Geocarto International

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgei20

Aboveground biomass estimation of an agro-pastoral ecology in semi-arid Bundelkhand region of India from Landsat data: a comparison of support vector machine and traditional regression models

Dibyendu Deb, Shovik Deb, Debashis Chakraborty, J. P. Singh, Amit Kumar Singh, Puspendu Dutta & Ashok Choudhury

To cite this article: Dibyendu Deb, Shovik Deb, Debashis Chakraborty, J. P. Singh, Amit Kumar Singh, Puspendu Dutta & Ashok Choudhury (2022) Aboveground biomass estimation of an agro-pastoral ecology in semi-arid Bundelkhand region of India from Landsat data: a comparison of support vector machine and traditional regression models, Geocarto International, 37:4, 1043-1058, DOI: 10.1080/10106049.2020.1756461

To link to this article: https://doi.org/10.1080/10106049.2020.1756461

GEOCARTO INTERNATIONAL
2022, VOL. 37, NO. 4, 1043–1058
https://doi.org/10.1080/10106049.2020.1756461

Taylor & Francis
Taylor & Francis Group

Aboveground biomass estimation of an agro-pastoral ecology in semi-arid Bundelkhand region of India from Landsat data: a comparison of support vector machine and traditional regression models

Dibyendu Deb[a], Shovik Deb[b], Debashis Chakraborty[c], J. P. Singh[a], Amit Kumar Singh[a], Puspendu Dutta[b] and Ashok Choudhury[b]

[a]Indian Grassland and Fodder Research Institute, Jhansi, India; [b]Uttar Banga Krishi Viswavidyalaya, Cooch Behar, India; [c]Indian Agricultural Research Institute, New Delhi, India

ABSTRACT

This study compared the traditional regression models and support vector machine (SVM) for estimation of aboveground biomass (ABG) of an agro-pastoral ecology using vegetation indices derived from Landsat 8 satellite data as explanatory variables . The area falls in the Shivpuri Tehsil of Madhya Pradesh, India, which is predominantly a semi-arid tract of the Bundelkhand region. The Enhanced Vegetation Index-1 (EVI-1) was identified as the most suitable input variable for the regression models, although the collective effect of a number of the vegetation indices was evident. The EVI-1 was also the most suitable input variable to SVM, due to its capacity to distinctly differentiate diverse vegetation classes. The performance of SVM was better over regression models for estimation of the AGB. Based on the SVM-derived and the ground observations, the AGB of the area was precisely mapped for croplands, grassland and rangelands over the entire region.

ARTICLE HISTORY
Received 9 December 2019
Accepted 28 March 2020

KEYWORDS
Aboveground biomass; agro-pastoral ecology; satellite data; Landsat data; enhanced vegetation index; regression model; support vector machine

1. Introduction

Estimation of the aboveground biomass (AGB) is necessary to understand the types of land use/cover, crop vigour and also to calculate the biosphere carbon (C) share in global C cycle (Baccini et al. 2008). Quantification of AGB on ground can be done through destructive methods or by estimation of biophysical variables like diameter at breast-height, canopy cover, tree density etc. and then using these values as input variables in allometric equations (Basuki et al. 2009; Datta and Deb 2017). For smaller plants, shrubs, herbs, volumetric method of biomass estimation is common (Muldavin et al. 2008).

Application of space-borne optical passive remote sensing to measure AGB is a good alternative with the advantages of high accuracy and is cheap and less labour intensive than any field based study (Yang et al. 2012; Gao et al. 2013). Launching of satellite series like Landsat and NOAA in 1970s initiated this approach and it became more effective

CONTACT Shovik Deb shovik@ubkv.ac.in

with betterment of spatial and spectral resolution of newer satellites (Anaya et al. 2009; Jaber 2019). The process involves establishment of a relationship of satellite derived data with the ground-based measurements and then use of that data to predict AGB over an area, using that equation (Jin et al. 2014; Deb et al. 2017). In most of the cases, vegetation indices based models are used in this purposes (Basuki et al. 2012; Yang et al. 2012). These indices are nothing but mathematical combination of reflectance spectra in red and near infrared bands (Viña et al. 2011). There are numbers of vegetation indices with varying degree of simplicity, reliability, inclusion and exclusion of factors and studies have used many of them comprehensively to estimate AGB of different land cover classes (Jiang et al. 2008; Bendig et al. 2015). Another approach is use of space-borne synthetic aperture (SAR) radar data (Neumann et al. 2012; Santi et al. 2017). Integration of optical and SAR data synergistically can help to reduce saturation error of optical data in high density vegetation area as well as miscalculation due to microwave backscattering from soil (Zhang et al. 2015; Shao and Zhang 2016).

Different studies have adopted several modelling techniques to establish relationship of vegetation indices to field observed AGB. Simple and multiple linear, and nonlinear regression models are effective and thus, have been adopted by many scientists (Zheng et al. 2004; McRoberts et al. 2013). Studies have found that nonlinear models can better estimate the AGB (Paine et al. 2012). Recently, machine learning techniques have been introduced in this context (Dang et al. 2019; Rehman et al. 2019). Machine learning algorithms like artificial neural network, support vector machine (SVM), random forests are gaining popularity in ecological modelling, and multiple studies explained these techniques as better alternative to the other modelling methods (Muttil and Chau 2007; Ibrahim and Bennett 2014; Lu et al. 2016; Dang et al. 2019). The SVM is a popular machine learning technique due to its stability, ease in tuning and precision in modelling with involvement of couple of parameters (Drake et al. 2006; Adam et al. 2014).

The present research was designed to understand the potential of traditional regression methods as well as SVM in estimating the AGB using space-borne data. Only passive optical data were considered as the mean ground measured AGB of the area was 14.85 T ha^{-1} only, far less than the potential saturation point for optical reflectance saturation (Zhao et al. 2016). To compare the models to predict AGB for different land cover classes, this study selected an area under agro-pastoral ecology, comprising diverse vegetation covers like croplands, grass and rangelands, and even discrete patches of dense vegetation/plantation and bare earth. We hypothesize that SVM is a better choice in order to predict AGB as compared to traditional regression models, using vegetation indices as input variables. Most of the earlier studies did not perform spatial interpolation and mostly restricted their estimation within pre-defined reference units thus limiting the model's applicability. To address this disadvantage, geostatistical interpolation was done in this study to map the vegetation cover of the entire area (Gerstmann et al. 2016). The objectives were set as (i) identification of the best model to predict AGB of sample locations using vegetation indices as explanatory variable and (ii) measuring the accuracy of that model by comparing the geostatistically interpolated layers of observed and estimated AGB.

2. Methododology

2.1. Study area

For this study, Shivpuri Tehsil (sub-district of Shivpuri district) of Madhya Pradesh, India was selected (Figure 1). It falls under the central highland physiographic zone of Bundelkhand

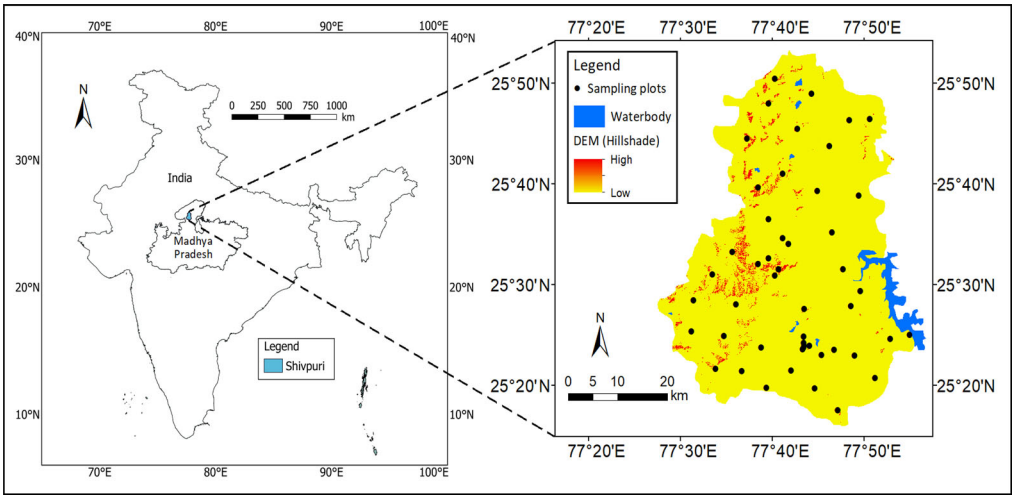

Figure 1. Location of the study area in Madhya Pradesh, India and distribution of the sampling points over the digital elevation model (DEM) Hill shade map of the area.

region and has semi-arid bio-climate (Bhattacharyya et al. 2008). The area is characterized by rugged up topography with ridges running from north to south with valleys in between. The tehsil has 190 villages and covers about 1956 km^2 area between 25°16′ N to 25°52′ N latitudes and 77°47′ E to 77°42′ E longitudes (Sheikh et al. 2011). The annual rainfall is approximately 100 cm and the maximum is concentrated during monsoon months only (ClimWAT software), resulting in an extreme weather condition like frequent droughts, (Gupta et al. 2014).

The area belongs to agro-pastoral land cover. Grasslands and rangelands, degraded woodlands with tall grasses and interspersed scrublands are common. Although harsh climatic condition restricts the agricultural activities of this region, a large part of land is under cultivation. As per the ground survey and Census of India, Madhya Pradesh (2011), major crops are rice, sorghum and wheat among cereals, gram and groundnut among the other crops, which are found in this region. The natural vegetation of the area is dominated by dry mixed deciduous trees, open woodlands etc. (Kale et al. 2004; Deb et al. 2017). *Acacia catechu, Buchanania lanzan, Anogeissus pendula, Azadirachta indica, Dalbergia sissoo, Tectona grandis,* etc. have been found as the dominant tree species.

2.2. Field survey and biomass estimation

A total of 100 ground control points (GCPs) were selected using a handheld 12 channel global positioning system receiver (Garmin Ltd., Schaffhausen, Switzerland). Along with ground observations, the Survey of India topographic map, Forest Survey of India (FSI) maps (ISFR. 2011), Google Earth Pro^{TM} data were used to obtain the land use/cover and vegetation distribution. Five vegetation types/classes were identified *viz.* barren land (class I), land under agricultural and allied activities (class II), grassland, rangeland, degraded natural vegetation (class III), sparse natural vegetation/plantation (class IV), and dense natural vegetation/plantation (class V) (Figure 2). In this region, intermittent scrubs are common near the cultivated areas, which usually spread into the agricultural lands during dry fallow season. Thus, they have also been considered within the class II vegetation types.

Out of 49 selected plots, nested sampling methodology was followed. The AGB of trees with > 35 cm diameter at breast height (DBH) (30 m × 30 m), 10-35 cm DBH

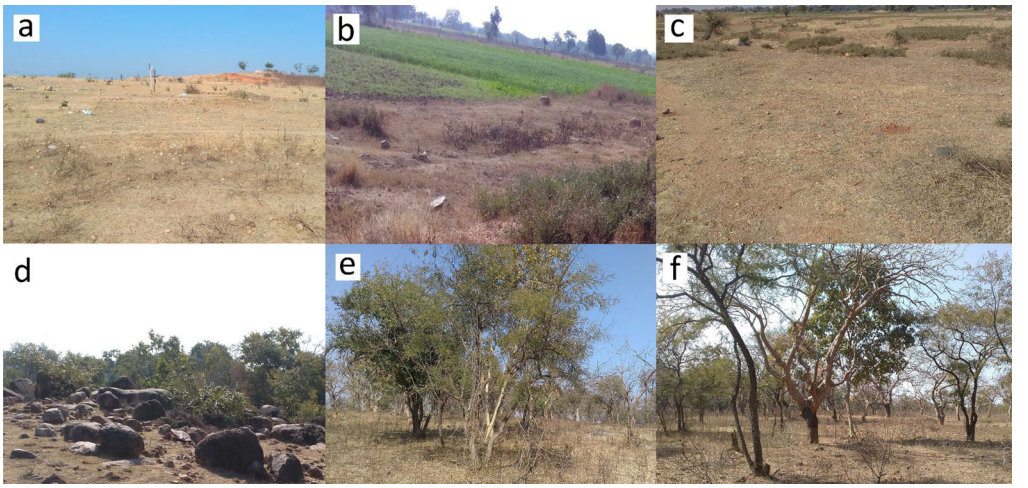

Figure 2. Types of vegetation present in the study area *viz.* (a) barren land (class I), (b) and (c) land under agricultural and allied activities (class II), (d) degraded natural vegetation (class III), (e) sparse natural vegetation/plantation (class IV) and (f) dense natural vegetation (class V).

(15 m × 15 m), and < 10 cm DBH (5 m × 5 m), and shrubs, herbs, grasses, agricultural crops (1 m × 1 m) were measured (Deb et al. 2017). To calculate tree biomass, individual tree volume was calculated (from the collected field data) using site-specific tree volume equations (ISFR 2003; Chave et al. 2008). For a few tree species, whose volume equations were not available, geometric relationships were used to approximate the volume of standing tree bole and were multiplied with specific gravity and expansion factor of each trees to estimate the biomass (Chave et al. 2008; Deb et al. 2017). The mean biomass expansion factor value of 1.5 (Deb et al. 2017) was used to calculate AGB of each plot by summing the phytomass of all the plant species (i.e. trees + shrubs + herbs).

2.3. Satellite data processing and selection of vegetation indices

Cloud free Landsat 8 image (30 m × 30 m) on 15th February, 2016, covering the study area, was obtained from the United States Geological Survey (https://earthexplorer.usgs.gov/). No geometric and radiometric corrections were required for this level 1 data (USGS 2018). The image was, however, processed for atmospheric correction through 'Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes' (FLAASH) module in ENVI 5.0 software (Deb et al. 2018). The image was resampled using Nearest Neighbor algorithm (Deb et al. 2014) in ERDAS Imagine 16 software. The digital numbers (DN) of the pixels were converted to spectral reflectance using rescaling factor following (USGS 2018):

$$\rho_\lambda{}' = M_\rho Qcal + A_\rho \tag{1}$$

Where, ρ_λ': planetary reflectance, without correction for solar angle, M_ρ : band-specific multiplicative rescaling factor from the metadata, A_ρ : band specific additive rescaling factor, and Qcal: Level 1 pixel value in DN.

The sun-angle was corrected as:

$$\rho_\lambda = \frac{\rho_\lambda{}'}{\text{coscos}\ (\theta_{SZ})} = \frac{\rho_\lambda{}'}{\text{sinsin}\ (\theta_{SE})_\rho} \tag{2}$$

ρ_λ : planetary reflectance, SZ: local solar zenith angle, SE: local sun elevation angle

The reflectance value of band 4 and 5 were used to generate vegetation indices like the Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Modified Soil Adjusted Vegetation Index (MSAVI) and Enhanced Vegetation Index 1 and 2 (EVI-1 and EVI-2). All these indices were used to contain the diverse vegetation cover of the area:

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{R}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{R}}} \quad \text{(Deb et al. 2017)} \tag{3}$$

$$\mathrm{SAVI} = \frac{1.5 \times (\rho_{\mathrm{NIR}} - \rho_{\mathrm{R}})}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{R}} + 0.5} \quad \text{(Ren and Zhou 2019)} \tag{4}$$

$$\mathrm{MSAVI} = 0.5 \times \left[(2 \times \rho_{\mathrm{NIR}} + 1) - \sqrt{(2 \times \rho_{\mathrm{NIR}} + 1)^2 - 8 \times (\rho_{\mathrm{NIR}} - \rho_{\mathrm{R}})} \right] \quad \text{(Jin et al. 2014)} \tag{5}$$

$$\mathrm{EVI} - 1 = 2.5 \times \left[\frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{R}}}{\rho_{\mathrm{NIR}} + 6 \times \rho_{\mathrm{R}} - 7.5 \times \rho_{\mathrm{B}} + 1} \right] \quad \text{(Jiang et al. 2008)} \tag{6}$$

$$\mathrm{EVI} - 2 = 2.5 \times \left[\frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{R}}}{\rho_{\mathrm{NIR}} + 2.4 \times \rho_{\mathrm{R}} + 1} \right] \quad \text{(Jin et al. 2014)} \tag{7}$$

Here, ρ_{NIR}, ρ_R and ρ_B indicated reflectance in near infrared, red and blue wavelength respectively.

2.4. Modelling approach

Modelling was done aiming to predict AGB using above-mentioned indices. Either the best fitted index or a combination of all the indices were used as the explanatory variable(s). Both traditional regression models and machine learning approaches were tried.

Traditional methods: At first, linear and nonlinear modelling techniques were used to determine AGB. While linear modelling is the simplest modelling technique, nonlinear modelling is well accepted, effective technique for curve fitting (Keshtegar et al. 2016) and successful in biomass estimation (Payandeh 1983; Tilly et al. 2015). Nonlinear relationship between vegetation and spectral reflectance data were also well documented (Thenkabail et al. 2000; Kross et al. 2015). Exponential model ($y = a{*}e^{bx}$) was selected as the preferred nonlinear model following its success in several earlier studies (Heath et al. 1996; Popescu 2007; Deb et al. 2017). As both the linear and exponential model can work with only one explanatory variable (Viña et al. 2011; Jin et al. 2014; Deb et al. 2017), association study was carried out between ground observed AGB and the values of different vegetation indices to determine the best vegetation index (Bendig et al. 2015; Ren and Zhou 2019).

This study further used multiple linear regression (MLR) model, using multiple vegetation indices together (Mutanga et al. 2012; Bendig et al. 2015). Forward selection approach was used to perform stepwise regression. The expression of MLR model can be expressed as:

$$\widehat{\mathrm{Y_n}} = \mathrm{a_0} + \mathrm{a_1X_{n1}} + \mathrm{a_2X_{n2}} + \ldots + \mathrm{a_iX_i} + \in_{\mathrm{n}} \rightarrow (\mathrm{X_{n1}} + \ldots + \mathrm{X_{ni}}) + \in_{\mathrm{n}} \tag{8}$$

Here, the random variable Y_n represented response for case n, a_0 was the intercept, i was no. of independent variables (vegetation indices), $a_{1\ldots i}$ and $X_{1\ldots i}$ were regression coefficient and values of independent variables respectively.

Support vector machine method: The SVM is a binary classifier for linear and nonlinear classification of regression problems with detection of outliers using an intuitive model representation (Drake et al. 2006; Lu et al. 2016). In this study, use of SVM

restricted the deviation between observed and estimated data to a small value ε, where the training data followed a pattern $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset X \times R$. Here x was function of independent variable, y was dependent variable, X and R were input patterns and set of real numbers, respectively. The e1071 package of R 3.4.4 was used to train the SVM (Liaw and Wiener 2002). The collected data was grouped into two sets namely training and testing datasets. The trained model was validated using fit statistics, verified against test data and then applied to predict AGB using the best fitted vegetation index.

2.5. Spatial distribution

Spatial distribution of AGB was mapped according to the vegetation classes by using ordinary kriging in Geostatistical Analyst of ArcGIS 10.1 platform (ESRI Inc., USA) (Chakraborty et al. 2017). In prediction output surface, spherical model was used to interpolate AGB of un-sampled locations. These un-known AGB values, $y(u)$, were estimated as per weighted linear combinations of known AGB values, $y(u_a)$, situated in a proximal neighbourhood (Santra et al. 2017). The following equation describes ordinary kriging:

$$y^*(u) = \sum_{n(u)}^{a=1} \lambda_a y(u_a) \text{ (Santra et al. 2017)} \quad (9)$$

Here λ_a is the weight allotted to value $y(u_a)$ situated within a neighborhood, centered on u. To measure spatial correlation, semi-variogram was used (Chakraborty et al. 2017). It included two standard spatial parameters *viz*. nugget and sill. While the nugget indicates the micro-scale variability measurement error, the partial sill describes amount of variation as per the spatial correlation structure (Santra et al. 2017).

3. Results and discussion

Ground survey of 49 sample plots showed a wide range of AGB spreading over a value as low as 0.17 to as high as 82.99 Mg ha^{-1}, which were distributed over croplands, grass and rangelands, and other vegetation classes. A total of 3, 9, 10, 23, and 4 plots were identified in class I, II, III, IV, and V, respectively. The ground observed AGB was correlated with satellite data derived vegetation indices. Apart from the most widely used NDVI, SAVI and MSAVI were chosen as they were better where barren land dominated in the background (Bendig et al. 2015; Ren and Zhou 2019). Further, EVI-1 and EVI-2 were selected for their better performance in areas dominated by dense vegetation/plantation and their capacity to curb background and atmospheric noises (Jiang et al. 2008; Liao et al. 2015).

3.1. Use of traditional regression models to determine AGB

To determine the best suitable vegetation index for linear and exponential models, values of vegetation indices were compared with ground observed AGB using Pearson product-moment correlation (Table 1). Both EVI-1 and EVI-2 showed significant correlations (p value ≤ 0.01) with ground-based AGB estimation under the vegetation class I; for rest of the classes, EVI-1 had the highest correlation (Table 1). This indicated the suitability of EVI-1 in traditional models for estimation of AGB. Figure 3 represented point-wise comparison of observed AGB through the ground survey with remote sensing derived EVI-1 values of the respective pixels. Here, the different coloured points and size of circles indicated range of satellite data derived EVI-1 values and ground observed AGB, respectively.

Table 1. Association study between aboveground biomass and vegetation indices.

			Pearson product-moment correlation				
Vegetation class	Mean	Variance	NDVI	SAVI	MSAVI	EVI-1	EVI-2
Barren land (class I)	0.93	0.34	0.05	0.04	0.31	0.98**	0.84**
Land under agricultural and allied activities (class II)	3.40	0.62	0.04	0.04	–0.46	0.93**	0.31
Grassland, rangeland, degraded natural vegetation (class III)	7.26	3.21	0.26	0.26	–0.64	0.95**	0.17
Sparse natural vegetation/plantation (class IV)	15.77	13.57	0.27	0.27	0.01	0.75*	0.30
Dense natural vegetation/plantation (class V)	46.67	441.43	0.43	0.43	0.31	0.83**	0.65*

** and * indicate p value $\leq$ 0.01 and $\leq$ 0.05 respectively. NDVI: Normalized Difference Vegetation Index, SAVI: Soil Adjusted Vegetation Index, MSAVI: Modified Soil Adjusted Vegetation Index, EVI-1: Enhanced Vegetation Index 1, and EVI-2: Enhanced Vegetation Index 2

A visual comparison of point colours and circle sizes indicated a high possible correlation within EVI-1 values and AGB.

For quantitative understanding of the relationship of EVI-1 with AGB, linear and exponential models were developed for each of the vegetation classes considering EVI-1 values as explanatory variable and AGB as response variable (Table 2). Statistical measures like regression coefficient (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE) were employed as evaluation indices for the accuracy of statistical models (Wallach and Goffinet 1989). While higher R^2 values indicated better statistical fit for a given dataset, it was just the opposite for RMSE and MAPE (Deb et al. 2017). Comparison of the R^2 values inferred that exponential model had better statistical fit for the vegetation classes I (barren land) (R^2: 0.97), II (land under agricultural and allied activities) (R^2: 0.97) and V (dense natural vegetation/plantation) (R^2: 0.81), while linear model was better for vegetation class IV (sparse natural vegetation/plantation) (R^2: 0.61). According to RMSE and MAPE values, linear model was statistically better than exponential model to predict AGB for all the vegetation classes using EVI-1 as explanatory variable. Therefore, the overall superiority of any of these models remained undeceive.

This study also considered use of MLR, which is a well adopted technique in environmental modelling and provides option to consider more than one input variable together (Mutanga et al. 2012). In this study, MLR technique was used considering the above mentioned vegetation indices as independent variables (Basuki et al. 2012). In stepwise regression, performance of model improved gradually with subsequent addition of variables which caused significant improvement in the fit. The statistical parameters (R^2, RMSE, and MAPE values) indicated a better fit of MLR models for AGB estimation under each of the individual vegetation classes in comparison to linear and exponential models (Table 3). Only exception was the class-II (land under agricultural and allied activities), for which exponential model fitted better than MLR in terms of R^2 values (0.97 and 0.91 in case of exponential and MLR respectively) while as per RMSE values, linear model was the best for class-II (0.26) followed by MLR (0.28) and exponential model (0.32). Based on the above, MLR appeared to be most acceptable for estimation of AGB using the satellite data derived vegetation indices.

3.2. Use of SVM to determine AGB

Use of SVM is ecological modelling is well established for its realistic computation, precise forecasting and high interpretability (Drake et al. 2006). Earlier researches indicated the potential of SVM to model AGB and other biophysical parameters of vegetation accurately (Gleason and Im 2012; Zhang and Liu 2013; Lu et al. 2016). To model AGB from satellite remote sensing by using SVM, this study used the most suitable vegetation

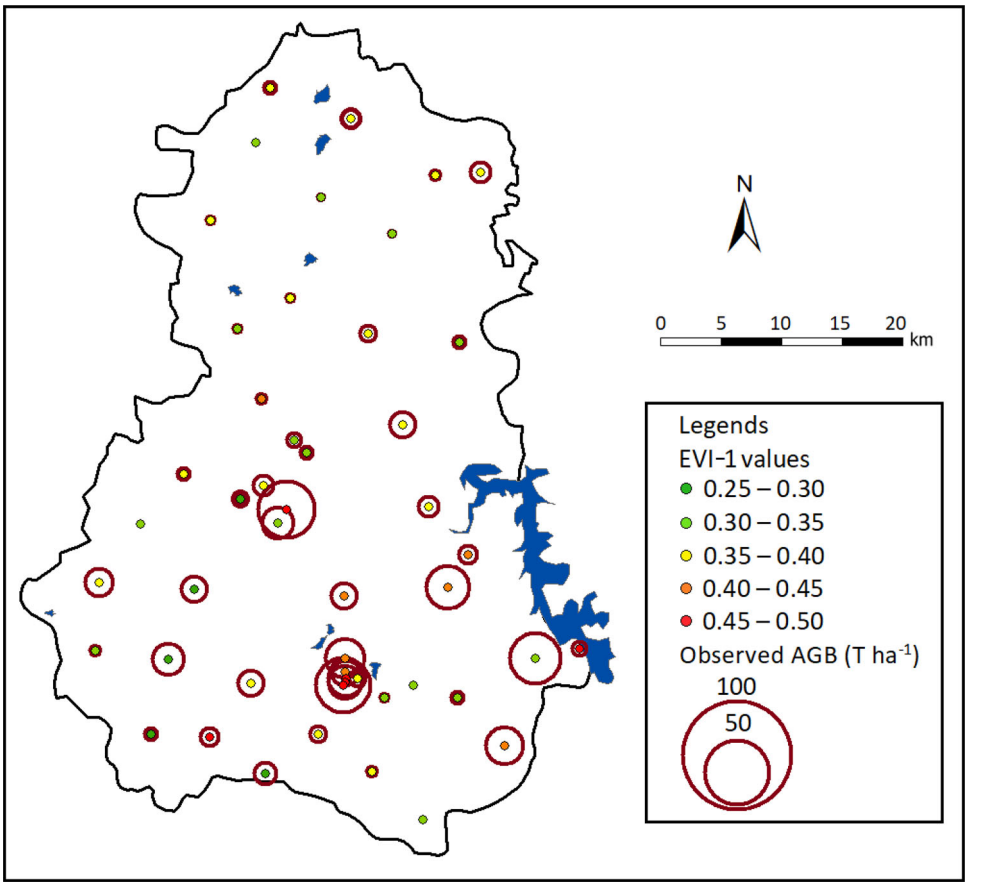

Figure 3. Comparison of point wise observed aboveground biomass (AGB) with enhanced vegetation index-1 (EVI-1) values of the sampling sites. Here, the AGB measured at ground survey and EVI-1 values were represented by ring diameter and colored points respectively.

index as explanatory variable. To choose it, relationship of all the vegetation indices were tested with the ground observed AGB of all the sites (Figure 4). As the figure indicated, croplands, grass and rangelands and other vegetation classes were most distinctly and separately presented with least number of overlapping points when the AGB was plotted against EVI-1. This could be a reasonable basis to select EVI-1 since maximization of differences between classes is the most important in SVM at conceptual level (Smola and Schölkopf 2004). This selection was further corroborated by highest R^2 value (0.73) and lowest residual standard error (RSE) (10.19) in the EVI-1 and observed AGB relationship. Among the other vegetation indices, MSAVI showed the least correlation with AGB ($R^2 = 0.01$, RSE = 19.54). Figure 4 indicated possible saturation problem of NDVI and SAVI with increase in the biomass density. Earlier studies also reported this limitation of NDVI, which was due to asymptotical way of saturation of red and near infrared band after a specific high vegetation density (Thenkabail et al. 2000; Mutanga and Skidmore 2004).

For application of SVM, the data set was divided into train and test data set comprising 80% and 20% of total data points respectively (Gleason and Im 2012). This calibration and validation were attempted using the model functions epsilon and cost (Mukherjee et al. 1997; Ding et al. 2016). Value range of 0 and 1 with 0.01 steps was used for epsilon while a range of values from 2 to 2^9 was used for cost. An iterative modeling approach

Table 2. Linear and exponential models for each vegetation classes using Enhanced Vegetation Index-1 as independent variable for determination of aboveground biomass.

		Fit statistics for linear model				Fit statistics for exponential model		
Vegetation classes	Linear model	R^2	RMSE	MAPE	Exponential model	R^2	RMSE	MAPE
Barren land (class I)	$-6.38 + 49.09 \times$ EVI-1	0.90	0.11	15.64	$0.0075 * e^{46.98 * EVI1}$	0.97	2.77	31.09
Land under agricultural and allied activities (class II)	$-13.70 + 94.31 \times$ EVI-1	0.88	0.26	6.12	$0.045 * e^{23.71 * EVI1}$	0.97	0.32	8.90
Grassland, rangeland, degraded natural vegetation (class III)	$-46.34 + 269.20 \times$EVI-1	0.92	0.86	5.86	$2.797 * e^{3.93 * EVI1}$	0.92	1.99	20.38
Sparse natural vegetation/plantation (class IV)	$-124.44 + 677.69 \times$ EVI-1	0.61	2.18	13.80	$0.0087 * e^{36.22 * EVI1}$	0.58	2.23	15.28
Dense natural vegetation/plantation (class V)	$-106.16 + 645.31 \times$ EVI-1	0.71	10.75	18.51	$4.866 * e^{9.42 * EVI1}$	0.81	12.37	22.69

RMSE and MAPE are abbreviation of root mean square error and mean absolute percentage error respectively. EVI-1: Enhanced Vegetation Index 1.

Table 3. Multiple linear regression model for each vegetation classes using combination of different vegetation indices as independent variables for determination of aboveground biomass.

		Fit statistics		
Vegetation classes	Multiple regression models	R^2	RMSE	MAPE
Barren land (class I)	$-5.67 + 61.44 \times$ EVI-1 $- 10.06 \times$ EVI-2 $- 2.77 \times$ SAVI $- 4.43 \times$ MSAVI	0.97	0.01	7.03
Land under agricultural and allied activities (class II)	$-17.87 + 105.02 \times$ EVI-1 $+ 9.44 \times$ EVI-2 $+ 7.06 \times$ MSAVI	0.91	0.28	5.89
Grassland, rangeland, degraded natural vegetation (class III)	$-297.94 + 1447.91 \times$ EVI-1 $+ 120.79 \times$ EVI-2 $- 13.53 \times$ MSAVI	0.94	0.74	5.16
Sparse natural vegetation/ plantation (class IV)	$-131.10 + 665.98 \times$ EVI-1 $+ 69.17 \times$ EVI-2	0.67	2.01	12.20
Dense natural vegetation/ plantation (class V)	$-126.40 + 776.86 \times$ EVI-1 $- 699.35 \times$ EVI-2 $+ 85.85 \times$ SAVI $+ 497.51$	0.84	10.28	16.01

RMSE and MAPE are abbreviation of root mean square error and mean absolute percentage error respectively. SAVI: Soil Adjusted Vegetation Index, MSAVI: Modified Soil Adjusted Vegetation Index, EVI-1: Enhanced Vegetation Index 1, and EVI-2: Enhanced Vegetation Index 2

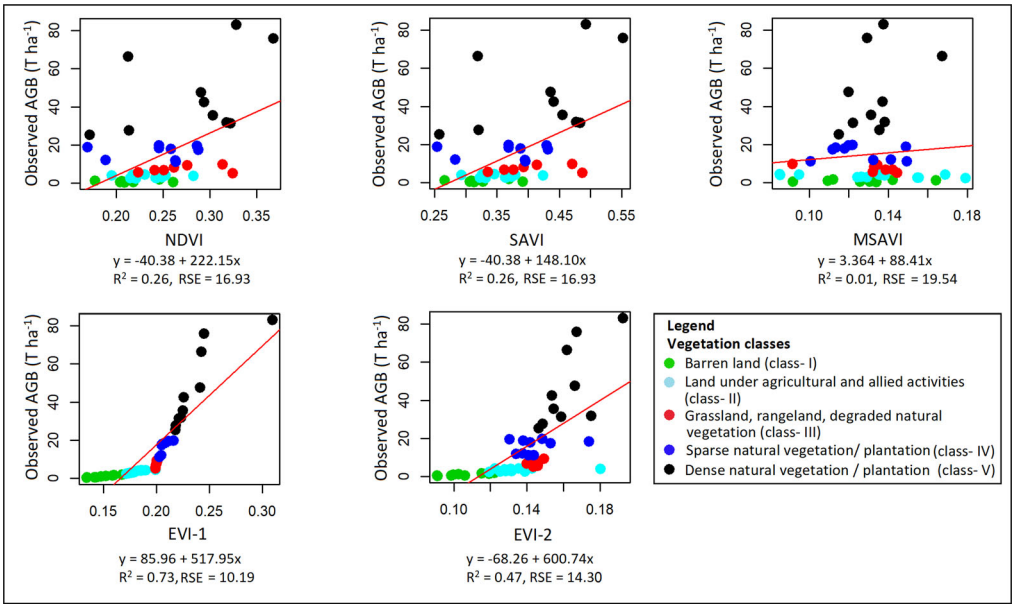

Figure 4. Relationship between different vegetation indices to aboveground biomass (AGB), different coloured points indicate different vegetation classes.

yielded a total of 909 (101×9) models to find out the best performing one. This whole set of models was generated through a grid search method using different values of epsilon and cost with an objective to minimize the error to the least. Model with epsilon and cost values of 0.04 and 512 respectively was selected as the best model. In SVM, support vector regression uses a kernel function to curb training error and model complexity (Axelsson et al. 2013; Ding et al. 2016). Here, simple linear model was selected as the kernel function to model AGB. The plotting of observed AGB (at ground) with predicted AGB (in SVM using EVI-1 as explanatory variable) indicated a high degree of correlation

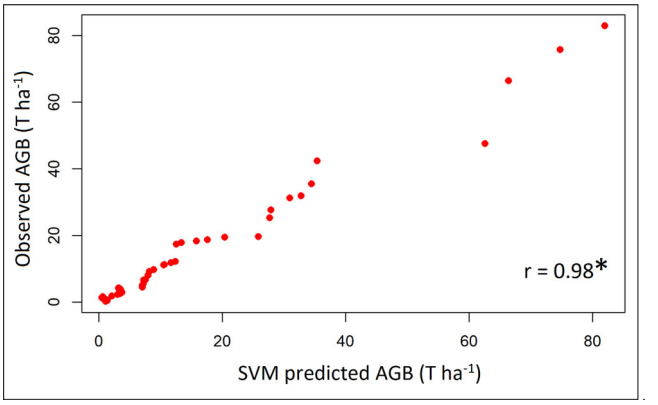

Figure 5. Correlation between observed and support vector machine (SVM) predicted aboveground biomass (AGB) (* indicates statistical significance at $p < 0.01$).

($r = 0.98$, $p < 0.01$) (Figure 5). For all the vegetation classes and overall vegetation of the area, SVM showed high R^2 (Class I: 0.98, II: 0.99, III: 0.98, IV: 0.80, V: 0.96), which were greater than all the previously discussed traditional models (Table 4). The SVM model also indicated better statistical fit for all the vegetation classes than traditional models in terms of RMSE and MAPE, except for class I (barren land). With respect to the first objective, these results thus established the SVM as best model to predict the AGB, using EVI-1 as an input variable. This success of SVM might partially be linked with its stability in presence of small training data vis-à-vis its ability to function with the outliers (Ali et al. 2015; Lu et al. 2016).

3.3. Spatial distribution pattern of vegetation

The heterogeneity of the land cover (from barren lands to dense vegetation/plantation) indicated the requirement of continuous land cover mapping. In this context, the ordinary kriging was applied to interpolate ground observed and SVM modeled AGB of the whole area using the discrete values of the studied sites (Figure 6) (Yang et al. 2008). Spherical model was used following its wide acceptability and popularity in mapping of natural resources (Yang et al. 2008; Santra et al. 2017). The model semi-variogram, which described the extent of the spatial correlation of any two sites, was described in Table 5. The nugget/sill ratio of the observed (0.52) and predicted AGB map (0.56) inferred a moderate spatial dependence and the suitability of the fitted semi-variogram model in generating the AGB pattern of the area (Chakraborty et al. 2017).

The geographical similarity of vegetation types and trends in the spatial interpolation maps was evident (Figure 6), which was supported by the alike RMSE value of both the maps (Table 5). For further understanding, area under each of the land cover classes were measured. Both the maps indicated approximately 6 km^2 area under dense natural vegetation/plantation (class V) while a large portion of the study area was identified as sparse natural vegetation/plantation (class IV) (1125 km^2 in observed and 1134 km^2 in predicted map). Ground survey also indicated sparse vegetation as a very usual land cover of this semi-arid area. The extent of grassland, rangeland and degraded natural vegetation (class III) was 343 km^2 and 374 km^2 in observed and predicted AGB maps respectively. Geostatistical analysis indicated a noteworthy land share by agricultural and allied activities (class II) (481 km^2 in observed and 442 km^2 in predicted map). However, observed

Table 4. Statistical measures of support vector machine model for each vegetation classes using Enhanced Vegetation Index-1 as the independent variable for determination of aboveground biomass.

	Vegetation classes				
Fit statistics	Barren land (class I)	Land under agricultural and allied activities (class II)	Grassland, rangeland, degraded natural vegetation (class III)	Sparse natural vegetation/ plantation (class IV)	Dense natural vegetation/ plantation (class V)
R^2	0.98	0.99	0.98	0.80	0.96
RMSE	0.09	0.10	0.45	1.60	6.00
MAPE	16.15	2.69	5.04	6.85	9.70

RMSE and MAPE are abbreviation of root mean square error and mean absolute percentage error respectively.

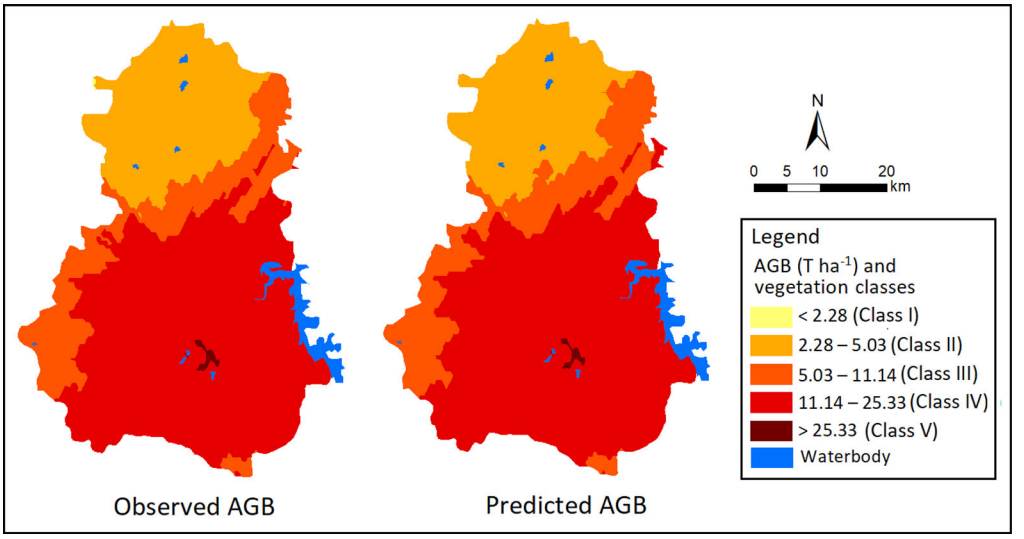

Figure 6. Predicted continuous surface maps of vegetation of the study area, produced by ordinary kriging based on a) observed aboveground biomass (AGB) and b) support vector machine predicted AGB.

Table 5. Ordinary kriging semi-variogram model parameters for the aboveground biomass (AGB).

	Model	Nugget	Sill	Nugget/Sill	RMSE
Observed AGB	Spherical	0.377	0.723	0.52	0.94
Predicted AGB	Spherical	0.398	0.708	0.56	0.95

AGB map indicated a small amount of barren land (class I) (~1 km^2) also, which was not present in the predicted AGB map. In response to the second objective, this quantitative comparison of vegetation distribution in observed and predicted AGB maps confirmed the high precision of SVM to model AGB successfully.

4. Conclusions

This study established SVM as a better option over the traditional regression models to simulate AGB of discrete plots of a Tehsil (sub-district), covering diverse range of land covers under agro-pastoral ecology. The best fitted vegetation index, derived from satellite data, was used as explanatory variable. This predicted dataset was further used to generate AGB map of the whole area. It could an important pilot scale project to initiate such work for a larger area, for better monitoring and management of natural resources. With

improved optimization and enhanced robustness, it can possibly be applied to model land cover ecologies with limited ground accessibility like mangroves, rain-forests etc. However, the passive optical remote sensing always has some limitations like spectral saturation, cloud cover etc. Thus, use of SAR data or airborne light detection and ranging (LiDAR) data can be considered in combination with optical space-data for estimation of AGB in future. Scientists have already initiated use of deep learning-based workflow or stacked sparse autoencoder network while estimating AGB using these input variables (Shao et al. 2017; Zhang et al. 2019). Further, comparative superiority of different machine learning techniques for AGB estimation using remotely sensed data will also be an important future concern.

Acknowledgement

Sincere thanks to all the people who helped during field data collection, the local villagers of the Buldelkhand region for their valuable assistance.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors are thankful to Indian Council of Agricultural Research and Director, Indian Grassland and Fodder Research Institute for funding this research work (Project No: CRSCIGFRISIL20150302).

References

Adam E, Mutanga O, Odindi J, Abdel-Rahman EM. 2014. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. Int J Remote Sens. 35(10):3440–3458.

Ali I, Greifeneder F, Stamenkovic J, Neumann M, Notarnicola C. 2015. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. Remote Sens. 7(12): 16398–16421.

Anaya JA, Chuvieco E, Palacios-Orueta A. 2009. Aboveground biomass assessment in Colombia: a remote sensing approach. For Ecol Manag. 257(4):1237–1246.

Axelsson C, Skidmore AK, Schlerf M, Fauzi A, Verhoef W. 2013. Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. Int J Remote Sens. 34(5):1724–1743.

Baccini A, Laporte N, Goetz SJ, Sun M, Dong H. 2008. A first map of tropical Africa's above-ground biomass derived from satellite imagery. Environ Res Lett. 3(4):045011.

Basuki TM, Skidmore AK, van Laake PE, van Duren I, Hussin YA. 2012. The potential of spectral mixture analysis to improve the estimation accuracy of tropical forest biomass. Geocarto Int. 27(4): 329–345.

Basuki TM, Van Laake PE, Skidmore AK, Hussin YA. 2009. Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests. For Ecol Manag. 257(8):1684–1694.

Bendig J, Yu K, Aasen H, Bolten A, Bennertz S, Broscheit J, Gnyp ML, Bareth G. 2015. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. Int J Appl Earth Obs Geoinf. 39:79–87.

Bhattacharyya T, Pal DK, Chandran P, Ray SK, Mandal C, Telpande B. 2008. Soil carbon storage capacity as a tool to prioritize areas for carbon sequestration. Curr Sci. 95:482–494.

Census of India, Madhya Pradesh. 2011. District census handbook Shivpuri. Madhya Pradesh: Directorate of Census Operations.

Chakraborty S, Man T, Paulette L, Deb S, Li B, Weindorf DC, Frazier M. 2017. Rapid assessment of smelter/mining soil contamination via portable X-ray fluorescence spectrometry and indicator Kriging. Geoderma. 306:108–119.
Chave J, Olivier J, Bongers F, Chatelet P, Forget PM, Van der Meer P, Norden N, Riera B, Charles-Dominique P. 2008. Aboveground biomass and productivity in a rain forest of eastern South America. J Trop Ecol. 24(4):355–366.
Dang ATN, Nandy S, Srinet R, Luong NV, Ghosh S, Kumar AS. 2019. Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam. Ecol Inform. 50:24–32.
Datta D, Deb S. 2017. Forest structure and soil properties of mangrove ecosystems under different management scenarios: experiences from the intensely humanized landscape of Indian Sunderbans. Ocean Coast Manage. 140:22–33.
Deb D, Singh JP, Deb S, Datta D, Ghosh A, Chaurasia RS. 2017. An alternative approach for estimating above ground biomass using Resourcesat-2 satellite data and artificial neural network in Bundelkhand region of India. Environ Monit Assess. 189(11):576.
Deb S, Ahmed A, Datta D. 2014. An alternative approach for delineating eco-sensitive zones around a wildlife sanctuary applying geospatial techniques. Environ Monit Assess. 186(4):2641–2651.
Deb S, Debnath MK, Chakraborty S, Weindorf DC, Kumar D, Deb D, Choudhury A. 2018. Anthropogenic impacts on forest land use and land cover change: modelling future possibilities in the Himalayan Terai. Anthropocene. 21:32–41.
Ding J, Li F, Yang GB, Chen LY, Zhang BB, Liu L, Fang K, Qin SQ, Chen YL, Peng YF, et al. 2016. The permafrost carbon inventory on the Tibetan Plateau: a new evaluation using deep sediment cores. Glob Change Biol. 22(8):2688–2701.
Drake JM, Randin C, Guisan A. 2006. Modelling ecological niches with support vector machines. J Appl Ecol. 43(3):424–432.
Gao Y, Liu X, Min C, He H, Yu G, Liu M, Zhu X, Wang Q. 2013. Estimation of the North-South transect of Eastern China forest biomass using remote sensing and forest inventory data. Int J Remote Sens. 34(15):5598–5610.
Gerstmann H, Doktor D, Gläßer C, Möller M. 2016. PHASE: a geostatistical model for the Kriging-based spatial prediction of crop phenology using public phenological and climatological observations. Comput Electron Agric. 127:726–738.
Gleason CJ, Im J. 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches. Remote Sens Environ. 125:80–91.
Gupta AK, Nair SS, Ghosh O, Singh A, Dey S. 2014. Bundelkhand drought: a retrospective analysis and way ahead. New Delhi: National Institute of Disaster Management.
Heath LS, Birdsey RA, Row C, Plantinga AJ. 1996. Carbon pools and fluxes in U.S. forest products. In: Apps M, Price D, editors. Forest ecosystems, forest management and the global carbon cycle. Berlin Heidelberg: Springer. p. 271–278.
Ibrahim AM, Bennett B. 2014. The assessment of machine learning model performance for predicting alluvial deposits distribution. Procedia Comput Sci. 36:637–642.
ISFR. 2003. Annexure III, Volume equations, State of forest report - 2003. Dehradun (India): Forest Survey of India, (Ministry of Environment and Forests).
ISFR. 2011. Chapter 2: Forest Cover. State of forest report-2011. Dehradun (India): Forest Survey of India, (Ministry of Environment and Forests).
Jaber SM. 2019. On the relationship between normalized difference vegetation index and land surface temperature: MODIS-based analysis in a semi-arid to arid environment. Geocarto Int. online first. DOI: 10.1080/10106049.2019.1633421.
Jiang Z, Huete AR, Didan K, Miura T. 2008. Development of a two-band enhanced vegetation index without a blue band. Remote Sens Environ. 112(10):3833–3845.
Jin Y, Yang X, Qiu J, Li J, Gao T, Wu Q, Zhao F, Ma H, Yu H, Xu B. 2014. Remote sensing-based biomass estimation and its spatio-temporal variations in temperate Grassland, Northern China. Remote Sens. 6(2):1496–1513.
Kale M, Singh S, Roy PS, Deosthali V, Ghole VS. 2004. Biomass equations of dominant species of dry deciduous forest in Shivpuri district, Madhya Pradesh. Curr Sci. 87:683–687.
Keshtegar B, Piri J, Kisi O. 2016. A nonlinear mathematical modeling of daily pan evaporation based on conjugate gradient method. Comput Electron Agric. 127:120–130.
Kross A, Mcnairn H, Lapen D, Sunohara M, Champagne C. 2015. International Journal of Applied Earth Observation and Geoinformation assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. Int J Appl Earth Obs Geoinf. 34:235–248.

Liao Z, He B, Quan X. 2015. Modified enhanced vegetation index for reducing topographic effects. J Appl Remote Sens. 9(1):096068.
Liaw A, Wiener M. 2002. Classification and regression by random forest. R News. 2/3:18–22.
Lu D, Chen Q, Wang G, Liu L, Li G, Moran E. 2016. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. Int J Digit Earth. 9(1):63–105.
McRoberts RE, Naesset E, Gobakken T. 2013. Inference for lidar-assisted estimation of forest growing stock volume. Remote Sens Environ. 128:268–275.
Mukherjee S, Osuna E, Girosi F. 1997. Nonlinear prediction of chaotic time series using support vector machines. In: Neural Networks for Signal Processing-VII. Proceedings of the IEEE Workshop. Amelia Island: IEEE. p. 511–520.
Muldavin EH, Moore DI, Collins SL, Wetherill KR, Lightfoot DC. 2008. Aboveground net primary production dynamics in a northern Chihuahuan Desert ecosystem. Oecologia. 155(1):123–132.
Mutanga O, Adam E, Cho MA. 2012. High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm. Int J Appl Earth Obs Geoinf. 18: 399–406.
Mutanga O, Skidmore AK. 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. Int J Remote Sens. 25(19):3999–4014.
Muttil N, Chau KW. 2007. Machine-learning paradigms for selecting ecologically significant input variables. Eng Appl Artif Intell. 20(6):735–744.
Neumann M, Saatchi SS, Ulander LMH, Fransson R. 2012. Assessing performance of L- and P-band polarimetric interferometric SAR data in estimating boreal forest above-ground biomass. IEEE Trans Geosci Remote Sens. 50(3):714–726.
Paine CET, Marthews TR, Vogt DR, Purves D, Rees M, Hector A, Turnbull LA. 2012. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. Methods Ecol Evol. 3(2): 245–256.
Payandeh B. 1983. Some applications of nonlinear regression models in forestry research. Forest Chron. 59(5):244–248.
Popescu SC. 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass Bioenergy. 31(9):646–655.
Rehman TU, Mahmud MS, Chang YK, Jin J, Shin J. 2019. Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. Comput Electron Agric. 156: 585–605.
Ren H, Zhou G. 2019. Estimating green biomass ratio with remote sensing in arid grasslands. Ecol Indic. 98:568–574.
Santi E, Paloscia S, Pettinato S, Fontanelli G, Mura M, Zolli C, Maselli F, Chiesi M, Bottai L, Chirici G. 2017. The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas. Remote Sens Environ. 200:63–73.
Santra P, Kumar M, Panwar NR, Das BS. 2017. Digital soil mapping and best management of soil resources: a brief discussion with few case studies. In: Rakshit A, Abhilash PC, Singh HB, Ghosh S, editors. Adaptive soil management: from theory to practices. Singapore: Springer Singapore. p. 3–38.
Shao Z, Zhang L. 2016. Estimating forest aboveground biomass by combining optical and SAR data: a case study in Genhe, Inner Mongolia, China. Sensors. 16(6):834.
Shao Z, Zhang L, Wang L. 2017. Stacked sparse autoencoder modeling using the synergy of airborne LiDAR and satellite optical and SAR data to map forest above-ground biomass. IEEE J Sel Top Appl Earth Observ Remote Sens. 10(12):5569–5582.
Sheikh MA, Kumar M, Bussman RW, Todaria NP. 2011. Forest carbon stocks and fluxes in physiographic zones of India. Carbon Balance Manag. 6(1):15.
Smola AJ, Schölkopf B. 2004. A tutorial on support vector regression. Stat Comput. 14(3):199–222.
Thenkabail PS, Smith RB, De Pauw E. 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. Remote Sens Environ. 71(2):158–182.
Tilly N, Aasen H, Bareth G. 2015. Fusion of plant height and vegetation indices for the estimation of barley biomass. Remote Sens. 7(9):11449–11480.
USGS. 2018. *Landsat mission* (online), Available from: https://landsat.usgs.gov [accessed January 2018].
Viña A, Gitelson AA, Nguy-Robertson AL, Peng Y. 2011. Comparison of different vegetation indices for the remote. Remote Sens Environ. 115(12):3468–3478.
Wallach D, Goffinet B. 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. Ecol Model. 44(3–4):299–306.
Yang F-g, Cao S-y, Liu X-n, Yang K-j. 2008. Design of groundwater level monitoring network with ordinary Kriging. J Hydrodyn. 20(3):339–346.

Yang J, Weisberg PJ, Bristow NA. 2012. Landsat remote sensing approaches for monitoring long-term tree cover dynamics in semi-arid woodlands: comparison of vegetation indices and spectral mixture analysis. Remote Sens Environ. 119:62–71.

Zhang L, Shao Z, Diao C. 2015. Synergistic retrieval model of forest biomass using the integration of optical and microwave remote sensing. J Appl Remote Sens. 9:96069.

Zhang L, Shao Z, Liu J, Cheng Q. 2019. Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data. Remote Sens. 11(12):1459.

Zhang Z, Liu X. 2013. Support vector machines for tree species identification using LiDAR-derived structure and intensity variables. Geocarto Int. 28(4):364–378.

Zhao P, Lu D, Wang G, Wu C, Huang Y, Yu S. 2016. Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. Remote Sens. 8(6):469.

Zheng D, Rademacher J, Chen J, Crow T, Bresee M, Le Moine J, Ryu SR. 2004. Estimating aboveground biomass using Landsat 7 ETM + data across a managed landscape in northern Wisconsin, USA. Remote Sens Environ. 93(3):402–411.