UN-REDD
PROGRAMME

Allometric equations at national scale for estimating tree and forest biomass in Viet Nam

Part B1: Equations for biomass of aboveground trees, branches and leaves biomass in Evergreen Broadleaved forests, and for aboveground biomass of six tree families in Evergreen and Deciduous forests

UN-REDD PROGRAMME
Viet Nam

July 2014
Ha Noi, Viet Nam

Allometric equations at national scale for estimating tree and forest biomass in Viet Nam

Part B1: Equations for biomass of aboveground trees, branches and leaves biomass in Evergreen Broadleaved forests, and for aboveground biomass of six tree families in Evergreen and Deciduous forests

By Bao Huy, Department of Forest Resources and Environment Management (FREM), Tay Nguyen University (TNU)

Editors: Gael Sola, Akiko Inoguchi, Matieu Henry (Food and Agriculture Organization of the United Nations)

Recommended citation:

Huy, B., 2014. Part B1: Equations for biomass of aboveground trees, branches and leaves biomass in Evergreen Broadleaved forests, and for aboveground biomass of six tree families in Evergreen and Deciduous forests. In: (eds) Sola, G. et al., (2014): Allometric equations at national scale for estimating tree and forest biomass in Viet Nam, UN-REDD Programme, Ha Noi, Viet Nam.

Acknowledgements:

The author would like acknowledge technical support during data analysis and model development with the R software provided by Gael Sola (FAO) for data contributed by Viet Nam Forestry University (VFU), Vietnamese Academy of Forest Sciences (VAFS), Forest Inventory and Planning Institute (FIPI) and TNU with funding from the UN-REDD Programme Viet Nam.

FOREWORD

This report is a part of a group of seven documents describing the context, rationale, detailed methodologies and results of the development of allometric equations at national scale to estimate tree biomass and other tree characteristics for the main forest types and eco-regions in Viet Nam. The recommended citation for the framework documents is:

Sola G., Phuong V.T., Huy B., Khoa P.V., Hung N.D., Xuan N.V., Inoguchi A. and Henry M., 2014. Allometric equations at national scale for estimating tree and forest biomass in Viet Nam, UN-REDD Programme, Ha Noi, Viet Nam.

This work was undertaken in 2014, with the support of the UN-REDD Viet Nam Phase II Programme and built on an extensive field measurement campaign supported by the UN-REDD Viet Nam Phase I Programme in 2012 and 2013. Four institutions collaborated on the field work and analysis: the Forest Inventory and Planning Institute (FIPI), Vietnamese Academy of Forest Sciences (VAFS), Viet Nam Forestry University (VFU) and Tay Nguyen University (TNU), with technical assistance from FAO.

In total, 139 models were developed to estimate tree aboveground biomass (AGB), biomass of tree branches and leaves, tree stem volumes, total tree heights and heights of first branches. Stem taper equations were also studied to better understand tree shapes and their influence on AGB.

This report, "Part B1: Equations for biomass of aboveground trees, branches and leaves biomass in Evergreen Broadleaved forests, and for aboveground biomass of six tree families in Evergreen and Deciduous forests", is intended to present a series of models to estimate tree characteristics in evergreen broadleaved forest and the detailed methodology for their development and validation. The output variables of the models included in this report are:

- Biomass of tree aboveground (AGB),
- Biomass of tree branches (B_{br}),
- Biomass of tree leaves (B_l),

For AGB, the general models are multispecies and additional models were developed for the main tree families.

For a better understanding of the context and the methodology, please see the report "Part A: Context, methodology and summary of the results". The report "Part C: Guidelines on the use of the allometric models" presents all the developed models in a simple format to help users find the most appropriate models to meet their needs.

A list of acronyms is also provided in the report: "Part A: Context, methodology and summary of the results".

EXECUTIVE SUMMARY

Context:

Evergreen broadleaf forest is the main forest type of Viet Nam thus no tools have been developped to estimate its carbon stock at national level. In the context of climate change mitigation through forestry sector and REDD+ mechanism, developing allometric equations with a dataset of tree biomass at national level could provide improved carbon stock estimates for Viet Nam, with a known accuracy.

Objectives:

FREM, with UN-REDD Viet Nam support, set out to develop a set of models to estimate biomass (aboveground biomass (AGB), branches, and leaves) in evergreen broadleaf forests (EBLF) at national level, AGB models for the main families in EBLF and Deciduous forests for comparison to pan-tropical and local models.

Methodology:

The dataset for EBLF included 860 trees located in five eco-regions of Viet Nam. DBH (diameter at breast height), H (tree height), WD (wood density) and all the combinations of these three tree characteristics were used as input variables. This dataset was merged with Deciduous forest data (969 tree in total) to select the families with more than 45 trees and another set of biomass models was established.

The modeling was performed by applying non-linear mixed effect models and power models on residuals, with or without random effects of eco-regions or environment variable to models' parameters. Indicators to select the best models were the Akaike information criterion (AIC), the sum of squared error (SSE), and R^2 as well as visual interpretation of each model. An independant dataset with 1 303 trees was used for validation, with bias (percent error of total trees, S%), efficiency factor (EF) and mean absolute percent error (percent error for individual tree, MAE%) the main indicators for validation of selected models and for comparison to pan-tropical and local models.

Results:

The power equation was appropriate for biomass models, with the overall **best model**[1] having three input variables: Biomass = a*[(DBH/100)^2*H*WD]^b. For other biomass models, random effects of eco-regions or WD classes improved estimates. The effect of Family on the subset of tree data (6 main families) was significant for models with two input variables (DBH and H or DBH and WD), but not for the model including three input variables, meaning that it adequately took into consideration the differences between families. The best models had EFs from 0.81-0.96. The models of AGB = a*DBH^b*WD and AGB = f(DBH,H,WD) had the highest accuracy with a S% bias under 3%, MAE% < 18% and EF of more than 0.95. By using the best models of this study, MAE% were reduced significantly by 14 percent compared to models from IPCC (2003) and Brown (1997) and 10-18 percent to Chave (2005, 2014). The selected models with eco-region and local models from previous studies had similar bias and EF.

Conclusion:

The models developed in this study are recommended for estimating forest carbon stocks in Viet Nam. The overall best model had three input vairables (DBH,H and WD) and should be prioritized. If wood density is not available, the models with specific parameters for each ecoregion are recommended.

---

[1] **The <u>best models</u> are the models with (1) no visuals issues in the graphs predictied values against observations and residuals againts predicted values, (2) the lowest AIC and SSE and (3) the simpliest from (lower number if input variables, lower number of parameters) for similar AIC or SSE.**

TABLE OF SELECTED ALLOMETRIC EQUATIONS

Input variables	Location	# trees	Equation
Aboveground biomass			
DBH	Central Highlands	114	$AGB = 0.198658 \times DBH^{2.415393}$
	North Central Coastal	331	$AGB = 0.121155 \times DBH^{2.415395}$
	Northeast	215	$AGB = 0.124830 \times DBH^{2.415395}$
	South Central Coastal	110	$AGB = 0.132507 \times DBH^{2.415395}$
	Southeast	110	$AGB = 0.120032 \times DBH^{2.415395}$
DBH and H	Central Highlands	114	$AGB = 363.43768 \times D2H^{0.94705}$
	North Central Coastal	331	$AGB = 254.49543 \times D2H^{0.94705}$
	Northeast	215	$AGB = 255.33956 \times D2H^{0.94705}$
	South Central Coastal	110	$AGB = 277.88007 \times D2H^{0.94705}$
	Southeast	110	$AGB = 235.21185 \times D2H^{0.94705}$
DBH and WD	Central Highlands	114	$AGB = 0.23342 \times DBH^{2.46615} \times WD$
	North Central Coastal	331	$AGB = 0.23342 \times DBH^{2.39720} \times WD$
	Northeast	215	$AGB = 0.23342 \times DBH^{2.39623} \times WD$
	South Central Coastal	110	$AGB = 0.23342 \times DBH^{2.40257} \times WD$
	Southeast	110	$AGB = 0.23342 \times DBH^{2.38600} \times WD$
DBH, H and WD	Nationwide	860	$AGB = 0.66609 \times D2HWD^{0.94304}$
Branches biomass			
DBH	Nationwide	860	$B_{branches} = 0.01339 \times DBH^{2.5601}$
DBH and H	Nationwide	860	$B_{branches} = 43.497 \times D2H^{0.9973}$
DBH and WD	Nationwide	860	$B_{branches} = 0.02426 \times DBH^{2.5494} \times WD$
Leaves biomass			
DBH	Central Highlands	114	$B_{leaves} = 0.06391 \times DBH^{1.71319}$
	North Central Coastal	331	$B_{leaves} = 0.06391 \times DBH^{1.57235}$
	Northeast	215	$B_{leaves} = 0.06391 \times DBH^{1.53303}$
	South Central Coastal	110	$B_{leaves} = 0.06391 \times DBH^{1.49935}$
	Southeast	110	$B_{leaves} = 0.06391 \times DBH^{1.38936}$
DBH and H	Central Highlands	114	$B_{leaves} = 14.23576 \times D2H^{0.61091}$
	North Central Coastal	331	$B_{leaves} = 9.33227 \times D2H^{0.62635}$
	Northeast	215	$B_{leaves} = 8.25209 \times D2H^{0.62959}$
	South Central Coastal	110	$B_{leaves} = 7.12003 \times DBH^{0.63342}$
	Southeast	110	$B_{leaves} = 4.68999 \times DBH^{0.64100}$
DBH and WD	Central Highlands	114	$B_{leaves} = 0.10691 \times DBH^{1.66816} \times WD$
	North Central Coastal	331	$B_{leaves} = 0.10691 \times DBH^{1.59387} \times WD$
	Northeast	215	$B_{leaves} = 0.10691 \times DBH^{1.53087} \times WD$
	South Central Coastal	110	$B_{leaves} = 0.10691 \times DBH^{1.50024} \times WD$
	Southeast	110	$B_{leaves} = 0.10691 \times DBH^{1.40358} \times WD$
DBH, H and WD	Central Highlands	114	$B_{leaves} = 0.22268 \times D2HWD^{0.62880}$
	North Central Coastal	331	$B_{leaves} = 0.17776 \times D2HWD^{0.62880}$
	Northeast	215	$B_{leaves} = 0.15789 \times D2HWD^{0.62880}$
	South Central Coastal	110	$B_{leaves} = 0.12845 \times D2HWD^{0.62880}$
	Southeast	110	$B_{leaves} = 0.08898 \times D2HWD^{0.62880}$

The models developed at Family levels aimed to better understand how the broader models competed with them but are not recommended for use as they are based on very small number of trees.

ACRONYMS

AGB: Aboveground Biomass, including biomass of stem, bark, branches and leave (kg/tree)

AIC: Akaike information criterion

BA: Basal area (m^2/ha)

Bba: Bark biomass (kg/tree)

Bbr: Branches biomass (kg/tree)

BEF: Biomass expansion factor

Bl: Leaves biomass (kg/tree)

Bst: Stem biomass (kg/tree)

Cm: Centimetres

DBH: Diameter at breast height (cm)

DE: Deciduous forest

EBLF: Evergreen broad leaf forest

EF: Efficiency factor

FAO: Food and Agriculture Organization of the United Nations

FIPI: Forest Inventory and Planning Institute

FREM: Department of Forest Resources and Environment Management

H: Tree height (m)

Ha: Hectares

IPCC: Intergovernmental Panel on Climate Change

M: Metres

Mm: Millimetres

MARD: Ministry of Agriculture and Rural Development

MRV: Measuring, reporting and verifying system

Nlme: Non-linear mixed effect models

REDD: Reducing Emissions from Degradation and Degraded Forest

TNU: Tay Nguyen University

VAFS: Vietnamese Academy of Forest Sciences

UNFCCC: United Nations Framework Convention on Climate Change

VFU: Viet Nam Forestry University

WD: Wood density (g/cm^3)

SSE: sum of squared errors

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

In the context of global climate change, forest management to mitigate climate change through CO_2 absorption by forest ecosystems deserves urgent attention from governments. To help support this need, the UN-REDD Programme has been taking action in developing countries and in Viet Nam since 2009. The IPCC, a scientific body set up under the auspices of the UN, in 1996, 2003 and 2006 also provided guidelines to measure and monitor forest carbon. However, there is a significant need globally and in Viet Nam to develop models for biomass and carbon estimations for national measuring, reporting and verifying (MRV) systems and produce accurate emission factors and reliable training datasets of biomass, carbon per hectare (ha) for activity data.

For rainforests over the tropics, authors have provided biomass models such as IPCC (2003), Brown et al. (1997), Chave et al. (2005, 2014) and Basuki et al. (2009). But, these developed models have no data on forest types, ecological zones and have not been evaluated for reliability in Viet Nam. In fact, some local equations were developed during Phase I of UN-REDD, but local models were found to be superior. However, these local models could not be effectively compared due to methodology differences used for their development. Thus, analysis of the national scale dataset was required.

The tree biomass samples used to develop equations during UN-REDD Phase I Programme were collected in forest types from different eco-regions. This study has further analysed this data to achieve the following objectives:

- Develop models to estimate biomass in EBLFs, including models for AGB, biomass parts of trees such as branches and leaves.
- Develop models to estimate biomass for most plant families and each main family in EBLF and Deciduous forests.
- Consider eco-regions and environment factors to set up models with adaptive parameters to increase the reliability of biomass estimates in different forest ecological conditions in Viet Nam.
- Consider forests and environmental parameters to establish models with parameter changes to improve the reliability of biomass estimates in different environmental conditions in the country.
- Validate the reliability and accuracy of selected models in this study and compare with local and pan-tropical models to provide a proposal to apply these models in the UN-REDD Programme in Viet Nam.

2 METHODOLOGY

2.1 Dataset description

For two forest types, Evergreen Broadleaf (EBLF) and Deciduous (DE), a harmonised dataset was created in regrouping data from five of the eight eco-regions classified by the Ministry of Agriculture and Rural Development (MARD). The data was collected with the support of UN-REDD Phase I Programme. The number of tree samples for each forest type, eco-region and the institutions responsible for collecting the data are shown in Table 1. More details on the methodology for data collection can be found in the Part A document: "Context, methodology and summary of the results".

Within each of the 1 ha sample plots, the number of trees sampled was determined based on the ratio of trees within each diameter class. In each sample plot, the fresh biomass of stems, branches and leaves from 55 trees were collected and prepared for biomass and WD calculations. To find the fresh to dry ratio of each tree to calculate the biomass, samples were taken from stem, branches and old and new leaves. For WD, samples were taken from every one-fourth or fifth of stem length. For further information on the data and methodology, refer to Part A of this document.

Table 1: Number of sampled trees in eco-zones and forest types

Eco-regions	DE	EBLF	Grand Total	Institution responsible for data collection
Central Highlands	54	114	168	VAFS
North Central Coastal		311	311	VFU
Northeast		215	215	FIPI
South Central Coastal		110	110	TNU
Southeast	55	110	165	FIPI
Grand Total	**109**	**860**	**969**	

2.2 Model development and selection

2.2.1 Model development

The development of biomass models included AGB, branches biomass (Bbr) and leaves biomass (Bl) with a single or group of input variables such as DBH (diameter at breast height, cm), H (tree height in m) and WD (wood specific gravity in $g.cm^{-3}$).

The list of models to test for each group of input variables was comprised of:

- DBH:
    - Biomass = a*DBH^b Eq. 1
    - Biomass = a + b*DBH + c*DBH^2 Eq. 2
    - Biomass = a + b*DBH + c*DBH^2 + d*DBH^3 Eq. 3
- DBH + H:
    - Biomass = a*DBH2H Eq. 4
    - Biomass = a*DBH2H^b Eq. 5
    - Biomass = a*DBH^b*H^c Eq. 6
- DBH + WD:

- Biomass = a*DBH^b*WD Eq. 7
- Biomass = a*DBH^b*WD^c Eq. 8

- DBH + H + WD:
  - Biomass = a*DBH2HWD^b Eq. 9
  - Biomass = a*DBH^b*H^c*WD^d Eq. 10

Group of input variables was calculated as follows:

$$\text{DBH2H (m3)} = \left(\frac{DBH}{100}\right)^2 \times \text{H} \quad \text{Eq. 11}$$

$$\text{DBH2HWD (kg)} = \text{DBH2H} \times \text{WD} \times 1000 \quad \text{Eq. 12}$$

All these models were developed and the best model forms selected (see the section on model selection below) for each combination of input variables. The effect of eco-regions and environmental variables was then tested on the selected models and specific model parameters were developed when the random effects improved the models.

2.2.2 Random effect tested

The effect of the following variables was tested:

i) Ecological zone:

- Vietnamese region (MARD agro-zones): Five ecological zones - Central Highlands, North Central Coastal, Northeast, South Central Coastal and Southeast.
- Biome WWF: Six zones - northern Annamites rainforests, northern Indochina subtropical forests, south China-Viet Nam subtropical evergreen forests, southeastern Indochina dry evergreen forests, southern Annamites mountain rainforests and southern Viet Nam lowland dry forests.
- Biome FAO: Two zones: Tropical moist Deciduous forest and tropical rainforest.
- Biome Holdridge: Four zones: Cool temperate wet forest, subtropical moist forest, subtropical wet forest and tropical dry forest

ii) Environmental variable:

- Basal area classes: Four classes - Poor: BA ≤ 10 m^2/ha, Medium: 10 < BA ≤ 20 m^2/ha, Rich: 20 < BA ≤ 30 m^2/ha and Very Rich: BA > 30 m^2/ha.
- WD classes: Four classes - WD1 = 0-0.4, WD2 = 0.4-0.6, WD3 = 0.6-0.8 and WD4 >0.8, in $g.cm^{-3}$,
- Soil type (from an old soil map of Indochina): Three types of soil - Cristalline shists, Igneous rocks and Sedimentary rocks
- Soil type from HWSD (Harmonised World Soil Database): Two types - Acrisols and Ferralsols
- The length of dry season (number of months with less than 60 mm rain): One, two, three or five months
- Rain classes: Five classes - Rain1 <1 400, Rain2 = 1 400-1 600, Rain3 = 1 600-1 800, Rain4 = 1 800-2 000, and Rain5 >2 000 mm/year.

The modelling was performed by applying non-linear mixed effect models (nlme) in R software with power models on residuals. More details on the statistical tools are provided in the report part A.

Example of R Script with model: AGB = a*DBH2H^b

Non-linear models - Maximum likelihood, no mixed effect: nlme

```
start <- coefficients(lm(log(AGB)~log(DBH2H_m3 ), data=t))
names(start) <- c("a","b")
start[1]<-exp(start[1])
Max_like1 <- nlme(AGB~a*DBH2H_m3^b, data=cbind(t,g="a"), fixed=a+b~1,
                  start=start, groups=~g, weights=varPower(form=~DBH2H_m3))
```

Non-linear models - Maximum likelihood, mixed effect on Region: nlme

```
start <- coefficients(lm(log(AGB)~log(DBH2H_m3), data=t))
names(start) <- c("a","b")
start[1]<-exp(start[1])
Max_like2 <- nlme(AGB~a*DBH2H_m3^b, data=t, fixed=a+b~1, random=a+b~1,
                  start=start, groups=~Region,
weights=varPower(form=~DBH2H_m3))
fixef(Max_like2)
ranef(Max_like2)
```

2.2.3 Indicators used for model selection

For each group of input variables, one was selected and considered the best model. The criteria used to select the best models were (Picard, Saint Andre et al. (2012)):

- **Visible issues in the three graphs**: predicted values and observations against input variables, predicted values against observations, and residuals (or weighted residuals) against predicted values.
- **AIC: Akaike information criterion**. The model with the smaller AIC value is preferred:

$$\mathrm{AIC} = -2\ln(\mathrm{L}) + 2\mathrm{p} \qquad \text{Eq. 13}$$

where, L is the likelihood of the fitted model and p is the total number of parameters in the model.

- **SSE: Sum of squared errors**. The model with the smaller SSE value is preferred:

$$\mathrm{SSE} = \sum_{i=1}^{n}(\mathrm{Yi} - \mathrm{Yipre})^2 \qquad \text{Eq. 14}$$

where, Yipre: the predicted biomass, Yi: the observed biomass, n = number of observations.

- **R^2: Coefficient of determination of the regression.** Generally, the highest R^2 value with statistical significance level exhibits the optimal model. In some cases, despite the R^2 value being high, the model is not optimal. Therefore, this criterion is considered after the other ones presented above.

Finally, when two models had very close values to all these indicators, the model with a smaller number of parameters was chosen even if it was slightly worse than another model with more parameters.

2.3 Model validation and comparison to local and pan-tropical models

An independent dataset for validation was set up. The independent data was synthesized by combination of volume dataset (from FIPI) and WD database per species (from VAFS, VFU and TNU). Some 1 303 independent trees were collected, mainly located in two eco-regions (North Central Coast and Northeast of Viet Nam) (Table 2). More information on this dataset can be found in the report part A.

Based on this independent data, all the selected models of this study were validated and compared with local models developed with support of UN-REDD Phase 1 and to pan-tropical models of Brown (1997), IPCC (2003) and Chave et al. (2005, 2014).

Table 2: Range of tree DBH in the independent dataset (in cm).

Region	Number of trees	Average DBH	Min DBH	Max DBH
North Central Coastal	520	37.3	4.8	125.0
Northeast	783	26.4	5.0	100.0
Total	1303	30.8	4.8	125.0

The validation indicators used Bias (S%), efficiency factor (EF) and mean absolute percent error (MAE%):

- **S%**: Bias. It stands for the percentage of average error (predicted values minus observations) for a group of trees. Smaller values are preferred (Chave et al., 2014):

$$S\% = 100 * \frac{\sum_1^n (\mathrm{Yipre} - \mathrm{Yi})}{\sum_1^n Yi} \qquad \text{Eq. 15}$$

where, Yipre: the predicted biomass, Yi: the observed biomass, n = number of trees for validation.

- **Efficiency Factor (EF)**: This is a dimensionless statistic which relates model predictions to observed data (Loague and Green (1991) referred by Mayer et al., 1993).

$$\mathrm{EF} = 1 - \frac{\sum_1^n (\mathrm{Yi} - \mathrm{Yipre})^2}{\sum_1^n (\mathrm{Yi} - \overline{\mathrm{Y}})^2} \qquad \text{Eq. 16}$$

where, Yipre: the predicted biomass, Yi: the observed biomass, $\overline{\mathrm{Y}}$ = the average of observations, n = number of trees for validation. A good fit model has an EF close to one.

- **MAE%**: Mean absolute percent error (or Average deviation percent), gives the average absolute error for a single tree prediction. Smaller MAE% value is preferred (Mayer et al., 1993):

$$\mathrm{MAE\%} = \frac{100}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{|\mathrm{Yipre} - \mathrm{Yi}|}{\mathrm{Yi}} \qquad \text{Eq. 17}$$

where, Yipre: the predicted biomass, Yi: the observed biomass, $\overline{\mathrm{Y}}$ = the average of observations and n = number of trees for validation.

Finally, the graphs of the observed data vs predicted by different models are presented. Table 3 presents the local and pan-tropical models validated and compared to the selected models in this study.

Table 3: Pan-tropical and local models compared with selected models in this study.

Author (year)	Models	Location, zone	Equation number
Brown (1997)	$AGB = \exp(- \ -2.134 + 2.530 * \log(DBH))$	Pan-tropical zone	Eq. 18
IPCC (2003)	AGB=exp(- 2.289 + 2.649*log(DBH) - 0.021*(log(DBH))^2)	Pan-tropical zone	Eq. 19
Chave I (2005)	$AGB = WD * \exp(-1.499 + 2.148 * \log(DBH) + 0.207 * (\log(DBH))^2 - 0.0281 * (\log(DBH))^3)$	Pan-tropical zone	Eq. 20
Chave II (2005)	AGB = exp(- 2.977 + log(WD*DBH^2*H)) = 0.0509*WD*DBH^2*H	Pan-tropical zone	Eq. 21
Chave III (2014)	AGB = 0.0673*(WD*DBH^2*H)^0.976	Pan-tropical zone	Eq. 22
Hung et al., FIPI, UN-REDD Viet Nam Phase I (2012)	AGB = 0.0547*D^2.1148*H^0.6131	Northeast, (NE)	Eq. 23
	AGB = 0.0421*(D^2H)^0.9440	North Central Coastal (NCC)	Eq. 24

3 ALLOMETRIC EQUATIONS DEVELOPMENT AND SELECTION

3.1 Characteristics of study forests

3.1.1 Characteristics of EBLF dataset

For EBLF, the total number of sampled trees was 860. The number of tree per DBH class and eco-regions is shown in Table 4. The DBH range of sampled trees was from 4.9 to 87.7 cm, with the average DBH 28.1 cm (Table 5).

Table 4: Number of sampled trees of EBLF per DBH Class and Region

DBH Class (cm)	Central Highlands	North Central Coastal	Northeast	South Central Coastal	Southeast	Total number of sampled trees
10	12	49	38	28	11	138
20	22	73	60	37	29	221
30	19	66	40	10	26	161
40	22	53	36	8	20	139
50	19	28	21	7	9	84
60	7	22	10	10	7	56
70	9	16	6	6	6	43
80	4	4	3	2	2	15
90	0	0	1	2	0	3
Total	114	311	215	110	110	860

Table 5: Range of DBH of sampled trees per region for EBLF

Eco-region	Average of DBH (cm)	Min of DBH (cm)	Max of DBH (cm)
Central Highlands	32.9	6.1	73.8
North Central Coastal	28.2	5.0	74.5
Northeast	26.2	5.4	81.8
South Central Coastal	25.7	4.9	87.7
Southeast	28.5	6.8	72.5
Grand Total	28.1	4.9	87.7

Most of the trees had a total height between 15 and 30 m, with a min of 4.7 and a maximum height 43.8, the average being 19.2 cm (Table 6).

Table 6: Range of H of sampled trees per region for EBLF

Eco-region	Average of H (m)	Min of H (m)	Max of H (m)
Central Highlands	23.2	6.7	43.8
North Central Coastal	18.4	6.5	36.2
Northeast	18.4	6.0	36.6
South Central Coastal	17.5	4.7	41.4
Southeast	20.3	6.8	38.0
Grand Total	19.2	4.7	43.8

3.1.2 Characteristics of dataset for main families in EBLF and Deciduous forests

To develop models for plant families and select families with more than 40 sampled trees in two forest types (Evergreen and Deciduous) with existing datasets, six families were selected. Of the six families, only the Dipterocarpaceae family had a almost the same number of trees in both forest types, other tree family were largely in EBLF.

Table 7: Number of sampled trees per main families in each forest type and eco-region

Family / Forest Type	Central Highlands	North Central Coastal	Northeast	South Central Coastal	Southeast	Total
Dipterocarpaceae	35	29	19	7	51	141
Deciduous	35				24	59
EBLF		29	19	7	27	82
Euphorbiaceae	1	32	15	4		52
EBLF	1	32	15	4		52
Fagaceae	39	25	24	7		95
Deciduous	8					8
EBLF	31	25	24	7		87
Lauraceae	6	25	30	3	2	66
Deciduous					2	2
EBLF	6	25	30	3		64
Leguminosae	3	34	19		28	84
Deciduous	2				9	11
EBLF	1	34	19		19	73
Myrtaceae	29	11	7	8	8	63
Deciduous					4	4
EBLF	29	11	7	8	4	59
Total	**113**	**156**	**114**	**29**	**89**	**501**

Sampled trees of the main families were located mostly in DBH classes from 20-60 cm with few trees In the 80-90 cm classes) and in 10-30 m height classes. Table 8 showed the range of DBH and H of sampled trees per main families. Generally, the DBH range is from 5.0-87.7 cm and H is from 5.7-41.4 m.

Table 8: Range of DBH and H of sampled trees per main families

Plant Family	Average of DBH (cm)	Min of DBH (cm)	Max of DBH (cm)	Average of H (m)	Min of H (m)	Max of H (m)
Dipterocarpaceae	31.4	5.0	87.7	18.8	5.7	41.4
Euphorbiaceae	32.2	5.9	79.0	20.4	6.9	34.5
Fagaceae	30.2	5.6	70.3	19.6	6.2	36.6
Lauraceae	24.2	5.2	72.2	18.1	7.0	35.5
Leguminosae	27.0	7.0	65.3	18.2	6.8	33.6
Myrtaceae	27.9	6.1	65.9	19.7	6.0	30.4
Grand Total	29.1	5.0	87.7	19.0	5.7	41.4

There was a difference in WD values among the six main families and between regions (Figure 1). This indicates the significance of WD variables in biomass models in tropical forests. Besides within one family, WD values were changed by region effect, so models with random effect on regions could become important to take into consideration these differences. This needs attention if the accuracy of biomass estimates is to be improved.

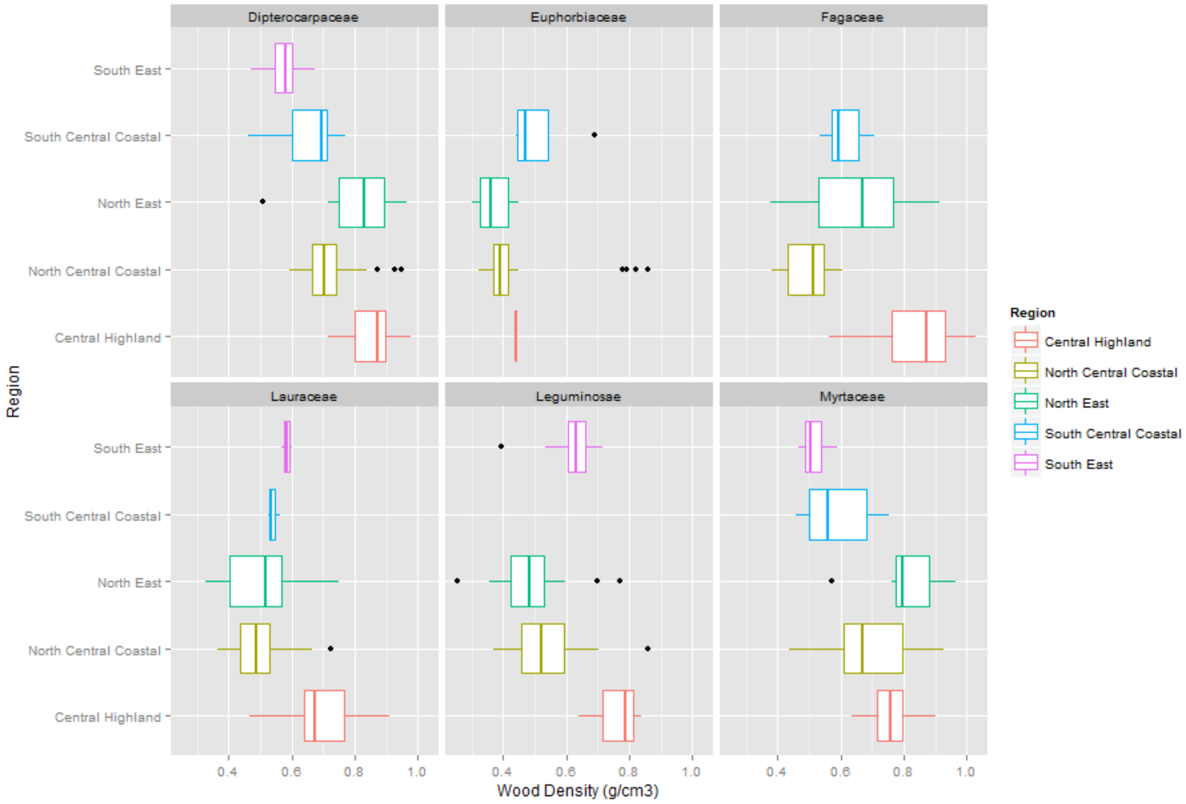


Figure 1: WD value of six main families and per eco-region

3.1.3 Wood density value of 10 main families in EBLF per eco-region

Table 9 offers descriptive statistics on the WD values of 10 main families in EBLF.

Table 9: Descriptive statistics of WD value of 10 main families of EBLF

Families	Number of trees	Average of WD (g/cm3)	Min of WD (g/cm3)	Max of WD (g/cm3)	StdDev of WD (g/cm3)
Burseraceae	32	0.595196	0.441000	0.824000	0.108154
Dipterocarpaceae	82	0.682868	0.462200	0.963556	0.129099
Euphorbiaceae	52	0.424995	0.299800	0.860600	0.128095
Fagaceae	87	0.668677	0.376556	0.959509	0.177933
Lauraceae	64	0.524767	0.328200	0.908841	0.120970
Leguminosae	73	0.548643	0.254600	0.859800	0.109279
Magnoliaceae	25	0.509447	0.292200	0.662690	0.110350
Myrtaceae	59	0.714358	0.437000	0.962400	0.128265
Theaceae	25	0.603433	0.390000	0.781610	0.113515
Ulmaceae	27	0.473996	0.384000	0.688600	0.065021
Grand Total	526	0.592622	0.254600	0.963556	0.158673

3.2 Aboveground biomass (AGB) in EBLF

This study tested AGB non-linear models per group of input variables based on DBH, H and WD. The best model forms were selected based on AIC, SSE and R^2 and random effects were tested on their parameters.

3.2.1 Model AGB = f(DBH): Comparison of model forms and selection of best equations

Alternative model forms were compared to select the optimum one, using AIC, SSE and R^2 to determine selection of equations. By comparing the three model forms in Table 10, the optimal result was generated through the model form: AGB = a*DBH^b.

When testing random effects on the power model, the results also indicated that eco-regions (MARD), Biome_WWF and four environment parameters including WD class, Soil type, Soil type_HWSD, Dry season length affected AGB = a*DBH^b model. Compared to the model without random effect, the models with random effect reduced SSE and AIC and increased R^2. This shows a lack of H or WD variables, random effect models on ecological zones and environment factors important to improve the reliability of biomass estimations.

Table 10: Comparison of different AGB = f(DBH) models with and without random effect

Id	Model form	Random effect	Residual function	AIC	SSE	Adj. R^2
1	AGB = a*DBH^b	No	1/DBH^k	10 201	151 350 333	0.847
2	AGB = a + b*DBH + c*DBH^2	No	1/DBH^k	10 215	161 457 982	0.837
3	AGB = a + b*DBH + c*DBH^2 + d*DBH^3	No	1/DBH^k	10 214	149 857 129	0.848
4(*)	**AGB = a*DBH^b**	**MARD**	**1/DBH^k**	**9 957**	**144 519 193**	**0.854**
5	AGB = a*DBH^b	WWF	1/DBH^k	9 952	143 869 350	0.855
6	AGB = a*DBH^b	WD classes	1/DBH^k	9 643	105 399 026	0.893
7	AGB = a*DBH^b	Soil type	1/DBH^k	10 184	154 018 354	0.844
8	AGB = a*DBH^b	Soil type HWSD	1/DBH^k	9 950	149 368 263	0.849
9	AGB = a*DBH^b	Dry season length	1/DBH2H^k	10 039	125 279 754	0.873

(*) Selected Model

The selected model with only DBH as input variable has a random effect of MARD eco-regions of MARD (in detailed equations are presented in Table 11) because of the simple ecological zoning, despite the AIC value being a little larger than with a random effect of WD classes. WD classes dependent models were not slected because WD classes are not commonly measured or reported, but the results show that WD improves more the quality of biomass models than ecological zoning. The models with random effect on Soil type (HWSD) gave similar AICs to the model with random effect on eco-zones and could be an artefact as there were only two classes and two plots from the Ferrasols class. More data from Ferrasols in other areas need to confirm this effect.

Table 11: The selected model AGB = a*DBH^b with random effect of eco-region

Id	Random effect class	N trees	Equation
1	All trees	860	AGB = 0.139436*DBH^2.415395
2	Central Highlands	114	AGB = 0.198658*DBH^2.415393
3	North Central Coastal	331	AGB = 0.121155*DBH^2.415395
4	Northeast	215	AGB = 0.124830*DBH^2.415395
5	South Central Coastal	110	AGB = 0.132507*DBH^2.415395
6	Southeast	110	AGB = 0.120032*DBH^2.415395

3.2.2 Model AGB = f(DBH, H): Comparison of model forms and selection of the best equations

On comparing the three model forms in Table 12, the optimal result was generated through the following model form: AGB = a*DBH2H^b with the lowest AIC and SSE values. DBH2H is calculated as the surrogate of tree volume:

DBH2H = (DBH/100)^2*H

With DBH in cm and H in m. DBH2H is in m^3.

Table 12: Comparison of different AGB = f(DBH, H) models with and without random effect

Id	Model form	Random effect	Residual function	AIC	SSE	Adj. R^2
1	AGB = a*DBH2H	No	1/DBH2H^k	10 020	153 249 974	0.845
2	AGB = a*DBH2H^b	No	1/DBH2H^k	9 972	135 723 726	0.863
3	AGB = a*DBH^b*H^c	No	1/DBH^k	10 042	135 479 428	0.863
4(*)	**AGB = a*DBH2H^b**	**MARD**	**1/DBH2H^k**	**9 816**	**121 080 777**	**0.878**
5	AGB = a*DBH2H^b	WWF	1/DBH2H^k	9 804	119 882 454	0.879
6	AGB = a*DBH2H^b	Biome Holdridge	1/DBH2H^k	9 957	131 955 659	0.867
7	AGB = a*DBH2H^b	WD classes	1/DBH2H^k	9 416	80 205 987	0.919
8	AGB = a*DBH2H^b	Soil type	1/DBH2H^k	9 962	135 756 150	0.863
9	AGB = a*DBH2H^b	Soil type HWSD	1/DBH2H^k	9 820	127 399 409	0.871
10	AGB = a*DBH2H^b	Dry season length	1/DBH2H^k	9 875	116 263 802	0.882
11	AGB = a*DBH2H^b	Rain classes	1/DBH2H^k	9 910	117 884 075	0.881

(*) Selected Model

The results indicated that eco-regions (MARD), Biome_WWF and Biome_Holdridge and five environment factors including WD class, Soil type, Soil type_HWSD, Dry season length, Rain_class affected AGB = a*DBH2H^b model. The models with random effect on these factors significantly reduced SSE and AIC values when compared with the model without random effect.

The two models with random effect on MARD and WWF classes had very close values of all these indicators. The model with random effect on MARD classes with less parameters was chosen (Table 13), despite being slightly worse than models with more parameters of WWF classes. The overall best model here was again the power model with random effect of WD classes. It was not selected as WD classes are not developed for most tree species, but it emphasizes the importance of Wood density for biomass models.

The best model with DBH and H had better AIC value than the best model with DBH only, indicating that tree total height is valuable information to be collected in the field to improve forest biomass estimates.

Table 13: The selected model AGB = a*DBH2H^b with random effect on eco-region

Id	Random effect class	N trees	Equation
1	All trees	860	AGB = 277.27292*DBH2H^0.94705
2	Central Highlands	114	AGB = 363.43768*DBH2H^0.94705
3	North Central Coastal	331	AGB = 254.49543*DBH2H^0.94705
4	Northeast	215	AGB = 255.33956*DBH2H^0.94705
5	South Central Coastal	110	AGB = 277.88007*DBH2H^0.94705
6	Southeast	110	AGB = 235.21185*DBH2H^0.94705

3.2.3 Model AGB = f(DBH, WD): Comparison of model forms and selection of the best equations

Among two model forms were tested, AGB = a*DBH^b*WD had the lowest AIC and SSE and was therefore selected (Table 14). DBH is expressed in cm and WD in $g.cm^{-3}$.

Table 14: Comparison of different AGB = f(DBH, WD) models with and without random effect

Id	Model form	Random effect	Residual function	AIC	SSE	Adj. R2
1	AGB = a*DBH^b*WD	No	1/DBH^k	9 536	87 118 189	0.912
2	AGB = a*DBH^b*WD^c	No	1/DBH^k	9 538	88 234 206	0.911
3(*)	**AGB = a*DBH^b*WD**	**MARD**	**1/DBH^k**	**9 404**	**80 739 721**	**0.918**
4	AGB = a*DBH^b*WD	WWF	1/DBH^k	9 414	77 821 331	0.921

(*) Selected Model

With random effect tested, the results indicated that eco-region (MARD), Biome_WWF affected parameters of the AGB = a*DBH^b*WD model. Compared to the model without random effect, models with random effect reduced SSE and AIC. When WD contributed to input variables, the eco-zone factor still affected AGB, because different conditions caused H variable changes within the DBH class.

Two models with random effect on MARD and WWF classes had very close values of all these indicators, with the model with random effect on MARD classes with the lowest AIC and less parameters chosen Table 15. The test results showed there were no random environment factors affects the models AGB = a*DBH^b*WD.

Table 15: The selected model AGB = a*DBH^b*WD with random effect on eco-region

Id	Random effect class	N trees	Equation
1	All trees	860	AGB = 0.23342*DBH^2.40963*WD
2	Central Highlands	114	AGB = 0.23342*DBH^2.46615*WD
3	North Central Coastal	331	AGB = 0.23342*DBH^2.39720*WD
4	Northeast	215	AGB = 0.23342*DBH^2.39623*WD
5	South Central Coastal	110	AGB = 0.23342*DBH^2.40257*WD
6	Southeast	110	AGB = 0.23342*DBH^2.38600*WD

The best model developed in this section had a significantly lower AIC value than the best model with DBH and H as input variable. This result confirms the importance of WD information to improve forest biomass estimates. Tree wood density is more important than tree height.

3.2.4 Model AGB = f(DBH, H, WD): Comparison of model forms and selection of the best equations

The tree biomass content may be different, even for trees with the same volume or the same DBH and H. This is due to the tree wood density, depending greatly on tree species. While it is difficult to develop models for each species in tropical forests, the variable WD is considered a representative factor, reflecting dry biomass stored in different species. Relationships between AGB with a combination of the three variables DBH, H and WD were examined through alternative models to obtain the optimum one as shown in

Table 16. Compared to the two forms, the optimum result for AGB selection was generated through the models with one variable combining DBH, H and WD:

AGB = a*DBH2HWD^b

With DBH2HWD = (DBH/100)^2*H*WD*1000, the surrogate of tree AGB in kg, DBH in cm, H in m and WD in $g.cm^{-3}$.

Table 16: Comparison of different AGB = f(DBH, H, WD) models with and without random effect

Id	Model form	Random effect	Residual function	N trees	AIC	SSE	Adj. R^2
1(*)	**AGB = a*DBH2HWD^b**	**No**	**1/DBH2HWD^k**	**860**	**9 200**	**65 965 670**	**0.933**
2	AGB = a*DBH^b*H^c*WD^d	No	1/DBH^k	860	9 304	67 510 566	0.932

(*) Selected Model

The selection form AGB = a*DBH2HWD^b was tested with random effect on eco-zones and environment parameters and the results indicated no factors to affect the AGB model. This showed that three variables together were able to account for AGB, reflecting tree size and biological characteristics of species in different site conditions. Moreover, this model had the lowest AIC of all the models tested with different combinations of input variables and random effects, meaning that it is the overall best model developed. As no random effect improved it is entirely based on the 860 trees of the EBLF dataset for evergreen forest, making it the most robust model as well.

The best model developed to predict tree AGB in evergreen broadleaved forest is:

AGB = 0.66609 *DBH2HWD^0.94304 **Eq. 25**

With:

AGB = tree aboveground biomass (kg),

DBH2HWD = (DBH/100)^2*H*WD*1000, surrogate of biomass (kg),

DBH = diameter at breast height (cm),

H = tree total height (m),

WD = wood density (g.cm^{-3}).

3.3 Biomass of branches (Bbr)

The method for selecting models for AGB was used to compare Bbr model forms and test random effects on the best ones (Table 17). No random effect of ecological and environmental factors improved the general models, thus they are not presented in this study.

Table 17: Comparison of alternative Bbr models with different input variables

Id	Input variables	Model form	Random effect	Residual function	AIC	SSE	Adj. R^2
1(*)	**DBH**	**Bbr = a*DBH^b**	**No**	**1/DBH^k**	**8 141**	**16 909 819**	**0.642**
2	DBH	Bbr = a + b*DBH + c*DBH^2	No	1/DBH^k	8 167	18 391 645	0.610
3	DBH	Bbr = a + b*DBH + c*DBH^2 + d*DBH^3	No	1/DBH^k	8 160	17 065 157	0.638
4	DBH and H	Bbr = a*DBH2H	No	1/DBH2H^k	8 164	17 854 035	0.622
5	DBH and H	Bbr = a*DBH2H^b	No	1/DBH2H^k	8 170	17 866 651	0.622
6	DBH and H	Bbr = a*DBH^b*H^c	No	1/DBH^k	8 147	17 011 151	0.639
7(*)	**DBH and WD**	**Bbr = a*DBH^b*WD**	**No**	**1/DBH^k**	**7 944**	**13 953 016**	**0.704**
8	DBH and WD	Bbr = a*DBH^b*WD^c	No	1/DBH^k	7 940	14 100 498	0.701
9	DBH, H and WD	Bbr = a*DBH2HWD^b	No	1/DBH2HWD ^k	8 002	14 969 735	0.683
10	DBH, H and WD	Bbr = a*DBH^b*H^c* WD^d	No	1/DBH^k	7 950	14 102 862	0.701

(*) Selected Model

The main result of the model comparison is that tree total height (H) as no influence on branches biomass. All the models with H as input variables gave worst results than the same models without it. On the contrary adding WD to DBH as input variable greatly improved the model (much lower AIC values). Adding a power parameter to WD improved slightly the model but not enough to be considered very important. Therefore the model Bbr = a*DBH^b*WD was selected as the best model. As WD is not easily measured in forest inventories, the model with DBH only as input variable was also selected (Table 17). On comparing to the selected AGB models, the relationship between Bbr and input variables had a lower coefficient R-squared (0.622-0.704). This means the biomass of branches in EBLF had a large variation.

The best models developed to predict tree branches biomass in evergreen broadleaved forest are: $$Bbr = 0.01339*DBH^{2.56010}$$ **Eq. 26** $$Bbr = 0.02426*DBH^{2.54939}*WD$$ **Eq. 27** With: Bbr = tree branches biomass (kg), DBH = diameter at breast height (cm), WD = wood density ($g.cm^{-3}$).

3.4 Biomass of leaves (Bl)

Employing the same method as above, the results of developing and selecting the biomass of leaves (Bl) models are presented in Table 18 and summarized as follows:

- The higher R-squared is below 0.60 meaning that biomass models explain just more than half of the leaves biomass variability.
- Power models and simpler forms gave the better results (associations of input variables instead of separated parameters for each input variable).
- The WWF and MARD eco-region had the highest effect on models, improving them significantly. AIC were from 100 to 200 lower when random effects on eco-regions were used.
- Other environmental factors such as soil type and rain improved the models but as much as eco-regions and were therefore not presented.
- The model with DBH, H and WD as input variables and MARD eco-regions as random effect is the overall best model for estimating biomass of tree leaves.
- For the model **Bl = a*DBH^b**, MARD classes were preferred over WWF, as the number of classes was smaller and the SSE was lower than WWF classes.

Table 18: Comparison of different leaves biomass models without random effect and of the two best leaves biomass models per group of input variables with random effect

Id	Model form	Random effect	Residual function	AIC	SSE	Adj. R^2
1	Bl = a*DBH^b	No	1/DBH^k	5 687	162 966	0.463
2	Bl = a + b*DBH + c*DBH^2	No	1/DBH^k	5 698	164 367	0.458
3	Bl = a + b*DBH + c*DBH^2 + d*DBH^3	No	1/DBH^k	5 700	160 207	0.471
4	Bl = a*DBH2H	No	1/DBH2H^k	5 938	254 798	0.162
5	Bl = a*DBH2H^b	No	1/DBH2H^k	5 645	163 302	0.464
6	Bl = a*DBH^b*H^c	No	1/DBH^k	5 691	162 598	0.464
7	Bl = a*DBH^b*WD	No	1/DBH^k	5 601	146 481	0.518
8	Bl = a*DBH^b*WD^c	No	1/DBH^k	5 608	146 508.2	0.517

Id	Model form	Random effect	Residual function	AIC	SSE	Adj. R^2
9	Bl = a*DBH2HWD^b	No	1/DBH2HWD^k	5 477	148 225	0.512
10	Bl = a*DBH^b*H^c* WD^d	No	1/DBH^k	5 616	145 753	0.519
11(*)	**Bl = a*DBH^b**	**MARD**	**1/DBH^k**	**5 558**	**135 806**	**0.553**
12	Bl = a*DBH^b	WWF	1/DBH^k	5 541	140 279	0.538
13(*)	**Bl = a*DBH2H^b**	**MARD**	**1/DBH2H^k**	**5 531**	**134 109**	**0.558**
14	Bl = a*DBH2H^b	WWF	1/DBH2H^k	5 543	137 629	0.547
15(*)	**Bl = a*DBH^b*WD**	**MARD**	**1/DBH^k**	**5 499**	**122 653**	**0.596**
16	Bl = a*DBH^b*WD	WWF	1/DBH^k	5 509	128 735	0.576
17(*)	**Bl = a*DBH2HWD^b**	**MARD**	**1/DBH2HWD^k**	**5 389**	**125 099**	**0.588**
18	Bl = a*DBH2HWD^b	WWF	1/DBH2HWD^k	5 398	130 764	0.570

(*) Selected Model

The full list of models and the value of the parameters for each class are presented in the Table 19.

Table 19: The selected models to estimate the biomass of tree leaves.

		Input variables of the models			
Random effect class	**N trees**	**DBH**	**DBH and H**	**DBH and WD**	**DBH, H and WD**
All trees	859	Bl = 0.06391*DBH^1.5415	Bl = 8.726*DBH2H^0.6283	Bl = 0.1069*DBH^1.5393*WD	Bl = 0.1551*DBH2HWD^0.6288
Central Highlands	114	Bl = 0.06391*DBH^1.7132	Bl = 14.2358*DBH2H^0.6109	Bl = 0.1069*DBH^1.6682*WD	Bl = 0.2227*DBH2HWD^0.6288
North Central Coastal	311	Bl = 0.06391*DBH^1.5724	Bl = 9.3323*DBH2H^0.6264	Bl = 0.1069*DBH^1.5939*WD	Bl = 0.1778*DBH2HWD^0.6288
Northeast	215	Bl = 0.06391*DBH^1.5330	Bl = 8.2521*DBH2H^0.6296	Bl = 0.1069*DBH^1.5309*WD	Bl = 0.1579*DBH2HWD^0.6288
South Central Coastal	109	Bl = 0.06391*DBH^1.4994	Bl = 7.1200*DBH2H^0.6334	Bl = 0.1069*DBH^1.5002*WD	Bl = 0.1285*DBH2HWD^0.6288
Southeast	110	Bl = 0.06391*DBH^1.3894	Bl = 4.6900*DBH2H^0.6410	Bl = 0.1069*DBH^1.4036*WD	Bl = 0.0890*DBH2HWD^0.6288

3.5 Aboveground biomass of the main families (Evergreen and Deciduous)

As six tree families had more than 40 trees when regrouping evergreen and deciduous forest (Dipterocarpaceae, Euphorbiaceae, Fagaceae, Lauraceae, Leguminosae and Myrtaceae) they were selected to test the influence of genetics and genetic + biomes on AGB models.

3.5.1 Influence of key tree families in the dataset – AGB models for six families in both forest types

The models form selected earlier were used to compare models without random effect to models with random effect on Family. As the number of trees used to develop these models is different from the previous sections, the models were developed again with the six family's dataset. Therefore the AIC results in this section are not comparable with the AIC of previous sections. These results are presented in Table 20.

Table 20: Influence of tree family on AGB models.

Id	Model form	Random effect	Residual function	N trees	AIC	SSE	R^2adj.
	AGB = a*DBH^b	Without	1/DBH^k	501	**6 221**	118 279 930	0.802
	AGB = a*DBH^b	Family	1/DBH^k	501	No significant family effect		
	AGB = a*DBH2H^b	Without	1/DBH2H^k	501	6 039	99 238 867	0.834
	AGB = a*DBH2H^b	Family	1/DBH2H^k	501	**5 922**	62 977 036	0.895
	AGB = a*DBH^b*WD	Without	1/DBH^k	501	6 002	55 693 032	0.907
	AGB = a*DBH^b*WD	Family	1/DBH^k	501	**5 947**	59 600 018	0.900
	AGB = a*DBH2HWD^b	**Without**	**1/DBH2HWD^k**	**501**	**5 590**	**40 005 105**	**0.933**
	AGB = a*DBH2HWD^b	Family	1/DBH2HWD^k	501	No significant family effect		

As a result, only two model forms were improved by the effect of tree family: the model with DBH+H and DBH+WD as input variables. The overall best model form, with DBH, H and WD as input variables was not improved by the effect of Family. The overall best model is still a model using a combination of DBH, H and WD without random effect.

3.5.2 Model per family and random effect on eco-region and forest type

This study also developed equations to estimate AGB for each of the six main families. Model forms of the best model selected with the whole dataset were used and the effect of MARD eco-region was tested as it was the dominant effect over the study. The effect of forest type was tested on the Dipterocarpaceae dataset as it was the only one with an equilibrate number of tree per forest type.

Table 21: Comparison of the AGB models for Dipterocarpaceae

Id	Model form	Random effect	Residual function	N trees	AIC	SSE	R^2 adj.
1	AGB = a*DBH^b	Without	1/DBH^k	141	1 739	19 541 519	0.913
2	AGB = a*DBH^b	MARD	1/DBH^k	141	1 640	14 188 540	0.937
3	AGB = a*DBH^b	Forest type	1/DBH^k	141	No significant random effect		
4	AGB = a*DBH2H^b	Without	1/DBH2H^k	141	1 656	16 431 686	0.927
5	AGB = a*DBH2H^b	MARD	1/DBH2H^k	141	1 628	12 741 520	0.943
6	AGB = a*DBH2H^b	Forest type	1/DBH2H^k	141	No significant random effect		
7	AGB = a*DBH^b*WD	Without	1/DBH^k	141	1 754	15 860 958	0.929
8	**AGB = a*DBH^b*WD**	**MARD**	**1/DBH^k**	**141**	**1 605**	**11 300 130**	**0.949**
9	AGB = a*DBH^b*WD	Forest type	1/DBH^k	141	1 673	15 408 547	0.931
10	AGB = a*DBH2HWD^b	Without	1/DBH2HWD^k	141	**1 629**	11 480 922	0.949
11	AGB = a*DBH2HWD^b	MARD	1/DBH2HWD^k	141	No significant random effect		
12	AGB = a*DBH2HWD^b	Forest type	1/DBH2HWD^k	141	No significant random effect		

For the Dipterocarpaceae, forest type only improved the biomass model AGB =f(DBH,WD) and was still worse than the model with random effect of the eco-region. Forest type had no generally influence on the models and models per eco-region are preferred (Table 21). However, the overall best model is now the model with DBH and WD as input variables and with the effect of eco-region on parameters a and b. The dataset is not big enough to make any final conclusion but this result is different from the models developed with the country level dataset.

For all other tree Families, the effect of forest type was not tested. Regarding the effect of eco-regions, three families out of five had a better model with the influence of eco-region. The number of tree per region is not homogenous for each family and the number of tree per family and region is quite low (Table 7). Still this could indicate that for several families, differences in terms of allometry between eco-regions are not taken into consideration by the association of DBH, H and WD in the models. The form factors could be one of the factors explaining these differences.

Table 22: Influence of tree Family on AGB models.

Id	Model form	Random effect	Residual function	N trees	AIC	SSE	R^2adj
	Euphorbiaceae						
1	AGB = a*DBH^b	Without	1/DBH^k	52	604	3 475 911	0.848
2	AGB = a*DBH^b	MARD	1/DBH^k	52	594	918 431	0.960
3	AGB = a*DBH2H^b	Without	1/DBH2H^k	52	581	2 378 909	0.896
4	AGB = a*DBH2H^b	MARD	1/DBH2H^k	52	554	186 144	0.992
5	AGB = a*DBH^b*WD	Without	1/DBH^k	52	571	1 493 156	0.935
6	AGB = a*DBH^b*WD	MARD	1/DBH^k	52	556	682 304	0.970
7	**AGB = a*DBH2HWD^b**	**Without**	**1/DBH2HWD^k**	**52**	**533**	**983 741**	**0.957**
8	AGB = a*DBH2HWD^b	MARD	1/DBH2HWD^k	52	No significant effect		
	Fagaceae						
1	AGB = a*DBH^b	Without	1/DBH^k	95	1 260	32 970 346	0.792
2	AGB = a*DBH^b	MARD	1/DBH^k	95	No significant random effect		
3	AGB = a*DBH2H^b	Without	1/DBH2H^k	95	1 226	20 285 364	0.872
4	AGB = a*DBH2H^b	MARD	1/DBH2H^k	95	1 187	8 563 283	0.946
5	AGB = a*DBH^b*WD	Without	1/DBH^k	95	1 216	26 541 504	0.832
6	AGB = a*DBH^b*WD	MARD	1/DBH^k	95	No significant random effect		
7	**AGB = a*DBH2HWD^b**	**Without**	**1/DBH2HWD^k**	**95**	**1 167**	**15 819 182**	**0.900**
8	AGB = a*DBH2HWD^b	MARD	1/DBH2HWD^k	95	No significant random effect		
	Lauraceae						
1	AGB = a*DBH^b	Without	1/DBH^k	66	720	5 963 874	0.774
2	AGB = a*DBH^b	MARD	1/DBH^k	66	No significant random effect		
3	AGB = a*DBH2H^b	Without	1/DBH2H^k	66	712	5 399 837	0.796
4	AGB = a*DBH2H^b	MARD	1/DBH2H^k	66	693	2 439 800	0.908
5	AGB = a*DBH^b*WD	Without	1/DBH^k	66	665	1 304 882	0.950
6	AGB = a*DBH^b*WD	MARD	1/DBH^k	66	No significant random effect		
7	AGB = a*DBH2HWD^b	Without	1/DBH2HWD^k	66	660	1 335 002	0.949
8	**AGB = a*DBH2HWD^b**	**MARD**	**1/DBH2HWD^k**	**66**	**649**	**749 681**	**0.972**
	Leguminosae						
1	AGB = a*DBH^b	Without	1/DBH^k	84	950	8 224 268	0.864
2	AGB = a*DBH^b	MARD	1/DBH^k	84	No significant random effect		
3	AGB = a*DBH2H^b	Without	1/DBH2H^k	84	929	5 405 166	0.911
4	AGB = a*DBH2H^b	MARD	1/DBH2H^k	84	928	5 267 347	0.913
5	AGB = a*DBH^b*WD	Without	1/DBH^k	84	921	4 881 444	0.919
6	AGB = a*DBH^b*WD	MARD	1/DBH^k	84	No significant random effect		
7	AGB = a*DBH2HWD^b	Without	1/DBH2HWD^k	84	879	3 418 814	0.944
8	**AGB = a*DBH2HWD^b**	**MARD**	**1/DBH2HWD^k**	**84**	**873**	**3 022 460**	**0.950**
	Myrtaceae						
1	AGB = a*DBH^b	Without	1/DBH^k	63	781	13 780 996	0.793
2	AGB = a*DBH^b	MARD	1/DBH^k	63	No significant random effect		
3	AGB = a*DBH2H^b	Without	1/DBH2H^k	63	740	9 891 846	0.851
4	AGB = a*DBH2H^b	MARD	1/DBH2H^k	63	721	3 945 596	0.941
5	AGB = a*DBH^b*WD	Without	1/DBH^k	63	744	7 597 079	0.886
6	AGB = a*DBH^b*WD	MARD	1/DBH^k	63	No significant random effect		
7	AGB = a*DBH2HWD^b	Without	1/DBH2HWD^k	63	691	3 966 744	0.940
8	**AGB = a*DBH2HWD^b**	**MARD**	**1/DBH2HWD^k**	**63**	**677**	**1 994 329**	**0.970**

AGB models (AGB = f(DBH,H,WD)) per family were compared to the general model developed in section 3.2.4, using S% bias, EF and MAE%. The results are shown in Table 23. Since the data used to compare the models were the dataset for each family, family models outperformed the national scale model, but the interesting result is that the national scale model had high bias only for two families: Euphorbiaceae and Fagaceae and with opposite signs. For the four other families the national scale model gave good results. This study is therefore not able to promote family scale models over multispecies models, but for specific families it seems that a general model could results in a high level of error. In the whole dataset the data from one family compensate the other but it would be interesting to see if these two families have almost the same importance in the country or not.

Table 23: Comparison of selected model to each family model

Tree dataset	N trees	Equations	S%	EF	MAE%
Dipterocarpaceae	141	AGB(Dipterocarpaceae) = 0.63320*DBH2HWD^0.94744	0.02	0.95	17.4
Dipterocarpaceae	141	AGB (national)= 0.66609*DBH2HWD^0.94304	1.40	0.95	17.9
Euphorbiaceae	52	AGB(Euphorbiceae) = 0.87695*DBH2HWD^0.88981	-0.26	0.96	12.7
Euphorbiaceae	**52**	**AGB (national)= 0.66609*DBH2HWD^0.94304**	**15.33**	**0.92**	**16.0**
Fagaceae	95	AGB(Fagaceae) = 0.65048*DBH2HWD^0.96474	1.80	0.90	23.0
Fagaceae	**95**	**AGB (national)= 0.66609*DBH2HWD^0.94304**	**-13.06**	**0.90**	**21.5**
Lauraceae	66	AGB(Lauraceae) = 0.73349* DBH2HWD^0.93392	1.20	0.95	15.9
Lauraceae	66	AGB (national)= 0.66609*DBH2HWD^0.94304	-1.43	0.95	15.1
Leguminoseea	84	AGB(Leguminosea) = 0.61749* DBH2HWD^0.95634	-0.01	0.95	16.0
Leguminoseae	84	AGB (national)= 0.66609*DBH2HWD^0.94304	-3.03	0.94	16.1
Myrtaceae	63	AGB(Myrtaceae) = 0.7223652* DBH2HWD^0.9332611	-7.18	0.93	13.4
Myrtaceae	63	AGB (national)= 0.66609*DBH2HWD^0.94304	-7.32	0.93	12.8

3.6 Biomass expansion factors

AGB also can be estimated through biomass expansion factors (BEF), with the conversion formula: AGB = BEF*Bst. Based on the whole dataset, descriptions of BEF were calculated for each eco-region and forest type and shown in Table 24. The overall BEF for EBLF is 1.25 with little variation between eco-regions.

Table 24: BEF description for two forest types and per eco-region

	Deciduous forest				EBLF			
Eco-region	**Average**	**Min**	**Max**	**StdDev**	**Average**	**Min**	**Max**	**StdDev**
Central Highlands	1.27	1.04	1.81	0.16	1.30	1.05	1.99	0.15
North Central Coastal					1.25	1.02	2.03	0.15
Northeast					1.24	1.02	1.83	0.15
South Central Coastal					1.22	1.00	1.70	0.15
Southeast	1.40	1.16	1.86	0.15	1.26	1.04	1.78	0.15
Total	**1.34**	**1.04**	**1.86**	**0.17**	**1.25**	**1.00**	**2.03**	**0.15**

4 VALIDATION OF THE SELECTED ALLOMETRIC EQUATIONS AND COMPARISON TO EXISTING MODELS

4.1 Validation of selected models and comparison to local models

The dataset of 1 303 independent sampled trees in EBLFs of two eco-regions (520 trees in North Central Coast (NCC) and 783 trees in Northeast (NE)) was used to validate selected models and compare to local models in the two regions. To do this, the AGB for each independent tree was predicted following selected and local models, developed in Phase I of UN-REDD Programme. They are plotted in **Error! Reference source not found.** with indicators for validation as the S% bias, efficiency factor (EF) and MAE% calculated for each model. The results are presented in Table 25.

As a result, all models validated had a good fit with a EF of 0.81-0.96 and a bias below 25 % for the models with DBH as input variable and below 3% form the best models. The models with DBH only as input variable had the lowest accuracy, with S% bias from -11.4 percent to -25.7 percent, and MAE% 30.9 to 33.6 percent. The model with the highest accuracy was the equation with the group of variables DBH2HWD, S% bias < 3% and MAE% < 18%, EF was over 0.95. The model with the group of variables "DBH^b*WD" also had good accuracy with validation indicators close to the model with the group of three variables "DBH2HWD".

Two local AGB models in the NCC and NE regions and selected national scale models had similar performances, the national scale model being slightly better. This indicates that national scale model with a combination of three input variables "DBH2HWD" could be used to estimate AGB in different eco-zones of Viet Nam. The locally developed models from UN-REDD phase I have a lower error at tree level but a higher bias, compared to the model developed at national scale. The national scale model is therefore more representative of the forest conditions in Viet Nam.

Table 25: Validation of selected models and comparison to local models

Equations	Scale/region of equation	Eco-region	N trees	S% Bias	EF	MAE %
AGB = 0.12115*DBH^2.41539	This study, equation for NCC	NCC	520	-11.4	0.87	33.6
AGB = 254.49543*DBH2H^0.94705	This study, equation for NCC	NCC	520	-8.7	0.92	30.1
AGB = 0.23342*DBH^2.39720*WD	This study, equation for NCC	NCC	520	-2.0	0.95	19.1
AGB = 0.66609*DBH2HWD^0.94304	**This study, national equation**	**NCC**	**520**	**2.2**	**0.97**	**17.6**
AGB = 0.0704*(DBH^2*H*WD)^ 0.93890	UN-REDD Phase I model for NCC	NCC	520	-9.5	0.95	14.6
AGB = 0.12483*DBH^2.41539	This study, equation for NE	NE	783	-25.7	0.82	30.9
AGB = 255.33956*DBH2H^0.94705	This study, equation for NE	NE	783	-20.7	0.85	26.8
AGB = 0.23342*DBH^2.39623*WD	This study, equation for NE	NE	783	-6.5	0.93	19.2
AGB = 0.66609*DBH2HWD^0.94304	**This study, national equation**	**NE**	**783**	**2.6**	**0.95**	**18.0**
AGB = 0.1173*(DBH^2*H^0.7*WD)^0.9898	UN-REDD Phase I model for NE	NE	783	-3.5	0.95	16.9

4.2 Comparison of the selected models to pan-tropical models

Data from 1 303 independent trees in EBLFs were employed to validate and compare the selected models to pan-tropical models. AGB for each independent tree was estimated following selected and pan-tropical models, plotted in Figure 2 and then S% bias, EF and MAE% were calculated as indicators to compare the models (Table 26).

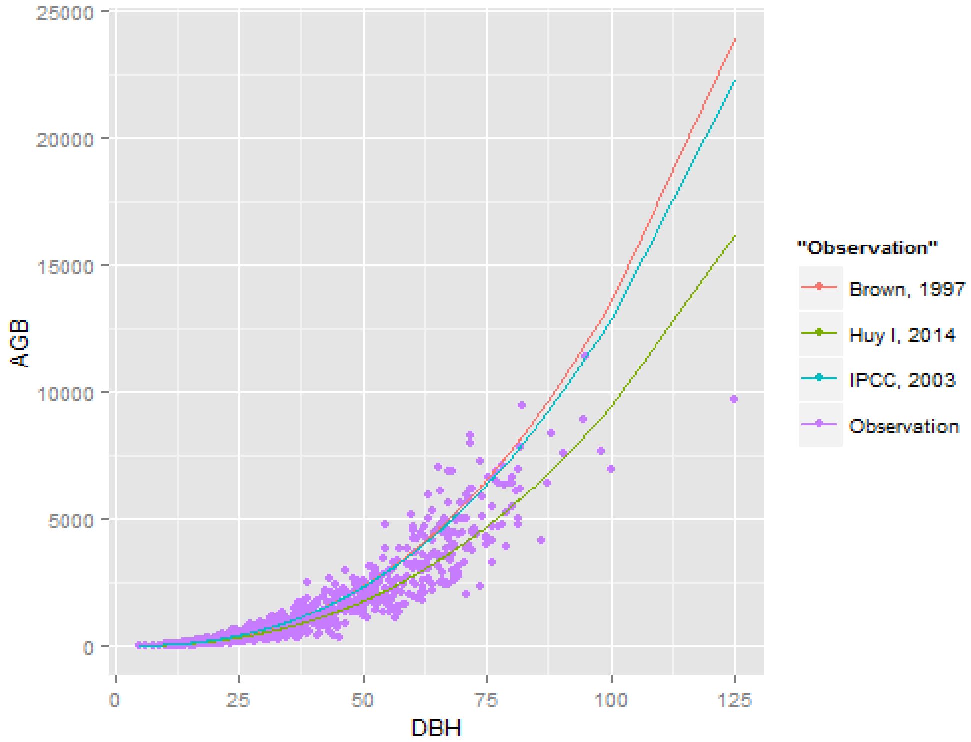
The comparisons were made as follows:

- Selected model at national scale AGB = a*DBH^b was compared to that of Brown (1997) and IPCC (2003) models with the same DBH variable. As a result, through validation indicators MAE% was reduced by 14 percent through use of this study's model, and the bias was reduced from 20 to 7 %.
- Selected model at national scale AGB = a*DBH^b*WD was compared to Chave I (2005) model with the same variables DBH and WD. The results indicated that by using the selected model, MAE% was reduced significantly by 30 percent.
- The best selected model at national scale AGB = a*DBH2HWD^b was compared to that of Chave II (2005) and Chave III (2014) models with the same variables DBH, H, WD. The results saw S% bias and MAE% reduced significantly by 22-26 percent and 10-18 percent, respectively by use of the best models of this study. The EF of the selected model had the highest value above 0.96.

These comparison results support the use of national allometric equations over generic global and regional tropical forest equations.

Table 26: Comparison of the selected models to pan-tropical models

Model	Author	Equation	N trees	S% Bias	EF	MAE%
AGB = f(DBH)	Brown (1997)	AGB = exp(-2.134 + 2.530*log(DBH))	1 303	24.6	0.73	49.7
	IPCC (2003)	AGB=exp(-2.289 + 2.649*log(DBH) - 0.021*(log(DBH))^2)	1 303	22.3	0.77	49.9
	This study	AGB = 0.1394363*DBH^2.4153948	1 303	-7.1	0.87	35.6
AGB = f(DBH, WD)	Chave I (2005)	AGB = WD*exp(-1.499 +2.148*log(DBH) + 0.207*(log(DBH))^2 - 0.0281*(log(DBH))^3)	1 303	50.9	0.54	50.5
	This study	AGB = 0.2334195*DBH^2.4096317*WD	1 303	0.9	0.94	20.5
AGB = f(DBH, H, WD)	Chave II (2005)	AGB = 0.0509*WD*DBH^2*H	1 303	28.3	0.82	27.5
	Chave III (2014)	AGB = 0.0673*(WD*DBH^2*H)^0.976	1 303	30.2	0.82	35.3
	This study	**AGB = 0.6660939 *DBH2HWD^0.9430468**	**1 303**	**2.4**	**0.96**	**17.8**



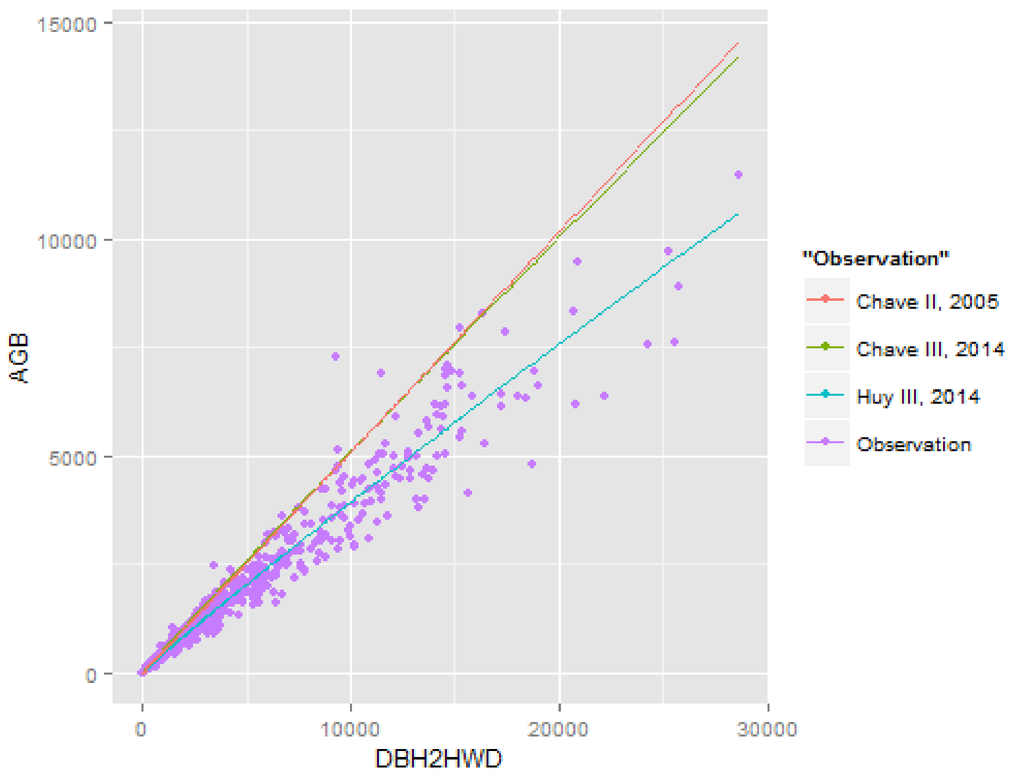


Figure 2: Graphs of the independent data and predictions by this study and different pan-tropical models.

5 CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

From the results of the dataset analysis of evergreen broadleaved forest and EBLF plus Deciduous for the main Families, national scale biomass models were developed. The main conclusions are:

- The power equation is appropriate for biomass models with one or groups of input variables with DBH, H and WD (Biomass = a*DBH^b, Biomass = a*DBH2H^b, Biomass = a*DBH^b*WD and Biomass = a*DBH2HWD^b).
- The increase in independent variables from one to three reduces the AIC and SSE of the estimates. The best option is biomass models with three variables DBH, H and WD using the equation Biomass = a*DBH2HWD^b. WD is more important than tree height for biomass models.
- Biomass models with random effect of eco-regions or environmental parameters improved the reliability of estimating biomass at a national scale. In particular the eco-region classification of MARD and WD classes significantly affected many models and improved the reliability and accuracy to estimate biomass at a national scale.
- The biomass of leaves and branches saw large deviations due to many effect factors. Tree height was not significant to estimated branches biomass. For leaves, many factors influenced the biomass of tree leaves but all together could not explain more than 60 % of its variability.
- Introducing the effect of Family in the models reduced the AIC and SSE for models with two input variables (DBH + H and DBH + WD) but not for the overall best model (AGB = f(DBH,H,WD)). For two families (Euphorbiaceae and Fagaceae), eco-region had no effect on the model and the model was very different from the general model (S% of the general model around 15%). For the three others families (not counting Dipterocarpaceae), even with three input variables (DBH, H and WD), the models were still improved by adding an effect of eco-regions. At a local scale, environmental conditions may change tree architecture to an extent not captured by the association of tree DBH, H and WD.
- All selected models were validated from independent datasets. The results showed that these models had good/perfect fits with an EF of more than 0.81-0.96. The models of AGB = a*DBH^b*WD and AGB = a*DBH2HWD^b with the highest accuracy with S% bias < 3% and MAE% < 18% and an EF of more than 0.95.
- AGB models at a national scale showed greater reliability than existing pan-tropical models through validations by S% bias, MAE% and EF indicators. By use of the best model of the current study, MAE% was reduced by 14 percent compared to models from IPCC (2003) and Brown (1997) and 10-18 percent from Chave (2005, 2014).
- The best models at a national scale with random effect on an eco-region had an equivalent reliability compared to local region models, the overall best model of this study being slightly better in terms of bias, and slightly worse in terms of MAE% or single tree prediction.

5.2 Recommendations for using the developed biomass models

- Biomass models with only one variable DBH should be used in rapid inventories with a participatory approach involving local communities. These models are proposed with specific parameters for each eco-region. For Northwest eco-region, models developed for North East are still valid.
- Biomass models with two variables DBH and H are appropriate for the existing national inventory.
- Biomass models with three variables DBH, H and WD should be applied for high reliability and where WD values of main species are available.

- Scientific studies and research experiments could help better understand how genetics affect the biomass allometry and find an adequate level of species identification to improve forest biomass estimates, given that the diversity is too high for species specific model development.
- National scale models with random effect on eco-regions should be used rather than pan-tropical models, due to improved accuracy and reliability.

REFERENCES

Basuki TM, Van Lake PE, Skidmore AK, Hussin YA, (2009). Allometric equations for estimating the above-ground biomass in the tropical lowland Dipterocarp forests. For. Ecol. and Manag. 257(2009): 1684-1694.

Brown S, (1997). Estimating biomass and biomass change of tropical forests: A Primer. FAO Forestry paper – 134. ISBN 92-5-103955-0. Available on web site: http://www.fao.org/docrep/W4095E/w4095e00.htm#Contents.

Chave J, Andalo C, Brown S, Cairns MA, Chambers JQ, Eamus D, Folster H, Fromard F, Higuchi N, Kira T, Lescure JP, Nelson BW, Ogawa H, Puig H, Rier B, Yamakura T, (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145 (2005): 87-99. DOI 10.1007/s00442-005-0100-x.

Chave J, et al. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. Global change biology (2014). Doi: 10.1111/gcb.12629

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, Ha Noi, Viet Nam.

Hung, N.D., Giang, L.T., Tu, D.N., Hung, P.T., Lam, P.T., Khanh, N.T., Thuy, H.M. (2012) Tree allometric equations in Evergreen broadleaf and Bamboo forests in the Northeast 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, Ha Noi, Viet Nam.

Huy, B. (2013). Developing biomass and carbon allometric equation and using remote sensing and GIS to estimate CO2 absorbed of ever-green broard leaf forests in the Central Higlands of Viet Nam. Sciences and Teniques Publisher, HCMC, pp 336.

Huy, B. (2014). Aboveground biomass allometric equations for evergreen broadleaf forests in South Central Coastal Eco-Region, Viet Nam. UN-REDD Viet Nam, FAO. (To be submitted to AFS Journal).

IPCC (2003, 2006). IPCC Guidelines for National Greenhouse Gas Inventories. Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T., Tanabe K., (eds). Published: IGES, Japan.

Mayer DG, Butler DG (1993). Statistical validation. Ecological Modelling, 68(1993): 21-32.

Picard N, Saint Andre L, Henry M, (2012). Manual for building tree allometric equations: from the field to the prediction, Food and Agriculture Organization of the United Nations, Centre de Coopération Internationale en Recherche Agronomique.

UNFCCC (1992). United Nations Framework Convention on Climate Change. United Nations.

UN-REDD (2011). Measurement, Reporting & Verification (MRV) Framework Document. UN-REDD Viet Nam Programme.