QUANTIFYING ABOVEGROUND BIOMASS OVER 50-HA TROPICAL FOREST DYNAMIC PLOT IN PASOH, MALAYSIA USING LIDAR AND CENSUS DATA

Hamdan Omar*, Muhamad Afizzul Misman and Yao Tze Leong

Forestry and Environment Division,
Forest Research Institute Malaysia (FRIM),
52109 Kepong, Selangor, Malaysia

Corresponding author : Hamdan Omar, Phone No.; +603-62797200,
Email : hamdanomar@frim.gov.my

ABSTRACT. *Airborne light detection and ranging (LiDAR) instruments have been widely used for quantification of forest biomass. This study investigated the relationships between LiDAR data and aboveground biomass (AGB). The study area is located at the 50-ha dynamic plot in a primary forest area of the Pasoh Forest Reserve, a lowland dipterocarp forest, a type of evergreen tropical moist forest. A number of variables have been produced from the LiDAR metrics. These variables were correlated with AGB that were derived from census data. The study found that the CHM and a few matrices are the best predictors for AGB and therefore used for the estimation of AGB in the entire study area. The estimated AGB ranged from 52 to 718 Mg ha^{-1}, with a root mean square error (RMSE) of about 59 Mg ha^{-1}. The study suggests that the AGB estimates produced by this study are the most accurate - with an accuracy of 83% based on the mean absolute percentage error (MAPE) - as compared to other remotely-sensed based estimates in the study area.*

KEYWORD. Center for Tropical Forest Science (CTFS); 50-ha dynamic plot; LiDAR; biomass

INTRODUCTION

Forest biomass is a key climate variable for the global carbon cycle and has attracted the attention of both scientists and policymakers. However, forest biomass is also an attribute that is very difficult to estimate. It can be estimated using field measurements such as tree height, stem diameter, and wood density by applying allometric models obtained from destructive sampling and weighing of dried vegetation (Zolkos *et al*., 2013 & Lu *et al*., 2016), and in situ measurements can be extended to larger areas using remote sensing methods (Frolking *et al*., 2009 & Houghton *et al*., 2009).

Light Detection and Ranging (LiDAR) has become a primary source of data for assessing aboveground biomass (AGB) in various climatic regions and types of forests. LiDAR emits laser pulses and measures the return time of scattered returns to estimate the height and vertical structure of forests directly (Drake *et al*., 2002 & Lefsky *et al*., 2002). LiDAR can be acquired at high sampling density with excellent geometric accuracy and reveal AGB variation at spatial scales (Reutebuch *et al*., 2010 & Mallet and Bretar, 2009). LiDAR is currently used to bridge the gap between field measurements and remotely sensed data. Therefore, LiDAR has emerged as the most promising technology for biomass estimation, especially over tropical forests when biomass density is high (Chen, 2013).

LiDAR remote sensing systems can be distinguished based on how return signals are recorded (discrete return or waveform), scanning pattern (profiling or scanning), platforms (airborne, spaceborne or ground based) and footprint sizes. The most common configurations of LiDAR systems are airborne small footprint discrete return scanning LiDAR and has been used in a large number of studies for mapping AGB of forest. However, since the LiDAR technology evolved recently, airborne waveform LiDAR is becoming more common and this kind of data is used in this study.

Estimating AGB using LiDAR can be carried out using two approaches (Dong and Chen, 2018): (i) area-based and (ii) individual tree-based methods. Due to the structural complexity of Malaysia's tropical rainforests, these two methods offer different capability in estimating AGB in terms of accuracies and require different levels of skills to process the LiDAR point-cloud. Therefore, comprehensive research is needed to identify the relationship between AGB measured in the field and LiDAR metrics, and to determine how these relationships impact the accuracy of predictive models. This study integrates tree level census data with LiDAR variables to predict AGB in lowland dipterocarp forest. The goal of this study was to model AGB based on trees that were measured and mapped in the field, with the intention that the model could later be applied to a wider area. The immediate objectives were to: (1) develop AGB models based on census data and LiDAR data and (2) validate the model in terms of accuracy and precision. This study concentrated in the area-based method as the other methods have been investigated previously in the same study area (Wan Shafrina *et al*., 2017 & Wan Shafrina *et al*., 2018).

MATERIAL AND METHODS

It appears that area-based method is commonly used for large areas as compared to individual-tree method. This usually involves five (5) major steps, which are: (i) a sample of forest plots are set up in the field, where forest attributes are measured at the tree level and summarized at the plot level; (ii) LiDAR metrics are extracted within these field plots; (iii) LiDAR metrics are extracted for the whole study area by partitioning the study area into grid cells of equivalent size to field plots and calculating LiDAR metrics within each grid cell; (iv) statistical models are developed to predict forest attributes using LiDAR metrics using the data at the plot level; and (v) the developed plot-level models are applied to each grid cell to predict and map forest attributes for the whole study area.

This study focuses on the development of LiDAR point-cloud based models for estimating AGB in the 50-ha Forest Dynamic Plot (FDP). The study also attempts to produce estimates at the highest accuracy since the census was carried out at 100% accuracy, where all trees with ≥ 1 cm diameter were measured and the locations are well mapped. Therefore, the distribution of the sample plots can be placed anywhere within the 50-ha plot. This is one of the advantages of having 100% census data. Figure 1 shows an overall methodology that was adopted in this study.

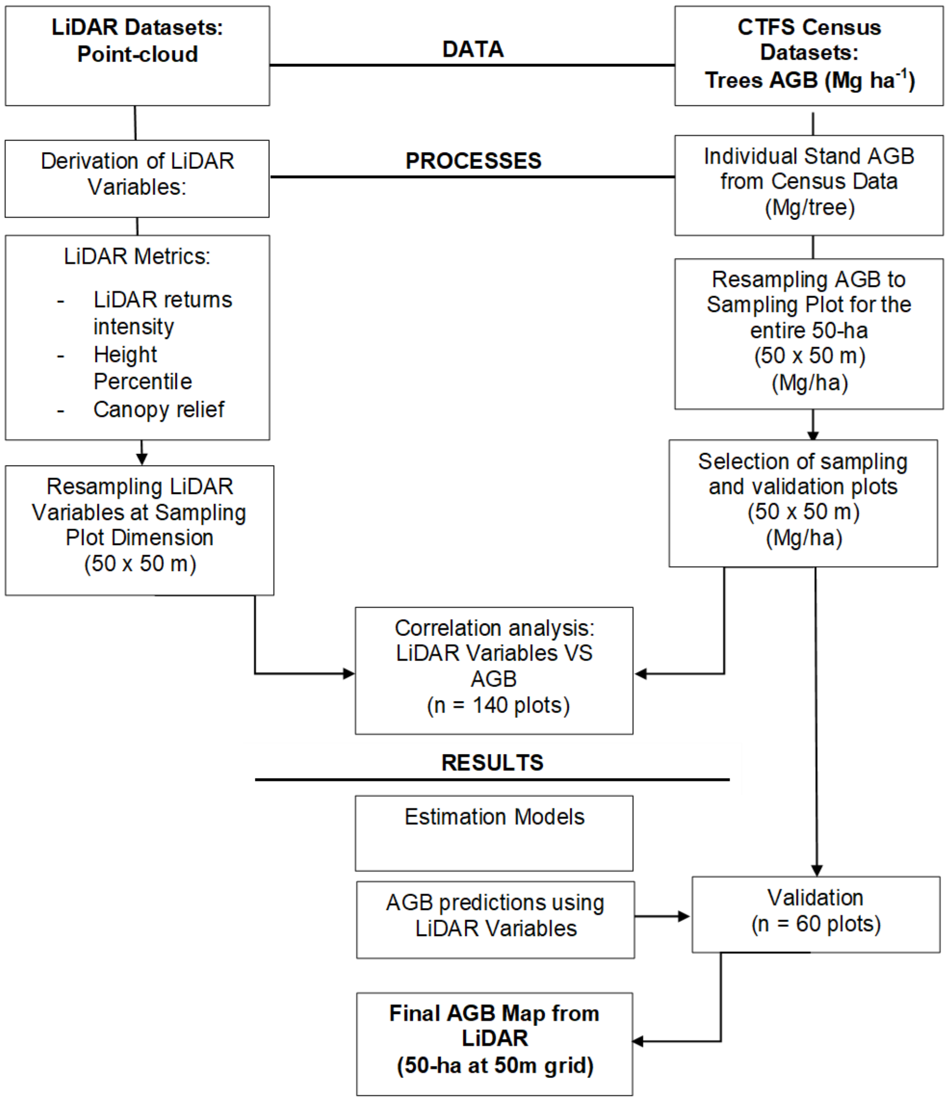

Figure 1: Flowchart of the methodology

The study area

Pasoh Forest Reserve (FR) is located in Jelebu District of Negeri Sembilan, Peninsular Malaysia (Figure 2). The FR sits in a rain-shadow valley (Ashton et al. 2003) with annual rainfall record varies between 1200–2400 mm. The forest is classified as south-central red *meranti-keruing* forest (Wyatt-Smith, 1987) characterized by the domination of Dipterocarpaceae family at the upper canopy. The main canopy formed at around 35 m height and the emergent trees reach 50–60 m tall (Manokaran *et al.*, 2003). An area totaled 1813 ha at the south-west corner of the forest reserve is designated as Pasoh Research Forest, where Pasoh 50-ha FDP lies. The plot is located in lowland primary tropical rain forest nestled within Pasoh Research Forest. The FDP was established between 1985 and 1988 to monitor long-term population and structural dynamics of primary forest (Kochummen *et al.*, 1990; Manokaran *et al.*, 1990) and re-census at five-year intervals ever since. To date, the plot sampled 435,591 trees (≥ 1 cm dbh) representing 819 species.

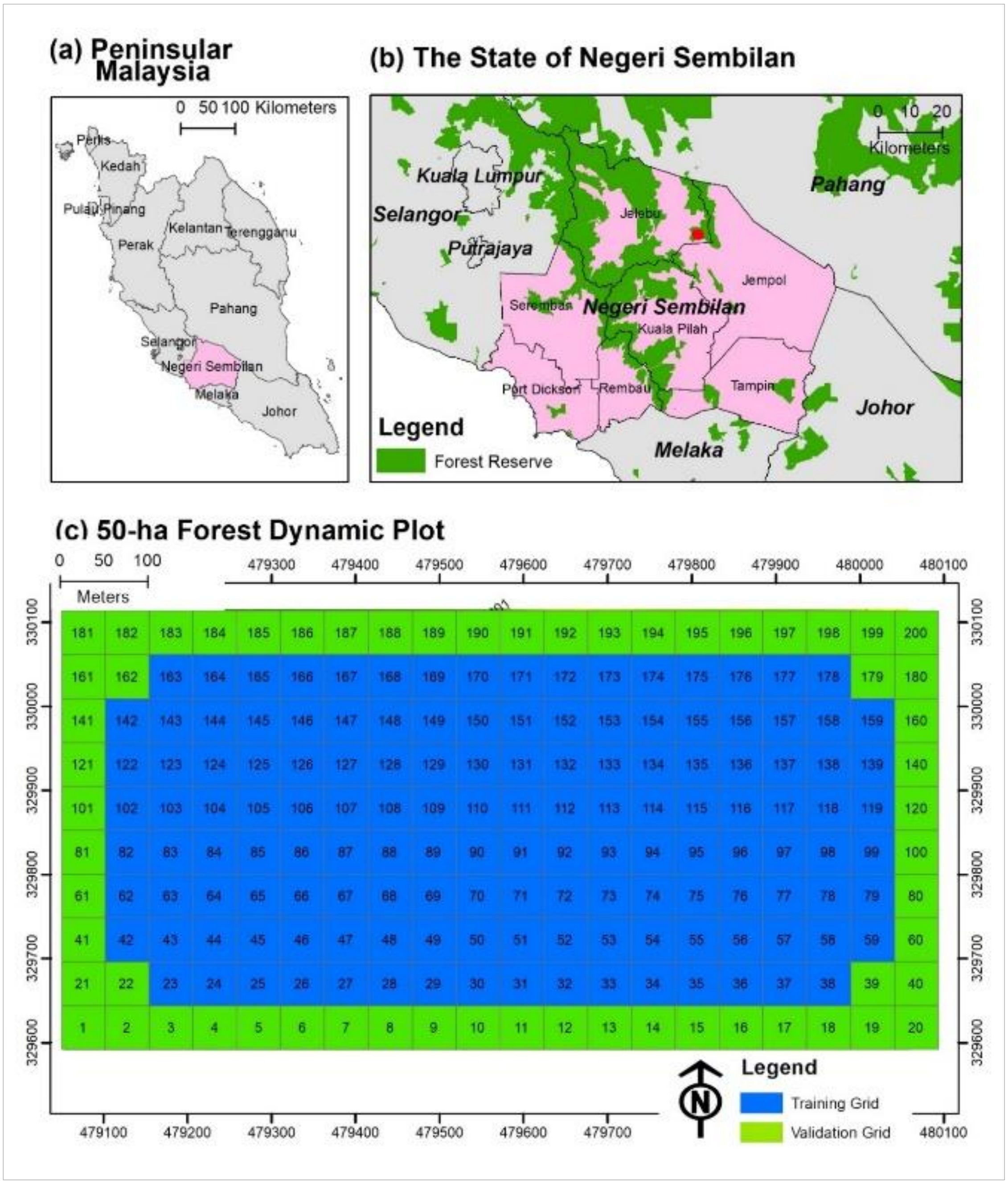

Figure 2: The location of the FDP, which is located in Negeri Sembilan, Peninsular Malaysia. The FDP measures 1000 × 500 m

Census approach

The FDP measured 1000 m length by 500 m width. Census of trees is based on protocols (Manokaran and LaFrankie, 1990 & Condit, 1998). All free-standing trees with stem diameter at breast height (DBH) measured 1 cm or greater are tagged with unique number, measured at DBH, mapped and identified to species. Tree height is not measured. The census data that was used in this study was the seventh (7^{th}) census. The census activity was started in 2015 and completed in 2018.

The 50-ha plot was divided into square plots measuring 50 × 50 m, which has made the total number of 200 plots. These plots served two purposes, i.e. one for sampling and another for validation. In this case, 140 plots were used as samples and the remaining 60 plots were used for validating the results (Figure 2-c). The AGB within each sample plot has been estimated based on the census data. The estimation were further divided into several diameter classes of trees, i.e. ≥ 5 cm, ≥ 10 cm, ≥ 15 cm, ≥ 20 cm, ≥ 25 cm, and ≥ 30 cm. Trees with dbh smaller than 5 cm are not taken into account in this study. These have produced 6 independent variables for AGB. The AGB for each tree was estimated base on allometric equation as found in (Chave *et al*., 2014).

LiDAR data

The flight mission was conducted on 26 April 2018. LiDAR data for the study area were obtained with a Riegl LMS-Q680i LiDAR system onboard EC120 Helicopter at 600 m flying altitude. The sensor scanning at ±30° scan angle at nadir with an average point density of 8.8 points per m^2 and a vertical accuracy of ±15 cm RMSE. LiDAR metrics have been derived based on the point-cloud by using R program (McGaughey, 2009). Generally LiDAR metrics are categorized into three categories, which are (i) related to height (a percentile variable), (ii) related to canopy cover and, (iii) describing the variation in the returns intensity (standard deviation or variance). Altogether 82 LiDAR metrics have derived from the LiDAR point-cloud and these were used as predictor variables for the AGB.

Statistical analysis

While there are a number of options for developing statistical models, this study focused primarily on linear regression. Existing studies indicate linear regression works best in conifer dominated forests; the predictions are not as good in tropical forests (Dong and Chen, 2018). The models cannot be applied universally and have to be developed on an area by area basis. The correlation analysis was carried out between the derived AGB (6 independent variables) and the LiDAR metrics (82 predictor variables). Linear regression method was adopted to correlate these variables and have produced 492 linear correlations. Later the best predictors out of the correlations were used to produce the best prediction models by using multiple linear regression method.

Accuracy assessment

The performance of the empirical models that have been developed from the correlations was assessed by using AGB values in validation plots. Root mean square error (RMSE) and mean absolute percentage error (MAPE) were calculated by comparison between the AGB estimated from census data and predicted from the models developed.

RESULT AND DISCUSSION

Summary of the census data

Each and every standing trees with dbh measuring ≥ 1cm was tagged as a unique ID in the census. Information such as tag number, dbh, species and position i.e. *x* and *y* coordinates are recorded in the census datasheet. Figure 3 summarizes the number of trees and the calculated AGB within the FDP. It is obvious that, although the number of small trees (dbh = 5.0 – 29.9) are plenty, the AGB within these trees are not that significant. However, larger trees (dbh ≥ 30 cm) indicated that although they small in terms of number, the AGB is substantially high. It means that the AGB is within this forest is stored predominantly in large trees.

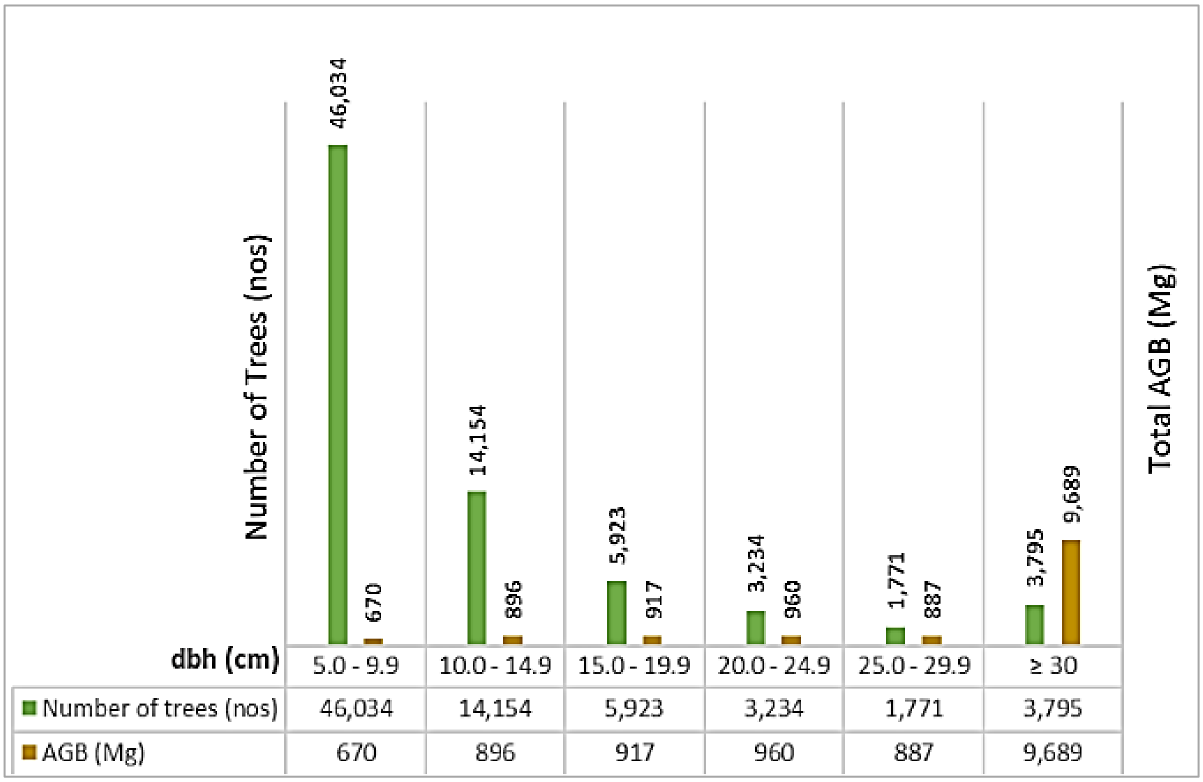

Figure 3: Summary of the census data

Altogether 74,910 trees (with dbh ≥ 5.0 cm) were taken into account for this study (Figure 4-a). Based on the census records, the remaining smaller trees (dbh ≤ 5.0 cm) was 167,451 stands, which is more than 2 folds greater that of trees ≥ 5.0 cm. However, these were not taken into account in this study. It can be concluded that the average number of trees and AGB within FDP was about 1498 trees ha^{-1} and 280 Mg ha^{-1}, respectively for trees ≥ 5.0 cm.

Estimation models

The objective of modelling is to define the best equation that represents the trend between the two sets of variables and represents reality. In this case, the LiDAR metrics were correlated with AGB to produce models. The first screening indicated that only few LiDAR metrics have linear correlation with AGB. Variables related to height and canopy relief, especially canopy height model (CHM) (Figure 4-b) are the best predictor for AGB in the study area. According to this finding, further analysis has been carried out to combine the best predictors into a single model to produce prediction functions. Multiple variables regression method was used to serve this purpose. Table 1 summarizes the models that have been derived from this process. The models represent the best-fit equation to estimate AGB for each dbh classes within the study area.

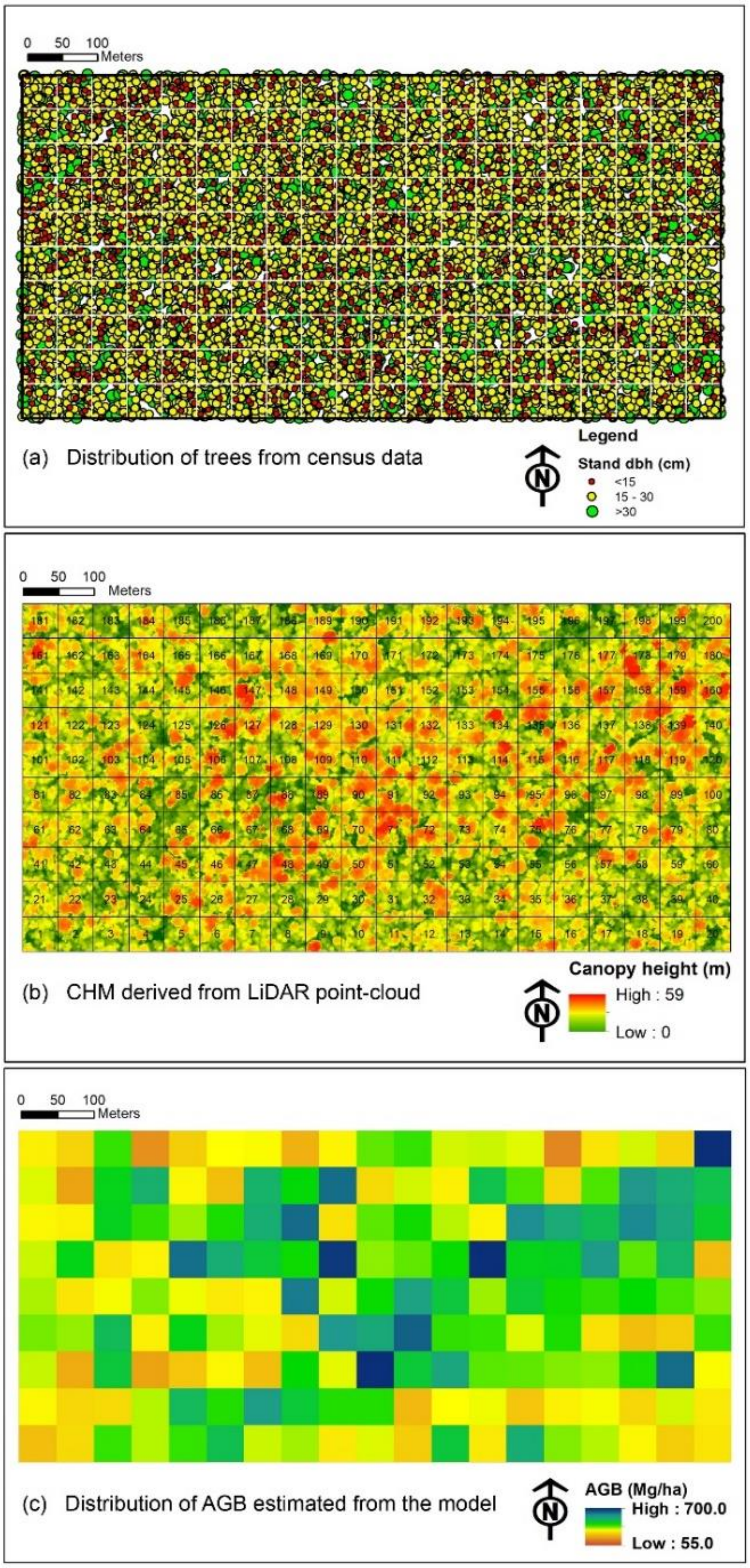

Figure 4: Spatially distributed map of census trees (a), canopy height model (b), and the estimated AGB (c).

Generally R^2 value is a good measure and commonly used for the model fit but the quest for the best R-squared can lead to over-fitting the data. The best model generally have the following principles (Dong and Chen, 2018): (i) models should have as few parameters as possible; (ii) linear models should be preferred to non-linear models; (iii) experiments relying on few assumptions should be preferred to those relying on many; (iv) models should be pared down until they are minimal but adequate, (v) simple explanations should be preferred to complex explanations. Most predictive LiDAR based models should not have more than three variables generally representing some form of the three metrics as mentioned earlier.

Referring to the Table 1, the study produced 6 different models to estimate AGB at different diameter classes. These models were validated and the accuracy was assessed. While the R^2 is showing good correlations, the RMSE and accuracy are not as good as those models of having lower R^2. These conditions indicate that the LiDAR returns interact across the stand structure, canopy conditions, and stand density of forests differently. The combination of several variables did not improve the correlation significantly as compared to that of using a single variable. However the combination of several variables can represent the reality better than that of using a single variable for estimating AGB within the study area. The total AGB within the entire study area ranged from 52 to 718 Mg ha^{-1} with an average of 284 Mg ha^{-1}. A spatially distributed map of AGB with a cell size of 50 m has been produced (Figure 4-c) and it the total AGB was estimated at 14,018 Mg for the entire FDP.

Table 1: Summary of the models developed in this study

Estimation Models	Variables	Coefficient	Adjusted R^2	RMSE (Mg ha^{-1})	MAPE	Accuracy (%)
Model 1 *(≥5 cm)*	*Intercept*	-284.45	0.48	61.68	0.18	82.41
	HSD	23.28				
	Canopy relief ratio	678.46				
Model 2 *(≥10 cm)*	*Intercept*	-293.49	0.49	61.05	0.17	83.36
	HVAR	1.44				
	HKUR	67.77				
	HSKE	-128.87				
	H05TH	2.24				
Model 3 *(>15 cm)*	*Intercept*	-154.6	0.55	59.01	0.18	82.41
	CHM	16.73				
Model 4 *(≥20 cm)*	*Intercept*	-96.15	0.56	57.96	0.19	80.77
	CHM	18.35				
	% of all returns above mean / total first return	-1.45				
Model 5 *(≥25 cm)*	*Intercept*	-191.78	0.55	59.03	0.21	78.72
	CHM	16.75				
Model 6 *(≥30 cm)*	*Intercept*	-343.52	0.5	60.71	0.25	75.23
	HVAR	1.44				
	HKUR	58.33				
	HSKE	-123.67				
	H05TH	2.23				

Another important requirement for an appropriate linear regression model is that the data are related linearly. If that assumption is violated, there may again be transformations that can be applied to create a linear relationship. Since the relationship is nonlinear and scale-dependent between LiDAR metrics and most forest attributes, the grid cell size for model prediction (i.e., minimal mapping unit) has to be equivalent to the field plot size. A critical issue is to choose the proper field plot size. When calculating plot-level forest attributes using tree-level attributes, the common practice is that a tree is included or excluded based on whether the tree trunk is inside or outside of the plot, even if the tree crown is partially inside the plot. However, the extraction of LiDAR metrics typically uses cookie-cutter approaches, so inconsistency exists between LiDAR metrics and field data for trees near the edge of plots.

Such “edge-effects” are more severe for small plots, leading to large errors in modeling forest attributes (Frazer *et al*., 2011). One way of reducing modeling errors is to use larger plots, which, however, also requires more fieldwork per plot. The choice of proper field plot size and, in general, field plot design, is an important yet difficult issue. LiDAR metrics at plot scale are commonly calculated based on either laser points or rasterized cells, even if the same formulas as listed in Table 1 are used. The metrics based on points could be sensitive to the flight conditions and the sensor setting (Roussel *et al*., 2017). In contrast, cell-based metrics can reduce such variations by focusing on only the canopy surface heights, which, however, miss the structural variations within canopy. Some researchers found that the different ways of generating metrics have small impacts on the performance of predicting forest attributes (Lu *et al*., 2012 & Chirici *et al*., 2016).

Third, although numerous LiDAR metrics can be generated from a point cloud or CHM, the metrics used to predict forest attributes should be carefully chosen to both increases the accuracy and enhance the interpretability of the models (Chen, 2013 & Magnussen *et al*., 2016).

CONCLUSION

This study has shown the potential use of waveform data in predicting structural and biophysical parameters of low land dipterocarp forest in Malaysia at high accuracy. The final models developed for AGB perform well with the adjusted R^2 ranging from 0.7 to 0.5. The lowest RMSE (59.01 Mg ha^{-1}) was derived from the model that was used to estimate AGB for tree ≥ 15 cm. However, the most accurate estimate, i.e. 83.36% was from the model that was used to estimate AGB for trees ≥ 10 cm. The study also demonstrated that estimating AGB in tropical forest by using waveform LiDAR can improve by reducing RMSE at about 40 Mg ha^{-1} as compared with other estimates from satellite imagery data (Hamdan *et al*., 2017 & Hamdan and Afizzul, 2018). Although limitations still exist, the information provided by the study can be a useful reference for other studies especially related to the applications of remotely sensed data for AGB estimations.

ACKNOWLEDGEMENT

This research was funded by the Forest Research Institute Malaysia, through the 11th Malaysian Plan for the development of Ecological Integrity Index (EII) for tropical forest ecosystems in Malaysia. The authors would like to acknowledge the Negeri Sembilan Forestry Department as the custodian of Pasoh Research Forest and Forestry Department Peninsular Malaysia and thanks to the FRIM Pasoh Research Station and all field staffs who contributed indirectly to this work.

REFERENCES

Ashton, P.S., Okuda, T. and Manokaran, N. (2003). *Pasoh Research, Past and Present*. In *Pasoh: Ecology and natural history of a Southeast Asian lowland tropical rain forest*. Okuda, T., Manokaran, N., Matsumoto, Y., Niiyama, K., Thomas, S.C., Ashton, P.S., Eds.; Springer: Tokyo, Japan. 1–13.

Chave, J., Rejou-Mechain, M., Burquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martinez-Yrizar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pelissier, R., Ploton, P., Ryan, C.M., Saldariagga, J.G. and Vielledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology,* **20(10)**: 3177-3190.

Chen, Q., (2013). Lidar remote sensing of vegetation biomass. *Remote Sensing of Natural Resources*. 399–420.

Chirici, G., McRoberts, R.E., Fattorini, L., Mura, M. and Marchetti, M. (2016). Comparing echo-based and canopy height model-based metrics for enhancing estimation of forest aboveground biomass in a model-assisted framework. *Remote Sensing of Environment*, **174**: 1–9.

Condit, R. 1998. *Tropical Forest Census Plots: Methods and Results from Barro Colorado Island, Panama and a comparison with other plots*. Springer: University of Michigan, USA.

Dong, P. and Chen, Q. (2018). *LiDAR Remote Sensing and Applications*. Taylor & Francis, Boca Raton, FL.

Drake, J.B., Dubayah, R.O., Knox, R.G., Clark, D.B. and Blair, J.B. (2002). Sensitivity of large-footprint LiDAR to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, **81**: 378–392.

Frazer, G.W., Magnussen, S., Wulder, M.A. and Niemann, K.O. (2011). Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR derived estimates of forest stand biomass. *Remote Sensing of Environment*, **115**: 636–649.

Frolking, S., Palace, M.W., Clark, D.B., Chambers, J.Q., Shugart, H.H. and Hurtt, G.C. (2009). Forest disturbance and recovery: A general review in the context of spaceborne remote sensing of impacts on aboveground biomass and canopy structure. *Journal of Geophysical Research: Biogeosciences*, **114**: G00E02.

Hamdan, O., Muhamad Afizzul, M. and Abd Rahman, K. (2017). Synergetic of PALSAR-2 and Sentinel-1A SAR Polarimetry for Retrieving Aboveground Biomass in Dipterocarp Forest of Malaysia. *Applied Sciences*, **7(675)**.

Hamdan, O. and Muhamad Afizzul, M. (2018). Time series maps of aboveground biomass in dipterocarps forests of Malaysia from PALSAR and PALSAR-2 polarimetric data. *Carbon Balance and Management*, 13:19.

Houghton, R.A., Hall, F. and Goetz, S.J. (2009). Importance of biomass in the global carbon cycle. *Journal of Geophysical Research*, **114**: G00E03.

Kochummen, K.M., LaFrankie, J.V. and Manokaran, N. (1990). Floristic Composition of Pasoh Forest Reserve a lowland rainforest in Peninsular Malaysia. *Journal of Tropical Forest Science*, **3**: 1–13.

Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A. and Gower, S.T. (2002). LiDAR remote sensing of aboveground biomass in three biomes. *Global Ecology and Biogeography*, **11**: 393–399.

Lu, D., Chen, Q., Wang, G., Liu, L., Li, G. and Moran, E. (2016). A survey of remote sensing based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth*, **9**: 63–105.

Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Vaglio Laurin, G. and Saah, D. (2012). Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates. *International Journal of Forestry Research*, Article ID 436537.

Magnussen, S., Næsset, E., Kändler, G., Adler, P., Renaud, J.P. and Gobakken, T. (2016). A functional regression model for inventories supported by aerial laser scanner data or photogrammetric point clouds. *Remote Sensing of Environment*, **184**: 496–505.

Mallet, C. and Bretar, F. (2009). Full-waveform topographic LiDAR: State-of-the-art. ISPRS *Journal of Photogrammetry and Remote Sensing*, **64**: 1–16.

Manokaran, N. and LaFrankie, J.V. (1990). Stand structure of Pasoh Forest Reserve, a lowland rainforest in Peninsular Malaysia. *Journal of Tropical Forest Science*, **3**: 14–24.

Manokaran, N., LaFrankie, J.V., Kochummen, K.M., Quah, E.S., Klahn, J., Ashton, P.S. and Hubbell, S.P. (1990). *Methodology for 50-ha research plot at Pasoh Forest Reserve*. FRIM Research Pamphlet, No. 104. FRIM: Kepong, Malaysia.

Manokaran, N., Quah, E.S., Ashton, P.S., LaFrankie, J.V., Nur Supardi, M.N., Wan Shukri, W.A. and Okuda, T. (2003). *Pasoh Forest Dynamic Plot, Peninsular Malaysia. In Tropical Forest Diversity and Dynamism, findings from a Large-Scale Plot Network*. (Eds. Losos, E.C. and Leigh, Jr. E.G.). The University of Chicago Press: Chicago, USA.

McGaughey, R. (2009). FUSION/LDV: *Software for LiDAR Data Analysis and Visualization*. US Department of Agriculture, Forest Service, Pacific Nortwest Research Station.

Reutebuch, S.E., McGaughey, R.J. and Strunk, J.L. (2010). *Sherman Pass LIDAR Forest Inventory Project*. United States Department of Agriculture, Forest Service. Pacific Northwest Research Station. 80.

Roussel, J.R., Caspersen, J., Béland, M., Thomas, S. and Achim, A. (2017). Removing bias from LiDAR-based estimates of canopy height: Accounting for the effects of pulse density and footprint size. *Remote Sensing of Environment*, **198**: 1–16.

Wan Shafrina, W.M.J., Woodhouse, I.H., Silva, C.A., Omar, H. and Hudak, A.T. (2017). Modelling Individual Tree Aboveground Biomass Using Discrete Return LiDAR in Lowland Dipterocarp Forest of Malaysia. *Journal of Tropical Forest Science* **29(4)**: 465–484.

Wan Shafrina, W.M.J., Woodhouse, I.H., Silva, C.A., Omar, H., Khairul Nizam, A.M., Hudak, A.T., Klauberg, C., Cardil, A. and Mohan, M. (2018). Improving Individual Tree Crown Delineation and Attributes Estimation of Tropical Forests Using Airborne LiDAR Data. *Forests*. **9(759)**.

Wyatt-Smith, J. (1987). *Manual of Malayan silviculture for inland forest, Part 3-Chapter 8. Red meranti-keruing forest*. FRIM Research Pamphlet No. 101; FRIM: Kepong, Malaysia.

Zolkos, S.G., Goetz, S.J. and Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment*, **128**: 289–298.