UN-REDD
PROGRAMME

FAO
UNDP
Empowered lives.
Resilient nations.
UNEP

PART B-3:

Tree allometric equations in Evergreen broadleaf forests in North Central Coastal 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 forests in Quang Binh Province.

By:

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

Msc. Nguyen Van Son, Division of Sciences, Techniques and Technologies

Dr. Nguyen Phu Hung, Deputy Director

Forest Inventory and Planning Institute

Edited by:

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

Recommended citation: Hung, D.N., Son, N.V., Hung, N.P. (2012) Tree allometric equations in Evergreen broadleaf forests in North Central Coastal region, Viet Nam, 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.

ACKNOWLEDGEMENTS

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 all staff members of Department of Agriculture and Rural Development of Quang Binh province, especially the Department of Forestry, Forest Protection Department, and the staff members of Long Dai Industrial-Forestry One-Member Limited Company, Quang Binh province 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

The initiative on reducing emission from deforestation and forest degradation (REDD) was proposed in COP 13 in Bali, Indonesia in 2007 and was formally adopted as a measure contributing to climate change mitigation. REDD is a mechanism being designed to provide financial rewards to forest owners and users for their efforts to protect and develop the forest. To implement this mechanism, countries will need to measure the GHG emissions from deforestation and forest degradation within their borders.

To support this initiative, the UN-REDD Programme is being carried out in a numbers of countries including Vietnam. The UN-REDD Vietnam Programme was started in 2009 and will move to its piloting phase soon. To prepare for the piloting phase, one of the tasks of the first phase is to develop a scientific base, including allometric equations and biomass conversion and expansion factors, for biomass estimation of major forest types in Vietnam.

This report describes the process of developing biomass allometric equations and biomass conversion and expansion factors for biomass estimation of the evergreen broadleaf forests in the Central Region of Vietnam. Destructive sampling was done to collect biomass data of 110 sample trees in two sample plots and these data were used as dependent variables in multiple regression analyses. Within the sample plot, DBH and species names (both scientific and Vietnamese) of all the living trees with DBH bigger than 5 cm were registered. DBH, total height, volume, length of tree bole and stump were measured for the 110 harvested sample trees. Afterwards, samples of stem, branches ad foliage were collected and immediately weighted. The dry weight of the samples was subsequently measured in the laboratory.

Totally nine statistical models were used for the regression analyses. Three regression approaches were applied. The first approach is to use the least squares optimization to the original models. The second approach is to use the least squares optimization to the logarithmically transformed forms of the original models. The third approach is to apply the maximum likelihood optimization to the original models. For equations developed using the least squares method (to both the original or transformed forms), the adjusted R^2 and SSE values are used to measure the goodness of fit. For equations developed using the maximum likelihood method, the Akaike Information Criterion with correction (AICc) is used to measure the goodness of fit.

The results of regression analyses of nine models which use various combinations of the input variables indicate that the inclusion of the height and wood density as additional input variables contributes to the improvement of the goodness of fit. Moreover, the inclusion of the wood density seems to improve the goodness of fit more than the inclusion of the height. Therefore, whenever these variables are available, equations that use them should be used to improve the accuracy and certainty of biomass estimation.

Cross validation tests were undertaken to assess the performance of the developed equations in practice and draw the ranges of errors for them. Based on the results of these tests, the following equations are suggested to be applied in practice:

No	Equation[1]	Expected value of error[2] (%)	Range of error[3]
1	$AGB = 0.1245 \times D^{2.4163}$	0.101	-16.96% ÷ 20.61%
2	$AGB = 0.0421 \times (D^2H)^{0.9440}$	-1.205	-16.67% ÷ 17.70%
3	$AGB = 0.2105 \times (D^{2.4}\rho)^{1.0025}$	0.600	-14.03% ÷ 17.67%
4	$AGB = 0.0704 \times (D^2H\rho)^{0.9389}$	-0.737	-12.32% ÷ 11.70%

[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 best allometric equation that uses DBH as the only input variable is then compared with the Brown (1997), Basuki *et al.* (2009) and Chave et al. (2005) equations. The results show that the Brown's equation tends to over-estimate the biomass of the sample trees in the studied region. The Basuki *et al.* equation slightly under-estimates the biomass of the large sample trees but the differences are not large.

These results lead us to the recommendation that countries need to develop their own specific allometric equations in order to improve the accuracy of biomass and carbon stock assessment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

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
CI	Confidence Interval
COP	Conference of the Parties
DBH	Diameter at Breast Height
EF	Emission Factors
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 Squares 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

In order to prepare for implementing REDD+ in Vietnam, the UN-REDD Vietnam Programme has supported four forestry-related institutions, namely, Forest Science Institute of Vietnam (FSIV), Forest Inventory and Planning Institute (FIPI), Vietnam Forestry University (VFU) and Tay Nguyen University (TNU) to carry out a Study on Development of Allometric Equations for Forest Biomass Estimation. The objective of this Study is to develop a scientific base for forest biomass estimation, including allometric equations and BCEF, for major natural forests types in Vietnam contributing to MRV under REDD+ and to compilation of the national GHG inventory in the land use, land use change and forestry sector.

The Study is implemented in two phases. The initial phase was carried out in 2011. The objectives include (i) reviewing literature relating to biomass estimation, (ii) developing a Guidelines on Destructive Measurement for Forest Biomass Estimation (UN-REDD Vietnam & FAO, 2012), and (iii) testing the Guidelines by carrying out destructive measurement on four pilot plots (two evergreen broadleaf forest plots and two bamboo forest plots).

The second phase is carrying out in 2012. One of the activities in this phase is to conduct forest biomass field measurements for selected forest types in 8 selected provinces. The Forest Inventory and Planning Institute (FIPI) has been assigned to conduct destructive measurement on two evergreen broadleaf forest in Quang Binh province of the Central Region.

Furthermore, FIPI has been assigned to carry out development of allometric equations for biomass estimation in the Central Region using data of its allocated plots. The objectives of the present work were to collect data for regression analysis and to develop biomass and volume tree allometric equations and Biomass expansion (BEF) and Biomass Conversion and Expansion Factor(BCEF) for evergreen broadleaf forests in Quang Binh Province.

FIPI is also responsible for carrying out error assessment for their developed allometric equations (AEs) 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 FIPI. 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. Section 4 presents the results together with discussion of these findings. Finally, conclusions and recommendations are provided in Section 5.

2 MATERIALS AND METHODS

The field measurement of forest biomass will be conducted through sample plots following latest version of the Guidelines on Destructive Measurement for Forest Biomass Estimation developed during the first phase of this Study[1].

2.1 Sampling strategy

2.1.1 Location and design of the plots

Criteria for sample plot establishment

The establishment of sample plots needs to meet the following criteria: (i) representativeness (based on assessment of experts) of the forest types being studied; (ii) representativeness for topographic conditions; (iii) covering a number of different tree sizes; and (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[2]).

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

Description of the sample plots

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

Table 1: Description of the sample plot QB-QN-01

Plot name:	QB-QN-01
Administrative location:	Compartment 281 – Truong Son commune - Quang Ninh district – Quang Binh province
Owner/manager:	Long Dai Industrial-Forestry One-Member Limited Company, Quang Binh province
Coordinate (VN2000 projection):	Longitude = 106°23’41” E; Latitude = 17°24’39” N
Altitude:	430 m
Slope:	15°
Plot area:	1 ha
Plot size:	100 x 100 (m)
Forest type:	Evergreen broadleaf forest

---

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

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

Forest status:	IIIB (forest that has been affected at medium-level; the structure of trees with DBH ≥ 40 cm has been changed)
Volume (estimated):	330 m^3/ha

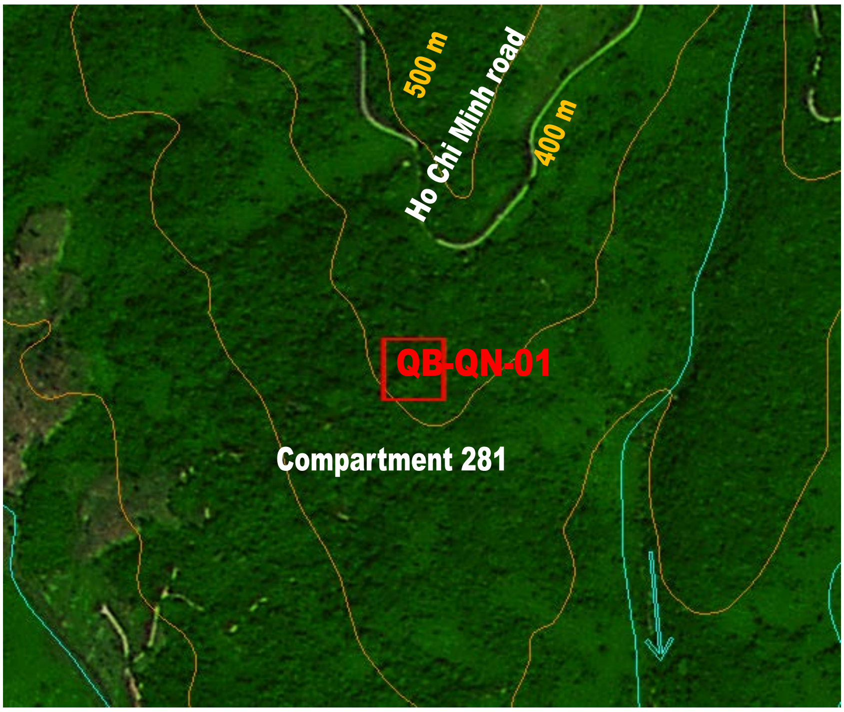


Figure 1: The position of the sample plot QB-QN-01 on satellite image

The information of the sample plot QB-QN-02 is given in Table 2 and its position on satellite image is given in Figure 2 below.

Table 2: Description of the sample plot QB-QN-02

Plot name:	QB-QN-02
Administrative location:	Compartment 329 – Truong Son commune - Quang Ninh district – Quang Binh province
Owner/manager:	Long Dai Industrial-Forestry One-Member Limited Company, Quang Binh province
Coordinate (VN2000 projection):	Longitude = 106°26’23” E; Latitude = 17°19’36” N
Altitude:	165 m
Slope:	20°
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):	250 m^3/ha

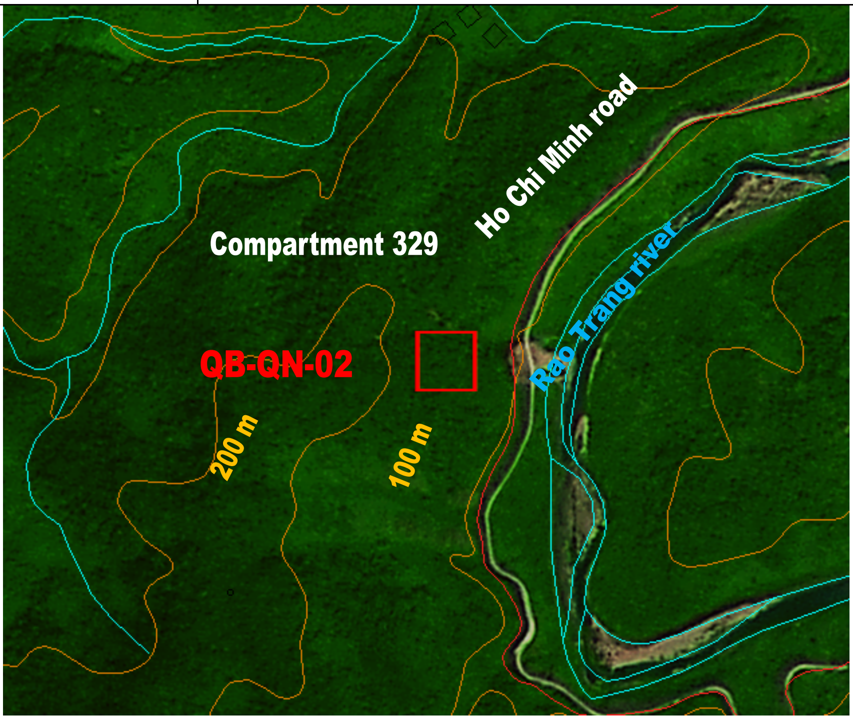


Figure 2: The position of the sample plot QB-QN-02 on satellite image

2.1.2 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 are 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. Select randomly the sample trees in each DBH class in the sample plots. The total number of sample trees for harvesting is 55 trees for 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. However, at least three sample trees should be harvested for each DBH class. Due to time and budget limitation, trees having DBH larger than 75 cm are basically not sampled in this study.

Totally, biomass data of 110 sample trees are collected. The numbers of felled trees for each species and each DBH class are given in Table 4 and Table 5, respectively.

Table 3: Number of felled trees divided by species in the two evergreen broadleaf sample plots

No	Local Name	Scientific Name	Number of felled trees		
			QB-QN-01	QB-QN-02	Total
1	Bông bạc	*Orthosiphon stamineus*		1	1
2	Bời lời	*Litsea sp.*	1	2	3
3	Bời lời đắng	*Litsea umbellata*		1	1
4	Bứa	*Garcinia oblongifolia*		1	1

No	Local Name	Scientific Name	Number of felled trees		
			QB-QN-01	QB-QN-02	Total
5	Bưởi bung	*Glycosmis citrifolia*	1	1	2
6	Chân chim	*Schefflera heptaphilla*		1	1
7	Chay	*Artocarpus sp.*		1	1
8	Chay lá to	*Artocarpus lakoocha*	1		1
9	Chẹo tía	*Engelhartia roxburghiana*		1	1
10	Cóc đá	*Garuga pierrei*	3	1	4
11	Chua khét	*Glenniea philippinensis*		1	1
12	Cò ke	*Microcos paniculata*	1		1
13	Đa	*Ficus sp.*		1	1
14	Dẻ	*Lithocarpus sp.*	1	2	3
15	Dung	*Symplocos sp.*	1	2	3
16	Giổi	*Michelia mediocris*	1		1
17	Gội	*Aphanamixis grandifolia*	2	1	3
18	Gõ mật	*Sindora siamensis*	2		2
19	Huỷnh	*Tarrietia javanica*	1	2	3
20	Kháo tía	*Machillus odoratissima*	1		1
21	Khổng	*Koilodepas longifolium*	2		2
22	Lá nến	*Macaranga denticulata*	2		2
23	Lim xanh	*Erythrophleum fordii*		2	2
24	Lim xẹt	*Peltophorum pterocarpum*		1	1
25	Máu chó	*Knema sp.*		3	3
26	Mít ma	*Ficus vasculosa*		1	1
27	Mít nài	*Artocarpus rigidus ssp*		1	1
28	Mò	*Cryptocarya sp.*		1	1
29	Nang	*Alangium ridleyi*	4	3	7
30	Nanh chuột	*Cryptocarya lenticellata*		1	1
31	Ngát	*Gironniera subaequalis*	3	2	5
32	Nhọ nồi	*Diospyros apiculata*	2	1	3
33	Nhọc lá to	*Polyathia lauii*	1	1	2
34	Ràng ràng	*Ormosia sp.*	2	1	3
35	Re	*Cinnamomum sp.*	1	1	2

No	**Local Name**	**Scientific Name**	**Number of felled trees**		
			QB-QN-01	**QB-QN-02**	**Total**
36	Re gừng	*Cinnamomum obtusifolium*		1	1
37	Rè	*Machilus sp.*	1	1	2
38	Rè vàng	*Machilus odoratissima*		1	1
39	Sang máu	*Horsfieldia amygdalina*		2	2
40	Tai chua	*Garcinia cowa*		1	1
41	Táu mật	*Vatica odorata ssp. brevipetiolata*	3	2	5
42	Thị rừng	*Diospyros sylvatica*	2		2
43	Trâm	*Syzygium sp.*	5	3	8
44	Trâm trắng	*Syzygium wightianum*	1		1
45	Trám	*Canarium sp.*	2	1	3
46	Trám đen	*Canarium tramdenum*	1	1	2
47	Trường	*Nephelium sp.*	5	3	8
48	Vạng trứng	*Endospermum chinense*	1	1	2
49	Xăng mả	*Carallia brachiata*	1		1
Total			***55***	***55***	***110***

Table 4: Number of standing and felled trees divided by DBH class in the two evergreen broadleaf sample plots

DBH class (cm)	**#of standing trees in the sample plot**		**# of felled trees for modeling**		**# of felled trees for validation**		**Total # of trees cut for modeling**	**Total #of trees cut for validation**
	QB-QN-01	**QB-QN-02**	**QB-QN-01**	**QB-QN-02**	**QB-QN-01**	**QB-QN-02**		
5 –15	634	921	23	24			47	
15 –25	141	205	10	9			19	
25 –35	43	107	7	8			15	
35 - 45	28	48	5	4			9	
45 - 55	17	10	4	4			8	
55 - 65	10	9	3	3			6	
65 - 75	9	5	3	3			6	
≥75	20	7	0	0			0	
Total	902	1312	55	55			110	

2.2 Variables measurement and calculation for volume and biomass

2.2.1 Field measurements

Measurement of tree DBH and names in sample plot

All live trees with DBH from 5 cm and above in the sample plots are measured. The information to collect 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 is marked, and then the tree is cut down at its base following logging procedures. Once the sample tree is felled down, using measuring tapes to accurately measure:

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 is separated into different components (stem, branches and leaves) and the weights of these components are weighed immediately in the field.

Collecting samples for analysis of dry oven mass and wood density

Sampling for dry mass analysis is taken immediately after completion of measurement of fresh weight of each tree components. The following steps are conducted for sampling:

1. Samples for dry mass analysis: for each tree, three samples (one for stem, one for branches and one for leaves) were collected. 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 are about 0.5 to 1.0 kg in weight. The samples of the leaves are about 0.3 – 0.5 kg in weight. The samples for dry mass analysis are weighted immediately in the field 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 determine accurately the fresh weight of each sample.
2. Samples for wood density analysis: five wood discs are taken from the stem. The sampling positions are at stump level (0.0 m), at 1/5; 2/5; 3/5; and 4/5 of the total tree height. The wood discs are 5 – 10 cm thick. The whole wood discs are taken for small discs or radial sections of the discs are taken for large discs.

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 are sent immediately to laboratories in FSIV for oven dry mass and wood density analyses. Dry mass of samples are determined using oven drier at a temperature of 105°C until the samples reach 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:

$$SWD = \frac{SDW}{SV} \qquad (1)$$

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 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:

$$TDW_c = TFW_c \frac{SDW_c}{SFW_c} \qquad (2)$$

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:

$$TDW_{tree} = TDW_{stem} + TDW_{branch} + TDW_{leave} \qquad (3)$$

Other variables

2.3 Model fitting and selection

Regression Analysis

The predictors include DBH (D, cm), height (H, m) and wood density (ρ, g/cm^3). The following statistical models are used:

Model no.	Model form
1	$AGB = aD^b$
2	$AGB = a(D^2H)^b$
3	$AGB = a(D^2H^{0.7})^b$
4	$AGB = aD^bH^c$
5	$AGB = a(D^{2.4}\rho)^b$
6	$AGB = aD^b\rho^c$
7	$AGB = a(D^2H\rho)^b$
8	$AGB = a(D^2H^{0.7}\rho)^b$
9	$AGB = aD^bH^c\rho^d$

Where *a*, *b*, and *c* are the coefficients needed to be found. Models 3, 5 and 8 are based on the results of previous analyses using other datasets that when undertaking 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).

Three approaches of regression analysis are 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.

Measuring the goodness of fit of developed equations

In order to evaluate the in-sample performance of the developed equations, we use three indicators: the adjusted R^2 value, the sum of squares error (SSE) and Akaike information criterion with correction (AICc, Burnham and Anderson 2002). The AICc is calculated by using the following formula:

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

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

The adjusted 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.

Cross validation and error assessment

To avoid over-fitting of the models, cross validation tests are 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-sets 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 developed equations are used to predict the total AGB of the testing sub-set. These predicted total AGB are then used to calculate the errors (in percentage) as compared to the measured total AGB of the testing sub-set. The above procedure is 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 5) in order to compare the performance of the equations in practice and estimate the confidence intervals of the error for each equation.

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

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.

Criteria for comparing with previously published equations

To compare the equations developed in this study with previously published equations, we used the two criteria: (i) average deviation $\bar{S}$ (%) and (ii) the total AGB error S (%), which are calculated as follows:

$$\overline{S}(\%) = \frac{100}{n}\sum_{i=1}^{n}\frac{\left|\hat{Y}_i - Y_i\right|}{Y_i} \quad (6)$$

$$S(\%) = 100\frac{\sum_{i=1}^{n}\hat{Y}_i - \sum_{i=1}^{n}Y_i}{\sum_{i=1}^{n}\hat{Y}_i} \quad (7)$$

Where *n* is the number of sample trees; $\hat{Y}_i$ and Y_i are the predicted and measured AGB of the *i*th tree, respectively.

3 RESULTS FOR EVERGREEN BROADLEAF FORESTS

3.1 Result 1: forest structure and trees characteristics

3.1.1 Species composition

There are totally 2,214 trees in the two studied sample plots (902 trees in QB-QN-01 and 1,312 trees in QB-QN-02). The average density is 1,107 trees/ha. Among these 2,214 trees, 107 species are found. There are 60 trees (N% = 2.7%) which names could not be identified. Table 3 provides a list of the 10 most dominant species. The full list of species is given in Annex 1.

Table 5: List of 10 most dominant species in the studied area

No.	Local name	Scientific Name	N	G	N%	G%	IV%
Total			***886***	***39.765***	***40.0***	***53.3***	***46.7***
1	Trâm	*Syzygium sp.*	230	4.084	10.388	5.473	7.931
2	Cóc đá	*Garuga pierrei*	36	8.403	1.626	11.261	6.444
3	Táu mật	*Vatica odorata ssp. brevipetiolata*	83	6.476	3.749	8.678	6.214
4	Trường	*Nephelium sp.*	128	4.040	5.781	5.414	5.598
5	Huỷnh	*Tarrietia javanica*	43	6.545	1.942	8.771	5.357
6	Nang	*Alangium ridleyi*	100	2.444	4.517	3.276	3.897
7	Trám	*Canarium sp.*	81	1.910	3.659	2.559	3.109
8	Khổng	*Koilodepas longifolium*	97	1.230	4.381	1.648	3.015
9	Vạng trứng	*Endospermum chinense*	30	2.844	1.355	3.811	2.583
10	Rè	*Machilus sp.*	58	1.789	2.620	2.398	2.509

Based on the importance value (IV%) index, which is calculated by taking the average of N% and G%,, the dominant species is *Syzygium sp.*, which has the IV% value of 7.93% and accounts for 10.4% of the total trees and 5.5% of the basal area. The second most dominant species is *Garuga pierrei*, which has the IV% value of 6.4%. Other species having IV% ≥ 5% include *Vatica odorata ssp. brevipetiolata* (IV% = 6.2%), *Nephelium sp.* (IV% = 5.6%) and *Tarrietia javanica* (IV% = 5.4%). The 10 most dominant species account for 40.0% of the total trees and 53.3% of the total basal area.

3.1.2 Forest structure

The N-D distribution of all the trees in the two studied sample plots is given in Figure 3. It can be seen that the number of trees is decreasing when DBH gets larger. The basal area distribution per diameter class in the two sample plots is given in Figure 4.

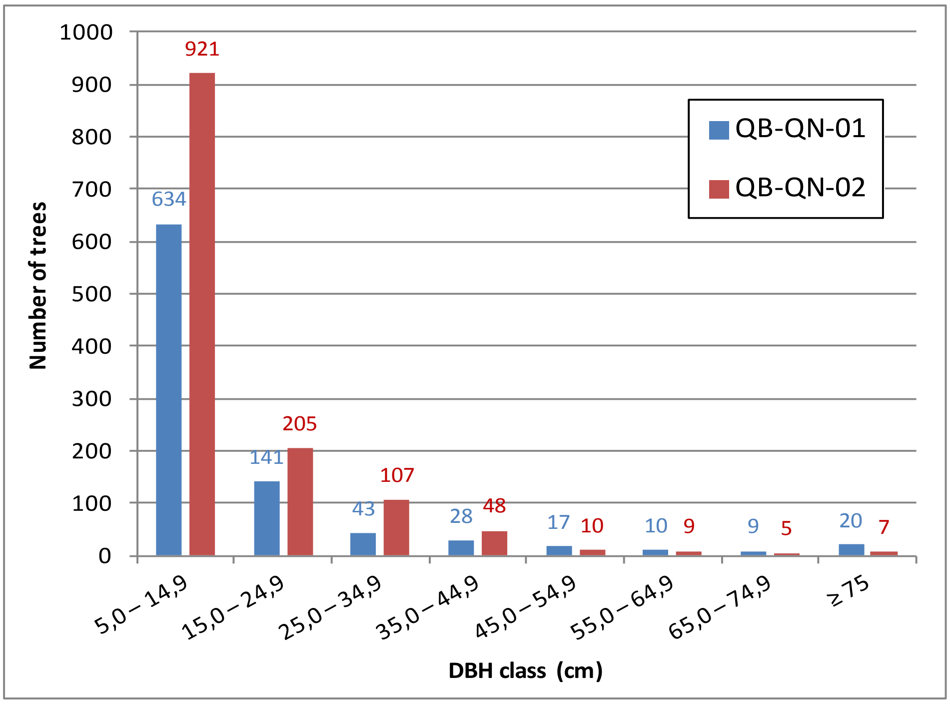


Figure 3: N-D distribution of the trees in two sample plots

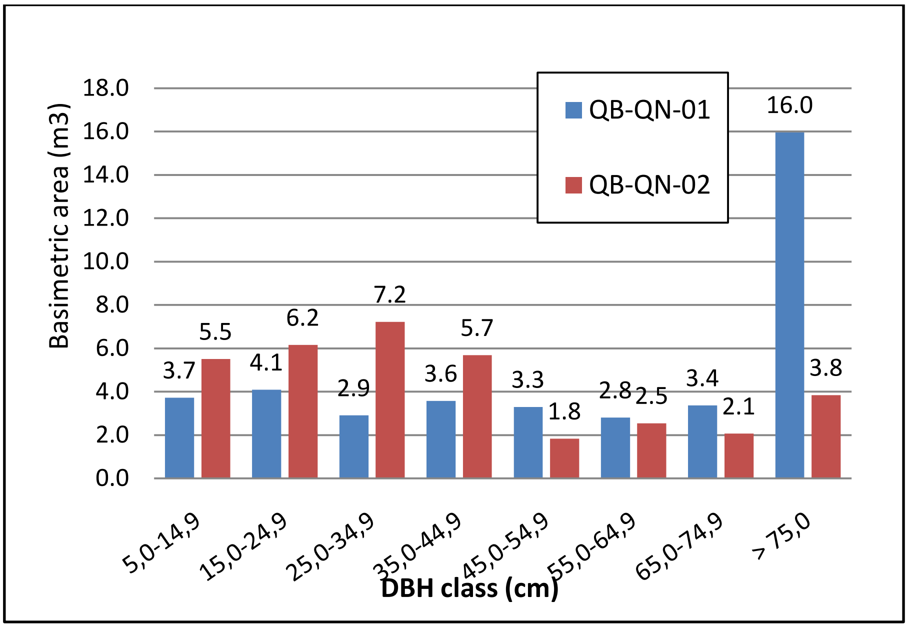


Figure 4 G-D distribution of the trees in two sample plots

3.1.3 Relation between H and diameter

We used the SAS software to do a regression analysis of the logarithm function correlating H(m) and DBH (cm). The resulted equation is H = 9.579 × ln(DBH) – 10.130 (R^2 = 0.896; F value = 926.34; p < 0.001; Figure 5). It can be observed that H correlates quite well with DBH in our dataset.

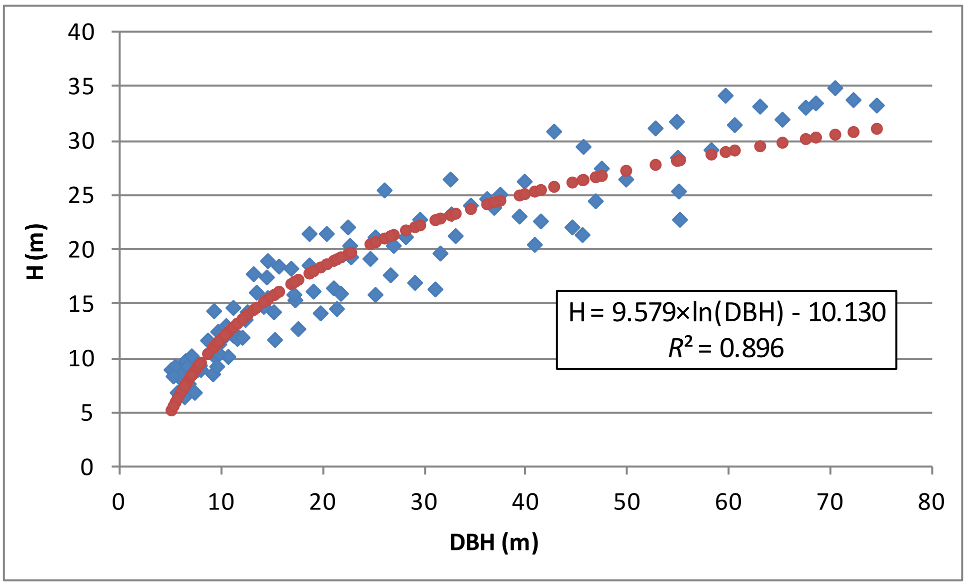


Figure 5: Correlation function between H (m) and DBH (cm)

3.1.4 Wood density analysis

Table 7 below provides a summary of wood density analysis results for the species in the sample dataset. It can be observed that the wood densities analyzed in this Study agree quite well with the previous known values. The wood densities of 110 trees vary largely from 0.386 to 0.946 with the average being 0.626 and the standard deviation being 0.143. This indicates that the inclusion of the wood density as an input variable should improve the accuracy and certainty of the biomass estimation.

Table 6: Results of wood density analysis for species in the two evergreen broadleaf sample plots

No	Local Name	Scientific Name	*N*	Wood density (g/cm^3)			
				Min	Max	Avg	Known value
1	Bời lời	*Litsea sp.*	3	0.451	0.665	0.545	0.560
2	Bời lời đắng	*Litsea umbellata*	1	0.437	0.437	0.437	
3	Bông bạc	*Orthosiphon stamineus*	1	0.661	0.661	0.661	
4	Bứa	*Garcinia oblongifolia*	1	0.667	0.667	0.667	0.710
5	Bưởi bung	*Glycosmis citrifolia*	2	0.462	0.671	0.567	
6	Chân chim	*Schefflera heptaphilla*	1	0.402	0.402	0.402	
7	Chay	*Artocarpus sp.*	1	0.446	0.446	0.446	0.430
8	Chay lá to	*Artocarpus lakoocha*	1	0.417	0.417	0.417	
9	Chẹo tía	*Engelhartia roxburghiana*	1	0.511	0.511	0.511	0.680

No	Local Name	Scientific Name	*N*	Wood density (g/cm^3)			
				Min	Max	Avg	Known value
10	Cóc đá	*Garuga pierrei*	4	0.595	0.639	0.618	0.680
11	Chua khét	*Glenniea philippinensis*	1	0.730	0.730	0.730	0.730
12	Cò ke	*Microcos paniculata*	1	0.637	0.637	0.637	
13	Đa	*Ficus sp.*	1	0.480	0.480	0.480	
14	Dẻ	*Lithocarpus sp.*	3	0.567	0.605	0.590	
15	Dung	*Symplocos sp.*	3	0.573	0.679	0.636	0.590
16	Giổi	*Michelia mediocris*	1	0.547	0.547	0.547	0.580
17	Gõ mật	*Sindora siamensis*	2	0.579	0.677	0.628	0.880
18	Gội	*Aphanamixis grandifolia*	3	0.553	0.777	0.682	0.670
19	Huỷnh	*Tarrietia javanica*	3	0.549	0.762	0.675	0.710
20	Kháo tía	*Machillus odoratissima*	1	0.722	0.722	0.722	0.750
21	Khổng	*Koilodepas longifolium*	2	0.792	0.861	0.826	0.960
22	Lá nến	*Macaranga denticulata*	2	0.780	0.823	0.801	0.580
23	Lim xanh	*Erythrophleum fordii*	2	0.583	0.860	0.722	0.930
24	Lim xẹt	*Peltophorum pterocarpum*	1	0.572	0.572	0.572	0.600
25	Máu chó	*Knema sp.*	3	0.497	0.669	0.555	0.650
26	Mít ma	*Ficus vasculosa*	1	0.448	0.448	0.448	
27	Mít nài	*Artocarpus rigidus ssp*	1	0.746	0.746	0.746	0.540
28	Mò	*Cryptocarya sp.*	1	0.443	0.443	0.443	0.480
29	Nang	*Alangium ridleyi*	7	0.548	0.765	0.662	
30	Nanh chuột	*Cryptocarya lenticellata*	1	0.496	0.496	0.496	
31	Ngát	*Gironniera subaequalis*	5	0.433	0.581	0.496	0.570
32	Nhọ nồi	*Diospyros apiculata*	3	0.568	0.646	0.607	0.810
33	Nhọc lá to	*Polyathia lauii*	2	0.386	0.435	0.410	
34	Ràng ràng	*Ormosia sp.*	3	0.456	0.537	0.510	0.560
35	Re	*Cinnamomum sp.*	2	0.473	0.526	0.500	0.510
36	Re gừng	*Cinnamomum obtusifolium*	1	0.535	0.535	0.535	0.530
37	Rè	*Machilus sp.*	2	0.443	0.538	0.490	
38	Rè vàng	*Machilus odoratissima*	1	0.442	0.442	0.442	
39	Sang máu	*Horsfieldia amygdalina*	2	0.465	0.494	0.480	0.590
40	Tai chua	*Garcinia cowa*	1	0.674	0.674	0.674	0.875
41	Táu mật	*Vatica odorata ssp.brevipetiolata*	5	0.643	0.946	0.825	0.860

No	Local Name	Scientific Name	*N*	Wood density (g/cm^3)			
				Min	Max	Avg	Known value
42	Thị rừng	*Diospyros sylvatica*	2	0.624	0.744	0.684	0.810
43	Trám	*Canarium sp.*	3	0.708	0.762	0.740	
44	Trám đen	*Canarium tramdenum*	2	0.508	0.528	0.518	0.650
45	Trâm	*Syzygium sp.*	8	0.521	0.927	0.749	
46	Trâm trắng	*Syzygium wightianum*	1	0.640	0.640	0.640	0.640
47	Trường	*Nephelium sp.*	8	0.650	0.942	0.802	0.910
48	Vạng trứng	*Endospermum chinense*	2	0.396	0.444	0.420	0.480
49	Xăng mả	*Carallia brachiata*	1	0.411	0.411	0.411	
All			***110***	***0.386***	***0.946***	***0.626***	

* These values are taken from the wood density database collected by RCFEE in the initial phase of this Study.

Table 7 Results of wood density analysis for families in the two evergreen broadleaf sample plots

Family	Average	n	St Dev	Min	Max
Actinidiaceae	0.661861	7	0.088290	0.548	0.765
Annonaceae	0.410481	2	0.034618	0.386	0.435
Arallaceae	0.402172	1	Na	Na	Na
Burseraceae	0.636763	9	0.089812	0.508	0.762
Caesalpinaceae	0.654150	5	0.122785	0.572	0.860
Cluciaceae	0.670615	2	0.005007	0.667	0.674
Dipterocarpaceae	0.824896	5	0.121410	0.643	0.946
Ebenaceae	0.638202	5	0.065799	0.568	0.744
Euphorbiaceae	0.682581	6	0.205881	0.396	0.861
Fabaceae	0.509716	3	0.046720	0.456	0.537
Fagaceae	0.589718	3	0.019865	0.567	0.605
Lauraceae	0.497194	12	0.065816	0.437	0.665
Magnoliaceae	0.546623	1	Na	Na	Na
Meliaceae	0.681893	3	0.115623	0.553	0.777
Moraceae	0.507419	5	0.135345	0.417	0.746
Myristicaceae	0.524919	5	0.081921	0.465	0.669
Myrtaceae	0.736768	9	0.140480	0.521	0.927
Rhizophoraceae	0.410700	1	Na	Na	Na

Rutaceae	0.566821	2	0.147533	0.462	0.671
Sapindaceae	0.794232	9	0.089909	0.650	0.942
Sterculiaceae	0.675087	3	0.111422	0.549	0.762
Symplocaceae	0.636347	3	0.127415	0.643	0.823
Tiliaceae	0.637248	1	Na	Na	Na
Ulmaceae	0.496287	5	0.068997	0.433	0.581
#N/A	0.631180	3	0.108584	0.511	0.722
Total	0.626	110	0.142988	0.386	0.946

3.2 Result 2: Modeling of the stem volume

The relation between volume and tree diameter at breast height (DBH) or tree height (H) are shown in Annex. From those graphs, several models could lead to a good prediction of stem volume and have been fitted. Only the model that allow a correction of the heteroscedasticity have been taken into account:

- $\ln(V) = b + c.\ln(DBH)$
- $V = b.DBH^c + \varepsilon$ avec $\varepsilon \sim N(\mu,\sigma^2.DBH^k)$
- $V = b.(D^2H)^c + \varepsilon$ avec $\varepsilon \sim N(\mu,\sigma^2.(D^2H)^k)$
- $V = b.D^c.H^d + \varepsilon$ avec $\varepsilon \sim N(\mu,\sigma^2.(DBH)^k)$

These models have been fitted using the 110 felled trees. The results are shown in the following table:

	Estimation of the parameters					Selection criteria		
Model	b	c	d	σ^2	k	RMSE	AICc	Adjust R^2
ln(V) = b + c.ln(DBH)	-8.57222	2.40883				0.03115		0.9911
$V=b.DBH^c + \varepsilon$ avec ε~N(μ,σ2.DBHk)	0.000200	2.396022		1.0000001E-8	4.104911		-2.15546	
$V= b.(D2H)^c + \varepsilon$ avec ε~N(μ,σ2.(D2H)k)	0.376022	0.940660		0.002114	1.948610		-2.58958	
$V=b.D^c.H^d + \varepsilon$ avec $\varepsilon \sim N(\mu,\sigma^2.(DBH)k)$	0.000059435	1.869203	0.983647	1.0000001E-8	3.912904		-0.25598	

According to the results, the log-log model is simple but offer an acceptable Mean squared error, 0.03 for values of ln(V) ranging from -5 to 2. The distribution of residuals against predicted values (see in Annex) shows no bias, that is confirmed by F tests.

For model fitted using the maximum likelihood procedure, the best model is the one fitting volume with D^2H, according to the AICc value. That means that the information provided by the height increase the accuracy of the model, compared to a model using DBH only. On the other hand, if we add another parameter for taking height into account the value of AIC and AICc is not as low as the model using D^2H. In conclusion, the following models can be used to assess stem volume (V) using the DBH or DBH and H:

- $\ln(V) = -8.57222 + 2.40883 \times \ln(DBH)$
- $V = 0.376022 \times (D^2H)^{0.940660}$

3.3 Result 3: Modeling of Aboveground biomass

3.3.1 Dry mass analysis

The results of dry mass analysis of 110 sample trees are given in Table 6. In average, stems have the highest ratio and branches rank second. The variation of the ratios is smallest in stems and highest in leaves.

Table 8: Ratio of dry biomass to fresh biomass of evergreen broadleaf forests

Statistical values	Dry to fresh mass ratio		
	Stem	Branch	Leaf
Min	0.408	0.338	0.186
Max	0.667	0.596	0.475
Avg	0.545	0.467	0.322
Stdev	0.062	0.059	0.075
S(%)	11.412	12.678	23.464

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 using the following formula (2) in Section 2 above. Data on species name, DBH, height, and fresh biomass of each component, dry-fresh mass ratios and the converted dry mass data of these 110 sample trees are given in the Annexes.

3.3.2 Modeling per tree compartments

Allometric equations for each component (stem, branches and leaves) of the tree are also developed. Only Model 1, which uses only the input variable D, is used here. The regression analyses using the second approach are done with the procedure NLIN in the SAS software. The results are given in Figure 10, Figure 11 and Figure 12 below.

Table 9: Results of the second regression approach relating dry biomass (in kg) of each part of the tree with DBH (cm) for Model 1

Part of tree	Parameter *a*				Parameter *b*				R^2	Pr > F
	Estimate	Std. err.	95% CL		Estimate	Std. err.	95% CL			
Stem	0.0962^{***}	0.0107	0.0750	0.1175	2.4250^{***}	0.0367	2.3521	2.4978	0.9130	<.0001
Branch	0.0140^{***}	0.0031	0.0078	0.0202	2.4932^{***}	0.0735	2.3476	2.6389	0.7354	<.0001
Leaf	0.0407^{***}	0.0085	0.0239	0.0576	1.6576^{***}	0.0687	1.5215	1.7937	0.6397	<.0001

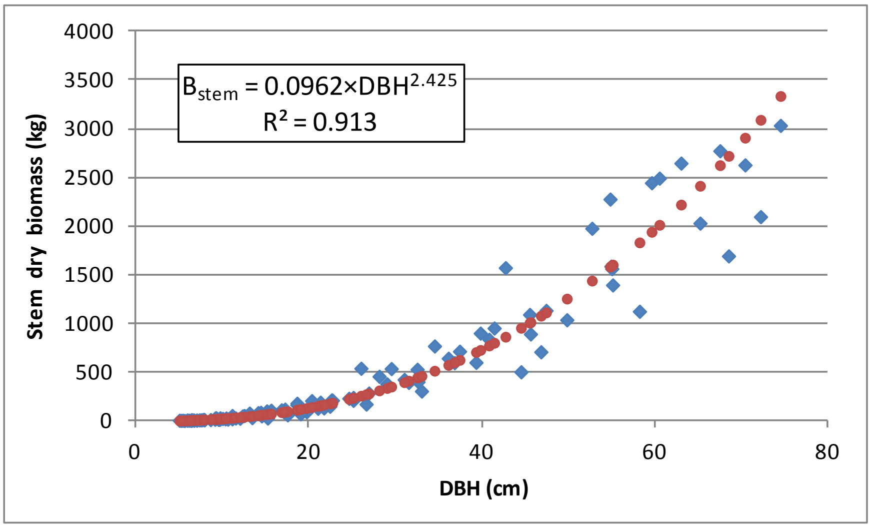


Figure 6: Allometric equation for estimating stem dry biomass (kg) from DBH (cm)

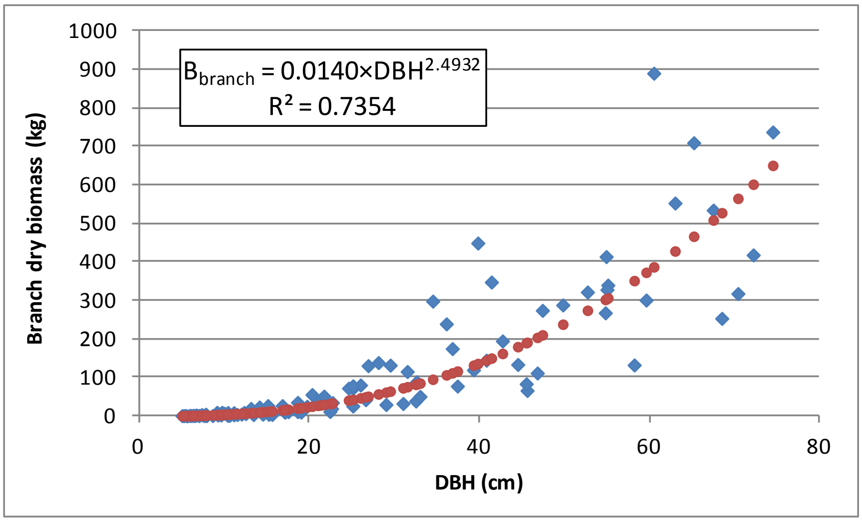


Figure 7: Allometric equation for estimating branch dry biomass (kg) from DBH (cm)

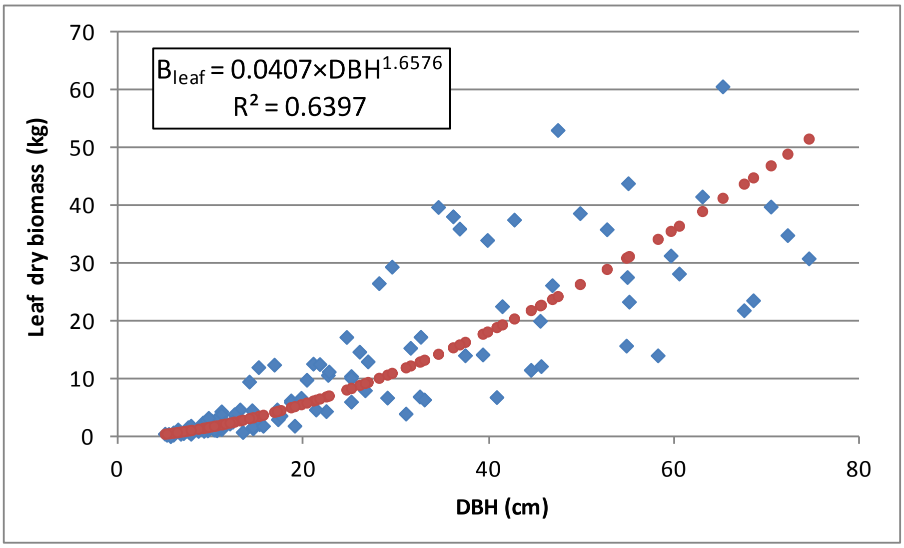


Figure 8: Allometric equation for estimating leave dry biomass (kg) from DBH (cm)

It can be observed from the above figures that stem biomass correlates strongest to the DBH (R^2 = 0.91), followed by branch biomass (R^2 = 0.74). Leaf biomass has weakest correlation with DBH (R^2 = 0.64).

3.3.3 Modeling of total aboveground biomass

First, regression analyses using the first approach (least squares optimization of the original equations) for the nine models described in Section 2.5 are applied using the procedure NLIN in the SAS software. The variables used for these analyses are: D (cm), H (m), and ρ (g/cm^3). The analyzed results are given in Table 8.

Table 10: Results of regression analyses using the first approach

No	Model	*a*	*b*	*c*	*d*	$\bar{R}^2$	SSE	Pr > F
1	$B = aD^b$	0.4282	2.1041			0.9174	7,852,784	<.0001
2	$B = a(D^2H)^b$	0.1610	0.8264			0.9289	6,778,576	<.0001
3	$B = a(D^2H^{0.7})^b$	0.2050	0.8838			0.9271	6,944,762	<.0001
4	$B = aD^bH^c$	0.0978	1.4121	1.2588		0.9297	6,641,430	<.0001
5	$B = a(D^{2.4}\rho)^b$	0.5689	0.8922			0.9363	6,065,094	<.0001
6	$B = aD^b\rho^c$	0.3911	2.2004	0.5891		0.9438	5,287,180	<.0001
7	$B = a(D^2H\rho)^b$	0.1644	0.8616			0.9565	4,147,323	<.0001
8	$B = a(D^2H^{0.7}\rho)^b$	0.2320	0.9147			0.9519	4,588,629	<.0001
9	$B = aD^bH^c\rho^d$	0.0952	1.5651	1.1836	0.6634	0.9600	3,736,551	<.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. Equation derived from Model 1 has the lowest $\bar{R}^2$, as it uses only the predictor D. Models that use H or ρ as an additional variable have higher $\bar{R}^2$ than Model 1. However, models that use ρ as an additional variable (Models 5 and 6) have higher $\bar{R}^2$ than models that use H as an additional variable (Models 2, 3 and 4), indicating that the inclusion of ρ can improve that prediction accuracy more than the inclusion of H. Among the three models that use H as an additional variable, Model 4 has the highest $\bar{R}^2$. Among the two models that use ρ as an additional variable, Model 6 has better $\bar{R}^2$. Models 7, 8 and 9, which use all three input variables, have the highest $\bar{R}^2$. Among these three, Model 9 has the highest $\bar{R}^2$.

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 9.

Table 11: Results of regression analyses using the second approach

No	Model	*a*	*b*	*c*	*d*	$\bar{R}^2$	SSE	Pr > F
1	$B = aD^b$	0.1245	2.4163			0.9736	10.158	<.0001
2	$B = a(D^2H)^b$	0.0421	0.9440			0.9775	8.652	<.0001
3	$B = a(D^2H^{0.7})^b$	0.0549	1.0112			0.9774	8.674	<.0001
4	$B = aD^bH^c$	0.0468	1.9426	0.8484		0.9771	8.632	<.0001
5	$B = a(D^{2.4}\rho)^b$	0.2105	1.0025			0.9891	4.198	<.0001
6	$B = aD^b\rho^c$	0.2129	2.4051	1.0198		0.9891	4.196	<.0001
7	$B = a(D^2H\rho)^b$	0.0695	0.9397			0.9915	3.254	<.0001
8	$B = a(D^2H^{0.7}\rho)^b$	0.0941	1.0057			0.9918	3.140	<.0001
9	$B = aD^bH^c\rho^d$	0.0922	2.0093	0.7096	0.9827	0.9918	3.137	<.0001

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

It can be observed that, with the same model, the coefficients estimated using the second approach are quite different from those estimated using the first approach. The order of the 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 is now has the highest $\bar{R}^2$ among the three models that use D and H as the input variables. Among the two models that use D and ρ as the input variables, Model 5 is now has the same $\bar{R}^2$ with Model 6.

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 10.

Table 12: Results of regression analyses using the third approach

No	Model	*a*	*b*	*c*	*d*	Log*L*	AICc
1	$B = aD^b$	0.1281	2.4218			-578.18	1,160.48
2	$B = a(D^2H)^b$	0.0420	0.9484			-567.47	1,139.05
3	$B = a(D^2H^{0.7})^b$	0.0557	1.0146			-568.00	1,140.12
4	$B = aD^bH^c$	0.0440	1.9181	0.9101		-567.45	1,143.28
5	$B = a(D^{2.4}\rho)^b$	0.2165	1.0010			-532.02	1,068.15
6	$B = aD^b\rho^c$	0.2146	2.4033	0.9860		-532.00	1,068.11
7	$B = a(D^2H\rho)^b$	0.0704	0.9389			-512.77	1,031.76
8	$B = a(D^2H^{0.7}\rho)^b$	0.0964	1.0036			-510.68	1,027.58
9	$B = aD^bH^c\rho^d$	0.0923	2.0030	0.7161	0.9572	-510.45	1,027.12

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

It can be seen 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 worst AICc value. (Note that for AICc, the lower is the better.) Models that use only two variables D and H (Models 2, 3 and 4) have lower AICc values as compared to Model 1. Among these three models, Model 4 has the lowest AICc. Models that use only two variables D and ρ (Models 5 and 6) have significantly lower AICc than Models that use only two variables D and H, suggesting that the inclusion of ρ is more important than the inclusion of H. Between these two, Model 6 performs slightly better in terms of AICc. Finally, Models 7, 8 and 9, which use all three input variables, have the lowest AICc. Among these three, Model 9 has the lowest AICc.

To avoid over-fitting of the models, we carried out cross validation tests as described in Section 2.6. Table 11 shows the properties of the approximated probability density functions of the total AGB error for every equations developed using the first approach.

Table 13: Properties of the probability density functions of the total AGB error (%) for the equations developed using the first regression approach

Model	α	σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
1	0.0136	0.1298	0.0356	3.315	2.669	1.395	0.0406	0.9994	-14.47	24.77	39.24
2	0.0154	0.1344	0.0460	3.668	3.052	1.837	0.0441	0.9995	-12.67	23.52	36.19
3	0.0150	0.1332	0.0438	3.597	2.978	1.756	0.0435	0.9995	-12.98	23.70	36.68
4	0.0142	0.1270	0.0381	3.324	2.732	1.563	0.0433	0.9995	-13.38	23.40	36.78
5	0.0118	0.1018	0.0339	3.394	2.937	2.029	0.0448	0.9992	-12.99	22.37	35.36
6	0.0132	0.1132	0.0350	3.201	2.697	1.699	0.0452	0.9993	-12.90	22.18	35.08
7	0.0068	0.0504	0.0188	2.960	2.772	2.394	0.0533	0.9988	-11.22	18.22	29.44
8	0.0080	0.0617	0.0226	3.101	2.857	2.371	0.0507	0.9989	-11.71	19.30	31.01
9	0.0073	0.0531	0.0230	3.378	3.181	2.787	0.0538	0.9995	-10.65	18.53	29.18

It can be 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. Equations developed using the first approach tend to over-estimate the total AGB by about 3.0% ÷ 3.7% (as indicated by the mean of the probability density functions). Model 1, which uses only D as the input variable, has the largest range of error (from -14.5% to 24.8%). Models that use H as an additional input variable (Models 2, 3 and 4) reduce the range of error by about 3% as compared to Model 1. Models that use ρ as an additional variable (Models 5 and 6) have slightly smaller ranges of error as compared to models that use H as an additional variable, confirming that the inclusion of ρ can improve the certainty of the biomass estimation more than the inclusion of H. Finally, Models 7, 8 and 9, which use all three predictors, have the smallest ranges of error. Among these three, Model 9 has a slightly smaller range of error. The probability density functions of error for some selected models are shown in Figure 6.

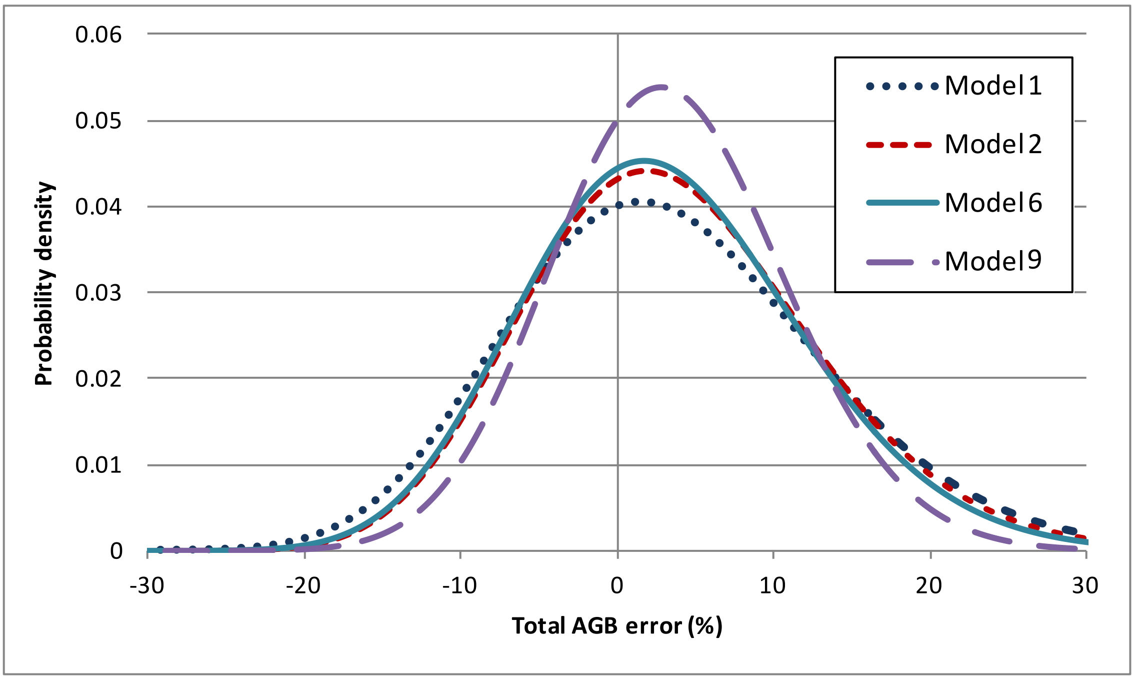


Figure 9: Probability density functions of the total AGB error (%) for some selected equations developed by using the first regression approach.

Next, the cross validation test is performed for equations derived using the second regression approach. The results are provided in Table 12.

Table 14: Properties of the probability density functions of the total AGB error (%) for the equations developed using the second regression approach

Model	α	σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
1	0.0133	0.1272	-0.0068	0.101	-0.505	-1.702	0.0424	0.9994	-16.96	20.61	37.57
2	0.0156	0.1391	-0.0287	-1.205	-1.810	-3.003	0.0465	0.9994	-16.67	17.70	34.37
3	0.0151	0.1362	-0.0209	-0.761	-1.364	-2.553	0.0457	0.9994	-16.53	18.44	34.97
4	0.0154	0.1388	-0.0259	-1.045	-1.656	-2.861	0.0460	0.9994	-16.69	18.08	34.77
5	0.0133	0.1065	0.0023	0.600	0.171	-0.680	0.0500	0.9994	-14.03	17.67	31.70
6	0.0142	0.1159	0.0043	0.775	0.300	-0.641	0.0491	0.9994	-14.03	18.29	32.32
7	0.0075	0.0478	-0.0096	-1.126	-1.278	-1.581	0.0630	0.9995	-13.14	11.75	24.89
8	0.0085	0.0575	-0.0070	-0.624	-0.817	-1.203	0.0594	0.9995	-13.29	13.14	26.43
9	0.0094	0.0642	-0.0056	-0.376	-0.596	-1.034	0.0585	0.9996	-13.17	13.67	26.84

It can be observed that equations developed using the second regression approach perform much better than those developed using the first regression approach. With the same model, the equation developed using the second approach is more accurate (i.e., the mean is closer to zero) and more robust (i.e., having smaller ranges of error). For example, the equation developed by using the second regression approach of Model 7 has the mean (or the expected value) of error -1.126 and the range of error from -13.14 to 11.75, while the corresponding one developed by using the first regression approach has the mean of error 2.96 and the range of error from -11.22 to 18.22.

Among the three models that use D and H as the input variables, Model 2 has the smallest range of error. Among the two models that use D and ρ as the input variables, Model 5 has smaller range of error. Among the three models that use all three input variables, Model 7 has the smallest range of error. The probability density functions of the equations derived from Models 1, 2, 5 and 7, which are the best for each group of input variables, are shown in Figure 6.

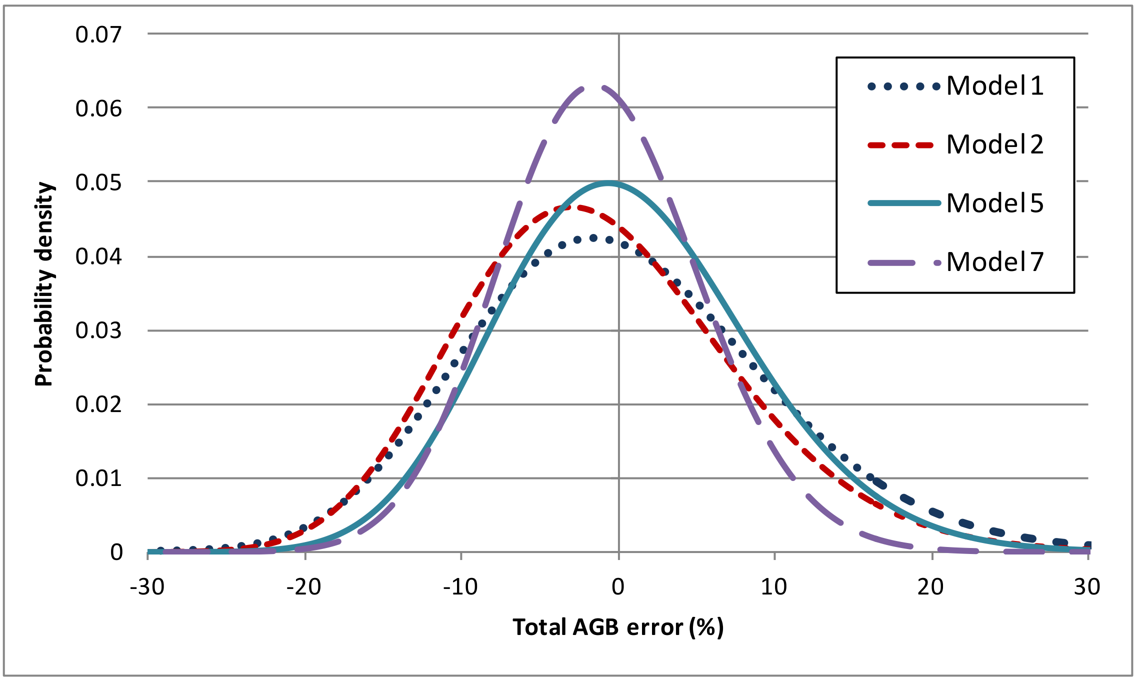


Figure 10: Probability density functions of the total AGB error (%) for some selected equations developed by the second regression approach.

Finally, the cross validation test is performed for equations developed using the third regression approach and the results are provided in Table 13.

Table 15: Properties of the probability density functions of the total AGB error (%) for the equations developed by the third regression approach

Model	α	σ	μ	Mean	Median	Mode	f_{max}	R^2	95% Confidence Interval		
									Lower	Upper	Range
1	0.0101	0.0948	0.0470	5.212	4.747	3.823	0.0409	0.9994	-12.78	25.85	38.63
2	0.0138	0.1193	0.0392	3.428	2.891	1.829	0.0448	0.9994	-12.78	22.69	35.47
3	0.0127	0.1101	0.0473	4.334	3.830	2.832	0.0440	0.9994	-12.25	23.78	36.03
4	0.0134	0.1161	0.0397	3.540	3.016	1.979	0.0446	0.9994	-12.75	22.81	35.56
5	0.0139	0.1125	0.0247	2.275	1.806	0.876	0.0483	0.9995	-12.83	20.05	32.88
6	0.0130	0.1037	0.0260	2.458	2.031	1.183	0.0488	0.9996	-12.54	19.88	32.42
7	0.0080	0.0491	-0.0071	-0.737	-0.888	-1.188	0.0653	0.9996	-12.32	11.70	24.02
8	0.0089	0.0584	-0.0033	-0.174	-0.365	-0.745	0.0612	0.9996	-12.46	13.20	25.66
9	0.0086	0.0575	0.0030	0.548	0.355	-0.032	0.0594	0.9996	-12.11	14.31	26.42

Equations derived from Model 1 using the third regression method tends to over-estimate the total AGB by about 5%. Equations derived from the models that use D and H as the input variable (Models 2, 3 and 4) tend to over-estimate the total AGB by 3-4%. Equations derived from the models that use D and ρ as the input variable (Models 5 and 6) tend to over-estimate the total AGB by about 2%. However, equations derived from Models 7, 8 and 9, which use all three variables, are quite accurate (the expected values of error are -0.74%, -0.17% and 0.55%, respectively).

Among the three models that use D and H as the input variable, Model 2 has the smallest range of error. Among the two models that use D and ρ as the input variables, Model 6 has the smaller range of error. Among the three models that use all three input variables, Model 7 has the smallest range of error. The equation that derived from Model 7 using the third regression method also has the smallest range of error (from -12.32 to 11.70) so far. The probability density functions of the equations derived from Models 1, 2, 6 and 7 using the third regression approach are shown in Figure 8.

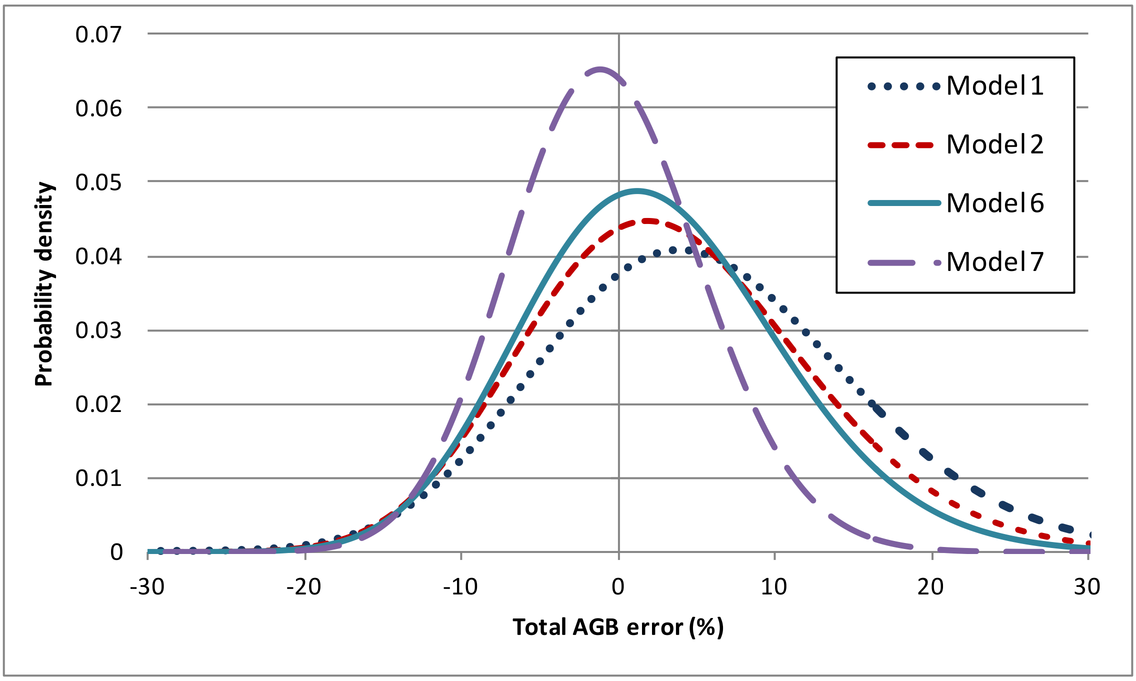


Figure 11: Probability density functions of the total AGB error (%) for some selected equations developed by using the third regression approach.

In order to find the best equations for each group of input variables, we did a comparison of the probability density functions of total AGB error across the three regression approaches. The results are shown in Figure 9.

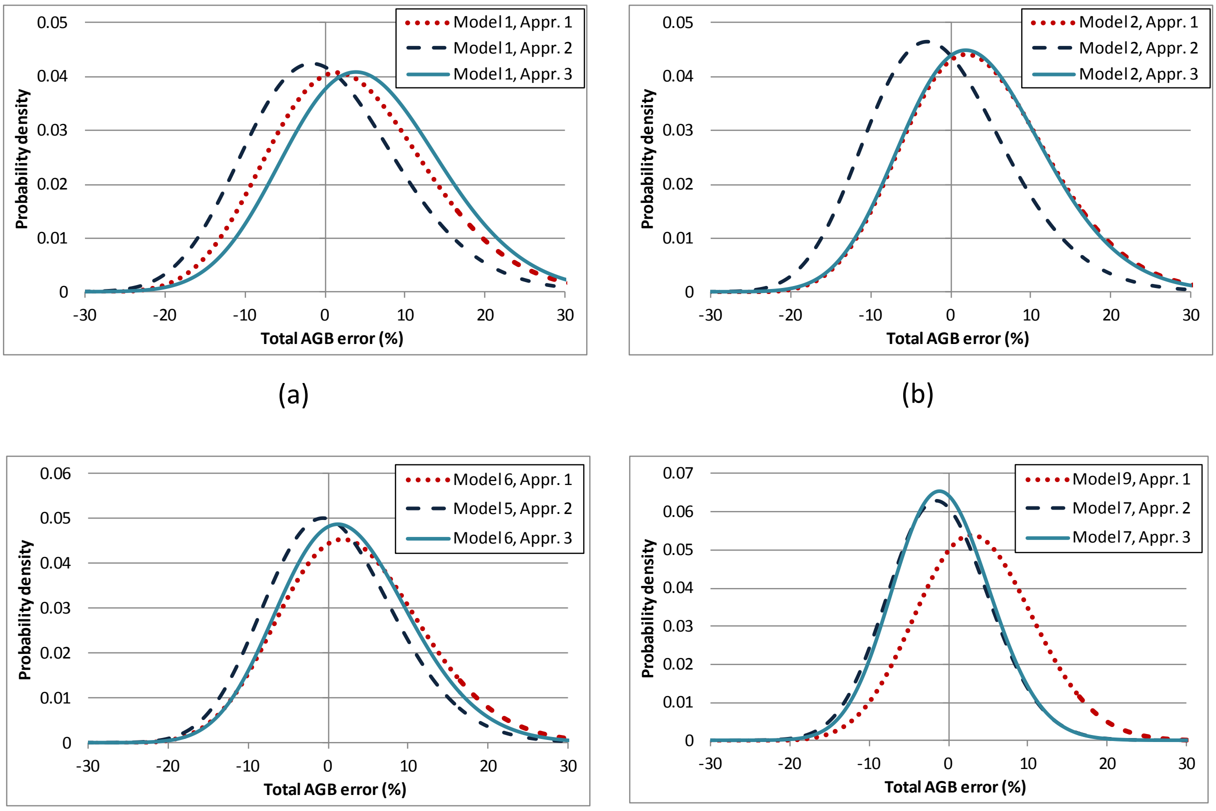
Based on the results of the comparison, we recommend to use (i) the equation derived from Model 1 using the second regression approach when D is the only input variable; (ii) the equation derived from Model 2 using the second regression approach when D and H are used as the input variables; (iii) the equation derived from Model 5 using the second regression approach when D and ρ are used as the input variables; and (iv) the equation derived from Model 7 using the third regression method when all three parameters D, H and ρ are used as the input variables. Specifically, the following equations are recommended to apply:

$$AGB = 0.1245 \times D^{2.4163} \tag{6}$$

$$AGB = 0.0421 \times (D^2H)^{0.9440} \tag{7}$$

$$AGB = 0.2105 \times (D^{2.4}\rho)^{1.0025} \tag{8}$$

$$AGB = 0.0704 \times (D^2H\rho)^{0.9389} \tag{9}$$



(a) (b)

(c) (d)

Figure 12: Comparison of the models across three regression approaches for each group of inputs. (a) models that use D as the only input variable; (b) models that use D and H as the input variables; (c) models that use D and ρ as the input variables; and (d) models that use all three input variables.

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. The equations from (6) to (9) 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. Our previous 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 the equations from (6) to (9) above when predicting the total AGB of 37 or more trees. The expected values and the ranges of error for equations from (6) to (9) are given in Table 14.

Table 16: Expected values and ranges of total AGB error for equations from (6) to (9) when predicting total AGB of 37 or more trees.

No.	Equation	Expected value of error (%)	Range of error (95% CL)
(6)	$AGB = 0.1245 \times D^{2.4163}$	0.101	-16.96% ÷ 20.61%
(7)	$AGB = 0.0421 \times (D^2H)^{0.9440}$	-1.205	-16.67% ÷ 17.70%

(8)	$AGB = 0.2105\times(D^{2.4}\rho)^{1.0025}$	0.600	-14.03% ÷ 17.67%
(9)	$AGB = 0.0704\times(D^{2}H\rho)^{0.9389}$	-0.737	-12.32% ÷ 11.70%

3.3.4 Modeling of ABG for the main tree families and species

3.3.5 Comparison with generic models

We did a comparison of our developed AE (6) (which uses DBH as the only input variable) with other two equations. The first equation is: AGB = exp(-2.134 + 2.530×ln(DBH)), which is developed by Brown (1997) for all tropical moist forests. The second equation is AGB = exp((-1.201 + 2.196×ln(DBH)), which is developed by Basuki et al. (2009) for mixed species in tropical lowland Dipterocarp forests. The result is shown in the Figure 13.

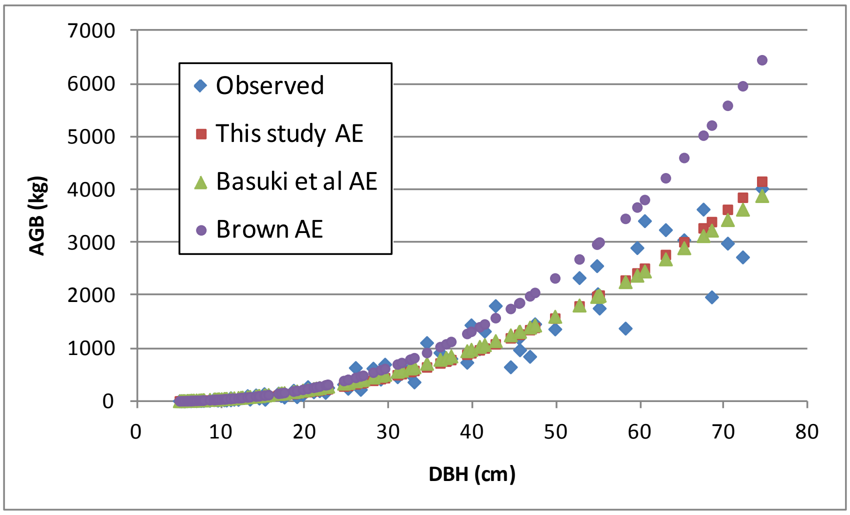


Figure 13: Comparison between the Model1 fitted AE using the second regression approach and the Basuki et al. (2009) AE and Brown (1997) AE

It can be observed that the Brown AE seems to over-estimate the AGB of trees in our dataset. Thus, it should be used with care when estimating forest biomass in Vietnam. The Basuki *et al.* equation, although closer to the equation developed in this Study, slightly under-estimate the AGB of large trees in our dataset. This is understandable as the Basuki *et al.* equation is developed specifically for the tropical lowland Dipterocarp forests.

Next, Eq. (9) was compared with the equation developed by Chave *et al.* (2005), which is AGB = $0.0509\times D^{2}H\rho$ (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.

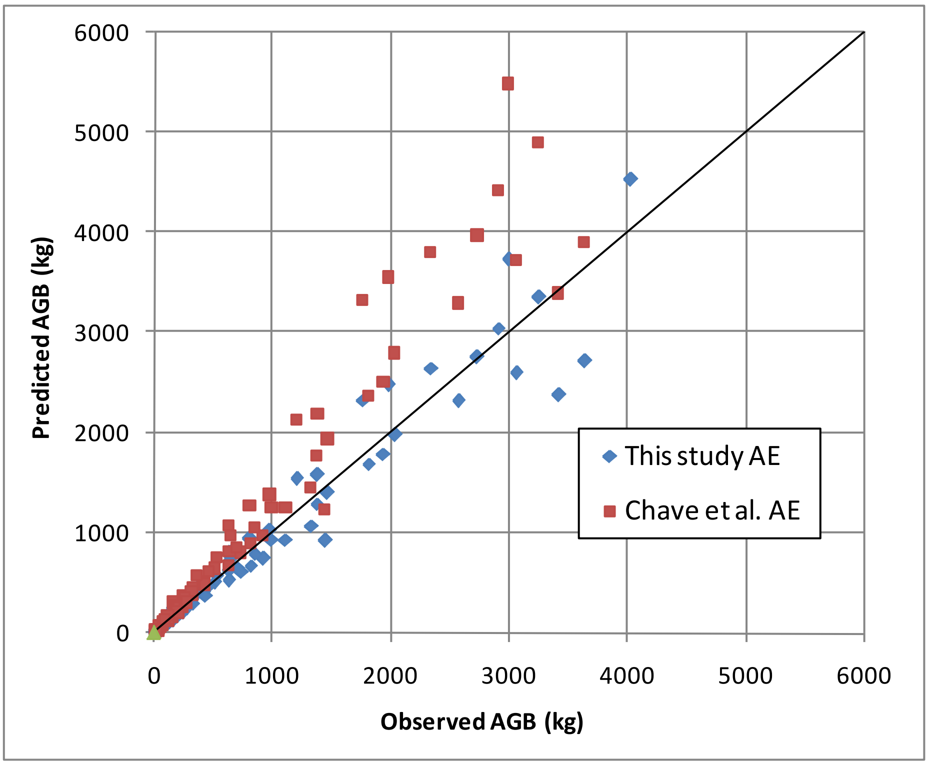


Figure 14: Comparison between Eq. (9) and the equation of Chave et al. (2005)

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:

$$\bar{S}(\%) = \frac{100}{n}\sum_{i=1}^{n}\frac{|\hat{Y}i - Yi|}{Yi} \qquad (10)$$

Where n is the number of sample trees; $\hat{Y}i$ and Yi are the predicted and measured AGB of the ith tree, respectively. The results are provided in Table 18.

Table 17: The standard deviation of different equations (evergreen broadleaf dataset)

Equation No.	Equation	S̄(%)	S(%)
Eq. (6)	$AGB = 0.1245 \times D^{2.4163}$	25.66%	-0.26%
Eq. (7)	$AGB = 0.0421 \times (D^2H)^{0.9440}$	24.67%	-1.83%
Eq. (8)	$AGB = 0.2105 \times (D^{2.4}\rho)^{1.0025}$	16.23%	2.26%
Eq. (9)	$AGB = 0.0704 \times (D^2H\rho)^{0.9389}$	13.73%	-0.84%
Eq. (Basuki)	$AGB = \exp((-1.201 + 2.196 \times \ln(D))$	43.11%	2.52%
Eq. (Brown)	$AGB = \exp(-2.134 + 2.530 \times \ln(D))$	45.39%	47.90%
Eq. (Chave)	$AGB = 0.0509 \times D^2H\rho$	26.51%	37.42%

As is observed, Eq. (9) from this study has the smallest S̄ (13.73%), followed by Eq. (8), (7) and (6), in that order. For the three previously developed equations, Eq. (Chave) performs the best in terms of S̄ value. 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. (6) in Figure 13, gives an S̄ value of 43.11%. Eq. (Brown) has the largest S̄ value (45.39%).

For the total AGB error *S*, equations developed in this study have the lowest values. Eq. (Basuki) overestimates the total AGB of the current dataset by 2.52%. Eq. (Chave) and Eq. (Brown) overestimate the total AGB by 37.42% and 47.90%, respectively.

3.4 Result 4: BCEF and BEF (AGBtotal/ABGstem)

The fractions of dry biomass for each component of the trees are given in Figure 14. It can be observed that stems occupy a large portion (81%) of the total above-ground biomass. Leaf biomass accounts for just 1.95% of the total above ground biomass.

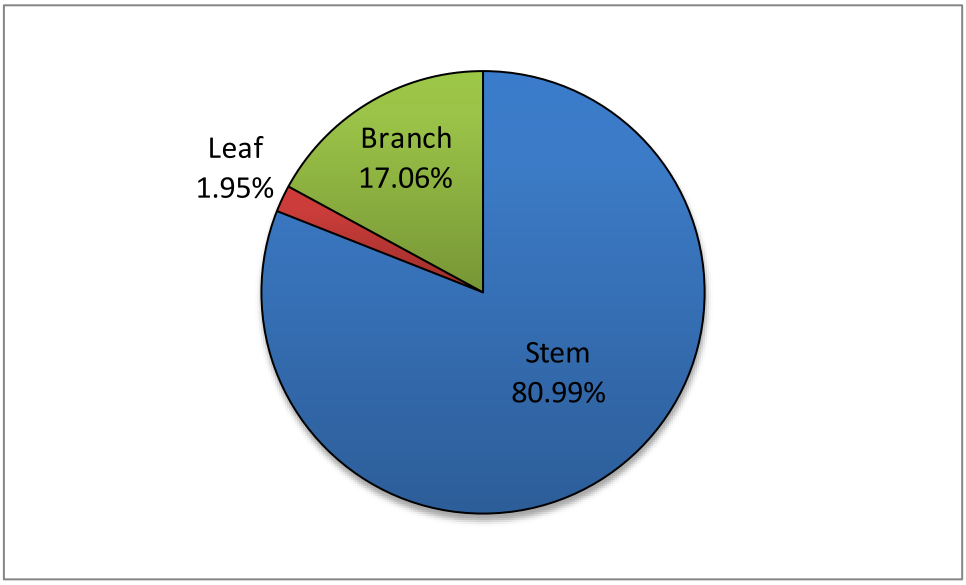


Figure 15: Fractions of dry biomass by tree components

Next, we calculated the BCEFs by dividing the dry biomass to the volume of each sample trees and examined the relationship between BCEF and DBH (Figure 14). The line in the figure shows the linear regression of the observed data. The very small value of R^2 (almost zero) indicates that BCEFs do not depend on DBH.

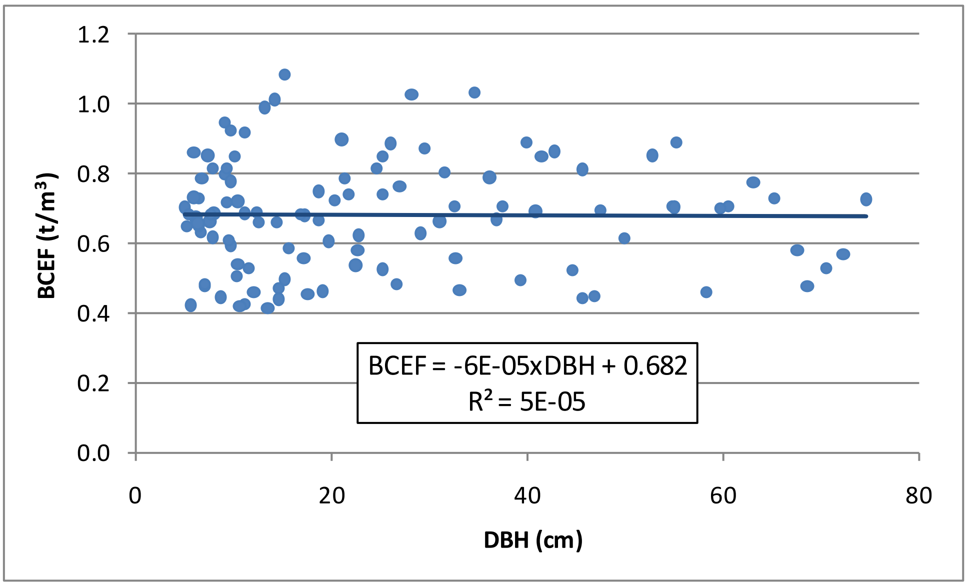


Figure 16: Relationship between BCEF (t/m3) and DBH (cm)

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 commercial 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:

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

The result for the 110 trees sampled in evergreen broadleaves forest is a BEF average value of 1.27 ± 0.13. The minimal value is 1.05 and the maximal is 1.69 (Figure 15).

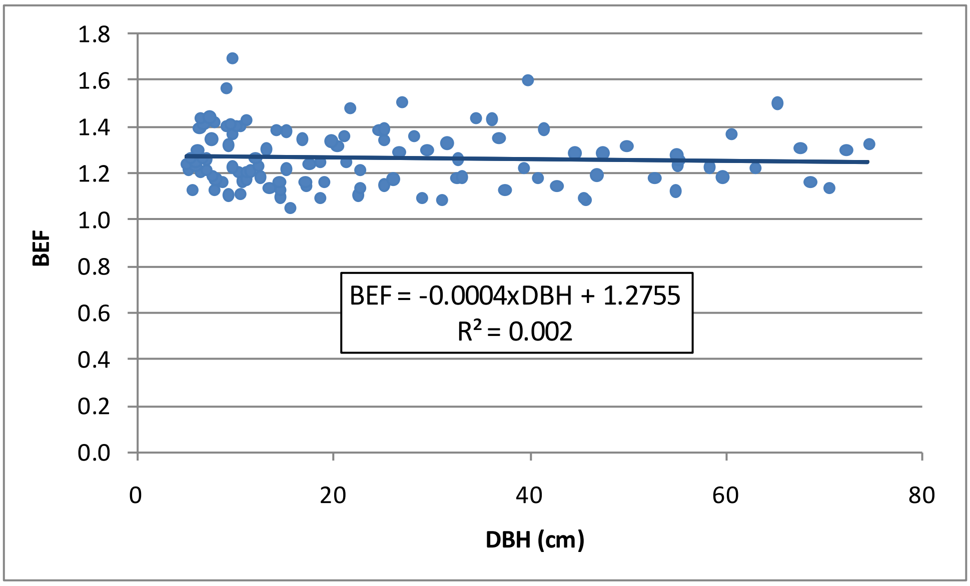


Figure 17: Relationship between BEF and DBH (cm)

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 forests in the Central Region of Vietnam. Destructive sampling was done to collect biomass data of 110 sample trees in two sample plots and these data were used as dependent variables in multiple regression analyses.

Totally eight statistical models were used for the regression analyses. Three regression approaches were applied. The first approach is to use the least squares optimization to the original models. The second approach is to use the least squares optimization to the logarithmically transformed forms of the original models. The third approach is to apply the maximum likelihood optimization to the original models. For equations developed using the least squares method (to both the original or transformed forms), the adjusted R squared and SSE values are used to measure the goodness of fit. For equations developed using the maximum likelihood method, the AICc was used to measure the goodness of fit.

The results of regression analyses of eight models which use various combinations of the input variables indicate that the inclusion of the height and wood density as additional input variables contributes to the improvement of the goodness of fit. Moreover, the inclusion of the wood density seems to improve the goodness of fit more than the inclusion of the height. Therefore, whenever these variables are available, equations that use them should be used to improve the accuracy and certainty of biomass estimation.

Cross validation tests were undertaken to assess the performance of the developed equations in practice and draw the ranges of errors for them. Based on the results of these tests, the following equations are suggested to be applied in practice:

Table 18: The recommended equations to be used in practice and their error assessment

No	Equation	Expected value of error* (%)	Range of error* (95% CI)
1	$AGB = 0.1245 \times D^{2.4163}$	0.101	-16.96% ÷ 20.61%
2	$AGB = 0.0421 \times (D^2H)^{0.9440}$	-1.205	-16.67% ÷ 17.70%
3	$AGB = 0.2105 \times (D^{2.4}\rho)^{1.0025}$	0.600	-14.03% ÷ 17.67%
4	$AGB = 0.0704 \times (D^2H\rho)^{0.9389}$	-0.737	-12.32% ÷ 11.70%

*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. The ranges of error here apply when predicting total AGB for datasets of 37 or more trees. For datasets with smaller number of trees, the ranges of error may be larger.

The recommended allometric equation which uses DBH as the only input variable is then compared with the Brown (1997) and Basuki et al. (2009) equations. The results show that the Brown's equation tends to over-estimate the biomass of the sample trees in the studied region. The Basuki et al. equation slightly under-estimates the biomass of the large sample trees but the differences are quite small. These results lead us to the recommendation that countries need to develop their own specific allometric equations in order to improve the accuracy of biomass and carbon stock assessment.

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.

Burnham, K. P., and Anderson, D.R. (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer-Verlag. ISBN 0-387-95364-7.

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 Vietnam, 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.

APPENDICES

Appendix 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[3].

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 gram/cm3 or ton/m3.

4. Biomass Expansion Factor (BEF)

A multiplication factor that expands 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.

5. 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).

6. 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.

7. Carbon stock

Carbon stock is the quantity of carbon in a pool.

8. Forest

---

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

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[4].

9. Root to shoot ratio (RS)

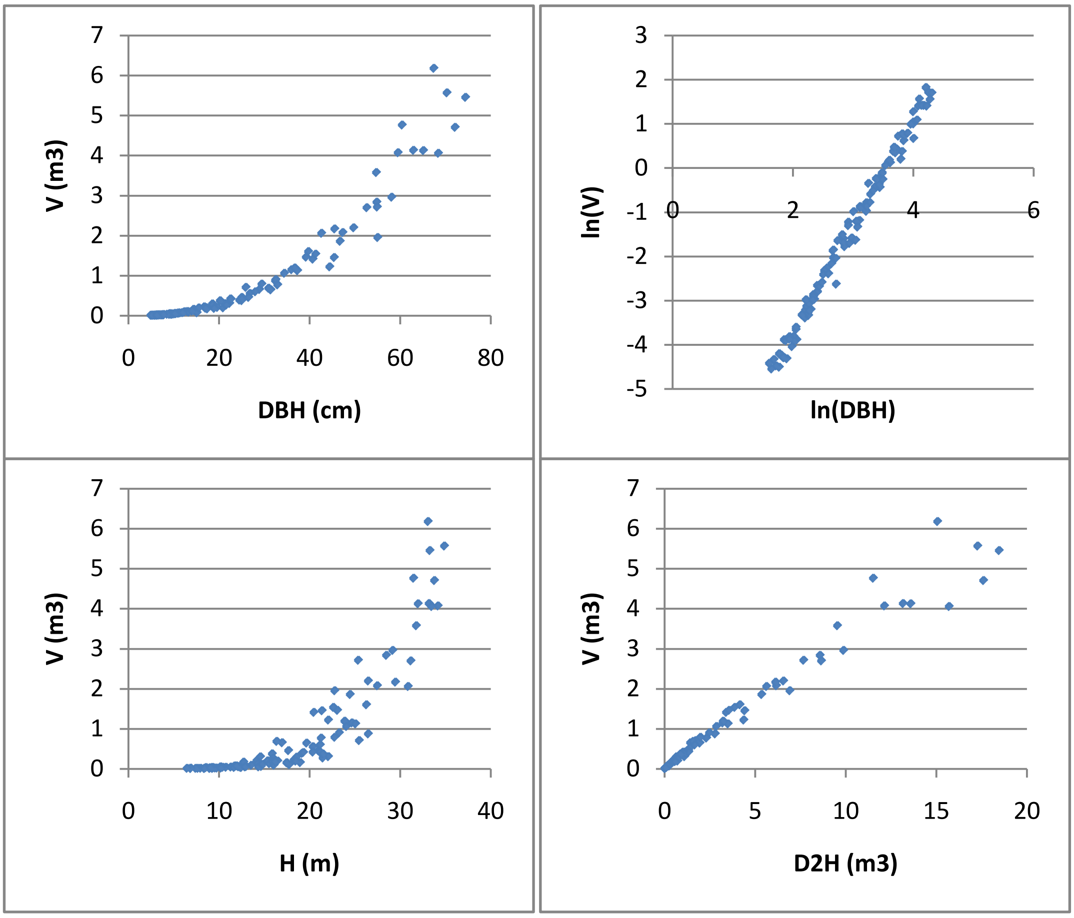
RS is defined as a ratio of below ground biomass of trees to above ground biomass of trees. RS is normally used to estimate below ground biomass of trees if above ground biomass of trees is known.

---

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

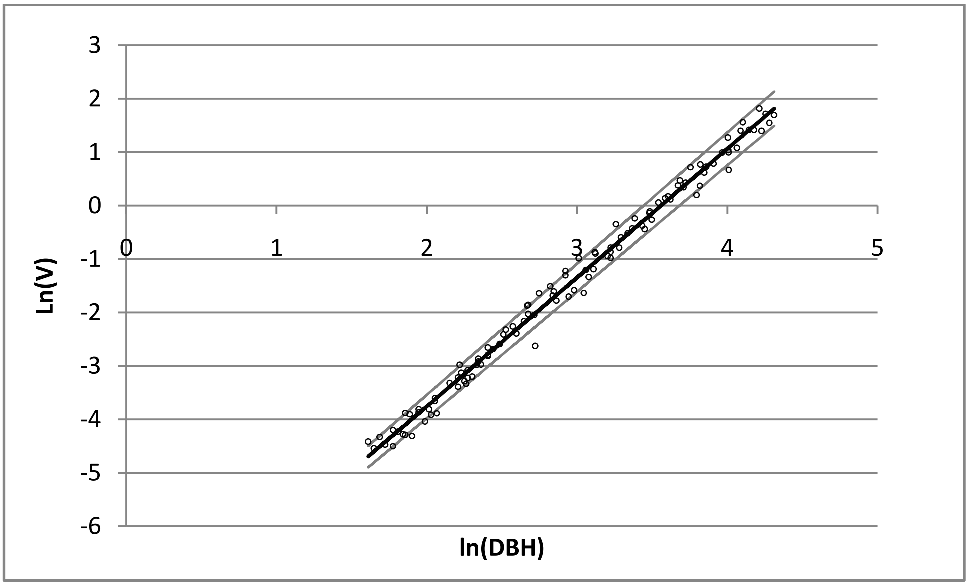
Appendix 6: graphs related to the modeling of stem volume

- Graphical exploration of the data

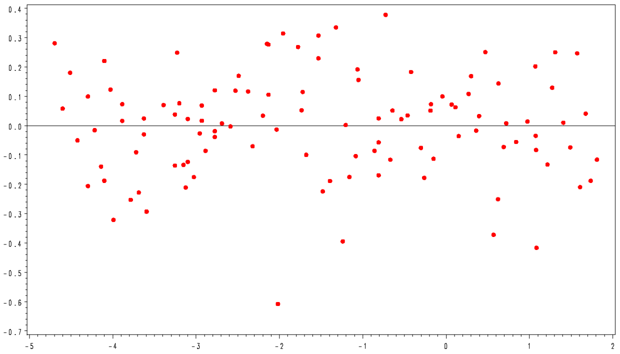


- Simple linear regression between ln(V) and ln (DBH):

$$\ln(V) = -8.57222 + 2.40883 \times \ln(DBH)$$



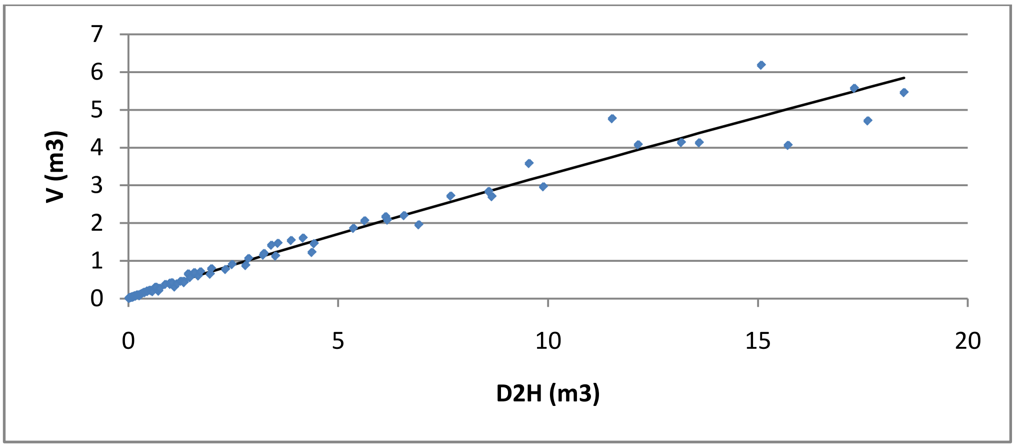
The points represent the felled trees, the black line represents the model and the grey lines its confidence interval at 95%.



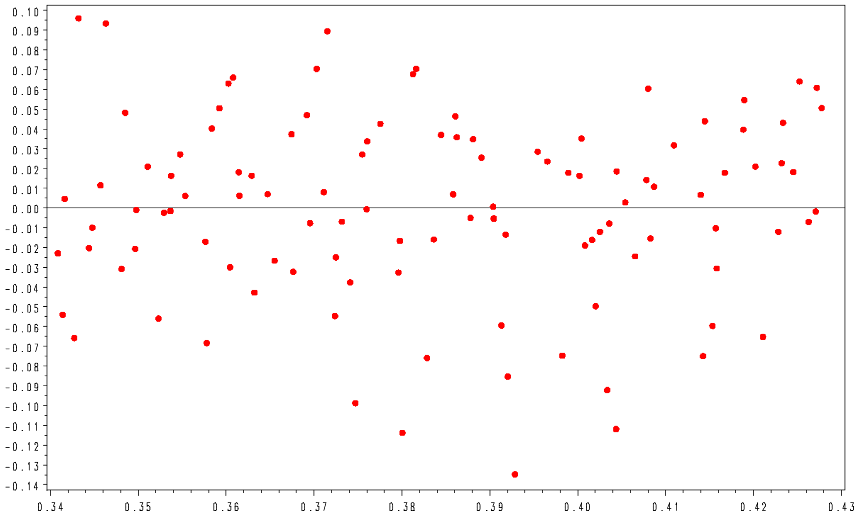
Residuals of the linear regression against the predicted values

- Non linear regression between V and D^2H:

$$V = 0.376022 \times (D^2H)^{0.940660} \text{ with } \varepsilon \sim N(\mu, 0.002114 \times (D^2H)^{1.948610})$$



The squares represent the felled trees and the black line represents the non linear model.



Weighted residuals against weighted predicted values of the non linear model