	0.0132	0.0956	-0.0408	-2.919	-3.279	-3.994	0.0576	0.9996	-15.47	11.96	27.43
Model (4)	0.0131	0.0950	-0.0399	-2.893	-3.255	-3.974	0.0574	0.9995	-15.49	12.06	27.55
Model (5)	0.0123	0.0699	-0.0146	-1.149	-1.378	-1.834	0.0717	0.9997	-11.40	10.55	21.95
Model (6)	0.0124	0.0719	-0.0171	-1.301	-1.531	-1.990	0.0705	0.9997	-11.74	10.59	22.33
Model (7)	0.0100	0.0578	-0.0242	-2.103	-2.256	-2.562	0.0712	0.9996	-12.79	9.26	22.05
Model (8)	0.0105	0.0575	-0.0205	-1.926	-2.093	-2.426	0.0742	0.9997	-11.93	9.23	21.16
Model (9)	0.0107	0.0597	-0.0222	-2.087	-2.268	-2.627	0.0729	0.9997	-12.20	9.33	21.53

It is observed that equations developed using the second regression approach tend to underestimate the total AGB by about 1.4-3.5%. However, they perform much better than the equations developed using the first regression approach in terms of robustness (i.e., with smaller ranges of error). For example, the equation developed by using the second regression approach of Model (8) has the range of error from -11.93 to 9.23, while the corresponding one developed by using the first regression approach has the range of error from -10.07 to 14.65.

Among the three models that use only variables D and H, Model (3) has the smallest range of error. Between the two models that use only variables D and ρ, Model (5) has smaller range of error. Among the three models that use all three input variables, Model (8) is the most accurate and robust. The probability density functions of the equations derived from Models (1), (3), (5) and (8), which are the best for each group of input variables, are shown in Figure 10.

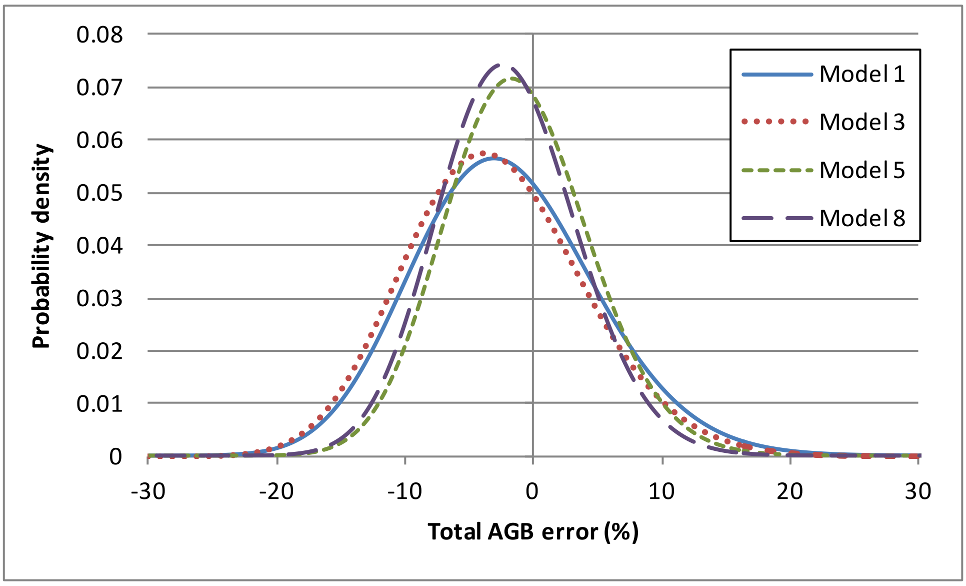


Figure 10: Probability density functions of total AGB error (%) for selected equations developed by the second regression approach (evergreen broadleaf)

Finally, the cross validation test was performed for equations developed using the third regression approach and the results are provided in Table 19.

Table 19: Properties of probability density functions of total AGB error (%) for equations developed by the third regression approach (evergreen broadleaf)

Model No.	α	σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0102	0.0746	0.0064	0.909	0.634	0.087	0.0544	0.9995	-12.78	16.16	28.94
Model (2)	0.0129	0.0941	-0.0254	-1.614	-1.951	-2.619	0.0562	0.9996	-14.72	13.40	28.12
Model (3)	0.0122	0.0872	-0.0095	-0.467	-0.777	-1.392	0.0565	0.9996	-13.54	14.36	27.90
Model (4)	0.0120	0.0867	-0.0103	-0.542	-0.852	-1.470	0.0560	0.9995	-13.74	14.43	28.17
Model (5)	0.0105	0.0599	0.0053	0.679	0.508	0.167	0.0700	0.9997	-10.05	12.38	22.43

Model (6)	0.0110	0.0636	0.0013	0.301	0.117	-0.248	0.0693	0.9997	-10.51	12.15	22.66
Model (7)	0.0106	0.0599	-0.0196	-1.663	-1.829	-2.161	0.0722	0.9997	-12.07	9.68	21.75
Model (8)	0.0097	0.0525	-0.0079	-0.666	-0.808	-1.089	0.0744	0.9997	-10.81	10.28	21.09
Model (9)	0.0097	0.0539	-0.0108	-0.961	-1.110	-1.407	0.0725	0.9997	-11.36	10.28	21.64

Equations developed using the third regression method are generally more accurate (i.e., their means of error are closer to zero) than those developed using the second approach. It is observed that all equations that use H as an additional variable tend to underestimate the total biomass. This may be due to the sample trees are not well sampled in the H variable. However, the fact that all equations that use H as an additional variable have smaller ranges of error as compared to the corresponding ones that do not use H, confirms that the inclusion of H can improve the robustness of the prediction. The inclusion of H, however, does not improve the accuracy of the prediction.

Among the three models that use D and H as the input variable, Model (3) has the smallest range of error. Between the two models that use D and ρ as input variables, Model (5) has a marginally smaller range of error. Among the three models that use all three input variables, Model (8) is the most accurate and robust. The probability density functions of the equations derived from Models (1), (3), (5) and (8) using the third regression approach are shown in Figure 11.

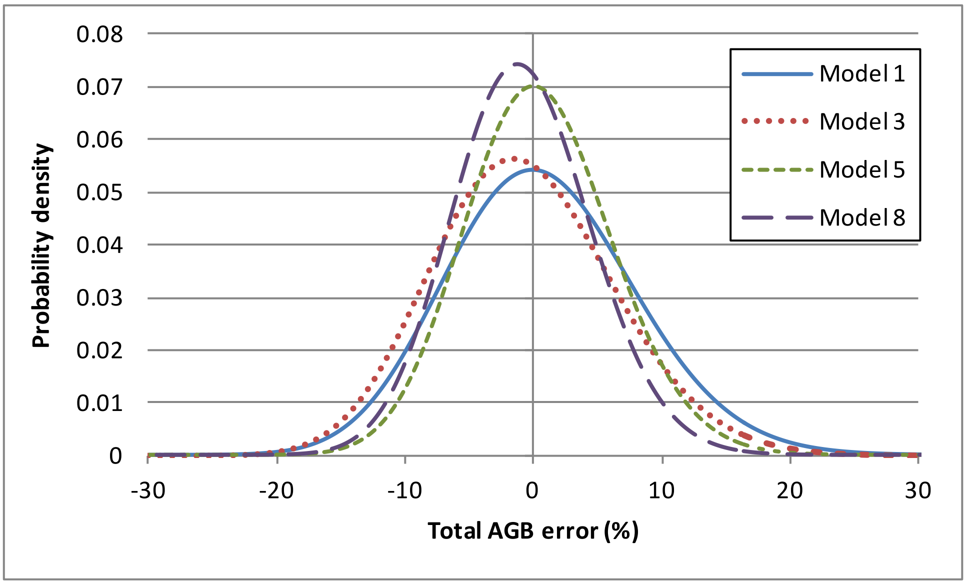


Figure 11: Probability density functions of total AGB error (%) for selected equations developed by the third regression approach (evergreen broadleaf)

In order to find the best equations for each group of input variables, a comparison of the probability density functions of total AGB error was conducted across the three regression approaches. The results are shown in Figure 12. The figure indicates that with the same model, equations developed using the first regression approach are the least robust (i.e., with largest ranges of error). Equations developed using the second regression approach have the smallest ranges of error. However, these equations tend to underestimate the total AGB by about 2-3%. Finally, equations developed using the third regression approach have slightly smaller ranges of error but they are in general more accurate than equations developed using the second approach.

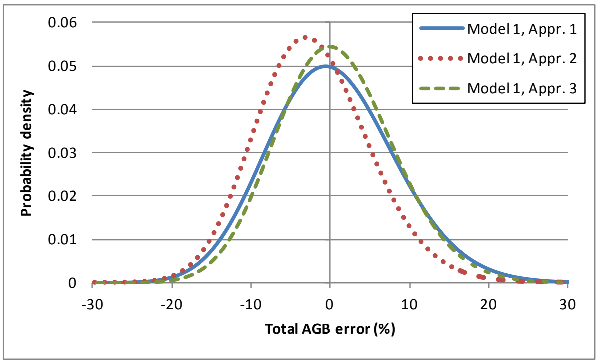
Based on the results of the comparison, the following are recommended; (i) application of the equation derived from Model (1) using the third regression approach when D is the only input variable; (ii) application of the equation derived from Model (3) using the third regression approach when D and H are used as input variables; (iii) application of the equation derived from Model (5) using the third regression approach when D and ρ are used as input variables; and (iv) application of the equation derived from Model (8) using the third regression approach when all three parameters D, H and ρ are used as input variables. Specifically, the following equations are recommended for application:

$AGB = 0.1277 \times D^{2.3943}$ Eq. (1)

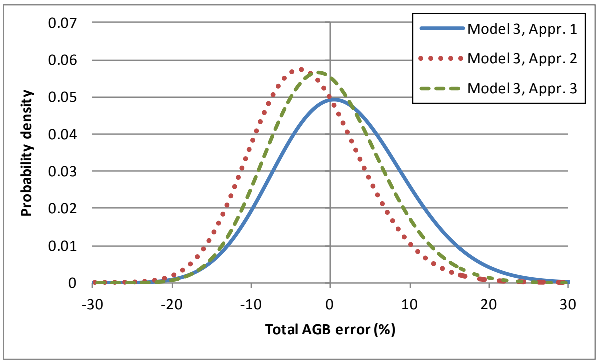
$AGB = 0.0530 \times (D^2H^{0.7})^{1.0072}$ Eq. (2)

$AGB = 0.2328 \times (D^{2.4}\rho)^{0.9933}$ Eq. (3)

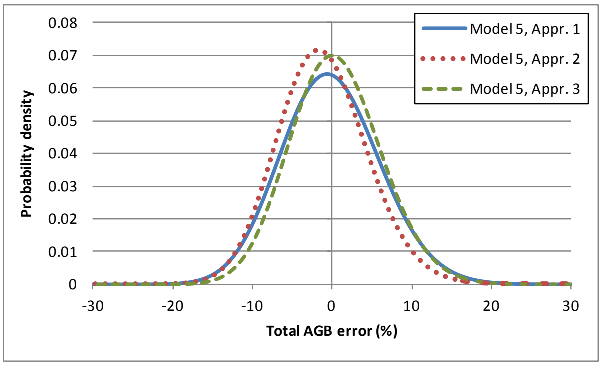
$AGB = 0.0968 \times (D^2H^{0.7}\rho)^{10037}$ Eq. (4)



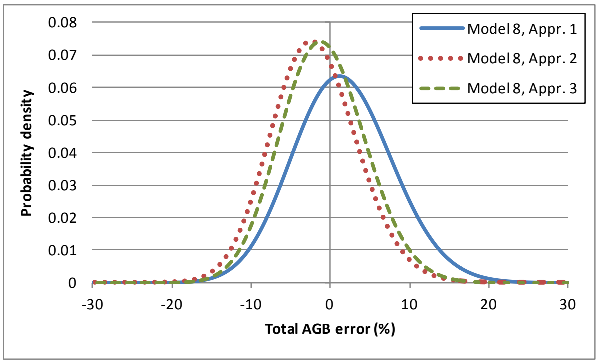
Models that use only variable D



(b) Models that use only D and H



Models that use only D and ρ



(d) Models that use all three input variables

Figure 12: Comparison of models across three regression approaches for each group of inputs (evergreen broadleaf)

Note that the probability density functions of the total AGB error reported in this section are for equations that are derived from a random dataset of 73 (two thirds of 110) trees and predict the total AGB of a random and independent dataset of 37 trees (one third of 110). Normally, the ranges of error of the equations decrease with the size of the training dataset. Eq. (1) to (4) are derived from the whole dataset (i.e., all 110 trees) so they should have smaller ranges of error (i.e., more robust) than those reported in this section. The ranges of error of the equations also depend on the size of the testing dataset. Previously carried out analyses to develop volume equations indicate that with a given model, when the size of the testing dataset increases, the expected value (i.e., the mean) of the error is almost unchanged while the range of error is narrowed (unpublished data). If this holds true for biomass equations, then it can be safe to use the expected values and the ranges of error reported in this section for Eq. (1) to (4) above when predicting the total AGB of 37 or more trees. The expected values and the ranges of error for these equations are given in Table 20.

Table 20: Expected values and ranges of total AGB error for Eq. (1)-(4) when predicting total AGB of 37 or more trees

Eq. No.	Expected value of error (%)	Range of error (95% CL)
Eq. (1)	0.909	-12.78% ÷ 16.76%
Eq. (2)	-0.467	-13.54% ÷ 14.36%
Eq. (3)	0.679	-10.05% ÷ 12.38%
Eq. (4)	-0.666	-10.81% ÷ 10.28%

3.3.3 Modeling of ABG for the main tree families and species

Due to the insufficient number a trees sampled per family and species, the development of allometric equations at family or species level would have lead to unrobust models and has not been done.

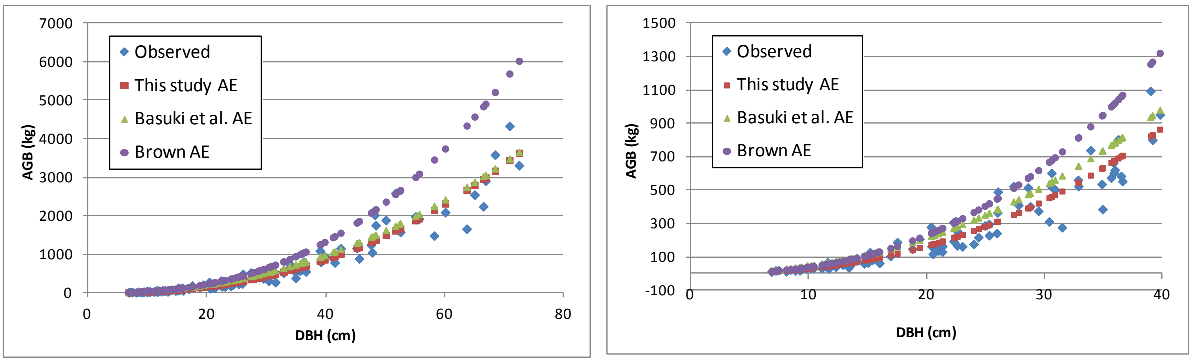
3.3.4 Comparison with generic models

A comparison was undertaken between Eq. (1) (which uses DBH as the only input variable) with two other equations. The first equation is that of Brown (1997) (Eq. (Brown)) for all tropical moist forests; and the second equation is that of Basuki et al. (2009) (Eq. (Basuki)) for mixed species in tropical lowland Dipterocarp forests.

$AGB = \exp(-2.134 + 2.530\times\ln(DBH))$ Eq. (Brown)

$AGB = \exp((-1.201 + 2.196\times\ln(DBH))$ Eq. (Basuki)

The result is shown in Figure 13.



For all trees

For trees with DBH < 40 cm

Figure 13: Comparison of the Model (1) fitted equation, with equations of Basuki *et al.* (2009) and Brown (1997) (evergreen broadleaf dataset)

- It is observed that the equation of Brown seems to significantly overestimate the AGB of trees for the dataset of this Study. Thus, it should be used with care when estimating forest biomass in Vietnam. The Basuki *et al.* equation, though closer to the equation developed in this Study, still overestimates the AGB of many trees for the dataset of this Study. The differences between results of the equation of Basuki et al. and Eq. (1) are larger for small trees (Figure 13b). This is explained by the fact that the Basuki *et al.* equation is developed specifically for the tropical lowland Dipterocarp forests.

- Next, Eq. (4) was compared with the equation developed by Chave *et al.* (2005) (Eq. (Chave)).

AGB = 0.0509×D^2Hρ Eq. (Chave)

- The result is shown in Figure 14. It is observed from the figure that the equation of Chave *et al.* also overestimates the AGB of trees for the dataset of this Study.

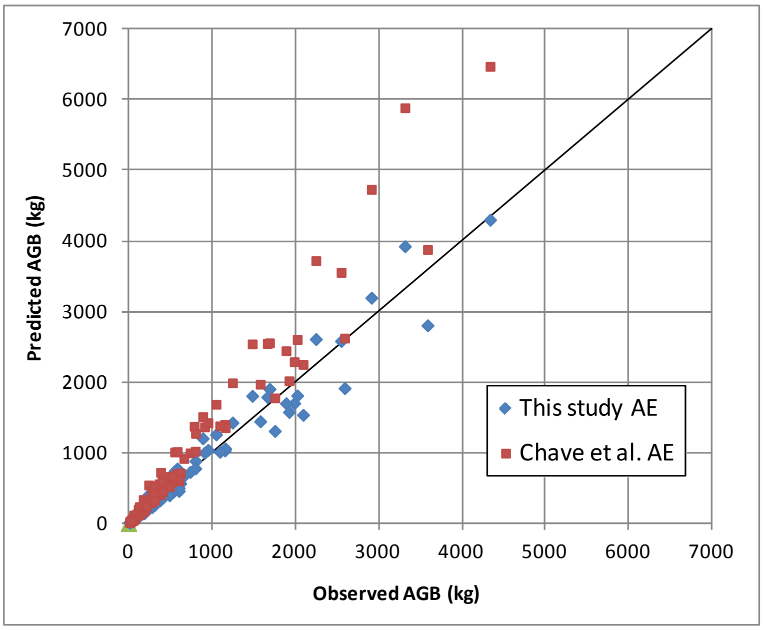


Figure 14: Comparison between Eq. (4) and the equation of Chave et al. (2005) (evergreen broadleaf dataset)

- Finally, the current dataset was used to calculate the average deviation S̄ (%) and the total AGB error S (%) for different equations, either developed in this study or previously developed. S̄ is calculated using Formula (6) below:

$$\bar{S}(\%) = \frac{100}{n}\sum_{i=1}^{n}\frac{|\hat{Y}i - Yi|}{Yi} \qquad \text{Formula (6)}$$

- Where *n* is the number of sample trees; Ŷ*i* and Y*i* are the predicted and measured AGB of the *i*th tree, respectively. The results are provided in Table 21.

Table 21: The standard deviation of different equations (evergreen broadleaf dataset)

Equation No.	Equation	S̄(%)	S(%)
Eq. (1)	AGB = 0.1277×$D^{2.3943}$	20.94	0.61
Eq. (2)	AGB = 0.0530×$(D^2H^{0.7})^{1.0072}$	19.40	-0.75
Eq. (3)	AGB = 0.2328×$(D^{2.4}ρ)^{0.9933}$	17.02	0.55
Eq. (4)	AGB = 0.0968×$(D^2H^{0.7}ρ)^{10037}$	15.46	-0.77
Eq. (Basuki)	AGB = exp((-1.201 + 2.196×ln(D))	39.28	11.17
Eq. (Brown)	AGB = exp(-2.134 + 2.530×ln(D))	53.15	57.16
Eq. (Chave)	AGB = 0.0509×D^2Hρ	30.25	35.75

-

- As is observed, Eq. (4) from this study has the smallest $\bar{S}$ (15.46%), followed by Eq. (3), (2) and (1), in that order. For the three previously developed equations, Eq. (Chave) performs the best. This is explained by its use of all three input variables while the other two use only the variable D. Eq. (Basuki) while appearing similar to Eq. (1) in Figure 13a, gives an $\bar{S}$ value of 39.28%. Eq. (Brown) has the largest $\bar{S}$ (53.15%).
- For the total AGB error *S*, equations developed in this study have the lowest values and are all under 1%. Eq. (Basuki) overestimates the total AGB of the current dataset by 11.17%. Eq. (Chave) and Eq. (Brown) overestimate the total AGB by 35.75% and 57.16%, respectively.

3.4 Result 4: BEF

The proportion of dry biomass for each component of the trees are given in Figure 15. It is observed that the stem accounts for a large proportion (77.8%) of the total above-ground biomass. Leaf biomass accounts for just 1.2% of the total above ground biomass.

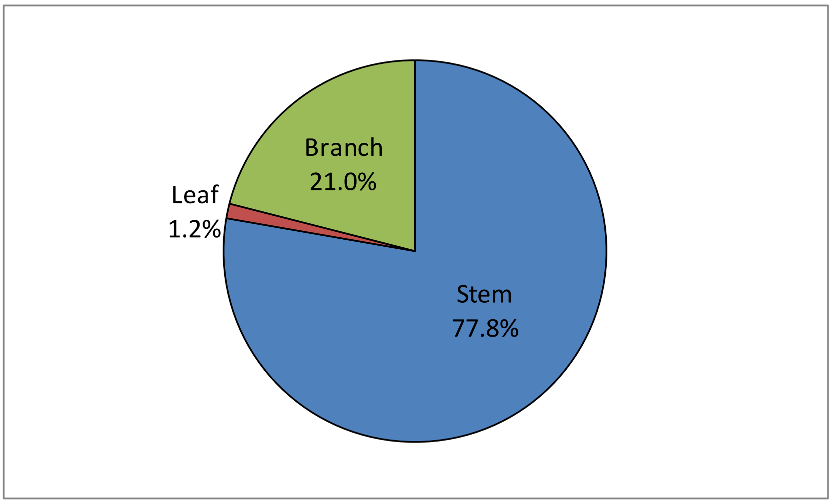


Figure 15: Proportion of dry biomass of each tree component (evergreen broadleaf)

Then the BEF has been calculated for each sampled tree. The result for the 110 trees sampled in evergreen broadleave forest is a BEF average value of 1.256 ± 0.145. The minimal value is 1.035 and the maximal is 1.856.

4 RESULTS AND DISCUSSIONS FOR DECIDUOUS FORESTS

4.1 Result 1: forest and trees characteristics

4.1.1 Species composition

A total of 29 species were found in the deciduous sample plot. Table 22 gives the list of the 10 most dominant species. The full list of species composition is provided in Annex 6. Based on the IV% index, the most dominant species are *Dipterocarpus intricatus,* which accounts for 58.6% of the number of trees and 65.7% of the total basal area*. Shorea obtusa* accounts for 15.3% of the total number of trees and 12.9% of the total basal area and is the second dominant species. Other important species include *Mangifera minitifolia* (IV% = 3.52%), *Terminalia alata* (IV% = 3.28%), *Xylia xylocarpa* (IV% = 3.08%), *Diospyros maritima* (IV% = 2.94%), and *Irvingia malayana* (IV% = 1.88%).

Table 22: Species composition (deciduous sample plot)

No.	Local Name	Scientific Name	N	G	%N	%G	IV%
1	Dầu lông	Dipterocarpus intricatus	430	13.613	58.58	65.69	62.14
2	Cà chắc	Shorea obtusa	112	2.684	15.26	12.95	14.11
3	Xoài rừng	Mangifera minitifolia	13	1.092	1.77	5.27	3.52
4	Chiêu liêu khế	Terminalia alata	23	0.709	3.13	3.42	3.28
5	Căm xe	Xylia xylocarpa	29	0.458	3.95	2.21	3.08
6	Cẩm thị	Diospyros maritima	29	0.400	3.95	1.93	2.94
7	Kơ nia	Irvingia malayana	11	0.468	1.50	2.26	1.88
8	Cà đuối	Cryptocarya petelotii	11	0.155	1.50	0.75	1.12
9	Trâm	Syzygium sp.	8	0.236	1.09	1.14	1.12
10	Thành ngạnh	Cratoxylon pruniflorum	9	0.067	1.23	0.32	0.77
Total					91.96	95.94	93.96

4.1.2 Forest structure

In the deciduous sample plot BT-PD-03, all live trees with DBH ≥ 5 cm are measured. There are a total of 734 trees in the sample plot. The N-D distribution of these trees is shown in Figure 16. It is observed that the number of trees decreases as the DBH increases.

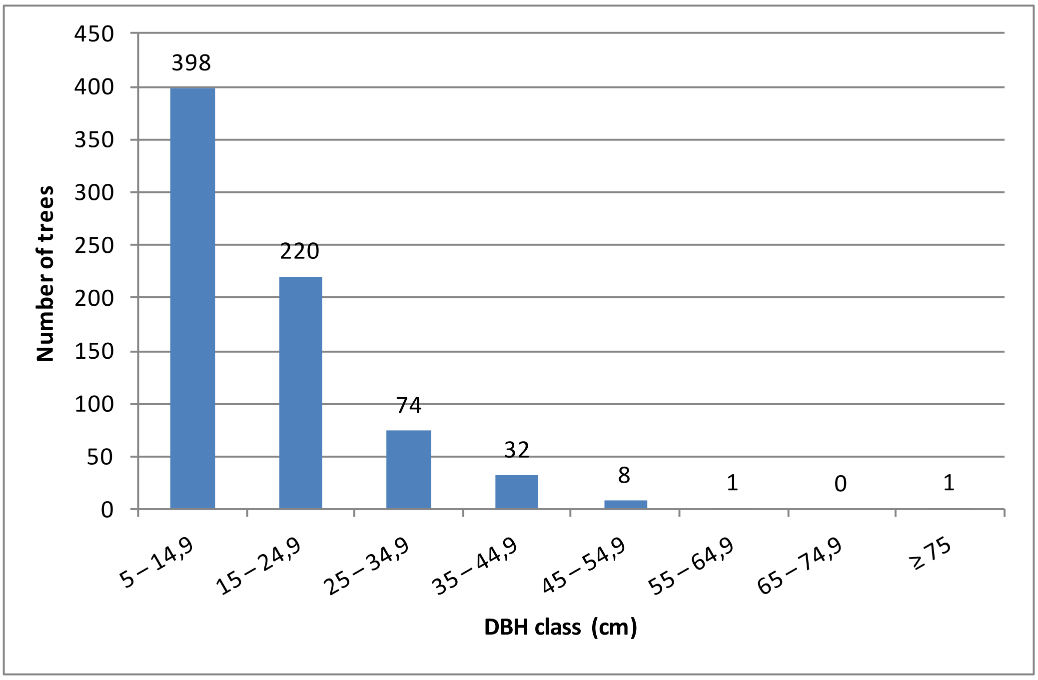


Figure 16: N-D distribution (deciduous sample plot)

4.1.3 Relation between H and diameter

Regression analysis for the logarithm function was done using the SAS software to develop the D-H correlation function of the 55 felled trees in the deciduous sample plot. The resulting equation is H = 7.866×ln(DBH) – 9.661 (R^2 = 0.873, F value = 365.9, p < 0.001; Figure 17). It is observed that there is a quite strong relationship between the two parameters. This means that the inclusion of H may not significantly improve the accuracy as well as the robustness of the prediction.

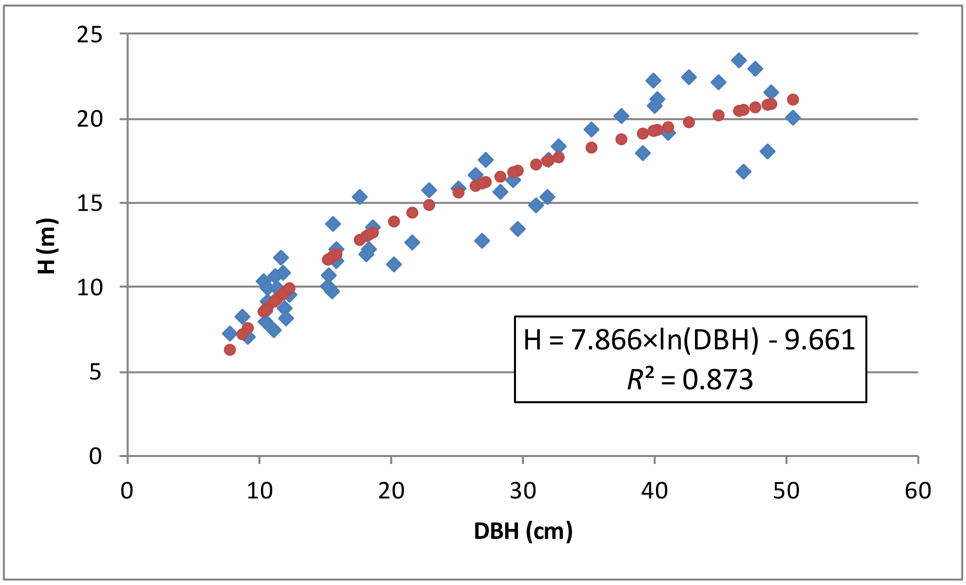


Figure 17: D-H correlation function of felled trees (deciduous)

4.1.4 Biomass of sample trees

The results of dry mass analysis of 55 sample trees are given in Table 23. Stems have the highest average ratio and branches rank second. The coefficient of variation (CV, %) of the ratios is smallest in stems and highest in leaves.

Table 23: Ratio of dry to fresh biomass (deciduous)

Statistical values	Dry to fresh mass ratio		
	Stem	Branch	Leaf
Min	0.464	0.447	0.286
Max	0.610	0.569	0.428
Avg	0.555	0.523	0.356
Stdev	0.032	0.031	0.041
CV(%)	5.692	5.870	11.558

From the data of fresh biomass and the dry-to-fresh mass ratio data, the dry biomass of each component of the trees are calculated. The fresh biomass, dry-to-fresh mass ratios and the converted dry biomass data of each tree component is given in Annex.

4.1.5 Wood density analysis

Table 24 below provides a summary of wood density analysis results for the species in the sample dataset. The average wood density for this dataset is 0.601±0.064 g/cm^3. It is observed that all the wood densities analyzed in this Study are marginally lower than the known values. Similar to that for evergreen broadleaf forests, one of the reasons may be that some of the wood densites collected in the initial phases[4] are dried wood densities (i.e., calculated by dividing oven-dried mass to dried volume) while in this Study the wood densities are calculated by dividing the oven-dried mass to the green volume over bark of the samples. Another reason may be due to differences in the method of analyzing the wood densities in different laboratories such as using different volume measurement methods. The wood densities of all 55 deciduous sample trees are given in Annex 7.

Table 24: Results of wood density analysis (deciduous)

No	Local Name	Scientific Name	N	Basic wood density (g/cm3)			Known value*
				Min	Max	Avg	
1	Bồ hòn	Sapindus saponaria	1	0.631	0.631	0.631	0.840
2	Căm xe	Xylia xylocarpa	6	0.608	0.689	0.651	1.150
3	Cà chắc	Shorea obtusa	3	0.583	0.650	0.615	1.060
4	Cà đuối	Cryptocarya petelotii	2	0.568	0.603	0.585	
5	Chiêu liêu khế	Terminalia alata	3	0.603	0.680	0.633	0.870

[4] By RCFEE.

No	Local Name	Scientific Name	N	Basic wood density (g/cm3)			Known value*
				Min	Max	Avg	
6	Dầu lông	Dipterocarpus intricatus	21	0.532	0.673	0.590	
7	Gáo vàng	Adina pilulifra	2	0.571	0.571	0.571	0.650
8	Kơ nia	Irvingia malayana	6	0.687	0.735	0.705	0.980
9	Nhàu	Morinda citrifolia	1	0.463	0.463	0.463	
10	Thành ngạnh	Cratoxylon pruniflorum	1	0.625	0.625	0.625	
11	Trắc	Dalbergia cochinchinensis	3	0.554	0.632	0.588	
12	Trâm	Syzygium sp.	4	0.464	0.514	0.487	
13	Vừng	Careya arborea	3	0.501	0.559	0.530	
All trees			56	0.463	0.735	0.601	

* Source: Wood density database collected by RCFEE in the initial phase of this Study.

4.2 RESULT 2: Modeling of the stem volume

The stem volume has not been measured during the field work so no model has been developed.

4.3 RESULT 3: Modeling of Aboveground biomass

4.3.1 Modeling per tree compartments

Allometric equations for each component (stem, branches and leaves) of the tree are also developed. Only the power model which uses the input variable D (Model (1)) is used here. The regression analyses were conducted with the SAS software. The results are given in Table 25, Figure 18, Figure 19, Figure 20 below.

Table 25: Results of second regression approach relating dry biomass (kg) of each part of tree with DBH (cm) for Model (1) (deciduous dataset)

Part of tree	Parameter *a*				Parameter *b*				R^2	Pr > F
	Est.	Std. err.	95% CL		Est.	Std. err.	95% CL			
Stem	0.0543	0.0079	0.0385	0.0700	2.5478	0.0465	2.4545	2.6411	0.9623	<.0001
Branch	0.0108	0.0031	0.0046	0.0170	2.7080	0.0919	2.5237	2.8923	0.8450	<.0001
Leaf	0.0123	0.0036	0.0051	0.0194	1.9918	0.0932	1.8047	2.1788	0.7810	<.0001

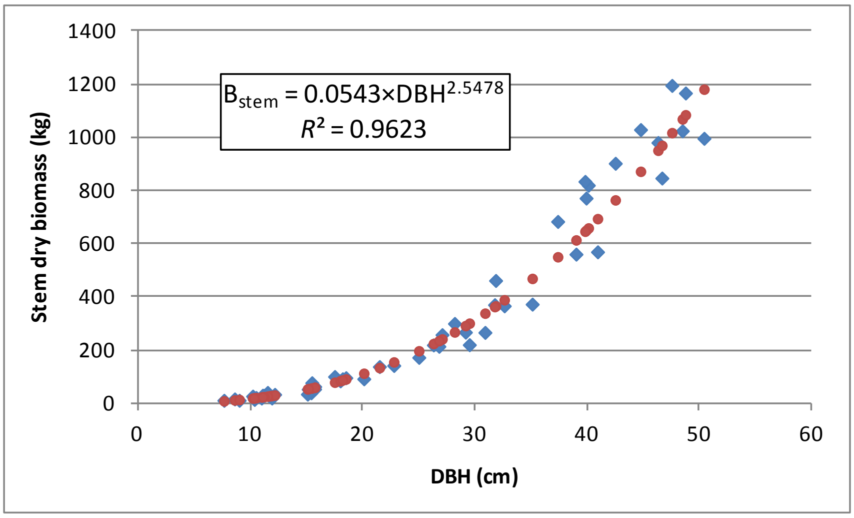


Figure 18: Equation for estimating dry stem biomass (kg) from DBH (cm) (deciduous dataset)

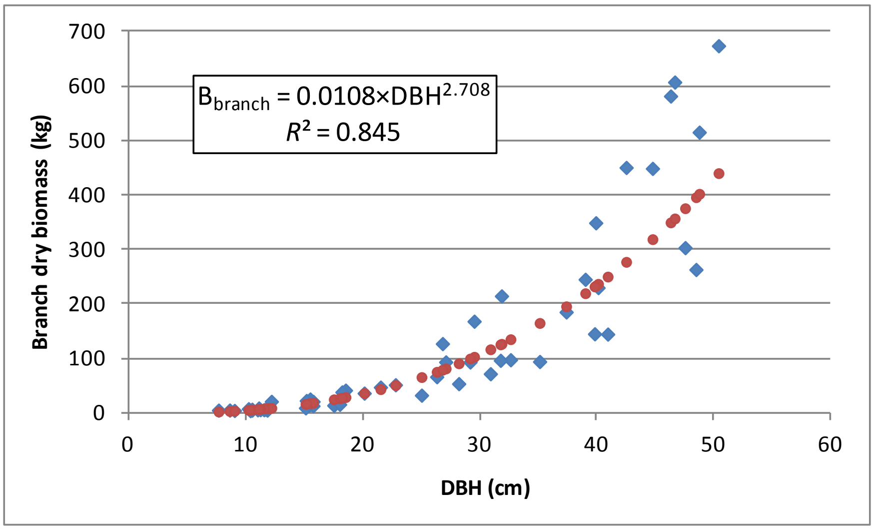


Figure 19: Equation for estimating dry branch biomass (kg) from DBH (cm) (deciduous dataset)

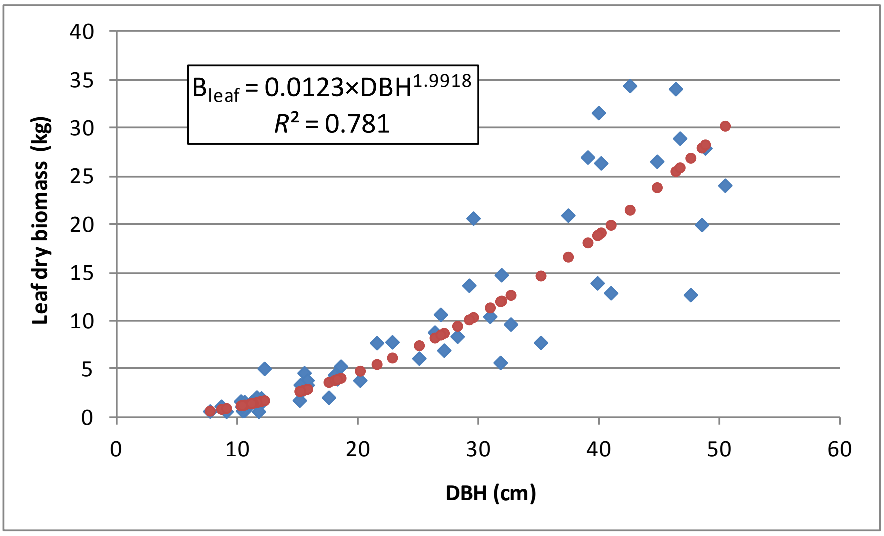


Figure 20: Equation for estimating dry leaf biomass (kg) from DBH (cm) (deciduous dataset)

It is observed that all equations have quite high R^2 values (ranging from 0.781 to 0.962) and are significant at p < 0.001 level and thus can be used in practice. Stem biomass correlates strongest to the DBH, followed by branch biomass. Leaf biomass has the weakest correlation with DBH.

4.3.2 Modeling of total aboveground biomass

Model fitting

First, regression analyses using the first approach for the nine statistical models described in the methodological section are applied using the procedure NLIN in the SAS software. The analyzed results are given in Table 26.

Table 26: Regression analyses using the first approach (deciduous)

Model No	a*	b*	c*	d*	$\bar{R}^2$	SSE	Pr > F
Model (1)	0.0510	2.6703			0.9668	502489.8	<.0001
Model (2)	0.0125	1.0880			0.9740	395145.7	<.0001
Model (3)	0.0154	1.1682			0.9766	356300.7	<.0001
Model (4)	0.0172	2.3932	0.7096		0.9763	352881.7	<.0001
Model (5)	0.0560	1.1655			0.9753	375397.5	<.0001
Model (6)	0.0549	2.7890	1.0595		0.9749	374122.5	<.0001
Model (7)	0.0130	1.1365			0.9827	266931.6	<.0001
Model (8)	0.0159	1.2275			0.9859	219138.6	<.0001
Model (9)	0.0178	2.5237	0.7202	1.1365	0.9858	212207.9	<.0001

* All parameters are significant at p < 0.001.

All equations have very high $\bar{R}^2$ values (ranging from 0.967 to 0.986), indicating that they can all be used to estimate biomass of deciduous forest. The equation derived from Model (1) has the lowest $\bar{R}^2$, as it uses only the predictor D. All models that use H as an additional input variable (i.e. Models (2), (3) and (4)) have a higher value $\bar{R}^2$ as compared to Model (1), suggesting that the inclusion of H can improve the prediction certainty. Among the three models that use only D and H, Model (3) has the highest $\bar{R}^2$. Models that use only D and ρ (i.e. Models (5) to (10)), unlike in evergreen broadleaf forests, do not have higher $\bar{R}^2$ as compared to models that use only D and H, indicating that the inclusion of ρ is less important for deciduous forests. This is understandable as the variation in ρ is smaller in deciduous forests than in evergreen forests. Between the two models that use only D and ρ, Model (5) performs marginally better. Models (7) to (9), which use all three input variables, have the highest $\bar{R}^2$. Among them, Model (8) has the highest $\bar{R}^2$.

Next, regression analyses using the second approach are performed using the procedure NLIN in the SAS software. The analyzed results are provided in Table 27.

Table 27: Regression analyses using second approach (deciduous)

Model No	a*	b*	c*	d*	$\bar{R}^2$	SSE	Pr > F
Model (1)	0.0691	2.5762			0.9819	2.1106	<.0001
Model (2)	0.0292	1.0017			0.9861	1.6223	<.0001
Model (3)	0.0358	1.0745			0.9862	1.6077	<.0001
Model (4)	0.0334	2.0998	0.8369		0.9860	1.6025	<.0001
Model (5)	0.1489	1.0418			0.9857	1.6656	<.0001
Model (6)	0.1352	2.5107	0.9170		0.9855	1.6572	<.0001
Model (7)	0.0613	0.9740			0.9895	1.2278	<.0001
Model (8)	0.0790	1.0425			0.9896	1.2165	<.0001
Model (9)	0.0653	2.0604	0.7972	0.8679	0.9893	1.1975	<.0001

* All parameters are significant at $p < 0.001$.

It is observed that the order of the models (ranked by $\bar{R}^2$) using the second approach is almost the same to that using the first regression approach. There are some small differences. Models that use only D and ρ now have lower $\bar{R}^2$ than models that use D and H, suggesting that for deciduous forests, the inclusion of H is more important than the inclusion of ρ. The best models for each group of input variables are the same with the first regression methods.

Finally, regression analyses using the third approach are done by the procedure NLP in the SAS software. The analyzed results are given in Table 28.

Table 28: Regression analyses using third approach (deciduous)

Model No.	a*	b*	c*	d*	Log*L*	AICc

Model No.	a*	b*	c*	d*	Log*L*	AICc
Model (1)	0.0670	2.5915			-273.16	550.55
Model (2)	0.0296	1.0012			-266.95	538.13
Model (3)	0.0358	1.0764			-266.42	537.08
Model (4)	0.0349	2.1333	0.7859		-266.41	539.29
Model (5)	0.1565	1.0363			-268.63	541.49
Model (6)	0.1256	2.5198	0.7960		-268.17	542.82
Model (7)	0.0658	0.9656			-259.93	524.10
Model (8)	0.0855	1.0330			-259.60	523.43
Model (9)	0.0694	2.0534	0.7831	0.8531	-259.18	527.16

* All parameters are significant at $p < 0.001$.

It is observed that Model (1), which uses only D as the input variable, has the highest AICc value. However, it is still a good predictor as the AICc is only 4.4% far from the best AICc model and it uses only variable D, which can be easily collected. Models that use H as an additional variable (i.e. Model (2), Model (3) and Model (4)) reduce AICc values more than models that use ρ as an additional variable (i.e. Model (5) and Model(6)). This implies that for deciduous forests, the variable H is more important than the variable ρ. This is understandable as in deciduous forests, the variation in ρ is not so large as compared to the variation in H. Among the three models that use only D and H, Model (3) has the lowest AICc. Between the two models that use only D and ρ, Model (5) performs marginally better. Models (7) to (9), which use all three variables, perform the best in terms of AICc. Among these three, Model (8) has the best AICc value and should be used.

Cross validation and error assessment

To avoid over-fitting of the models, cross validation tests were conducted. Table 29 shows the properties of the approximated probability density functions of the total AGB error for every equations developed using the first approach.

Table 29: Properties of probability density functions of total AGB error (%) for equations developed using first regression approach (deciduous)

Model No	α	σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0150	0.1049	0.0042	0.651	0.283	-0.448	0.0572	0.9995	-12.13	15.53	27.66
Model (2)	0.0097	0.0615	-0.0081	-0.642	-0.836	-1.223	0.0634	0.9996	-12.47	12.29	24.76
Model (3)	0.0142	0.0856	-0.0130	-0.657	-0.913	-1.422	0.0671	0.9997	-11.68	11.82	23.50

Model (4)	0.0140	0.0924	-0.0179	-0.968	-1.269	-1.867	0.0617	0.9995	-12.92	12.69	25.61
Model (5)	0.0128	0.0765	-0.0088	-0.454	-0.681	-1.133	0.0676	0.9997	-11.46	11.84	23.30
Model (6)	0.0141	0.0908	0.0000	0.296	0.003	-0.579	0.0623	0.9996	-11.55	13.80	25.35
Model (7)	0.0003	0.0014	-0.0005	-1.688	-1.691	-1.699	0.0775	0.9996	-11.77	8.42	20.19
Model (8)	0.0042	0.0194	-0.0079	-1.850	-1.895	-1.985	0.0863	0.9998	-10.79	7.34	18.13
Model (9)	0.0132	0.0721	-0.0197	-1.280	-1.472	-1.857	0.0749	0.9997	-11.24	9.78	21.02

In this table, the means (or expected values) of error indicate the accuracy while the ranges of error show the robustness of the models. Model (1), which uses only D as the input variable, is quite accurate (mean = 0.651) but has the largest range of error (-12.13% ÷ 17.30%). Models that use H as an input variable (all models except Models (1), (5) and (6)) tend to underestimate the total AGB (i.e., with negative means). Models that use only D and H have similar ranges of error with models that use only D and ρ. However, they seem to be a little less accurate than models that use only D and ρ. Among the three models that use only D and H, Model (3) is the most robust. Between the two models that use only D and ρ, Model (5) has smaller range of error. Finally, Models (7) to (9), which use all three predictors, are the most robust. However, they tend to underestimate the total AGB by about 1.3-1.8% and are the least accurate models. Among these three, Model (8) has the smallest range of error. The probability density functions of total AGB error for the best models of each group of input variables are shown in Figure 21.

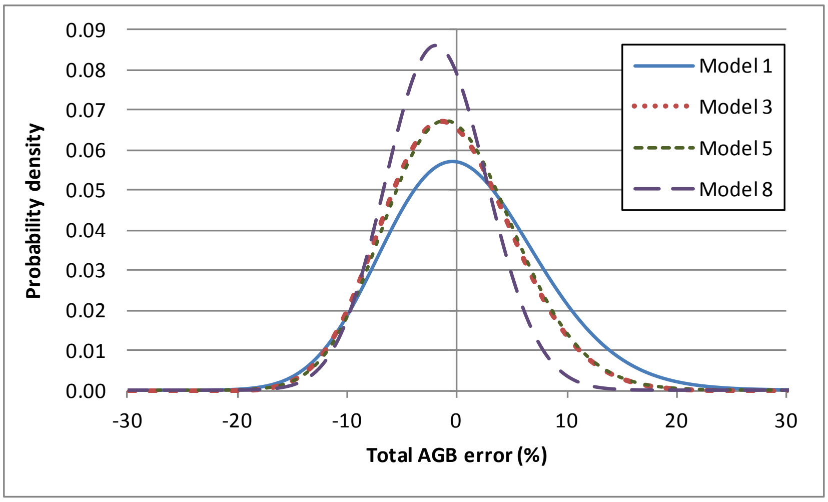


Figure 21: Probability density functions of the total AGB error (%) for selected equations developed by first regression approach (deciduous)

Next, the cross validation test is performed for equations derived using the second regression approach. The results are provided in Table 30.

Table 30: Properties of probability density functions of total AGB error (%) for equations developed using second regression approach (deciduous)

Model No	α	Σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0198	0.1363	-0.0772	-3.315	-3.751	-4.611	0.0632	0.9996	-14.70	10.56	25.26
Model (2)	0.0187	0.1217	-0.0746	-3.471	-3.839	-4.567	0.0667	0.9996	-14.35	9.50	23.85
Model (3)	0.0217	0.1364	-0.0857	-3.388	-3.784	-4.564	0.0698	0.9997	-13.71	9.18	22.89
Model (4)	0.0197	0.1282	-0.0817	-3.596	-3.981	-4.744	0.0671	0.9996	-14.37	9.37	23.74
Model (5)	0.0196	0.1283	-0.0976	-4.369	-4.752	-5.509	0.0676	0.9997	-15.06	8.50	23.56
Model (6)	0.0197	0.1298	-0.0904	-3.992	-4.384	-5.159	0.0669	0.9997	-14.80	9.04	23.84
Model (7)	0.0186	0.1147	-0.0944	-4.514	-4.836	-5.474	0.0717	0.9997	-14.67	7.48	22.15
Model (8)	0.0209	0.1250	-0.1072	-4.518	-4.855	-5.520	0.0750	0.9998	-14.18	7.06	21.24
Model (9)	0.0200	0.1233	-0.0978	-4.312	-4.658	-5.342	0.0719	0.9998	-14.39	7.74	22.13

It is observed that equations developed using the second regression approach tend to underestimate the total AGB by about 3.8-4.8%. Models (1), (2), (3), (4) and (6) have smaller ranges of error than when using the first regression method, but Models (5), (7), (8) and (9) have larger ranges of errors. Among the three models that use only variables D and H, Model (3) has the smallest range of error. Between the two models that use only variables D and ρ, Model (5) has a marginally smaller range of error. Among the three models that use all three input variables, Model (8) is the most robust. The probability density functions of the equations derived from Models (1), (3), (5) and (8), which are the best for each group of input variables, are shown in Figure 22.

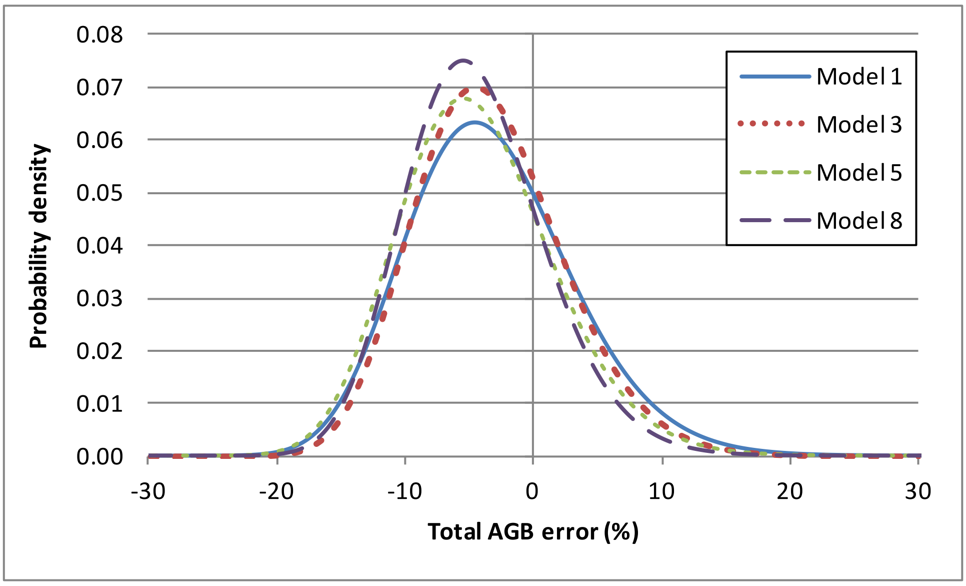


Figure 22: Probability density functions of total AGB error (%) for selected equations developed by the second regression approach (deciduous)

Finally, the cross validation test is performed for equations developed using the third regression approach and the results are provided in

Table 31.

Table 31: Properties of probability density functions of total AGB error (%) for equations developed by the third regression approach (deciduous)

Model No	α	Σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0199	0.1354	-0.0272	-0.895	-1.345	-2.233	0.0609	0.9995	-12.73	13.50	26.23
Model (2)	0.0215	0.1432	-0.0691	-2.656	-3.103	-3.983	0.0649	0.9995	-13.71	10.95	24.66
Model (3)	0.0220	0.1390	-0.0497	-1.784	-2.205	-3.033	0.0669	0.9995	-12.53	11.36	23.89
Model (4)	0.0198	0.1320	-0.0513	-2.105	-2.524	-3.352	0.0636	0.9995	-13.45	11.63	25.08
Model (5)	0.0195	0.1387	-0.0924	-4.075	-4.527	-5.417	0.0621	0.9996	-15.65	10.07	25.72
Model (6)	0.0143	0.1017	-0.0454	-2.747	-3.093	-3.779	0.0592	0.9996	-15.14	11.61	26.75
Model (7)	0.0180	0.1270	-0.1162	-5.700	-6.101	-6.893	0.0640	0.9997	-17.01	7.89	24.90
Model (8)	0.0180	0.1246	-0.1076	-5.263	-5.651	-6.417	0.0649	0.9997	-16.43	8.11	24.54

Model No	α	Σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (9)	0.0119	0.0834	-0.0579	-4.469	-4.746	-5.297	0.0603	0.9995	-16.75	9.38	26.13

Equations derived using the third regression method also tend to underestimate the total AGB by about 0.9-5.7%. They are in general less robust than those derived using the first and second approaches. Models that use all three variables (i.e. Models (7) to (9)), have smaller AICc values (see Table 27), but are less accurate as compared to other models. Their robustness is not better than other models either. One reason may be because maximum likelihood optimization does not perform well when the number of sample trees is small. Similar to the first and second approaches, Models (1), (3), (5) and (8) have the smallest ranges of error within each group of input variables. Their probability density functions are shown in Figure 23.

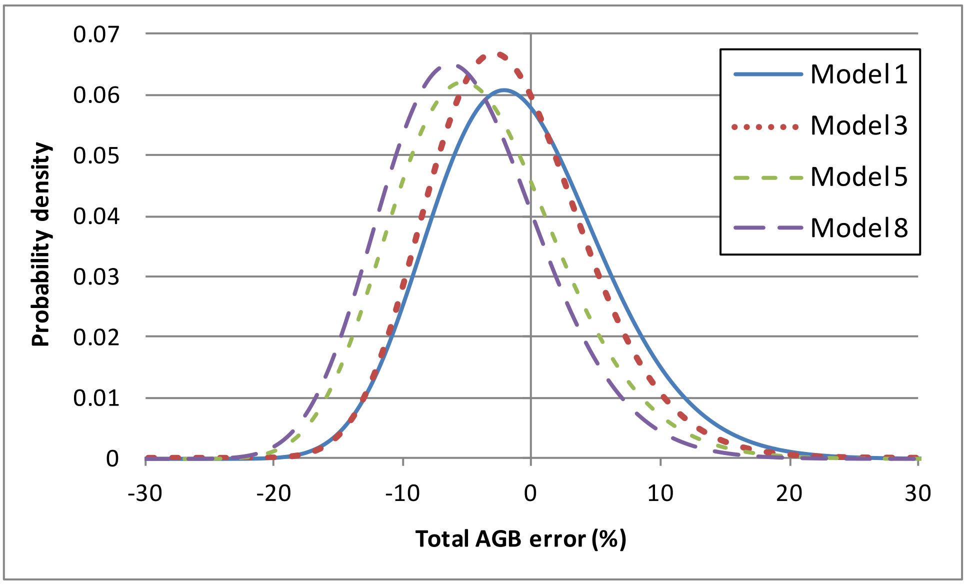
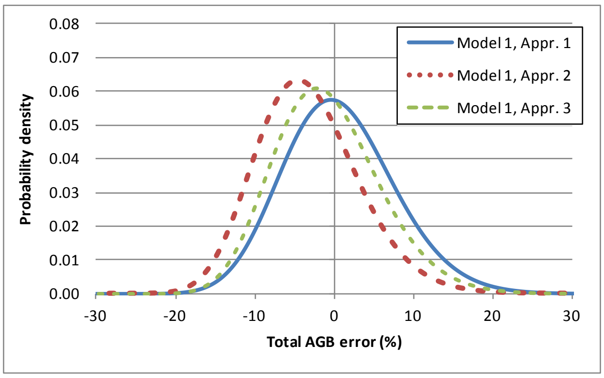
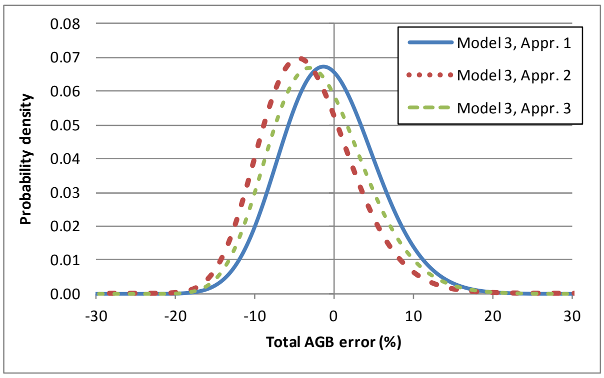


Figure 23: Probability density functions of total AGB error (%) for selected equations developed by the third regression approach (deciduous)

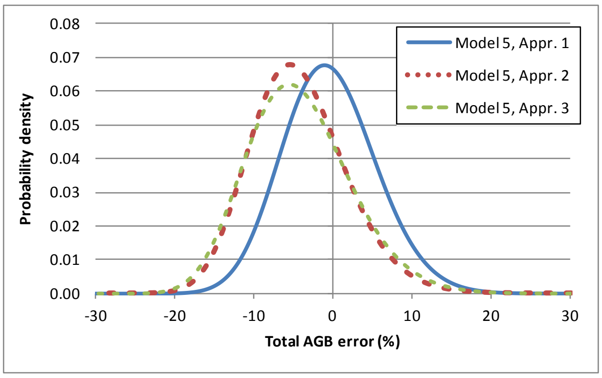
In order to find the best equations for each group of input variables, a comparison of the probability density functions of total AGB error across the three regression approaches were carried out. The results are shown in Figure 24.



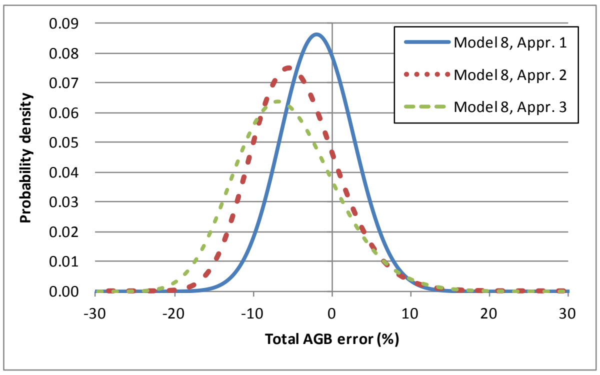
Models that use only variable D



Models that use only D and H



Models that use only D and ρ



Models that use all three input variables

Figure 24: Comparison of models across three regression approaches for each group of inputs (deciduous)

It is observed that, with the same model, equations developed using the first regression approach are the most accurate (i.e., their means of error are closer to zero). When D is used as the only input variable, the second approach is a compromise between accuracy and robustness and should be chosen. For other group of input variables, equations developed using the first regression approach are recommended. Specifically, the following equations are recommended for application:

$AGB = 0.0670 \times D^{2.5915}$ Eq. (5)

$AGB = 0.0154 \times (D^2H^{0.7})^{1.1682}$ Eq. (6)

$AGB = 0.0560 \times (D^{2.4}\rho)^{1.1655}$ Eq. (7)

$AGB = 0.0159 \times (D^2H^{0.7}\rho)^{1.2275}$ Eq. (8)

With similar arguments with the ones in Section 4.1, it is safe to use the expected values and the ranges of error reported in this section for Eq. (5) to (8) above when predicting the total AGB of 19 (⅓ of 55 trees) or more trees. Their expected values and ranges of error are given in Table 32.

Table 32: Expected values and ranges of total AGB error for Eq. (5)-(8) when predicting total AGB of 19 or more trees.

Eq. No.	Expected value of error (%)	Range of error (95% CL)
Eq. (5)	-1.082	-12.92% ÷ 13.33%

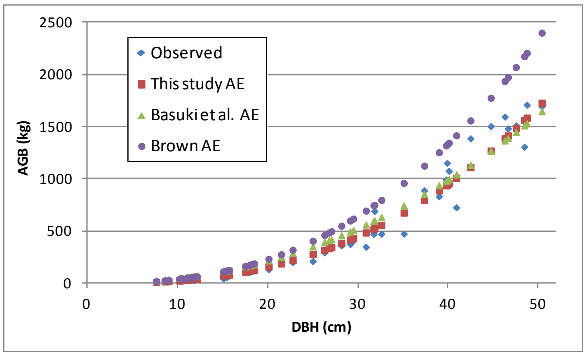
Eq. (6)	-0.913	-11.68% ÷ 11.82%
Eq. (7)	-0.681	-11.46% ÷ 11.84%
Eq. (8)	-1.850	-10.79% ÷ 7.34%

4.3.3 Modeling of ABG for the main tree families and species

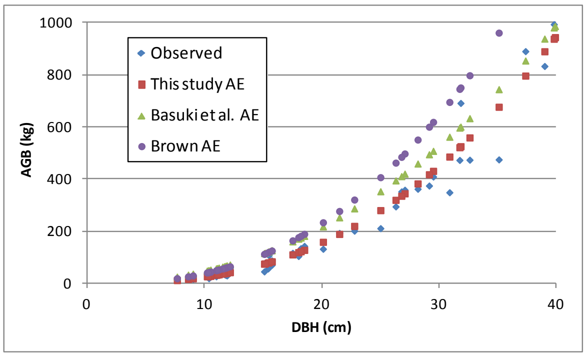
Due to the insufficient number a trees sampled per family and species, the development of allometric equations at family or species level would have lead to unrobust models and has not been done.

4.3.4 Comparison with generic models

Firstly, Eq. (5) (which uses DBH as the only input variable) was compared with two published equations of Eq. (Brown) for all tropical moist forests and Eq. (Basuki) for mixed species in tropical lowland Dipterocarp forests. The result is shown in the Figure 25.



(a) For all trees



(b) For trees with DBH < 40 cm

Figure 25: Comparison between Model (1) fitted equation, and equations of Basuki et al. (2009) and Brown (1997) (deciduous dataset)

It is observed that the Eq. (Brown) significantly overestimates the AGB of trees for the dataset of our Study. Eq. (Basuki), though closer to Eq. (5), predicts larger AGB for trees with DBH < 40 cm (Figure 25b) and smaller AGB for trees with DBH > 45 cm as compared to the equations developed through this Study. One of the reasons for the difference between Eq. (5) and Eq. (Basuki) may be that the dataset contains 11 (20%) hollow trees, which in general have lower AGB as compared to normal trees. Note that for these hollow trees, their appearance are similar to normal trees before cutting down.

Next, Eq. (8), which uses all three input variables, was compared with Eq. (Chave). The result is shown in Figure 26. It is observed that Eq. (Chave) predicts similar values as compared to Eq. (8).

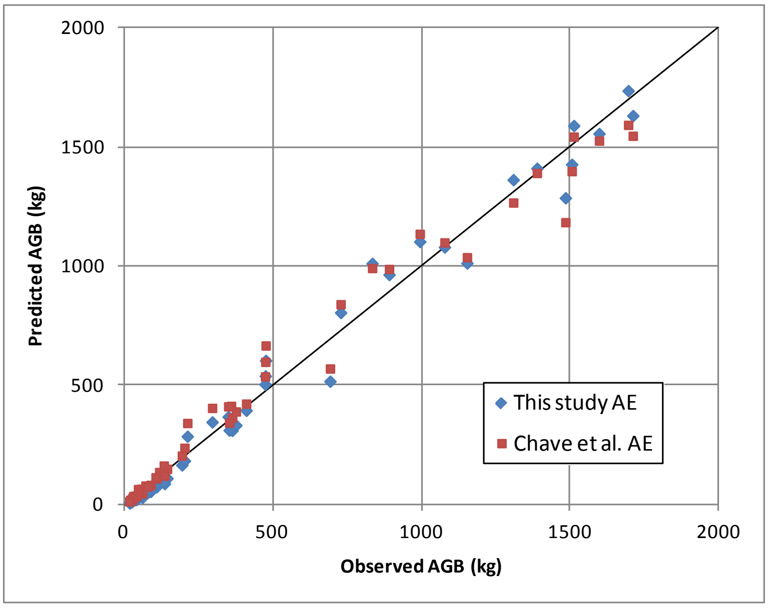


Figure 26: Comparison between Eq. (8) and equation of Chave *et al.* (2005) (deciduous dataset)

Finally, average deviation $\bar{S}$ (%) and the total AGB error S (%) for different equations were calculated. The results are provided in Table 33.

Table 33: The average deviation of different equations for the deciduous dataset

Eq. No.	Equation	$\bar{S}$(%)	S(%)
Eq. (5)	$AGB = 0.0670 \times D^{2.5915}$	16.65	-1.19
Eq. (6)	$AGB = 0.0154 \times (D^2H^{0.7})^{1.1682}$	16.11	-0.79
Eq. (7)	$AGB = 0.0560 \times (D^{2.4}\rho)^{1.1655}$	19.90	-0.75
Eq. (8)	$AGB = 0.0159 \times (D^2H^{0.7}\rho)^{1.2275}$	23.18	-1.53
Eq. (Basuki)	$AGB = \exp((-1.201 + 2.196 \times \ln(D))$	46.19	6.65
Eq. (Brown)	$AGB = \exp(-2.134 + 2.530 \times \ln(D))$	51.78	39.68
Eq. (Chave)	$AGB = 0.0509 \times D^2H\rho$	12.48	1.67

As is observed, Eq. (5) to (8) have relatively small $\bar{S}$ values (ranging from 16.11% to 23.18%). Eq. (Basuki), though appearing similar to Eq. (5) in Figure 24, still gives an $\bar{S}$ value of 46.19%. Eq. (Brown) has the largest $\bar{S}$ (51.78%). Eq. (Chave), has the smallest $\bar{S}$ (12.48%). Furthermore, Eq. (6) to (8) tend to underestimate the AGB of small trees and this seems to be a characteristic of the first regression approach. Noting that the selection of the optimal equations is based on the total AGB error, the first regression approach has performed the best for the subject dataset. In the event the selection of equations is based on $\bar{S}$ (i.e., average absolute error of every single tree) then the equations developed using the third regression approach should be chosen. For example, the equation derived from Model (8) has the $\bar{S}$ value of 12.13%, which is slightly better than Eq. (Chave).

4.4 Result 4: BEF

The proportion of dry biomass for each tree component are given in Figure 27. Proportion of dry biomass of stem, branches and leaves in the dataset are 70.2%, 27.6% and 2.2%, respectively.

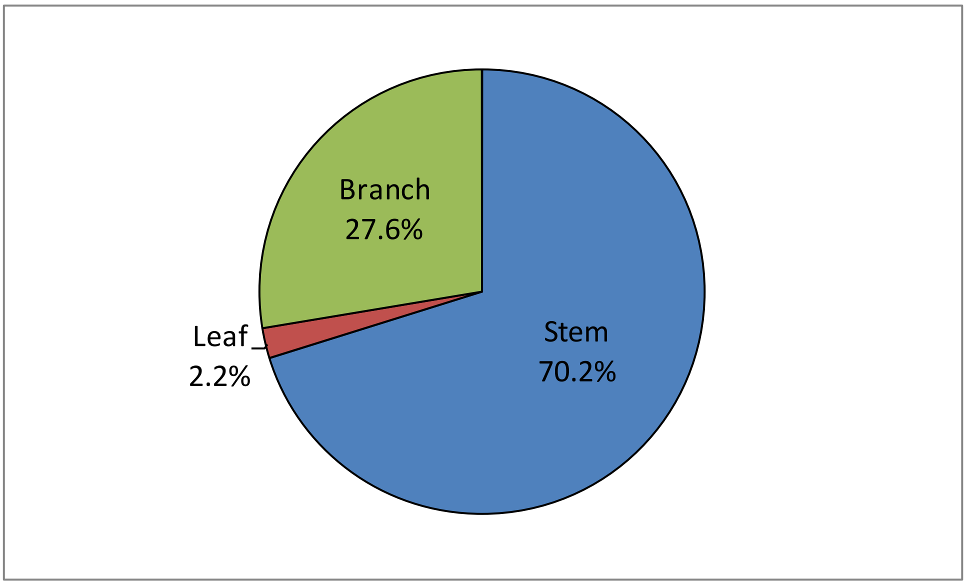


Figure 27: Proportion of dry biomass of each tree component (deciduous dataset)

Then BEF has been calculated for all sampled trees. The result for the 115 trees sampled in evergreen broadleave forest is a BEF average value of 1.396 ± 0.151. The minimal value is 1.160 and the maximal is 1.856.

5 RESULTS FOR BAMBOO (BAMBUSA BALCOA)

5.1 Result 1: forest and trees characteristics

5.1.1 Forest structure

In the *Bambusa balcoa* sample plot BT-PD-04, four sub-plots, each has a size of 20m x 20m, were established and all live bamboos with DBH ≥ 2 cm in the sub-plots were measured. There were a total of 658 bamboos in these 4 sub-plots. Thus, the estimated density of the *Bambusa balcoa* forest is 4,100 bamboos/ha. The N-D distribution of these bamboos is shown in Figure 28. This is a single peak distribution with the peak at the 6.0-6.9 cm DBH class.

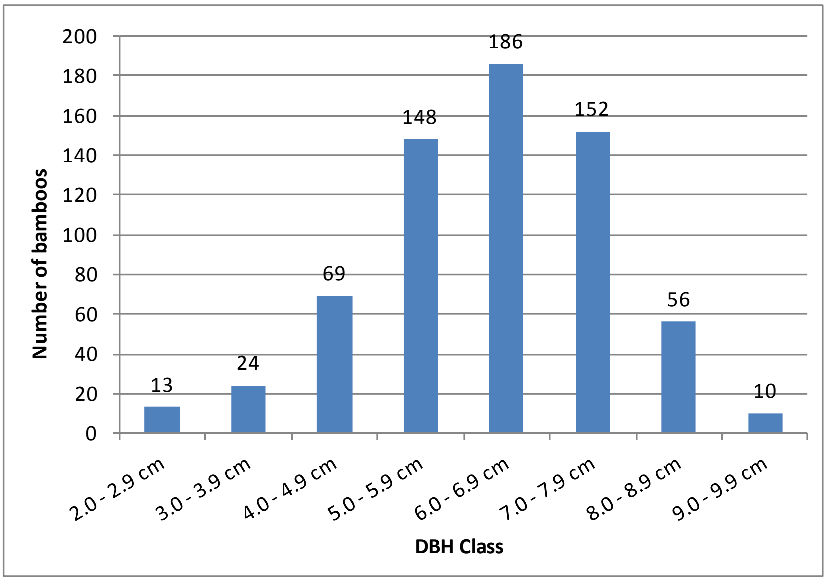


Figure 28: N-D distribution (bamboo sample plot)

5.1.2 Proportion of age classes

The proportion of age classes of the 658 bamboos is given in Figure 29. It is observed that each age class accounts for approximately one third of the total number of bamboos.

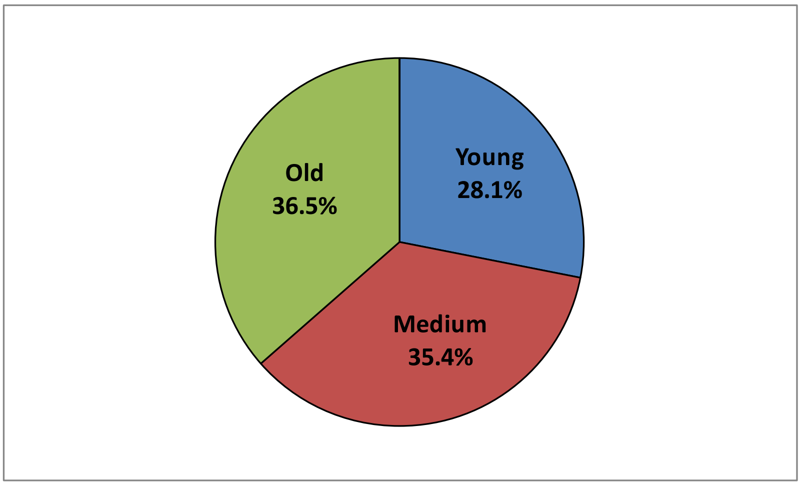


Figure 29: Proportion of age classes (bamboo sample plot)

5.1.3 Relation between H and diameter

A total of 120 bamboos were felled for destructive biomass measurement. The D-H correlation function of the 120 felled bamboos in the sample plot PD-04 is shown in Figure 30. As is observed, the correlation coefficient R^2 of the regression equation is not as high as compared the the D-H correlation functions of evergreen broadleaf and deciduous forests.

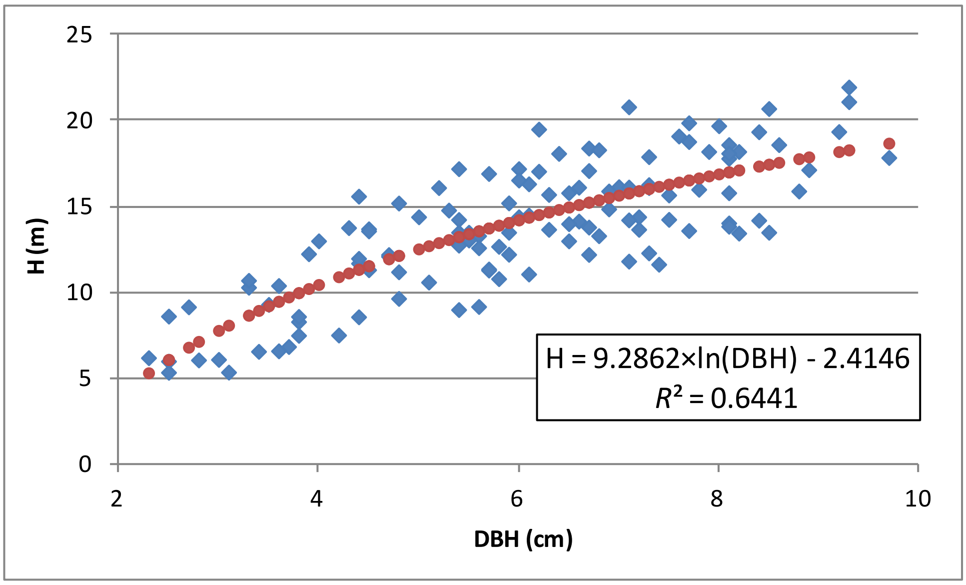


Figure 30: D-H correlation function of the felled bamboos

5.1.4 Biomass of sample trees

Among the 120 bamboos sampled for fresh biomass, 70 were sampled for dry mass analysis. The results of dry mass analysis are given in Table 34 and Table 35. It is observed in Table 34 that in each age class, the dry mass ratio of the stem is always the largest; the ratio of the branch ranks second; and the ratio of the leaf is the smallest. Per tree component (i.e., stem, branch, leaf), the dry mass ratios among the different age classes rank in the order of old, medium, then young. For the 50 trees that were not sampled for dry mass analysis, their dry to fresh biomass ratios are taken from the averages of each age class.

Table 34: Ratio of dry to fresh biomass of different components (bamboo)

Age class	Bamboo part	N	Min	Max	Avg	Stdev	CV(%)
Old	Stem	24	0.511	0.581	0.542	0.020	3.721
	Branch	24	0.453	0.541	0.507	0.022	4.286
	Leaf	24	0.306	0.433	0.358	0.033	9.242
Medium	Stem	24	0.493	0.593	0.538	0.025	4.585
	Branch	24	0.385	0.539	0.483	0.035	7.226
	Leaf	24	0.283	0.422	0.349	0.041	11.840
Young	Stem	22	0.452	0.543	0.512	0.023	4.503
	Branch	22	0.163	0.479	0.388	0.096	24.675
	Leaf	22	0.236	0.387	0.317	0.053	16.699
All	Stem	70	0.452	0.593	0.531	0.026	4.880

Age class	Bamboo part	N	Min	Max	Avg	Stdev	CV(%)
	Branch	70	0.163	0.541	0.461	0.077	16.758
	Leaf	70	0.236	0.433	0.342	0.046	13.368

Table 35 shows the dry to fresh mass ratios at different positions of the stem. It is interesting to note that for all age classes, the ratios are increasing with the heights of the positions. Because the diameter of the stem is larger at lower position, if the dry to fresh mass ratio of the stem is taken as the average ratio of the four positions (as implemented in this study), the dry mass of the stem may be overestimated. In this case, it is better to take the weighted average but the problem of how to select the weighting for each position requires further research.

Table 35: Ratio of dry biomass to fresh biomass of different positions along the stem (bamboo)

Age class	Position	N	Min	Max	Avg	Stdev	CV(%)
Old	0/4L	24	0.456	0.552	0.494	0.026	5.310
	1/4L	24	0.474	0.591	0.532	0.028	5.331
	2/4L	24	0.524	0.623	0.563	0.025	4.424
	3/4L	24	0.542	0.628	0.581	0.024	4.047
Medium	0/4L	24	0.447	0.555	0.491	0.027	5.578
	1/4L	24	0.465	0.595	0.524	0.028	5.417
	2/4L	24	0.489	0.618	0.554	0.036	6.427
	3/4L	24	0.514	0.640	0.581	0.028	4.905
Young	0/4L	22	0.419	0.518	0.465	0.025	5.389
	1/4L	22	0.450	0.554	0.511	0.025	4.874
	2/4L	22	0.481	0.588	0.532	0.031	5.742
	3/4L	22	0.458	0.592	0.541	0.038	7.065
All	0/4L	70	0.419	0.555	0.484	0.029	5.982
	1/4L	70	0.450	0.595	0.523	0.028	5.413
	2/4L	70	0.481	0.623	0.550	0.033	5.985
	3/4L	70	0.458	0.640	0.568	0.035	6.196

5.2 RESULT 2: Modeling of the stem volume

The volume of the bamboo trees has not been measured.

5.3 RESULT 3: Modeling of Aboveground biomass

5.3.1 Modeling per tree compartments

Regression analyses using Model (1) and the third approach were undertaken to develop the equations for calculating the dry biomass of stem, branch and leaf of the bamboos. Cross validation tests are also conducted to estimate the means and ranges of error for these equations. The results are given in Table 36 and Figure 31 and Figure 32 below. The results show that the allometric equations for estimating branch and leaf biomass from DBH have very large ranges of error and should be used with care in practice. An attempt to develop age-class-specific equations for branch and leaf has been made but no clear improvement on the accuracy as well as robustness of prediction was observed (data not shown).

Table 36: Regression analyses using Model (1) and the third regression approach for tree components (bamboos)

Component	Parameter *a*				Parameter *b*				Error mean (%)	Range of error (%)
	Est.	Std. err.	t value	Pr > |t|	Est.	Std. err.	t value	Pr > |t|		
Stem	0.0803	0.0112	7.163	< 0.001	2.2872	0.0730	31.323	< 0.001	0.703	-5.96 ÷ 7.80
Branch	0.0164	0.0043	3.765	< 0.001	1.7734	0.1485	11.939	< 0.001	1.519	-19.09÷ 27.27
Leaf	0.0123	0.0037	3.335	< 0.005	1.4138	0.1705	8.294	< 0.001	1.572	-21.57÷ 31.06

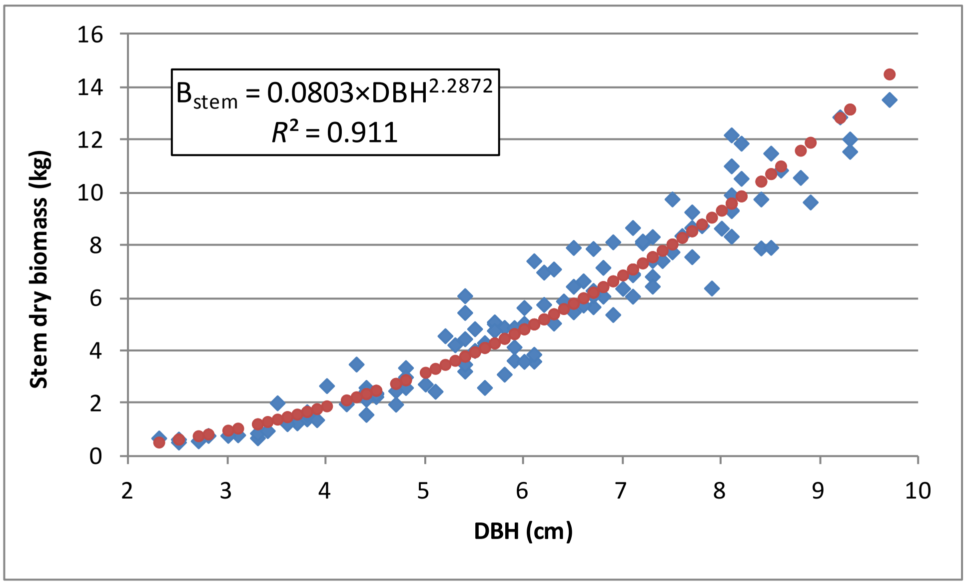


Figure 31: Equation relating dry stem biomass (kg) with DBH (cm) for all age classes (bamboo)

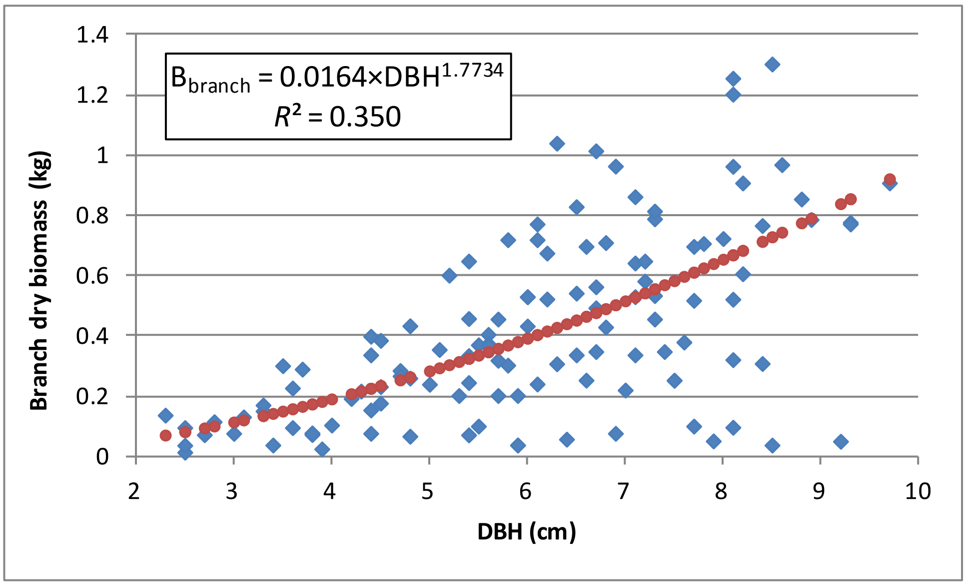


Figure 32: Equation relating dry branch biomass (kg) with DBH (cm) for all age classes (bamboo)

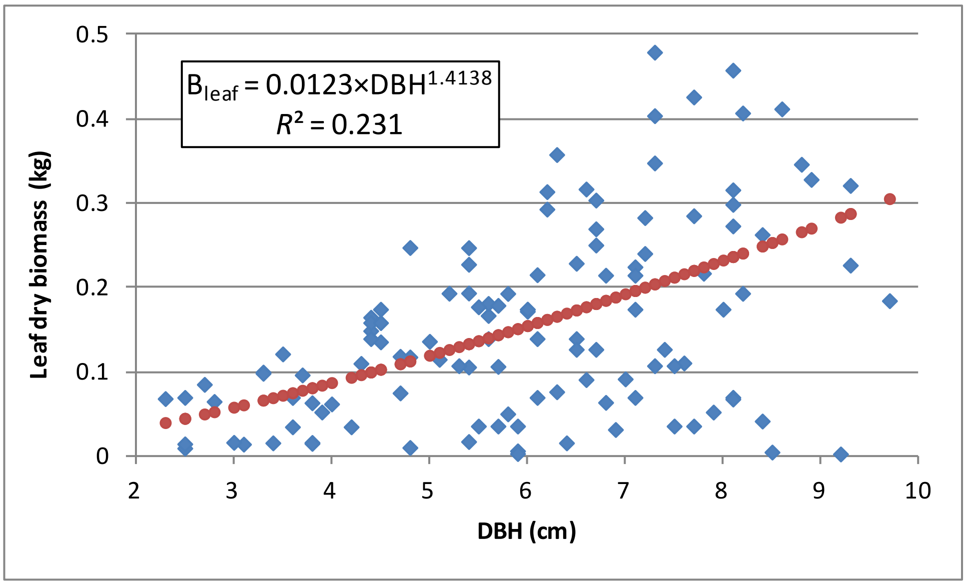


Figure 33: Equation relating dry leaf biomass (kg) with DBH (cm) for all age classes (bamboo)

5.3.2 Modeling of total aboveground biomass

Model fitting

First, regression analyses using the first approach for the four statistical models (Models (1)-(4) in Section 2.5) were undertaken using the procedure NLIN in the SAS software. The analyzed results are given in Table 37.

Table 37: Regression analyses using first approach (bamboo dataset)

Model No	Model	a*	b*	c*	$\bar{R}^2$	SSE	Pr > F
Model (1)	$B = aD^b$	0.1442	2.0365		0.8998	150.26	<.0001
Model (2)	$B = a(D^2H)^b$	0.0607	0.7278		0.8863	170.56	<.0001

Model (3)	$B = a(D^2H^{0.7})^b$	0.0707	0.8050		0.8978	153.35	<.0001
Model (4)	$B = aD^bH^c$	0.0998	1.8580	0.2619	0.9050	141.32	<.0001

* All coefficients are significant at $p < 0.001$.

All equations have a relatively high $\bar{R}^2$ value, indicating that they can all be used to estimate forest biomass. Equations derived from Models (2) and (3), although using both variables D and H, have lower $\bar{R}^2$ as compared to the equation derived from Model (1), which uses only variable D. This indicates that the D^2H and $D^2H^{0.7}$ forms are not suitable for bamboo forest. The equation derived from Model (4) has the highest $\bar{R}^2$.

Next, regression analyses using the second approach were performed using the procedure NLIN in the SAS software. The analyzed results are provided in Table 38.

Table 38: Regression analyses using second approach (bamboo dataset)

Model No	a*	b*	c*	$\bar{R}^2$	SSE	Pr > F
Model (1)	0.0878	2.2877		0.9367	4.5839	<.0001
Model (2)	0.0368	0.8012		0.9305	5.0298	<.0001
Model (3)	0.0436	0.8862		0.9382	4.4734	<.0001
Model (4)	0.0589	2.0224	0.3345	0.9422	4.1475	<.0001

* All coefficients are significant at $p < 0.001$.

It is observed that, when using the same model, the estimated coefficients vary considerably from the results of the first approach. The order of preference of the models (ranked by $\bar{R}^2$) using the second approach is similar to that using the first approach, with a difference in the order between Model (3) and Model (1), with Model (3) with the higher $\bar{R}^2$ value.

Finally, regression analyses using the third approach was condutcted by the procedure NLP in the SAS software. The analyzed results are given in Table 39.

Table 39: Regression analyses using third approach (bamboo dataset)

Model No	a*	b*	c*	Log*L*	AICc
Model (1)	0.1006	2.2220		-163.47	331.05
Model (2)	0.0380	0.7965		-167.37	338.84
Model (3)	0.0467	0.8755		-161.18	326.46
Model (4)	0.0644	1.9696	0.3426	-157.54	321.28

* All coefficients are significant at $p < 0.001$.

It is observed from the table that the values of the coefficients estimated using the third approach are quite close to those estimated using the second approach. The order of the models ranked by AICc is the same with the order ranked by $\bar{R}^2$ of the second regression approach. Among the four models, Model (4) has the lowest (i.e., best) AICc.

Cross validation and error assessment

To avoid over-fitting of the models, cross validation tests were carried out. Table 40 shows the properties of the approximated probability density functions of the total AGB error for every equations developed using the first approach.

Table 40: Properties of probability density functions of total AGB error (%) for equations developed using the first regression approach (bamboo)

Model No.	α	Σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0125	0.0456	0.0075	0.690	0.606	0.439	0.1087	0.9998	-6.28	8.14	14.42
Model (2)	0.0107	0.0418	0.0071	0.743	0.661	0.497	0.1017	0.9998	-6.72	8.68	15.40
Model (3)	0.0112	0.0414	0.0071	0.712	0.635	0.481	0.1074	0.9998	-6.36	8.22	14.58
Model (4)	0.0121	0.0435	0.0071	0.671	0.592	0.435	0.1103	0.9998	-6.21	8.00	14.21

In this table, the means (or expected values) of error indicate the accuracy while the ranges of error show the robustness of the models. All models tend to overestimate the total AGB by about 0.7%. Model (1), although using only variable D, is more robust than Models (2) or (3), which use both variables D and H. Model (4) is the most accurate and robust but the differences between Model (4) and Model (1) are quite small. The probability density functions of total AGB error for these models are shown in Figure 34.

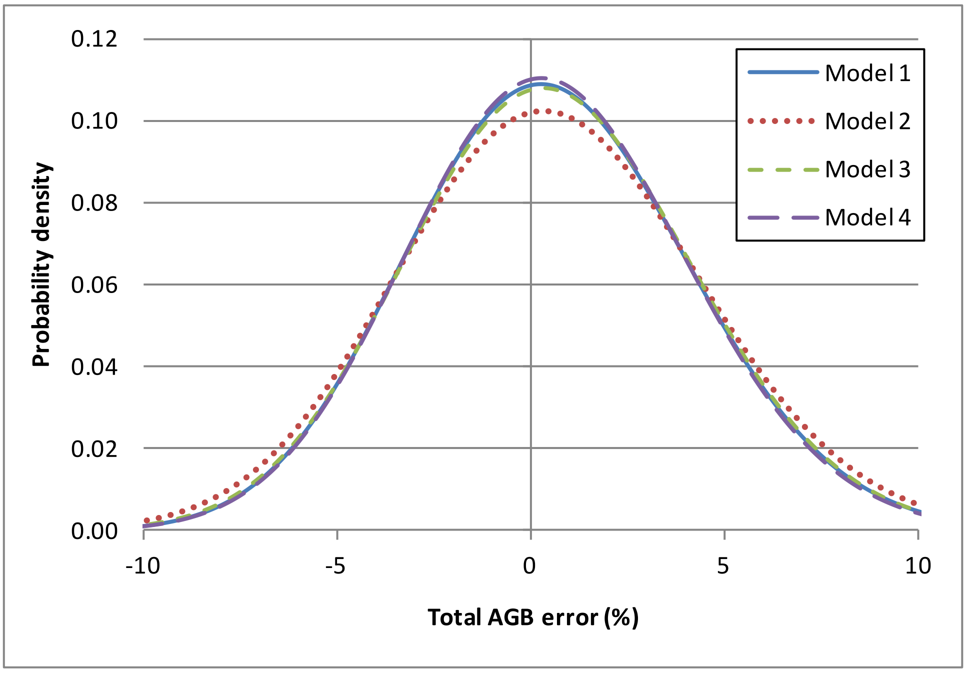


Figure 34: Probability density functions of total AGB error (%) for equations developed by the first regression approach (bamboo)

Next, the cross validation test is performed for equations derived using the second regression approach. The results are provided in

Table 41.

Table 41: Properties of probability density functions of total AGB error (%) for equations developed using the second regression approach (bamboo)

Model No	α	σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0110	0.0431	-0.0136	-1.142	-1.225	-1.391	0.1036	0.9999	-8.47	6.66	15.13
Model (2)	0.0129	0.0536	-0.0169	-1.189	-1.298	-1.516	0.0980	0.9998	-8.88	7.13	16.01
Model (3)	0.0128	0.0507	-0.0148	-1.044	-1.143	-1.340	0.1025	0.9998	-8.41	6.89	15.30
Model (4)	0.0118	0.0460	-0.0127	-0.985	-1.073	-1.250	0.1037	0.9999	-8.29	6.83	15.12

It is observed that equations developed using the second regression approach tend to underestimate the total AGB by about 1%. Once again, Model (1) seems to outperform Models (2) and (3) on the robustness aspect. Model (4) performs marginally better than Model (1) on both accuracy and robustness aspects. The probability density functions of the four models are shown in Figure 35.

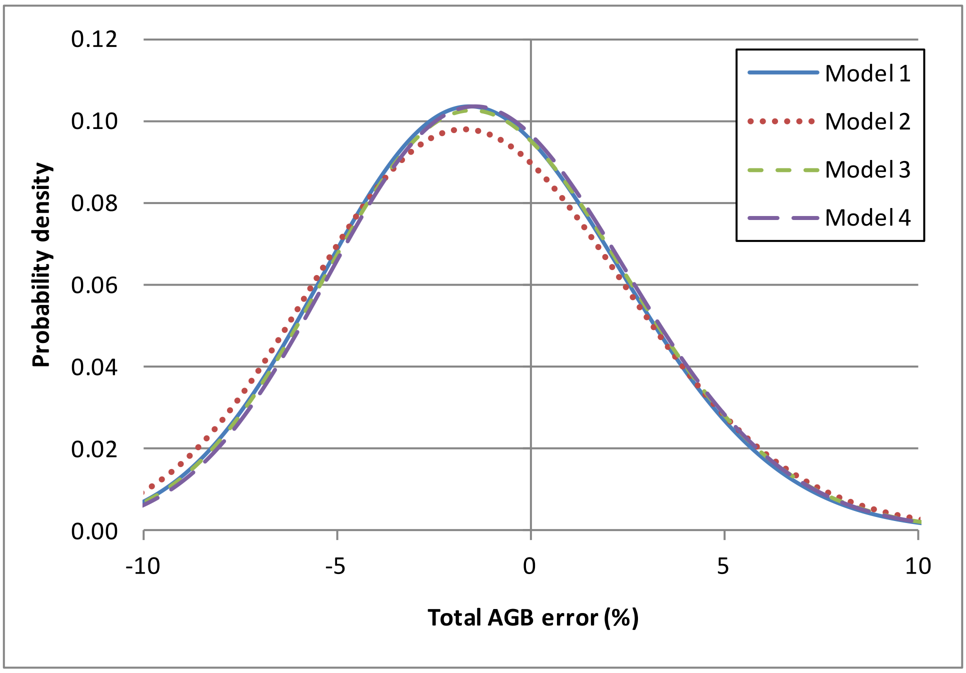


Figure 35: Probability density functions of total AGB error (%) for selected equations developed by the second regression approach (bamboo)

Finally, the cross validation test is performed for equations developed using the third regression approach and the results are provided in

Table 31.

Table 42: Properties of probability density functions of total AGB error (%) for equations developed by the third regression approach (bamboo)

Model No.	α	σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0107	0.0398	0.0027	0.327	0.253	0.105	0.1073	0.9998	-6.76	7.84	14.60
Model (2)	0.0117	0.0457	-0.0173	-1.379	-1.468	-1.643	0.1037	0.9999	-8.69	6.43	15.12
Model (3)	0.0122	0.0456	-0.0059	-0.395	-0.479	-0.648	0.1078	0.9999	-7.43	7.12	14.55
Model (4)	0.0110	0.0400	0.0021	0.265	0.192	0.047	0.1098	0.9998	-6.66	7.61	14.27

Equations derived from Models (1), (3) and (4) are very accurate. Their expected values of error are 0.33%, -0.40% and 0.27%, respectively. Model (4) also has the smallest range of error (i.e., the most robust). The probability density functions of the total AGB error for these equations are shown in Figure 36.

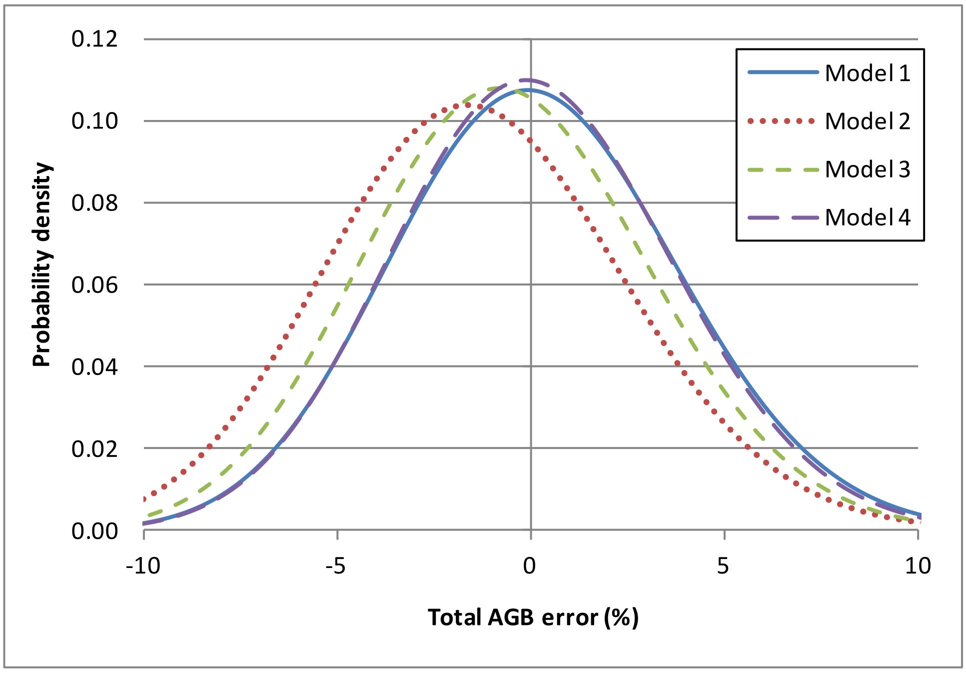
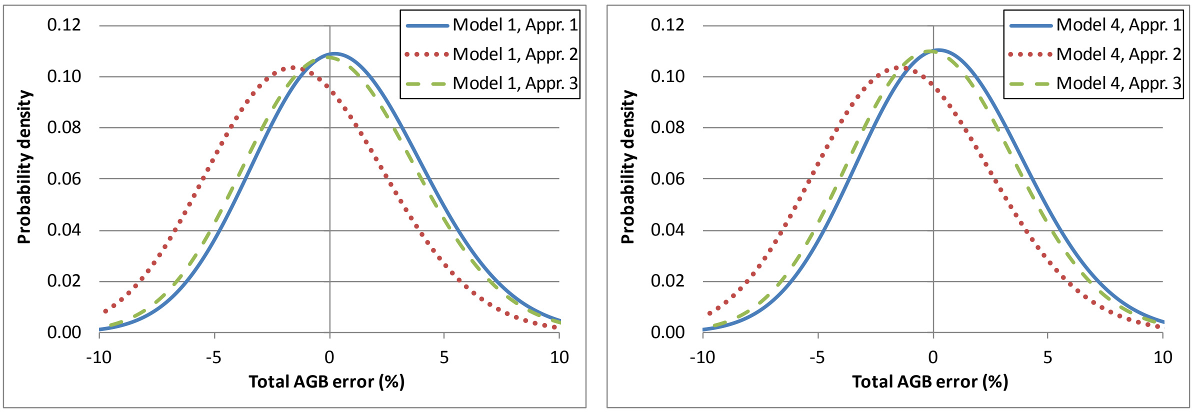


Figure 36: Probability density functions of total AGB error (%) for equations developed using the third regression approach (bamboo)

In order to find the best equations for each group of input variables, a comparison of the probability density functions of total AGB error across the three regression approaches were conducted. The results are shown in Figure 37.



Models that use only variable D (b) Models that use D and H

Figure 37: Comparison of models across three regression approaches for each group of inputs (bamboo)

It is observed from the figure that, with the same model, equations developed using the second regression approach are the least accurate and the least robust. Equations derived using the first regression method are the most robust. Equations developed by the third regression approach, have only marginally larger ranges of error, but are more accurate than those developed by the first regression approach. Therefore, it is recommended to choose both equations developed by the third approach. Specifically, the following equations are recommended for application:

$$AGB = 0.1006 \times D^{2.2220} \quad \text{Eq. (9)}$$

$$AGB = 0.0644 \times D^{1.9696} H^{0.3426} \quad \text{Eq. (10)}$$

Applying the same arguments as for evergreen broadleaf forests, it is considered safe to use the expected values and the ranges of error reported in this section for Eq. (9) and (10) when predicting the total AGB of 40 (⅓ of 120 trees) or more trees. Their expected values and ranges of error are given in Table 43.

Table 43: Expected values and ranges of total AGB error for Eq. (9) and (10) when predicting total AGB of 40 or more trees

Eq. No.	Expected value of error (%)	Range of error (95% CL)
Eq. (9)	0.327	-6.76% ÷ 7.84%
Eq. (10)	0.265	-6.66% ÷ 7.61%

It is observed from the table that the inclusion of H only marginally improves the accuracy as well as the robustness of the prediction. Heights of standing bamboos, however, are difficult to measure with accuracy. Therefore, it is recommended that for bamboo forests, it is not necessary to use the parameter H in biomass prediction.

5.3.3 Modeling of ABG for each age class

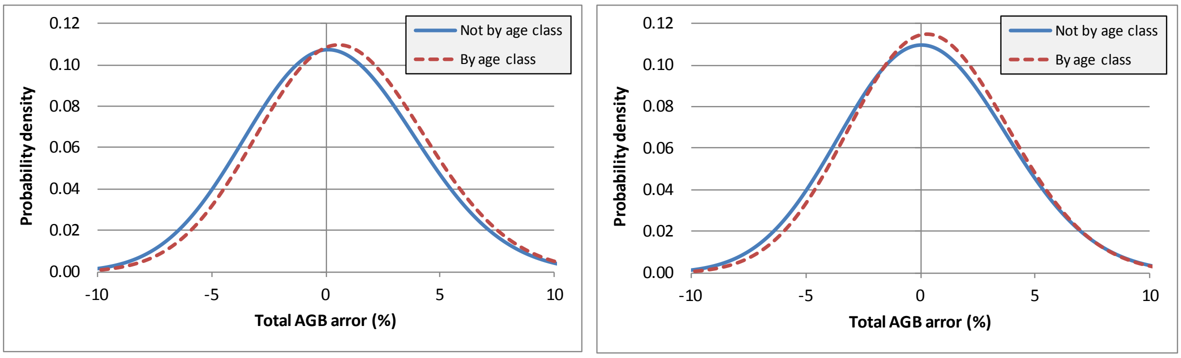
In order to see whether the development of allometric equations specified for each age group of the bamboos can improve the biomass prediction, an experiment to generate the probability density functions for Models (1) and (4) using the third regression approach (since they are proved to be the best for bamboos) for each bamboo age class was conducted. The results are provided in Table 44.

Table 44: Properties of the probability density functions of the total AGB error (%) for equations developed for each age class using the third regression approach (bamboo)

Model No.	α	σ	µ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
Model (1)	0.0111	0.0401	0.0078	0.775	0.702	0.556	0.1095	0.9998	-6.17	8.14	14.31
Model (4)	0.0119	0.0412	0.0051	0.499	0.427	0.284	0.1148	0.9998	-6.12	7.53	13.65

It is observed from the table that when developing equations specified to each age class of bamboos, the ranges of error for Models (1) and (4) has been narrowed from 14.60% and 14.27% (see Table 42) to 14.31% and 13.65%, respectively. This means that the robustness of the prediction has been improved. However, on the accuracy aspect, the approach that use equations specified to each age class is less accurate, with expected values for Models (1) and (4), respectively, 0.78% and 0.50% (as compared to 0.33% and 0.27% when using a general equation for all age classes). There is a tradeoff between accuracy and robustness when using equations developed specifically for each bamboo age class (Figure 38).

It should be noted that the above results are for a dataset that has a quite balanced proportions of bamboos in each age class. In the case of predicting biomass for bamboo forests with an un-balanced proportion of bamboos in each age class, age-class specific equations are recommended to be applied.



Model (1) Model (4)

Figure 38: Comparison of two approaches: (i) using one equation for all age classes and (ii) using three equations specified for each age class (the third regression approach) (bamboo)

The age-class specific equations relating AGB and DBH (developed using Model (1) and the third regression approach) are given in Table 45 and Figure 39. It is observed that with the same DBH, the total AGB of bamboos tends to be highest in the old class, followed by the medium-aged class. The young class has lowest AGB. However, the differences are not substantial.

Table 45: Regression analyses (using Model (1) and third regression approach) divided by each age class (bamboo dataset)

Age class	Parameter *a*				Parameter *b*			
	Estimate	Std. err.	t value	Pr > |t|	Estimate	Std. err.	t value	Pr > |t|
All classes	0.1006	0.0134	7.5221	<0.001	2.2220	0.0702	31.6338	<0.001
Old	0.1428	0.0433	3.2991	<0.005	2.0744	0.1573	13.1877	<0.001
Medium	0.1066	0.0165	6.4606	<0.001	2.2013	0.0842	26.1565	<0.001
Young	0.0645	0.0083	7.7927	<0.001	2.4057	0.0732	32.8663	<0.001

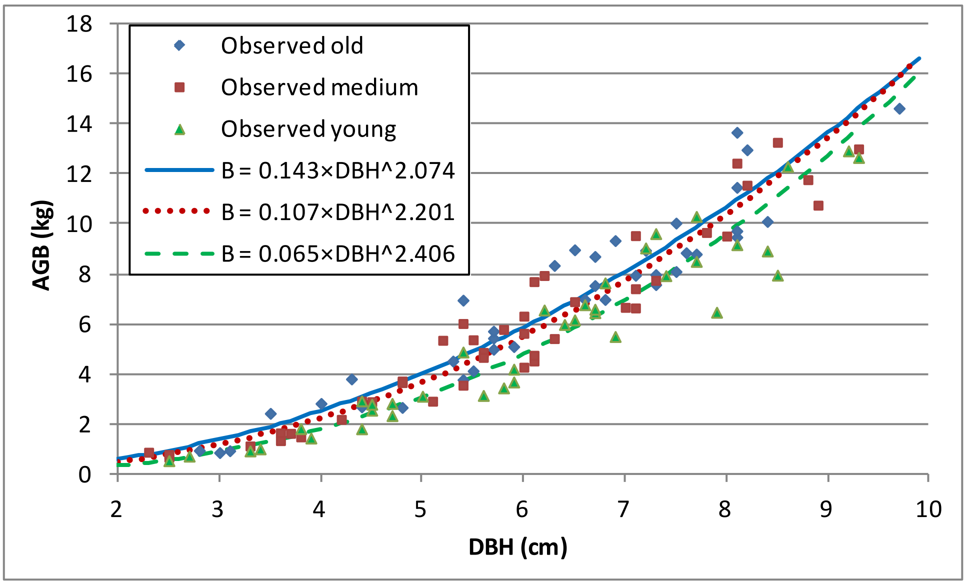


Figure 39: Equations relating AGB (kg) with DBH (cm) for each age class (bamboo)

6 CONCLUSIONS AND RECOMMENDATIONS

This report describes the process of developing biomass allometric equations and biomass conversion and expansion factors for biomass estimation of the evergreen broadleaf, deciduous and bamboo forests in the South East Region of Vietnam. Destructive sampling was done to collect biomass data of sample trees and use these data as dependent variables in the multiple regression analyses. Equations from various different statistical models and regression approaches were developed and compared. For equations developed using the least squares approach, the adjusted R^2 was used for comparison. For equations developed using the maximum likelihood approach, the AICc was used as for comparison. Cross validation tests were conducted to assess the errors of prediction and compare the equations across different regression approaches. For woody forests, the best chosen equations were compared with previouly published equations, including those of Basuki *et al.* (2009), Brown (1997) and Chave *et al.* (2005).

For evergreen broadleaf forests, results of analysis of nine statistical models using three regression approaches have generated the following four equations, as the optimal for each group of input variables:

Equation[1]	Expected value of error[2] (%)	Range of error[3] (95% CL)
$AGB = 0.1277 \times D^{2.3943}$	0.909	-12.78% ÷ 16.76%
$AGB = 0.0530 \times (D^2H^{0.7})^{1.0072}$	-0.467	-13.54% ÷ 14.36%
$AGB = 0.2328 \times (D^{2.4}\rho)^{0.9933}$	0.679	-10.05% ÷ 12.38%
$AGB = 0.0968 \times (D^2H^{0.7}\rho)^{10037}$	-0.666	-10.81% ÷ 10.28%

[1] AGB is the above-ground biomass in kg; D is the diameter at breast height in cm; H is the height in m; and ρ is the wood density in g/cm^3 of the tree.

[2] The error here means the error (in percentage) of the predicted total AGB as compared to the measured total AGB of a set of trees.

[3] These ranges of error apply when predicting the total AGB for datasets of 37 or more trees. For datasets with smaller number of trees, the ranges of error may be larger.

The results also indicated that the inclusion of height and wood density as additional input variables contributes to the improvement of prediction. Therefore, whenever data for these variables are available, the equations using them as variables should be applied. Moreover, the inclusion of wood density improves the robustness of prediction much more than the inclusion of height so wood density should be given the first priority when considering additional variables. The comparison with previously published equations has shown that all three previously published equations tend to overestimate the AGB of trees in the dataset in this study. The total AGB errors of the Eq. (Basuki), Eq. (Brown) and Eq. (Chave) for the current dataset are 11.2%, 57.2% and 35.8%, respectively. This indicates that countries need to develop their own specific equations in order to improve the certainty of biomass prediction and carbon stock assessment.

An attempt was also made to estimate BCEF and BEF for evergreen broadleaf forests. The results show that BCEF and BEF do not depend on DBH but vary around a constant, which is 0.715 for BCEF and 1.256 for BEF.

For deciduous forests, results of analysis of nine statistical models using three regression approaches generated the following four following equations, as the optimal equation for each group of input variables:

Equation	Expected value of error (%)	Range of error[1] (95% CL)
$AGB = 0.0670 \times D^{2.5915}$	-1.082	-12.92% ÷ 13.33%
$AGB = 0.0154 \times (D^2H^{0.7})^{1.1682}$	-0.913	-11.68% ÷ 11.82%
$AGB = 0.0560 \times (D^{2.4}\rho)^{1.1655}$	-0.681	-11.46% ÷ 11.84%

$AGB = 0.0159 \times (D^2H^{0.7}\rho)^{1.2275}$	-1.850	-10.79% ÷ 7.34%

[1]These ranges of error apply when predicting the total AGB for datasets of 19 or more trees. For datasets with smaller number of trees, the ranges of error may be larger.

The inclusion of both height and wood density improves the biomass prediction. Unlike evergreen broadleaf forests, the role of wood density in biomass prediction of deciduous forests is less important. This is because the number of species in deciduous forests is smaller and therefore the variation in wood density is smaller than in evergreen broadleaf forests. Comparison with published equations indicate that while the equations of Basuki *et al.* and Brown overestimate the total AGB of the current dataset by, respectively, 6.7% and 39.7%, the equation of Chave *et al.* adapts very well (the total AGB error is 1.7%) with the current dataset.

Similar to evergreen broadleaf forests, the calculation of BCEF and BEF for deciduous forest shows that they do not depend on DBH but vary around a constant, which is 0.834 for BCEF and 1.396 for BEF.

For bamboo forests, results of analysis of four statistical models using three regression approaches generated the following two equations as the optimal equations:

Equation	Expected value of error (%)	Range of error[1] (95% CL)
$AGB = 0.1006 \times D^{2.2220}$	0.327	-6.76% ÷ 7.84%
$AGB = 0.0644 \times D^{1.9696}H^{0.3426}$	0.265	-6.66% ÷ 7.61%

[1]These ranges of error apply when predicting the total AGB for datasets of 40 or more trees. For datasets with smaller number of trees, the ranges of error may be larger.

The results also show that the inclusion of H only slightly improves the accuracy as well as the robustness of the prediction. Because heights of standing bamboos are difficult to measure with accuracy, for bamboo forests, it is not necessary to include the variable H in biomass prediction.

Age-class specific equations were also developed for bamboos. The analyzed results show that although age-class specific equations help to improve the robustness of biomass prediction, the accuracy is degraded. Therefore, the general equations developed for all age classes should be used for bamboo forests with balanced proportion of bamboos in each age class (i.e., each age class accounts for about one third of the bamboos) and age-class specific equations should be used otherwise.

In order to improve the certainty of biomass prediction in the studied region, the next studies should concentrate on the development of equations and BCEFs specified to each tree family or wood density class. Since the ranges of error of the best models for evergreen broadleaf and deciduous forests are still large (±10%), destructive sampling of more sample trees is also recommended.

REFERENCES

Basuki, T.M., Van Lake, P.E., Skidmore, A.K., Hussin, Y.A., 2009. Allometric equations for estimating the abobe-ground biomass in the tropical lowland Dipterocarp forests. Forest Ecology and Management 257, 1684-1694.

Brown, S. 2002. Measuring carbon in forests: current status and future challenges. Environmental Pollution, 3(116), 363–372.

Brown, S., 1997. Estimating biomass and biomass change of tropical forests: a primer. FAO. Forestry Paper 134, Rome.

Brown, S. and Iverson, L. R., 1992. Biomass estimates for tropical forests. World Resources Review 4, 366-384.

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

Chave, J., Condit, R., Aguilar, S., 2004. Error propagation and scaling for tropical forest biomass estimates. Phil. Trans. R. Soc. Lond. B 359, 409–420.

Dietz, J., Kuyah, S., 2011. Guidelines for establishing regional allometric equations for bimass estimation through destructive sampling. World Agroforestry Center (ICRAF).

FAO, 1998. FRA 2000 Terms and Definition. FRA Working Paper 1. FAO Forestry Department.

Gibbs, H.K., Brown, S., Niles, J.O., Foley, J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environmental Research Letters 2, 13.

Henry, M., Benard, A., Asante, W.A., Eshun, J., Adu-Bredu, S., Valentini, R., Bernoux, M., Saint-Andre, L., 2010. Wood density, phytomass variations within and among trees, and allometric equations in s tropical rainforest of Africa. Forest Ecology and Management Journal 260, 1375-1388.

ICRAF, 2011. Guidelines for establishing regional allometric equations for biomass estimation through destructive sampling.

IPCC, 2006. IPCC Guidelines for National Greenhouse Gas Inventories. Prepared by the Natinal Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T., Tanabe K., (eds). Published: IGES, Japan.

IPCC, 2003. Annex A Glossary. In: Good Practice Guidance for Land Use, Land Use Change and Forestry. Institute for Global Environmental Strategies (IGES). Japan.

Ketterings, Q.M., Richard, C., Meine van N., Ambagau, Y., Palm, C.A., 2001. Reducing uncertainty in the use of allometric biomass equations for predicting above ground tree biomass in mixed secondary forests. Forest Ecology and Management 146, 199-209.

UN-REDD Vietnam & FAO, 2012. Guidelines on Destructive Measurement for Forest Biomass Estimation. Draft version. UN-REDD Vietnam, Hanoi.

UN-REDD, 2011: Measurement, Reporting & Verification (MRV) Framework Document. UN-REDD Vietnam Programme. UN-REDD Vietnam, Hanoi.

Yamakura, T., Hagihara, A., Sukardjo, S., Ogawa, H., 1986. Aboveground biomass of tropical rainforest stands in Indonesian Borneo. Plant Ecology 68 (2), 71–82.

ANNEXES

6.1 Annex 1. Glossary of basic terms

A glossary of the following key terms is adapted from Good Practice Guidance for Land Use, Land Use Change and Forestry[5].

1. Biomass

Organic material both above ground and below ground, and both living and dead, e.g., trees, crops, grasses, tree litter, roots etc. Biomass includes the pool definition for above and below ground biomass.

2. Biomass of forests

Biomass is defined as the total amount of aboveground living organic matter in trees expressed as oven-dry tons per unit area (tree, hectare, region, or country). Forest biomass is classified into above ground biomass and below ground biomass.

Above ground biomass is living biomass above the soil including stem, stump, branches, bark, seeds, and foliage.

Below ground biomass is all living biomass of live roots. Fine roots of less than (suggested) 2 mm diameter are sometimes excluded because these often cannot be distinguished empirically from soil organic matter or litter.

3. Basic wood density

Ratio between oven dry mass and fresh stem wood volume without bark. It allows the calculation of woody biomass in dry matter mass. Basic wood density is normally expressed in g/cm3 or ton/m3.

4. Biomass Conversion and Expansion Factor (BCEF)

Ratio between above-ground biomass in tonnes and growing stock in m^3.

5. Biomass Expansion Factor (BEF)

Ratio between above-ground biomass and biomass of growing stock. This factor is often used to expand biomass of growing stock, or commercial round wood volume, or growing stock volume increment data, to account for non-merchantable biomass components such as branches, foliages, and non-commercial trees.

6. Carbon fraction

Carbon fraction is a carbon content expressed in per cent (%) in dry oven mass of certain component of forests (stem, branches, foliage, root, etc).

7. Carbon pools

Carbon pool is reservoir containing carbon. There 5 carbon pools in a forests considered for forest carbon estimation that are: carbon in live trees (above and below ground), carbon in dead trees and wood, carbon stock in under-storey vegetation (seedlings, shrubs, herbs, grasses), carbon stock in forest floor (woody debris, litter, humus) and soil organic carbon.

8. Carbon stock

Carbon stock is the quantity of carbon in a pool.

---

[5] IPCC, 2003. Annex A Glossary. In: Good Practice Guidance for Land Use, Land Use Change and Forestry. Institute for Global Environmental Strategies (IGES). Japan.

9. Forest

Forest is a minimum area of land of 0.05 – 1.0 hectares with tree crown cover (or equivalent stocking level) of more than 10 – 30 per cent with trees with the potential to reach a minimum height of 2 – 5 meters at maturity in situ (in place). A forest may consist either of closed forest formations where trees of various stories and undergrowth cover a high portion of the ground or open forest. Young natural stands and all plantations which have yet to reach a crown density of 10 – 30 per cent or tree height of 2 – 5 meters are included under forest, as are areas normally forming part of the forest area which are temporarily un-stocked as a result of human intervention such as harvesting or natural causes but which are expected to revert to forest.

FAO provides the definition of a forest which is land spanning more than 0.5 hectares with trees higher than 5 meters and a canopy cover of more than 10 percent, or trees able to reach these thresholds in situ. It does not include land that is predominantly under agricultural or urban land use[6].

10. Root to shoot ratio (RS)

RS is defined as a ratio of below ground biomass to above ground biomass of trees. RS is normally used to estimate below ground biomass of trees if above ground biomass of trees is known.

[6] FAO, 1998. FRA 2000 Terms and Definition. FRA Working Paper 1. FAO Forestry Department.