

# MODELLING INDIVIDUAL TREE ABOVEGROUND BIOMASS USING DISCRETE RETURN LIDAR IN LOWLAND DIPTEROCARP FOREST OF MALAYSIA

WS Wan-Mohd-Jaafar<sup>1</sup>\*, IH Woodhouse<sup>1</sup>, CA Silva<sup>2</sup>, H Omar<sup>3</sup> & AT Hudak<sup>4</sup>

<sup>1</sup>School of Geosciences, University of Edinburgh, Drummond Street, EH8 9XL, Edinburgh, United Kingdom

<sup>2</sup>Department of Natural Resources and Society, University of Idaho, 875 Perimeter Drive, Moscow, ID 83844, United States

<sup>3</sup>Forest Research Institute Malaysia, 52109 FRIM, Kepong, Selangor, Malaysia

<sup>4</sup>US Forest Service (USDA), Rocky Mountain Research Station, RMRS, 1221 South Main Street, Moscow, Idaho, USA – 83843

\*W.B.Wan-Mohd-Jaafar@sms.ed.ac.uk

Submitted November 2016; accepted March 2017

Light Detection and Ranging (LiDAR) has become a common means for predicting key forest structural attributes. The aim of this study was to explore the relationship between individual tree LiDAR-based metrics and field data on tree attributes from a tropical rainforest in Peninsular Malaysia, to assess the correlation between LiDAR and field data at the individual-tree level for aboveground biomass (AGB) estimates. The model was developed using multiple regression analysis, with a non-linear power model being used to fit the predictive models. The AGB model was developed based on estimated AGB at the field site and LiDAR data using the following methodology; (1) pooling of both field sample and LiDAR data, using ANCOVA to justify this approach, (2) selection of independent variables, (3) regression model development and (4) model assessment and validation. LiDAR height percentile (h80) and crown width (CW) measurement were found to best fit the data as evidenced by Adj-R<sup>2</sup> value of 0.63, root mean square error (RMSE) of the model of 14.8% and analysis of the residuals. This study provides an analytic framework for developing a predictive LiDAR-AGB model at tree level as LiDAR derived information helps natural resource managers to provide details of forest that could be derived from the biomass assessment to improve management decisions.

Keywords: LiDAR-AGB model, tree level, tropical rainforest, height percentile, crown width

## INTRODUCTION

Tropical forests in Southeast Asia have declined acutely over the past several decades (William 2007). In particular, according to a new global forest map in partnership with Google, Malaysia had the world's highest rate of forest loss between 2000 and 2012 (Butler 2013). Reducing emissions from deforestation and forest degradation (REDD+) is the framework for conserving and enhancing carbon stocks of forested area in the tropics (UNFCCC 2007). For REDD+ implementation, accurate estimation and monitoring of carbon stocks are required at the national and subnational levels. To establish robust and transparent monitoring systems, a combination of ground-based sampling and remote sensing approaches was recommended (UNFCCC 2009). Aboveground biomass (AGB) of trees in tropical forests account for significant part of the total carbon pool (Houghton et al. 2001). Therefore, estimating AGB is critical

to accurately quantifying carbon stocks in the tropics (Gibbs et al. 2007).

Tropical forests are known for their complex stand structure and abundant diversity in species composition (Steininger 2000) and estimating AGB from remote sensing data in this dense forest is challenging. Satellite-mounted optical sensors have been widely used to estimate AGB (Anaya et al. 2009). However, optical sensors acquire information from the upper canopy and are unable to measure the three-dimensional structure, including canopy height and sub-canopy topography (Lu 2006), which limits their utility to quantify AGB in tropical forests with complex canopy structures. Radar sensors (e.g., ALOS/PALSAR) use active microwave signals to generate an image, and these can be used to determine forest vertical structure (Gibbs et al. 2007), even in areas of high cloud cover such as the tropics. Radar sensors can be used for

relatively young or homogeneous forests, but their accuracy and sensitivity decrease in old-growth forests unless longer wavelengths are used (Hamdan et al. 2015).

Light Detection and Ranging (LiDAR) emits laser pulses and measures the return time of scattered returns to directly estimate the height and vertical structure of forests (Dubayah & Drake 2000, Lefsky et al. 2002). LiDAR can be acquired at high sampling density with excellent geometric accuracy and reveal AGB variation at fine spatial scales (Reutebuch et al. 2005, Mallet & Bretar 2009). LiDAR is therefore well placed to bridge the scale gap between satellite observations and field measurements (Asner 2009).

LiDAR remote sensing systems can be distinguished based on the way in which 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 is airborne small footprint discrete return scanning LiDAR, as used in this study. Airborne discrete return LiDAR has been used in a large number of studies for mapping biomass mainly using two approaches: (i) area-based and (ii) individual tree-based methods.

In this study, the individual tree-based method was considered at Pasoh Forest Reserve in Peninsular Malaysia. This site contains mixed species and is dominated by trees from the Dipterocarpaceae family, which is common in lowland dipterocarp forest. Lowland dipterocarp forest is one of the most species-rich communities in the world, with more than 200 species per hectare (Symington 1943, Wyatt-Smith 1964). Individual tree detection is seen as the most relevant approach to extract tree structural attributes in tropical rainforest characterised by a complex three-dimensional structure. LiDAR forest inventory methodologies based on individual tree detection have been widely studied, but are not widely used in practice, due to the difficulties of tree detection in various forest conditions, especially in dense, closed-canopy tropical forests (Kaartinen et al. 2012).

The most widely used LiDAR metrics for AGB prediction are various height metrics that are associated with field measurements through empirical models (Kaartinen et al. 2012). LiDAR metrics can be calculated based

on first return, last return or all of the returns (Qi 2013). In this study all returns were used to maximise the information content. Unlike most of the published algorithms that detect individual trees from a LiDAR-derived raster Canopy Height Model (CHM), this study worked directly with the LiDAR point cloud data and field data to distinguish individual trees and to estimate individual tree metrics. The CHM is a raster image interpolated from LiDAR points depicting the top of the vegetation canopy (Khosravipour et al. 2014). As a result, the CHM can have inherent errors and uncertainties from a number of sources (Khosravipour et al. 2014). However, by directly interpolating the raw LiDAR point cloud to extract individual trees, the measurements are not affected by the errors associated with interpolation, and the important 3D forest parameters can be extracted directly from the LiDAR returns that make up each tree (Li et al. 2012). To our knowledge, the use of direct LiDAR point cloud to detect individual trees and extraction of LiDAR height metrics in tropical rainforest of South East Asia has been little studied and this is one of the first studies to implement this approach for individual tree LiDAR-AGB modelling in Malaysia.

Due to the structural complexity of tropical rainforests, further research is needed to identify the relationship between AGB measured in the field and LiDAR height metrics, and to determine how these relationships impact the accuracy of predictive models. This research integrates, tree-level field-sample data with LiDAR variables to predict AGB in tropical rainforest. The goal of this study was to model individual tree AGB based on trees that were 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 a non-linear AGB model based on field sample plot and LiDAR data and (2) validate the model in terms of accuracy and precision.

## MATERIALS AND METHODS

### Site description

The Pasoh Forest Reserve (PFR) in Peninsular Malaysia was selected as the study area for this research because of its species and structural diversity. The PFR study site (2.98 N 102.31 E) is located about 8 km from Simpang Pertang, Negeri

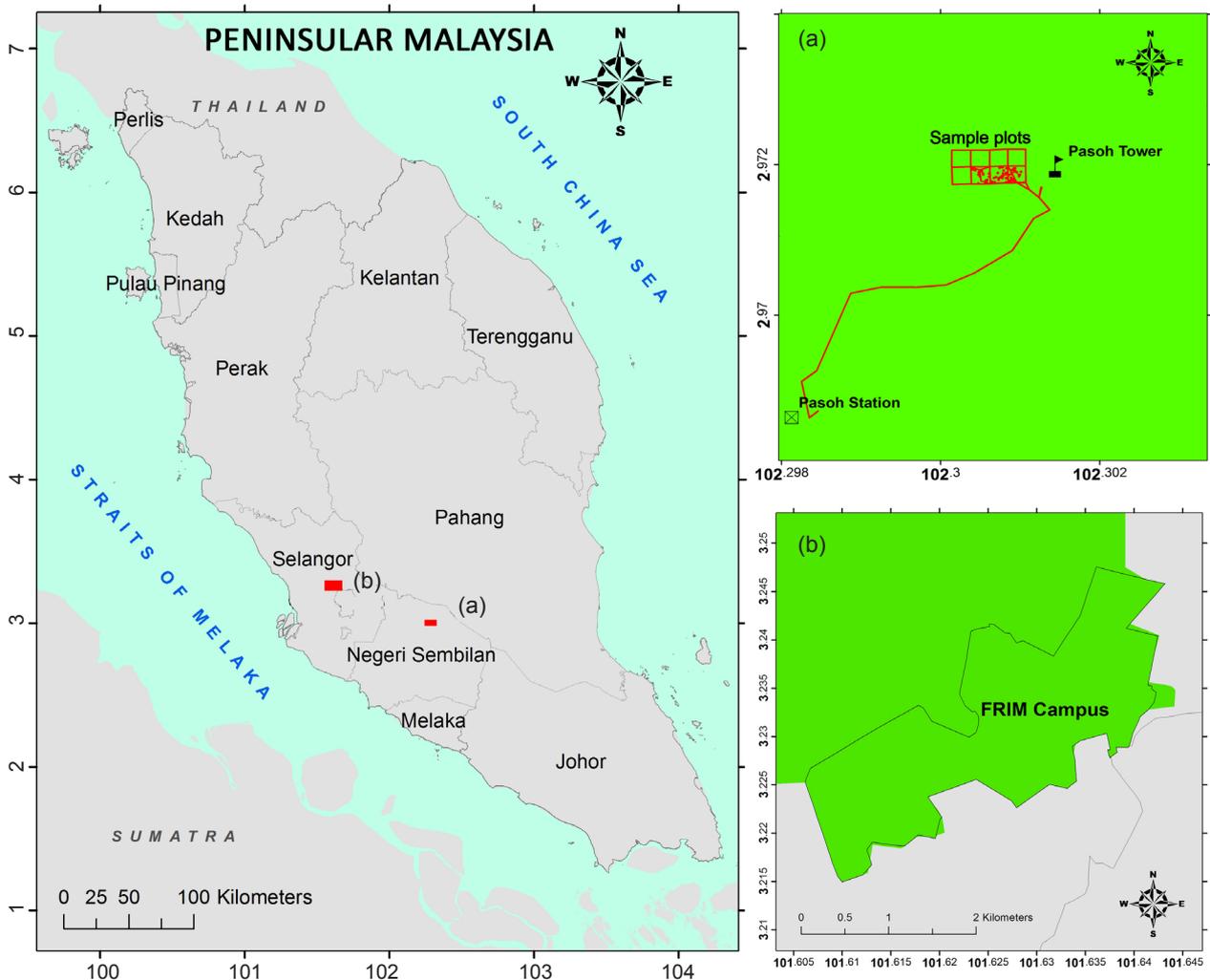
Sembilan. PFR has an area of approximately 140 km<sup>2</sup>, and is mainly covered with lowland dipterocarp forest, with upland dipterocarp forests near the north-eastern boundary. The core area of old growth forest is approximately 600 ha. Most of the surrounding area has been logged in the past, resulting in several areas of regenerating lowland forest. The PFR is one of the most species-rich forest communities in the world with 340,000 trees, ≥ 1 cm diameter at breast height (DBH), consisting of 818 species. The field measurements for this study were collected within the 6-ha International Biological Programme (IBP) plot (known as the ‘ecological plot’) in which all trees, ≥ 5cm in DBH, have been measured and mapped since the early 1970s.

In order to represent a wide sampling of lowland dipterocarp forest, the LiDAR-AGB model also incorporated field-sampled tree data

recorded from the Forest Research Institute Malaysia (FRIM) Forest Reserve. FRIM is located at Kepong, Selangor, approximately 16 km north-west of the capital city of Kuala Lumpur and 140 km away from the main test site of PFR. This 600 ha tropical forest contains approximately 15,000 species of plants. This forest has a similar structure to the study plot at PFR, where the area was stripped of its original forest cover and logged over 100 years prior, providing a generation almost as old as the primary forest. Both forest study sites are categorised as lowland forest (Figure 1).

**Field data collection**

The data used in this study include a vegetation field sample data collected in 2014 and LiDAR data collected in 2012 at PFR, as well as the



**Figure 1** Study area (a) PFR, 6-ha IBP plot location and (b) 600-ha FRIM located inside FRIM campus

vegetation field sample data and LiDAR data in FRIM both collected in 2014.

Fieldwork was conducted at PFR, Negeri Sembilan, Malaysia from 28<sup>th</sup> to 30<sup>th</sup> October 2014. Six main forest parameters were collected; horizontal position (x, y) of individual trees, DBH, total height, bole height, crown diameter and tree species. Stratified random sampling of a rectangular area of 50 m × 100 m (0.5 ha) was split into 8 experimental plots with dimension of 25 m × 25 m (Figure 2.). Each trees with DBH > 10 cm was measured within plots A1, A2, A3 and A4. The criteria was changed for plot B1 to B4 such that in these plots only emergent trees with height > 20 m were measured. The measurement strategy was designed in such a way as to facilitate the mobility of sampling work in the field. A total of 105 individual trees were used for final assessment.

Considering that the distribution of the AGB sample data deviates far from a normal distribution, the sample size of the selected 105 trees is possibly too small to build a reliable regression model representative of all species across both study areas. To address this issue, the field sample acquired in early 2014 from FRIM were evaluated to determine whether it could be incorporated into the regression

model. Since both PFR and FRIM are under the authority of FRIM, a similar sampling method and plant measurement protocol was applied at FRIM as at PFR. The FRIM is similar to PFR in terms of topography, vegetation composition and structure and land use history, because both forest are managed under the authority of FRIM. Some environmental and successional differences exist between PFR and FRIM. However, the AGB characteristics are likely to have more important impact on the LiDAR metrics than the environmental factors, assuming a relationship exists between the AGB and the LiDAR variables. To test this assumption, an analysis of covariance (ANCOVA) was conducted in this study as detailed in section ANCOVA.

The trees locations were determined using geographic coordinates of the plot centres, and the direction and distance of trees, relative to the plot centre. The plot centres were measured with a handheld global positioning system (GPS) device and the locations were post processed with local base station data, resulting in an average error of approximately 0.5 m horizontally. Tree heights were measured using a hypsometer and the DBHs and crown diameter were measured using a diameter tape (d-tape). Crown diameter was measured in four cardinal directions with

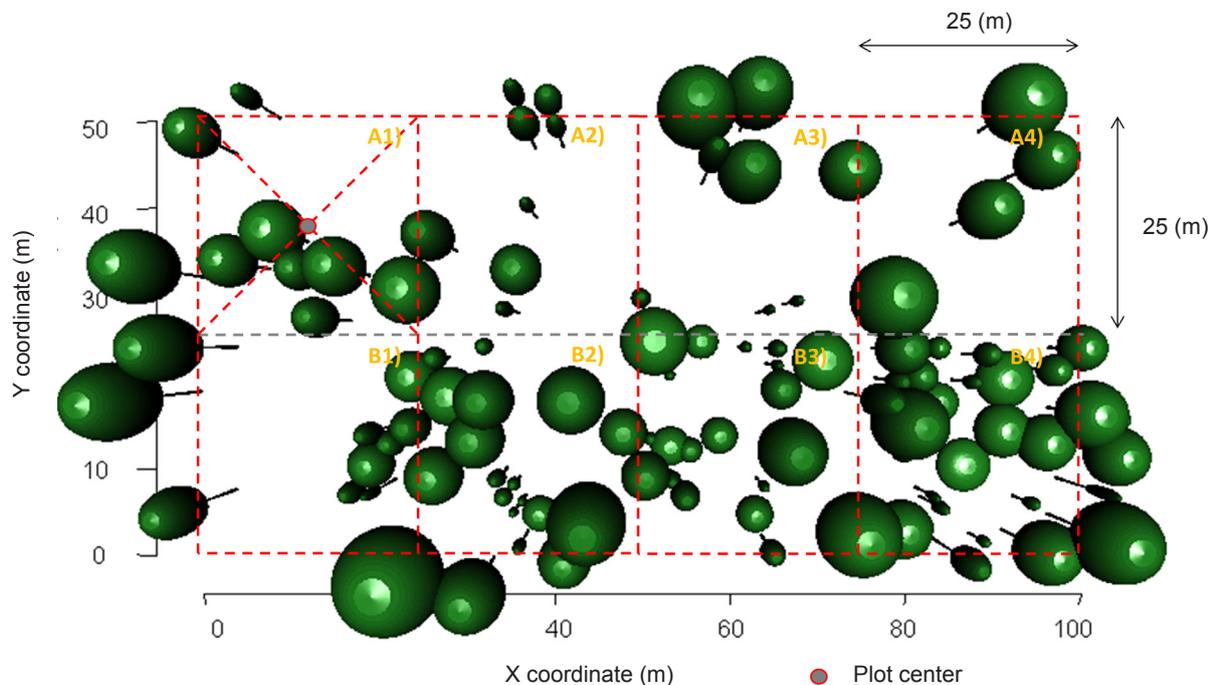


Figure 2 Dimensions of plots created within 0.5-ha plot in PFR

respect to tree trunk. Statistics describing the trees are in Table 1.

**Aboveground biomass (AGB) estimation**

An abundance of allometric models to calculate AGB have been developed in South East Asian tropical, secondary and Dipterocarp forests (Basuki et al. 2009, Niiyama et al. 2010). The selection of allometric equation for AGB are dependent on the characteristics and composition of the study area. In this study allometric equations for calculating AGB from field measurement were selected from Chave et al. 2014. This is a pan-tropical multispecies allometric equation whereby the study site is one of the many test sites of the Center for Tropical Forest Science (CTFS), used to develop the allometry model. Moreover, the chosen equation is a tree parameter incorporating DBH, height and wood density as predictor variables, describing the shape of the tree. Wood density is an important determinant of AGB, especially when a broad range of vegetation type is considered, and this model has been improved by including twice the number of trees as previous studies (Chave et al. 2014). A careful selection of allometric equations is important to reduce the uncertainty in estimating AGB:

$$AGB = 0.0673 \times (\rho D^2 H)^{0.976} \quad (1)$$

where  $\rho$  = wood density ( $g/cm^3$ ),  $D$  = diameter at breast height (DBH) (cm) and  $H$  = tree height (m).

**LiDAR operations**

The LiDAR data were obtained from a private Malaysian airborne company, using an IGI LiteMapper-5600 system with a Riegl Q560 LiDAR sensor scanning at a  $\pm 22.5^0$ , at a line rate of 60 line  $s^{-1}$ . Data processing steps include the production of radiometrically calibrated data (level 1), traceable to national standards for derived geophysical data products (level 2), which followed the application of an atmospheric correction. Finally, data were geometrically rectified to the local geo-reference co-ordinate system with user-defined ground control points (level 3). The data supplied were checked for quality, and delivered as classified and unclassified point clouds in both ASCII XYZ and LAS formats with a projection of UTM\_Zone\_48N. The data have been validated and quality checked and any possible low points have been removed. The root mean square (RMS) achieved by specific land class of forest was 0.092 m, and proved that the accuracy of LiDAR data is within tolerance of 0.15 m in vertical offset. The average point density was 8.8 points  $m^{-2}$ . The LiDAR data for FRIM was collected by FRIM. The data supplied was the same standard as the data received from PFR, with the average point density 7.5 points per square meter. Specification of the LiDAR system used at PFR and FRIM are summarized in Table 2.

**Data post-processing**

FUSION software was used to process the LiDAR data to generate three main products: the digital

**Table 1** Characteristics of individual tree field samples from both sites

Forest site	Parameters	Min	Max	Mean	SD
PFR	Tree Height (m)	7.30	39.90	18.16	6.87
FRIM		12.90	49.60	38.93	9.36
PFR	DBH (cm)	10.10	82.30	29.84	19.33
FRIM		10.00	82.20	37.92	20.86
PFR	Volume ( $m^3$ )	0.02	8.00	1.03	1.70
FRIM		0.12	9.14	2.46	2.47
PFR	AGB (kg)	8.70	8067.16	918.74	2219.91
FRIM		122.1	7771.90	2421.80	2162.92

PFR = Pasoh Forest Reserrve, FRIM = Forest Research Institute Malaysia, DBH = diameter at breast height, AGB = aboveground biomass, SD = standard deviation

**Table 2** LiDAR acquisition parameters

LiDAR sensor	IGI Lite Mapper-5600 Riegl Q560	ALTM Gemini laser system
Pulse Rate	Range between 70 KHz to 240 KHz	70 KHz
Scan Angle	$\pm 22.5^{\circ}$	$\pm 25^{\circ}$ , in increments of $\pm 1^{\circ}$
Scan Pattern	Regular	Regular
Effective rate	46,667 Hz	33- 167 KHz
Line s <sup>-1</sup>	Max 160	Max 160
Flying height	700-1000m	150- 4000 m nominal
Laser points m <sup>-2</sup>	8.8 points m <sup>-2</sup>	7.5 points m <sup>-2</sup>
Max Above Ground Level	1040m (3411ft)	5000 m (16404ft)
Data Format	ASCII XYZ and LiDAR exchange format (LAS)	LiDAR exchange format (LAS)

terrain model (DTM), the digital surface model (DSM) and the canopy height model (CHM). The ‘catalog’ function was used to evaluate the LiDAR characteristics.

A digital terrain model (DTM) was created in two steps from the discrete return LiDAR data: (1) the data were filtered to remove the above-ground returns using algorithm an adapted from Kraus and Pfeifer (2001), performed by using the ‘groundfilter’ function and (2) the DTM was created by calculating the average elevation from the remaining (ground) LiDAR returns within a cell (cells that contain no points were filled by interpolation using neighbouring cells) by using function ‘gridsurfacecreate’. The digital surface model (DSM) which represents the earth’s surface and includes the trees and other objects on it, was created using ‘canopymodel’. The discrete return point clouds were then normalised against the ground surface height and extracted for each plot using the coordinates of the lower left and upper right plot corners by using the ‘clipdata’ function. After heights were normalised, the canopy height model (CHM), which represents the height of the forest was generated by using ‘canopymodel’ function. The CHM was created for visualisation and image interpretation purposes and for manual co-registration between field sample data and LiDAR data, but not for LiDAR metrics extraction. Rather, individual tree LiDAR metrics were extracted directly from the normalised point cloud data. All point cloud data processing was performed using FUSION software (McGaughey 2014).

### Individual tree extraction—co-registering LiDAR and field sample data

The development of the co-registration procedure was based on field sample data collected during

fieldwork. The attributes relevant for this study are given in Table 3. The position of the plot edges were georeferenced with a total station. Coordinates of each tree were determined by manually seeking the optimum fit between tree positions and heights measured by forest sampling and CHM. To do this, the absolute positions of the trees within each plot were calculated from the geographical coordinates of the sample plot centres and the coordinates of the individual trees measured from field. To obtain an interpretable best fit of the tree pattern, these coordinates were then converted into ArcGIS shapefiles, which, in combination with the field height of each tree and crown diameter measured in the field, formed a polygon representing the crown dimensions of each tree. Each of the tree crown polygons was assigned a unique identification number (ID) and projected to the same projection as the LiDAR (UTM\_Zone\_48N).

To facilitate visual comparison with the LiDAR CHM, the field tree crown polygon was visualised and manually moved to best fit the shape and height of the CHM. Errors of this manual co-registration method are expected to lie at the subpixel level (i.e. < 1.0 m). Out of the 142 individual trees, 3 trees were initially removed due to the condition of being broken at first branch, which left 139 individual trees to be assessed manually. From these, 105 could be unambiguously manually co-registered. After manual co-registration, the estimated absolute planimetric accuracy of the tree location was  $\pm 0.5$  m. After carefully co-registering the individual tree polygons with the LiDAR data, using the X, Y coordinates and the crown diameter measurement for each individual tree, the tree crown polygons were used to clip the LiDAR point cloud data such that the points within each

polygon clouds were assigned the same ID as the individual tree crown polygon ID. Individual tree LiDAR metrics were then computed using the rMetrics function in the rLiDAR package (Silva et al. 2015). The generated metrics from LiDAR were used to model the individual tree heights represented by the LiDAR height metrics (Figure 3).

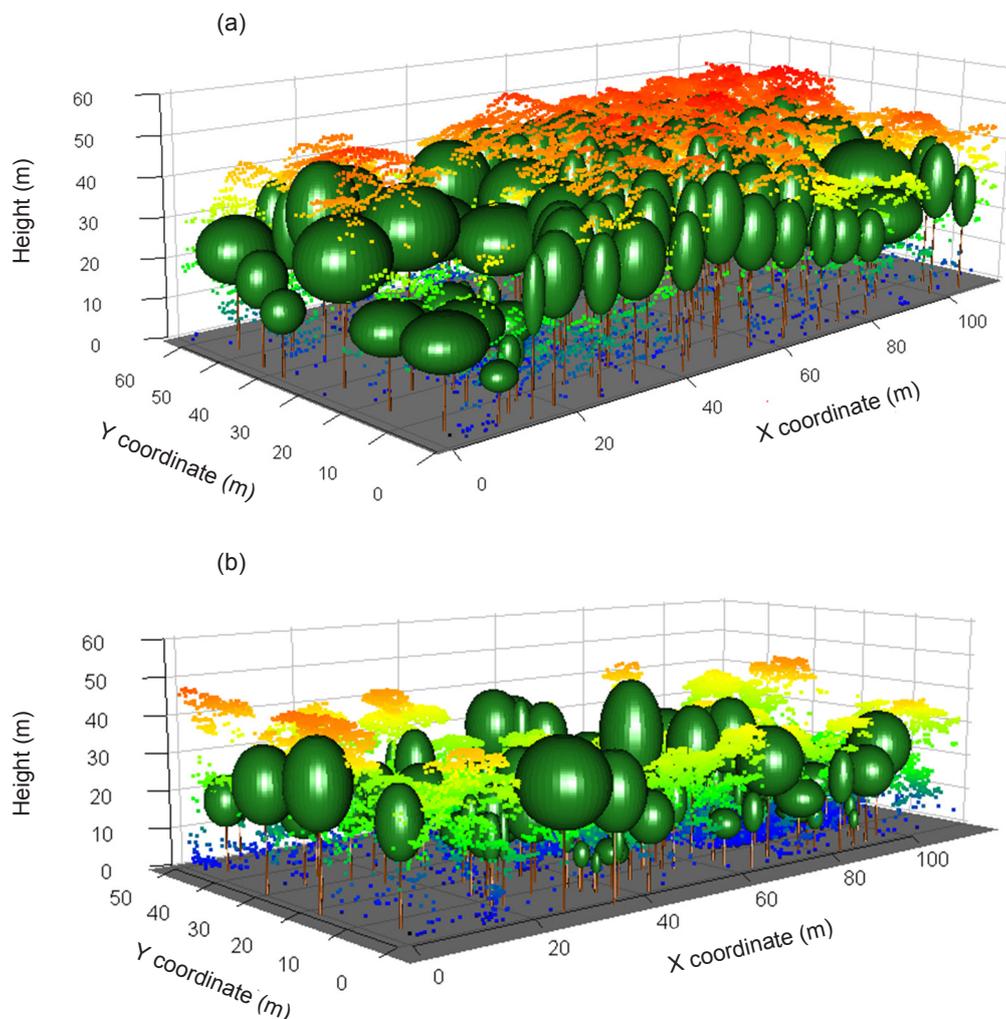
**LiDAR metrics**

Discrete return LiDAR metrics are descriptive structure statistics, calculated from the measurement of height normalised point cloud in three-dimensional space (Lefsky et al. 2005). In this study, 30 metrics for each individual tree were calculated including: (1) selected percentile heights (i.e., 5,10,...99 denoted as  $h_5, h_{10}, \dots, h_{99}$ , respectively), maximum height ( $h_{max}$ ), crown

base height (CBH), mean height ( $h_{mean}$ ), median height ( $h_{med}$ ), mode height ( $h_{mode}$ ) and (2) variability of height measures i.e., coefficient of variations of height ( $h_{cv}$ ), variance ( $h_{var}$ ), kurtosis ( $h_{kurtosis}$ ), skewness ( $h_{skewness}$ ) and standard deviation ( $h_{sd}$ ). The CW was computed using the chullLiDAR 2D function after computing the canopy area. A summary of the metrics with corresponding descriptions is shown in Table 3. To examine the impact of laser returns on AGB estimation, LiDAR metrics point clouds based on all returns were generated.

**Statistical analysis**

Assumption of homogeneity of regression slopes in analysis of covariance (ANCOVA) was tested before developing the full model. This was to ensure that field sample data from FRIM



**Figure 3** 3-dimensional tree height model; individual tree crowns extracted from the LiDAR point cloud (overlaid) at (a) FRIM and (b) PFR; trees measured in Pasoh site were based on selected criteria, not all trees were measured, thus, some trees were represented by point cloud

**Table 3** List of LiDAR metrics

Metrics	Description
Total number of returns	Total number of discrete LiDAR measurements for individual tree
Crown base height (CBH)	Points height above the crown base height
Maximum height (hmax)	Maximum height above ground of all LiDAR returns for individual tree
Mean height (hmean)	Mean height above ground of all LiDAR returns for individual tree
Median height (hmed)	Median height above ground of all LiDAR returns for individual tree
Mode height (hmode)	Mode height above ground of all LiDAR returns for individual tree
Standard deviation (hsd)	Standard deviation of heights of LiDAR returns for individual tree
Percentile height (h5, h10, h20, h25, h30, h40, h50, H60, h70, h75, h80, h90, h95, h99)	The percentiles of the canopy height distributions (5th, ..., 99th) of all returns
hcv	Coefficient of variation of heights of all LiDAR returns
hvar	Distribution of average height variance derived from all LiDAR returns
hkurtosis	Distribution of height kurtosis
hskewness	Distribution of height skewness
CW	Crown width

could be pooled with PFR for regression model development. To estimate the AGB in PFR, regression model that represents the relationship between the selected LiDAR metrics and the AGB was developed. The statistical analysis involved three steps: (1) selection of independent variables, (2) regression model development and (3) model assessment.

### ANCOVA

Since AGB at FRIM differed from PFR due to the different stand conditions, it is important that data pooling does not compromise the LiDAR-AGB relationship for the regression model in the study area. This can be tested using an ANCOVA approach. The AGB was estimated using the selected appropriate allometric equations and field data. Initially a linear regression model was chosen to observe any non-random patterns in the residuals. A curve was fitted to the trends of residuals for both forests.

ANCOVA was conducted with a general regression model having one continuous outcome variable and one or more factor variables. It tests whether certain factors have an effect on the outcome variable after adjusting for the effects of confounding factors. In this study, ANCOVA was used to determine whether the sampled data from PFR and FRIM forest could be pooled for the AGB regression model in representing both study areas through testing the assumption of homogeneous regression slopes. This assumed

that the relationship between the AGB and the LiDAR metrics in PFR and FRIM were independent of the regional conditions. To test this assumption, interaction effects were added into the regression model by adding the product of covariates and the regional factor. A power transformation was used to test this assumption because it best fit the sample data in the study:

$$Y = (\beta_0 + \beta_{\text{site}}) + (\beta_1 + \beta_{1\text{site}}) \text{Ln CW} + (\beta_2 + \beta_{2\text{site}}) \text{Ln (h80)} \quad (2)$$

where Y = field values of AGB (kg tree<sup>-1</sup>); site is a dummy variable for representing the different site (1 = FRIM, 0 = PASOH) and CW and h80 is the independent variables. When site is zero, the intercept and the slopes become  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  respectively in the model, which is derived only from PFR. When site is 1, the intercept and the slopes of the model include the addition of field sample data from FRIM. This approach was used to determine whether adding the sampled data from FRIM would impact the intercept and the slopes of the model significantly. The null hypothesis for the F test that  $\beta_0 = 0$ ,  $\beta_1 = 0$  and  $\beta_2 = 0$ ; adding the FRIM data changed neither the intercept nor the slopes.

### Selection of independent variables

The strength of relationships between AGB and the 30 LiDAR metrics were tested with the coefficient of determination ( $R^2$ ). Regression

models with highly correlated independent variables are not stable from a statistical perspective and are hard to interpret from a biological perspective (Li et al. 2008). Following the variables selection method of Næsset et al. (2005) who minimised the number of LiDAR metrics to avoid information redundancy and promote parsimony, the original LiDAR metrics were reduced to non-correlated principle components. It is difficult to interpret the principle components themselves because they are in linear combinations of the original LiDAR metrics and do not themselves have a clear physical meaning (Li et al. 2008). Some of the metrics selected introduced serious multicollinearity with VIF value greater than 3. The solution was to shift to the variable selection method of (Næsset & Gobakken, 2008) where all of the independent variables were included as possible predictor variables for selection using both stepwise variable selection with an  $R^2$  improvement technique.

The model simulates the relationship between field AGB and LiDAR variables in all sites and in combination. The estimated AGB, a dependent variable, was calculated based on the converted ground measurements taken during October, 2014 at PFR and April, 2014 at FRIM. All 30 variables of LiDAR metrics were used as potential model-independent variables.

### Regression model development

Power functions have been used in a large number of studies for AGB estimation, probably because most allometric equations for calculating AGB in the field are power functions. The relationship between AGB and forest height has been well described (Morel et al. 2011, Wang et al. 2013), and the power function is widely accepted (Niklas & Enquist 2001) as:

$$B = a (H)^b \quad (3)$$

where  $B$  = plot AGB ( $\text{mg ha}^{-1}$ ),  $H$  = field tree height (m) and  $a$  &  $b$  = coefficients.

The allometric equations used to derive AGB from field data in this study were based on a power function developed by Chave et al. (2014) and thus, provide a solid justification for the chosen model. Since most allometric equations for calculating tree-level AGB from field measurements are power models, AGB

and LiDAR metrics were log transformed to fit a regression model. This is to reduce the heterogeneity of the regression residual variance. A multiplicative model formulated as equation (3) was selected, which can be translated into a linear form according to equation (4). This type of model has been used successfully by others to model various forest biophysical properties (Naeset 2002, Lim et al. 2003). The natural logarithm transformations required in equation (4) also ensure that, in most cases, regression assumptions are not violated. In the regression analysis, a linear multiplicative model was used which correspond to multiplicative power transformation, as used successfully by others to estimate various forest attributes. All derived variables were transformed to the natural logarithm. First, AGB was regressed against height metrics and AGB as a dependent variable. Multiple linear regression analysis using all independent variables was then carried out for both sites independently and in combination. Stepwise selection using Akaike Information Criterion (AIC) was performed in R to select variables included in the final model. The AIC is a goodness of fit measure that favours smaller residual error in the model, but penalises the inclusion of predictors and helps avoiding overfitting. At each step, individual variables were either added or deleted and the next model with lowest AIC was retained in the next step. The coefficient of determination ( $R^2$ ), the adjusted coefficient of determination ( $\text{Adj-}R^2$ ) and the root-mean-square error (RMSE) were calculated for model comparisons.

$$Y = \beta_0 + \beta_1 CW + \beta_2 h_{\max} + \beta_3 CBH + \beta_4 h_5 + \dots + \beta_{30} h_{99} \quad (4)$$

$$\ln Y = \beta_0 + \beta_1 \ln CW + \beta_2 \ln h_{\max} + \beta_3 \ln CBH + \beta_4 \ln h_5 + \dots + \beta_{30} \ln h_{99} \quad (5)$$

where  $Y$  = field values of AGB ( $\text{mg ha}^{-1}$ ),  $h_{\max}$  = maximum height of canopy,  $CBH$  = crown base height,  $h_5, \dots, h_{99}$  = percentiles corresponding to 5, ..., 99% of the laser canopy height (m),  $CW$  = crown size width. Both stepwise variable selection and the maximum  $R^2$  variable selection techniques were applied to select LiDAR variables, to be included in the models. When using stepwise selection method, no independent variables were left in the models with a partial  $F$  statistic significance level greater than 0.05. The

best fitting models were selected based on the lowest AIC value. The variance inflation factor (VIF) was used to address multicollinearity issues by calculating and monitoring the size of the condition number. Because the best models suggested by the stepwise procedure might have many independent variables that could introduce multicollinearity problems, the maximum  $R^2$  improvement technique searches for the best one-variable model.

In accordance with the objectives of this study, the influence of forest site on the estimated AGB models were assessed by extending the preliminary regression models derived above with dummy variables representing these factors. Since both sites were lowland dipterocarp forest with almost similar species, the homogeneity test to assess whether both datasets can be pooled for model development was only assessed on variables related to site properties. To assess the effects of different forest site, the dummy variables were assigned a value of 0 for sampled data in Pasoh and a value of 1 for sampled data in FRIM. The result are discussed in ANCOVA analysis in ANCOVA.

### Model assessment

Two questions were essential in assessing the model: (a) how well does the model fit the sampled data (model fitting analysis) and (b) is the model generalisable outside the sampled data? In the model fitting analysis, potentially influential data outliers were identified using the studentised deleted residual and Cook's distance statistic to assess a data point's influence on the regression coefficients of the regression model; generally, it should be considered a potentially influential point when its value exceeds 1. The goodness of fit of the model was evaluated using an adj- $R^2$  and the RMSE, and standardised residuals were used to check the following model assumptions: (1) the equal variance for all independent data and (2) no systematic pattern between the regression model residuals and the predictions. The  $R^2$  evaluates the fit of a linear regression model to the sampled data. However, the impact of the degree of freedom on the model accuracy is well known; the more predictive variables in the model, the higher the  $R^2$ . Clearly, the  $R^2$  is not a good index to assess the model's goodness of fit. An Adj- $R^2$  removes the impact of the degree of freedom, and thus

provides a more conservative measure of the model's goodness of fit.

The assumption of normality of error terms were assessed using the Shapiro-Wilk test (Shapiro & Wilk 1965) and heteroscedasticity were tested using the Breusch-Pagan test (Breusch & Pagan 1979). Comparisons between models were based on their predictive capabilities with respect to the coefficient of determination ( $R^2$ ), root mean square error (RMSE) and AIC between observed and predicted values. Once the best model was chosen, 10-fold cross validation was performed. RMSE were used as a way to evaluate the measured values and to estimate the true value. RMSE was calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

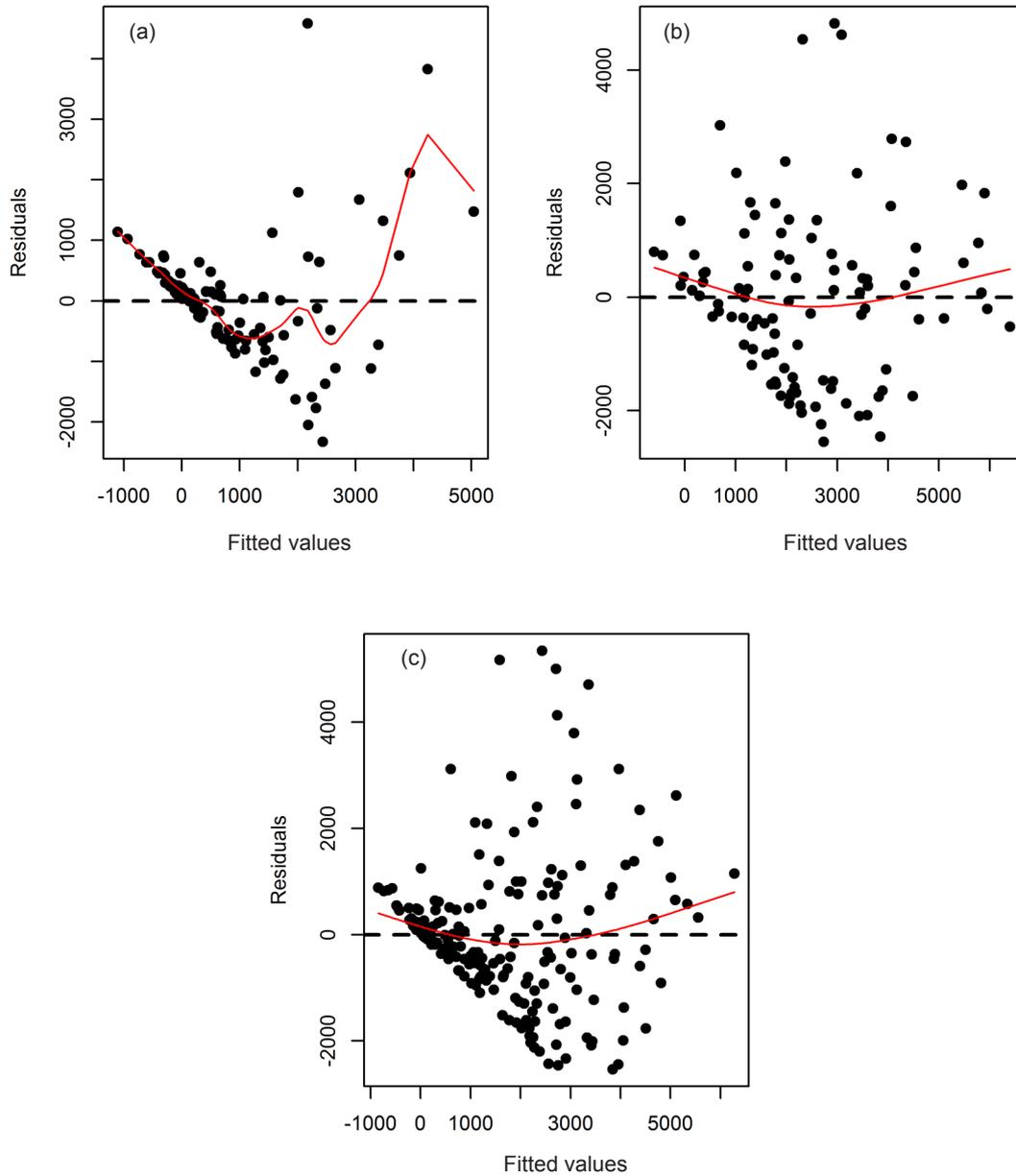
where  $y_i$  = observed values and  $\hat{y}_i$  = predicted values for the  $i$ th compound, respectively, and  $n$  = number of samples in the training set.

### RESULTS

A linear regression model was chosen to evaluate whether a relationship exists between AGB and LiDAR variables. Residuals versus modeled AGB were plotted to examine the assumption of linearity. To help observe patterns of residuals, a curve was fitted to the trends of residuals (Figure 4). The plot clearly displayed non-linear trends in the residuals.

### Independent variables selection

Regression models with log-transformed variables were estimated with LiDAR derived metrics as the only independent variables. The models were selected by a stepwise regression procedure using all possible independent variables from both sites independently and combined. These are summarised in Table 4. To simplify the models, the maximum  $R^2$  improvement technique was used to find the best variable combination and the best-two variable model and so-forth, with the benchmark of the best fitting model being the one with the lowest AIC value. Multicollinearity was further evaluated to confirm that all independent variables had correlations below 0.90, a VIF value below 5, and high correlation with AGB.



**Figure 4** Residuals on fitted linear regression, (a) PFR site, (b) FRIM site and (c) pooled data between PFR and FRIM site, all indicating randomness existed in their relationship

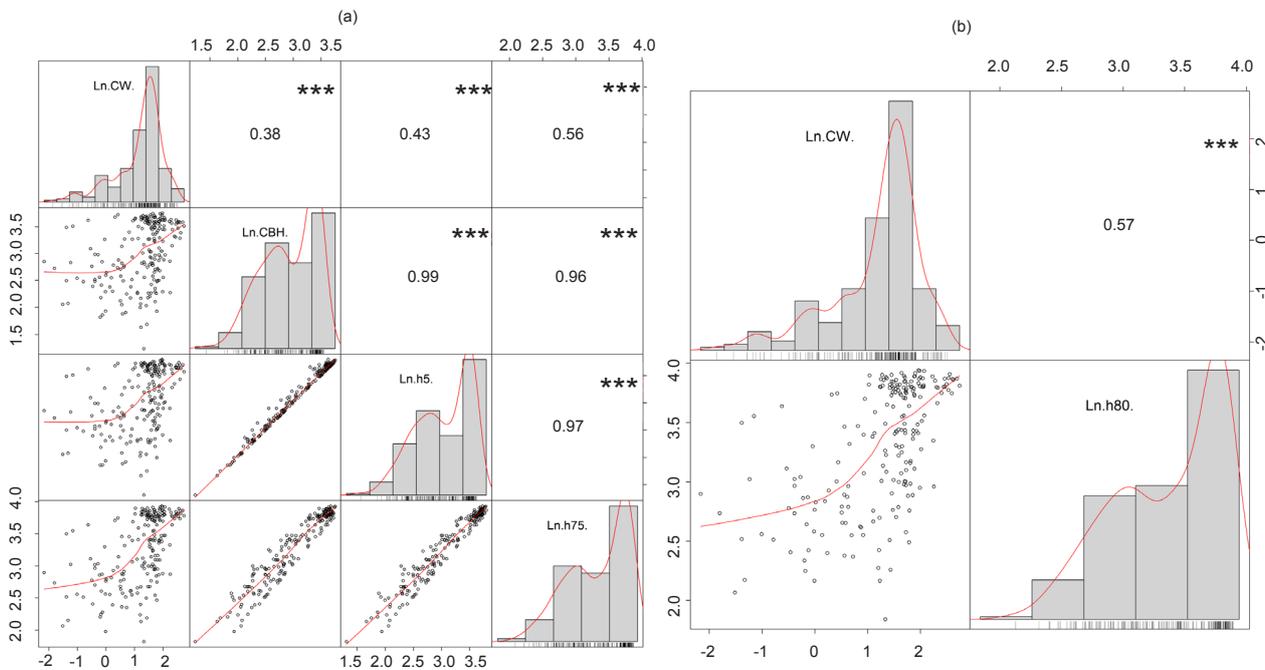
Based on high Adj-R<sup>2</sup> and low rRMSE values, the predictive models, combining data from two forest sites, proved as an efficient strategy to predict AGB of the region. The best model was selected based on lowest AIC value, however, the selection of the best variables was determined by the correlation coefficients and p-values of the partial F statistic of selected metrics. Figure 5 shows a scatterplot matrix of the two models with lowest AIC value. The model combining four variables (CW, CBH, h5 and h75) gave the lowest AIC value and highest Adj-R<sup>2</sup>, but

this combination introduced multicollinearity with VIF value > 5 and high inter-correlated independent variables with  $r > 0.90$ . This value can be summarised through Figure 5(a). The model combining the CW and h80 variables, however, passes all the diagnostic tests with  $VIF < 3$  and low correlation between predictors with  $r < 0.70$  used for further analysis in this study, as summarised in Figure 5(b). The histogram of each metric are drawn in the diagonal line. The kernel density overlaid and the significant asterisks shows the level of significant (\*0.05,

**Table 4** Top four best variable combinations, based on lowest AIC values from two forest sites considered independently and in combination

Study site and model	Selected independent variables (natural log transformed)	Adj-R <sup>2</sup>	AIC	RMSE (kg tree <sup>-1</sup> )	rRMSE
PFR	1) CW, CBH, h70	0.74	256.45	0.79	13.65%
	2) CW, h70	0.74	256.72	0.79	13.74%
	3) CW, CBH, h5, h70	0.74	257.05	0.78	13.61%
	4) h90	0.61	298.71	0.97	16.84%
FRIM	1) CBH, hmode	0.25	300.10	1.00	13.73%
	2) CW, hvar, hcv, h90	0.22	305.85	1.01	13.96%
	3) h90	0.11	316.81	1.09	14.96%
	4) hmedian	0.12	317.70	1.08	14.86%
Pooled Data site	1) CW, CBH, h5, h75	0.63	579.65	0.95	14.52%
	2) CW, h80	0.63	581.90	0.96	14.68%
	3) CW, CBH, h5	0.63	591.15	0.97	14.96%
	4) h90	0.53	627.68	1.12	16.41%

PFR = Pasoh Forest Reserve, FRIM = Forest Research Institute Malaysia, AIC = akaike information criterion CW = crown width, RMSE = root mean square error, CBH = crown base height, h = percentiles



**Figure 5** The matrix of scatter plots (lower panel) and the correlation coefficients (r) (upper panel) for all detected trees from all possible best metrics in pooled data site; the correlations are presented as (a) high correlation and (b) low correlation

\*\*0.01 and \*\*\*0.001). Further model assessment will focus on the model based on the combined sites instead of the individual test sites.

From the selected independent variables, an empirical approach was employed to identify the most appropriate curve that fitted the data. Linear, cubic, quadratic and power curve functions were fitted to the data. The R<sup>2</sup>, Adj-R<sup>2</sup> and RMSE were used to determine the most

appropriate model. Of the four non-linear functions tested, the power function was found to best fit the sample data (Table 5).

**ANCOVA for pooling data from different forest sites**

The regression output of equation (1) fit to the 209 individual trees from both sites and the F

**Table 5** Comparison of the curve fitted transformation in nonlinear regression model from pooled sample data

Independent variable	Model	R <sup>2</sup>	Adj-R <sup>2</sup>	RMSE (kg tree <sup>-1</sup> )
CW	Linear	0.385	0.382	1745.172
	Cubic	0.395	0.386	1739.572
	Quadratic	0.386	0.380	1748.366
	Power	0.450	0.447	1.161
h80	Linear	0.245	0.241	1933.415
	Cubic	0.254	0.244	1930.686
	Quadratic	0.251	0.244	1929.913
	Power	0.532	0.530	1.071

CW = crown width, h = percentiles, R<sup>2</sup> = coefficient of determination, Adj-R<sup>2</sup> = adjusted coefficient of determination, RMSE = root-mean-square error

test for the overall model was statistically highly significant ( $p \leq 0.001$ ) and provided a good fit to the data (Table 6). It was found that all the interaction effects, Ln (CW) \* dummy site and Ln (h80) \* dummy site were not statistically significant according to their p values (0.782 and 0.339 respectively), which supported the null hypothesis: the response variables of the AGB in PFR and FRIM were independent of the site factor. Therefore, there were no statistically significant changes in the regression slopes of the model in the presence of different data from FRIM site. This result confirmed an early assumption that the vegetation structure was likely to have more impact on the LiDAR metrics than site-level environmental factors. Therefore, the data from FRIM could be pooled with the PFR data to develop the regression models and to mitigate the issue of the limited sample size in this study.

### Data outlier and influential analysis

The results from the regression are shown in Table 7. The p-value and R<sup>2</sup> value for the model were  $\leq 0.001$  and 0.6112, respectively, thus the model provided a good fit to the data.

### Regression model fitting analysis

Multicollinearity is a serious issue that must be considered when using regression models. Multicollinearity was controlled by checking the VIF of the models. All independent variables and the model as a whole had a VIF below 5. Normal probability plots of the residuals suggested that the residuals were normally distributed. Besides

the normal Q-Q plot distribution, residuals from the prospective model were also tested for normality by using the Shapiro-Wilk test and analysis on heteroscedasticity by using Breusch-Pagan test. Both tests accept the null hypothesis that residuals from the model are normally distributed by reporting  $p = 0.13$  for Shapiro-Wilk test, and the variance of the error term is constant (homoscedasticity) with  $p = 0.34$ .

By adopting the power transformation into the model's design, the residual variances were reduced and the nonlinearity was addressed. Further, the linearity was not violated according to the normal probability plot of the residuals and the power function model was best fit to the sample data. Next, how well the proposed model could predict the outcome in different datasets (known as model generalisability) need to be assessed. The generalisation of the proposed regression model was evaluated using the 95% confidence intervals of the prediction in addition to cross-validation. The model performed well for small predicted values (Figure 6).

The mean square error (MSE) from a ten-fold cross validation was 0.907 compared to 0.856 that was estimated from the 105 samples. Since the two results were reasonably close, the proposed model can be considered generalisable. Figure 7, shows the model's predictive results for both the training and calibration data sets in each pass. Prediction of the omitted test data are used to assess the predictive accuracy. The dashed lines are parallel and close to each other. The model's curve were scattered gradually in the low AGB values and close to each other towards the high AGB range, slightly showing a positive trend of errors in the low AGB range.

**Table 6** Result of analysis of covariance derived from 209 individual trees

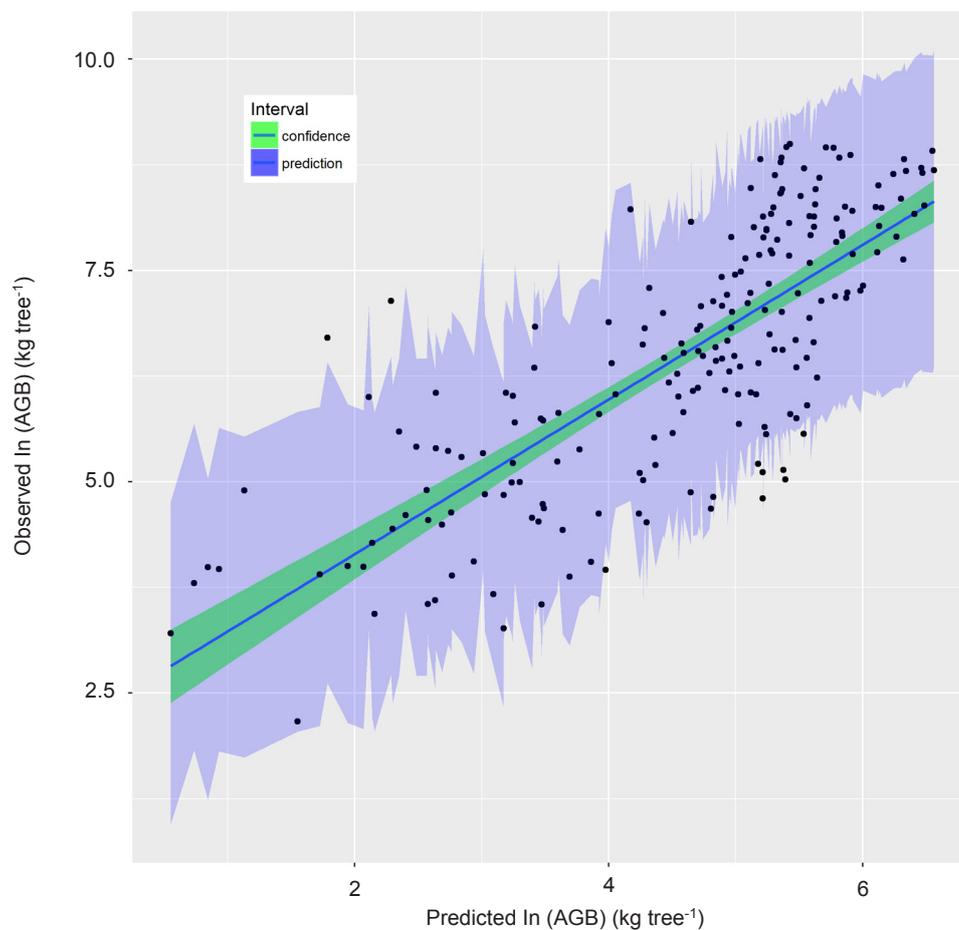
Source	Sum of squared (type III)	Df	F value	Significance
Intercept	0.044	1	0.049	0.8250
Ln (CW)	39.546	1	43.861	3.04e-10
Ln (h80)	71.087	1	78.844	3.41e-16
Ln (CW) * dummy site	0.069	1	0.077	0.782
Ln (h80) * dummy site	0.827	1	0.917	0.339

\*Residual standard error = 0.9495 on 205 degrees of freedom, Adj-R<sup>2</sup> = 0.630, F-statistic = 90.04 on 4 and 205 DF, p-value < 2.2e-16, CW = crown width, h = percentiles, Ln = natural logarithm

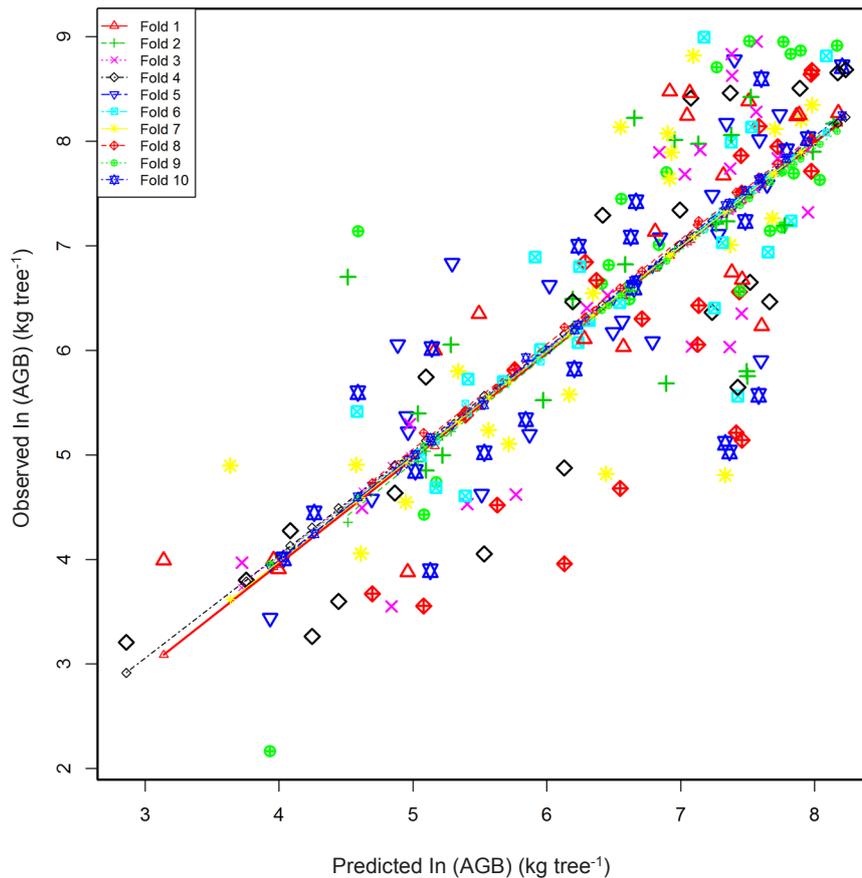
**Table 7** Results of nonlinear power model

	Coefficients	Estimated standard error	t value	Pr (>  t  )
(Intercept)	0.612	0.463	1.324	0.187
Ln (CW)	0.629	0.085	7.375	3.9e-12
Ln (h80)	1.540	0.154	10.012	<2e-16

Ln = natural logarithm



**Figure 6** The 95% confidence intervals of the predictions of power regression model, 95% probability that the true best-fit line for aboveground biomass (AGB) lies within the confidence interval and 95% of the y-values to be found for a certain x-value within the interval range around the linear regression line; ln = natural logarithm



**Figure 7** Summary of 10-fold cross validation for evaluating aboveground biomass (AGB) power model, the line is fitted to the training data set in each pass, leaving out corresponding test data set; ln = natural logarithm

AGB from at the two sites was well estimated using the final prediction model predicting AGB from two LiDAR metrics (i.e. CW and h80), with Adj-R<sup>2</sup> of 0.63 and RMSE of 14.68% (Figure 8).

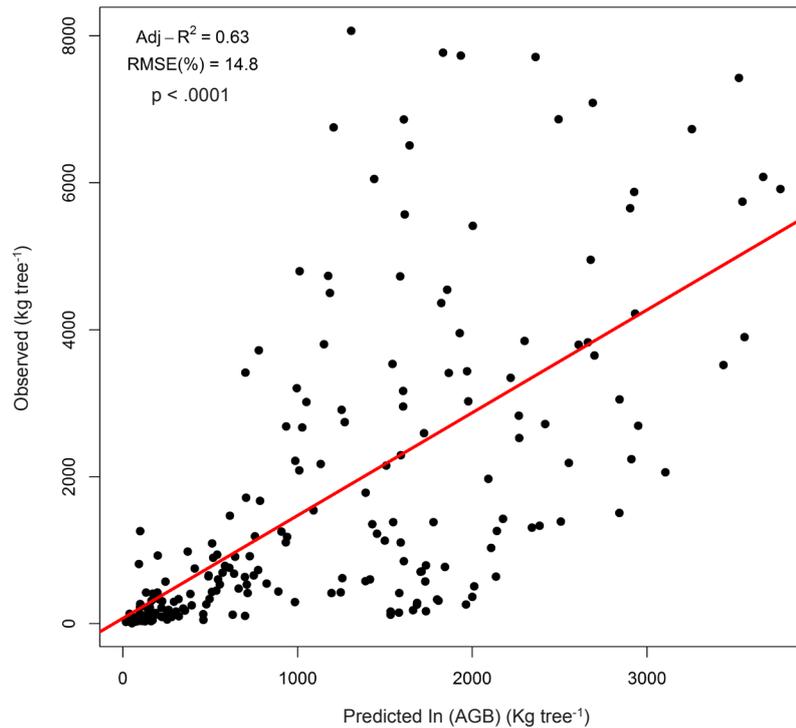
## DISCUSSION

Malaysia's interest in participating in the World Bank's Forest Carbon Partnership Facility (FCPF) (Hamdan 2012) requires a baseline study to calculate carbon, so as to present the country's current status of carbon stored in forest. The development of biomass model for carbon stock estimation will help to support the research development for carbon monitoring methodology in Malaysia, with aims to prepare forested countries for REDD+ implementation in Malaysia (FRIM 2011, FRIM 2012). To best of our knowledge, this is one of the first studies to develop a model using individual tree and LiDAR-derived crown metrics at tree level in tropical rainforest. The increasing importance

of accurate biomass estimation to support the REDD+ implementation, has created a critical need to understand, evaluate and improve current tree biomass prediction methods by adopting state-of-the-art analytical and statistical techniques. In this study, LiDAR metrics that correlate very well with field AGB was extracted based on trees that were mapped in the field as the main point to develop the LiDAR-AGB model.

### Field-LiDAR AGB model

Field sample data can be valuable for evaluating LiDAR-based and other remotely sensed AGB maps, as plots are systematically arranged to provide a spatially unbiased estimate of forest AGB over an area, followed by well-documented measurement protocols that are quality controlled (Johnson et al. 2014). Findings from this study show that the LiDAR metrics and stem-localised AGB correlated very well



**Figure 8** Scatter plot of final observed *versus* predicted aboveground biomass (AGB) from the combination at both sites

with < 15% error in the residuals. An interesting finding from this research was that CW was one of the best LiDAR metrics for predicting AGB. Incorporating both crown size and tree height, particularly of large trees, may improve estimates from remote sensing data for both standing carbon stocks and carbon stock changes and this may be especially applicable to methods based on small footprint LiDAR (Goodman et al. 2014). A recent published study by Ferraz et al. (2016) also incorporates CW, from decomposed entire point cloud into 3-Dimensional (3D) clusters that correspond to individual tree crown. The 3D clusters were modelled using a convex hull to calculate the crown area (CA) and the crown volume (CV) using the same tools developed by Silva et al. (2015), as used in this study. However, the method allows for the use of many forest parameters in existing field allometrics equations to estimate the AGB. This method is only appropriate for forest areas with well-established field AGB allometry. The model developed was based on a geolocated stem map that represent most of the species which exist for lowland dipterocarp forest for building up LiDAR-AGB model with an adequate comparison of individual tree at stem level. Developing a regression model for AGB estimation at stem level is important to

derive the most important LiDAR metrics that correlate with field AGB, before implementation at landscape level. The initial approach for quality assessment was presented in this study.

Herein CW is related to DBH, direct retrieval of DBH from LiDAR point cloud at tree level is possible at temperate forest but very little has been addressed (Maltamo et al. 2009, Vauhkonen et al. 2010). However, it is not possible to apply the approach to tropical rainforest. Instead, an interpolation through the relationship with crown size is needed. Given that extraction of CW can be achieved through LiDAR, it is possible that future studies could retrieve DBH using similar methods. The CW and DBH relationships of tropical forest has previously been reported (Perez 1970), the CW and DBH relationships has also been reported (Wile 1964, Roberts & Ross 1965, Bonner 1968) for a number of conifer species. All of these studies found a strong relationship between DBH and CW ( $R^2 = 0.6-0.9$ ).

### Metrics selection and their explanation

The analysis involved all independent variables in the initial search, finding correlations among independent variables, starting the analysis

using the less correlated metrics ( $r < 0.7$ ) and discarding those with least importance until the classification accuracy became stable. The top-ranked metrics were selected based on the lowest AIC value as the first criteria. To make sure the best model with the lowest AIC value is stable from the statistical perspective, multicollinearity was tested using VIF. The best model based on AIC contained CW, CBH, 5<sup>th</sup> percentile and 75<sup>th</sup> percentile. However, it also introduced high multicollinearity based on VIF value  $> 5$  and the  $p$ -values were not significant for CBH (0.06) and 5<sup>th</sup> percentile (0.32) variables. There were also strong high correlations among independent variables as shown in Figure 5(a). Regression models with highly correlated independent variables are not stable and hard to interpret from either statistical or biological perspectives (Naesset & Gobakken 2008). The second best model with the lowest AIC value was the model containing CW and 80<sup>th</sup> percentile height as the independent variables for predicting AGB. This model passed all the diagnostic tests, with all variables having a VIF less than 5, suggesting no serious multicollinearity in the model. In addition, all variables were significant in terms of  $p$ -values, and there were no statistically significant changes in the regression slopes of the model in the presence of combining data from FRIM field sample plot.

There have been other studies reporting different height metrics that correlate very well with AGB that varies not just depending on forest type and location, but also on model and data processing procedures (Thapa et al. 2015). In this study, multiplicative models using power functions performed best for predicting AGB from metrics generated from the raw point cloud LiDAR. AGB is usually non-linearly related to remote sensing variables, therefore, nonlinear transformations such as a power function where the response and explanatory variables were log transformed (Hall et al. 2005) reduces the heterogeneity of the regression residual variance. On the plus side, power functions have been used in a large number of studies for AGB estimation, probably because most allometric equations for calculating AGB in the field are power models (Qi 2013). An empirical approach has been implemented in this study to identify the most appropriate curve that fitted the data and power model was found to be the best fit for the sample data, providing justification for the model specification.

## Analysis of covariance and regression model

Although environmental and successional differences existed between PFR and FRIM, the regression output of ANCOVA demonstrated that the relationship between AGB and LiDAR metrics was independent of these two sites by confirming the assumption of homogeneous slopes. High AGB is related to the size and structure of trees. The smaller the sample size, the greater the errors will be; limited sample size was overcome by combining the sample data between two sites similar in nature. The 105 samples from PFR were probably too small to represent the tree species and structural diversity exist in tropical rainforest. To overcome the issue of uncertainty in estimating the regression coefficients and intercepts, pooling sample data from similar forest types is a means to overcome uncertainty in AGB estimation. The final selected model, did not violate assumptions of equal variance and normality. However, there was a slight positive trend in the lower predicted AGB errors confirmed in both the variances of residuals and the ten-fold cross-validation analysis. It implies that the model might provide a better AGB estimation with a higher AGB than in the lower AGB.

This study provided direct retrieval of individual metrics (e.g. tree height and crown size) that allow testing of the hypothesis that LiDAR-based AGB models can replace or complement ground-based AGB models. Ferraz et al. (2016) implemented a similar approach through direct retrieval of LiDAR individual tree metrics, but the model assessment was dependent on a ground-based AGB model, which is an approach only appropriate for a forest area with well-established field AGB allometry. In contrast, the current approach used a precisely geolocated stem map that represented the majority of tropical rainforest species to develop the AGB-LiDAR model. It is much more reliable as it showed that important LiDAR metrics correlated very well with field-based AGB measurements, before implementing at a landscape level.

## CONCLUSIONS

ANCOVA analysis confirmed that the relationship between AGB and LiDAR were independent of the forest site, supporting an early assumption that the characteristics of the vegetation structure were likely to have more dominant influence on

LiDAR metrics than environmental and other factors between the study areas. Combining data from two forest sites has two benefits. First tropical forest is a very complicated ecological system with diverse species and different biophysical structures, therefore the 105 tree sampled at PFR were possibly too small to build a reliable regression model to represent the whole population that exist in the region. Since PFR and FRIM forest is similar in terms of topographic conditions, vegetation cover and land use history, it is useful to combine the datasets for model calibration and validation. Second, in terms of statistics, combining data is a great way to understand the result of applying multiple models and approaches. The regression model in this study indicated a good correlation between LiDAR predictors and AGB, as in many prior studies.

A power function was identified as the best solution to fit the modelled data. Of 30 LiDAR metrics, h80 and CW were identified as the best independent variables to predict AGB. An interesting finding from this study, supporting earlier findings of Goodman et al. (2014), was the importance of incorporating crown size and height to improve estimates of AGB carbon in tropical forest especially, also based on small footprint LiDAR. This research provided an analytic framework for developing a predictive AGB model from LiDAR and field plot data and will be of value especially for forest resource managers for estimating the AGB in lowland dipterocarp forest to improve management decisions.

It should stress that, the method used to extract LiDAR data for the trees was not an automatic individual tree detection method, but based on trees that were mapped in the field. The results of this study, may have been affected by the manual tree delineation, and individual tree position on the field was recorded by handheld GPS, however this approach tends to give an over-promising impression of the methods. It is suggested that an automatic procedure with focus to derive multi-layered crown delineation will be a promising avenue for future research. The usefulness of producing accurate tree-level data by means of LiDAR should therefore be assessed carefully with respect to alternative methods, model improvement and the costs involved.

## ACKNOWLEDGEMENTS

The authors would like to thank Airborne Informatics Malaysia, vendor of LIDAR data, and Geoinformation Programme team, FRIM, for providing access to Pasoh Forest Reserve, field assistance in data collection and fieldwork instruments and diligence with plot geolocation setup. The authors would also like to thank the survey team from the Faculty of Architecture, Planning and Surveying, MARA University of Technology, Malaysia for setting up GPS plot location during fieldwork campaign. Utmost thanks to Majlis Amanah Rakyat (MARA), Malaysia (330408224838) for supporting the PHD research, and Lamb Fund, Geography Centenary, School of Geosciences, University of Edinburgh, UK (E08802) to have supported the research during fieldwork campaign.

## REFERENCES

- ANAYA JA, CHUVIECO E & PALACIOS-ORUETA A. 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management* 257: 1237–1246.
- ASNER GP. 2009. Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters* 3: 1748–9326
- BASUKI TM, VAN LAAKE PE, SKIDMORE AK & HUSSIN YA. 2009. Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests. *Forest Ecology and Management* 257: 1684–1694.
- BONNER GM. 1968. Stem diameter estimates from crown width and tree height. *Commonwealth Forestry Review* 47: 8–13.
- BREUSCH TS & PAGAN AR. 1979. A Simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47: 1287–1294.
- BUTLER RA. 2013. Malaysia has the world's highest deforestation rate, reveals Google forest map. <https://news.mongabay.com/2013/11/malaysia-has-the-worlds-highest-deforestation-rate-reveals-google-forest-map/>.
- CHAVE J, REJOU-MECHAIN M, BURQUEZ A ET AL. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology* 20: 3177–3190.
- DAWKINS HC. 1963. Crown diameters: their relation to bole diameter in tropical forest trees. *The Commonwealth Forestry Review* 42: 318–333.
- DUBAYAH RO & DRAKE JB. 2000. LiDAR remote sensing for forestry. *Journal of Forestry* 98: 44–46.
- FERRAZ A, SAATCHI S, MALLET C & MEYER V. 2016. Lidar detection of individual tree size in tropical forests. *Remote Sensing of Environment* 183: 318–333.
- FRIM. 2011. *Research on Development of Carbon Monitoring Methodology for REDD+ in Malaysia*. Forest Research Institute Malaysia, Kepong.

- FRIM. 2012. *Research on Development of Carbon Monitoring Methodology for REDD+ in Malaysia*. Forest Research Institute Malaysia, Kepong.
- GIBBS HK, BROWN S, NILES JO & FOLEY JA. 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters* 2: 045023.
- GOODMAN RC, PHILLIPS OL & BAKER TR. 2014. The importance of crown dimensions to improve tropical tree biomass estimates. *Ecological Applications* 24: 680–698.
- HALL SA, BURKE IC, BOX DO, KAUFMANN MR & STOKER JM. 2005. Estimating stand structure using discrete-return LiDAR: an example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management* 208: 189–209.
- HAMDAN O. 2012. *An Overview of REDD+ Activities in Malaysia*. Forest Research Institute Malaysia, Kepong.
- HAMDAN O, MOHD-HASMADI I, KHALI-AZIZ H, NORIZAH K & HELMI-ZULHAIDI MS. 2015. Determining L-band saturation level for aboveground biomass assessment of dipterocarp forests in Peninsular Malaysia. *Journal of Tropical Forest Science* 27: 1–12.
- HOUGHTON RA, LAWRENCE KT, HACKLER JL & BROWN S. 2001. The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates. *Global Change Biology* 7: 731–746.
- JOHNSON KD, BIRDSEY R, FINLEY AO ET AL. 2014. Integrating forest inventory and analysis data into a LiDAR-based carbon monitoring system. *Carbon Balance and Management* 9:1–11.
- JOHNSON KD, BIRDSEY R, FINLEY AO ET AL. 2014. Integrating forest inventory and analysis data into a LiDAR-based carbon monitoring system. *Carbon Balance and Management* 9:1–11.
- KAARTINEN H, HYPPIÄ J, YU X ET AL. 2012. An international comparison of individual tree detection and extraction using airborne laser scanning. *Remote Sensing* 4: 950–974.
- KHOSRAVIPOUR A, SKIDMORE AK, ISENBURG M, WANG TJ & HUSSIN YA. 2014. Generating pit-free canopy height models from airborne LiDAR. *Photogrammetric Engineering and Remote Sensing* 80: 863–872.
- KRAUS K & PFEIFER N. 2001. Advanced DTM generation from LiDAR data. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences* 34: 23–30.
- LEFSKY MA, COHEN WB, HARDING DJ, PARKER GG, ACKER SA & GOWER ST. 2002. Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography* 11: 393–399.
- LEFSKY MA, HUDAK AT, COHEN WB & ACKER SA. 2005. Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sensing of Environment* 95: 517–531.
- LI WK, GUO QH, JAKUBOWSKI MK & KELLY M. 2012. A new method for segmenting individual trees from the LiDAR point cloud. *Photogrammetric Engineering and Remote Sensing* 78: 75–84.
- LI YZ, ANDERSEN HE & MCGAUGHEY R. 2008. A comparison of statistical methods for estimating forest biomass from Light Detection and Ranging data. *Western Journal of Applied Forestry* 23: 223–231.
- LIM K, TREITZ P, WULDER M, ST-ONGE B & FLOOD M. 2003. LiDAR remote sensing of forest structure. *Progress in Physical Geography* 27: 88–106.
- LU DS. 2006. The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing* 27: 1297–1328.
- LU D, CHEN Q, WANG G ET AL. 2012. Aboveground forest biomass estimation with landsat and LiDAR data and uncertainty analysis of the estimates. *International Journal of Forestry Research* 2012: 1–16.
- MALLET C & BRETAR F. 2009. Full-waveform topographic LiDAR: state-of-the-art. *Isprs Journal of Photogrammetry and Remote Sensing* 64: 1–16.
- MALTAMO M, PEUHKURINEN J, MALINEN J, VAUHKONEN J, PACKALÉN P & TOKOLA T. 2009. Predicting tree attributes and quality characteristics of Scots pine using airborne laser scanning data. *Silva Fennica* 43: 507–521
- MCGAUGHEY RJ. 2014. Fusion/LDV: software for LiDAR data analysis and visualization version 3.50. [http://forsys.cfr.washington.edu/fusion/FUSION\\_manual.pdf](http://forsys.cfr.washington.edu/fusion/FUSION_manual.pdf).
- MOREL AC, SAATCHI SS, MALHI Y ET AL. 2011. Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. *Forest Ecology and Management* 262: 1786–1798.
- NAESSET E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80: 88–99.
- NAESSET E, BOLLANDSAS OM & GOBAKKEN T. 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sensing of Environment* 94: 541–553.
- NAESSET E & GOBAKKEN T. 2008. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment* 112: 3079–3090.
- NIYAMA K, KAJIMOTO T, MATSUURA Y ET AL. 2010. Estimation of root biomass based on excavation of individual root systems in a primary dipterocarp forest in Pasoh Forest Reserve, Peninsular Malaysia. *Journal of Tropical Ecology* 26: 271–284.
- NIKLAS KJ & ENQUIST BJ. 2001. Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proceedings of the National Academy of Sciences of the United States of America* 98: 2922–2927.
- PEREZ JW. 1970. Relation of crown diameter to stem diameter in forest of Puerto Rico, Dominic and Thailand. Pp B105–B122 in Odum HT & Pigeon RF (eds) *A Tropical Rainforest – A Study of Irradiation and Ecology at El Verde*. U.S. Atomic Energy Commission, Division of Technical Information, Springfield.
- QI C. 2013. *LiDAR Remote Sensing of Vegetation Biomass*. *Remote Sensing of Natural Resources*. CRC Press and Taylor & Francis Group, Boca Raton.
- REUTEBUCH SE, ANDERSEN HE & MCGAUGHEY RJ. 2005. Light Detection and Ranging (LiDAR): an emerging tool for multiple resource inventory. *Journal of Forestry* 103: 286–292.
- SHAPIRO SS, & WILK MB. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52: 591–611.
- SILVA CA, CROOKSTON NL, HUDAK AT & VIERLING LA. 2015. LiDAR: An R package for reading, processing and visualizing LiDAR (Light Detection and Ranging) data, version 0.1. <https://cran.rproject.org/web/packages/rLiDAR/index.html>.

- STEININGER MK. 2000. Satellite estimation of tropical secondary forest above-ground biomass: data from Brazil and Bolivia. *International Journal of Remote Sensing* 21: 1139–1157.
- SYMINGTON CF. 1943. *Foresters' Manual of Dipterocarps*. University of Malaya, Kuala Lumpur.
- THAPA RB, WATANABE M, MOTOHKA T, SHIRAIISHI T & SHIMADA M. 2015. Calibration of aboveground forest carbon stock models for major tropical forests in central Sumatra using airborne LiDAR and field measurement data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing* 8: 661–673.
- UNFCCC 2007. Reducing emissions from deforestation in developing countries: approaches to stimulate action, conference of the parties. Paper presented at The United Nations Framework Convention in Climate Change Conference, Bonn.
- UNFCCC 2009. Recommendations on future financing options for enhancing the development, deployment, diffusion and transfer of technologies under the convention. Paper presented at The United Nations Framework Convention in Climate Change Conference, Bonn.
- VAUHKONEN J, KORPELA I, MALTAMO M & TOKOLA T. 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment* 114: 1263–1276.
- WILLIAM FL. 2007. Forest destruction in tropical Asia. *Current Science* 93: 1544–1550.
- WANG X, OUYANG S, SUN OJ & FANG J. 2013. Forest biomass patterns across northeast China are strongly shaped by forest height. *Forest Ecology and Management* 293: 149–160.
- WILE BC. 1964. *Crown Size and Stem Diameter in Red Spruce and Balsam Fir*. Department of Forestry Publication, Ottawa.
- WYATT-SMITH J. 1964. A preliminary vegetation map of Malaya with descriptions of the vegetation types. *Journal of Tropical Geography* 18: 200–213.