CANOPY COVER ESTIMATION BASED ON SUPPORT VECTOR REGRESSION DERIVED FROM LIDAR & LANDSAT 8 OLI

Sahid Hudjimartsu (1,2), Felliks Tampinongkol (1), Yoga Rudianto (1), Yudi Setiawan (1), Lilik Budi Prasetyo (1)

[1] Forests2020 Programme, Environmental Analysis and Spatial Modelling Laboratory
Department of Forest Resource Conservation, Faculty of Forestry,
Bogor Agricultural University Bogor, West Java, Indonesia
[2] Geoinformatic of Informatics Engineering Study Program, Faculty of Engineering and Science, Ibn Khaldun University, Jalan KH Soleh Iskandar KM 2, Tanah Sereal, Bogor, West Jawa, Indonesia
Email: hudjimartsu_syabil02@apps.ipb.ac.id; felliks_ilkom2016@apps.ipb.ac.id; yoga_rudianto@apps.ipb.ac.id; setiawan.yudi@apps.ipb.ac.id; lbprastdp@apps.ipb.ac.id

ABSTRACT: Light detection and ranging (LiDAR) provides geospatial information with massive amounts of data to use in a variety of applications. Canopy cover percentage of tree crown is one application that is estimated through LiDAR data. This study aims to estimate a canopy cover derived from LiDAR used as reference data with multispectral Landsat 8 OLI. Statistical or machine learning approaches can be used to estimate a canopy cover. In statistical inference, the form of distribution chosen by the analyst and parameter is estimated from the data. This will be a problem in the construction function for the National scale, due to many ecosystems and then will have many functions for each ecosystem. Machine learning (ML) methods, in contrast, are various kinds of algorithms that are used to learn the mapping function or classification rules inductively, directly from the training data. ML can solve problems related to big data and global method and it will be used as a methodical approach to estimate canopy cover on the National scale. In this study, we use a support vector regression (SVR) as ML method to estimate the canopy cover. The estimated percentage of canopy cover will be used as one of the parameters to classify a forest.

KEY WORDS: Canopy cover, Light detection and ranging (LiDAR), National scale, Statistical, support vector regression (SVR)

1. INTRODUCTION

Canopy cover is one of the parameters to classify an area as a forest. According to (MoEF, 2015), a forest is defined as land spanning more than 0.25 hectares with trees height more than 5 meters and a canopy cover of more than 30 percent. Canopy cover is a floor area covered by the vertical projection of tree crowns (Jennings *et al*, 1999). Measurement of canopy cover can be estimated by remote sensing, remote sensing can provide spatially representative characters of investigated forest stands in a more efficient manner (Wallace *et al*, 2016; Liu *et al*, 2017).

Light Detection and Ranging (LiDAR) is one of the technical approaches in remote sensing to estimate canopy cover percentage (Hyde *et al*, 2005; Smith *et al*, 2009; Hall *et al*, 2011). LiDAR has been the primary data source for three-dimensional information on forest vertical structure or tool for forest inventories (Leeuwen and Niuewenhuis, 2010; Liu *et al*, 2017).

The research of canopy cover using LiDAR in Indonesia seems limited, due to a very expensive and limited covered area by LiDAR data. To solve the limitation is integrating LiDAR data with Landsat 8 OLI imagery. Landsat 8 OLI is free data and covered all around the world. By integrating LiDAR and Landsat 8 OLI, we can estimate a canopy cover derived from LiDAR used as reference data with multispectral Landsat 8 OLI.

The estimation of canopy cover can be calculated by using statistical or machine learning approaches. In statistical inference, the form of distribution chosen by the analyst and parameter is estimated from the data. This will be a problem in the construction function for the National scale, due to many ecosystems in Indonesia and then will have many functions for each ecosystem.

Machine learning (ML) methods, in contrast, are various kinds of algorithms that are used to learn the mapping function or classification rules inductively, directly from the training data (Simeone, 2018). ML can solve problems related to big data and global method and it will be used as a methodical approach to estimate canopy cover on the National scale.

Support Vector Machine (SVM) is a family of classification and regression techniques in ML based on statistical learning theory (Walton, 2008; Shataee *et al*, 2012). This study applied SVM for regression or called Support Vector Regression (SVR) that can overcome the overfitting, resulting in a good performance and can be applied to various cases with continuous data (Agustina *et al*, 2018) such as canopy cover values.

The main objective of this research is to apply SVR to estimate the canopy cover with Landsat 8 OLI as parameters and the model can be used in an area beyond the ecosystem.

2. STUDY SITE AND DATA

2.1 Study Site

Three ecosystems were selected as the study sites, namely agroforestry, mangrove, and mountainous ecosystems. There are located on Sumatra Island (mangrove and mountainous) and Java Island for the agroforestry ecosystem (see Figure 1). The various ecosystems have been used in this study only three due to available data and limited area.

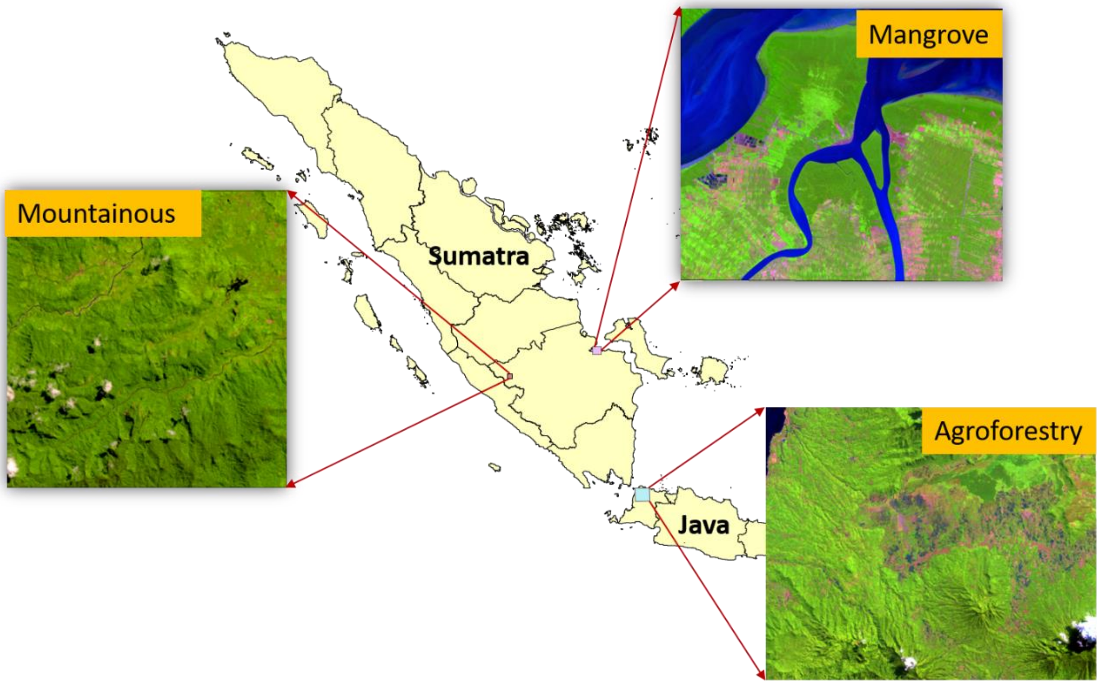


Figure 1. The Study Sites Based on Various Ecosystems In Indonesia

2.2 Landsat 8 data

Landsat 8 data is obtained from the United States Geological Survey (USGS) Landsat archive (http://earthexplorer.usgs.gov). The data over study sites with an acquisition date on 27th August 2017 for agroforestry (path/row: 123/64), 16th July 2014 for mountainous (path/row: 125/061), and 25th July 2014 for mangrove (path/row: 124/061). Meanwhile, due to various terrain elevation in study sites cause shading that can introduce errors in image understanding in general or differences illumination conditions due to solar position with respect to slope and aspect that can produce reflectance bias of pixel in the same category. To avoid reflectance bias from terrain effects due to shading, we applied an illumination algorithm to correct the reflectance from terrain effect (Tan *et al*, 2013; Hudjimartsu *et al*, 2017; Hudjimartsu *et al*, 2018).

2.3 LiDAR Data

LiDAR data were taken at different times for each ecosystem. LiDAR data for agroforestry was acquired on 11th September 2017, in October 2014 for mountainous and mangrove. The raw LiDAR data was taken by PT ASI Pudjiastuti Geosurvey and provided by GIZ-Bioclime project (mountainous and mangrove). To create forest structure (canopy cover), the data was processed using R programming and was used as reference data in estimating canopy cover using Landsat 8 OLI.

3. METHODS

This study was divided into several stages: data collection, preprocessing data, implementation SVR algorithm to estimate canopy cover and validation (see Figure 2).

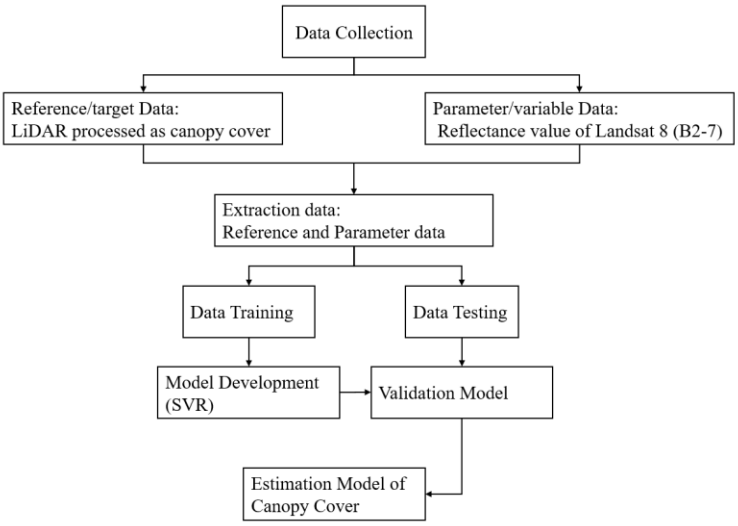


Figure 2 Research Flow

3.1 Data Collection

Landsat 8 and LiDAR data have a time acquisition. The time acquisition of Landsat 8 selected is close to the time of LiDAR data acquisition. Furthermore, extraction Landsat 8 data reflectance (band 2 -7) based on LiDAR data (overlay/extract value).

3.2 Preprocessing Data

Preprocessing data is divided into three following phases:

1. Removing Outlier
   The data was extracted will be eliminated for outlier data using the DBSCAN algorithm.
2. Stratified Data
   At this phase, the data selected depend on the class canopy cover to represent all of the class data.
3. Data Division
   The data is divided into training data and testing data with K-fold cross-validation. Training data was used to analyze the model equations while the testing data was used to test the model equations. K-fold cross-validation is a method used to evaluate the learning algorithm by dividing the data into K-fold, as K-1 fold is used as training data and 1 fold as the testing data (Liu & Özsu 2009; Hudjimartsu, 2017). Data division with K-fold cross-validation can be

illustrated as in Figure 3.

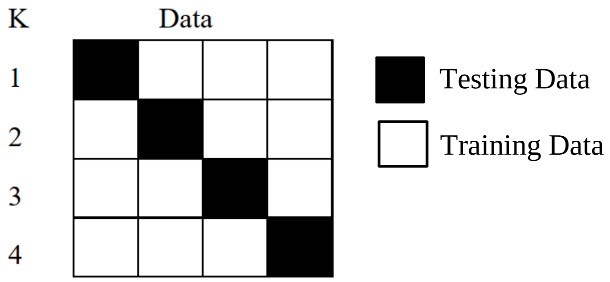
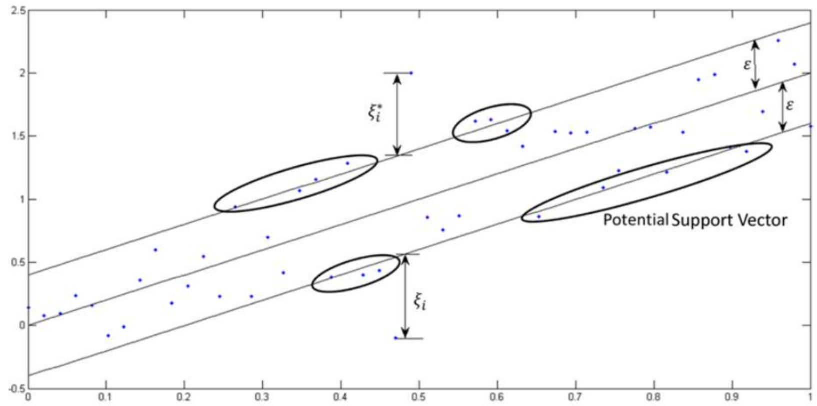


Figure 3 Dividing Data with K-fold Cross-Validation

3.3 Implementation SVR algorithm

SVM for regression (SVR) is an application for continuous data cases. SVR will search a function



of f(x) having the largest deviation ε of the reference or actual target data for all training data (Agustina *et al*, 2018). Furthermore, by using SVR, when the value of ε is equal to 0, the perfect function of regression will be obtained. SVR illustration can be seen in Figure 4.
Figure 4. SVR Illustration with One Dimensional Linear SVR (source: Awad and Khanna, 2015)

3.4 Validation

Model validation is to evaluate the performance model in estimating canopy cover by calculating the Root Mean Square Error (RMSE). RMSE measures how much error between two data sets. In other words, it compares an estimated value and an observed or actual value.

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

Where,
y = Actual or reference data
$\hat{y}$ = Estimation data
n = Total data

4. RESULT AND DISCUSSION

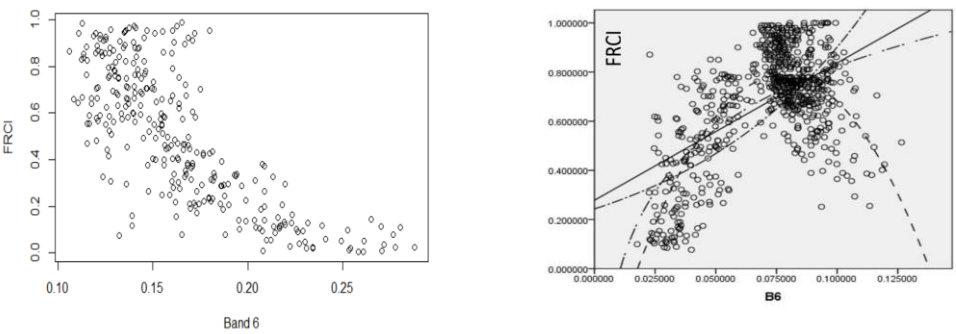
Estimation canopy cover using Landsat 8 data were analyzed using SVR, we have a 1047 sample of canopy cover percentage derived from LiDAR processed as reference data. The data was selected for the SVR model as are agroforestry ecosystem due to various types of land use compared to mangrove and mountainous ecosystems. Before the data were analyzed by SVR, we did data preprocessing as stated in the Methods section (extraction data) and can be seen in Table 1. After the extraction, the next step is stratified and removing the outlier data to represent all of the class and noise data. The results showed a reduction in sample data to 580 samples.

The sample data have been processed will be divided into training and testing data using K-fold CV. This study conducted experiments with K values to obtain better results in the distribution of the training and testing data. Based on the results of the experiment K value on K-fold CV, it was found that the number of K = 5 has better accuracy. The results of data division with K = 5 in cross-validation showed that there are 406 training data and 174 testing data. The training data were used as data for modeling using SVR.

The use of SVR (ML approach) in order to develop an estimation model for various ecosystems due to statistically will have a model for each ecosystem. Figure 5 illustrates the comparison distribution of data between agroforestry and mangrove.

Table 1. Result of Extraction Data

CC %	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
0	0.091	0.079	0.053	0.324	0.187	0.090
1	0.100	0.088	0.085	0.242	0.271	0.168
6	0.093	0.084	0.059	0.329	0.224	0.107
2	0.097	0.083	0.079	0.226	0.234	0.144
67	0.085	0.067	0.039	0.310	0.128	0.049
62	0.092	0.076	0.048	0.325	0.168	0.072
7	0.089	0.077	0.051	0.362	0.181	0.080
4	0.108	0.100	0.105	0.233	0.221	0.142
62	0.082	0.062	0.038	0.243	0.106	0.044
99	0.090	0.067	0.043	0.280	0.077	0.024
97	0.090	0.067	0.044	0.281	0.077	0.024
70	0.089	0.071	0.048	0.279	0.115	0.052
63	0.089	0.072	0.043	0.333	0.143	0.058
97	0.091	0.069	0.045	0.279	0.081	0.026
98	0.091	0.067	0.044	0.283	0.080	0.026
100	0.089	0.067	0.040	0.351	0.071	0.020
94	0.090	0.066	0.042	0.280	0.072	0.024

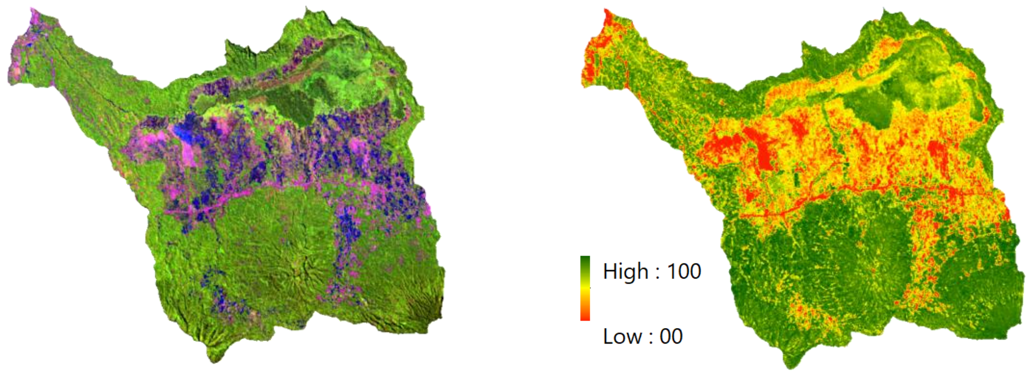


(a) (b)

Figure 5 Comparasion distribution data of agroforestry (a) and mangrove ecosystem (b)

From Figure 5 is explained that will be a different model for each ecosystem, in agroforestry the model is exponential while the quadratic model in the mangrove ecosystem (Prasetyo *et al*, 2018). Hence, SVR (ML approach) can solve that problem and the model can be applied to various ecosystems. The result of the SVR model indicates the ability to estimate the canopy cover using

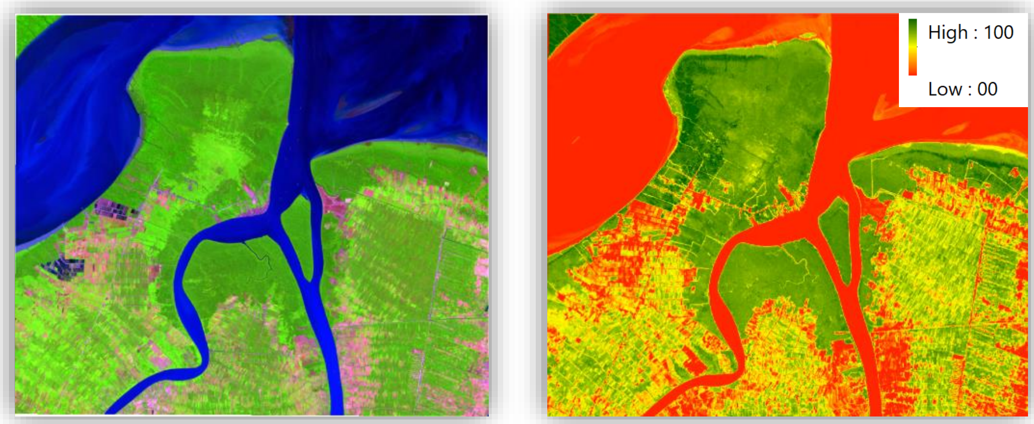
Landsat 8 data (see Figure 6).



(a) (b)

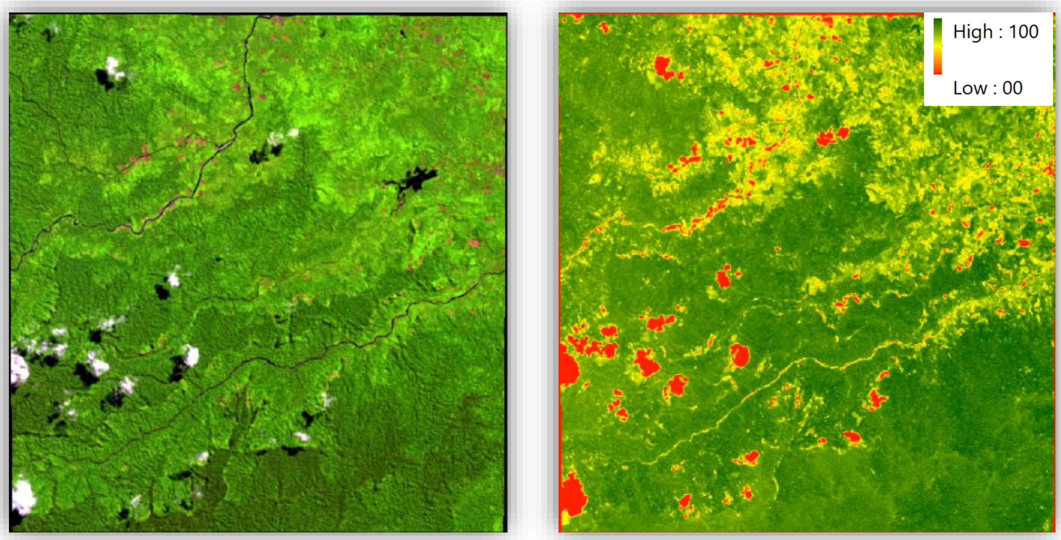
Figure 6 Combination Landsat Band 654 Vegetation (a) and Estimated Canopy Cover (b)

The evaluation model was done to find out how proper the model was able to estimate canopy cover percentage by calculating RMSE. The model evaluation used testing data and was obtained an RMSE value 0.214. This indicates the model was quietly good to estimate a canopy cover percentage. Following the SVR model was developed, we used it to estimate other ecosystems (mangrove and mountainous). Figure 7 (mangrove) and Figure 8 (mountainous) showed the model was able to estimate canopy cover with various ecosystems.



(a) (b)

Figure 7 Combination Landsat Band 654 Vegetation (a) and Estimated Canopy Cover (b)



(a) (b)

Figure 8 Combination Landsat Band 654 Vegetation (a) and Estimated Canopy Cover (b)

Validation of estimated canopy cover based on RMSE calculations for the ecosystem: mangrove obtained RSME value of 0.271 and mountainous of 0.275.

5. CONCLUSIONS

Canopy cover estimation with Landsat 8 data and LiDAR data derived the percentage of canopy cover as reference data used to develop the estimation model. The model was developed in this study uses the SVR algorithm to estimate canopy cover in various ecosystems. The results of analysis in this study revealed some conclusions which are the model quietly good for estimation of canopy cover in various ecosystems. Based on validation with RMSE calculations for the ecosystems: agroforestry obtained RMSE value of 0.214, mangrove of 0.271, and mountainous of 0.275.

ACKNOWLEDGMENT

This research was supported by the UK Space Agency (UKSA) International Partnership Program (IPP) under the Forests2020 program coordinated by Ecometrica, Ltd. The Forests 2020 is a collaborative program to advance Earth Observation applications to forest monitoring.

REFERENCES

Agustina SD, Mustakim, Okfalisa, Bella C and Ramadhan MA 2018 *J. Phys.: Conf. Ser*. 1028 012240.

Hall FG, Bergen KJB, Blair R, Dubayah R, Houghton, Hurtt G. 2011. Characterizing 3D vegetation structure from space: mission requirements. *Remote Sensing of Environment*. 115, pp. 2753–2775.

Hudjimartsu, S., Prasetyo, L. B., Setiawan, Y., and Suyamto, D. 2017. “Illumination Modelling for topographic correction of Landsat 8 and Sentinel 2A imageries”. EMS2017: European Modelling Symposium. Manchester, United Kingdom, November 20-22 2017.

Hudjimartsu, S., Prasetyo, L. B., Setiawan, Y., Suyamto, D. Ikbal, W, Kustiyo. 2018. “Automated Landsat 8 data preprocessing for national forest monitoring system”. RSCy2018: Sixth International Conference on Remote Sensing and Geoinformation of the Environment.

Hyde P, Dubayah R, Peterson B, Blair JB, Hofton M, Hunsaker C, Knox R, Walker W. 2005. Mapping forest structure for wildlife habitat analysis using wave form lidar: Validl ation of montane ecosystems. *Remote Sensing of Environment*. 96, pp. 427 – 437.

Jennings SB, Brown NB, Sheil D. 1999. Assesing forest canopies and understory illumination: canopy closure, canopy cover and other measures. Forestry. 72(1), pp. 59-73.

Liu L, Özsu MT. 2009 . *Encyclopedia of database systems*. Springer Berlin, Heidelberg, Germany.

M. Awad, R. Khanna. 2015 Support Vector Regression, Berkeley, CA:Apress, pp. 67-80.

O. Simeone. 2018. "A very brief introduction to machine learning with applications to communication systems", *IEEE Trans. on Cogn. Commun. Netw*., vol. 4, no. 4, pp. 648-664.

Smith AMS, Falkowski MJ, Hudak AT, Evans JS, Robinson AP, Steele CM. 2009. A cross-comparison of field, spectral, and lidar estimates of forest canopy cover. *Canadian Journal of Remote Sensing*, 35, pp. 447–459.

Prasetyo, L. B., Ikbal, W, Setiawan, Y, Rudianto, Y, Wikantika, K. 2018. “Canopy cover of mangrove estimation based on airborne LIDAR & Landsat 8 OLI”. ICDALC 2018: International Conference On Digital Agriculture From Land to Consumers.

Prasetyo, L. B., Ikbal, W, Setiawan, Y, Rudianto, Y, Wikantika, K. 2018. “Canopy cover of mangrove estimation based on airborne LIDAR & Landsat 8 OLI”. ICDALC 2018: International

Conference On Digital Agriculture From Land to Consumers.
Van Leeuwen M. and Nieuwenhuis M. 2010. Retrieval of forest structural parameters using LiDAR remote sensing. Eur J For Res 129, pp. 749–770.
Wallace, L., A. Lucieer, Z. Malenovský, D. Turner, and P. Vopěnka. 2016. "Assessment of Forest Structure Using Two UAV Techniques: A Comparison of Airborne Laser Scanning and Structure from Motion (SfM) Point Clouds." Forests 7 (3), pp. 1–16.
Walton, Jeffrey T. 2008. Subpixel urban land cover estimation: comparing cubist, random forests, and support vector regression. Photogrammetric Engineering & Remote Sensing. 74(10), pp. 1213-1222.