Media Konservasi, 29 (2) 249-262

MEDIA KONSERVASI
SCIENTIFIC JOURNAL IN CONSERVATION, ENVIRONMENT & ECOTOURISM

RESEARCH ARTICLE

Check for updates

Article Info:
Received 04 February 2021
Revised 26 July 2023
Accepted 27 July 2023

Corresponding Author:
Muhammad Buce Saleh
Department of Forest Management, Faculty of Forestry and Environment, IPB University
E-mail: buce_w@apps.ipb.ac.id

Canopy Density Estimation Model in Peat Swamp Forest Using LiDAR Data and Landsat 8 OLI Satellite Imagery

Muhammad Buce Saleh[a], Malta Daerangga[b], Lilik Budi Prasetyo[b], Yudi Setiawan[b], Sahid Agustian Hudjimartsu[c], and Arif Kurnia Wijayanto[b]

[a]Department of Forest Management, Faculty of Forestry and Environment, IPB University
[b]Department of Forest Resources Conservation and Ecotourism, Faculty of Forestry and Environment, IPB University
[c]Informatics Engineering Study Program, Faculty of Engineering and Science, Ibn Khaldun University

Abstract

Canopy density is one of the important parameters in measuring the forest conditions. Canopy density can be estimated by using a remote sensing technology system. Light Detection and Ranging (LiDAR) is an active remote sensing system which uses a laser that is emitted by a sensor to the objects on the earth surface. For a wide area, image utilization which solely relies on LiDAR is still relatively expensive, so it is necessary to develop a method that combine LiDAR data with other medium resolution images such as Landsat 8 OLI imagery. Therefore, this research was conducted to obtain the canopy density estimation model from LiDAR and Landsat 8 OLI data. The results showed that the best estimation model at the study site, PT Global Alam Lestari's peat swamp forest was FRCI = - 0.0171 + 8.691 GRVI. The equation model had coefficient of determination (R^2) of 50.2%, standard deviation value (s) of 0.101, aggregate deviation (SA) value of 0.459, and correlation coefficient (r) between the actual FRCI and the estimation FRCI (best model) of 0.503.

Keywords: canopy density, Landsat 8 OLI, LiDAR, vegetation

1. Introduction

Peat swamp forests are unique and vital ecosystems found in several countries across the world, including Indonesia. These forests are characterized by the presence of waterlogged, acidic, and nutrient-poor peat soils, which contribute to their distinct ecological features. Indonesia, as one of the countries with the largest extent of peat swamp forests, holds significant ecological and environmental importance. The peat swamp forests of Indonesia are predominantly located in the low-lying coastal areas of Sumatra and Kalimantan (Borneo), with smaller areas found in Papua and other regions. These forests are home to a rich biodiversity, supporting numerous endemic and endangered species, including the critically endangered Bornean orangutan, Sumatran tiger, and many unique bird species.

The peat swamp forests provide a wide range of ecosystem services, playing a crucial role in climate regulation, carbon sequestration, and water regulation. They act as immense carbon sinks, storing substantial amounts of carbon dioxide and helping mitigate global climate change. Additionally, the forests regulate water flows by acting as natural sponges, absorbing and releasing water slowly, thus reducing the risk of flooding during heavy rainfall and ensuring a steady water supply during dry seasons.

Indonesia's peat swamp forests also contribute significantly to the livelihoods and cultures of local communities. These forests provide essential resources such as timber, non-timber forest products, and traditional medicinal plants. They support traditional practices, cultural beliefs, and indigenous knowledge systems, fostering a strong connection between the local communities and their natural surroundings.

Despite their ecological importance, peat swamp forests in Indonesia face numerous threats and challenges. The expansion of agricultural activities, particularly palm oil plantations, logging, and drainage for industrial purposes, has led to widespread deforestation and degradation of these fragile ecosystems, especially land/forest fire. The conversion of peatlands for agricultural use disrupts their hydrological balance, accelerates peat decomposition, and releases significant amounts of greenhouse gases, exacerbating climate change.

Recognizing the critical value of peat swamp forests, the Indonesian government, along with various international organizations and stakeholders, has implemented conservation and restoration initiatives. Efforts are being made to protect remaining peat swamp forest areas, restore degraded ones, and promote sustainable land-use practices that balance environmental conservation with socio-economic development. Accurate assessment of canopy density in these forests is crucial for effective management and conservation efforts.

Canopy density is a ratio between the area of canopy with a certain area, and canopy diameter is calculated as the average canopy diameter [1]. Estimation of forest canopy density is determined by field measurements that requires considerable time, therefore, it is difficult to collect good data in estimating the canopy density in a large area. The use of remote sensing technology has been applied in the forestry sector such as land cover mapping, evaluation of cover changes, and forest land use [2]. The rapid development of remote sensing technology has facilitated various parties, specifically in the forestry sector to monitor forests well. One remote sensing technology that has been developed for quite a long time and has incredibly high accuracy is Light Detection and Ranging (LiDAR).

LiDAR data has several functions, including being able to display three-dimensional (3D) data of a vegetation, measuring the height of vegetation canopy, forming DEM/DTM, and estimating canopy cover [3]. However, LiDAR data obtained in a wide area is still relatively expensive, so other data collection that can produce information in a large area is needed, such as the use of Landsat 8 OLI imagery. It aims at providing data availability in a wide scope which makes it an important data source in estimating canopy density.

Forest development in Indonesia by using LiDAR and Landsat 8 OLI remote sensing data is still very rarely made in the sustainable planning and management of forests. Phu La et al. (2014) found a strong relationship between canopy cover (LiDAR) and vegetation index (NDVI) from Landsat TM imagery by 0.69. This paper aims to contribute to the conservation and management of peat swamp forests in Indonesia by developing a canopy density estimation model using LiDAR data and Landsat 8 OLI satellite imagery. By accurately assessing the canopy density, the model can provide valuable insights into the health and structure of peat swamp forests, supporting targeted conservation efforts and sustainable forest management practices.

2. Research Method

This research was conducted from February to August 2018 which was located on LiDAR flight path in the work area of PT Global Alam Lestari, Bayung Lendir Subdistrict, Musi Banyuasin Regency, South Sumatra Province. Data needed in this research included primary and secondary data. The primary data included the results of the lowest tree canopy height obtained from preliminary observations in the field Error! Reference source not found..

Table 1. Data of the research

No	Data Types	Year	Source
1	LiDAR Point Clouds	2014	Peatland Restoration Agency
2	Landsat 8 OLI path/row 124/61	2015	Earthexplorer.usgs.gov
3	Concession Boundary Map of PT Global Alam Lestari	-	PT Global Alam Lestari
4	Canopy height	2018	Field measurement

Data Processing

The Landsat 8 OLI image data processing used Erdas Imagine 9.1 and ArcMap 10.3 software. LiDAR data was obtained in the form of point clouds with standard LAS data format (.las) and the processing used RStudio 3.4.4 software. The process carried out could be seen in the figure below.

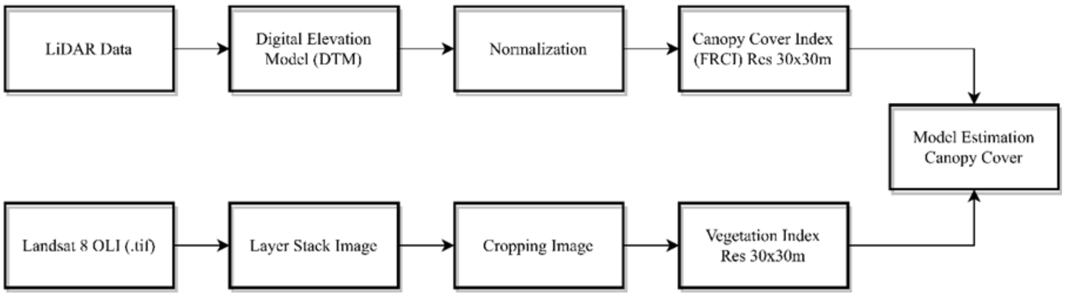


Figure 1. General workflow of the research

On the LiDAR data preprocessing, to obtain the value of canopy cover based on the extraction of FRCI values, several stages were carried out as follows: determining the Digital Terrain Model (DTM) and normalizing tree height so that it could produce the actual tree height. Furthermore, the calculation of canopy cover with a resolution of 30m x 30m was also carried out.

The Landsat 8 OLI satellite image pre-processing included data format conversion, image stack (layer stack), geometric correction, and cropping, then followed by geometric correction. The geometric corrections were carried out so that the coordinates in the image matched the geographical coordinates [4]. The Landsat 8 OLI images here had been orthorectified of 1T- precision level which meant that the Digital Elevation Model (DEM) rectification data from Global Land Surveys 2000 had been done. Since the Landsat data had been orthorectified, then the pre-processing was only image reprojection to change the image projection from geographic coordinate of World Geodetic System (WGS) 84.to Universal Transverse Mercator (UTM) zone 48S.

2.1. Lidar Data Processing

2.1.1. Making of Digital Terrain Model (DTM)

Terrestrial model or DTM is a digital surface model without objects located above the earth surface [3]. Typically, DTMs can be formed based on LiDAR last return laser data (pulses). Last Return is the last return point from a LiDAR laser that is emitted from the sensor to the objects above the earth surface. The illustration of DTM is presented in **Figure 2**.

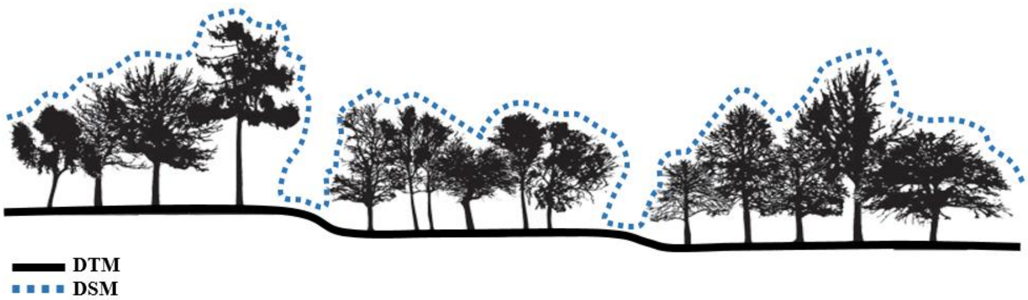


Figure 2. Illustration of terrestrial/DTM model (black)

2.1.2. Data Normalization

Data normalization is intended to get the actual canopy height. This is because the height of a canopy is calculated based on the DTM. Digital Surface Model or DSM is a digital surface model with objects (buildings, trees) located on the earth's surface [3]. Typically, DSM can be formed based on first return laser data (pulses) or even all LiDAR return data. First return is the first return point from LiDAR laser that is fired at objects on the earth surface, while all return is all return points from LiDAR laser fired at objects on the earth surface. Normalization of LiDAR data is presented in **Figure 3**.

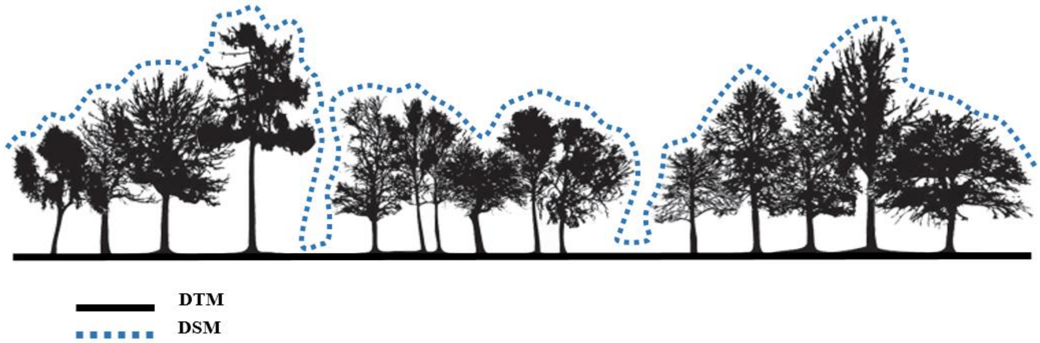


Figure 3. Illustration of LiDAR data normalization

2.1.3. Canopy Density Percentage

LiDAR-based method in the form of points to estimate canopy density can be divided into two methods, which are All Return Cover Index (ARCI) and First Return Cover Index (FRCI). ARCI is the ratio between all LiDAR laser return points in the canopy with the total of all laser return points fired in both vegetation and non-vegetation, while FRCI calculates the ratio only from the first return point and single return from the laser fired because the last and intermediate return point data from the LiDAR data provides little information about the canopy cover [5]. According to the research of Ma et al. [6], FRCI equation was closer to the results of field measurements than the ARCI equation, so this research only used FRCI method where the height taken was above 11 m and the spatial resolution adjusted the Landsat 8 OLI imagery (30x30m). The ARCI and FRCI equations are as follows:

$$ARCI = \sum All_{canopy} \;/\; \sum All_{total}$$

$$FRCI = \frac{\sum First_{canopy} + \sum Single_{canopy}}{\sum First_{total} + \sum Single_{total}}$$

All canopy is all reflections about the canopy, *all total* is the sum of all reflections, *first canopy* is the first reflection about the canopy, *single canopy* is the single reflection about the canopy*, first total* is the sum of all first reflections, and *total single* is the sum of all single reflections.

2.2. Landsat 8 OLI Vegetation Index

Vegetation index is a value obtained from mathematical operations using pixels derived from several channels contained in the image [4]. This research used several vegetation indexes from Landsat 8 OLI images which were used as free variables. The vegetation index used is presented in **Table 2**.

Table 2. Vegetation index used

No	Vegetation Index	Formula
1.	Soil Adjusted Vegetation Index (SAVI)	$SAVI = \frac{NIR - RED}{NIR + RED + 0.5}(1 + 0.5)$
2.	Green Red Vegetation Index (GRVI)	$GRVI = \frac{GREEN - RED}{GREEN + RED}$
3.	Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{NIR - RED}{NIR + RED}$

4.	Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = \frac{NIR - GREEN}{NIR + GREEN}$
5.	Difference Vegetation Index (DVI)	$DVI = NIR - RED$

Determining Sample Plot

The sample plot collection was determined by using a purposive sampling method. The sample plots were selected based on the LiDAR path with consideration of not being covered by clouds, the level of vegetation density, no forest damage, and not about canals. The sample plots selected were 395 with the plot size adjusting the spatial resolution of Landsat 8 OLI imagery, 30x30m. Data taken in each plot were LiDAR canopy density value (FRCI) and OLI Landsat 8 vegetation index value. The distribution of sample plots is presented in **Table 3**.

Table 3. Number of sample plot based on density class distribution

Density class	Percentage	Number of plot
Low dense	10% - 39.9%	61
Moderately dense	40% - 69.9%	290
High dense	> 69.9%	44
	Total sample plot points	395

Modelling Procedure

2.3. Variable Selection

When evaluating high-dimensional data, where multicollinearity analysis is frequently used, variable selection is a typical approach [7,8]. Multicollinearity, which happens when two or more variables have a high degree of correlation, can result in a model that is unreliable and has low predictive power [8–18]. As a result, the degree of collinearity might affect both the estimation of model coefficients and the model's interpretation [19]. To solve this issue, we used Pearson correlation to apply variable selection to all the indicated covariates, including climatic, anthropogenic, and biophysical factors. When choosing the variables, we used a 0.7 Pearson correlation threshold as a criterion to eliminate strongly correlated predictors.

2.4. Regression Model

The construction of regression equation model that is built previously can be seen from the scatter diagram between dependent variable (FRCI) and independent variable (vegetation index) that forms a certain pattern. Regression model that only uses one independent variable to predict an independent variable is called simple linear regression models, while a regression model that uses two or more independent variables in estimating the dependent variable is called a multiple regression model [20]. The regression equation model used in estimating the canopy density was based on LiDAR data with free variables of vegetation index values in Landsat 8 OLI image.

2.5. Regression Model Testing

After the regression model had been obtained, then a testing was conducted on several regression parameters which were F test, coefficient of determination (R^2), adjusted coefficient of determination (R^2adj), and standard deviation (s). The stages of preparing and testing the regression equation model were assisted using Minitab 18 software.

4.a F Test

Testing was carried out by comparing F-count value generated in ANOVA table with F-table with a confidence level of 95% (α = 0.05). If F-count value is higher than the F-table then H0 is rejected, which means one or more free variables in the model significantly affects a certain level (α) [20]. The decision rules are presented in **Table 4**.

Table 4. Decision rules that would be used in the research

Decision rules	Description
Fcount > Ftable	Accept H_0
Fcount ≤ Ftable	Reject H_0

4.b Coefficient of Determination

Coefficient of determination is a measure of independent variable diversity that can be explained by the independent variable diversity. The coefficient of determination can indicate the accuracy and closeness of the regression model relationships that have been made. According to Tiryana [20], the coefficient of determination (R^2) is a value that explains the amount of diversity in dependent variable (Y), which can be explained by independent variable (X). The value of R^2 is expressed as a percentage with the equation to calculate R^2 as follows:

$$R^2 = \frac{JKR}{JKT} \, .100\ \%$$

Description: JKR is the number of regression square, JKT is the total square. JKR and JKT are obtained from the ANOVA table.

4.c Standard Deviation

Standard deviation is the root from the middle of the deviation square from the median or the average deviation root [21]. Standard deviation (s) indicates that the smaller the value, the better the result is, so that the estimate value will be more accurate. Standard deviation values are calculated using the following equation [22]:

$$s = \sqrt{S^2} = \sqrt{\frac{JRS}{(n-p)}}$$

Description: s is standard deviation, JKS is the sum of residual square, $(n–p)$ is the residual degree of freedom.

2.6. Validation Test

Models that have met the criteria for the best modelling through the coefficient of determination (R^2), standard deviation (s), and the F test need to be tested for validation. This research used several calculations, namely aggregate deviation (SA) and correlation test (r).

Aggregate deviation (SA) is the difference between the number of actual values and the estimate value as proportional to the estimate value or the difference between the total actual value and the sum of the estimate values. An equation is good if you have SA values between -1 and 1. SA values can be calculated by using the following formula:

$$SA = \frac{\sum FRCIt - \sum FRCIa}{\sum FRCIt}$$

Description: FRCIt is the estimate canopy density value, FRCIa is the actual canopy density value.

Correlation test (r) in this research was intended to see the correlation coefficient between the actual canopy density (FRCI) and the estimate canopy density. Correlation coefficient values range from -1 to 1 which shows the size of the independent variable (presumed FRCI) in explaining the dependent variable (actual FRCI). In this research, a good correlation test had a correlation coefficient (r) that was close to 1.

2.7. Selection of the Best Regression Model

The best model is obtained from the ranking of the model with reference to the validation test criteria. The ranking is done by giving a score/eight to the models obtained, then the best model that can be used according to the existing criteria is the selected model that has the smallest aggregate deviation value and the largest correlation coefficient. Scoring can use the following formula:

$$SA\ Score = \left(\frac{NSA - \max SA}{\min SA - \max SA} x(n-1)\right) + 1$$

$$r\ score = \left(\frac{Nr - \min r}{\max r - \min r} x(n-1)\right) + 1$$

Description: NSA is the aggregate deviation value, Nr is the correlation value, max SA/r is the largest value of each criterion, min SA/r is the smallest value of each criterion.

Canopy Density Display

The best model chosen is displayed in the form of raster data and visually observed with actual canopy density (FRCI) in the form of raster data as well and see if there is a difference between the displays. If the raster display between the actual canopy density and the estimate canopy density is different, then calculate the amount of error using the regression model. Regression model is made based on the value of the actual canopy density as dependent variable and the value of the estimate canopy density as independent variable.

3. Results and Discussion

We recommend separate approach for results and discussion. In separated approach, you present the results first, providing an objective description of your findings, and then follow with a separate discussion section where you interpret and discuss the implications of those findings.

3.1. Field Observation

Field observations were carried out in the work area of PT Global Alam Lestari by determining the lowest height of the tree canopy from various types of density classes. Three types of density class found in the field were shrub class, low density class, and medium density class.

(a) (b) (c)

Figure 4. Three types of canopy density at the study area: (a) shrub, (b) low density, and (c) medium density

The shrub class had a height of 2 m. The low-density class had the lowest tree canopy height of 8.2 m and the medium density class had the highest tree canopy height reaching 30 m. Furthermore, field observations were used as a basis for determining canopy density from LiDAR data where canopy densities were calculated from the lowest canopy height of a tree with a height taken 11 m and not including shrub.

3.2. Modelling

3.2.1. Multicollinearity Test

Correlation testing aims to determine the correlation between dependent variables (FRCI) and independent variables (vegetation index). The correlation coefficient values are presented in **Table 5**.

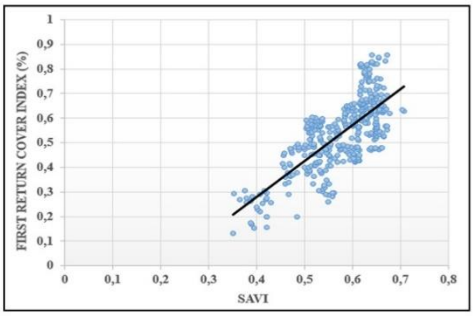
Table 5. Results of multicollinearity analysis

	SAVI	GRVI	NDVI	GNDVI	DVI
SAVI		0.891	1.000	0.992	0.981
GRVI	0.891		0.891	0.829	0.801
NDVI	1.000	0.891		0.992	0.981
GNDVI	0.992	0.829	0.992		0.994
DVI	0.981	0.801	0.981	0.994	

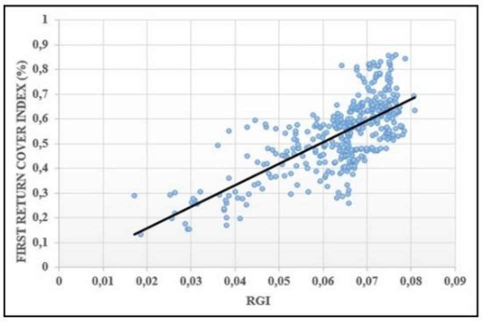
Table 5 shows that the correlation between independent variables and other independent variables has a large correlation coefficient, so when using two or more independent variables in making models, it will experience multicollinearity. Multicollinearity causes the regression coefficient variance that is not minimal, so the regression model is unstable [23]. Therefore, it is necessary to use only one independent variable in building the regression model.

3.2.2. Canopy Density Estimation Model

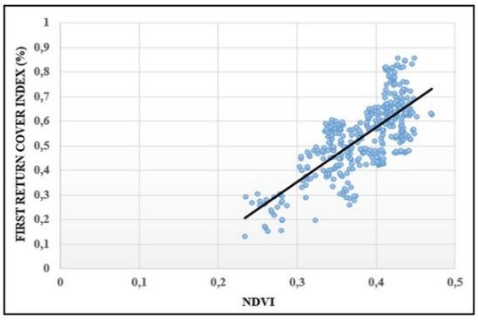
The canopy density estimation model was made using 395 sample plots taken from LiDAR canopy density data (FRCI) and vegetation index from Landsat 8 OLI satellite imagery. The canopy density estimation equations can be constructed based on scatter diagram patterns. A scatter diagram of the relationship between LiDAR canopy density (FRCI) and vegetation index from Landsat 8 OLI satellite imagery is presented in **Figure 5**.



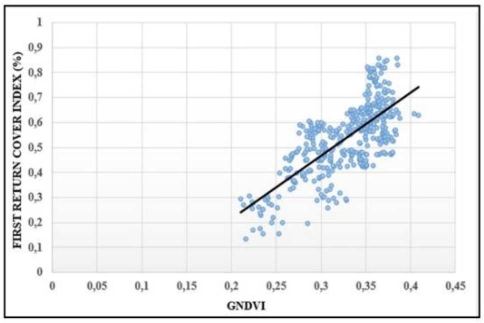
(a)



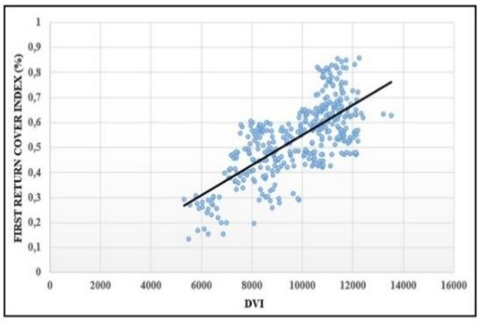
(b)



(c)



(d)



(e)

Figure 5. Relationship between (a) SAVI and FRCI; (b) GRVI and FRCI; (c) NDVI and FRCI; (d) GNDVI and FRCI; (e) DVI and FRCI

Scatter diagrams between canopy density (FRCI) as dependent variable (Y) and independent variable vegetation indices (X) show that the regression analysis conducted can use a linear regression model. The equation model made was 15. According to Tiryana [20], a good equation model is a model with a high coefficient of determination (R^2) near 100%. Therefore, the R^2 value exceeding 50% was taken. Furthermore, from the construction of canopy density estimation models it was obtained 12 regression models consisting of linear, logarithmic, and quadratic models. The equation models obtained are presented in **Table 6**.

Table 6. Canopy density estimation equation model (FRCI)

Type	Code	Vegetation Index	Equation	S	R^2 (%)	Fcount	Ftable
Linear	M1	SAVI	Y = -0.3070 + 1.4684 SAVI	0.097	54.45	469.88	3.87
	M2	GRVI	Y = -0.0171 + 8.691 GRVI	0.101	50.24	396.80	3.87
	M3	NDVI	Y = -0.3070 + 2.203 NDVI	0.097	54.45	469.88	3.87
	M4	GNDVI	Y = -0.2892 + 2.520 GNDVI	0.100	51.63	419.44	3.87
Logarithmic	M5	SAVI	Y = 0.9773 + 0.7827 ln SAVI	0.097	54.56	471.97	3.87
	M6	NDVI	Y = 1.2947 + 0.7828 ln NDVI	0.097	54.56	471.97	3.87
	M7	GNDVI	Y = 1.4116 + 0.7783 ln GNDVI	0.099	52.02	426.09	3.87
	M8	DVI	Y = -4.544 + 0.5538 ln DVI	0.101	50.79	405.55	3.87
Quadratic	M9	SAVI	Y = 0.0889 + 1.3353 SAVI²	0.098	53.61	454.09	3.87
	M10	GRVI	Y = 0.2077 + 78.33 GRVI²	0.101	50.69	403.97	3.87
	M11	NDVI	Y = 0.0889 + 3.004 NDVI²	0.098	53.61	454.09	3.87
	M12	GNDVI	Y = 0.1029 + 3.976 GNDVI²	0.101	50.69	404.05	3.87

A good regression model equation has large coefficient of determination (R^2), F-count value is higher than F-table value, and small standard deviation (S). Based on **Table 7**, it shows that the equation model produces a coefficient of determination (R^2) ranging from 50.24% - 54.56%. The highest coefficient of determination occurs in the logarithmic model with SAVI and NDVI vegetation index. The value of R^2 respectively in the model was 54.56% which meant that the independent variable (vegetation index) could explain the dependent variable (FRCI) of 54.56%. Based on the F test, it could be seen that the F-count value of each model was higher than the F-Table value at 5% significance level ($\alpha = 0.05$). This showed that the independent variable of vegetation index significantly affected canopy density (FRCI) at 5% significance level ($\alpha = 0.05$).

Standard deviation values range between 0.097 - 0.101. A good standard deviation has the smallest value [20]. The lowest standard deviation values were the linear and logarithmic models with a vegetation index of SAVI and NDVI by 0.097.

3.2.3. Heteroscedasticity Test

After the canopy density models had been obtained, heteroscedasticity testing was conducted by using Glejser method. The results of the heteroscedasticity test are presented in **Table 7**.

Table 7. Results of heteroscedasticity test

Vegetation index	p-value	Description
SAVI	0.740	No heteroscedasticity
GRVI	0.316	No heteroscedasticity
NDVI	0.741	No heteroscedasticity
GNDVI	0.558	No heteroscedasticity
DVI	0.768	No heteroscedasticity

Based on the results of heteroscedasticity test if the p-value > 0.05, it can be concluded that the relationship between one variable and another does not experience heteroscedasticity or has similarities between the variants from the residuals of one observation to another observation. Based on Table 6, the relationship between the independent variable (vegetation index) and the dependent variable (FRCI) has a p-value of more than 0.05, thus indicating that heteroscedasticity does not occur, or it does not pass the heteroscedasticity test.

3.2.4. Validation Test

A validation test was conducted using census method on all data and as many as 6250 plots were obtained to see the extent of accuracy of an equation. All the equation models that have been built can be used in estimating canopy density. However, the use of all models in estimating canopy density is still ineffective, so it is necessary to conduct a validation test. The validation test of this canopy density equation model uses aggregate deviation (SA) and correlation test (r) criteria. The aggregate deviation looks at the difference between the estimated canopy density and the actual canopy density. The correlation test looks at the relationship between actual canopy density and estimate canopy density. The results of canopy density estimation model calculation are presented in **Table 8**.

Table 8. Validation test results. Asterisk (*) depict the best model with the lowest SA and the largest correlation (r)

Code	Validation Test Criteria	
	SA	r
M1	0.498	0.461
M2	0.459*	0.503*
M3	0.499	0.461
M4	0.502	0.441
M5	0.489	0.460
M6	0.489	0.460
M7	0.493	0.439
M8	0.512	0.292
M9	0.507	0.462
M10	0.469	0.502
M11	0.507	0.462
M12	0.511	0.444

Based on the results of model validation test in **Table 8**, the results of aggregate deviation (SA) validation show that the entire model meets the requirements with aggregate deviation values ranging from -1 to 1. A good SA value that is the closest to zero, is shown in M2 model with a GRVI vegetation index equal to 0.459 which means that the estimate canopy density value overestimated the actual canopy density value. A good correlation coefficient value that has the highest correlation coefficient is shown in the M2 model also with a vegetation index of 0.503 GRVI.

3.2.5. Selection of the Best Regression Model

The best equation model can be determined by using the number of scores on each criterion used in the validation test. The criteria used are aggregate deviation values and correlation coefficient values. The model that has the highest score is used as the best model. The results of the calculation of the total scores are presented in **Table 9**.

Table 9. Score for determining the best equation model

Code	Score		Total	Ranking
	SA	r		
M1	0.74	9.61	10.35	5
M2	2.86	12.00	14.86	1
M3	0.73	9.61	10.34	6
M4	0.51	8.47	8.98	10
M5	1.25	9.55	10.81	4
M6	1.25	9.55	10.81	3
M7	1.01	8.36	9.37	9
M8	0.00	0.00	0.00	12
M9	0.27	9.67	9.94	8

M10	2.34	11.9	14.24	2
M11	0.27	9.67	9.94	7
M12	0.04	8.64	8.68	11

Based on **Table 9**, it shows that the best LiDAR canopy density estimation model is a simple linear model on M2 model where FRCI = - 0.0171 + 8.691 GRVI. That is because the simple linear model with GRVI vegetation index has the highest score compared to other models that is equal to 14.86.

3.3. Canopy Density Display

From the best equation model that had been chosen, a comparison between the displays in the form of a raster and the existing FRCI raster was conducted. The comparison of raster data display between the best canopy density estimates model and the canopy density (FRCI) taken from the two sample segmentations is presented in **Figure 6** and **Figure 7**.

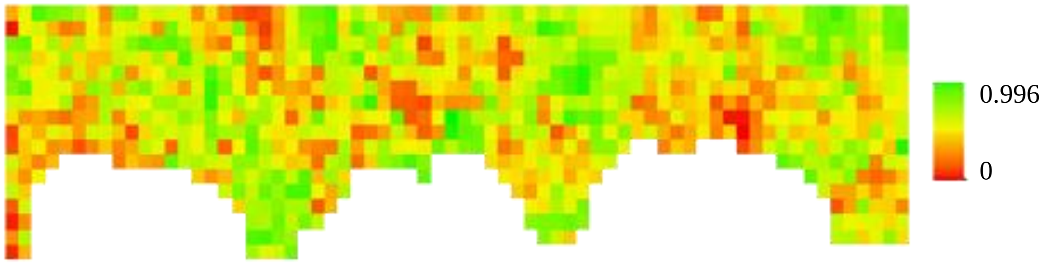


Figure 6. LiDAR-based canopy density

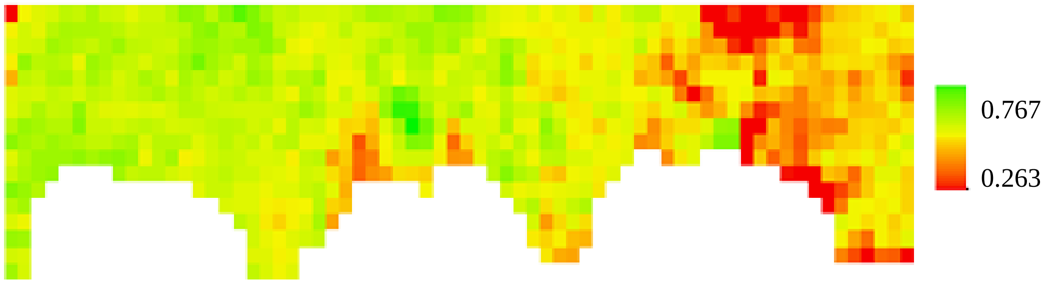


Figure 7. Model-based canopy density

Based on the comparison of canopy density displays that are visually observed in the form of raster data, it shows that the value of canopy density between the best model (M2) and FRCI experiences discrepancy. Canopy density (FRCI) values range from 0 to 0.996, while the canopy density (M2) values range from 0.263 to 0.767. In this research, regression modeling was conducted between the canopy density (FRCI) as dependent variable and estimate canopy density (M2) as the independent variable. The results of making a regression model were the actual FRCI = -1.3512 + 2.7073 estimates FRCI, where the equation had R^2 value of 30.43% and standard error value (s) of 0.2451. Based on the results of the regression, it shows that the estimate FRCI model has an error of 49.50%. This happens due to differences in data collection for each image and the level of spatial resolution of each image as presented in **Table 10**.

Table 10. Comparison of displays based on image capture time

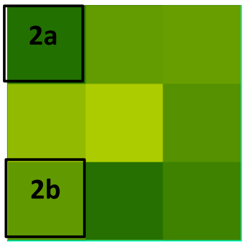
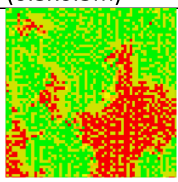
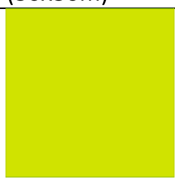
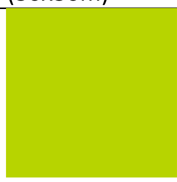
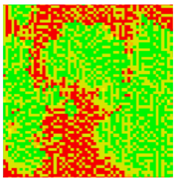
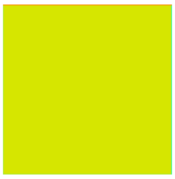
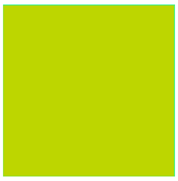
No	Accuracy Object	LiDAR (November 2014)	Estimation Model (April 2015)
1.	Image capture time	 1a: 0.794 1b: 0.949	 1a: 0.439 1b: 0.339
2.		 2a: 0.797 2b: 0.939	 2a: 0.244 2b: 0.105

Table 11. Differences in display based on spatial resolution

No	Accuracy Object	LiDAR		The Best Model
		(0.5x0.5m)	(30x30m)	(30x30m)
1.	Spatial resolution differences	 0.516	 0.586	 0.662
2.		 0.493	 0.572	 0.656

The difference in acquisition time of each image is one of the causes that makes the canopy density of the best model different from the actual canopy density. LiDAR was acquired in November 2014 and Landsat 8 OLI satellite imagery was acquired in April 2015. Based on Table 9, it is showed that the canopy density values in pixels 1a and 1b from LiDAR are 0.794 and 0.949, while in the best model of Landsat 8 OLI imagery, it is showed that the value of the canopy density at pixels 1a and 1b is equal to 0.439 and 0.339. This showed that the forest canopy in November 2014 was still classified as high dense, while in April 2015 the forest canopy density decreased due to forest fires or tree logging.

The difference in the spatial resolution of each image is also one of the causes of differences in the canopy density value. LiDAR resolution is 0.5x0.5m, while Landsat 8 OLI image resolution is 30x30m. Based on Table 10, it is showed that the average value of LiDAR canopy density with a spatial resolution of 0.5x0.5m adjusted to the Landsat 8 OLI image resolution is 0.516, and the value of the LiDAR canopy density with a spatial resolution of 30x30m is 0.586, while the best model of the imagery Landsat 8 OLI is 0.662.

4. Conclusions

The LiDAR canopy density estimation model using Landsat 8 OLI satellite imagery at PT Global Alam Lestari is FRCI = -0.0171 + 8.691 GRVI. The equation model has coefficient of determination (R^2) of 50.24%, standard deviation value (s) of 0.101, aggregate deviation (SA) value of 0.459, and correlation coefficient (r) between the actual FRCI and the estimate FRCI (M2) is 0.503. Regression between canopy density (FRCI) and canopy density (M2) has an error of 49.50%. The display of canopy density (FRCI) and canopy density (M2) when it is compared visually shows slightly different canopy density values. This is due to differences acquisition time in data collection and the level of spatial resolution of each image.

Suggestion

Further research needs to be conducted regarding the canopy density estimation model with a large enough LiDAR data to represent the forest condition. It is also recommended to use High Resolution Satellite Imagery or Very High-Resolution Satellite Imagery and the same data acquisition.

References

1. Priyanto *Model Penduga Biomassa Tegakan Hutan Rawa Gambut Menggunakan Citra SPOT Pankromatik [Tesis]*; IPB Press: Bogor, 2011;
2. Lu, D. The Potential and Challenge of Remote Sensing-Based Biomass Estimation. *Int J Remote Sens* **2006**, *27*, 1297–1328.
3. Prahasta, E. *Pengolahan Data Sistem LiDAR (Light Detection and Ranging) Di Bidang Teknik Geodesi Dan Geomatika*; Informatika Bandung: Bandung, 2015;
4. Jaya, I.N.S. *Analisis Citra Digital: Perspektif Penginderaan Jauh Untuk Pengelolaan Sumberdaya Alam*; Fakultas Kehutanan IPB: Bogor, 2010;
5. Korhonen, L.; Korpela, I.; Heiskanen, J.; Maltamo, M. Airborne Discrete-Return LiDAR Data in the Estimation of Vertical Canopy Cover, Angular Canopy Closure and Leaf Area Index. *Journal of Remote Sensing of Environment* **2011**, *115*, 1065–1080.
6. Ma, Q.; Su, Y.; Guo, Q. Comparison of Canopy Cover Estimation from Airborne LiDAR, Aerial Imagery, and Satellite Imagery. *Journal of Selected Topic in Applied Earth Observations and Remote Sensing* **2017**, *10*, 4225–4236.
7. Mahecha, M.D.; Martínez, A.; Lischeid, G.; Beck, E. Nonlinear Dimensionality Reduction: Alternative Ordination Approaches for Extracting and Visualizing Biodiversity Patterns in Tropical Montane Forest Vegetation Data. *Ecological Informatics* **2007**, *2*, 138–149, doi:10.1016/J.ECOINF.2007.05.002.
8. Zhao, N.; Xu, Q.; Tang, M.L.; Jiang, B.; Chen, Z.; Wang, H. High-Dimensional Variable Screening under Multicollinearity. *Stat* **2020**, *9*, doi:10.1002/sta4.272.
9. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; et al. Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Ecography* **2013**, *36*, doi:10.1111/j.1600-0587.2012.07348.x.
10. Toothaker, L.E.; Aiken, L.S.; West, S.G. Multiple Regression: Testing and Interpreting Interactions. *The Journal of the Operational Research Society* **1994**, *45*, doi:10.2307/2583960.
11. Farrar, D.E.; Glauber, R.R. Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics* **1967**, *49*, doi:10.2307/1937887.
12. O'Brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality and Quantity* **2007**, *41*, doi:10.1007/s11135-006-9018-6.
13. Tu, Y.K.; Clerehugh, V.; Gilthorpe, M.S. Collinearity in Linear Regression Is a Serious Problem in Oral Health Research. *European Journal of Oral Sciences* 2004, *112*.
14. Graham, M.H. Confronting Multicollinearity in Ecological Multiple Regression. *Ecology* **2003**, *84*, doi:10.1890/02-3114.
15. Mason, C.H.; Perreault, W.D. Collinearity, Power, and Interpretation of Multiple Regression Analysis. *Journal of Marketing Research* **1991**, *28*, doi:10.2307/3172863.

16. Fekedulegn, B.D.; Colbert, J.J.; Hicks, R.R.; Schuckers, M.E.; Service, F.; Fekedulegn, B.D.; Colbert, J.J.; Hicks, R.R.; Schuckers, M.E. Coping with Multicollinearity : An Example on Application of Principal Components Regression in Dendroecology. *United States Department of Agriculture* **2002**.

17. Stewart, G.W. Collinearity and Least Squares Regression. *Statistical Science* **1987**, *2*, doi:10.1214/ss/1177013439.

18. Gordon, R.A. Issues in Multiple Regression. *American Journal of Sociology* **1968**, *73*, doi:10.1086/224533.

19. Feng, X.; Park, D.S.; Walker, C.; Peterson, A.T.; Merow, C.; Papeş, M. A Checklist for Maximizing Reproducibility of Ecological Niche Models. *Nature Ecology & Evolution* **2019**, *3*, 1382–1395, doi:10.1038/s41559-019-0972-5.

20. Tiryana, T. *Panduan Praktis Analisis Regresi Linear Dengan Program Minitab For Windows*; Fakultas Kehutanan IPB: Bogor, 2008;

21. Santi, R.C.N.; Eniyati, S. Implementasi Statistic Dengan Database Mysql. *Jurnal Teknologi Informasi DINAMIK* **2015**, *20*, 132–139.

22. Draper, N.R.; Smith, H. *Analisis Regresi Terapan Edisi 2 Terjemahan*; Gramedia: Jakarta, 1992;

23. Maubanu, E.; Kartiko Analisis Komponen Utama Untuk Mengatasi Multikolinearitas Pada Faktor-Faktor Yang Mempengaruhi Tingkat Pengangguran Terbuka Di Provinsi Nusa Tenggara Timur. *Jurnal Statistik Industri dan Komputasi* **2018**, *3*, 21–30.

