

# ESTIMATION OF MANGROVE FOREST ABOVEGROUND BIOMASS USING MULTISPECTRAL BANDS, VEGETATION INDICES AND BIOPHYSICAL VARIABLES DERIVED FROM OPTICAL SATELLITE IMAGERIES: RAPIDEYE, PLANETSCOPE AND SENTINEL-2

A. B. Baloloy <sup>a\*</sup>, A. C. Blanco <sup>a,b</sup>, C. G. Candido <sup>a</sup>, R. J. L. Argamosa <sup>a</sup>, J. B. L. C. Dumalag <sup>a</sup>, L. L. C. Dimapilis <sup>a</sup>, E. C. Paringit <sup>b</sup>

<sup>a</sup> Training Center for Applied Geodesy and Photogrammetry, University of the Philippines, Diliman, Quezon City, 1001, Philippines  
- alvinbbaloloy@gmail.com, acblanco@upd.edu.ph

<sup>b</sup> Department of Geodetic Engineering, University of the Philippines, Diliman, Quezon City, 1001, Philippines

## Commission III, WG III/10

**KEY WORDS:** Aboveground Biomass, RapidEye, Sentinel 2, PlanetScope, Mangroves

### ABSTRACT:

Aboveground biomass estimation (AGB) is essential in determining the environmental and economic values of mangrove forests. Biomass prediction models can be developed through integration of remote sensing, field data and statistical models. This study aims to assess and compare the biomass predictor potential of multispectral bands, vegetation indices and biophysical variables that can be derived from three optical satellite systems: the Sentinel-2 with 10m, 20m and 60m resolution; RapidEye with 5m resolution and PlanetScope with 3m ground resolution. Field data for biomass were collected from a *Rhizophoraceae*-dominated mangrove forest in Masinloc, Zambales, Philippines where 30 test plots (1.2 ha) and 5 validation plots (0.2 ha) were established. Prior to the generation of indices, images from the three satellite systems were pre-processed using atmospheric correction tools in SNAP (Sentinel-2), ENVI (RapidEye) and python (PlanetScope). The major predictor bands tested are Blue, Green and Red, which are present in the three systems; and Red-edge band from Sentinel-2 and Rapideye. The tested vegetation index predictors are Normalized Differenced Vegetation Index (NDVI), Soil-adjusted Vegetation Index (SAVI), Green-NDVI (GNDVI), Simple Ratio (SR), and Red-edge Simple Ratio (SRre). The study generated prediction models through conventional linear regression and multivariate regression. Higher coefficient of determination ( $r^2$ ) values were obtained using multispectral band predictors for Sentinel-2 ( $r^2 = 0.89$ ) and PlanetScope ( $r^2 = 0.80$ ); and vegetation indices for RapidEye ( $r^2 = 0.92$ ). Multivariate Adaptive Regression Spline (MARS) models performed better than the linear regression models with  $r^2$  ranging from 0.62 to 0.92. Based on the  $r^2$  and root-mean-square errors (RMSE's), the best biomass prediction model per satellite were chosen and maps were generated. The accuracy of predicted biomass maps were high for both Sentinel-2 ( $r^2 = 0.92$ ) and RapidEye data ( $r^2 = 0.91$ ).

## 1. INTRODUCTION

Mangroves have a wide range of economic, social and environmental benefits often referred to as ecosystem services. Like other vegetated coastal ecosystems, mangroves are important blue carbon sinks with a storage capacity between 990 and 1074 t C ha<sup>-1</sup> (Donato et al., 2011). In the tropics, mangroves are among the carbon-rich forests with an average storage of 1023 t C ha<sup>-1</sup> (Laffoley & Grimsditch, 2009). The greatest carbon pool in a tree is the aboveground biomass which refers to the living biomass above the soil including the stems, bark, branches, foliage, and seeds. It is usually measured for carbon flux monitoring (Vashum & Jayakumar, 2012), carbon stock quantification (Kumar and Mutanga, 2017) and for developing carbon policies and forest management protocols.

Traditional approach to field biomass estimation of mangroves is limited to the spatial constraints of data collection and inaccessibility of mangroves stands. A common non-destructive approach is the use of allometric equations derived from parameters such as diameter at breast height (DBH). Remote sensing served as a non-destructive alternative for a more robust, continuous and spatially explicit biomass assessment (Herold and Johns, 2007). The availability of different remote sensing systems led to increased capability for biomass estimation. Optical remote sensing systems offers global coverage which is

often cost effective. For regional scale, aboveground biomass estimation is usually carried using optical platforms such as Landsat (Shao & Zhang, 2016; Gleason & Im, 2011), IKONOS and MODIS (Yin et al, 2015). With newer moderate resolution satellite systems, plot-level biomass estimate can also be achieved through improved imaging sensors with shorter revisit time. Among these new platforms are RapidEye (2008), Sentinel-2 (2015, 2017) and PlanetScope (2014). Sentinel-2 is a land-monitoring constellation of two identical satellite with novel spectral capabilities with a swath width of 290 km and a frequent revisit time of 5 days. The optical payload it carries has visible, near-infrared and infrared sensors, which provide a total of 13 spectral bands with 10m, 30m and 60m ground spatial resolution (ESA). Compared to Sentinel-2, RapidEye has higher resampled spatial resolution of 5 meters with revisit time of just one day. It is known as the first commercial satellite with a red-edge band in addition to the blue, green, red, and NIR bands. Prediction models using RapidEye bands were found to explain biomass variation better than Landsat (Ramoelo and Cho, 2014). PlanetScope has the least number of bands (blue, green, red, and NIR) but it has the highest spatial resolution of 3m. Fewer studies on biomass estimation were conducted using PlanetScope data compared to the other satellite imageries. No previous studies have compared the performance of these three satellite data using prediction models developed from the same field data, with focus on the common bands, indices, and biophysical factors that can be derived from these systems.

\* Corresponding author

This study aimed to evaluate the biomass prediction efficiency of multispectral bands, vegetation indices and biophysical variables derived from RapidEye, PlanetScope and Sentinel-2. Specifically, different biomass prediction models using linear regression and non-linear multivariate regression algorithms were developed in this study. Furthermore, the accuracy of each prediction model as well as the accuracy of the predicted aboveground biomass maps were assessed using field validation plots.

## 2. DATA AND METHODS

### 2.1 Study Site

The test area is a mangrove plantation located in the village of Baloganon, Masinloc in the province of Zambales (Figure 1). The site is dominated by Rhizophora species such as *R. mucronata* and *R. apiculata*. Diameter at breast height (DBH), tree heights and other field data needed to compute the aboveground biomass were collected in November 2015. A total of 1.2 ha consisting of thirty 20m x 20m plots were selected as the training data while another 0.2 ha were set aside as map validation data.

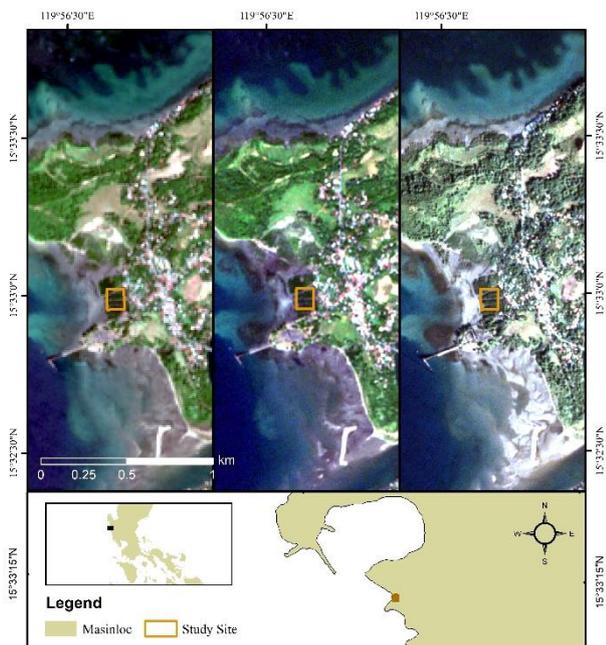


Figure 1. Location of the study site in Masinloc, Zambales, Philippines. A subset area is shown in RGB composite: Sentinel-2 (left), RapidEye (center) and PlanetScope (right)

### 2.2 Satellite Data Collection and Pre-processing

The available Sentinel-2, PlanetScope and RapidEye data acquired closest to the field data were selected (see Table 1). The Sentinel-2 Multispectral Imager Instrument (MSI) Level 1-C image covering Baloganon, Masinloc was downloaded from Sentinel Scientific Data Hub (ESA). The product is already orthorectified, georeferenced, and radiometrically calibrated into top-of-atmosphere (ToA) reflectance data. Atmospheric correction was carried using Sen2Cor standalone tool, but can be processed alternatively in the S2A Toolbox of the Sentinel Application Platform (SNAP). This processor uses image-based retrievals with Look-Up tables (LUTs) pre-calculated from the *libRadtran* model to minimize or remove atmospheric effects from level 1-C images (Main-Knorn et al., 2015). All Level-2A

bands were stacked and resampled to 10m pixel size using SNAP (ver. 5.0) geometric operation tool.

Satellite	Product level	Acquisition date
Sentinel-2A	Multispectral image-1C	Feb. 11, 2016
RapidEye	Ortho-3A	Nov. 14, 2015
PlanetScope	Ortho scene-3B	Dec. 25, 2015

Table 1. Product levels and satellite acquisition dates

The downloaded RapidEye level 3A orthoproduct has undergone radiometric, sensor and geometric correction using Ground Control Points (GCPs) and fine Digital Elevation Models (DEMs). The image was then atmospherically corrected using Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) in ENVI 5.0. Image center, illumination azimuth angle, spacecraft view angle and other correction parameters are incorporated in the RapidEye image. The PlanetScope image was downloaded as an Ortho Scene product which is orthorectified, scaled Top of Atmosphere Radiance image product (Level 3B), and delivered as analytic 4-band product (Planet Team, 2017). Conversion to ToA reflectance image were made using a Planet Labs python guide ([www.planet.com/docs/guides/quickstartndvi](http://www.planet.com/docs/guides/quickstartndvi)).

### 2.3 Multispectral Bands

There are 4 common multispectral bands (Blue, Green, Red and NIR) among Sentinel-2, RapidEye and PlanetScope; and 5 bands (+Red-edge) between Sentinel-2 and RapidEye. The Red-edge 1 of Sentinel-2 (705nm central wavelength) was chosen being the closest to the wavelength values of RapidEye's Red-edge band (690nm – 730nm). Additional bands of Sentinel including Red-edge 1-3 and SWIR 1-2 were tested as a separate group of predictor input.

### 2.4 Vegetation indices

Indices that were selected as biomass model inputs are Normalized Differenced Vegetation Index (NDVI), Soil-adjusted Vegetation Index (SAVI), Green-NDVI (GNDVI), Simple Ratio (SR) and Red-edge Simple Ratio (SRre). The first four indices were generated from the three satellite data while SRre can only be generated from Sentinel-2 and RapidEye since it requires a Red-edge band. These indices are combinations of visible, red-edge and NIR bands. The formula for each index is shown in Table 2.

Vegetation index	Formula	Reference
Normalized Difference Vegetation Index (NDVI)	$(NIR-R) / (NIR + R)$	Rouse et al, 1973
Green NDVI (GNDVI)	$(NIR-G) / (NIR + G)$	Gitelson et al, 1996
Soil Adjusted Vegetation Index (SAVI)	$\frac{(NIR-Red)}{(NIR+R+L)} \times (1 + L)$	Huete, 1988
Simple ratio (SR)	$NIR/R$	Jordan (1969)
Red-edge simple ratio (SRre)	$NIR/R-edge$	Gitelson & Merzlyak, 1994

Table 2. Equations for the vegetation indices used in the biomass models

### 2.5 Biophysical Variables

Three biophysical variables were generated from the atmospherically corrected and resampled (10m) Sentinel-2 data using SNAP tool: Leaf area index (LAI), Fractional vegetation cover (FVC), and Leaf chlorophyll content ( $C_{ab}$ ). The biophysical

variables describe the spatial distribution of vegetation state and dynamics, thus, are useful for biomass estimation (Widłowski et al., 2004). LAI, Fraction of absorbed photosynthetically active radiation (FAPAR), and FVC are the main variables computed by the SNAP toolbox using tested, generic algorithms based on specific radiative transfer models. The generation of the variables were composed of three main steps: (1) normalization of the inputs, (2) implementation of the artificial neural network (ANN) algorithm and (3) denormalization of the output and (4) generation of quality indicator (Weiss and Baret, 2016).

Rapideye and Planetscope has no biophysical variables included in their products. To facilitate comparison of these variables among the three satellite data, products such as LAI, FAPAR and  $C_a$  were derived for RapidEye and PlanetScope using available equations. LAI was obtained using Equation 1 (Zeng et al., 2000) as implemented by previous studies (Ali, 2015; Zeng et al., 2003). The  $NDVI_s$  (bare soil) and  $NDVI_v$  (dense vegetation) were selected through histogram evaluation, scaled between the lowest  $NDVI$  ( $NDVI_s$ ) and highest  $NDVI$  ( $NDVI_v$ ).

$$FVC_{NDVI} = \frac{NDVI - NDVI_s}{NDVI_v - NDVI_s} \quad (1)$$

where  $FVC_{NDVI}$  = fractional vegetation cover (NDVI-derived)  
 $NDVI_s$  = NDVI values for bare soil  
 $NDVI_v$  = NDVI values for vegetation

The output FVC derived from NDVI was then used for computing the LAI using a logarithmic equation (Norman et al., 1995) tested on large-scale field experiments:

$$LAI_{NDVI} = \frac{-\log(1 - FVC_{NDVI})}{k(\theta)} \quad (2)$$

where  $LAI_{NDVI}$  = NDVI-derived leaf area index  
 $FVC_{NDVI}$  = fractional vegetation cover  
 $k(\theta)$  = light extinction coefficient for a given solar zenith angle

The light extinction coefficient  $k(\theta)$  is a measure of attenuation of radiation in the canopy determined by the angle and spatial arrangement of the leaves. It can range between 0.4 and 0.65 in a variety of mangroves canopies. The average value between this ranges, approximately 0.5, was used in this paper as suggested by Clough (1997) and Perera et al. (2013).

Alternative data for comparison with Sentinel's Chlorophyll-a ( $C_a$ ) was obtained by using the Green Chlorophyll Index model ( $CI_{GREEN}$ ) developed by Gitelson (2003):

$$CI_{Green} = (NIR - Green) - 1 \quad (3)$$

Equations 1-3 were also applied to the Sentinel bands to compare the correlation between the modeled and the SNAP-generated biophysical products. These were labelled Sentinel-2s (SNAP generated) and Sentinel-2m for the modeled variables.

## 2.6 Model Development, Regression and Analysis

Mean data were obtained from the bands, vegetation indices and biophysical variables through zonal statistics extracted using thirty 20m x 20m polygons. The first analysis was done using linear regression models between the measured AGB and the biomass estimation predictors. The coefficient of determination values and RMSE's were recorded and were compared between input groups and among the satellite data.

The second analysis was carried using a Multivariate Adaptive Regression Splines (MARS). MARS is a regression and data mining technique developed by Friedman in 1991. This method uses basis functions in modeling the predictor and response variables. The generated basis functions will then be used as the new set of predictor variables to generate the final model. The initial step of MARS is a forward algorithm, which selects all possible basic functions and their corresponding knots. Then, a backward algorithm will discard the basis functions which do not contribute significantly to the accuracy of the fit (Friedman, 1991). The final model of MARS consists of a collection of basis functions including nonlinear and interaction relationships among the predictor variables (Bilgili et al., 2010).

Friedman (1991) suggested a maximum of 15 basis functions. In this study, a maximum of 10 basis functions was set to avoid overcomplexity of the models. Standard MARS parameter values were used for the three satellite data. Other MARS parameters are the minimum observation between knots (1), maximum interaction (10), ridge value (-7), degree of freedom for knot optimization (1) and speed factor (1). This was implemented in Salford Predictive Modeler v.8 (Salford Systems, San Diego, California, USA). The  $r^2$  and RMSE per test input group were obtained.

## 2.7 Aboveground Biomass Map and Accuracy Assessment

The best linear or multivariate biomass prediction model per satellite data were chosen based on  $r^2$  and RMSE values. The basis functions per model were converted as individual bands in ENVI 5.0 by applying the mathematical equation to the important variables per model. The stacked basis functions bands were used for the final equation. Ratio and multipliers were used to convert the output biomass product (Mg/plot) to biomass per hectare (Mg ha<sup>-1</sup>) for each pixel of the satellite data. This data was used in creating the aboveground biomass maps for Balaganon, Masinloc. The accuracy was obtained using five 20m x 20m validation plots located outside the training plots used for model development. The correlation between measured AGB from the validation plots with the predicted AGB generated using RapidEye, PlanetScope, and Sentinel-2 models was examined.

## 3. RESULTS AND DISCUSSION

The study compared the correlation between field-measured AGB and predicted AGB of three satellite data using two statistical regression methods: linear correlation and MARS algorithm.

### 3.1 Linear Regression of Aboveground Biomass

There are six input groups tested for the regression analysis: PlanetScope multispectral bands, PlanetScope derived indices, RapidEye bands, RapidEye derived indices, Sentinel-2 derived indices and Sentinel-2 bands. The best linear regression models for each satellite data are the PlanetScope SR-based model ( $r^2 = 0.56$ ), RapidEye NIR-based model ( $r^2 = 0.71$ ) and Sentinel-2 SR-based model (Table 3, Figure 2). The highest  $r^2$  and lowest RMSE (9.75 Mg ha<sup>-1</sup>) among all data inputs were obtained with the NIR band of RapidEye. SR was seen to be an efficient biomass predictor index, providing the highest  $r^2$  for both Sentinel-2 and PlanetScope data.

Among the multispectral bands input, the highest coefficient of determination values was obtained with RapidEye ( $r^2 = 0.71$ ) while Sentinel-2 and PlanetScope have equal coefficient values,

$r^2 = 0.49$ . NIR band is the most effective predictor band for both RapidEye and PlanetScope data. The SWIR 1, SWIR 2 and Red-edge bands of Sentinel data gave the lowest coefficients of determination ( $r^2 = 0.03, 0.003, 0.013$ ). The same result were reported by Castillo et al. (2017) where negative correlation ( $r$ ) with biomass were observed. Higher  $r^2$  were obtained with the Blue, Red and NIR bands. The green band gave the lowest coefficient of determination value for the PlanetScope and RapidEye, with  $r^2 = 0.20$  and  $0.06$ , respectively.

The efficiencies of all satellite data as biomass predictors are relatively higher with the use of vegetation indices with mean increase of 0.2, 0.14 and 0.19 for PlanetScope, RapidEye and Sentinel-2, respectively. This is driven by the potential of the vegetation indices to highlight plant intrinsic properties that are well related with biomass accumulation, such as leaf greenness and vigor. Each index has its specific expression which can represent green vegetation properties better than using individual bands.

Modelling Group	Predictor	Correlation, $r$	Correlation, $r^2$	(Mg/ha-1)
PlanetScope derived indices	NDVI	0.72	<b>0.51</b>	12.50
	SAVI	0.72	<b>0.52</b>	12.25
	GNDVI	0.71	<b>0.50</b>	12.50
	SR	0.75	<b>0.56</b>	11.75
PlanetScope Multispectral Bands	Blue	0.56	0.31	14.75
	Green	0.44	0.20	16.00
	Red	0.53	0.29	15.00
	NIR	0.70	<b>0.49</b>	12.75
RapidEye Derived Indices	NDVI	0.71	<b>0.50</b>	12.75
	SAVI	0.71	<b>0.50</b>	12.75
	GNDVI	0.72	<b>0.52</b>	12.25
	SR	0.78	<b>0.61</b>	11.25
	SRre	0.71	<b>0.51</b>	12.50
RapidEye Multispectral Bands	Blue	0.59	<b>0.35</b>	14.5
	Green	0.25	0.06	17.25
	Red	0.50	0.25	15.50
	Red-edge	0.76	<b>0.57</b>	11.75
Sentinel-2 Derived Indices	NIR	0.84	<b>0.71</b>	9.75
	NDVI	0.70	<b>0.49</b>	12.75
	SAVI	0.70	<b>0.50</b>	12.75
	GNDVI	0.70	<b>0.47</b>	13.00
Sentinel-2 Multispectral Bands	SR	0.73	<b>0.53</b>	12.25
	SRre	0.62	<b>0.38</b>	14.00
	Blue	0.69	<b>0.49</b>	13.00
	Green	0.51	0.26	15.25
	Red	0.67	<b>0.45</b>	13.25
	Red-edge1	0.11	0.01	17.75
	Red-edge2	0.62	<b>0.38</b>	14.00
	Red-edge3	0.62	<b>0.39</b>	14.00
	Red-edge4	0.62	<b>0.39</b>	14.00
	NIR	0.66	<b>0.44</b>	13.25
SWIR 1	0.18	0.03	17.50	
SWIR 2	0.02	0.003	17.75	

Table 3. Linear correlation of measured above-ground biomass and satellite-based predictors. Values of  $r^2$  in bold text are significant with  $p > 0.001$ .

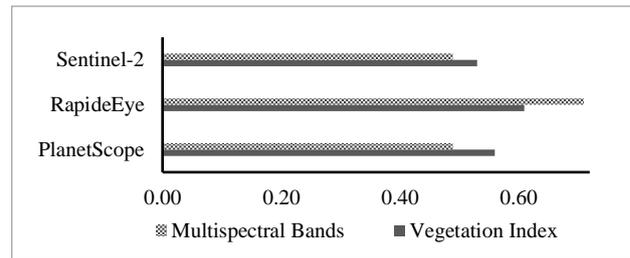


Figure 2. Highest obtained linear correlation ( $r^2$ ) of observed aboveground biomass and satellite data

### 3.2 MARS Aboveground Biomass Models

Vegetation indices, multispectral bands and an additional set of Sentinel-2 bands consisting of Red-edge 2-4 and two SWIR bands were used as inputs for MARS. Sensitivity analysis was performed to determine the most effective parameter values for the multivariate modeling. This set of parameters was applied to the different inputs group with a total of 30 samples per predictor variable.

Input Data	Satellite Data	$r^2$	RMSE (Mg ha <sup>-1</sup> )	Important Variables
Vegetation Indices (NDVI, SAVI, GNDVI, SR, SRre)	PlanetScope	0.80*	7.68	GNDVI, NDVI, SAVI
	Rapideye	0.82	7.24	GNDVI, SR, NDVI
	Sentinel-2	0.89*	5.69	SR, SRre, NDVI
Bands (B, G, R, NIR, RE**)	PlanetScope	0.80	7.78	NIR, Red, Green
	Rapideye	0.92*	4.96	NIR, Blue, Red
	Sentinel-2 Set A	0.62	10.66	Blue, Red-edge1
Additional Bands (Set A + RE2-4, SWIR 1,2)	Sentinel-2 Set B	0.84	6.94	Blue, SWIR1, Red-edge1

\*Best model per satellite data \*\* Not applicable for PlanetScope

Table 4. Important variables and correlation of measured above-ground biomass and satellite-based predictors using Multivariate Adaptive Regression Splines (MARS)

The best predictive models were obtained with the multispectral bands of Rapideye with the highest  $r^2$  value of 0.89 and smallest RMSE of 4.96 Mg ha<sup>-1</sup> (Table 4). The three important variables of RapidEye are NIR (100%), Blue (42%) and Red bands (38%). This conforms to the results of Huang et al. (2017) which reported that NIR was the most important RapidEye band for biomass estimation.

The best model for Sentinel-2 was generated with vegetation indices SR, SRre and NDVI ( $r^2 = 0.89$ ; RMSE = 5.69 Mg ha<sup>-1</sup>) (Figure 3 - top). Simple ratio has 100% variable importance in the final model. Also known as Ratio Vegetation Index, SR do not have normal distribution compared to indices such as NDVI. High efficiency of SR as biomass predictor is commonly observed in areas with closed and dense vegetation cover, including biomass estimates in tropical forests (Clerici et al., 2016). SR is also an important variable in the RapidEye index-based model and in the two best linear regression models. Meanwhile, NDVI is known to perform best in estimation of leaf biomass (Kross et al., 2015) and is usually a successful biomass predictor for a wide range of satellite data.

The vegetation index of PlanetScope generated the best model for this satellite system ( $r^2 = 0.80$ ;  $RMSE = 7.68 \text{ Mg ha}^{-1}$ ) although the coefficient of determination is lower than that of Sentinel-2 and PlanetScope. Among the important variables, GNDVI provided 100% importance to the model followed by NDVI (89%) and SAVI (56%). GNDVI was also significant for the RapidEye model; while NDVI was consistently an important index for the three satellite data.

### 3.3 Regression of Biophysical Variables

Linear regression of LAI, FVC and  $CI_{Green}$  for the three satellite systems resulted to low  $r^2$  values of LAI and FVC with measured AGB (LAI  $r^2 = 0.01$  to  $0.44$ ; FVC  $r^2 = 0.01$  to  $0.45$ ) while high  $r^2$  values were obtained between AGB and  $CI_{Green}$  ( $r^2 = 0.42$  to  $0.69$ ). The weak  $r^2$  values between AGB and LAI and between AGB and FVC were considered to be affected by the presence of undergrowth vegetation and layering in the study site. This was also the reason observed by Russel and Tompkins (2005) and Heiskanen (2006) in their test areas. It is important to note that among the modeled LAI and FVC data, PlanetScope produced higher  $r^2$  than RapidEye while RapidEye has higher  $r^2$  than Sentinel 2-m. This observation proves that canopy and understory mixing was partly reduced by using higher resolution data. Unlike these two parameters, Chlorophyll-a values derived for each pixel may not have been significantly affected by undergrowth vegetation, thus, less errors were introduced. The highest coefficient of determination among the  $CI_{Green}$ -based models was generated with RapidEye data ( $r^2 = 0.69$ ). Sentinel 2-m models performed better than Sentinel 2-s models for both LAI and FVC inputs. Between the modeled and the SNAP-generated biophysical variables, only the  $CI_{Green}$  values have high correlation ( $r = 0.81$ ,  $p > 0.001$ ).

Input Variables	Satellite Data	$r^2$	Important Variables
	Planetscope	0.47	$CI_{Green}$ , LAI, FVC,
LAI, FVC, Cab	Rapideye	0.71	$CI_{Green}$
	Sentinel 2-m	0.45	$CI_{Green}$
	Sentinel 2-s	0.53	$CI_{Green} / C_a$

Table 5. Important variables and correlation of measured above-ground biomass and biophysical variables using MARS

MARS regression of AGB and the biophysical variables resulted to higher  $r^2$  values (Table 5) than linear regression results, except for Sentinel-2m where correlation between AGB and  $CI_{Green}$  is lower.  $CI_{Green}$  was the sole important variable for Sentinel and RapidEye while all three variables were used in PlanetScope.

### 3.4 Basis Functions and Final Models

The best biomass prediction models for each satellite data were chosen based on the highest  $r^2$  and lowest RMSE value. These models are the Sentinel-2 index-based model, PlanetScope index-based model and the RapidEye multispectral bands-based model (Figure 3). The basis functions and final equation were generated by MARS. Table 6 shows the basis functions and final models for predicting the biomass using the observed AGB data.

BFs are functions used to demonstrate each distinct interval of the predictors in the form below:

$$BF_n = \max(0, x-k) \text{ or } = \max(0, k-x) \quad (4)$$

where  $BF_n$  = basis function number

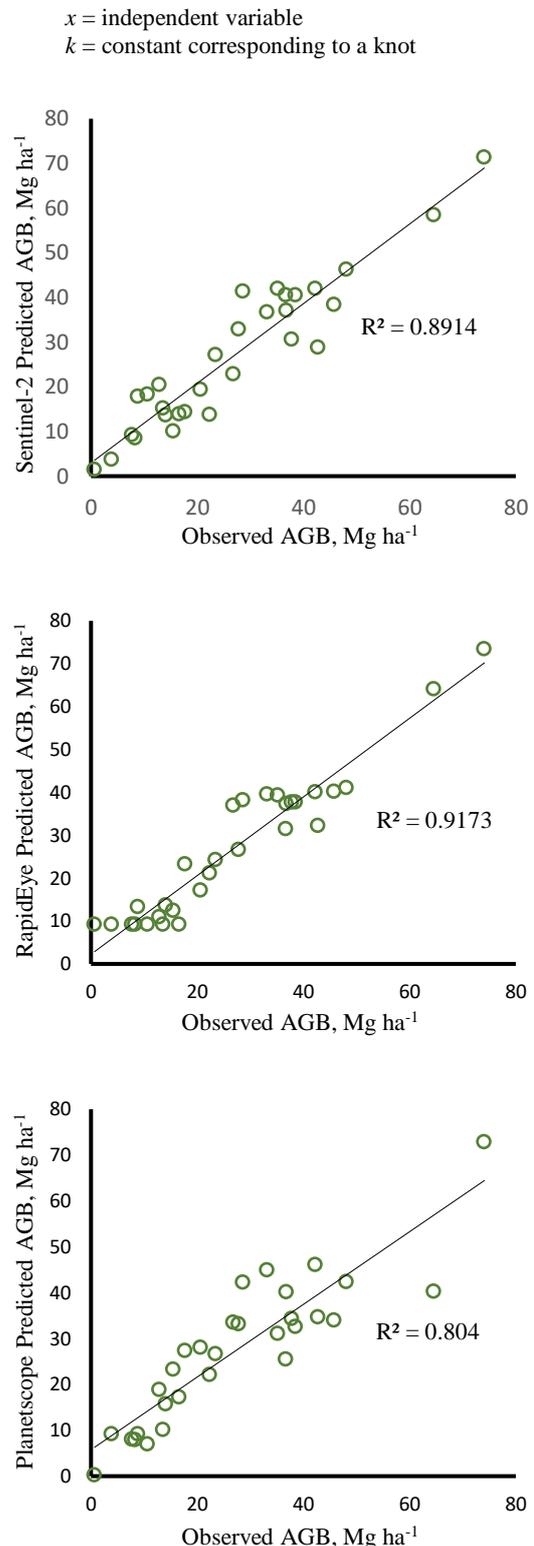


Figure 3. Best MARS regression models showing relationship of observed and predicted AGB from Sentinel (top), RapidEye (middle) and PlanetScope (bottom).

All MARS models performed better than the linear regression models. The result of MARS is consistently high using the vegetation indices of Sentinel-2, RapidEye and PlanetScope. This is partially attributable to the close association between the indices. Bilgili et al. (2010) suggested that the accuracy of MARS

models depends more on the range of data values and their degree of autocorrelation rather than the effect of spectral pre-treatment on the band reflectance data. The advantages of MARS compared to other machine algorithms includes predictive accuracy, computational speed and simplicity of interpretation (Mina and Barrios, 2010). Compared to linear regression, MARS can transform variables and identify higher order interactions between variables. MARS also performed better than multiple linear regression when applied to LiDAR-based biomass estimation (Laurin et al., 2016).

Satellite Data	Model and Basis Functions
Sentinel-2	$y = 5.54388 - 0.384761 * BF2 - 1.25011 * BF3 - 1.37807 * BF4 + 9.34749 * BF6 - 0.0339066 * BF7 - 0.69426 * BF8 + 23.3997 * BF9 + 82.4958 * BF10$ $BF2 = \max(0, 19 - SR);$ $BF3 = \max(0, SRRE - 2.75);$ $BF4 = \max(0, 2.75 - SRRE);$ $BF6 = \max(0, 0.747658 - NDVI);$ $BF7 = \max(0, SR - 10.5) * BF3;$ $BF8 = \max(0, 10.5 - SR) * BF3;$ $BF9 = \max(0, NDVI - 0.747658) * BF3;$ $BF10 = \max(0, 0.747658 - NDVI) * BF8$
RapidEye	$y = 0.366588 + 0.00333015 * BF2 + 5.50533e-005 * BF4 + 1.19337e-005 * BF5 - 6.37222e-007 * BF6 - 3.51081e-006 * BF7 - 0.0027732 * BF8$ $BF2 = \max(0, NIR - 1682.25);$ $BF4 = \max(0, BLUE - 247.25) * BF2;$ $BF5 = \max(0, 247.25 - BLUE) * BF2;$ $BF6 = \max(0, RED - 245) * BF4;$ $BF7 = \max(0, 245 - RED) * BF4;$ $BF8 = \max(0, NIR - 1906.25)$
PlanetScope	$Y = 4.34818 - 23.786 * BF2 + 20.8186 * BF4 + 71.1518 * BF5 - 16.1731 * BF7 - 779.811 * BF9$ $BF2 = \max(0, 0.613744 - NDVI);$ $BF4 = \max(0, 0.339722 - GNDVI);$ $BF5 = \max(0, GNDVI - 0.494629);$ $BF7 = \max(0, SAVI - 0.223856);$ $BF8 = \max(0, 0.223856 - SAVI);$ $BF9 = \max(0, NDVI - 0.366894) * BF8$

Table 6. Basis functions and final model generated for each satellite data to predict aboveground biomass of mangroves

### 3.5 Aboveground Biomass Maps

The basis functions of the best model per satellite data were applied to either bands or vegetation indices. The maps were generated using the native resolution of the input data (Figure 4) after converting the aboveground biomass per plot to AGB per hectare for each generated pixel. The ideal output of the pixel-based maps is that the total biomass per hectare of four Sentinel pixels will be equal to the total biomass of 16 RapidEye pixels and 44.44 PlanetScope pixels, having a similar area of 20m x 20m. Aboveground biomass totals of 690 Mg ha<sup>-1</sup>, 613 Mg ha<sup>-1</sup>, and 793 Mg ha<sup>-1</sup> were recorded for Sentinel-2, RapidEye and PlanetScope data, respectively.

### 3.6 Map Accuracy Assessment

Accuracy of maps generated by MARS algorithm were assessed by using five validation plots with an area of 20m x 20m each. These plots were not included as training data to test the biomass predictive mapping efficiency of the generated models when predictors are correlated to a response variable outside the training sample. The accuracy of predicted biomass maps cannot be evaluated by inter-comparison of maps, thus validation data is needed. The RMSEs and coefficients of determination (r<sup>2</sup>) between predicted values and field measurements were recorded (Table 7).

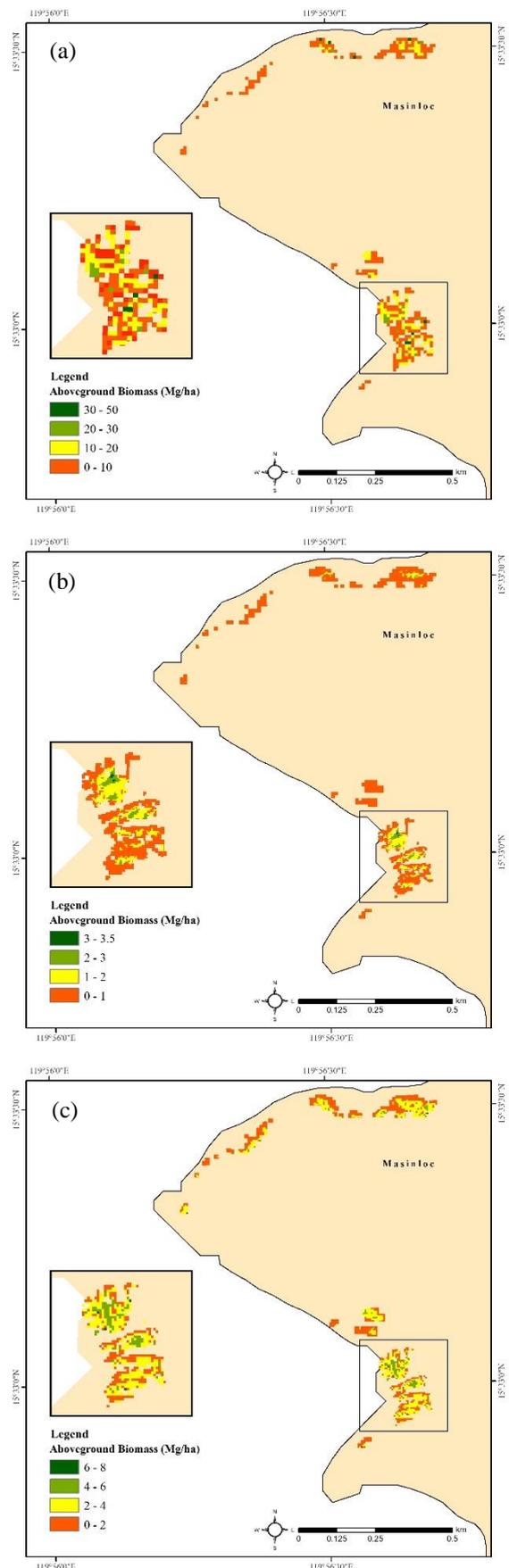


Figure 4. Predicted Aboveground Biomass Map using data from a) Sentinel-2, b) RapidEye and c) PlanetScope Imageries

Satellite Data	Correlation ( $r^2$ )	RMSE ( $\text{Mg ha}^{-1}$ )
Sentinel-2	0.92	8.79
RapidEye	0.91	4.27
PlanetScope	0.40	13.75

Table 7. Result of accuracy assessment through regression of the predicted AGB raster data and the field validation dataset

For models developed through MARS, a suggested validation technique is to test the applicability of the models to a new set of target variables. This technique will determine if the model generated are over-fitted or under-fitted with the training samples. Among the three satellite data, only the coefficient of determination values of PlanetScope reflected possible overfitting with a relatively higher coefficient in the training dataset. The accuracy, however, can be affected by the fewer number of samples used as validation data. Lower values of important vegetation indices derived from PlanetScope, significantly different with RapidEye and Sentinel-2 ( $p < 0.001$ ), may have affected the knot value computed for each basis function in the PlanetScope model.

The current study showed satisfactory result for RapidEye and Sentinel-based predictive mapping and estimation of aboveground biomass. Multivariate regression model with Sentinel-2 generated an RMSE of  $5.69 \text{ Mg ha}^{-1}$ , one of the lowest RMSE's comparable to recent studies (Castillo et al., 2016; Clerici et al, 2016). RapidEye result is satisfactory with an RMSE values of  $4.96 \text{ Mg ha}^{-1}$  and  $4.27 \text{ Mg ha}^{-1}$  from the training and validation data, respectively.

#### 4. CONCLUSION

Reliable biomass estimates are essential for obtaining the net primary production in ecological studies. Forest aboveground biomass is one of the baseline data needed for carbon stocks assessment and climate change studies.

The relationships between AGB and the set of predictor variables were established. This study have demonstrated the efficiency of the multispectral band, vegetation indices, and biophysical variables derived from three novel optical satellite imageries: Sentinel-2, PlanetScope and RapidEye. Multispectral bands are the preferred input predictors for RapidEye while derived vegetation indices are recommended when Sentinel-2 and PlanetScope were employed. Simple Ratio consistently provided high  $r^2$  for RapidEye and Sentinel both through linear and multivariate regression. NIR band is the most effective predictor band for RapidEye and PlanetScope. For Sentinel-2, the important bands are Blue and Red-edge 1. Weak linear correlations were observed between AGB and the other Red-edge bands and the two SWIR bands. However, addition of these bands (Sentinel2 Set-B) increased the coefficients from 0.62 to 0.84 in the case of MARS regression. The biophysical variables generated from Sentinel-2, PlanetScope, and RapidEye generated low coefficient of determination values except for the Green Chlorophyll Index ( $\text{CI}_{\text{Green}}$ ).

The multivariate non-parametric MARS is a robust classification method that can be used in remote sensing analysis. It is efficient in determining the relevant variables with good predictive accuracy, computational speed and flexibility with the choice of parameter values for penalty parameter and degree of freedom for knot optimization, among others. In this study, MARS models performed better than the linear regression results. As MARS was reported to be sensitive to data size and outliers, we would recommend further studies to introduce more test and validation

datasets to the algorithm. This paper is one of the few studies on mangroves biomass prediction conducted using PlanetScope data. Improvement of result with the generated PlanetScope-based models is also recommended such as assessing other vegetation indices and band ratios.

Overall, the study recommends both Sentinel-2 and RapidEye for mangrove biomass prediction due to consistently high coefficient of determination and low RMSE values based on test and validation data. Through the methods employed in this study, a plot level and pixel-based aboveground biomass estimates can be generated which can aid in mangrove management and conservation.

#### ACKNOWLEDGEMENTS

This research is funded by the DOST Philippine Council for Industry, Energy and Emerging Technology Research and Development (PCIEERD) under the Integrated Assessment and Modelling of Blue Carbon Ecosystems for Conservation and Adaptive Management (IAMBlueCECAM) Project 1: Mangrove Remote Sensing using LiDAR, RADAR, Multispectral and Hyperspectral Data (MaRS). This project is being implemented in close cooperation with The Project for Comprehensive Assessment and Conservation of Blue Carbon Ecosystems and their Services in the Coral Triangle (BlueCARES). Utmost acknowledgement is given to Nationwide Detailed Resource Assessment Using LiDAR (Phil-LiDAR 2), Project 3 for the field inventory data used in this study and PlanetLabs for providing the Planet data.

#### REFERENCES

- Ali, M., Montzka, C., Stadler, A., Menz, G., Thonfeld, F., Vereecken, H., 2015. Estimation and Validation of RapidEye-Based Time-Series of Leaf Area Index for Winter Wheat in the Rur Catchment (Germany). *Remote Sensing*. 7(3), pp2808-2831.
- Bilgili, Volkan A., van Es, H.M., Akbas, F., Durak, A., Hively, W.D., 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *Journal of Arid Environments*. 74, pp229–238.
- Castillo, J.A.A., Apan, A. A., Maraseni, T.N., Salmo, S.G. III., 2017. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*. 134: pp70-85. doi: 10.1016/j.isprsjprs.2017.10.016.
- Clerici, N., Rubiano, K., Abd-Elrahman, A., Hoestettler, J.M.P., Escobedo, F.J. 2016. Estimating Aboveground Biomass and Carbon Stocks in Periurban Andean Secondary Forests Using Very High Resolution Imagery. *Forests*. 7(7) 138p.
- Clough, B. F., Dixon, P., Dalhaus. O., 1997. Allometric relationships for estimating biomass in multi-stemmed mangrove trees. *Australian Journal of Botany*. 45(6), pp1023.
- Donato, D., Kauffman, J.B., Murdiyarso, D., Kurnianto, S., Stidham, M., Kanninen, M., 2011. Mangroves among the most carbon-rich forests in the tropics. *Nature Geoscience*. 4(5), pp.293-297
- European Space Agency. [http://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Copernicus/Sentinel-2](http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2) (December 2017).

- European Space Agency. The Copernicus Open Access Hub. Available online: <https://scihub.copernicus.eu/> (November 2017).
- Friedman, J.H., 1991. Multivariate adaptive regression splines. *Annals of Statistics*. 19, pp1-67.
- Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*. 160, pp271–282.
- Gitelson, A., Merzlyak, M.N., 1994. Spectral reflectance changes associated with autumn senescence of *A. hippocastanum* L. and *A. platanoides* L. Leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*. 143, pp286.
- Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*. 58, pp289–298.
- Gleason, C.J., 2011. In: J.H. A Review of Remote Sensing of Forest Biomass and Biofuel: Options for Small-Area Applications. *GIScience & Remote Sensing*. 48(2), pp141-170.
- Hassan, R., Scholes, R., Ash, N., 2005. Millennium Ecosystem Assessment. In *Ecosystems and Human Well-Being: Current State and Trends: Findings of the Condition and Trends Working Group*; Island Press: Washington, DC, USA.
- Heiskanen, J., 2006. Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *International Journal of Remote Sensing*. 27, pp1135–1158.
- Herold, M. and Johns, T., 2007. Linking requirements with capabilities for deforestation monitoring in the context of the UNFCCC-REDD process. *Environmental Research Letters*, 2(4), 045025.
- Huang, S., Miao, Y., Yuan, F., Gny, M.L., Yao, Y., Cao, Q., Wang, H., 1, Lenz-Wiedemann, V.I.S., Bareth, G., 2017. Potential of RapidEye and WorldView-2 Satellite data for improving rice nitrogen status monitoring at different growth stages. *Remote Sensing*. 9(3), 227. doi:10.3390/rs9030227
- Huete, A.R., 1988. A soil adjusted vegetation index (SAVI). *Remote Sensing of Environment*. 25, pp295-309.
- Jordan, C.F., 1969. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*. 50, pp663–666.
- Kross, A., McNairn, H., Lapen, D., Sunohara, M., Champagne, C., 2015. Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *Int J. Appl Earth Obs Geoinf*. 34, pp235–248.
- Kumar, L., Mutanga, O., 2017. Remote Sensing of Above-Ground Biomass. *Remote Sensing*. 9, 935.
- Laffoley, D.d'A., Grimsditch, G., (eds). 2009. The management of natural coastal carbon sinks. A short summary. IUCN, Gland, Switzerland. 8p.
- Laurin, G. V., Puletti, N., Chen, Q., Piermaria, C., Papale, D., and Valentini, R., 2016. Above ground biomass and tree species richness estimation with airborne Lidar in tropical Ghana forest. *International Journal of applied Earth Observation and Geoinformation*. 52, pp371-379.
- Main-Knorn, M., Pflug, B., Debaecker, V., Louis, J., 2015. Calibration and validation plan for the L2A processor and products of the Sentinel-2 Mission. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 40, pp1249-1255.
- McLeod, E., Chmura, G.L., Bouillon, S., Salm, R., Björ, M., Duarte, C.M., Lovelock, C.E., Schlesinger, W.H., et al. A blueprint for blue carbon: Toward an improved understanding of the role of vegetated coastal habitats in sequestering CO<sub>2</sub>. *Frontiers in Ecology and the Environment*. 9, pp552–560.
- Mina, C.D., Barrios, E.B., 2010. Profiling Poverty with Multivariate Adaptive Regression Splines. 2010. *Philippine Journal of Development*. 37.
- Norman, J.M., Kustas, W.P., Humes, K.S., Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface-temperature. *Agricultural and Forest Meteorology*. 77, pp263–293.
- Perera, K.A.R.S., Amarasinghe, M.D., Somaratna, S., 2013. Vegetation structure and species distribution of mangroves along a soil salinity gradient in a micro tidal estuary on the North-Western coast of Sri Lanka. *American Journal of Marine Science*. 1, pp7–15.
- Planet Team, 2017. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. <https://api.planet.com>.
- Ramoelo, A., Cho, M.A., 2014. Dry season biomass estimation as an indicator of rangeland quantity using multi-scale remote sensing data. AARSE. University of Johannesburg, 27-31.
- Rouse, J.W., Has, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. In *Proceedings of the Third ERTS Symposium* (NASA), Washington, DC, USA.
- Russell, W., Tompkins, R., 2005. Estimating biomass in coastal *Baccharis pilularis* dominated plant communities. *Fire Ecology*. 1, pp 20-27.
- Weiss, M., Baret, F., 2016. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER", Version 1.1. ESA. S2 Toolbox Level 2 Product algorithms v1.1. Issued 05 February 2016.
- Widłowski, J.L., Pinty, B., Gobron, N., Verstraete, M.M., Diner, D.J., Davis, A.B., 2004. Canopy structure parameters derived from multi-angular remote sensing data for terrestrial carbon studies. *Climatic Change*. 67(2-3), pp403-415.
- Zeng, X.B., Rao, P., DeFries, R.S., Hansen, M.C., 2003. Interannual variability and decadal trend of global fractional vegetation cover from 1982 to 2000. *Journal of Applied Meteorology*. 42, pp1525–1530.
- Zeng, X., Dickinson, R. E., Walker, A., Shaikh, M., DeFries, R. S., Qi, J. 2000. Derivation and evaluation of global 1-km fractional vegetation cover data for land modeling. *Journal of Applied Meteorology*. 39, pp826–839.
- Zhao, S., L. Zhang., 2016. Estimating Forest Aboveground Biomass by Combining Optical and SAR Data: A Case Study in Genhe, Inner Mongolia, China. *Sensors*. 16(6), 834.