

# Modeling Forest Carbon Estimation Using Sentinel-2 Derived Indices in Yayu Afro-Montane Forest, South West Ethiopia

Seid Muhe

Semera University

Mekuria Argaw (✉ [mekuria.argaw@aau.edu.et](mailto:mekuria.argaw@aau.edu.et))

Addis Ababa University <https://orcid.org/0000-0001-6697-6709>

---

## Research

**Keywords:** Above ground biomass, Sentinel-2 imagery, Carbon stock, Vegetation indices, Biophysical variables

**Posted Date:** October 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1003735/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Empirical analyses were common methods for forest carbon estimation. Lately, satellite images are popularly used to study different attributes of forest vegetation. Sentinel-2 image provides a significant improvement in spectral coverage, spatial resolution and temporal frequency to assess forest biomass. This study assessed the potential of vegetation indices and biophysical variables derived from Sentinel-2 images in modeling above ground biomass (AGB) and carbon stock in the Yayu forest biosphere reserve.

**Method:** About twenty variables extracted from the Sentinel-2 image were used in this study. Forest stand parameters such as DBH and tree height were used to calculate AGB using allometric equations. The correlation between the biomass values measured from plots and the variables extracted from Sentinel-2 images were examined using the Pearson correlation coefficients. A regression analysis was applied to select determinant variables for predicting AGB. The regression model results were validated based on the coefficients of determination between the observed and the predicted values.

**Results:** A strong correlation ( $r = 0.65 - 0.74$ ) was found between the biophysical variables from Sentinel-2 image and AGB measured from sampling plots. The multispectral (MS) Band 4, the biophysical vegetation variables from Sentinel-2 images were strongly correlated with the AGB. The variables MS Band 4, IRECI, LAI, FCOVER and FAPAR are good predictors of the forest AGB. The model goodness of fit between the observed and predicted values of the AGB showed a coefficient of determination ( $r^2$ ) value of 0.74 and root mean square error (RMSE) of 0.16 ton C/pixel.

**Conclusion:** The developed AGB prediction model was applied to successfully quantify and map the AGB and carbon stock of the forest in the biosphere reserve. Vegetation indices from Sentinel-2 images can effectively predict AGB in forest landscapes and can avoid costly ground surveys to quantify AGB and carbon stock in difficult terrains.

## Background

Forests play an important role in the global carbon cycle and climate change mitigation by reducing atmospheric CO<sub>2</sub> concentration (Alkama and Cescatti 2016; Georgia et al. 2017). Holding 40 % of the global terrestrial carbon, sustainable management of tropical forests is crucial for mitigating climate change and conserving biodiversity (Canadell and Raupach 2008; Mauya et al. 2015; Schuit et al. 2021). Data on forest productivity assessment, total biomass production, growth prediction and ecosystem services valuation are essential for forest management planning and utilization (Zianis and Mencuccini 2004; Soenen et al. 2010). However, data accuracy and collection methods have remained serious methodological challenges (Powel et al. 2010). Accurate data on forest biomass are needed for appropriate management decision making and monitoring. Data accuracy is a key factor for forest carbon accounting for successful implementation of carbon market mechanisms such as the REDD<sup>+</sup> (Herold et al. 2011). Techniques that facilitate rapid and accurate forest biomass estimation across

spatial and temporal scales are very useful in reducing the level of uncertainty in carbon stock assessments and for informing strategic forest management plans (Soenen et al. 2010; Mascaro et al. 2011; Pan et al. 2011; Dou and Yang 2018).

In Ethiopia, the last remaining patches of dry and moist afro-montane forests are situated in mountainous terrains and valley gorges where access is very limited (Kebede et al. 2013). Forest biomass quantification methods are mostly through direct measurement of tree parameters from sampling plots and general allometric equations (Yohannes et al. 2015; Siraj 2019; Dibaba et al. 2019). For some lowland woodlands, species-specific equations are developed through extensive inventory and using destructive methods (MoA 2000). These techniques are costly, labor intensive and time consuming. Although the results could provide more accurate estimates of the biomass, such methods have limitations for inaccessible terrains and results may be questionable for representativeness (Zianis and Mencuccini 2004; Shrestha and Nandin-Erdene 2011; Vashum and Jayakumar 2012). Ethiopia has committed ambitious national emission reduction target and needs accurate carbon accounting method, particularly for its forestry sector. Effective and efficient method of carbon estimation is instrumental for successful implementation of the REDD+ program (MEFCC 2016). Combining remote sensing data and measurements from representative sample plots (Pertille et al. 2019) on the ground has become a common approach to generate spatially explicit estimations of forest biomass (McRoberts et al. 2013). Although satellite images are utilized to quantify land use land cover changes over time, biomass quantification using vegetation indices derived from high resolution satellite images is a new development. Landsat images have been freely accessible and widely used for vegetation assessment and biomass estimation (Lyon et al. 1998; Timothy et al. 2015; Georgia et al. 2017). Nevertheless, the data saturation problems in Landsat images often caused under-estimation of forest biomass (Lu et al. 2014).

The release of freely accessible Sentinel-2 images from the European Space Agency (ESA) hub has opened new opportunity for use of vegetation indices and spectral bands for biomass quantification (Zhang et al. 2017; Castillo et al. 2019). Sentinel-2 satellite has multispectral instrument (MSI) sensor yielding image with better spectral coverage (e.g., red-edge band, shortwave infrared bands), high spatial resolution (e.g., 10m, 20m 60m) (Shoko 2017), and increased temporal frequency compared to the Landsat series (Gómez 2017; Pandit et al. 2018; Isbaex and Coelho 2020). The red-edge band in Sentinel-2 is most suitable for assessing, mapping and monitoring of vegetation characteristics (Ramoelo et al. 2015; Shoko 2017; Pertille 2019). Vegetation indices, multispectral bands and biophysical variables derived from Sentinel-2 have been tested for estimating nutrients in herbaceous biomass (Ramoelo et al. 2015), mapping of land use land cover (Forkuor et al. 2017), and mapping of biomass by coupling with data from field plot measurements (Powell et al. 2010). Studies have demonstrated that based on the strength of the relationship between directly measured biomass data from a field sampling and a remotely sensed spectral indices from satellite images, a spatially explicit above ground biomass can be modeled for an entire forest cover of a landscape (Castillo et al. 2017; Chen et al. 2018, Pandit et al. 2018). A range of multispectral bands, vegetation cover biophysical variables and indices extracted from satellite images can be used to construct a biomass estimation model by selecting best predictor

variables (Georgia et al. 2017; Pertille et al. 2019). In this study, we used correlation analysis and regression algorithm to establish above ground biomass estimation model in a natural forest. The technique has been tested in other studies and yielded best results in modeling above ground biomass estimation (Lu 2006). This study was aimed at examining the relationship between directly measured above ground forest biomass and vegetation indices as well as biophysical variables derived from Sentinel-2 Multispectral Instrument (MSI) image, and to identify best predictor variables so as to develop carbon stock predictor model for the study forest types.

## Materials And Methods

### Description of the study area

The Yayu afro-montane forest is found in the Illubabor Zone, southwest of the country at about 550 Km from the capital, Addis Ababa. The geographic location is between 8°4' 56.05"- 8° 24' 40.46" N latitude and 35° 44'53.85"- 36° 5'12.23" E longitudes (Fig. 1). Large part of the Yanu afro-montane forest is protected as a Forest Biosphere Reserve. The forest is part of the last remaining intact patches of natural forests in the southwest region. The forest has multiple economic, social and environmental benefits. It provides non-timber forest products, mainly spices, honey, and herbal medicine to rural communities for their livelihoods. The forest contains one of the largest forest biomass in the country and hence significantly contributes to climate change mitigation. Besides, the Yayu forest is one of the last remaining montane-rainforests containing wild *Coffea arabica* gene pool populations in Ethiopia. The forest site is effectively serving as an *in situ* conservation forest for the wild *Coffea arabica* population gene pool (Gole et al. 2008; Schuit et al. 2021). Coffee makes the largest share of living for the local communities. The climate is characterized by hot and humid tropical climate with a mean annual temperature of 25°C, varying between 12.7°C and 26.1°C. The region receives high mean annual rainfall of about 2100 mm, with high annual variability ranging from 1400 to 3000 (Gole et al. 2008).

The topography is complex with undulating hills and valleys dissected by several small streams draining into the Geba and Dogi Rivers. The elevation ranges between 1217 m.a.s.l at the valley bottom to 2583 m.a.s.l at the highest point in the watershed (Fig. 2). The valley gorges and the mountains are steep slopes and not easily accessible. As result, the dense and large patches of the forests are located in these parts of the landscape.

The land use land cover was mapped from a Landsat 8 dry season imagery of 2018. The forest land constitutes the largest cover with about 62 % followed by cultivated agricultural land constituting about 30 % of the total cover. The rest of the landscape is covered with shrub lands (3 %), settlements (2.7 %) and wetlands (2.3 %) (Fig. 3). Although the forest area is registered as a National Forest Priority Area and a Biosphere reserve, the local communities are highly dependent on the forest mainly for harvesting natural coffee, spices and honey production. Thus, the Biosphere reserve forest has three functional zones allowing farmers to harvest non-timber forest products in the transition and buffer zones while leaving the core zone as access-restricted conservation zone (located primarily in the valleys and

mountains). As shown in Figure 3 below, the dark green covers are the dense forests designated as core zones in the inaccessible high altitude steep mountain and in the low altitude river valleys in the Yayu forest. The landscape in the middle altitude landscape are the buffer and transitions zones, where agricultural cultivation is practices with strict management (Gole et al. 2008)

## Data sources

Three data sources were used for the study. The Landsat 8 image, dated February 2018, was used to classify the land use land cover map and extract the forest cover area of the Yayu forest biosphere reserve. Vegetation parameter data for biomass estimation were directly measured in the field using vegetation sampling plots. Vegetation indices (VIs), biophysical variables (BPVs) and relevant bands were derived from Sentinel-2 imagery (Fig. 4).

The Sentinel-2 satellite imagery, taken in the dry season of February 2018, was downloaded from the open access European Space Agency (ESA) hub. The images were pre-processed using the Sentinel Application Platform (SNAP) and quantum GIS (QGIS). The Sentinel-2 Multispectral instrument (MSI) with swath width of 290 km was Ortho-rectified to UTM Zone 37N projection and a radiometric correction was done to reduce atmospheric and sun angle effects (Baillarin et al. 2012). The image was transformed from radiance to surface reflectance by applying the Dark Object Subtraction (DOS) method using the semi-automatic classification plugin (SCP) in QGIS software. The DOS method removes the darkest pixel in each band that might be affected by atmospheric scattering (Chavez 1988). The blue, green, red and near infrared bands, with 10 m resolution, were resampled into a 20 m resolution using ArcGIS software to correspond with the 20 m vegetation sampling plot size of the field data measurement. The Sentinel-2 MSI was used for deriving multi-spectral bands, vegetation indices (VIs) and biophysical variables (BPVs) (Fig. 4).

## Vegetation indices (VI) extraction

The vegetation indices for biomass estimation in this study were extracted from the Sentinel-2 image (Table 1). In a remotely sensed data, a vegetation index is a spectral transformation of two or more bands designed to enhance the contribution of vegetation properties and allow reliable spatial and temporal inter-comparisons of terrestrial photosynthetic activity and canopy structural variations (Huete et al. 2000). Vegetation indices extracted from Satellite data have emerged as important tools in monitoring, mapping and managing terrestrial vegetation as the indices provide radiometric measurement of the quantity, structure and condition of vegetation, and effectively serve as useful indicators of seasonal and inter-annual variations.

There are many VIs with similar functionality and most of them use the inverse relationship between red and near-infrared reflectance associated with healthy green vegetation. The measurements of vegetation attributes include leaf area index (LAI), green leaf area index (GLAI), percent green cover or fractional green cover, chlorophyll content, green biomass and absorbed photosynthetically active radiation (APAR). According to Bannari et al. (1995), VIs are normally classified based on a range of attributes such as the number of spectral bands (2 or greater than 2); the method of calculations (ratio or orthogonal),

depending on the required objective; and the historical development (as first generation VIs or second generation VIs). In order to compare the effectiveness of different VIs, Lyon et al. (1998) classified seven types of VIs based on their computational methods (Subtraction, Division or Rational Transform). With the advancement in hyper-spectral remote sensing technology, high-resolution reflectance spectrums are now available to be used along the traditional multispectral VIs. Besides, VIs have also been developed to be specifically used with hyper-spectral data such as the use of Narrow Band Vegetation Indices.

## **Biophysical variables (BPVs) extraction**

Surface biophysical or canopy properties provide an understanding of the physics of the interactions between solar radiation and vegetation elements (Asrar et al. 1989). Surface parameter retrieval from satellite remote sensing data has been one of the major sources to obtain surface parameters because it relates the vegetation characteristics to its spectral signature or reflectance value thereby providing reasonable estimates of vegetation properties across various spectral, spatial and temporal scales (Asrar et al. 1989). According to Widlowski et al. (2004), biophysical variables describe the spatial distribution of vegetation state and dynamics, thus, are useful for biomass estimation. The vegetation indices and biophysical variables were computed using the ArcGIS and SNAP software. The indices were selected based on their performances in biomass estimation in earlier studies (Table 1). The vegetation index map layers were produced using QGIS and ArcMap (Fig. 5 and 6).

Table 1: List of selected multispectral bands, derived VIs and BPVs from Sentinel-2 image

Multispectral bands and description	
B2	Blue, 490 nm(10m) _Resampled to 20m
B3	Green, 560 nm(10m) _Resampled to 20m
B4	Red, 665 nm(10m) _Resampled to 20m
B5	Red edge, 705 nm(20m)
B6	Red edge, 749 nm(20m)
B7	Red edge, 783 nm(20m)
B8	Near Infrared, 842 nm(10m) _Resampled to 20m
B8a	Near Infrared, 865 nm(20m)
B11	Short Wave IR, 1610 nm(20m)
B12	Short Wave IR, 2190 nm(20m)
Vegetation Indices (Vis)	
IRICI (Inverted Red-Edge Chlorophyll Index)	$(\text{Band 7} - \text{Band 4}) / (\text{Band 5} / \text{Band 6})$
NDVI (Normalized Difference Vegetation Index)	$(\text{Band 8} - \text{Band 4}) / (\text{Band 8} + \text{Band 4})$
TNDVI (Transformed Normalized Difference Vegetation Index)	$[(\text{Band 8} - \text{Band 4}) / (\text{Band 8} + \text{Band 4}) + 0.5]^{1/2}$
NDVI45 (Normalized Difference Vegetation Index with band 4 and 5)	$(\text{Band 5} - \text{Band 4}) / (\text{Band 5} + \text{Band 4})$
SAVI (Soil Adjusted Vegetation Index)	$(\text{Band 8} - \text{Band 4}) / (\text{Band 8} + \text{Band 4} + 0.5) * 1.5$
Biophysical Vegetation Variables (BPVs)	
LAI	Leaf Area Index
FCOVER	Fraction of Vegetation Cover
FPARA	Fraction of Absorbed Photosynthetically Active Radiation
Cab	Chlorophyll content in the leaf

## Vegetation parameter measurement from sampling plots

A total of 20 randomly drawn sample plots were used to measure the AGB biomass samples from the forest. The vegetation parameter (tree parameters) such as Diameter at Breast Height (DBH) and height (H) were measured in a 20 m X 20 m (400m<sup>2</sup>) sampling plot, which were randomly generated from the forest map using ArcGIS. The sampling plot coordinates were used as references to locate the plots on the ground and within the transitional, buffer and core zones of the Yuyu Biosphere reserve forest. Within each plot, all trees with  $\geq 5$  cm diameter and H of  $> 1.3$  m were recorded and measured for DBH and height. The DBH was measured using diameter tape while H was measured using Sunnto clinometer. The field data were used for validating the biomass modeling outputs and to serve as a ground truth data. For most vegetation types in the tropics, a relationship is established for measurable tree parameters and forest stand parameters such as volume and biomass, which are often difficult for a direct measurement

(Husch et al. 2003). Hence, already established allometric equations are often used to estimate the biomass by using tree parameter data.

## Extraction of the pixel values of predictor variables

The pixel values for each variable derived from the Sentinel-2 image were extracted using zonal statistics in ArcGIS. The field plot geographical location (latitude and longitude) points were used as references to match the pixels as shown in the figure below (Fig. 7). The extracted pixel values for each variable were exported in CSV (comma separated variable) data formats.

## Above ground biomass and carbon stock estimation

The above ground biomass and carbon stock were quantified using an allometric equation with input data from the tree parameter measurements such as DBH and H in the field. Besides, specific wood density, which is the dry mass of a unit volume of fresh wood of trees, is used to convert the wood volume into biomass and carbon estimate. The allometric equation selected for this study was established for tropical forest biomass estimation and has been widely applied in similar studies (Chave et al. 2014). The selected equation was applicable for the Yayu forest because of the climatic conditions (mainly rainfall), which is a key parameter for allometric equation determination and vegetation biomass development. The wood density values were species specific and obtained from a secondary source (Gisel et al. 1992).

$$AGB = 0.0673 \times (\rho D^2 H)^{0.976} \text{ Equ. 1}$$

Where, AGB is Above-ground biomass (g),  $\rho$  is specific wood density (g/cm<sup>3</sup>), D2 is diameter at breast height (DBH) (cm); H is height of tree (m). The above-ground biomass was converted into carbon equivalent using the biomass conversion factor or carbon fraction of 0.47 IPCC (2006).

$$C = AGB \times CF \text{ Equ. 2}$$

Where, C is Carbon stock (g), and CF is Carbon Fraction of above ground biomass

## Data analysis (Correlation, regression analysis and model development)

The forest biomass data measured from the field and the extracted variables from the Sentinel-2 images were organized into a spreadsheet with CSV format. Correlation between the biomass estimates from the field and variables from the Sentinel images were tested using SPSS software. Those variables having significant correlation with the measured biomass data were identified, selected and a regression analysis was performed between the measured biomass and the vegetation indices to develop a biomass prediction model.

The model was then evaluated based on the magnitude of the Root Mean Square Error (RMSE) and coefficient of determination ( $r^2$ ). The best model was developed by integrating those variables with high

r<sup>2</sup> and a low RMSE. The equation obtained from the regression model was then used to estimate AGB. The r<sup>2</sup> was preferred since it has a standard measure with values ranging from 0 to 1. The r<sup>2</sup> also shows the percentage of the variability explained by the model (Husch et al. 2003). This makes it easy to understand the relationship between the independent (indices) and dependent variable (biomass) (Peters 2007). The significance of the model was assessed using the P-Value at α = 0.05. For those significant indices, the equation obtained from the regression model was then used to estimate AGB.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (AGB_o - AGB_p)^2}{n}} \dots\dots\dots \text{Equ.3}$$

Where  $AGB_o$  is observed AGB value,  $AGB_p$  is predicted AGB value, and  $n$  is number of samples

## Results

### Above Ground Biomass of the Forest

The results from the plot measurements data showed that the highest amount of forest above ground biomass (ABG) was concentrated in those plots located in the dense part of the forest or the core zone of the biosphere reserve, in which large sized trees are found in large numbers (Table 2, plots shown in bold). The least amount of above ground biomass were recorded in those plots located in the disturbed and semi-disturbed parts of the forest or in the buffer and transition zones of the biosphere reserve (Table 2, plots in red and bold). This is perhaps directly linked to the degree of human impact on the forest since the core is protected while the buffer and transition zones are open for community access that might lead to selective removal of mature trees from the forest.

Table 2: Forest above ground biomass estimated in each sample plot

<b>Sample</b>	<b>UTM location</b>		<b>Plot Area</b>	<b>Above ground biomass</b>	
<b>Plot No.</b>	<b>Latitude (m)</b>	<b>Longitude (m)</b>	<b>Area (m2)</b>	<b>Kg/plot</b>	<b>ton/plot</b>
1	<b>0147368</b>	<b>0931779</b>	<b>400</b>	<b>196.66</b>	<b>0.19666</b>
2	<b>0145903</b>	<b>0932027</b>	<b>400</b>	<b>975.2</b>	<b>0.9752</b>
3	0147329	0934143	400	489.16	0.48916
4	<b>0147023</b>	<b>0934962</b>	<b>400</b>	<b>237.39</b>	<b>0.23739</b>
5	<b>0150452</b>	<b>0925206</b>	<b>400</b>	<b>104.46</b>	<b>0.10446</b>
6	<b>0147389</b>	<b>0929212</b>	<b>400</b>	<b>927.31</b>	<b>0.92731</b>
7	<b>0147854</b>	<b>0927192</b>	<b>400</b>	<b>859.69</b>	<b>0.85969</b>
8	0155815	0923434	400	807.59	0.80759
9	<b>0150437</b>	<b>0922525</b>	<b>400</b>	<b>194.95</b>	<b>0.19495</b>
10	0149495	0922560	400	403.98	0.40398
11	0149358	0922303	400	617.37	0.61737
12	0151628	0937211	400	390.43	0.39043
13	<b>0147518</b>	<b>0926112</b>	<b>400</b>	<b>1104.14</b>	<b>1.10414</b>
14	0147484	0926386	400	618.4	0.6184
15	0145704	0922617	400	481.18	0.48118
16	0146176	0922736	400	461.42	0.46142
17	0146521	0922727	400	383.69	0.38369
18	0163083	0926702	400	337.88	0.33788
19	0822091	0917444	400	686.18	0.68618
20	0151444	0937851	400	754.02	0.75402

## Correlation between AGB and the predictor variables from Sentinel-2 image

The result of a correlation analysis between the measured above ground biomass and the predictor variables extracted from the Sentinel-2 images showed that there is strong correlation between the observed AGB and the predictor variables, with a correlation coefficient (r) value ranging from 0.36 to 0.74. Among the predictor variables, NDVI (r = 0.36), IRECI (r = 0.5), NDVI45 (r = 0.40), LAI (r = 0.74), FAPAR (r = 0.7), FCOVER (r = 0.64) and Cab (r = 0.69) were strongly correlated with the AGB (Table 3). IRECI from the vegetation indices and LAI from the biophysical variables were found to be best correlated with the observed/measured AGB. Among the different variables, biophysical variables were strongly correlated with the above ground biomass (r = 0.65-0.74).

Table3: Correlation between field measured above ground biomass and selected vegetation index predictor variables

	AGB (ton/plot)	IRECI	NDVI	NDVI 45	LAI	FAPAR	FCOVER	Cab	B4
AGB (ton/plot)	1.00								
IRECI	0.52	1.00							
NDVI	0.36	0.89	1.00						
NDVI45	0.40	0.88	0.98	1.00					
LAI	0.74	0.90	0.92	0.94	1.00				
FAPAR	0.71	0.92	0.86	0.96	0.98	1.00			
FCOVER	0.65	0.96	0.97	0.95	0.95	0.98	1.00		
Cab	0.69	0.94	0.91	0.94	0.90	0.92	0.88	1.00	
B4	-0.44	-0.67	-0.65	-0.60	-0.78	-0.68	-0.66	-0.74	1.00

## Relationship between measured biomass and derived indices

The results from the regression analysis revealed that there is a positive linear relationship between forest biomass and vegetation indices extracted from the Sentinel-2 satellite images (Fig. 8;  $r^2 = 0.017 - 0.27$ ). Similarly, the forest biomass has showed a strong and linear relationship with the biophysical variables drawn from the satellite images (Fig. 9;  $r^2 = 0.42 - 0.54$ ). From the MSI bands, Band 4 performed better than other Sentinel-2 bands ( $r = -0.44$  and  $r^2 = 0.2$ ) which is selected for developing the AGB prediction regression model. The best predictor variables were selected for the biomass prediction model development based on the strength of the relationship between the indices and the measured above ground biomass.

## Modeling AGB Biomass Prediction from vegetation indices

From the regression analysis, those variables with high values of coefficient of determination were selected for the above ground biomass prediction. Those with low values of coefficient of determination and those showing multicollinearity were excluded from the model. Only five variables were selected to develop the model and others were excluded because of very low values of coefficient of determination and problem of multi-collinearity (Table 4). As a result, LAI, FCOVER and FAPAR from the biophysical variables, IRECI from the vegetation indices and Band 4 from the MSI bands were selected for the model development (Table 4). The results show that the biophysical variables are better suited for developing forest biomass prediction model compared to other types of vegetation indices.

Table 4: Linear regression result of the observed AGB with selected predictor variables

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.856 <sup>a</sup>	.733	0.638	0.1714926

a. Predictors: (Constant), B4, FCOVER, IRECI, LAI, FAPAR

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t-value	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.282	.905		-2.522	.024
	IRECI	-6.307	3.138	-1.195	-2.010	.064
	LAI	13.452	6.069	2.433	2.217	.044
	FAPAR	-6.180	7.340	-1.364	-.842	.414
	FCOVER	6.633	7.471	1.172	.888	.390
	B4	20.176	12.105	.507	1.667	.118

a. Dependent Variable: AGB\_T

Based on the above regression results, the biomass prediction equation was developed to estimate the above ground biomass of the forest vegetation using Sentinel-2 image extracted variables as shown in the equation below.

$$\text{AGB} = (20.176 * B4) + (6.633 * \text{FCOVER}) - (6.180 * \text{FAPAR}) + (13.452 * \text{LAI}) - (6.307 * \text{IRECI}) - 2.282 \text{ Equ. 4}$$

Where B4 is Band 4, LAI is Leaf area index, IRECI is Inverted Red-Edge Chlorophyll Index, FCOVER is Fraction of vegetation cover, FAPAR is Fraction of Absorbed Photo-synthetically Active Radiation. These indices can be derived from any Sentinel-2 image and can be used to predict forest above ground biomass using the prediction equation. The prediction model was validated using the measured or observed values of above ground biomass from the field (Table 5). The measure of the goodness of fit between the observed and predicted values showed a strong linear relationship with a coefficient of determination of  $r^2 = 0.73$  (Fig. 10).

Table 5: Values of the observed and predicted above ground biomass

Plot Code	Latitude (m)	Longitude (m)	Observed AGB (ton/plot)	Predicted AGB (ton/pixel)
1	0147368	0931779	0.20	0.29
2	0145903	0932027	0.98	0.78
3	0147329	0934143	0.49	0.45
4	0147023	0934962	0.24	0.29
5	0150452	0925206	0.10	0.39
6	0147389	0929212	0.93	0.91
7	0147854	0927192	0.86	0.90
8	0155815	0923434	0.81	0.60
9	0150437	0922525	0.19	0.18
10	0149495	0922560	0.40	0.40
11	0149358	0922303	0.62	0.56
12	0151628	0937211	0.39	0.36
13	0147518	0926112	1.10	0.89
14	0147484	0926386	0.62	0.50
15	0145704	0922617	0.48	0.66
16	0146176	0922736	0.46	0.60
17	0146521	0922727	0.38	0.50
18	0163083	0926702	0.34	0.14
19	0822091	0917444	0.69	0.94
20	0151444	0937851	0.75	0.67

## Discussion

The plot measurement results correspond with the forest biosphere strata of the Yayu forest, illustrating the different management zones of the forest. The magnitude of the Biomass measured from those plots located in the buffer zone was lower than those measured from plots located in the core zone. Unlike the buffer and transition zones, the core zone is protected for biodiversity reserve and conservation (Gole et al. 2008; Schuit et al. 2021). Besides, it is inaccessible as well as the density and size of trees is relatively high compared to the transition zones. On the contrary, the buffer and transition zones are freely accessible for agricultural production and managed for coffee cultivation (Schuit et al. 2021). Trees are sparse and hence, biomass density is low. The average biomass recorded per unit area agrees with earlier studies of similar vegetation types in the Central highlands afro-montane forests (Yohannes et al. 2015; Dibaba et al. 2019; Eshetu and Hailu 2020). These studies used similar coupled methods of sample plot measurements and allometric equations for the biomass estimation. From the different vegetation indices derived from Sentinel-2 image, IRECI showed strong correlation with measured biomass values. This is because IRECI uses the red-edge bands (Band 6 and Band 7), that are well related to vegetation biomass (Castillo et al. 2017). On the contrary, most of the vegetation biophysical indices were strongly

correlated with the measured biomass. This suggests that vegetation biophysical indices are better predictor variables for above ground biomass estimation than the other vegetation indices and multispectral bands. A study by Chen et al. (2018) revealed that the vegetation biophysical variables of Sentinel-2 were the most relevant and important predictors for explaining the observed variability of AGB. This is possibly because the Sentinel-2 product has comparatively large spatial coverage and high resolution to perform efficiently for estimation of biomass than other open-source sensors data products (Juniansah et al. 2018). LAI was found to be the single most important predictor by showing strong correlation with measured biomass. Similar observation has been reported by Castillo et al. (2017). LAI is known to be a very good indicator of vegetation density and can accurately describe state of vegetation compared to other canopy variables (Dusseux et al. 2015; Castillo et al. 2017). Those variables with high coefficient of determination (i.e., most of the vegetation biophysical indices) and those without multicollinearity (e.g, LAI, FCOVER, FAPAR, IRECI Band 4) were found to be suitable predictors of biomass in the regression model. This is consistent with a study by Pandit et al. (2018), who reported that Sentinel-2 images yielded reliable estimates of forest biomass and carbon stock using prediction algorithms, after observing a strong relationship between selected vegetation biophysical variables and measured biomass from the forest with a high measure of goodness of fit ( $R^2=0.81$ ). This also agrees with the model goodness of fit for the values between the predicted and observed values in this study (Fig. 10 above).

Using the raster calculator of ArcGIS and the forest layer thematic map, the above ground biomass was mapped by applying the prediction model (Fig. 11). The result well corresponded with the biosphere structure zones. The highest amount of the AGB is in the range of 6 to 10 ton per pixel or 150 to 250 ton/ha, which are those areas closer to the core conservation zone of the forest biosphere reserve. In the transitional and the buffer zone of the forest, where access roads are available and where agricultural activities are permitted, the concentration of biomass is below 6 ton per pixel (Fig. 11). Within the core zone, a predicted value of biomass higher than 10 ton per pixel (250 ton/ha) has been recorded in scattered pocket areas of the forest reserve. These spots are located in the steepest and most inaccessible parts of the forest, in which anthropogenic activities are restricted and very minimum.

The above ground forest biomass was converted to the carbon equivalent using carbon conversion factor (a default value of  $CF = 0.47$ ) and mapped using the raster calculator in ArcGIS. The carbon stock map is similar to the biomass distribution map and the inaccessible areas were found to have higher amount of carbon stock with a value of 7.05 ton/pixel or 176.25 ton/ha. Likewise, the lowest amount of carbon stock, i.e., 2.82 ton/pixel or 70.5 ton/ha, was recorded in the accessible and transitional zone of the biosphere reserve forest (Fig. 12).

## Conclusion

Forest biomass and carbon stock estimation techniques using remotely sensed data are becoming more reliable due to improvements in spectral and spatial resolutions of products from different sensors. Sentinel-2 optical data are increasingly applied for biomass and other vegetation attributes estimation.

This study assessed the strength of correlation and relationship between forest above ground biomass directly measured from sampling plots and vegetation indices as well as biophysical variables extracted from Sentinel-2 optical images so as to establish a forest biomass prediction model by identifying the determinant vegetation indices. Those variables and bands with high coefficient of determination were selected for a multi-linear regression analysis. The input variables for the prediction model were again selected based on the magnitude of their coefficients and the root mean square error. The values of the coefficients of determination of the variables ranged from 0.31 to 0.74 between Sentinel-2 image extracted indices and the above ground biomass. A strong correlation was observed for most of the vegetation biophysical variables ( $r = 0.65 - 0.74$ ) than for the vegetation indices with the measured AGB data from the field. As a result, the variables such as the LAI, FCOVER, FAPAR, IRECI and Band 4 were main inputs for developing the above ground biomass prediction model. The model was validated by considering the strong correlation coefficient of 0.738 and the root mean square error of 0.16 between the observed and predicted values of the forest above ground biomass. Hence, we can conclude that vegetation biophysical variables derived from Sentinel-2 optical images are highly suitable for forest above ground biomass prediction.

## Abbreviations

AGB: Above Ground Biomass; Cab: Chlorophyll content in the leaf; CF: Carbon Fraction; CSV: Comma Separated Variable; DBH: Diameter at Breast Height; DOS: Dark Object Subtraction; ESA: European Space Agency; FAPAR: Fraction of Absorbed Photo-synthetically Active Radiation; FCOVER: Fraction of Vegetation Cover; IPCC: Intergovernmental Panel for Climate Change; IRECI: Inverted Red-Edge Chlorophyll Index; LAI: Leaf Area Index; MoA: Ministry of Agriculture; MEFCC: Ministry of Environment, Forestry and Climate Change; MS: Multispectral; Multispectral Instrument; SNAP: Sentinel Application Platform; QGIS: Quantum GIS; REDD: Reduced Emission from Deforestation and forest Degradation; RMSE: Root Mean Square Error,

## Declarations

### Acknowledgements

The authors would like to thank the Environment and Coffee Forest Forum for allowing the researchers to use the sampling sites in the Yayu Forest Biosphere Reserve. The School of Earth Science at the College of Natural Science is highly appreciated for providing the GIS lab facility during the research.

### Authors' contribution

The authors contributed equally to the research design and writing of the manuscript. SM has collected and analyzed the field data, conducted image analysis and data analysis. MA designed the paper structure, composed the main text and mainly contributed to the final writing of the manuscript. All authors have read and approved the manuscript.

## Authors' information

Seid Muhe is a GIS and Remote Sensing expert, lecturer at Semera University, in the Department of Geography and Environmental Studies, in Semera Town, Ethiopia. Dr. Mekuria Argaw is an associate professor of Ecology and Environmental Sciences, at the Center for Environmental Science, College of Natural and Computational Science, Addis Ababa University, Addis Ababa, Ethiopia.

## Funding

No funding was provided to this research.

## Availability of data and materials

The data set used/or analyzed during the current study are available with the authors. Data will be provided upon reasonable request to the corresponding author.

## Ethics approval and Consent to participate

Not applicable

## Consent for publication

Not applicable

## Competing interests

Authors declare that they don't have competing interests.

## References

1. Alkama R, Cescatti A (2016) Biophysical climate impacts of recent changes in global forest cover. *Science* 351: 600-604. DOI: 10.1126/science.aac8083.
2. Asrar G, Myneni R, Kanemasu T (1989) Measuring and modeling spectral characteristics of a tall grass prairie. *Rem Sens Envir* 27: (2), 143-155. doi.org/10.1016/0034-4257(89)90014-X.
3. Baillarin SJ, Meygret A, Dechoz C, Petrucci B, Lacherade S, Tremas T, Spoto F (2012) Sentinel-2 level 1 products and image processing performances. *Int Geoscie and Rem Sens Symposium (IGARSS)* 39(B1): 197-202.
4. Bannari A, Morin D, Bonn F, Huete AR (1995) A review of vegetation indices. *Rem Sens Revi.* 13 (1–2): 95–120.
5. Canadell JG and Raupach MR (2008) Managing forests for climate change mitigation. *Science*, 320 (5882): 1456–1457. doi:10.1126/science.1155458
6. Chave JR, Condit SA, Hernandez ASL, and Perez R (2004) Error propagation and scaling for tropical forest biomass estimates. *Philos. Trans. R. Soc. London, Ser. 359 (1443):*409– 420

7. Chavez PS (1988) An Improved Dark-Object Subtraction Technique for Atmospheric scattering Correction of Multispectral Data. *Remote Sens. Environ.* 24(3): 459–479.
8. Chen L, Ren C, Zhang B, Wang Z, and Xi Y (2018) Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery. *Forests* 9:1-20.
9. Dibaba A, Soromessa T, Workineh B (2019) Carbon stock of the various carbon pools in Gerba-Dima moist Afromontane forest, South-western Ethiopia. *Carb Bala and Mana* 14 (1): doi:10.1186/s13021-019-0116-x.
10. Dou X, Yang Y (2018) Estimating forest carbon fluxes using four different data-driven techniques based on long-term eddy covariance measurements: Model comparison and evaluation. *Scie of The Total Envir* 627: 78–94. doi:10.1016/j.scitotenv.2018.01.
11. Dusseux P, Hubert-Moy L, Corpetti T, Vertès F (2015) Evaluation of SPOT imagery for the estimation of grassland biomass. *Int J Appl Earth Obs Geoinf* 38: 72–77.
12. Eshetu EY, Hailu TA (2020). Carbon sequestration and elevational gradient: The case of Yegof mountain natural vegetation in North East, Ethiopia, implications for sustainable management. *Cogent Food & Agri* 6(1): 1733331. doi:10.1080/23311932.2020.17333
13. Forkuor G, Dimobe K, Serme I, Tondoh J (2017) Landsat-8 vs. Sentinel-2: Examining the added value of Sentinel-2's red-edge bands to land-use and land cover mapping in Burkina Faso. *GISci Remo Sens* 2: 1–24. <https://doi.org/10.1080/15481603.2017.1370169>.
14. Georgia G, Dimitris Z, Ioannis G, Kalliopi R, Vassilia K, Maria TS, Iain W, Giorgos M (2017) Vegetation biomass estimation with remote sensing: focus on forest and other wooded land over the Mediterranean ecosystem. *Int J of Remo Sens* 38 (7): 1940-1966. DOI: 10.1080/01431161.2016.1266113
15. Gisel R, Sandra B, Jonathan C, Ariel E (1992) Wood Densities of Tropical Tree Species. New Orleans, Louisiana. [https://www.srs.fs.usda.gov/pubs/gtr/gtr\\_so088.pdf](https://www.srs.fs.usda.gov/pubs/gtr/gtr_so088.pdf) accessed on 28.03.2019.
16. Gole TW, Borsch T, Denich M, Teketay D (2008) Floristic composition and environmental factors characterizing coffee forests in southwest Ethiopia. *Fores Ecol Managt* 255: 2138–2150. <https://doi.org/10.1016/j.foreco.2007.12.028>
17. Gómez M (2017) Joint Use of Sentinel-1 and Sentinel-2 for Land Cover Classification: A Machine Learning Approach. MSc Thesis, Lund University, Lund, Sweden.
18. Herold M, Román-Cuesta RM, Mollicone D, Hirata Y, Van Laake P, Asner GP, Souza C, Skutsch M, Avitabile V and MacDicken K (2011) Options for monitoring and estimating historical carbon emissions from forest degradation in the context of REDD+. *Carb Bala and Manat*, 6 (13): 1-7
19. Huete A, Didan K, Miura T, Rodriquez EP, Gao X, Ferreria LG (2000). "Overview of the radiometric and biophysical performance of the MODIS vegetation indices". *Remo Sens Environ* 83 (5): 195–213.
20. Husch B, Beers TW, Kershaw JA (2003) *Forest mensuration* 4<sup>th</sup> ed. Hoboken, Wiley and Sons, Inc New York.

21. Isbaex C, Coelho AM (2020) The Potential of Sentinel-2 Satellite Images for Land-Cover/ Land-Use and Forest Biomass Estimation: A Review. IntechOpen.  
DOI: <http://dx.doi.org/10.5772/intechopen.93363>.
22. Juniansah A, Tama GC, Febriani KR, Baharain MN, Kanekaputra T, Wulandari WS, Kamal M (2018) Mangrove Leaf Area Index Estimation Using Sentinel 2A Imagery in Teluk Ratai, Pesawaran Lampung IOP Conf. Series. Earth and Envir Science 165: 012004. doi :10.1088/1755-1315/165/1/012004.
23. Kebede M, Kanninen M, Yirdaw E, Lemenih M (2013) Vegetation structural characteristics and topographic factors in the remnant moist Afro-montane forest of Wondo Genet, south central Ethiopia. *J of For Res* 24 (3): 419–430. doi:10.1007/s11676-013-0374-5
24. Lu D (2006) The Potential and Challenge of Remote Sensing-based Biomass Estimation. *Inter J of Remo Sens* 27 (7): 1297–1328. doi:10.1080/01431160500486732.
25. Lyon JG, Yuan D, Lunetta RS, Elvidge CD (1998) A Change Detection Experiment Using Vegetation Indices. *Photo Engin & Remo Sens* 64 (2): 143-150.
26. Mascaro J, Detto M, Asner GP, Muller-Landau HC (2011) Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remo Sens of Environ*, 115 (12): 3770–3774.  
doi:10.1016/j.rse.2011.07.019
27. Mauya EW, Hansen E, Gobakken T, Bollandsås M, Malimbwi E, Næsset E (2015) Effects of field plot size on prediction accuracy of aboveground biomass airborne laser scanning- assisted inventories in tropical rain forests of Tanzania. *Carb Bal and Managt* 10: 1–14.
28. McRoberts RE, Næsset E, Gobakken T (2013) Inference for lidar-assisted estimation of forest growing stock volume. *Remo Sens Envir* 128: 268–275.
29. MEFCC (Ministry of Environment, Forest and Climate Change) (2016) Ethiopia's forest reference level submission to the United Nations framework convention for climate change. Addis Ababa, Ethiopia.
30. Melese B, Kelbessa E, Soromessa T (2014) Forest carbon stocks in woody plants of Arba Minch ground water forest and its variations along environmental gradients. *Scie Tech and Art Rese J* 3(2): 141–147. doi:10.4314/star.v3i2.18
31. MoA (Ministry of Agriculture) (2000) Woody Biomass Inventory and Strategic Planning Project (WBISPP), Manual for woody biomass inventory, Ministry of Agriculture, Addiss Ababa, Ethiopia.
32. Pan Y, Birdsey R, Fang J, Houghton R, Kauppi P, Kurz W, Phillips O, Shvidenko A, Lewis SL, Canadell J (2011) A large and persistent carbon sink in the world's forests. *Science* 333: 988–993.
33. Pandit S, Tsuyuki S, Dube T (2018) Estimating above-ground biomass in sub-tropical buffer zone community Forests, Nepal, using Sentinel 2 data. *Remo Sens* 10 (4): 601.
34. Pertille CT, Marcos FN, Larissa RT, Thiago F (2019) Biomass quantification of Pinus taeda L. from remote optical sensor data. *Adva in Fore Scien* 6 (2): 603-610.
35. Peters AJ (2007) Performance evaluation of spectral vegetation indices using a statistical sensitivity function. *Remo Sens Envir* 106 (1): 59–65.

36. Powell SL, Cohen WB, Healey SP, Kennedy RE, Moisen GG, Pierce KB, Ohmann JL (2010) Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remo Sens of Enviro* 114 (5): 1053–1068. doi:10.1016/j.rse.2009.12.018
37. Ramoelo A, Cho M, Mathieu R, Skidmore A (2015) Potential of Sentinel-2 spectral configuration to assess rangeland quality. *J. Appl Remote Sens Environ* 124: 516–533.
38. Schuit P, Moat J, Gole TW, Challa ZK, Torz J, Macatonia S, Cruz G, Davis AP (2021) The potential for income improvement and biodiversity conservation via specialty coffee in Ethiopia. *PeerJ* 9:e10621 DOI 10.7717/peerj.10621
39. Shoko C, Mutanga O (2017) Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species. *ISPRS J Photogr Remo Sens* 129:32–40.
40. Shrestha SK (2011) Carbon stock estimation using very high-resolution satellite imagery and individual crown segmentation. (A case study of broadleaved and needle leaved forest of Dolakha, Nepal. MSc Thesis, ITC - University of Twente, Enschede.
41. Siraj M (2019) Forest carbon stocks in woody plants of Chilimo-Gaji Forest, Ethiopia: Implications of managing forests for climate change mitigation. *Sou Afri J of Bota*, 127: 213–219. doi:10.1016/j.sajb.2019.09.003
42. Soenen SA, Peddle DR, Hall RJ, Coburn CA, Hall FG (2010). Estimating aboveground forest biomass from canopy reflectance model inversion in mountainous terrain. *Remo Sens of Envir* 114 (7): 1325–1337. doi:10.1016/j.rse.2009.12.012
43. Sun X, Guicai L, Meng W, Zemeng F (2019) Analyzing the Uncertainty of Estimating Forest Aboveground Biomass Using Optical Imagery and Space-borne LiDAR. *Remo Sens* 11: 722; doi:10.3390/rs11060722
44. Tadesse H, Zerihun A, Burud I, Terje G, Hans O, Øystein BD, Erik N (2020) Use of Remotely Sensed Data to Enhance Estimation of Aboveground Biomass for the Dry Afro-montane Forest in South-Central Ethiopia. *Remo Sens* 12: 3335. doi:10.3390/rs12203335
45. Timothy D, Onisimo M, Riyad O (2015) Evaluating the utility of the medium-spatial resolution Landsat 8 multi-spectral sensor in quantifying aboveground biomass in Umgeni catchment, South Africa, *ISPRS J. Phot Rem Sens* 101: 36–46.
46. Timothy D, Onisimo M, Riyad O (2016) Quantifying aboveground biomass in African environments: A review of the trade-offs between sensor estimation accuracy and costs. *Tro Eco* 57 (3): 393-405.
47. Vashum KT, Jayakumar S (2012) Methods to Estimate Above-Ground Biomass and Carbon Stock in Natural Forests - A Review. *J Ecosyst Ecogr* 2:116. doi:10.4172/2157-7625.1000116
48. Widlowski JL, Pinty B, Gobron N, Verstraete MM, Diner DJ, Davis AB (2004) Canopy structure parameters derived from multi-angular remote sensing data for terrestrial carbon studies. *Clim Change* 67 (2-3): 403-415.

49. Yohannes H, Soromessa T, Argaw M (2015) Carbon stock analysis along altitudinal gradient in gedo for-est: Implications for forest management and climate change mitigation. *Ame J of Enviro Prot*, 4 (5): 237–244. doi:10.11648/j.ajep.20150405.14
50. Zianis D, Mencuccini M (2004) On simplifying allometric analyses of forest biomass. *Fore Ecol and Mana* 187: (2-3), 311–332. doi:10.1016/j.foreco.2003.07.00

## Figures

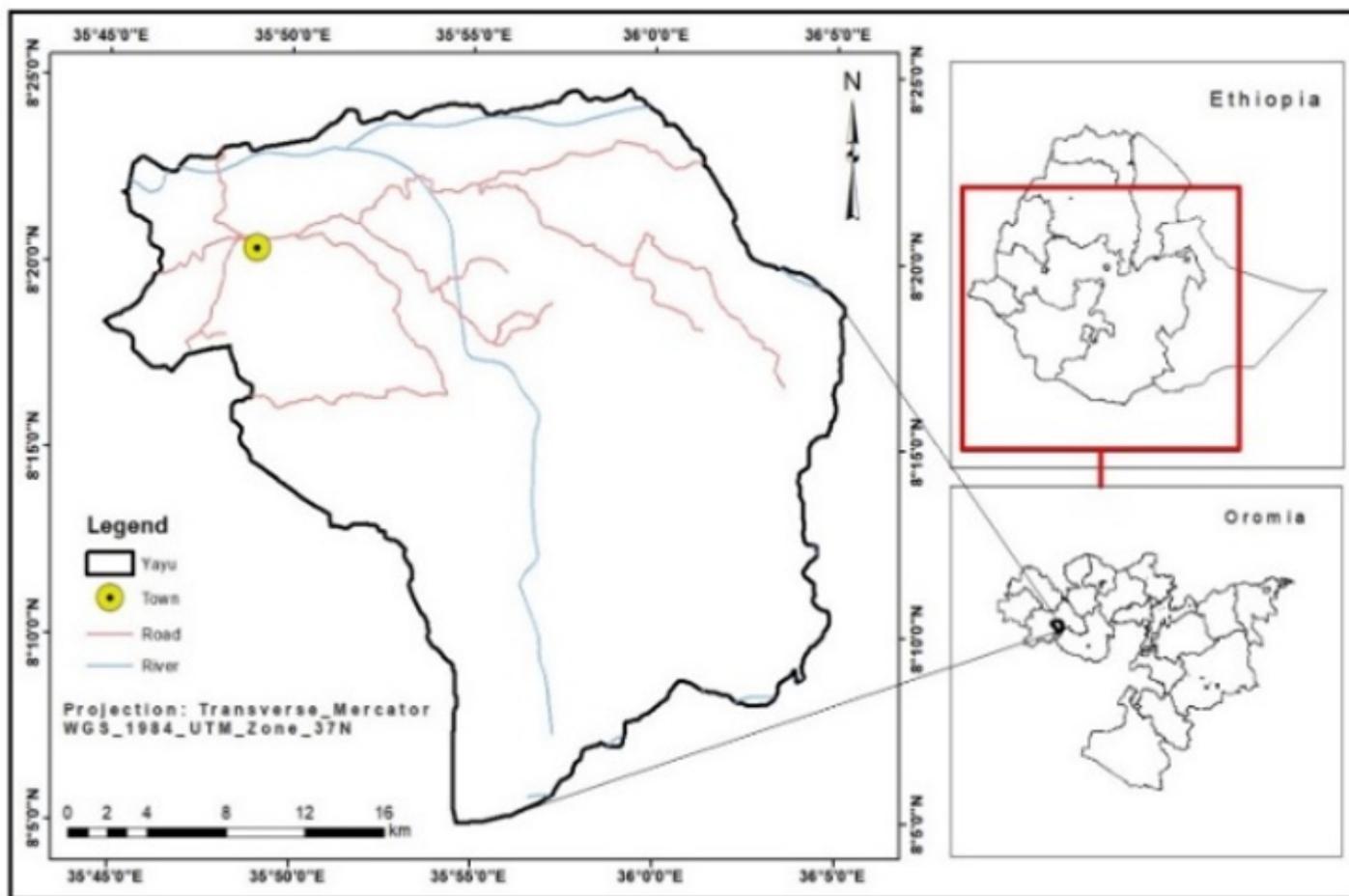
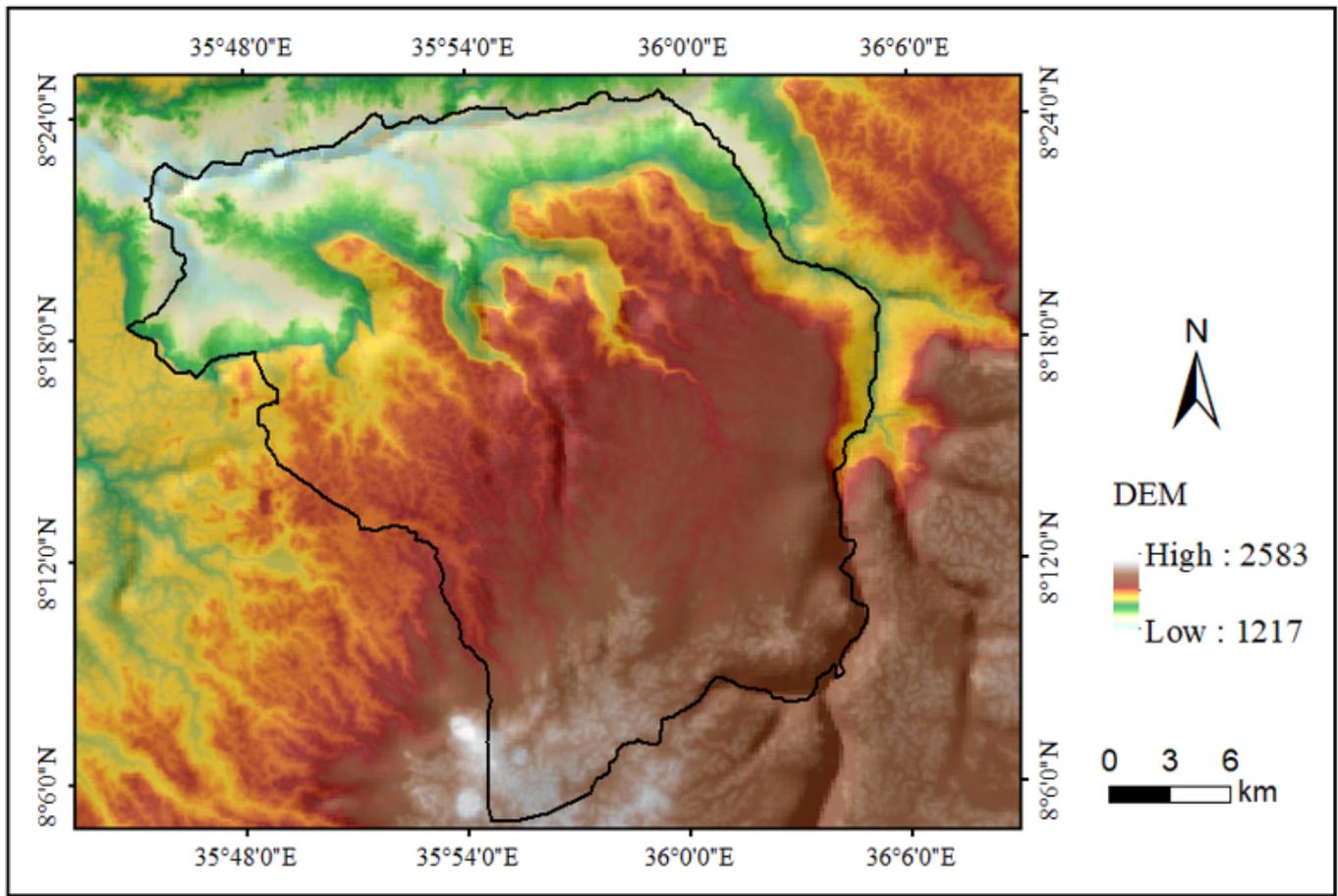


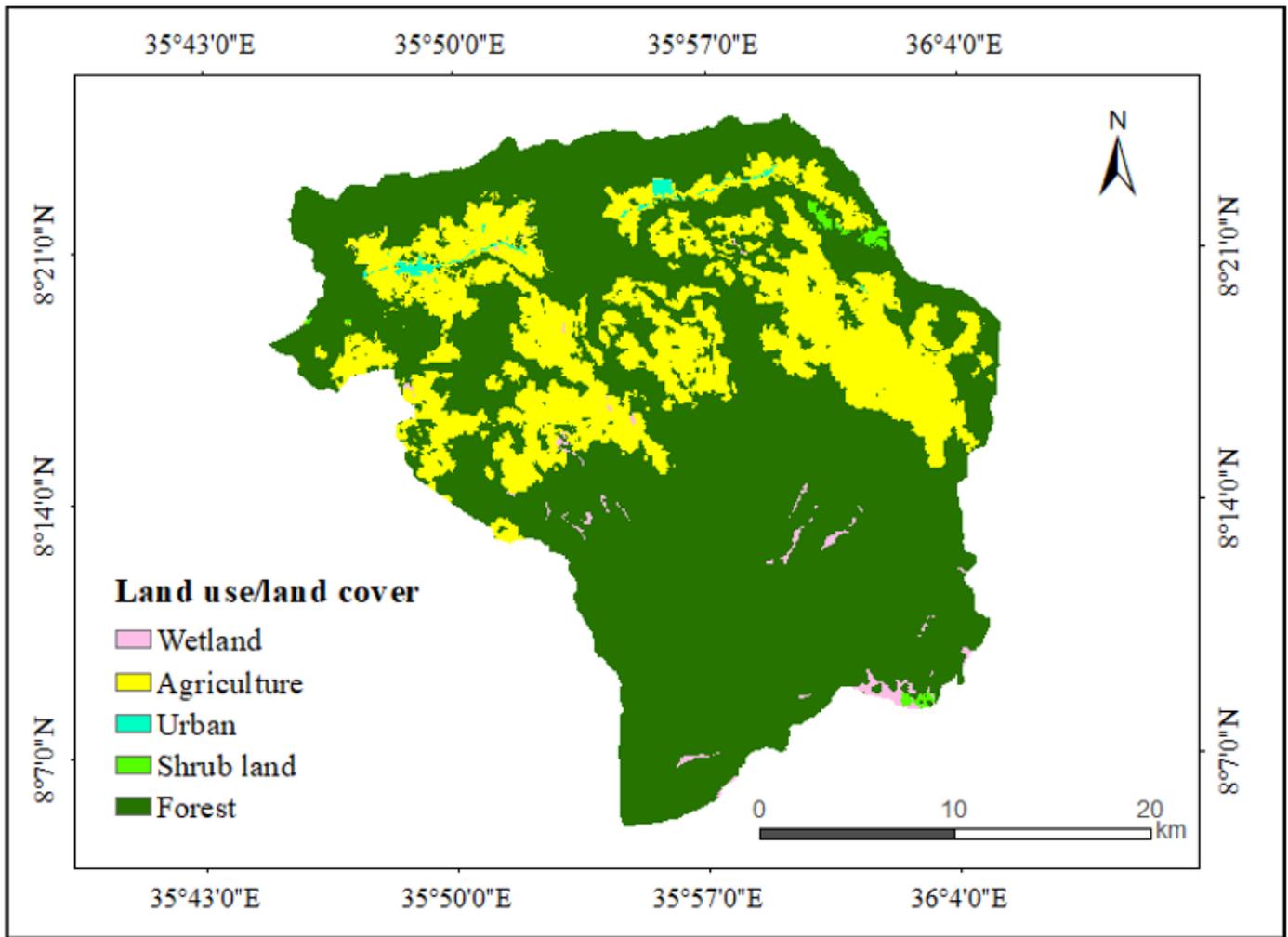
Figure 1

Location map of the study area



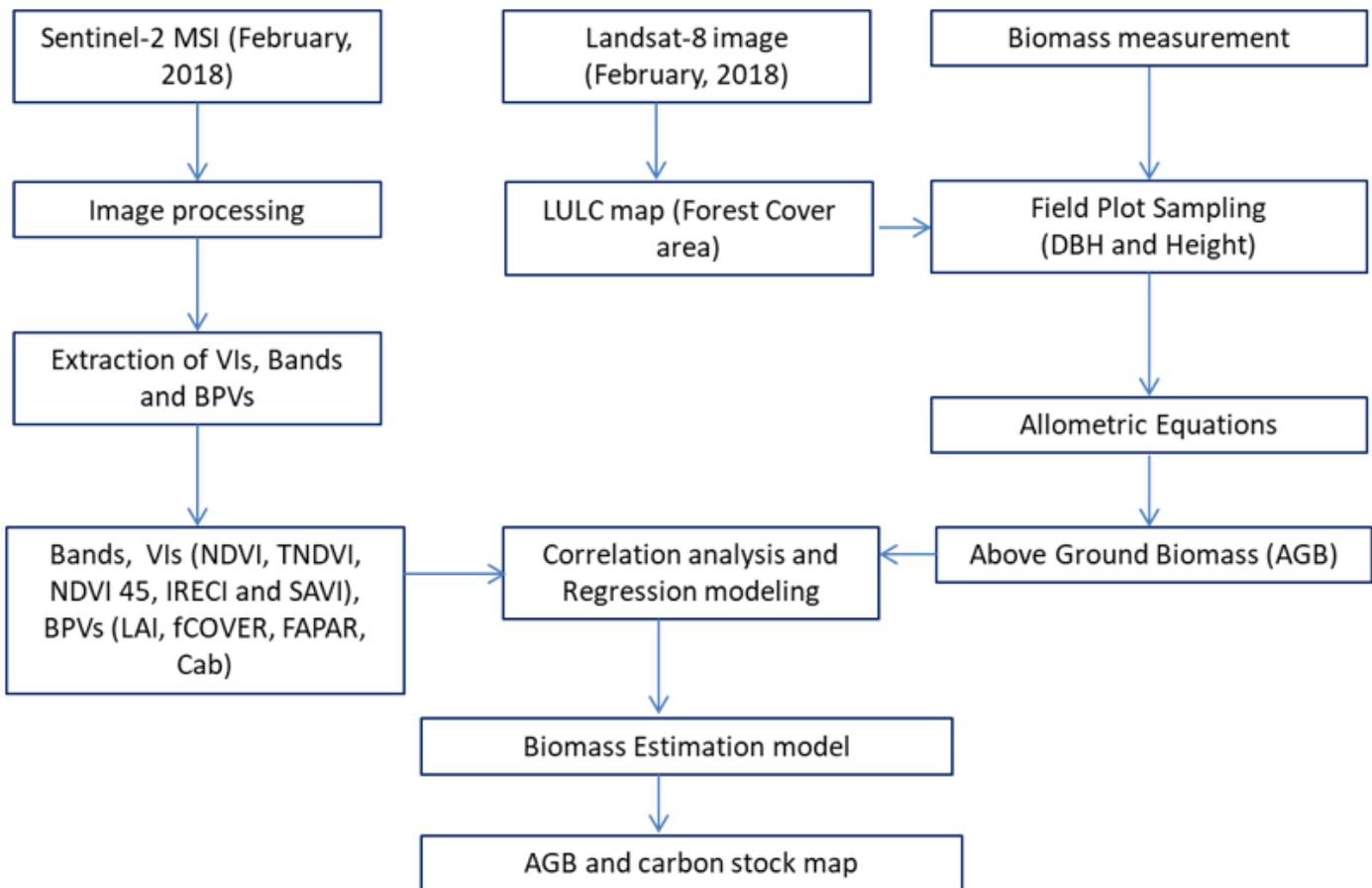
**Figure 2**

Elevation map of the study area



**Figure 3**

Land use land cover map of the study area



**Figure 4**

Methodological flow chart

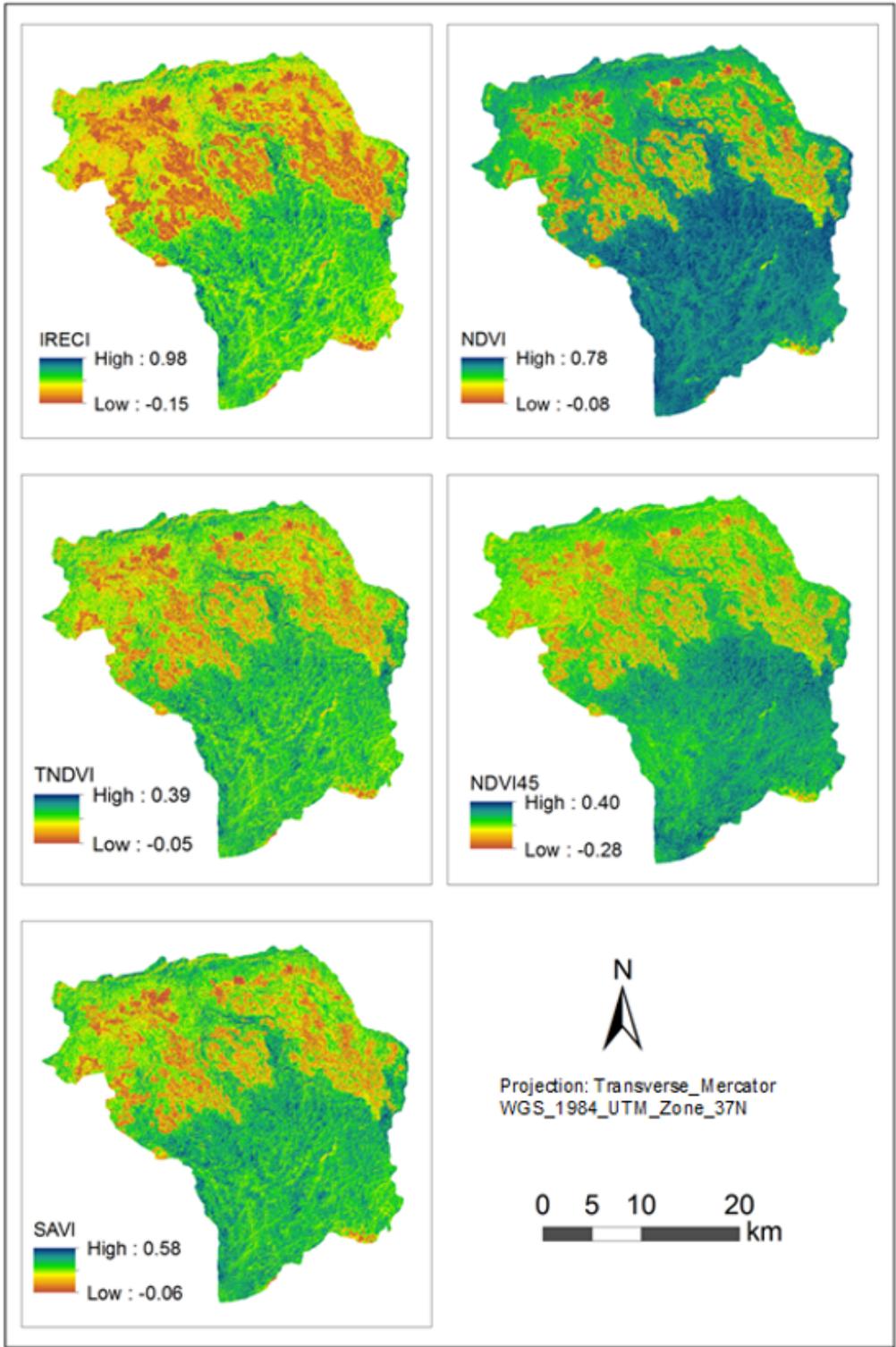


Figure 5

Map of the vegetation indices

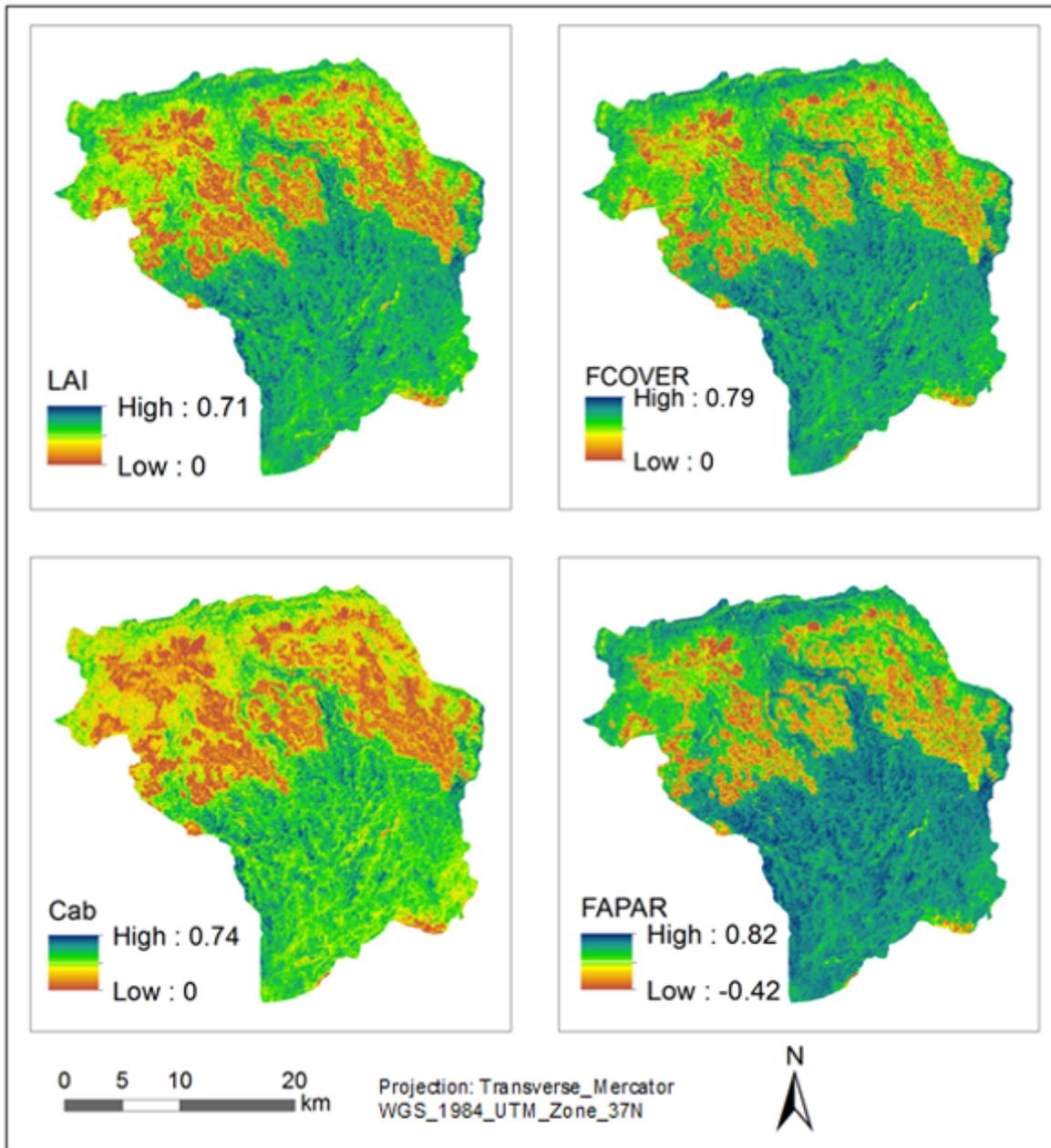
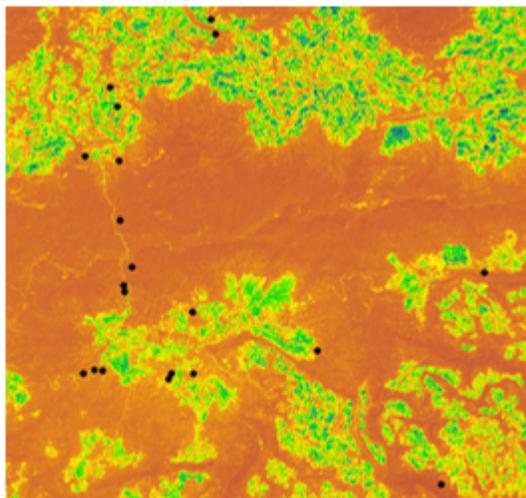
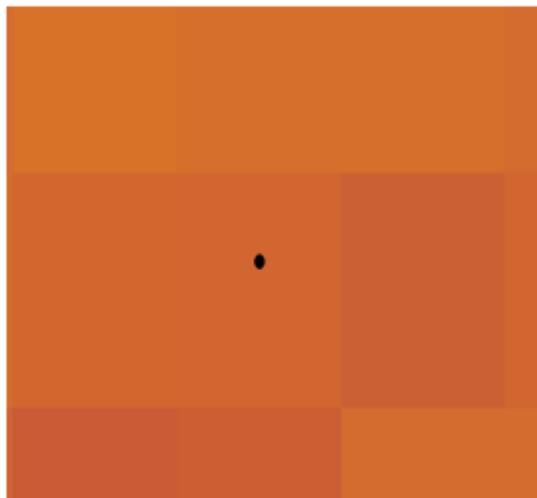


Figure 6

Map of the vegetation biophysical variable



a) Field sample plots



b) Sample plot overlaid on pixels

Figure 7

Pixel values extraction by overlaying field samples plots on image pixels

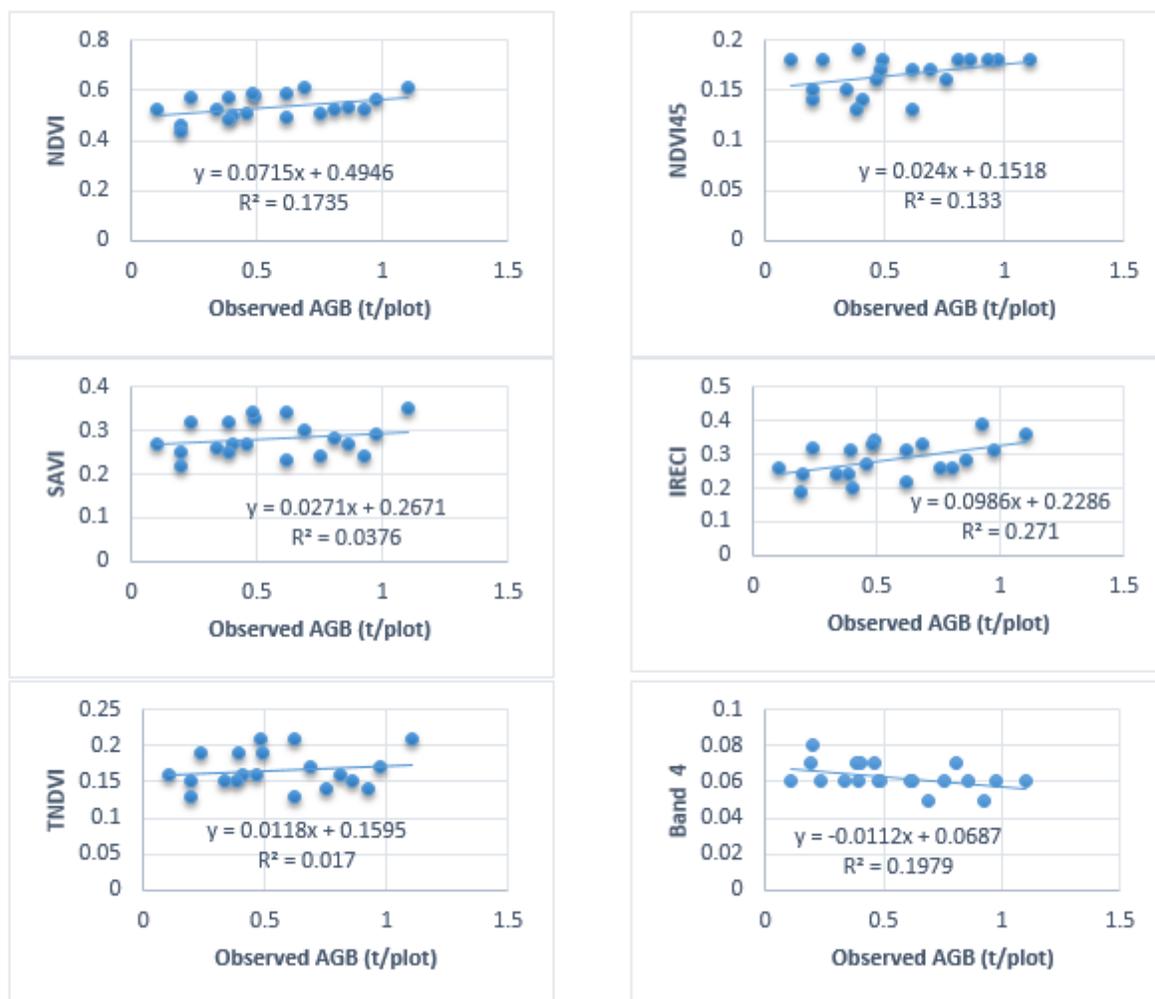


Figure 8

Linear relationship between Observed AGB and Vegetation Indices extracted from Sentinel-2 MSI

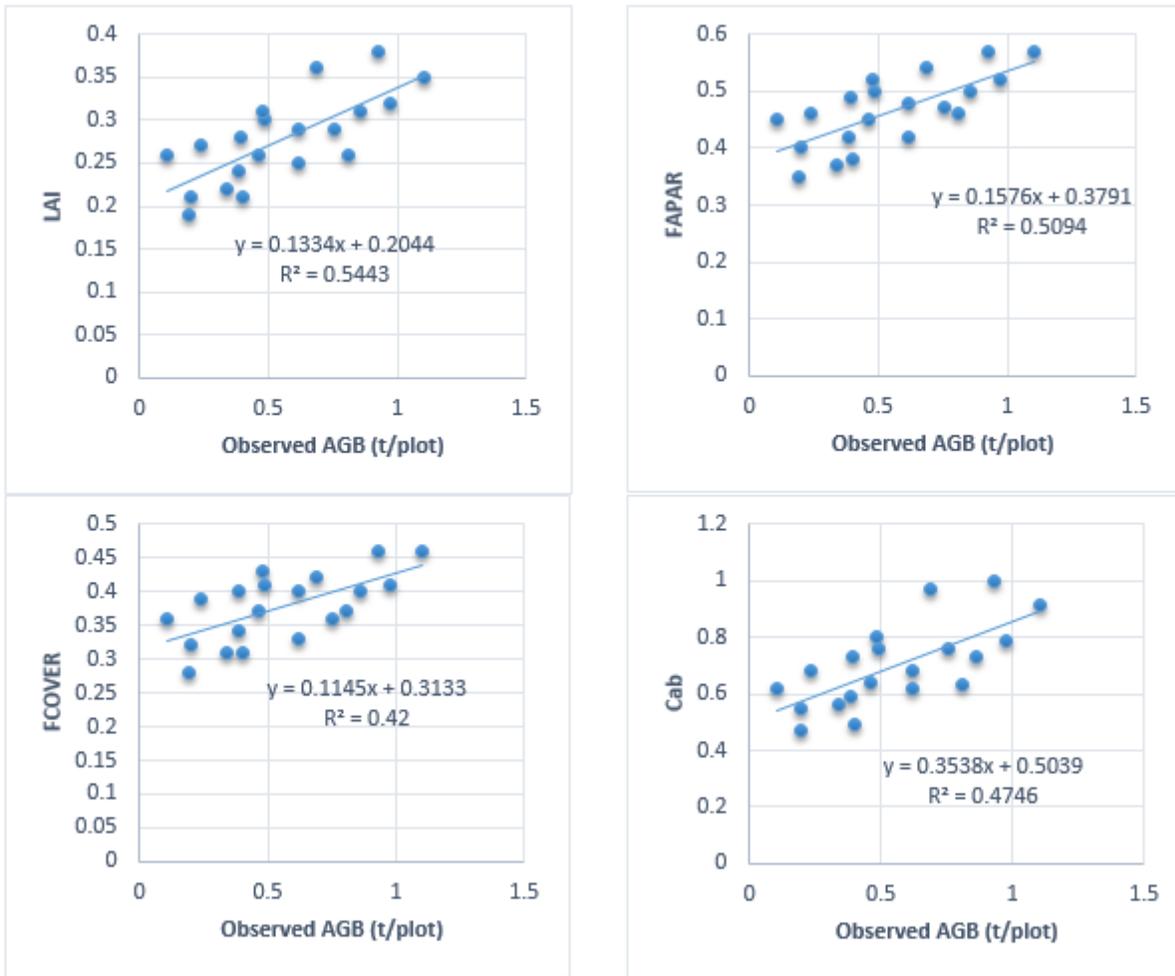


Figure 9

Linear relationship between Observed AGB and Vegetation Biophysical Variables extracted from Sentinel-2 MSI

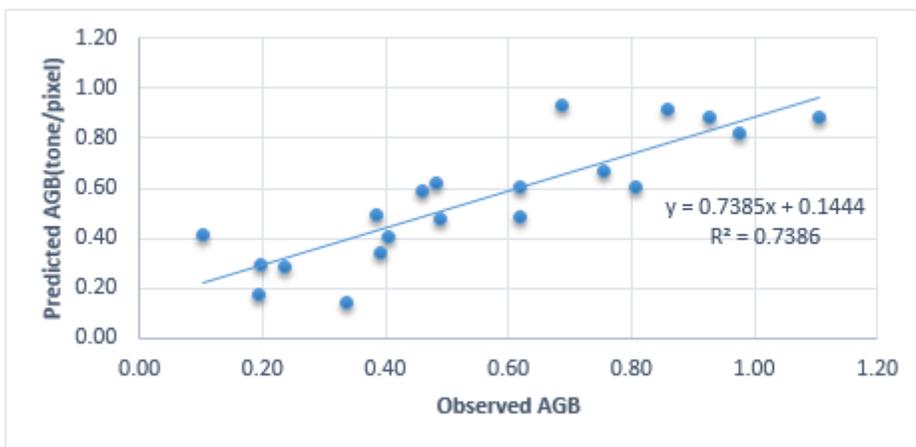


Figure 10

Scatter plot showing goodness of fit between the observed and predicted values of the above ground biomass

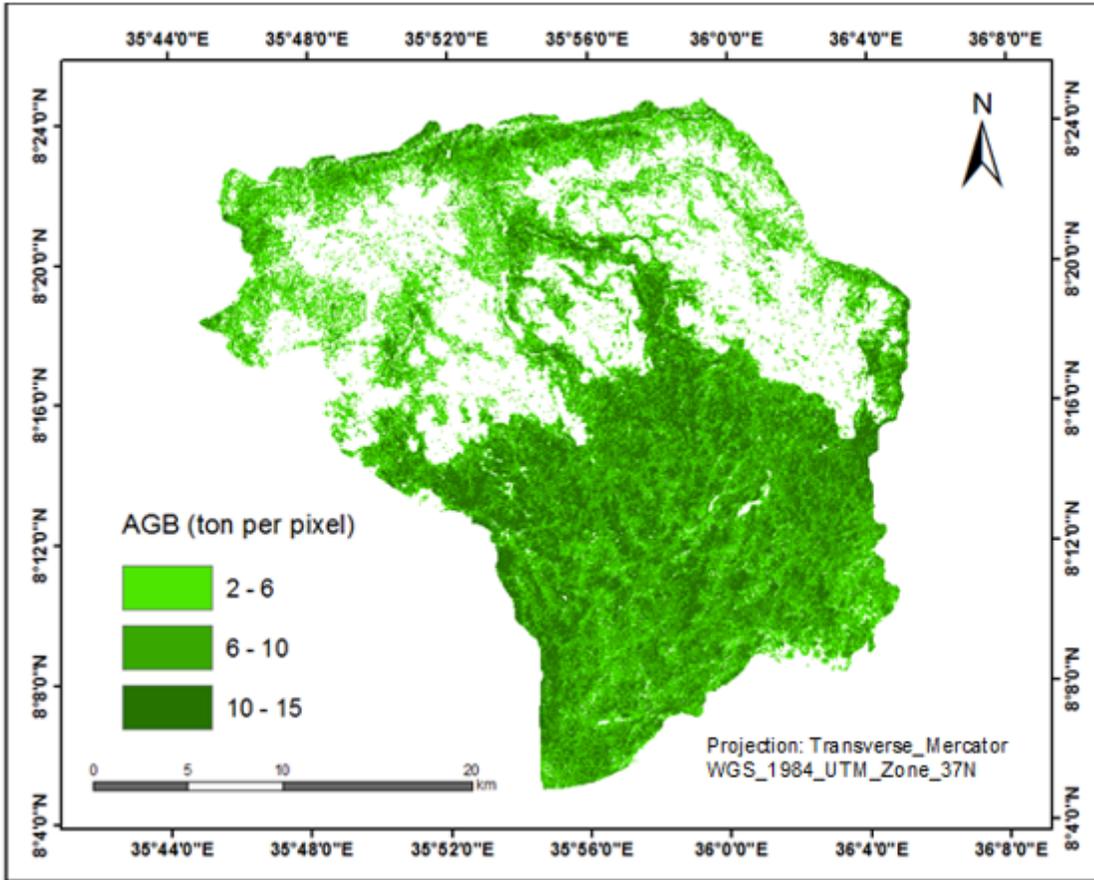


Figure 11

Map of the predicted values of the above ground biomass of Yayu forest biosphere reserve

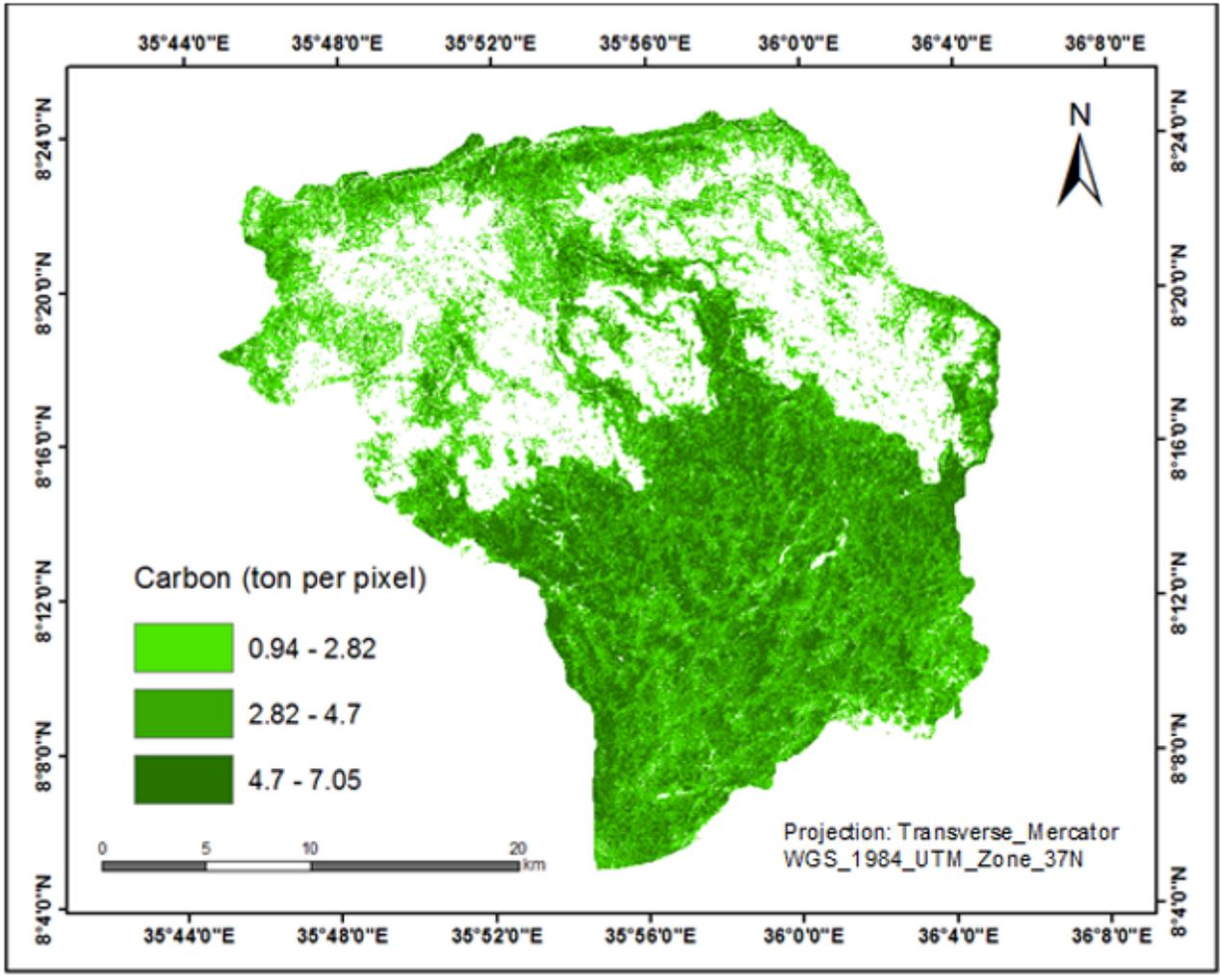


Figure 12

Map of the predicted carbon stock in the biosphere reserve of Yayu forest