EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology, 2025, 26(2), 178–192
https://doi.org/10.12912/27197050/197151
ISSN 2719–7050, License CC-BY 4.0

Received: 2024.11.21
Accepted: 2024.12.21
Published: 2025.01.01

Aerial imagery and UAV-LiDAR data fusion for quantifying aboveground carbon stock of teak plantation

Emma Soraya[1*], Deha Agus Umarhadi[1,2], Wahyu Wardhana[1], Senawi[1], Fiqri Ardiansyah[1]

[1] Faculty of Forestry, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2] Department of Biology, Ludwig Maximilian University of Munich, Munich, Germany
* Corresponding author's e-mail: esoraya@ugm.ac.id

ABSTRACT

This study examines four regression models to estimate tree-level above-ground carbon stock (AGC) in the clonal teak plantations of Wanagama Forest, Indonesia using a data fusion of UAV-LiDAR and RGB aerial imagery. Data collected in the field were diameter at breast height and tree height. Vegetation indices were derived from the visible bands of a georeferenced orthomosaic captured by DJI Mavic 2 Pro. The LiDAR data was obtained using GeoSun GS-100M device mounted on a quadcopter drone. Spectral features included RGB bands and vegetation indices, while vertical features consisted of a canopy height model (rasterized of normalized LiDAR point clouds) and relative height of LiDAR data. Additionally, the teak crown diameter was derived from object-based segmentation. Our study shows that crown diameter, Z max, and one of the vegetation indices are the three most influential features in estimating the AGC of individual tree in all models. Multiple linear regression (MLR) with eleven independent variables outperformed other models, showing a determination coefficient of 0.73 and providing the most accurate predictions of individual tree AGC, with MAE = 28.35 kg and WAPE = 34.13%. The teak stand predicted using MLR shows an average AGC of 64.15 Mg ha^{-1}. The regression models developed are accurate for clonal teak plantations and provide a non-destructive way to monitor and manage carbon stocks as the basis for forest management and conservation strategies. However, the LiDAR and RGB spectral features may not capture all relevant tree characteristics, since this study acquired the airborne data during a teak leaf-off season, which may be challenging to distinguish between individual tree causing overestimated tree numbers. Teak sheds its leaves during drought thus similar studies at leaf-on conditions might adequately represent the characteristics of the teak tree.

Keywords: regression, point clouds, visible vegetation indices, clonal teak, spectral.

INTRODUCTION

Tropical forests are the largest forests in the world, accounting for nearly half (45%) of the total global forest area and contributing 200–300 Pg C from standing trees, equivalent to one-third of the atmospheric carbon (Mitchard, 2018; Sarre, 2020). Nevertheless, environmental issues relating to forest loss and degradation occurred worldwide including that in Indonesia leading to an increase in greenhouse gas (GHG) emissions (Margono et al., 2014; Potapov et al., 2022; Tsujino et al., 2016). The country has committed to participating in reducing GHG emissions through the FOLU Net Sink 2030 program to achieve carbon sequestration equal to or even higher than the carbon emissions produced (Nurbaya, 2023). Some actions include the establishment of plantation forests and forest rehabilitation with long rotation species, for instance teak (*Tectona grandis*).

Teak plantation forests cover an extensive area in lowland monsoon forest in Java Island, Indonesia, mostly planted by a state forest enterprise, Perhutani (Darmawan et al., 2015). Teak is a hard wood species which has a high value in the global market due to its durability for construction and furniture (Chayaporn et al., 2021). Teak genetic improvement has been implemented

in the last two decades to produce tree clones that can increase forest productivity, so that the fast-growing clonal teak acts in creating greater carbon sink over time (Behera and Priyadarshini, 2015; Wirabuana et al., 2022). Therefore, teak plantations play a double pivotal role: not only in the economy, but also in the environment by sequestering carbon from the atmosphere.

Effective and accurate measurement and mapping of this plantation forest biomass stock at various scales has never been more critical as this data is essential. First, for monitoring progress towards emission reduction targets and developing more effective climate mitigation policies and strategies. Second, for reporting the state's greenhouse gas emissions and removal to meet its climate commitments. Third, for carbon trading schemes and other economic incentives/financial support that reward carbon sequestration efforts such as forest conservation and restoration projects.

Traditionally, tree biomass or so-called aboveground biomass (AGB) and AGC are measured directly in the field with laborious and time-consuming works, yet unable to reflect the spatial aspect of an extensive area. Remote sensing-based estimation has been widely used as the primary means to spatially measure AGB and AGC, while not taking aside the field data for the ground truth data source (Chave et al., 2019; Lu et al., 2016). Remote sensing offers carbon stock mapping at multiple levels, from global scale based on satellite imageries to parcel scale using sensors mounted on unmanned aerial vehicle (UAV). UAV enables the acquisition of a very high-resolution image for large scale biomass estimates at individual tree levels (Lin et al., 2018).

The most common approach to spatially estimate carbon stock and biomass based on optical imagery is by means of vegetation indices since vegetation properties can be enhanced by an image transformation (Kamal et al., 2022; Sudarma et al., 2024). Some studies improved the satellite based-AGC mapping by involving vertical features derived from airborne Light detection and ranging (LiDAR) and synthetic aperture radar because it reflects on the tree height as the main parameter of tree biomass (Narine, et al., 2020; Schlund et al., 2015). Having the advantage of multi-viewing image captures, UAV generates point clouds explaining 3-dimensional structures of objects. However, the resulted 3D model only depicts the canopy surface that can lead to inaccurate tree height extraction despite being terrain corrected, especially in dense vegetation and rough topography (Umarhadi et al., 2023). Compared to a normal red-green-blue (RGB) camera, LiDAR sensor allows a comprehensive measurement of the biophysical parameters of trees owing to its active system that can be mounted on UAVs. LiDAR pulse, where the forest canopy is not too dense, can penetrate through the vegetation canopy to capture the understory and the underlying ground (Fekry et al., 2022), thus allowing it to provide more accurate vertical structure of vegetation.

UAV LiDAR is widely utilized as a standalone tool for AGB and AGC mapping primarily through the derivation of 3D point clouds within the delineated tree crowns (Du et al., 2023; Gleason and Im, 2012; Lefsky et al., 2002). To improve accuracy, some research has adopted a multi-sensor approach by combining LiDAR data with multispectral and hyperspectral sensors (Balestra et al., 2024; Qin et al., 2021a; Su et al., 2023). This fusion of data sources has shown promising results, enhancing the precision of AGB estimates. However, the limited availability of multispectral and hyperspectral sensors often restricts the feasibility of this approach. Alternatively, several studies have explored visible-based vegetation indices and have adapted traditional infrared-reliant indices for use with visible bands only, such as excess green minus excess red (ExG–ExR) and visible normalized difference vegetation index (vNDVI) (Costa et al., 2020; Meyer and Neto, 2008). Therefore, it highlights the potential of standard RGB cameras as complementary tools to LiDAR data in vegetation analysis.

Multi-linear regression has traditionally been used to predict AGB and AGC due to the assumed linear relationships between dependent and independent variables (Qin et al., 2021). However, machine learning (ML)-based regression algorithms (e.g., decision trees, random forests, support vector machines, and artificial neural networks) are gaining attention for their potential to outperform traditional linear regression models (Maesano et al., 2022; Zheng et al., 2022). Thus, it is worthwhile to compare regression models to attain the most optimum results for teak plantation AGC estimation. This study aims to examine multi-linear regression, random forest, support vector regression, and XGBoost regression to estimate tree-level AGC in the clonal teak plantation of Wanagama Forest, Indonesia using a fusion of LiDAR and RGB aerial images.

MATERIALS AND METHODS

The workflow of this study is illustrated in Figure 1, and the explanation is given in the following sections.

Study area

This study took place in Wanagama Forest (UTM Zone 49M 448041–448565 mE and 9127080–9125999 mN) which is located in Yogyakarta, Indonesia (Figure 2). Wanagama Forest is designated as an education forest managed by Universitas Gadjah Mada with a total area of 622.25 ha. Our focus is on the monoculture area of improved teak stands in Compartment 13 covering 5675 ha (Na'iem et al., 2020). The improved clonal teak, planted in 2004, has superior growth characteristics in diameter, height, and stem alignment as compared to unimproved or conventional teak (Hidayati et al., 2016).

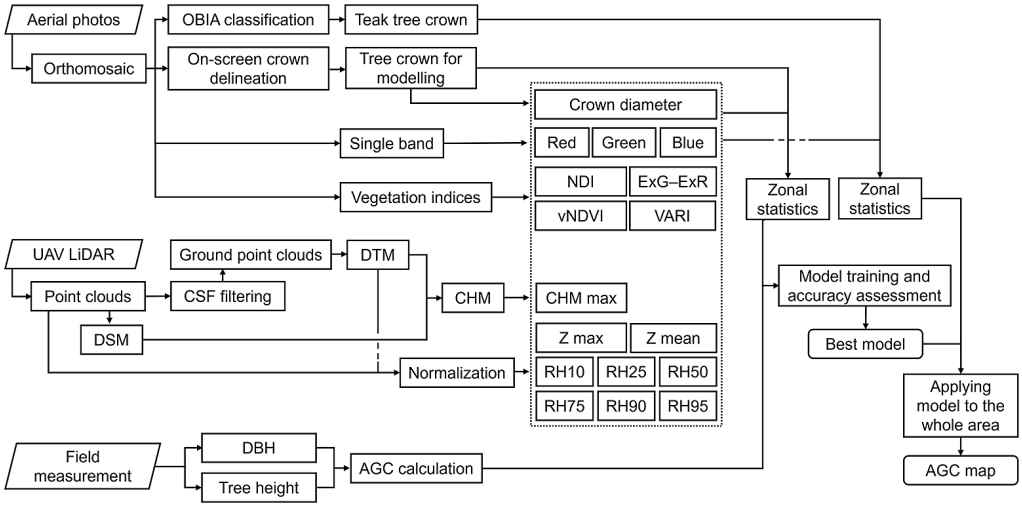

Figure 1. Flowchart of the study

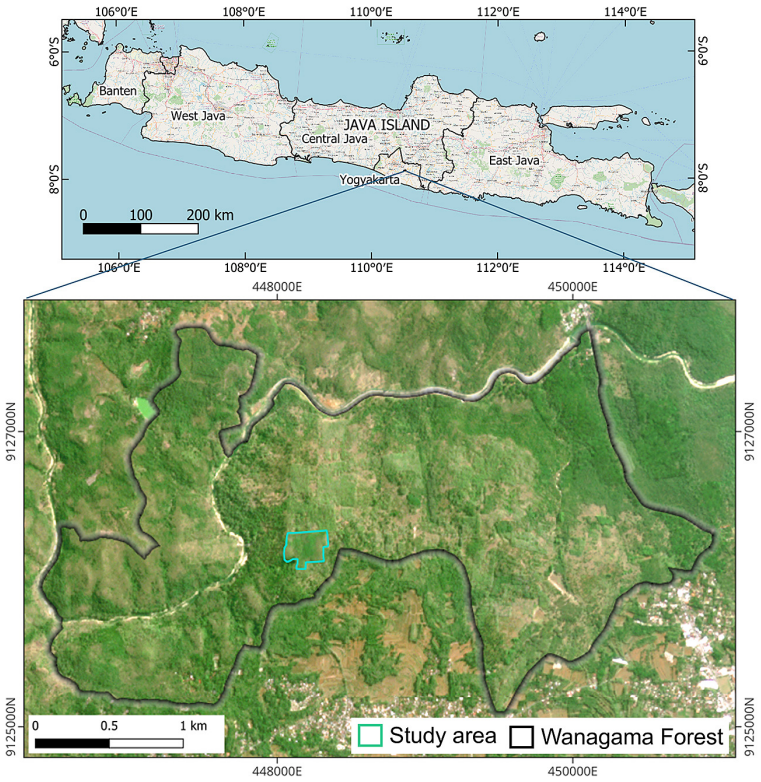

Figure 2. Study area with a base map of monthly Planet image of July 2023

In-situ data collection

The in-situ data collected in the field consists of diameter at breast height (DBH) and tree height. DBH was measured using diameter tape, while tree height was measured by using the 2-points measurement mode of Nikon Forestry Pro II laser rangefinder on every individual tree. We measured the total height of individual trees between the topmost part of the tree and its base (Larjavaara and Muller-Landau, 2013). The tree height survey was conducted in August 2023; however, DBH measurement was taken place in September 2024. As all remote sensing data dated in 2023, we normalized the DBH values based on the annual diameter increment value of 1.4 cm in the same study area (Seta et al., 2021), making both DBH and tree height represent the actual condition in 2023. The total number of trees surveyed is 101.

An allometric equation was used to model teak biomass (Equation 1) referred to a study by Purwanto and Silaban (2011) who built an allometric equation based on the destructive method of teak trees in Karanganyar Regency, about 50 km away from Wanagama forest. The allometric equation is as follows:

$$AGB = 0.0149 \times (D^2H)^{1.0855} \quad (1)$$

where: *AGB* denotes the total of aboveground biomass (kg), *D* denotes diameter at breast height (cm), and *H* denotes tree height (m).

The aboveground carbon stock is the result of multiplication between AGB and a factor of 0.47 according to Indonesian National Standards Number 7724:2011 (Purnamasari et al., 2021).

Aerial imagery

The aerial photographs in this study were acquired using DJI Mavic 2 Pro drones in July 2023. The drones were programmed to capture the study area at 150 m altitude relative to the ground station with the same overlap and sidelap of 80%. The aerial photographs were then processed by means of the Structure from Motion (SfM) algorithm in Agisoft Metashape which has some advantages compared to other processing software (Fraser and Congalton, 2018). The main output is a georeferenced orthomosaic with a resolution of 4 cm, comprising three bands, i.e., red, green, and blue. Vegetation indices were subsequently derived from the visible bands to enhance vegetation features, consisting of normalized difference index (NDI), excess green minus excess red (ExG–ExR), visible atmospherically resistant index (VARI), and visible normalized difference vegetation index (vNDVI). NDI and ExG–ExR can greatly separate the values between vegetation features and the background (Meyer and Neto, 2008; Pérez et al., 2000). vNDVI was developed to replicate NDVI, the most general-purpose vegetation index, for visible aerial imagery (Costa et al., 2020). VARI is capable in overcoming the saturation issue of NDVI, while it can also minimize the atmospheric effect (Gitelson et al., 2002). The equations of the above-mentioned vegetation indices are as follows:

$$NDI = \frac{G-R}{G+R} \quad (2)$$

$$E \times G - E \times R = \\ = (2.0 \times G - R - B) - (1.4 \times R - G) \quad (3)$$

$$vNDVI = \\ 0.5847 \times (R^{-0.1294} \times G^{0.3389} \times B^{-0.3118}) \quad (4)$$

$$VARI = \frac{G-R}{G+R+B} \quad (5)$$

where: *B*, *G*, and *R* respectively denote blue, green, and red bands.

LiDAR data

The LiDAR data acquisition was conducted on 23 August 2023 using GeoSun GS-100M device mounted on a customized quadcopter drone. GS-100M occupied Livox Avia laser sensor, working on 905 nm wavelength. The LiDAR survey covered the study area with a flight altitude of 130 m following the terrain condition and a speed of 6 m/s, thus resulting in a density of 190 points/m^2.

The Z dimension of point clouds derived from the LiDAR data describes the elevation of all objects captured, thus we classified LiDAR point clouds through 2 stages consisting of ground points classification and non-ground points classification in Global Mapper software. Multiscale curvature classification (MCC) method was implemented to distinguish between ground and non-ground point clouds by applying a curvature threshold from the interpolated thin-plate spline (Evans and Hudak, 2007). Prior to MCC implementation, a morphological filtering procedure was applied to identify potential non-ground points using parameters as follows: maximum

height delta = 50 m; expected terrain slope = 7.5°; and maximum building width = 100 m. The parameters used in MCC method included grid bin size and curvature which are set to 2 m and 0.5 m, respectively. Afterward, non-ground points classification was applied to automatically assign vegetation class to the point clouds. Noise elimination was also performed to discard the remnants of non-classified points.

All point clouds regardless of classes were rasterized to generate digital surface model (DSM) by taking the maximum Z values, while the ground class points were rasterized to produce digital terrain model (DTM) at 25 cm resolution. Canopy height model (CHM) was resulted by subtracting DTM from DSM. To minimize the extreme outliers, mean filtering (3 × 3 window size) was performed on the CHM. Normalization was also carried out by subtracting the Z values of point clouds with DTM. The percentile of relative height (RH) was quantified for each tree crown.

Object-based classification

Object-based classification routine is intended to generate teak crowns as the main object in this study while eliminating other objects (Figure 1). Image segmentation was applied with the approach of the multi-resolution segmentation method in eCognition software using the input of orthomosaic (Baatz and Schäpe, 2000). The parameters used are as follows: scale = 40; shape = 0.5; and compactness = 0.5 with the same weight between bands (blue, green, and red). We applied a binary classification, i.e., tree crown and non-tree crown, as the study area consists of a monoculture teak plantation. Support vector machine (SVM) algorithm with a linear kernel was applied based on the average value of respective bands from the segmentation result (Zylshal et al., 2016). The crown diameter was calculated based on the area of segmented as shown in Equation 6.

$$Crown\ diameter = \sqrt{\frac{Area}{\pi}} \times 2 \qquad (6)$$

Table 1. Independent variables used in carbon stock modelling

Sensor	Variable	Processing
UAV LiDAR	CHM max	Rasterized LiDAR points
	Z max	Relative height of LiDAR points
	Z mean	
	RH10	
	RH25	
	RH50	
	RH75	
	RH90	
	RH95	
Aerial image	Crown diameter	Multi-resolution segmentation
	Red mean	Spectral value
	Green mean	
	Blue mean	
	ExG–ExR mean	Vegetation index
	NDI mean	
	VARI mean	
	vNDVI mean	

Carbon stock modelling

Tree-level carbon stock modeling used the dependent variable of allometric-based carbon stock values as response of the independent variables consisting of crown diameter, spectral values (i.e., red, green, blue, NDI, ExG-ExR, vNDVI, and VARI), and LiDAR features (i.e., CHM max, Z max, Z mean, RH10, RH25, RH50, RH75, RH90, and RH95) as summarized in Table 1. Zonal statistics were executed to calculate the mean values of spectral values and respective LiDAR features for each corresponding field-surveyed tree crown. All variables were transformed to natural logarithm to handle the non-linear relationship and the skewed data distribution. The data were split into training (70%) and testing (30%).

Four regression models were used in this study, comprising multiple linear regression (MLR), support vector regression (SVR), random forest regression (RFR), and extreme gradient boost regression (XGBR) (Figure 3). MLR estimates the linear relationship between a response variable and more than one independent variable by the approach of ordinary least squares (Tranmer et al., 2020). SVR is a regression-based support vector machine that estimates a continuous-valued multivariate function by constructing a hyperplane and a tube around it (Awad and Khanna, 2015). The objective is to minimize the width of the tube while ensuring that most data points lie within it, formulated as a convex optimization problem. RFR and XGBR use an ensemble learning strategy to solve regression problems. RFR performs by creating multiple decision trees and aggregation on the trees in the forest that were bootstrapped (Segal, 2004). XGBR is described as a scalable end-to-end

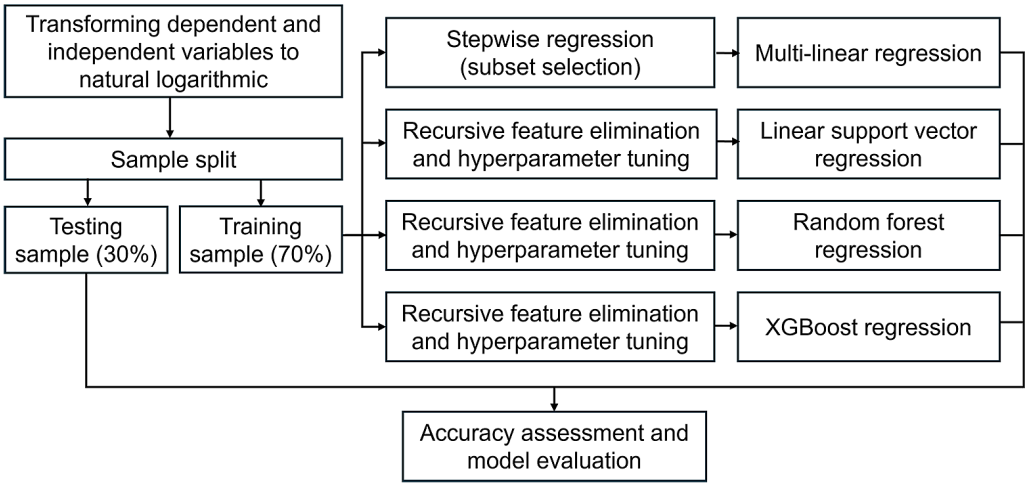

Figure 3. Detailed flowchart of model training, comprising four regression models that are examined

tree boosting method that builds models incrementally by focusing on errors from previous models (Chen and Guestrin, 2016). It entails regularization to prevent overfitting and is highly scalable, making it suitable for large and sparse datasets.

Strong relationships were found for spectral and LiDAR feature variables (Figure 4). However, it may lead to data redundancy and multicollinearity that affect prediction performance (Han et al., 2019). Variable selection methods, such as stepwise regression and recursive feature elimination (RFE), are capable of improving regression performance by finding the optimal subset of predictor variables. The backward method of stepwise regression was implemented for MLR. This method iteratively processes regression models starting by including all candidate variables and then removing one least significant variable until all involved variables are significant (Ruengvirayudh and Brooks, 2016). For predictions using SVR, RFR, and XGBR, RFE was applied together with hyperparameter tuning using GridSearchCV.

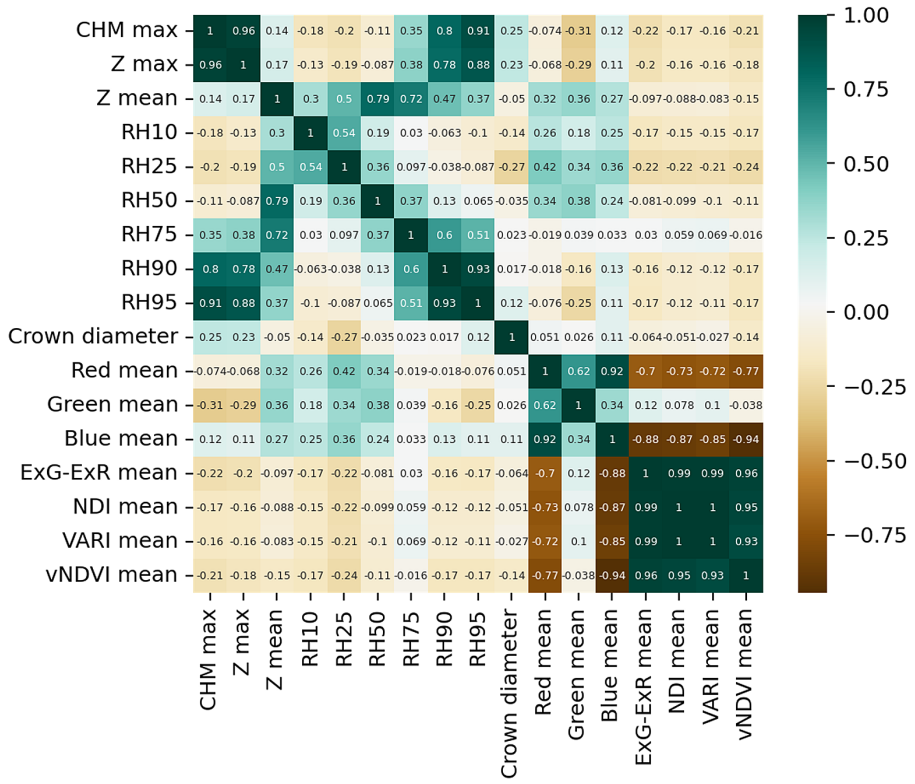

Figure 4. Correlation matrix between independent variables

In principle, RFE is similar to stepwise regression, working by repeatedly training the model and selecting predictors that contribute most to model's performance. The values of hyperparameters tuned are listed in Table 2. Each regression method has a different approach to calculate feature importance values. We used the t-value resulting from stepwise MLR regression, where the larger value indicates the more important variable. The importance values in SVR were interpreted from the magnitude of coefficients, converted into absolute values. The mean decrease in impurity and the average gain of splits are used for identifying the feature importance of RFR and XGBR, respectively.

Accuracy assessment

An accuracy assessment was carried out to evaluate the performance of the models. As the data were split into training and testing, the number of samples for accuracy assessment was 30 data. Two metrics were quantified, i.e., mean absolute error (MAE) and weighted average percentage error (WAPE).

$$MAE = \sum_{i=1}^{n} |y - \hat{y}| \quad (7)$$

$$WAPE = \frac{\sum_{i=1}^{n} |y - \hat{y}|}{\sum_{i=1}^{n} |y|} \quad (8)$$

where: y, $\hat{y}$, and $\bar{y}$ denote observed, predicted, and mean of observed values, respectively.

Table 2. Configurations of hyperparameters for machine learning methods

Algorithm	Hyperparameters tuned	Hyperparameter configurations
SVR	gamma	scale, 0.01, 0.1, 1
	C	1, 10, 100
RFR	n_estimators	10, 50, 100, 200
	max_features	sqrt, log2, None
XGB R	n_estimators	10, 50, 100, 200
	learning_rate	0.01, 0.1, 0.2

RESULTS AND DISCUSSION

Field observation

We collected 101 ground truth tree data during the fieldwork. The samples were located in three clusters of a plot about 20 × 20 m, representing different terrain conditions and vegetation density. The measured DBH showed a normal distribution, ranging from 10.10 to 19.70 cm (mean = 14.80 cm) (Figure 5). On the other hand, the tree height is skewed as only 3 trees are taller than 35 m with a relatively DBH above average. When calculated to AGC, the data skewness is much more apparent (Figure 5b). The average AGC value is 103.01 kg, while AGC of the 3 tallest trees are 771.89, 489.79, and 428.04 kg, respectively. Thus, the AGC values were transformed

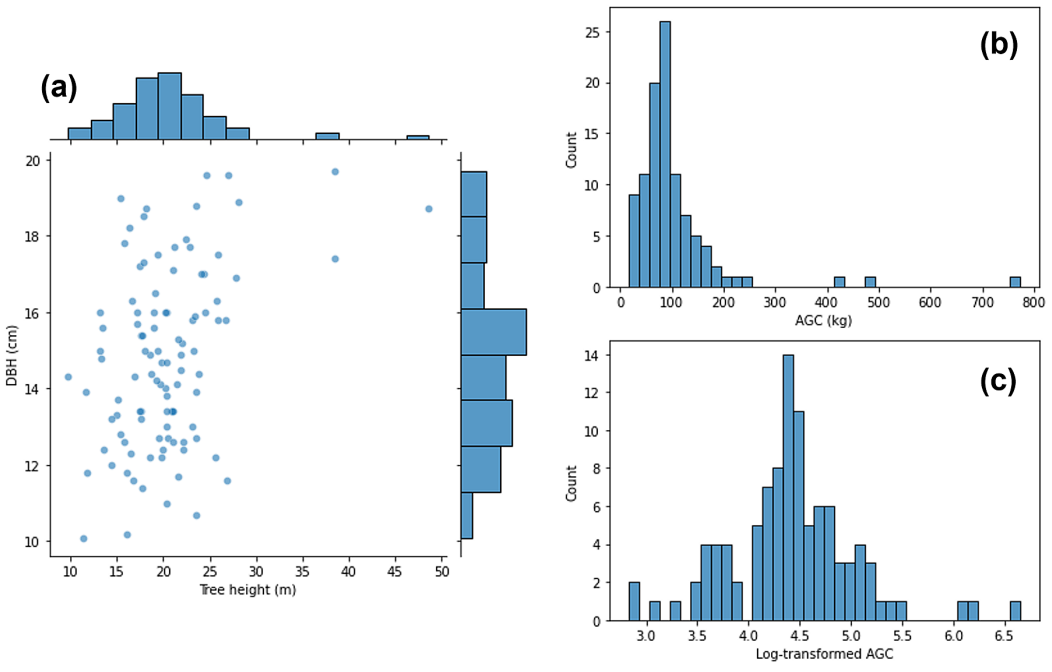

Figure 5. Data distribution of (a) ground truth data consisting of tree height and diameter at breast height, (b) calculated aboveground carbon stock, and (c) log-transformed of aboveground carbon stock

into natural logarithmic as it can solve skewed data to be appropriately used for further statistical analysis (Benoit, 2011).

Teak crown segmentation

Figure 6 shows the results of multi-resolution segmentation, separating crowns for the corresponding tree. Since the study area is a monoculture plantation forest, there are no other tree species than teak, however, shrubs exist but were classified as non-crown class. There are a total of 7062 crowns segmented with an average diameter of 2764 m. However, 22 crowns are found invalid having no vegetation point clouds.

Our results revealed a density of 1244 trees/ha, showing an overestimate compared to the tree inventory in the same study area, i.e., 665 trees/ha (Afafi et al., 2022). Object-based segmentation segregates crowns, nevertheless, teak trees may have more than one crown visible from above, leading to overestimation. In this study, we have explored the well-known watershed segmentation based on point clouds-based segmentation approaches (Heurich, 2008). However, due to the leaf-off season during the data acquisition, it showed poor segmentations. As LiDAR has somehow limited penetration in dense canopy, data acquisition during leaf-off season on the other hand has advantages in capturing more ground surface points. Chen et al. (2022) reported that the combination of leaf-on and leaf-off data acquisition increased the accuracy of individual tree segmentation, particularly for small trees. Thus, future studies should consider conducting LiDAR acquisition in the leaf-on season as well.

LiDAR processing

CHM was generated by subtracting the rasterized maximum value of point clouds with the rasterized ground surface point clouds. After being calculated using zonal statistics for each crown segment by taking the maximum height value, the average CHM max is 17.57 m with the highest value of 33.20 m. LiDAR data also revealed the rugged condition of topography in the study area. As shown in Figure 7, LiDAR pulse can measure the ground surface conditions well. The lack of point clouds between the canopy and the ground indicates that, although it describes better than camera-based point clouds, it is not as good at capturing the features of lower canopy structures (Cao et al., 2019). Combining UAV LiDAR with terrestrial laser scanner can produce more detailed tree vertical structures as demonstrated by Lian et al. (2022). Nevertheless, it can only be effectively applied in a small area, in contrast, covering a vast area would need a lot of intensive work.

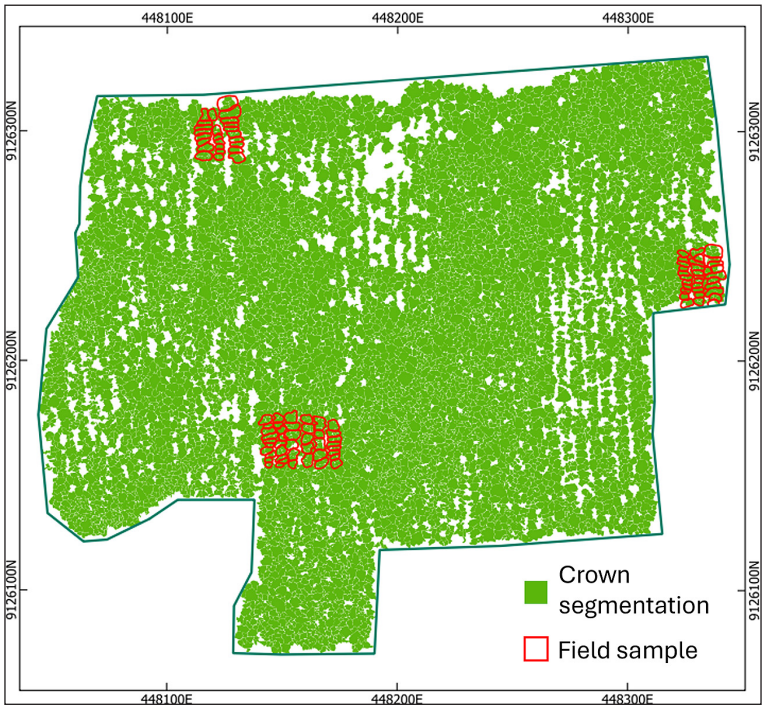

Figure 6. Crown segmentation results and location of field observations

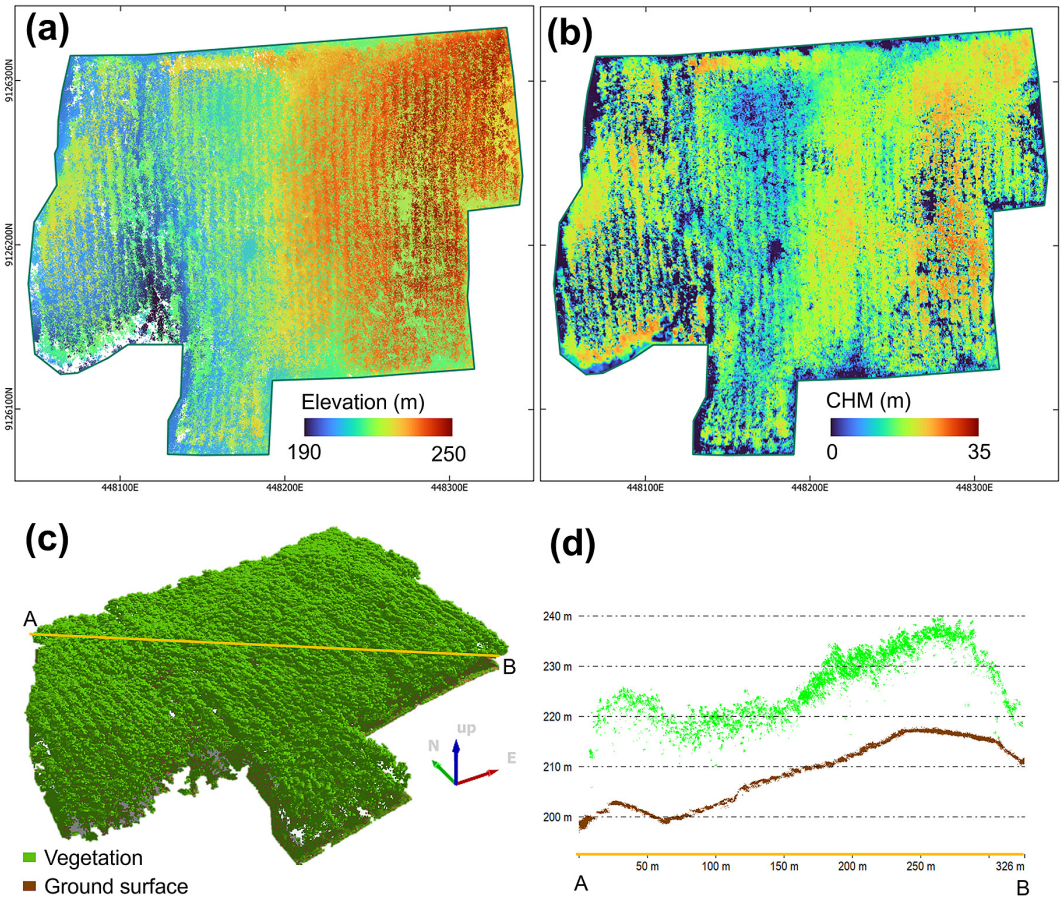

Figure 7. (a) LiDAR point clouds color-coded according to Z-values, (b) generated canopy height model, (c) 3D view of point clouds classified into vegetation and ground surface classes, (d) profile showing the vertical cross-section of the study area

Feature selection and model training

As mentioned in Methods, feature selection strategies were implemented for model training which used stepwise regression for MLR and RFE for RFR, SVR, and XGBR methods. All algorithms exhibit different selected variables for further AGC training. Figure 8 showed that 11 out of 17 variables were significant as inputs for linear regression based on backward stepwise regression. The number of selected variables was the same for XGBR (11 variables), while RFR and SVR included 5 and 9 variables, respectively.

Crown diameter was the most influential feature among the others, except for XGBR which ranked number three. Only two features, i.e., crown diameter and Z max, were involved in all methods and ranked relatively higher than other features. Meanwhile, vegetation indices were among the top influential variables, although, it is very between VARI, ExG-ExR, and NDI. As seen in Figure 4, correlation coefficients exceeding 0.9 were observed among vegetation indices which indicates strong multicollinearity. Similarly, strong relationships were also found between CHM max, Z max, and RH95. Multicollinearity affects model training in regression analysis, thus this feature optimization is important (Qin et al., 2021; Wang et al., 2024). Stepwise regression and RFE prioritize features that most contribute to predictive performance, therefore reducing redundancy and multicollinearity. During the feature elimination process, highly correlated variables were evaluated and often substituted each other by keeping one while removing the others to avoid redundancy.

It is noteworthy that crown diameter, height metrics, and spectral features were all important in AGC modeling, especially crown diameter as shown by the feature importance and it has a high correlation with DBH. The strong relationship between DBH and delineated crown diameter from UAV imaging has also been reported by previous studies (Fu et al., 2024; Popescu, 2007). Qin et

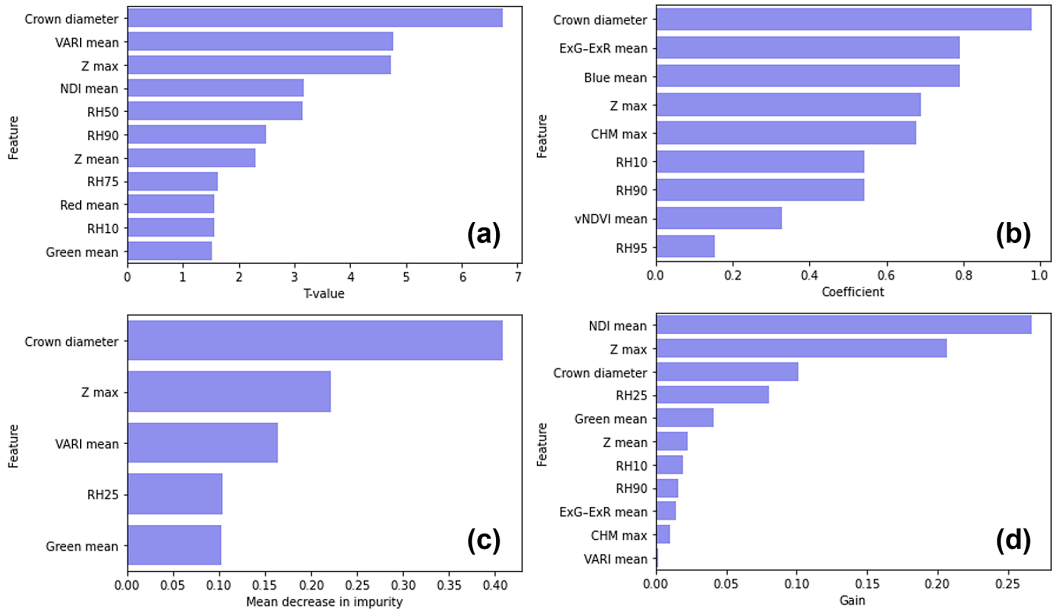

Figure 8. Feature importance values of model training using (a) stepwise-linear regression, (b) random forest regression, (c) support vector regression, and (d) XGBoost regression

al. (2021) reported that vertical structures derived from LiDAR and hyperspectral-based features on the other hand are more significant compared to crown diameter. This might be due to the tree species heterogeneity in their study area, while we only focused on the homogenous teak plantation.

This study demonstrated the fusion of an RGB camera and UAV LiDAR. Spectral features with the inclusion of red-edge and infrared wavelengths and narrower spectrums derived from hyperspectral images have the ability to describe carbon coefficient across tree species (Laurin et al., 2014; Qin et al., 2021). Therefore, as multispectral and hyperspectral cameras will become more affordable in the future, we recommend using them as excellent supplemental tools for LiDAR data in AGC mapping, especially when utilized in mixed-species forests. Besides feature elimination, hyperparameters were also tuned within the GridSearchCV routine of RFR, SVR, and XGBR. XGBR managed to have an almost perfect coefficient of determination ($R^2 = 0.999$) with n_estimators and learning_rate of 100 and 0.2, respectively. It is followed by RFR with an R^2 of 0.954 (n_estimators = 200; max_features = None) and MLR with an R^2 of 0.727. SVR ranked the last compared to other algorithms ($R^2 = 0.611$; gamma = scale; C = 100). However, the coefficient of determination in model training does not fully represent the model's accuracy since it has not been evaluated using the unseen data.

Accuracy assessment and model evaluation

The models were evaluated using 30 testing samples. Figure 9 depicts the scatterplots between predicted and actual AGC values with the accuracy metrics. All models managed to have an MAE between 28 and 37 kg. MLR demonstrated the most accurate among the others (MAE = 28.35 kg; WAPE = 34.13%), followed by RFR (MAE = 30.12 kg; WAPE = 36.27%) and XGBR (MAE = 33.27 kg; WAPE = 40.06%). Meanwhile, SVR was the least accurate with an MAE of 36.39 kg and 43.82%. Our results showed that linear regression outperformed other methods. XGBR, having an R^2 of 0.999 in the model training, possessed a lower accuracy than MLR and RFR. It is mainly due to the small size of training samples (n = 61) leading to overfitting in non-parametric ML algorithms. ML models such as RFR, XGBR, and SVR generally require larger datasets in order to learn complex patterns and generalize well (Zhou et al., 2017). On the other hand, MLR can perform better on smaller datasets, even more when the linear or near-linear relationship exists between AGC and the independent variables involved. Wang et al. (2024) also reported that MLR with feature optimization can accurately estimate AGB using a small dataset (n = 53). While advanced machine learning techniques were also employed in this study, our MLR model showed to be more practical for estimating AGC with limited available training data.

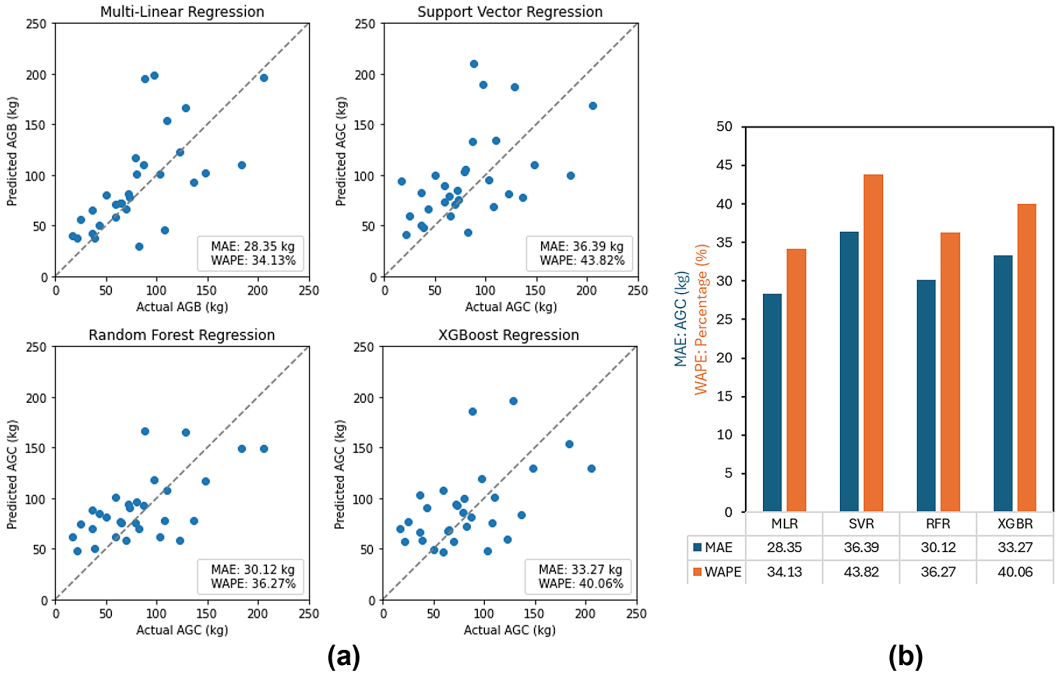

Figure 9. (a) Scatterplots showing the predicted versus actual above-ground carbon values, (b) plots of accuracy assessment results

Spatial distribution of AGC stock

Based on the accuracy assessment, MLR managed to produce the most accurate AGC estimates compared to other methods. We then generated the final AGC stock map (Figure 10) using the MLR equation, showing the total AGC stock of 364.04 Mg with an average of 51.67 kg per individual tree. With respect to the area (5675 ha), the resulting AGC stock is 64.15 Mg·ha^{-1}. This is a slightly higher carbon stock compared to the field inventory in the same stand conducted by Afafi et al. (2022) in 2021 who reported the AGC stock was 59.98 Mg·ha^{-1}, indicating an annual growth of 2.08 Mg·ha^{-1}. The growth is slightly lower than clonal teak plantation in Kendal, Central Java, Indonesia with carbon stock of 2.68 Mg·ha^{-1} (Wirabuana et al., 2022).

Very high AGC stocks were observed having values above 100 kg in the western and

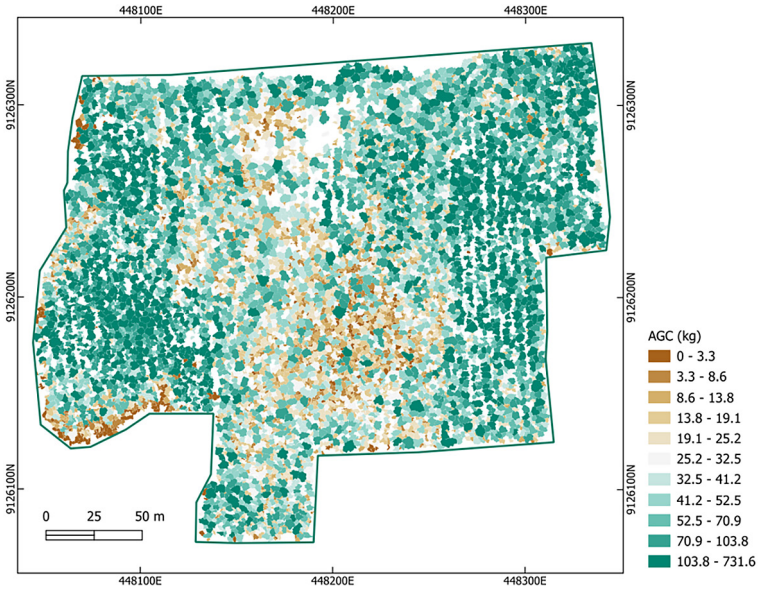

Figure 10. Aboveground carbon stock map of the study area using multiple linear regression

eastern side of the study area. The western area has a topographical condition as a valley which contains more deposit of soil nutrients resulted from erosion in the upper area. The availability of nutrients affects the growth performance of teak trees by physiological cycle. Additionally, the eastern side has a relatively flat topography, therefore the erosion is low, and soil nutrients are available to support tree growth. Intensive silvicultural treatments such as fertilizing, thinning, and pruning are applied to formulate best management practices which might also result in AGC variation.

CONCLUSIONS

This study examines the fusion of nine LiDAR features and aerial image's crown diameter, RGB spectral values, and visible vegetation indices by evaluating multi-linear regression, random forest, support vector regression, and XGBoost regression to estimate tree-level AGC in the clonal teak plantation of Wanagama Forest, Indonesia. Our results showed that crown diameter, Zmax, and one of the vegetation indices are three most influential features among other features in estimating the AGC of individual tree in all models.

However, object-based segmentation using a multi-resolution approach that was employed to generate teak crowns in this study, resulted in an overestimation of the number of trees compared to field inventory. This might be attributed to some teak trees having more than one crown visible from above. Zmax (maximum height of LiDAR point cloud) has a strong correlation with, but more significant than, maximum Canopy Height Model. The four visible vegetation indices indicate multicollinearity. Multiple linear regression with eleven independent variables outperformed other models by showing value of determination coefficient of 0.73 but and most accurate in predicting individual tree's AGC with MAE = 28.35 kg; WAPE = 34.13%. The stand AGC based on MLR shows 64.15 $Mg \cdot ha^{-1}$. We demonstrated the accurate tree-level AGC estimation by integrating spectral and vertical features. Our methods can be potentially improved in future studies by considering both leaf-on and leaf-off seasons, also involving multispectral and hyperspectral sensors that have not been implemented for teak plantations.

Acknowledgements

The authors would like to thank Universitas Gadjah Mada for funding this study under the Academic Excellence Grant Number 6529/UN1.P1/PT.01.03/2024 and the Young Lecturer Research Grant Number 5985/UN1.P.II/Dit-Lit/PT.01.03/2023. We are thankful to the assistants of Spatial Information System and Forest Mapping Laboratory, Faculty of Forestry, Universitas Gadjah Mada for their help in data acquisition and pre-processing.

REFERENCES

1. Afafi, S. N., K. I. Supartha, H. Fatmawati, N. H. E. Sari, F. Y. Al-Husna, F. D. Himawan, M. Aulia, M. B. Ardiansyah, and B. Mulyana. (2022). Carbon Storage of Superior Clonal Teak Stand in Special Purpose Forest Area of Wanagama, Special Region of Yogyakarta. Jurnal Galam 2(2), 66–76.
2. Awad, Mariette, and Rahul Khanna. (2015). Support Vector Regression. 67–80 in Efficient Learning Machines. Berkeley, CA: Apress.
3. Baatz, M., and A. Schäpe. (2000). Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image Segmentation. 12–23 in Angewandte Geographische Informations-Verarbeitung XII, edited by J. Strobl, T. Blaschke, and G. Griesebner. Wichmann Verlag, Heidelberg.
4. Balestra, M., Marselis, S., Sankey, T.T., Cabo, C., Liang, X., Mokroš, M., Peng, X., Singh, A., Stereńczak, K., Vega, C., Vincent, G. and Hollaus, M. (2024). LiDAR data fusion to improve forest attribute estimates: A review. *Current Forestry Reports 10*(4), 281–97. https://doi.org/10.1007/s40725-024-00223-7
5. Behera, M., and Mohapatra, N.P. (2015). Biomass accumulation and carbon stocks in 13 different clones of teak (*Tectona Grandis Linn*. F.) in Odisha, India. *Current World Environment 10*(3), 1011–16. https://doi.org/10.12944/CWE.10.3.33
6. Benoit, K. (2011). Linear Regression Models with Logarithmic Transformations.
7. Cao, L., Liu, H., Fu, X., Zhang, Z., Shen, X. and Ruan, H. (2019). Comparison of UAV LiDAR and digital aerial photogrammetry point clouds for estimating forest structural attributes in subtropical planted forests. *Forests 10*(2), 145. https://doi.org/10.3390/f10020145
8. Chave, J., Davies, S.J., Phillips, O.L., Lewis, S.L., Sist, P., Schepaschenko, D., Armston, J., Baker, T.R., Coomes, D., Disney, M., Duncanson, L., Hérault, B., Labrière, N., Meyer, V., Réjou-Méchain, M.,

Scipal, K. and Saatchi, S. (2019). Ground data are essential for biomass remote sensing missions. *Surveys in Geophysics 40*(4), 863–80. https://doi.org/10.1007/s10712-019-09528-w

9. Chayaporn, P., Sasaki, N., Venkatappa, M. and Abe, I. 2021. Assessment of the overall carbon storage in a teak plantation in Kanchanaburi Province, Thailand – Implications for carbon-based incentives. *Cleaner Environmental Systems 2*, 100023. https://doi.org/10.1016/j.cesys.2021.100023
10. Chen, Q., Gao, T., Zhu, J., Wu, F., Li, X., Lu, D. and Yu, F. (2022). Individual tree segmentation and tree height estimation using leaf-off and leaf-on UAV-lidar data in dense deciduous forests. *Remote Sensing 14*(12), 2787. https://doi.org/10.3390/rs14122787
11. Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. 785–94 in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA: ACM.
12. Costa, L., Nunes, L. and Ampatzidis, Y. (2020). A new visible band index (vNDVI) for estimating NDVI values on RGB images utilizing genetic algorithms. *Computers and Electronics in Agriculture 172*, 105334. https://doi.org/10.1016/j.compag.2020.105334
13. Darmawan, W., Nandika, D., Sari, R.K., Sitompul, A., Rahayu, I., and Gardner, D. (2015). Juvenile and mature wood characteristics of short and long rotation teak in Java. *IAWA Journal 36*(4), 428–42. https://doi.org/10.1163/22941932-20150112
14. Du, L., Pang, Y., Wang, Q., Huang, C., Bai, Y., Chen, D., Lu, W. and Kong, D. (2023). A LiDAR Biomass Index-Based Approach for Tree- and Plot-Level Biomass Mapping over Forest Farms Using 3D Point Clouds. *Remote Sensing of Environment 290*, 113543. https://doi.org/10.1016/j.rse.2023.113543
15. Evans, J.S., and Hudak, A.T. (2007). A multiscale curvature algorithm for classifying discrete return LiDAR in forested environments. *IEEE Transactions on Geoscience and Remote Sensing 45*(4), 1029–38. https://doi.org/10.1109/TGRS.2006.890412
16. Fekry, R., Yao, W., Cao, L. and Shen, X. (2022). Ground-Based/UAV-LiDAR data fusion for quantitative structure modeling and tree parameter retrieval in subtropical planted forest. Forest *Ecosystems 9*, 100065. https://doi.org/10.1016/j.fecs.2022.100065
17. Fraser, B., and Congalton, R. (2018). Issues in unmanned aerial systems (UAS) data collection of complex forest environments. *Remote Sensing 10*(6), 908. https://doi.org/10.3390/rs10060908
18. Fu, H., Zhao, H., Jiang, J., Zhang, Y., Liu, Ge., Xiao, W., Du, S., Guo, W. and Liu, X. (2024). Automatic detection tree crown and height using mask R-CNN based on unmanned aerial vehicles images for biomass mapping. *Forest Ecology and Management 555*, 121712. https://doi.org/10.1016/j.foreco.2024.121712
19. Gitelson, Anatoly A., Yoram J. Kaufman, Robert Stark, and Don Rundquist. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment 80*(1), 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9
20. Gleason, C.J., and Im, J. (2012). Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment 125*, 80–91. https://doi.org/10.1016/j.rse.2012.07.006
21. Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., Li, Z. and Yang. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods 15*(1), 10. https://doi.org/10.1186/s13007-019-0394-z
22. Heurich, M. (2008). Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the bavarian forest national park. *Forest Ecology and Management 255*(7), 2416–33. https://doi.org/10.1016/j.foreco.2008.01.022
23. Hidayati, F., Fajrin, I. T., Ridho, M. R., Nugroho, W. D., Marsoem, S. N. and Na'iem, M. (2016). Sifat fisika dan mekanika kayu jati unggul 'Mega' dan kayu jati konvensional yang ditanam di hutan Pendidikan Wanagama, Gunungkidul, Yogyakarta. *Jurnal Ilmu Kehutanan 10*(2).
24. Kamal, M., Hidayatullah, M.F., Mahyatar, P. and Ridha, S.M. (2022). Estimation of aboveground mangrove carbon stocks from WorldView-2 imagery based on generic and species-specific allometric equations. *Remote Sensing Applications: Society and Environment 26*: 100748. https://doi.org/10.1016/j.rsase.2022.100748
25. Larjavaara, M., and Muller-Landau, H.C. (2013). Measuring tree height: A quantitative comparison of two common field methods in a moist tropical forest edited by J. Metcalf. Methods in *Ecology and Evolution 4*(9), 793–801. https://doi.org/10.1111/2041-210X.12071
26. Laurin, G.V., Chen, Q., Lindsell, J.A., Coomes, D.A., Del Frate, F., Guerriero, L., Pirotti, F. and Valentini, R. (2014). Above ground biomass estimation in an african tropical forest with lidar and hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing 89*, 49–58. https://doi.org/10.1016/j.isprsjprs.2014.01.001
27. Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G.,Acker,S.A.andGower,S.T.(2002).Lidarremote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography 11*(5), 393–99. https://doi.org/10.1046/j.1466-822x.2002.00303.x
28. Lian, X., Zhang, H., Xiao, W., Lei, Y., Ge, L., Qin,

K., He, Y., Dong, Q., Li, L., Han, Y., Fan, H., Li, Y., Shi, L. and Chang, J. (2022). biomass calculations of individual trees based on unmanned aerial vehicle multispectral imagery and laser scanning combined with terrestrial laser scanning in complex stands. *Remote Sensing 14*(19), 4715. https://doi.org/10.3390/rs14194715

29. Lin, J., Wang, M., Ma, M., and Lin, Y. (2018). Aboveground tree biomass estimation of sparse subalpine coniferous forest with UAV oblique photography. *Remote Sensing 10*(11), 1849. https://doi.org/10.3390/rs10111849
30. 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*(1), 63–105. https://doi.org/10.1080/17538947.2014.990526
31. Maesano, M., Santopuoli, G., Moresi, F., Matteucci, G., Lasserre, B. and Mugnozza G.S. (2022). Above ground biomass estimation from UAV high resolution RGB images and LiDAR data in a pine forest in southern Italy. *iForest - Biogeosciences and Forestry 15*(6), 451–57. https://doi.org/10.3832/ifor3781-015
32. Margono, B.A., Potapov, P.V., Turubanova, S., Stolle, F. and Hansen, M.C. (2014). Primary forest cover loss in Indonesia over 2000–2012. *Nature Climate Change 4*(8), 730–35. https://doi.org/10.1038/nclimate2277
33. Meyer, G.E., and Neto, J.C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture 63*(2), 282–93. https://doi.org/10.1016/j.compag.2008.03.009
34. Mitchard, E.T.A. (2018). The tropical forest carbon cycle and climate change. *Nature 559*(7715), 527–34. https://doi.org/10.1038/s41586-018-0300-2
35. Na'iem, M., Rudian, P.A., Hasibuan, S.M., Idhom, A.M., Mustaqim, A., Sutriyati, and CahyonoM.F. (2020). Wanagama: kisah terciptanya hutan pendidikan, konservasi dan kesejahteraan sosial ekonomi bagi rakyat sekitar. Yogyakarta, Indonesia: Samudra Biru.
36. Narine, Lana L., Sorin C. Popescu, and Lonesome Malambo. (2020). Using ICESat-2 to Estimate and Map Forest Aboveground Biomass: A First Example. *Remote Sensing 12*(11), 1824. https://doi.org/10.3390/rs12111824
37. Nurbaya, S. (2023). *Folu net sink: Indonesia's Climate Actions Towards 2030*. Jakarta, Indonesia: The Ministry of Environment and Forestry of the Republic of Indonesia.
38. Pérez, A.J., López, F., Benlloch, J.V. and Christensen, S. (2000). Colour and Shape Analysis Techniques for Weed Detection in Cereal Fields. Computers and Electronics in Agriculture 25(3):197–212. https://doi.org/10.1016/S0168-1699(99)00068-X
39. Popescu, Sorin C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy 31*(9), 646–55. https://doi.org/10.1016/j.biombioe.2007.06.022
40. Potapov, P., Hansen, M.C., Pickens, A., Hernandez-Serna, A., Tyukavina, A., Turubanova, S., Zalles, V., Li, X., Khan, A., Stolle, F., Harris, N., Song, S.-P., Baggett, A., Kommareddy, I. and Kommareddy, A. (2022). The global 2000–2020 land cover and land use change dataset derived from the landsat archive: First results. *Frontiers in Remote Sensing 3*, 856903. https://doi.org/10.3389/frsen.2022.856903
41. Purnamasari, E., Kamal, M. and Wicaksono, P. (2021). Comparison of vegetation indices for estimating above-ground mangrove carbon stocks using planet scope image. *Regional Studies in Marine Science 44*, 101730. https://doi.org/10.1016/j.rsma.2021.101730
42. Purwanto, R.H., and Silaban, M. (2011). Inventore biomasa dan karbon jenis jati (*Tectona Grandis* L.f.) di hutan rakyat Desa Jatimulyo, Karanganyar. *Jurnal Ilmu Kehutanan 5*(1), 40. https://doi.org/10.22146/jik.581
43. Qin, H., Zhou, W., Yao, Y. and Wang, W. (2021). Estimating aboveground carbon stock at the scale of individual trees in subtropical forests using UAV LiDAR and hyperspectral data. *Remote Sensing 13*(24), 4969. https://doi.org/10.3390/rs13244969
44. Ruengvirayudh, P., and Brooks, G.P. (2016). Comparing stepwise regression models to the best-subsets models, or, the art of stepwise. *General Linear Model Journal 42*(1), 1–14.
45. Sarre, A., ed. (2020). *Global forest resources assessment*, 2020: Main Report. Rome: Food and Agriculture Organization of the United Nations.
46. Schlund, M., Von Poncet, F., Kuntz, S., Schmullius, C. and Hoekman, D.H. (2015). TanDEM-X data for aboveground biomass retrieval in a tropical peat swamp forest. *Remote Sensing of Environment 158*, 255–66. https://doi.org/10.1016/j.rse.2014.11.016
47. Segal, M.R. (2004). Machine learning benchmarks and random forest regression. San Francisco: University of California.
48. Seta, G.W., Widiyatno, Hidayati, F., and Na'iem, M. (2021). Impact of thinning and pruning on tree growth, stress wave velocity, and pilodyn penetration response of clonal teak (*Tectona Grandis*) plantation. *Forest Science and Technology 17*(2), 57–66. https://doi.org/10.1080/21580103.2021.1911865
49. Su, R., Du, W., Ying, H., Shan, Y. and Liu, Y. (2023). Estimation of aboveground carbon stocks in forests based on LiDAR and multispectral images: A case study of duraer coniferous forests. *Forests 14*(5), 992. https://doi.org/10.3390/f14050992

50. Sudarma, I. Made, Moh Saifulloh, I. Wayan Diara, and Abd. Rahman As-syakur. (2024). Carbon stocks dynamics of urban green space ecosystems using time-series vegetation indices. *Ecological Engineering & Environmental Technology 25*(9), 147–62. https://doi.org/10.12912/27197050/190515

51. Tranmer, M., Murphy, J., Elliot, M. and Pampaka, M. (2020). *Multiple Linear Regression*. 2nd Edition.

52. Tsujino, R., Yumoto, T., Kitamura, S., Djamaluddin, I. and Darnaedi, D. (2016). History of forest loss and degradation in Indonesia. *Land Use Policy 57*, 335–47. https://doi.org/10.1016/j.landusepol.2016.05.034

53. Umarhadi, D.A., Senawi, Wardhana, W., Soraya, E., Jihad, A.N. and Ardiansyah, F. (2023). Can iPhone/iPad LiDAR data improve canopy height model derived from UAV? edited by M. Z. Abidin, R. Shamsudin, P. Tongdeenok, S. T. Sobue, A. Young, and A. Hoque. *BIO Web of Conferences* 80, 03003. https://doi.org/10.1051/bioconf/20238003003

54. Wang, L., Ju, Y., Ji, Y., Marino, A., Zhang, W. and Jing, Q. (2024). Estimation of forest aboveground biomass in the study area of greater khingan ecological station with integration of airborne LiDAR, Landsat 8 OLI, and hyperspectral remote sensing data. *Forests 15*(11), 1861. https://doi.org/10.3390/f15111861

55. Wirabuana, P.Y.A.P., Hendrati, R.L., Baskorowati, L., Susanto, M., Mashudi, Sulistiadi, H.B.S., Setiadi, D., Sumardi, and Alam, S. (2022a). Growth performance, biomass accumulation, and energy production in age series of clonal teak plantation. *Forest Science and Technology 18*(2), 67–75. https://doi.org/10.1080/21580103.2022.2063952

56. Zheng, C., Abd-Elrahman, A., Vance Whitaker, V. and Dalid, C. (2022). Prediction of strawberry dry biomass from UAV multispectral imagery using multiple machine learning methods. *Remote Sensing 14*(18), 4511. https://doi.org/10.3390/rs14184511

57. Zhou, L., Pan, S., Wang, J., and Vasilakos, A.V. (2017). Machine learning on big data: opportunities and challenges. *Neurocomputing 237*, 350–61. https://doi.org/10.1016/j.neucom.2017.01.026

58. Zylshal, S.S., Yulianto, F., Nugroho, J.T. and Sofan, P. (2016). A support vector machine object based image analysis approach on urban green space extraction using Pleiades-1A imagery. *Modeling Earth Systems and Environment 2*(2), 54. https://doi.org/10.1007/s40808-016-0108-8