Canopy Cover Estimation in Lowland Forest in South Sumatera, Using LiDAR and Landsat 8 OLI imagery

Muhammad Buce Saleh[*1], Rosima Wati Dewi [1], Lilik Budi Prasetyo[2], Nitya Ade Santi[1]

[1]Departement of Forest Management, Faculty of Forestry and Environment, IPB University, Academic Ring Road, Campus IPB Dramaga, Bogor, Indonesia 16680
[2]Department of Forest Resource Conservation and Ecotourism, Faculty of Forestry and Environment, IPB University, Academic Ring Road, Campus IPB Dramaga, Bogor, Indonesia 16680

Abstract

Canopy cover is one of the most important variables in ecology, hydrology, and forest management, and useful as a basis for defining forests. LiDAR is an active remote sensing method that provides the height information of an object in three-dimensional space. The method allows for the mapping of terrain, canopy height and cover. Its only setback is that it has to be integrated with Landsat to cover a large area. The main objective of this study is to generate the canopy cover estimation model using Landsat 8 OLI and LiDAR. Landsat 8 OLI vegetation indices and LiDAR-derived canopy cover estimation, through First Return Canopy Index (FRCI) method, were used to obtain a regression model. The performance of this model was then assessed using correlation, aggregate deviation, and raster display. Lastly, the best canopy cover estimation was obtained using equation, FRCI = 2.22 + 5.63Ln(NDVI), with R^2 at 0.663, standard deviation at 0.161, correlation between actual and predicted value at 0.663, aggregate deviation at -0.182 and error at 56.10%.

Keywords: *canopy cover, Landsat 8 OLI, LiDAR, vegetation indices*

*Correspondence author, email: *buce.saleh@gmail.com*

Introduction

Canopy hydrology cover is one of the most important parameters used in ecology and forest management (Nakamura et al., 2017). It is defined as an area of land covered by vertical projections of canopies, which describes the structural conditions of a forest (Jennings et al., 1999). Furthermore, it is used as a critical variable for the definition and evaluation of forest gain (recovery) and loss (degradation and deforestation) (FAO, 2000). The measurement is useful for forest management applications such as for classification of forest structure, characterization of carbon sinks, forest fire behavior and fuel models, as well as the estimation of canopy light transmission(Ahmed et al., 2014).

The measurement of a canopy cover can be obtained either directly on the field or based on remote sensing technology. Field measurement provides more accurate data but relatively laborious and expensive. Meanwhile, remote sensing technology such as LiDAR offers large spatial coverage of an area, more efficient in time, less financial cost, and less human resources, and also more powerful for the detection and segmentation of trees —(Zhen et al., 2016). The technology is expected to replace the field measurement method (Hyyppa et al., 2001; Kim et al., 2016; Jeronimo et al., 2018; Irlan et al., 2020). LiDAR is an active remote sensing technology that provides height information of an object (Jakubowski et al., 2013). It calculates a distance by emitting a laser pulse and calculating the return time from the laser to its sensor after being reflected. LiDAR is well-suited for canopy cover estimation since it can penetrate through the canopy (Korhonen & Morsdorf, 2014). Other benefits of using this system are: can be used during day and night, more effective and efficient in operational cost compared to terrestrial surveys, can provide high precision and accurate elevation data, and can process a large amount of data in a short period of time (Jakubowski et al., 2013). Although this technology has promising benefits, it is often limited in spatial coverage, and is relatively expensive due to its acquisition cost. To overcome this limitation, it can be integrated with another remote sensing technology.

The integration LiDAR with Landsat 8 OLI is used to determine the potential of Landsat imagery, to accurately estimate forest canopy cover, in order to expand the measurement area (Ahmed et al., 2014).This integration could make LiDAR more cost-effective over larger areas (Hudak et al., 2002; Chen et al., 2012). The estimation of forest canopy cover can be obtained by using a regression between forest canopy cover, and vegetation indices that have been generated from LiDAR data and Landsat 8 OLI Imagery respectively. Vegetation indices are a mathematical combination or a transformation of spectral bands that

accentuate the spectral properties of green plants: therefore, they can appear distinct from one another in features. Another research showed a result if vegetation indices gave a good result for detecting growth using remote sensing with standards error 9.16% (Jaya et al., 2019).

The vegetation indices that have been proposed for determining the vigor and health of vegetation is intended to create better indices, which means that they take into account many factors, such as soil reflectance, atmosphere and vegetation density. All these improvements and modifications aim to get more reliable information about vegetation, based on their reflectances (Jakubowski et al., 2013). Canopy cover measurement is difficult. LiDAR makes measurement is easier and faster, as well as the coverage can be wider. So that building models and validation through Landsat/SPOT will be more effective. Therefore, the objective of this research is to examine the capability of vegetation indices to estimate forest canopy cover in dry lowland forests, in Sumatera Selatan, Indonesia.

Methods

Location This research was carried out in Harapan Forest of PT REKI, South Sumatera (Figure 1), which astronomically is located at E103°7'55"–S103°27'39" and S2°2'16"–S2°21'14" with elevation range of 30–120 m above the sea level. Administratively, it is located between Soralangun District and Batanghari District. This forest is a dry lowland type and is characterized by secondary forest, shrubs, opened area, and mixed plantations. The area has a humid tropical climate with even precipitation all year round and is divided into three categories: high, medium, and low secondary forest. In high secondary forest, the stratum consists of seedling (height < 1.5 m), sapling (heights > 1.5 m and diameter > 20 cm), pole (10 cm < diameter < 20 cm), and mature tree (diameter > 20 cm). The canopy cover in this category has a range between 71%–100%, and is dominated by *Shorea* sp., *Litsea* sp., and *Palaquium* sp. The medium is a transition between the high and low secondary forests. It has a canopy cover range of less than 40%, is dominated by pole (diameter 10–20cm), and is mostly covered by *Shorea* sp., *Litsea* sp., and *Koompasia excelsa*. The low secondary forest has a canopy cover of less than 40%, and is categorized as a very degraded forest (REKI, 2020). Shrubs and poles are commonly found in this region, with the low secondary forest particularly on the burnt areas. The LiDAR footprint located in the Harapan rainforest strips from west to east. LiDAR aquisition mode using full waveform dan discrete return with 800 m flying height and 500 Khz laser pulse frequency. LiDAR point density has 8–15 points m^{-2} for the full waveform LiDAR and 6–8 points m^{-2} for the discrete return LiDAR. For scanning angle we used field of View (FOV) 56, swath width 851 m, and forward overlap 6% for full waveform and 80% forward overlap for discrete return LiDAR.

Data LiDAR data was acquired using an airborne method on October, 2014, and then formatted in LASer (LAS). LAS is a file format for the interchange of 3-dimensional point cloud data. Afterwards, LiDAR pulses were converted into point clouds which were used to calculate the canopy cover.

Preceding the calculation, the quality of the point clouds was investigated to make sure that they had a complete classification, and their heights represented the entire population in the area. This quality check was conducted because classification and heights were the most important variables. The points cloud which had a usual height was filtered out, and used to generate the digital terrain model (DTM). Normalization was then performed to get the actual height of the canopy cover. Canopy cover estimation was calculated using first return cover index (FRCI) method as shown in Equation *[1]* (Ma et al., 2017), which utilized only the first and the single returns, because it was assumed that the last and intermediate were only providing little additional information (Korhonen et al., 2011).

$$\text{FRCI} = \frac{\sum First\ canopy + \sum Single\ Canopy}{\sum First\ Canopy + \sum Single\ Canopy} \qquad [1]$$

The laser pulse transmitted by the LiDAR sensor can return as one (single return) or multiples returns (first, second, third, fourth and last return). Either the first canopy return or single canopy return is the most significance return, which is associated with the highest feature (tree canopy). To estimate the canopy cover, threshold for tree classification was set at 7 m, based on the field observation. So, object below this height was not included in the calculation. The result of the estimation was a point feature generated from 30 m × 30 m grid area, following Landsat 8 OLI image resolution. This point feature was then overlaid with Landsat vegetation indices. The canopy cover estimation value ranged between 0–100%.

Landsat 8 OLI is a satellite imagery that offers large spatial coverage, and wide range of spectral resolution. The Landsat imagery of path 125, row 6 was obtained in July 2014 from the United States Geological Surveys (USGS) (https://usgs.earthexplorer.com). This acquisition occurred three months before the LiDAR was obtained. Pre-processing data such as geometric and topographic correction were applied based on illumination correction (Tan et al., 2013). Geographic correction was conducted to correct distortions and assign the properties (and practical value) of the map to the images (Green et al., 2000). It was performed by reprojecting the initial projection from WGS 1984 Zone 48 N to WGS 1984 Zone 48 S. Then, topographic correction was performed using terrain illumination correction model by combining the mathematical model and empirical rotation model (Hudjimartsu et al., 2017). The model was the improvement of topographic influence correction (Tan et al., 2013). In mountainous areas, topography strongly influences the signal recorded by spaceborne optical sensors. In particular, the same surface cover slopes oriented away from and towards the sun which would appear darker and brighter, respectively, when compared to a horizontal geometry (Richter et al., 2009).

After the pre-processing had been carried out, the vegetation indices were calculated. Vegetation indices were calculated using raster calculator in Arcmap. Vegetation indices are defined as the mathematical transformation involving several bands from the optical sensors' imagery, which are used in creating new imagery that are more representative in presenting aspects related to vegetation.

Figure 1 Study area map.

Vegetation indices were also derived from satellite data and have been widely used to assess variations in the physiological state and biophysical properties of vegetation (Huete et al., 2002). This study used normalized difference vegetation index (NDVI), green normalized vegetation index (GNDVI), and simple ratio vegetation index (SRVI) as predictors to predict the canopy cover estimation as shown in Equation *[2]*, Equation *[3]*, and Equation *[4]*.

$$NDVI = \frac{NIR-RED}{NIR+RED} \text{ (Rouse et al., 1974)} \qquad [2]$$

$$GNDVI = \frac{NIR-GREEN}{NIR+GREEN} \text{ (Gitelson et al., 1996)} \qquad [3]$$

$$SRVI = \frac{NIR}{RED} \text{ (Rouse et al., 1974)} \qquad [4]$$

note: Green = band with a wavelength of 500–600 nm; Red = band with a wavelength of 600–700 nm; NIR = band with a wavelength of 700–1000 nm.

Statistical methods: Building the regression model
Regression model was used in canopy cover equation, with FRCI as dependent variable and vegetation indices as independent variable. Before developing the regression model, several tests such as classic assumption and correlation test were carried out. The classic assumption test is a prerequisite in developing a regression model. For constructing this model, data should: 1) have a normal distribution, 2) have a homoscedasticity of errors, 3) have no autocorrelation occurrences, and 4) have no multicollinearity for the model estimation by using more than one independent variable. This research was only using one variable and a single date imagery, therefore, only the normality and heteroscedasticity of data were tested. The normality was tested using Kolmogorov-Smirnov test, which was first introduced by A. Kolmogorov (Kolmogorov, 1933) and later modified and proposed as a test by N. Smirnov (Smirnov, 1948).

$$D = Sup_x |F_n(X) - F(X, \mu, \sigma)| \qquad [5]$$

Refers to Equation *[5]*, $F(X, \mu, \sigma)$ is the theoretical cumulative distribution function of the normal distribution function and (X) is the empirical distribution function of the

data. When the above equation results in large values of *D,* it indicates the data are not normal (Das & Imon, 2016). Heteroscedasticity test to find out whether there are differences in residual variance, from one observation to another. It is performed by the Glejser Test, using the SPSS program. The Glejser test is a well-known test for heteroscedasticity (Glejser, 1969), which is based on weak assumptions, and is very easy to implement. It checks for the presence of a systematic pattern in the variances of the errors, by estimating the auxiliary regression, where the absolute value of the residuals of the main equation is the dependent variable (Furno, 2005). The test decision is conducted when the significance value between the dependent variable and the residual value (Sig) is higher than 0.05. It means that there is no heteroscedasticity found in the data. Correlation test is used to determine the direction and strength of the relationship between two variables. While correlation measures the monotonic association between two variables. This monotonic relationship refers to a situation when both variables increases simultaneously, or when as one variable increases, the other decreases (Schober et al., 2018). The correlation test uses the equation as shown in Equation *[6]*.

$$r = \frac{\sum_{i=1}^n x_i y_i \left(\frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}\right)}{\sqrt{\left(\sum_{i=1}^n x_i^2 \frac{(\sum_{i=1}^n x_1)^2}{n}\right)\left(\sum_{i=1}^n y_i^2 \frac{(\sum_{i=1}^n y_1)^2}{n}\right)}} \text{ (Drapper \& Smith, 1966)} \qquad [6]$$

Note: r is the correlation coefficient, x_i is the independent variable (vegetation index), and y_i is the dependent variable (FRCI). Correlation coefficient values range between $-1 \leq r \leq 1$. A correlation coefficient closes to -1 indicates that there is a close relationship between the independent variable, and it has a negative slope for data distribution. Meanwhile, when it is close to +1, it shows that there is a close relationship between the free variable and the independent variable, and it has a positive data distribution slope. A regression model is prepared after the classic assumption and correlation test were done.

The canopy cover estimation model was prepared using the equation models as shown in Equation *[7–11]*

Linear	: $y = a + bx$	*[7]*
Exponential	: $y = ae^{bx}$	*[8]*
Logarithmic	: $y = a + b\ln(x)$	*[9]*
Power	: $y = aX^b$	*[10]*
Quadratic	: $y = a + bx + cx^2$	*[11]*

note: y = the percentage of canopy cover (FRCI); x = vegetation indices

After building the regression model, an equation was selected using the coefficient of determination (R^2), the standard error value (SE), and the analysis of variance. R^2 shows the ability of the independent variable to explain the dependent variable. Furthermore, it describes and expresses the level of accuracy and closeness of the independent to the dependent variable as a percentage. R^2 is calculated using the formula as shown in Equation *[12]* (Draper & Smith, 1998).

$$R^2 = \frac{MSS - RSS}{TSS} \times 100\% \qquad [12]$$

note: *MSS* is the model sum of squares (also known as explained sum of squares/*ESS*) of the linear regression

prediction minus the mean for that variable. *TSS* is the total sum of squares associated with the outcome variable obtained from the sum of the squares of the measurements minus their mean. *RSS* is the sum of the squares of the measurements minus the prediction from the linear regression.

The Standard Error value indicates the average distance of observational data from the regression line. The smaller the SE value, the more precise the model is in predicting. The Standard Error value is calculated using the formula as shown in Equation *[13]*.

$$S_{Y.X} = \sqrt{\frac{\sum (Y-\hat{Y})^2}{(n-2)}} \qquad [13]$$

note: $S_{Y.X}$ is the standard error value, Y is the actual value of canopy cover, and N-2 is the degree of freedom.

Model validation of regression All data in the LiDAR strip were utilized during the validation process using the census methods. The plots for validation must not be covered by clouds. Validation testing was performed using the correlation criteria between LiDAR FRCI and estimated FRCI value, aggregate deviation, and visual comparison of raster model estimation with LiDAR raster data display.

The correlation test between LiDAR FRCI value and the value of estimated FRCI from regression model was conducted by testing the correlation between the percent cover values of the estimator canopy model with the percent cover value of the LiDAR. The greater the correlation value, the more it shows that the value of the percent cover estimation model created has a close relationship with the percent cover value of the LiDAR header data (Drapper and Smith, 1992).

Aggregate deviation is the difference between the sum of the canopy density of estimator model and the percentage of canopy cover from LiDAR data. Good aggregate deviation values range from -1 to 1. Aggregate deviation values according to Drapper and Smith (1992) recalculated using a formula as shown in Equation *[14]*.

$$SA = \frac{\sum_{i=1}^{n} Ct - \sum_{i=1}^{n} Ca}{\sum_{i=1}^{n} Ct} \qquad [14]$$

note: *Ct* = canopy cover percentage from model estimation; *Ca* = canopy cover percentage from LiDAR

Visual comparison of the raster model It is comparative assessment of raster model estimation display and LiDAR raster display performed by visualization. The LiDAR FRCI raster display was better than the raster model estimation when visually they have a similar appearance

Selection of the best regression model It is obtained by scoring the correlation criteria and aggregate deviation values. Meanwhile, scoring is made by weighing the correlation value and the aggregate deviation. The best model has the highest scoring value. Meanwhile, the criteria for this highest score is a large correlation value and a small aggregate deviation value. Equation *[15]* and Equation *[16]* are the scoring formula (Jaya, 2010).

$$\text{Correlation score} = \left[\frac{\text{Correlation value} - \text{Min correlation value}}{\text{Max correlation value} - \text{Min correlation value}} \times (N-1) \right] + 1 \qquad [15]$$

$$\text{Aggregate deviation score} = \left[\frac{\text{Aggregate deviation value} - \text{Min aggregate deviation value}}{\text{Min aggregate deviation value} - \text{Max aggregate deviation value}} \times (N-1) \right] + 1 \qquad [16]$$

Results and Discussion

Plot sample identification The sample point was selected using purposive sampling by considering the presence of clouds, stand density based on canopy cover value, and canopy height (canopy height model). These points are scattered on the LiDAR line, in the work area of PT REKI. The distribution of sample points on the LiDAR footprint-path at PT REKI is presented in Table 1.

LiDAR data canopy cover calculations were performed by calculating the canopy cover in one Landsat pixel size of 30 m × 30 m. The percentage calculation of canopy cover was performed using the FRCI method which is based on the reflection of the first and single return that hit the trees. These returns were defined as the point cloud having a minimum height of 7 m. The number of sample observations used for generating the model were 338 sample plots with the high density sample class contributing the highest at 125 plots or 85.76% of the canopy cover per pixel. While the lowest contributor was the low density class which contributed only 99 plots and had an average of 23.67% canopy cover per pixel.

The FRCI value represents the percentage of canopy cover in one pixel with a size of 30 m × 30 m. It is strongly influenced by the specified height of the trees and the scan angle (Ma et al., 2017). The accuracy of FRCI values is influenced by the scan angle, because a large scan angle can cause inequality in tree canopy sampling. The error in its estimation can also be caused by the density of the LiDAR point. However, increasing this density does not necessarily increase the accuracy of FRCI estimation due to excessive data and the effects of classification on LiDAR points (Ma et al., 2017).

Correlation analysis Correlation analysis was performed to determine the relationship between the independent and the dependent variable. Furthermore, the independent variable was the vegetation index consisting of NDVI, SRVI, and GNDVI, while the dependent was the percent cover value of the LiDAR data from FRCI. Correlation value was obtained through Pearson correlation value. Pearson correlation test results between FRCI with vegetation index are presented in Table 2.

The correlation value for all vegetation indices on the overall index is positive, meaning that there is a positive linear relationship between independent and dependent variables. Therefore, an increase in the value of the independent variable will be followed by a simultaneous increase in the value of the dependent variable. Based on Table 2, there is a strong correlation between the independent variables, therefore only one was used in the preparation of the model to avoid multicollinearity.

Model equation of canopy cover estimation and validation based on Figure 2 and the relationship between FRCI and vegetation indices show positive connection. This means that an increase in the value of the vegetation indices

will cause a simultaneous increase in the FRCI value. However, this only applies to certain intervals of vegetation value. Based on Figure 2, it can be seen that the NDVI value is above 0.75, whatever the FRCI value might be, the index value will always be in the range of 0.75. When the SRVI value is above 7, regardless the FRCI value might be, the index value will always be in the range of 7. When GNDVI value is in the range of 0.65, whatever FRCI value, the index will always be the same, in the range of 0.65. This can be due to the effect of the spectral sensor response factor on the vegetation.

An increase in the value of the vegetation index is not always followed by an increase in the value of FRCI, especially in a vegetation that has reached stagnant growth. The research of Yengoh et al. (2015) said, when photosynthesis has reached its maximum, there will be no change in the NDVI value because all the visible light has been absorbed.

The percentage of cover estimator model was made through a regression model with the dependent variable is the percentage of the cover canopy from LiDAR (FRCI) and the independent variable is the vegetation indexes. The regression analysis results of the percentage canopy cover estimator model are presented in Table 3.

The result of the validation test in the correlation criteria between canopy cover value from LiDAR and canopy cover estimation from models shows that they have an average correlation of 0.554. The highest value was recorded in the R7 model with a correlation value of 0.663, while the lowest was in the R12 model with a value of 0.389.

Correlation value reflects the relationship between percentage of canopy cover from models and LiDAR. When

Table 1   Plot sample distribution

Canopy cover class	Number of sample	Average canopy cover (% pixel^{-1})
Low (10%−39.6%)	99	23.67
Medium (39.6%−69.26%)	114	55.48
High (>69.26%)	125	85.76
Total number of samples	338	

Table 2   The correlation between the independent and dependent variable

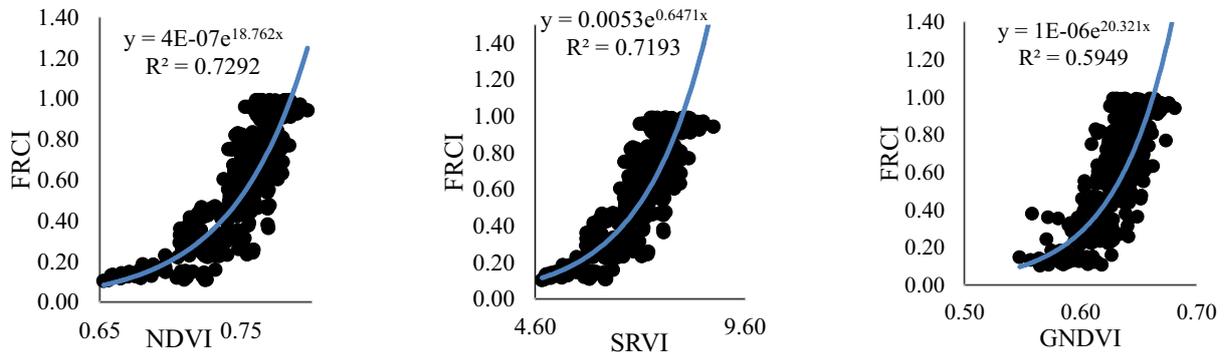|	FRCI	NDVI	SRVI	GNDVI
FRCI	1.00			
NDVI	0.80	1.00		
SRVI	0.82	0.99	1.00	
GNDVI	0.75	0.92	0.93	1.00

Figure 2 Correlation between FRCI and vegetation indices: (a) NDVI, (b) SRVI, (c) GNDVI.

Table 3 Canopy cover model equation and validation test

Code	Equation	s	R^2 (%)	F_{actual}	Correlation*	Aggregate deviation*
R1	$FRCI = -5.24 + 7.77NDVI$	0.159	64.70	615.69	0.658	-0.191
R2	$FRCI = -1.35 + 0.275SRVI$	0.153	67.30	691.64	0.605	-0.143
R3	$FRCI = -4.88 + 8.69GNDVI$	0.177	56.30	432.44	0.576	-0.187
R4	$FRCI = -2.42 + 5.34NDVI^2$	0.157	65.40	635.58	0.651	-0.184
R5	$FRCI = -0.440 + 0.0204SRVI^2$	0.152	67.50	698.04	0.562	-0.132
R6	$FRCI = -2.20 + 7.03GNDVI^2$	0.176	56.50	436.54	0.563	-0.184
R7	$FRCI = 2.22 + 5.63Ln(NDVI)$	0.161	63.90	594.03	0.663	-0.182
R8	$FRCI = -2.94 + 1.81Ln(SRVI)$	0.156	66.10	656.35	0.639	-0.169
R9	$FRCI = 3.07 + 5.36Ln(GNDVI)$	0.177	55.90	426.61	0.588	-0.200
R10	$FRCI = 0.0000004e^{(18.8NDVI)}$	0.316	72.90	904.65	0.497	-0.080
R11	$FRCI = 0.00530026e^{(0.647NDVI)}$	0.322	71.90	860.83	0.395	-0.090
R12	$FRCI = 0.0000014e^{(20.3X)}$	0.387	59.50	493.51	0.389	-0.221
R13	$FRCI = 26.576NDVI^{890911.17}$	0.318	72.70	895.63	0.515	-0.169
R14	$FRCI = 0.00011SRVI^{75.19}$	0.317	72.90	902.39	0.454	-0.113
R15	$FRCI = 174.164GNDVI^{296558.57}$	0.386	59.60	496.29	0.406	-0.180

Note: Ftable = 3.87 ; *=Validation test criteria

this value is close to 1, it indicates a close relationship which means that the estimator value is getting closer to the actual value, showing that the model is getting better.

Aggregate deviation is the difference between the percentage of canopy cover from models and LiDAR data. Deviation values have an average of -0,162. The largest aggregate deviation value was recorded in the R10 model with a value of -0.08, while the lowest was in the R12 model with a value of -0.221. Negative values have a tendency causing a deviation which tends to be lower than the actual value (underestimation). The raster display shows that of the 15 models built, only 4 (R1, R3, R7 and R9) have a raster look that resembles a LiDAR raster.

Selection of the best regression model The best model was selected by scoring using correlation criteria and aggregate deviation values. The best model has the highest correlation value criteria and the smallest aggregate deviation value. The results of the scoring are presented in Table 4.

The best canopy density estimation model according to Table 4 is the R7 model because, in addition with the NDVI, it has the highest score compared to other models. Furthermore, R7 has the equation, $FRCI = 2.22 + 5.63Ln$ (*NDVI*). Prasetyo et al., (2019) found that NDVI has a good correlation with FRCI. Regression between LiDAR FRCI and FRCI models is intended to see how far FRCI values can explain the LiDAR FRCI and the errors generated from the regression model. The regression obtained the equation LiDAR FRCI = 0.332 + 0.712 FRCI R7, with 0.0150 as standard error and R^2 at 43.9%. This shows that the LiDAR FRCI can be explained by the best FRCI model as 43.9%, and the resulting error is at 56.3%. These results indicate that there are still sizable errors because the FRCI variable can only be explained with LiDAR FRCI variable at 43.9%.

Discussion

Based on the results of model testing and validation, the $FRCI = 2.22 + 5.63Ln$ (*NDVI*) model was obtained as the best to explain the relationship between LiDAR FRCI and vegetation index. The equation has an R^2 of 43.9%. Based on this, it can be indicated that there are still some errors in the equation model. The error is allegedly caused by several factors such as the difference in acquisition time and spatial differences between LiDAR and satellite imagery. LiDAR was acquired on October 2014, while the Landsat imagery, which was used to create a vegetation index, was acquired on July on the same year.

Knowingly that PT REKI is a secondary forest, it is very possible that changes in land cover will occur, especially in the bushes. LiDAR and Landsat have spatial resolutions of 0.5 and 30 m respectively. In addition, very detailed spatial resolution in LiDAR data enables a more accurate calculation of percentage of cover canopy and better identification between tree canopy and shrubs. However, calculating vegetation index values at 30 m will result resolution in more general values and cannot separate tree and non-tree class of vegetation.

Comparison between LiDAR raster display and the best chosen model This comparison was made to see the accuracy of the model in estimating the percentage of canopy cover. Figure 3 shows a comparison of the best LiDAR raster and landscape model displays in some regions in the LiDAR pathway.

Figure 3 shows the inconsistency of the model in making estimations. This can be seen in some areas that have underestimated and overestimated values, indicated by red and blue circles, and which result in a low correlation value between LiDAR FRCI and the best FRCI model. Overall, the

Table 4 Scoring result

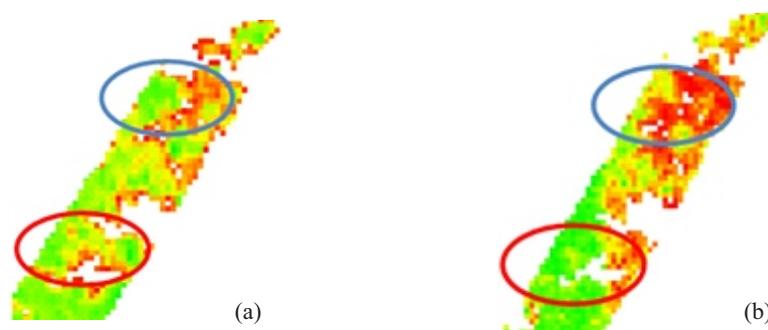| Code	Aggregate deviation score	Correlation score	Total score	Ranking
R7	4	4	8	1
R1	3.832	2.590	6.422	2
R3	1	3.215	4.215	3
R9	1.404	1	2.404	4

Figure 3 Comparison of inconsistent pixel display between the best models and LiDAR. (a) The best raster model segment 15 in LiDAR line; (b) LiDAR raster segment of LiDAR line.

best model was underestimated and therefore had low accuracy. Furthermore, its accuracy and consistency are influenced by several factors, including the difference in spatial resolution and in acquisition time between LiDAR and Landsat.

The effect on different spatial resolution between LiDAR and landsat imagery To see the effects, it was performed by comparing several pixels on Landsat used manual interpretation with those on LiDAR, with the same land cover. For example, several pixels were taken in the segment 1 of the LiDAR line. A comparison of the effects of spatial resolution between LiDAR and Landsat is presented in Figure 4.

Figure 4 shows some differences between the pixels found in LiDAR, and in Landsat, especially in the model. The appearance of raster in the model image tends to have a more yellowish color when compared to the LiDAR raster. This indicates that the estimated value of the model was actually underestimated when compared to LiDAR. Besides the different raster displays, the effects of different spatial resolutions are also seen through the FRCI values on LiDAR, which are calculated using different resolutions, a resolution of 30 m and 0.5 m. A comparison of FRCI values is presented in Table 5.

Table 5 shows the differences between LiDAR FRCI and FRCI values in Landsat. Furthermore, FRCI values in LiDAR have differences, when calculated at different resolutions. LiDAR FRCI values at 30 m resolution tend to have higher values when compared to the same values at a resolution of 0.5 m. In addition, the value of FRCI on LiDAR with a resolution of 0.5 m tends to have values that are closer to the FRCI model. This proves that the LiDAR spatial resolution used can affect the accuracy value of the model. The higher the spatial resolution of the image, the lower the effect of the problem of mixing pixels and has the potential to extract much more detailed information about land use or land cover structures.

The effect on differeny acquisition date between LiDAR and landsat imagery To see the effect of the different acquisition date on the model prediction, several pixels in Landsat were compared with those in LiDAR raster on different land covers. For example, several pixels were taken in the segment 1 of the LiDAR footprint-path. This comparison is shown in Figure 5.

Figure 5 shows the difference between LiDAR and Landsat pixels. Landsat and LiDAR pixels were acquired on July and October 2014, this times differences resulted in differences in land cover between LiDAR and Landsat

(a) LiDAR raster (b) Landsat 8 OLI in 6-5-4 composite (c) Best chosen model raster
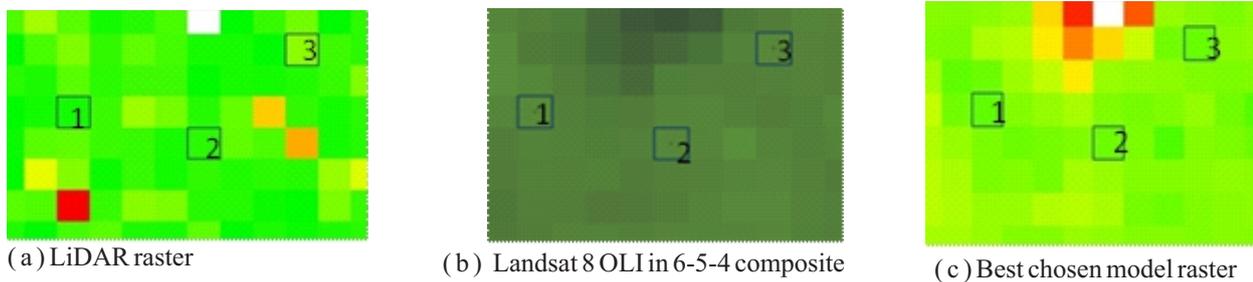
Figure 4 Pixel comparison between LiDAR and Landsat in an effect of a different spatial resolution

Table 5   Comparison of LiDAR FRCI and Landsat at different spatial resolutions

Pixel	FRCI value from LidaR (%)		FRCI from the best chosen model (%)
	30 m resolution	0.5 m resolution	
1	99.80	88.30	85.30
2	99.50	88.40	80.90
3	96.10	69.40	76.80

(a) LiDAR raster (b) Landsat 8 OLI in 6-5-4 composite (c)Best chosen model raster
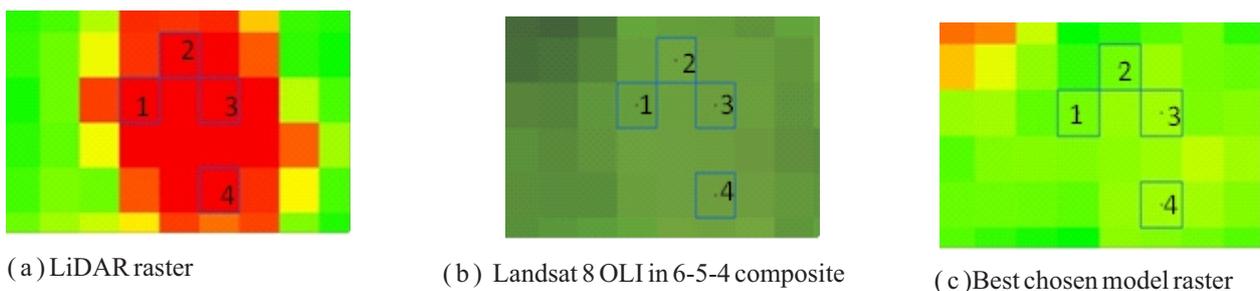
Figure 5 Pixel comparison between LiDAR and Landsat in an effect of a different acquisition date.

Table 6 Comparison of LiDAR FRCI and Landsat at different acquisition times

Pixel	FRCI value from LidaR (%)		FRCI from the best choosen model (%)
	30 m resolution	0.5 m resolution	
1	0	0	72.00
2	0	0	72.80
3	0	0	58.70
4	0	0	57.50

pixels. In the case of the Landsat pixels, Figure 5 (b) shows a green color that indicates land cover with vegetation, while in Figure 5 (a), the red colored indicates a bare land cover. This results in differences in the value of FRCI on LiDAR and Landsat.

The followings are the differences in FRCI values due to differences in acquisition time in Table 6.

Table 6 shows that the value of LiDAR FRCI, with a spatial resolution of 30 and 0.5 m does not have a different FRCI value, but when it is compared to the FRCI values from the model, the difference is quite significant. The difference FRCI value between Lidar and model is due to the difference in acquisation time, as within that period, significant changes occurred in the level of vegetation in the forest.

Conclusion

Canopy cover measurement in the field is difficult. LiDAR makes the measurement easier and faster, as well as the coverage can be wider. So that building models and validation through Landsat/SPOT could be extended to larger areas. The research is set to examine the capability of vegetation indices in estimating forest canopy cover in dry lowland forests in Sumatera Selatan, Indonesia. This research concludes that the best regression model is the R7 model. The equation of the regression is $FRCI = 2.22 + 5.63Ln(NDVI)$. It has R^2_{adj} of 63.80%, standard error of 0.161, correlation between actual FRCI and the alleged FRCI of 0.663, aggregate deviation of -0.182, and an error of 56.10%. This model was underestimated and, therefore, has low accuracy. Furthermore, its accuracy and consistency are influenced by the differences in spatial resolution and acquisition time between LiDAR and Landsat.

Acknowledgment

References

Ahmed, O. S., Franklin, S. E., & Wulder, M. A. (2014). Integration of LiDAR and Landsat data to estimate forest canopy cover in coastal British Columbia. *Photogrammetric Engineering Remote Sensing*, *80*, 953–961. https://doi.org/10.14358/PERS.80.10.953

Chen, G., Hay, G. J., & St-Onge, B. (2012). A GEOBIA framework to estimate forest parameters from LiDAR transects, Quickbird imagery and machine learning: A case study in Quebec, Canada. *International Journal of Applied Earth Observation Geoinformation*, *15*, 28–37. https://doi.org/10.1016/j.jag.2011.05.010

Das, K. R., & Imon, A. (2016). A brief review of tests for normality. *American Journal of Theoretical Applied Statistics*, *5*, 5–12. http://doi.org/10.11648/j.ajtas.20160 501.12

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3th ed.). New York: John Wiley & Sons.

[FAO] Food and Agriculture Organization. (2000). *On definitions of forest and forest change*: *FRA Working Paper 33*. Rome: FAO.

Furno, M. (2005). The Glejser test and the median regression. *Sankhyā: The Indian Journal of Statistics*, *67*, 335–358.

Glejser, H. (1969). A new test for heteroscedasticity. *Journal of the American Statistical Association*, *64*, 316–323.

Green, E. P., Mumby, P. J., Edwards, A. J., & Clark, C. D. (2000). *Remote sensing handbook for tropical coastal management* (A. J. Edwards, Ed.). Paris: United Nations Educational, Scientific and Cultural Organization (UNESCO).

Hudak, A. T., Lefsky, M. A., Cohen, W. B., & Berterretche, M. (2002). Integration of LiDAR and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, *82*, 397–416. https://doi.org/10.1016/S0034-4257(02)00056-1

Hudjimartsu, S., Prasetyo, L. B., Setiawan, Y., & Suyamto, D. (2017). Illuminating modelling for topographic correction of Landsat 8 and Sentinel-2A imageries. *European Modelling Symposium* (*EMS*), *2017*, 95–99, https://doi.org/10.1109/EMS.2017.27

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Hyyppa, J., Kelle, O., Lehikoinen, M., & Inkinen, M. (2001). A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience Remote Sensing*, *39*, 969–975. http://doi.org/10.1109/36.921414

Irlan, Saleh, M. B., Prasetyo, L. B., & Setiawan, Y. (2020).

Evaluation of tree detection and segmentation algorithms in peat swamp forest based on LiDAR point clouds data. *Jurnal Manajemen Hutan Tropika, 26*, 123–132. https://doi.org/10.7226/jtfm.26.2.123

Jakubowski, M. K., Guo, Q., & Kelly, M. (2013). Tradeoffs between LiDAR pulse density and forest measurement accuracy. *Remote Sensing of Environment, 130*, 245–253. https://doi.org/10.1016/j.rse.2012.11.024

Jaya, I. N. S. (2010). *Analisis citra digital: Perspektif penginderaan jarak jauh untuk pengelolaan sumberdaya alam*. Bogor: IPB Press.

Jaya, I. N. S., Saleh, M. B., Noventasari, D., Santi, N. A., Anggraini, N., Sutrisno, D., ..., & Qian, L. (2019). The examination of the satellite image-based growth curve model within mangrove forest. *Jurnal Manajemen Hutan Tropika, 25*, 44–50. https://doi.org/10.7226/ jtfm.25.1.44

Jennings, S., Brown, N., & Sheil, D. (1999). Assessing forest canopies and understorey illumination: Canopy closure, canopy cover and other measures. *Forestry: An International Journal of Forest Research, 72*, 59–74.

Jeronimo, S., Kane, van R., Churcill, D. J., Mcgaughey, R. J., & Franklin, J. F. (2018). Applying LiDAR individual tree detection to management of structurally diverse forest landscapes. *Journal of Forestry, 116*(4), 336–46. https://doi.org/10.1093/jofore/fvy023

Kim, E., Lee, W. K., Yoon, M., Lee, J. Y., Lee, E. J., & Moon, J. (2016). Detecting individual tree position and height using airborne LiDAR data in Chollipo arboretum, South Korea. *Terrestrial, Atmospheric and Oceanic Sciences, 27*(4), 593–604. https://doi.org/10.3319/TAO.2016. 03.29.01(ISRS)

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari, 4*, 83–91.

Korhonen, L., Korpela, I., Heiskanen, J., & Maltamo, M. (2011). Airborne discrete-return LiDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment, 115*, 1065–1080. https://doi.org/10.1016/ j.rse.2010.12.011

Korhonen, L., & Morsdorf, F. (2014). *Forestry applications of airborne laser scanning*. New York: Springer.

Ma, Q., Su, Y., & Guo, Q. (2017). Comparison of canopy cover estimations from airborne LiDAR, aerial imagery, and satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing, 10*, 4225–4236. http://doi.org/10.1109/JSTARS.2017. 2711482

Nakamura, A., Kitching, R. L., Cao, M., Creedy, T. J., Fayle, T. M., Freiberg, M., ..., & Ma, K. (2017). Forests and their canopies: Achievements and horizons in canopy science. *Trends in Ecology Evolution, 32*, 438–451. https://doi.org/10.1016/j.tree.2017.02.020

Prasetyo, L. B., Nursal, W. I., Setiawan, Y., Rudianto, Y., Wikantika, K., & Irawan, B. (2019). Canopy cover of mangrove estimation based on airborne LiDAR & Landsat 8 OLI. *IOP Conference Series: Earth and Environmental Science, 335*, 012029. https://doi.org/ 10.1088/1755-1315/335/1/012029

[REKI] Restorasi Ekosistem. (2020). *RKUPHHK-Restorasi Ekosistem 2011–2020*. Jambi: PT REKI.

Richter, R., Kellenberger, T., & Kaufmann, H. (2009). Comparison of topographic correction methods. *Remote Sensing, 1*, 184–196. https://doi.org/10.3390/rs1030184

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. In *Third Earth Resources Technology Satellite-1 Symposium* (pp. 309–317). NASA Goddard Space Flight Center.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia Analgesia, 126*, 1763–1768. https://doi.org/ 10.1213/ANE.0000000000002864

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics, 19*, 279–281.

Tan, B., Masek, J. G., Wolfe, R., Gao, F., Huang, C., Vermote, E. F., ..., & Ederer, G. (2013). Improved forest change detection with terrain illumination corrected Landsat images. *Remote Sensing of Environment, 136*, 469–483. https://doi.org/10.1016/j.rse.2013.05.013

Yengoh, G. T., Dent, D., Olsson, L., Tengberg, A. E., & Tucker III, C. J. (2015). *Use of the normalized difference vegetation index (NDVI) to assess land degradation at multiple scales: Current status, future trends, and practical considerations*. New York: Springer.

Zhen, Z., Quackenbush, L. J., & Zhang, L. (2016). Trends in automatic individual tree crown detection and delineation Evolution of LiDAR data. *Remote Sensing, 8*, 333. https://doi.org/10.3390/rs8040333