

Assessment of aboveground biomass using data from Unmanned Aerial Vehicles (UAVs) and Terrestrial Laser Scanning (TLS) in the Ban Nong Hai community forest project area in Chiang Mai Province.

Kairop Pongphiboonkiat¹*, Kampanat Deeudomchan¹, Kativich Kantha¹

¹ Geo-Informatics and Space Technology Development Agency (Public Organization), Geo-informatic Applications and Service Office,

120 The Government Complex Commemorating His Majesty The King's 80 th Birthday

Anniversary, 5th December, B.E.2550(2007) Ratthaprasasanabhakti Building 6th and 7th Floor,

Chaeng Wattana Road, Lak Si, 10210 Bangkok, Thailand.

*kairop@gistda.or.th

*Abstract***:** *Due to the problem of climate change affecting all life on earth. Thailand has realized and prepared to reduce this impact. Therefore, the goal is to increase forest areas and green areas. The objective of the research is to develop an aboveground biomass model in the Ban Nong Hai Community Forest Project area in Chiang Mai Province using the machine learning model method from the application of unmanned aerial vehicles (UAVs) data that is equipped with various sensors, such as multispectral and processed together with data obtained from LiDAR cameras. Including comparison the model developed using data for training between forest inventory data and Terrestrial Laser Scanning (TLS) data from 6 sample plots, size 40 x 40 meters, then utilized to develop a machine learning model. The model's accuracy is compared under two scenarios: one using manual labor for model development and the other employing TLS data. The model's accuracy is quantified using the coefficient of determination (R²), Root Mean Square Error (RMSE), and Relative Root Mean Square Error (rRMSE). For the manual labor approach, the R², RMSE, and rRMSE values are 0.56, 114.7 kg/100 m², and 8.8%, respectively. In contrast, utilizing TLS data yields values of 0.61, 79.8 kg/100 m², and 4.5%. Both models have a very high level of accuracy, given that the rRMSE is below 10%. The aboveground biomass model derived from the UAVs-TLS model estimates biomass accumulation at 2,825.4 tons, equivalent to 4,868.9 tCO2e in carbon dioxide absorption. The UAVs-Forest inventory model estimate is 2,805.4 tons, corresponding to 4,834.5 tCO2e.*

Keywords: Aboveground Biomass, Machine learning Model, Terrestrial laser scanner (TLS), Unmanned Aerial Vehicles (UAVs)

INTRODUCTION

The problem of forest resource conservation to prevent deforestation, forest degradation and wildfires are challenges for every country. Due to the increase in population, economic growth, including the development of basic public utilities, it causes a direct impact on the forest area, forest degradation and a permanent change of the forest area to other areas. This will also lead to climate change in the future.

Therefore, this study focuses on the application of current technology to examine the appropriateness of the method to be used to monitor the abundance of Ban Nong Hai

community forest, Chiang Mai province. By using remote sensing technology, such as data from Unmanned Aerial Vehicles (UAVs) that are equipped with sensors, including multispectral sensors and LiDAR cameras, as well as data obtained from the survey with the Terrestrial laser scanner (TLS) to use the data to develop a model to estimate carbon accumulated in the current area and in the future, in the next 5-10 years. When checking back again, it will be possible to calculate the increase in carbon accumulation that can be used to buy and sell as income for the community and also to preserve the forest area forever.

STUDY AREA

Ban Nong Hai Community Forest Area is located in Khuang Pao Subdistrict, Chom Thong District, Chiang Mai Province, which is in the north of Thailand. The coordinates of latitude 18 deg 28 min 2.0244 sec to 18 deg 28 min 31.2384 sec and longitude 98 deg 43 min 10.675 sec to 98 deg 43 min 58.462 sec have an area of about 0.74 square kilometers. Mean sea level is about 300–400 meters. The area is flat and mountainous. The soil condition is loamy mixed with gravel. Most of the land use in the study area is forest.

Figure 1: Boundary of Ban Nong Hai Community Forest, Chiang Mai Province.

METHOD

1. Equipment Preparation.

- 1.1 Equipment for Unmanned Aerial Vehicles.
	- DJI Phantom 4 Model equipped with multispectral sensors.
	- DJI Matrice 300 RTK Model equipped with DJI Zenmuse L1 (LiDAR).
	- PIX4D mapper and PIX4D field are the software used to process data.

- 1.2 Equipment used to survey in the sample plots.
	- Terrestrial laser scanner (TLS): Faro S70 Model
	- Equipment for sample plots such as ropes, distance tapes, water level meters, compasses, etc.
	- GNSS: Greenvalley LiBase Model
- 1.3 Processing Software
	- Scene
	- LiDAR360
	- ArcGIS
	- Maximum Entropy (MaxEnt)

2. Operating diagram.

In this study, the design and development concept for developing the aboveground biomass model obtained from forest inventory data and TLS data in the community forest area of Ban Nong Hai, Chiang Mai Province, was developed with the following operational diagram. As shown in Figure 2.

Figure 2: Operational diagram.

3. Landuse classification

Landuse classification is performed using visual interpretation from coordinatecorrected and mosaicked UAV images.

4. Number of sample plots

For calculating the appropriate number of sample plots according to the A/R Methodology Tool "Calculation of the number of sample plots for measurements within A/R CDM project activities" in cases where samples are less than 5 percent of the total area (UNFCCC, 2022). The equation that calculates the appropriate number of sample plots is as follows:

$$
n = \frac{N \times t_{VAL}^2 \times (\sum_i w_i \times s_i)^2}{N \times E^2 + t_{VAL}^2 \times (\sum_i W_i \times s_i^2)}
$$
 Equation 1

$$
n = \left(\frac{t_{VAL}}{E}\right)^2 \times \left(\sum_i w_i \times s_i\right)^2
$$
Equation 2

When $n =$ Suitable Number of Sample Plots

 N $=$ The number of probability of all sampling plots of the area.

 t_{VAL} = Critical values of probability distributions

 W_i = The proportion of the area in stratum i to the total area.

 s_i = Standard deviation of stratum i.

Е $=$ Level of confidence

= Estimation of biomass in stratum $1,2,3...$ i

The equation, the stratum classification, is based on GISTDA's 2022 biomass assessment data to calculate the statistical value and calculate the appropriate number of sample plots according to Equations 1-2.

5. Forest inventory

In forest inventory, 40x40 meters of random samples were divided into 16 subplots with a size of 10x10 meters to collect the diameter at breast height (DBH) and height of all trees in the sample plots, using manual labor to measure and calculate the resulting biomass by the equation for estimating biomass accumulation in deciduous mixed forests in Thailand. (Ogawa et al., 1965)

- $Ws = stem \, biomass (kg)$
- $Wb = branch \, biomass \, (kg)$
- $W1 = leaf \text{biomass} (kg)$
- Wtc = Total biomass (stem+ branch+ leaf) (kg)
- $=$ DBH (Diameter at Breast Height 1.3 m.) D.
- $=$ Tree height H

6. Terrestrial Laser Scanner (TLS)

The survey using a terrestrial laser scanner (TLS), which survey using 3D laser scanning to all the trees in the plot and sampling at the same location as the sample plots that forest inventory. The scanned data in the sample plots will be recorded at each tree's location, as well as the reference geographic coordinate system automatically. The data will be stored in the form of a point cloud, which can use point cloud data to measure DBH and height of trees from the LiDAR 360 program. Then used to measure DBH, and the height of trees in the sample plots was calculated using the equation for estimating biomass accumulation in deciduous mixed forests in Thailand.

7. Unmanned Aerial Vehicles (UAVs)

- 7.1 Multispectral sensors.
	- UAV is equipped with multispectral sensors that can record data in multiple wavelengths, including: blue band, green band, red band, red-edge band, near infrared band, and RGB. Then, the wavelength values are used to find various index values for model development.
	- Factors obtained from the multispectral sensors can calculate various index values. By using PIX4D field software, including: Blue band, Green band, Red band, Near infrared band, Red Edge band, Blue Normalized Difference Vegetation Index (BNDVI), Green Normalized Difference Vegetation Index (GNDVI), Visible Atmospherically Resistant Index (VARI), Triangular Greenness Index (TGI), Normalized Difference Red Edge (NDRE), Normalized Difference Vegetation Index (NDVI), Modified chlorophyll index (MCARI), Structure Intensive Pigment Index (SIPI), and Leaf Chlorophyll Index (LCI).
- 7.2 LiDAR cameras.
	- Factors obtained from LiDAR cameras can analyze the Canopy Height Model (CHM) by analyzing the Digital Terrain Model (DTM) and the Digital Surface

Model (DSM) together and then using the Canopy Height Model (CHM) to analyze the canopy cover percentage of the study area.

8. Model development with machine learning model.

- 1) Prepare various factors obtained from multispectral sensors and LiDAR cameras, such as the Normalized Difference Vegetation Index, Triangular Greenness Index, Blue band, and Modified Chlorophyll Index, etc., to be used in the development of an aboveground biomass model.
- 2) Prepare computed biomass data from forest inventory data and TLS data, which will be used as data for training and testing. By dividing the biomass data into two sets, data for the training model and data for testing the accuracy of the model in the ratio of 80/20. Then use the training dataset as a representative to train the machine learning model.
- 3) Then, the aboveground biomass values for training dataset and various factors obtained from multispectral sensors and LiDAR cameras were imported into the machine learning model.
- 4) The model development in this study chose Maximum Entropy or MaxEnt (Saatchi et al., 2008), because MaxEnt is popular for assessing biomass accumulation in forest areas and providing an uncertainty map for the model calculated from the machine learning model. The probability of aboveground biomass in each stratum is divided according to the amount of aboveground biomass accumulation in each subplot. Therefore, when calculating all stratum of aboveground biomass, it will be combined into a map of the biomass of the study area with equation 6. Including calculating the uncertainty caused by the model at every image point with Equations 7 and 8. (Saatchi et al., 2011)

$$
\widehat{AGB} = \frac{\sum_{l=1}^{N} P_l^n A G B_l}{\sum_{l=1}^{N} P_l^n}
$$
 Equation 6

$$
\varepsilon_{prediction} = \frac{\sigma_{\widehat{AGB}}}{\widehat{AGB} \times 100}
$$

$$
Equation\ 7
$$

$$
\sigma_{\widehat{AGB}} = \sqrt{\frac{\sum_{i=1}^{N} (AGB_i - \widehat{AGB})^2 P_i}{\sum_{i=1}^{N} P_i}}
$$
 Equation 8

When \widehat{AGB} = AGB values obtained from the model at every pixel of image.

- P_i $=$ Probability value of AGB class range calculated from the machine learning model.
- AGB_i = Average of AGB in each class.

9. Model accuracy

Validate the model from the aboveground biomass values for testing dataset, approximately 20% of the total number of subplots. By plotting the biomass values obtained from the field survey plots (x) and the biomass value calculated from the model (y). The accuracy of the model with statistics is R2 root mean square error (RMSE) as in Equation 9 and relative root mean square error (rRMSE) as in Equation 10.

$$
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{H_d^{i,m} - H_d^{i,c}}{H_d^{i,m}}\right)^2}
$$
 Equation 9

$$
rRMSE = \frac{\sqrt{\sum_{i=1}^{n} (H_d^{i,m} - H_d^{i,c})^2}}{\sqrt{\sum_{i=1}^{n} (H_d^{i,m})^2}} \times 100
$$
 Equation 10

 $\overline{H}^{i,m}_{d}$ = Aboveground biomass accumulation values from the model.

 $\bar{H}^{i,c}_a$ $=$ Aboveground biomass accumulation values from field survey (testing plots).

For consideration, how reliable is the model. The statistical value of the rRMSE model, according to the study by Despotovic et al. (2016), is presented as follows:

> $rRMSE < 10\%$ = the model is very accurate. 10% < rRMSE < 20% = the model is well accurate. 20% < rRMSE < 30% = the model is fairly accurate. $rRMSE > 30\%$ = the model is low accurate.

RESULTS AND DISCUSSION

1. Results of landuse classification.

Based on the UAV image data for landuse classification by visual interpretation. The results show that most of the land use is covered by deciduous mixed forests, with an area

n $=$ The weight of the probability values calculated from the model, with the probability value is highest and closest to the actual biomass measured in the field, while the other probability values have low values.

of approximately 72.48 hectares (97.76% of the total area), the longan plantation area was 0.83 hectares (1.12% of the total area), the water source area was 0.5 hectares (0.66% of the total area), the mango plantation area was 0.21 hectares (0.29% of the total area), and the grassland area was 0.13 hectares (0.17% of the total area), as shown in Table 1 and Figure 3.

Figure 3: Landuse classification based on UAV image data.

2. Results of the appropriate number of sample plots.

Calculate the statistical value of the biomass accumulated in each stratum and the study area were calculated to be used in calculating the appropriate number of sample plots in each using Equations 1 and 2, and set a representative area of less than 5% of the study area. From the statistical calculation, it was found that the total number of sample plots to be sampled was distributed throughout the study area, totaling 6 sample plots. Therefore, the sample plots will be defined to cover areas with low biomass density to high biomass values

to be used as representative (training/testing plots) for assessing biomass using machine learning models, as well as to verify the accuracy of the models, as shown in Figure 4.

Figure 4: Stratified maps for use in defining random sample plots to collect field data.

3. Results of development the aboveground biomass model.

Development of aboveground biomass model with machine learning model and using survey data with TLS in 6 sample plots (40X40 meters) and the aboveground biomass data in each subplot of 10X10 meters, a total of 96 subplots. Then, the data was then divided for training and testing the model at a ratio of approximately 80:20. The first stage of the model was tested using 15 variables obtained from the use of UAVs, including multispectral sensors and LiDAR cameras.

3.1 Development of aboveground biomass model from forest inventory data.

The results showed that the percentage values for importance of 15 factors that affected the calculation of the aboveground biomass model are shown in Table 2. When considering the selection of factors to be used in the final model calculation with percentage value of importance for the model more than 1 percent (sorted from highest to lowest), it consisted of Triangular Greenness, Blue band, Canopy Height Model, RedEdge band, Modified Chlorophyll Index, Green band, Normalized Difference Vegetation Index, and Structure Intensive Pigment Index with the percentage importance of each factor to the model being 23.84, 22.43, 15.34, 9.98, 9.63, 6.93, 6.45, and 3.00 respectively.

Table 2: The percentage values for importance of 15 factors to develop a preliminary

aboveground biomass model from forest inventory data.

Then take 8 factors that have the percentage importance of the model, were used to develop the model again. It was found that the percentage value of each factor had an effect on the calculation of the aboveground biomass model in each biomass class, as shown in Figure 5 and Table 3.

Table 3: The percentage values for importance of factors to develop the final aboveground biomass model from forest inventory data.

The unit is kilograms per 100 square meters.

Class4 (>450 kg/ 100 m²)

Figure 5: Probability of estimating biomass in each biomass class (kg/100 m²) using machine learning model from forest inventory data to serve as training data for the model.

The results of the machine learning model include a probability map of the estimation of aboveground biomass for each biomass class (kg/100 m²) with 4 classes. Then, the probability map of all classes were combined and the results of the aboveground biomass map of the Ban Nong Hai community forest project area (Figure 6) and the uncertainty map of the model (Figure 7) were obtained.

Results of aboveground biomass model development with machine learning model, from factors of multispectral sensors and LiDAR cameras using training data of the forest inventory found that the amount of aboveground biomass accumulation using forest inventory data has a total accumulation value of 2 ,805.4 tons or equivalent to carbon dioxide absorption of 4,834.5 tCO2e.

Figure 6: Aboveground biomass distribution map (kg/100 m²) using forest inventory data for training in the machine learning model.

Figure 7: Uncertainty map of aboveground biomass (kg/100 m²) using forest inventory data for training in the machine learning model.

3.2 Development of aboveground biomass model from TLS data.

The results showed that the percentage values for importance of 15 factors that affected the calculation of the aboveground biomass model are shown in Table 4. When considering the selection of factors to be used in the final model calculation with percentage value of importance for the model more than 1 percent (sorted from highest to lowest), it consisted of the Modified Chlorophyll Index, Blue band, Green band, Red band, RedEdge band, Canopy Height Model, Blue Normalized Difference Vegetation Index, Visible Atmospherically Resistant Index, Triangular Greenness Index, Normalize Difference Red Edge Index, and Normalized Difference Vegetation Index, with the percentage importance of each factor to the model being 26.27, 18.89, 14.26, 8.86, 8.66, 7.36, 6.15, 3.29, 3.20, 2.09, and 0.91, respectively.

Table 4: The percentage values for importance of 15 factors to develop a preliminary aboveground biomass model.

N ₀	Factors used in model development	Sum of factor	Percentage of	
		importance	importance	
		values	for factors to	
		(4 classes)	models	
$\mathbf{1}$	Modified Chlorophyll Index	105.1	26.27	
$\overline{2}$	Blue band	75.6	18.89	
3	Green band	57.0	14.26	
$\overline{4}$	Red band	35.4	8.86	
5	RedEdge band	34.6	8.66	
6	Canopy Height Model	29.5	7.36	
7	Blue Normalized Difference Vegetation Index	24.6	6.15	
8	Visible Atmospherically Resistant	13.2	3.29	
9	Triangular Greenness	12.8	3.20	
10	Normalize Difference Red Edge	8.3	2.09	
11	Normalized Difference Vegetation Index	3.6	0.91	
12	Structure Intensive Pigment	0.3	0.06	
13	Near infrared band	$\overline{0}$	$\overline{0}$	
14	Green Normalized Difference Vegetation	$\overline{0}$	$\overline{0}$	
15	Leaf Chlorophyll Index	$\overline{0}$	$\overline{0}$	
	Total	Total	400	

Then take 11 factors that have the percentage importance of the model, were used to develop the model again. It was found that the percentage value of each factor had an effect on the calculation of the aboveground biomass model in each biomass class, as shown in Figure 8 and Table 5.

Table 5: The percentage values for importance of factors to develop the final aboveground biomass model from TLS data.

Factors used in	Class 1	Class 2	Class 3	Class 4	Sum of	The
model	$(0-225)$	$(225-300)$	$(300-450)$	(>450)	importance	importance
development					value of the	percentage
					factors	of factor
Blue band	75.2	14.2	04	0.0	90.3	22.6
Modified	0.2	0.0	0.0	85.1	85.2	21.3
chlorophyll						
index						
Green band	9.6	0.3	57.0	0.0	66.9	16.7
Canopy Height	3.9	9.6	23.8	3.5	40.9	10.2
Model						
Red band	4.0	54.5	0.0	0.0	58.6	14.6
RedEdge band	0.0	15.0	0.0	10.7	25.7	6.4
Triangular	6.2	5.8	1.8	0.0	13.8	3.5
Greenness						
Blue	0.0	0.0	7.7	0.7	8.4	2.1
Normalized						
Difference						
Vegetation						
Index						
Normalize	0.0	0.0	5.3	0.0	5.3	1.3
Difference Red						
Edge						
Normalized	0.0	0.0	3.9	0.0	3.9	$1.0\,$
Difference						
Vegetation						
Index						

The unit is kilograms per 100 square meters.

Class3 (300-450 kg/100m²)

Class4 (>450 kg/ 100 m²)

Figure 8: Probability of estimating biomass in each biomass class (kg/100 m²) using machine learning model from TLS data to serve as training data for the model.

The results of the machine learning model include a probability map of the estimation of aboveground biomass for each biomass class (kg/100 m²) with 4 classes. Then, the probability map of all classes were combined and the results of the aboveground biomass map of the Ban Nong Hai community forest project area (Figure 9) and the uncertainty map of the model (Figure 10) were obtained.

Results of aboveground biomass model development with machine learning model, from factors of multispectral sensors and LiDAR cameras using training data of TLS found that the amount of aboveground biomass accumulation using TLS data has a total accumulation value of 2,825.4 tons or equivalent to carbon dioxide absorption of 4,868.9 tCO2e.

Figure 9: Aboveground biomass distribution map (kg/100 m²) using TLS data for training in the machine learning model.

Therefore, the aboveground biomass and its equivalent to carbon dioxide absorption of the study area can be summarized and compared using forest inventory data, and TLS data are shown in Table 6.

Table 6: Comparison of aboveground biomass accumulation between models using forest inventory data and TLS data.

4. Accuracy of aboveground biomass models.

4.1 Accuracy of the aboveground biomass model from forest inventory data.

The model validation using the testing dataset for accuracy verification and plotting the graph between the biomass detected in the field and the biomass obtained from the model found that the coefficient of determination $(R2)$ was 0.56 , the root mean square error (RMSE) was 114.7 kg/100 m², and the Relative root mean square error (rRMSE) was 8.8%, as shown in Figure 11.

4.2 Aboveground biomass from TLS data.

The model validation using the testing dataset for accuracy verification and plotting the graph between the biomass detected in the field and the biomass obtained from the model, found that the coefficient of determination (R2) was 0.61, the root mean square error (RMSE) was $79.8 \text{ kg}/100 \text{ m}^2$, and the relative root mean square error (rRMSE) was 4.5%, as shown in Figure 12.

Actual aboveground biomass ($kg/100m^2$)

Figure 12: The graph shows the accuracy of the aboveground biomass model using UAVs data and TLS data.

CONCLUSION AND RECOMMENDATION

Land use in the Ban Nong Hai community forest area, Chiang Mai province, found that most of the area is deciduous mixed forest, accounting for an area of about 73 hectares or 97.76 percent of the total area. In developing models from aboveground biomass data for training and various factors obtained from multispectral sensors and LiDAR cameras using machine learning models, It was found that the UAVs-Forest inventory model obtained the total accumulated aboveground biomass in the study area were 2,805.4 tons or equivalent to carbon dioxide absorption of 4,834.5 tCO2e. For the UAVs-TLS model obtained the total accumulated aboveground biomass in the study area were 2,825.4 tons or equivalent to carbon dioxide absorption of 4,868.9 tCO2e. The accuracy results of both models are very

high, but UAVs-TLS model is more accurate than the UAVs-Forest inventory model. For this study presents the following recommendations:

- 1) There should be a study of other indices that may affect the assessment of biomass density in forest areas, in order to develop a more accurate model.
- 2) The data obtained from TLS is reliable in terms of measurement accuracy. But the disadvantage of TLS is that the equipment and software are quite expensive. To analyze the data, it requires experts and takes a long time to survey.
- 3) In the development of aboveground biomass models. The more training/testing datasets there are, the more accurate the assessment of aboveground biomass.

ACKNOWLEDGMENTS

We express our gratitude to the Royal Forest Department's executive and operational levels for their joint participation in this research project and for their support of the field survey.

REFERENCES

Despotovic, M., Nedic, V., Despotovic, D. and Cvetanovic, S., (2016). Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. Renewable and Sustainable Energy.

Ogawa, H.; Yoda, K.; Ogino, K; Kira, T., (1965). Comparative ecological studies on three main types of forest vegetation in Thailand. II. Plant Biomass. Nature and Life in Southeast Asia.

Saatchi, S., Buermann, W., Ter Steege, H., Mori, S. and Smith, T.B., (2008). Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. Remote Sensing of Environment.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. Proceeding of the National Academy of Sciences.

UNFCCC. Calculation of the number of sample plots for measurements within A/R CDM project activities. Retrieved December 14, 2022, from https://cdm.unfccc.int/methodologies/ ARmethodologies/tools/ar-am-tool-03-v2.1.0.pdf/history_view.