Applied Geography 96 (2018) 29–40

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest

Sujit Madhab Ghosh*, Mukunda Dev Behera

Centre for Oceans, Rivers, Atmosphere and Land Sciences, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, 721302, India

ARTICLE INFO

Keywords:
Aboveground biomass
SAR remote sensing
Random forest
Stochastic gradient boosting
Data synergy

ABSTRACT

Forest aboveground biomass (AGB) is an important factor for tracking global carbon cycle to tackle the impact of climate change. Among all available remote sensing data and methods, Synthetic Aperture Radar (SAR) data in combination with decision tree based machine learning algorithms has produced favourable results in estimating higher biomass values. Suitability of this method for dense tropical forests has not been properly checked with an adequate number of studies. In this study, aboveground biomass was estimated for two major tree species, *Shorea robusta,* and *Tectona grandis*, of Katerniaghat Wildlife Sanctuary, a tropical forest situated in northern India. Aboveground biomass was estimated by combining C-band SAR data from Sentinel-1A satellite, texture images generated from Sentinel-1A data, vegetation indices produced using Sentinel-2A data and ground inventory plots. Decision tree-based machine learning algorithms were used in place of parametric regression models for establishing a relationship between fields measured values and remotely sensed parameters. Using random forest model with a combination of vegetation indices with SAR backscatter as predictor variables shows the best result for *S. robusta* forest, with a coefficient of determination value of 0.71 and an RMSE value of 105.027 t/ha. In *T. grandis* forest best result can be found in the same combination but for stochastic gradient boosted model with a coefficient of determination value of 0.6 and an RMSE value of 79.45 t/ha. This study shows that Sentinel series satellite data has exceptional capabilities in estimating dense forest AGB and machine learning algorithms can be very helpful to do so.

1. Introduction

Tropical forests are currently the primary source of terrestrial carbon to the atmosphere (Houghton, Byers, & Nassikas, 2015). Therefore, they play a crucial role in the global carbon cycle. Tropical forests also act as a significant carbon sink and have the potential to become an essential factor in climate change mitigation if appropriately managed (Chazdon, Broadbent, Rozendaal, Bongers, & et al., 2016; Xu et al., 2017). A crucial step in this process will be the measurement of biophysical indicators of forest carbon content, e.g., tree canopy height or above ground biomass (AGB) (Houghton, Hall, & Goetz, 2009). Global initiatives developed under the United Nations Framework Convention on Climate Change (UNFCCC) to mitigate climate change also depends on carbon content information of forests (Campbell, 2009).

Forest biomass is mainly estimated by either field-based measurements (Salunkhe, Khare, Sahu, & Singh, 2016) or remotely sensed methods (Lu, 2006). However, measuring forest biomass on a regional scale using field-based methods is not feasible, as it will take enormous resources and consume too much time. Remote sensing systems, both optical and active sensors, has been proved to be an effective alternative for measurement and monitoring of forest biomass at various scales and landscapes (Ali, Greifeneder, Stamenkovic, Neumann, & Notarnicola, 2015; Babcock et al., 2015; Kaasalainen et al., 2015; Su et al., 2016). Active remote sensing methods like Radio Detection and Ranging (RaDAR) and Light Detection and Ranging (LiDAR) have been regularly used in recent aboveground biomass estimation studies (Carreiras, Melo, & Vasconcelos, 2013; Carreiras, Vasconcelos, & Lucas, 2012; Cartus, Santoro, & Kellndorfer, 2012; Chen, 2015; Deng, Katoh, Guan, Yin, & Li, 2014; Li, Glenn, Olsoy, Mitchell, & Shrestha, 2015; Puliti et al., 2017; Takagi et al., 2015; Wang et al., 2016; Zhao et al., 2016). While RaDAR beams directly interacts with trunk and other parts containing biomass (Joshi et al., 2017), LiDAR measures the height of vegetation (Ghosh & Behera, 2017; Ho Tong Minh et al., 2016) which is a proxy to vegetation biomass (Erten, Lopez-Sanchez, Yuzugullu, & Hajnsek, 2016). However, Synthetic Aperture Radar (SAR) is favoured over LiDAR due to its wall-to-wall coverage which is absent in all LiDAR systems (Su et al., 2016).

* Corresponding author.
E-mail address: sujitmghosh@iitkgp.ac.in (S.M. Ghosh).

https://doi.org/10.1016/j.apgeog.2018.05.011
Received 22 March 2018; Received in revised form 11 May 2018; Accepted 11 May 2018
Available online 16 May 2018
0143-6228/

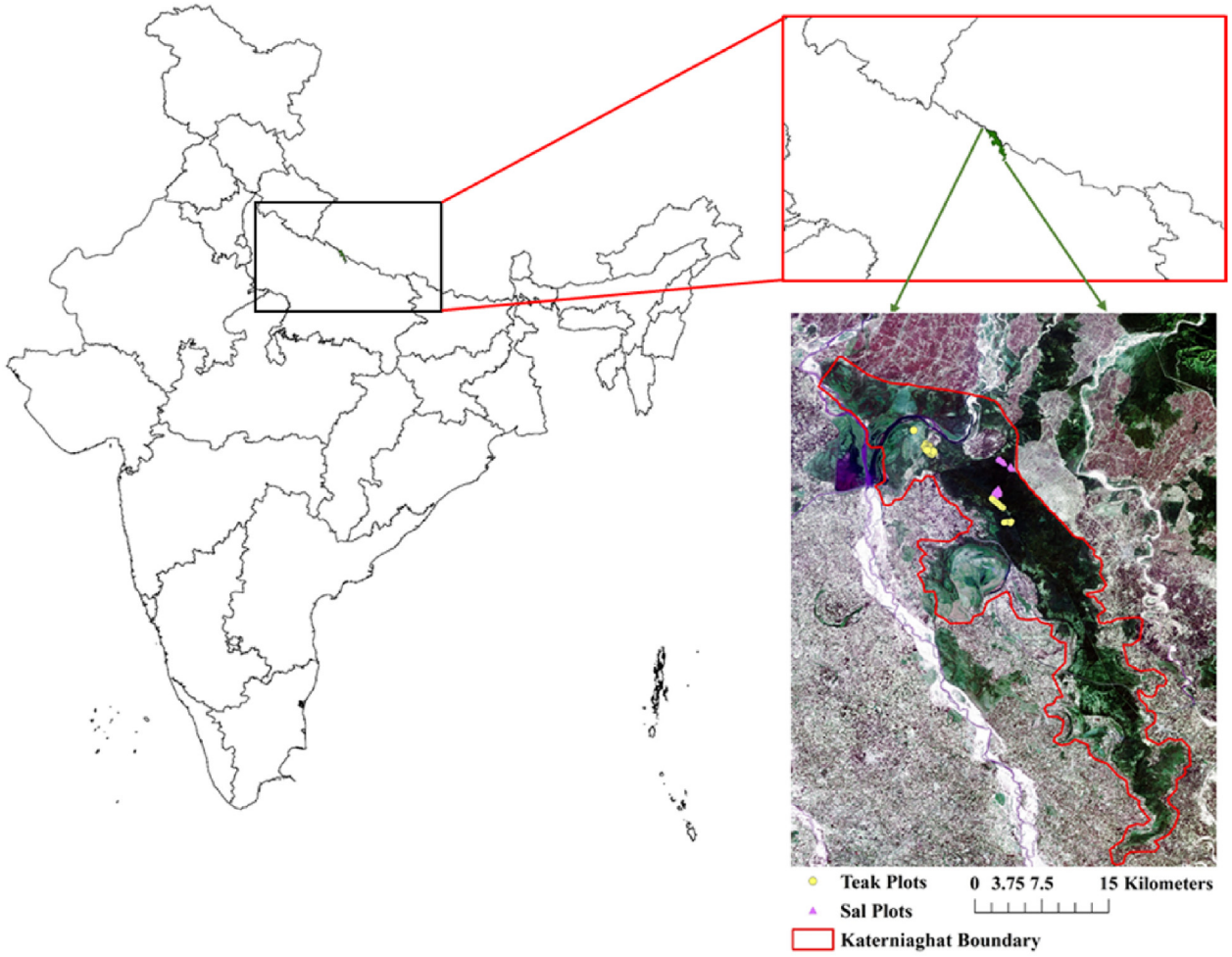

Fig. 1. The geographical location of study area.

Longer wavelength SAR bands are more suitable for biomass estimation as they can penetrate the canopy and directly interacts with trunk (Ouchi, 2013). However, working with longer wavelength data is not always possible due to less number of operational satellites and the high cost of the data. European Space Agency's Sentinel-1 mission has made high spatial resolution C-band SAR data freely available in the global scale. Still, this data alone will not be sufficient for estimating forest biomass due to C-band's low penetrating power. Beside SAR data, other remotely sensed data products such as vegetation indices (Das & Singh, 2012; Gasparri, Parmuchi, Bono, Karszenbaum, & Montenegro, 2010; Günlü, Ercanli, Başkent, & Çakır, 2014; Maynard, Lawrence, Nielsen, & Decker, 2007) and SAR image texture (Kuplich, Curran, & Atkinson, 2003; Sarker, Nichol, Ahmad, Busu, & Rahman, 2012; Sarker, Nichol, Iz, Ahmad, & Rahman, 2013) have showed their potential to be a good indicator of above ground biomass. Recent studies have concentrated on synergetic use of data from different sources (Castillo, Apan, Maraseni, & Salmo, 2017; Chang & Shoshany, 2016; Omar, Misman, & Kassim, 2017; Shen, Li, Huang, & Wei, 2016; Sinha et al., 2016; Su et al., 2016; Vaglio et al., 2017) rather than using only SAR data (Antropov, Rauste, Häme, & Praks, 2017; Ningthoujam et al., 2017; Solberg, Hansen, Gobakken, Næssset, & Zahabu, 2017) to improve the biomass estimation accuracy. Till date, only a few studies (Chang & Shoshany, 2016; Omar et al., 2017; Sinha et al., 2016) explored the effectiveness of multi-sensor data synergy in a dense tropical forest. Omar et al. (2017) combined PALSAR-2 and Sentinel-1A data for dipterocarp forests of Malaysia using multiple regression to found a maximum correlation value of 0.356 between observed and predicted biomass. In the Indian context, Sinha et al. estimated biomass for a tropical mixed deciduous forest and got a correlation value of 0.892 when they combined PALSAR backscatter with LANDSAT thematic mapper data derived vegetation indices and texture matrices. Chang and Shoshany (2016) fused backscatter and its ratio from Sentinel-1 and NDVI from Sentinel-2 to improve the biomass estimation accuracy of Mediterranean shrublands by 14%. Performance of Sentinel-1 and Sentinel-2 data combination for tropical forests of India has not been validated yet with any study.

The tropical forests of India are expected to go through most rapid and significant climate and vegetation changes over the next decades (Ravindranath, Joshi, Sukumar, & Saxena, 2006). This trend is a topic of substantial concern worldwide because approximately one-fifth of the global carbon stock is stored in tropical forests and they also act as a reservoir of almost one half of the above-ground carbon, stored in the vegetation of all biomes (Hunter, Keller, Victoria, & Morton, 2013). While the biomass of most temperate and boreal zones has been systematically inventoried at least once (Goodale et al., 2002), tropical regions suffer from operational limitations and consequent lack of data (Houghton et al., 2009). Some field-based and remote sensing data based study had tried to estimate forest biomass for some parts of India. Salunkhe et al. (2016) estimated biomass for the forests of central India using non-destructive field techniques and found the average biomass to be 31.8 t/ha for dry deciduous forest and 20.7 t/ha for a mixed deciduous forest. Thumaty et al. estimated biomass for the same region with L-band SAR data and found the average biomass to be 58 t/ha (Thumaty et al., 2016) which is almost twice of field estimated value. There is a need for the development of new methodologies using the latest dataset to remove or reduce the ambiguity between ground

measured values and remotely estimated values.

In this study, aboveground biomass of dense tropical forest has been estimated using C-band SAR data in combination with optical datasets and the efficiency of the method was checked. Two machine learning algorithms; Random Forest (RF) and Stochastic Gradient Boosting (SGB) were used in the study to predict inter-species aboveground biomass in a tropical forest in the Terai region of Uttar Pradesh, India.

2. Materials and methods

2.1. Study area

Katerniaghat Wildlife Sanctuary was chosen as the study area. It is situated in Uttar Pradesh, India, along with the Indo-Nepal border (Fig. 1) spreading over an area of 400 km^2 (Behera et al., 2015). The sanctuary has a typical Terai ecosystem, which is distinguished by alluvial plains inhabited by moist forests, wetlands, woodlands, and grasslands. The Terai landscape is widely known for its distinct biodiversity and high productivity, and it is considered among the prime ecoregions of the world (Behera et al., 2015). The geographical extent of the area lies between latitudes 81°24′ E and 81°19′ E and longitudes 28°60′ and 28°24′ N. The main vegetation of the sanctuary are *S. robusta* forests and *T. grandis* plantations (Chitale & Behera, 2014).

2.2. Forest inventory data collection and AGB estimation

The AGB of two major forest types, i.e., *S. robusta* and *T. grandis* were calculated based on ground inventory plots. Diameter at breast height (DBH) of the trees of ≥10 cm DBH were measured within 20 m × 20 m (0.04 ha) area in *S. robusta* forest and *T. grandis* plantation. Although increase in plot size improves biomass estimation efficiency (Kachamba, Ørka, Næsset, Eid, & Gobakken, 2017; Mauya et al., 2015; Næsset, Bollandsås, Gobakken, Solberg, & McRoberts, 2015), we limited it to 20 m × 20 m, so that it matches with the resolution of satellite data used in this study. The number of plots we generated within our permitted time is limited to 36 for *S. robusta* and 34 for *T. grandis*. Region and species-specific volume equations developed by the Forest Survey of India (FSI, 1996) were used to estimate the stand volumes from the inventory data. The stand volumes were converted to AGB using wood specific gravity information (Forest Research Institute, 1996) showed in Table 1.

AGB = Stem volume × Specific Gravity
AGB = above-ground biomass; Specific Gravity = tree-wise standard specific gravity value

The basic statistics for field-measured *S. robusta* and *T. grandis* biomass are shown in Table 3. It shows that *T. grandis* has higher average biomass than *S. robusta*.

2.3. Satellite data

The C band SAR data used for this study was obtained from Sentinel-1A satellite, which operates at a frequency of 5.405 GHz. Sentinel-1A is the first satellite of Sentinel series of Earth-observation satellites, which is designed by European Space Agency (ESA). ESA has specifically considered major observational requirements of end users, such as revisit period, coverage and quick data distribution while designing the system (Torres et al., 2012). As a result, Sentinel-1 data has a 12 day revisit period at the global scale, and the data is made available within hours of acquisition. The images for the study area was acquired in dual polarization (VV and VH) Interferometric Wide swath mode with a near incidence angle of 30.7139 and far incidence angle of 46.1383. Acquisition date was 3rd March 2017. Vegetation indices images were generated using ESA's Sentinel-2A satellite data. Sentinel-2 is designed as a part of Sentinel series to perform terrestrial observations in support of forest monitoring, land cover changes detection, and natural disaster management (Drusch et al., 2012). Sentinel-2 is a symbol of major advancement of optical remote sensing. With features like 13 bands covering the visible and NIR regions, resolutions as high as 10 m and a revisit period of 6 days for the constellation, it transcends all its predecessors. Both Sentinel-1 and Sentinel-2 data were downloaded freely from ESA Copernicus hub. Acquisition date of Sentinel-2A was 23rd February 2017. The use of these satellite data products is outlined in Fig. 2. It consists of mainly three parts. First the pre-processing of the data. Then, generation of derivative images such as texture parameters and vegetation indices. Finally, field measured biomass were correlated with the images using RF and SGB models. As the end product, a biomass map of two vegetation species is generated for the study area.

Table 1
Species specific volume equations as per Forest Survey of India (FSI, 1996).

Species Botanical name	Volume equation	Specific gravity
Shorea robusta	$\sqrt{V} = 0.16306 + 4.8991D + 1.57402\sqrt{D}$	0.726
Tectona grandis	$V = 0.08847 - 1.46936D + 11.98979D^2 + 1.970560D^3$	0.578

2.4. SAR data processing

Sentinel-1A data procured for the study was single look complex (SLC) slant range images. Sentinel Application Platform (SNAP) was used for preliminary processing of Sentinel-1A data. The images were first calibrated and resampled to a pixel size of 20 m, as this is the same size as the plot size taken for the field data. Speckle noise in the image was suppressed using Gamma MAP filter with a window size of 9 by 9 pixel. The Gamma MAP filter is more suitable for vegetated areas as it removes the speckle while preserving the finer details (Huang & Genderen van, 1996; Lopes, Nezry, Touzi, & Laur, 1990). The filtered image was rectified and geo-coded using 30 m Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM). Besides generating VH and VV images, their ratio image and the square root of their product image were also produced. Texture features were calculated using Grey Level Co-occurrence Matrix (GLCM) algorithm (Haralick, Shanmugan, & Dinstein, 1973) with a five by five-pixel window and for all four directions. Geo-coded and rectified image of each polarization was used as the input image for texture features calculation. Among all the textural features Contrast, Entropy and Sum Average were calculated as Kuplich et al. found that these texture matrices are closely related to biomass (Tatiana M Kuplich et al., 2003).

2.5. Optical data processing

Sentinel-2A data were first resampled to 20 m pixel size so that it matches the field plot size. Then it was atmospherically corrected via SEN2COR tool of SNAP, and then vegetation indices images were generated using atmospherically corrected data. The vegetation indices which were considered for this study were Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Renormalized Difference Vegetation Index (RDVI), and Soil-Adjusted Vegetation Index (SAVI). Equations used for band arithmetic are given in Table 2.

2.6. Extraction of remotely sensed parameters at field point locations

The centre coordinates of each ground plot were acquired through handheld Global Positioning System (GPS) receiver. Carreiras et al. (Carreiras et al., 2013) in their study had observed that an error of 8–10 m is quite common in GPS surveys with no differential correction.

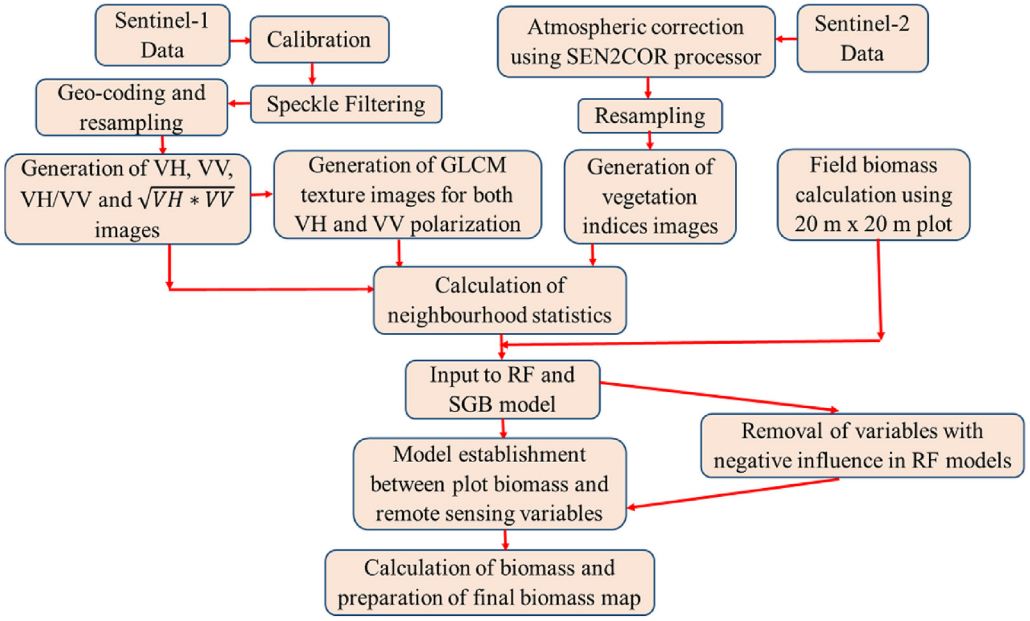

Fig. 2. Methodology flow diagram of data processing and data synergy.

Table 2
Vegetation indices used in analysis.

Index	Formulation (R = reflectance)	Reference
NDVI	$(R_{NIR} - R_{Red})/(R_{NIR} + R_{Red})$	(Rouse, Hass, Schell, & Deering, 1973)
GNDVI	$(R_{NIR} - R_{Green})/(R_{NIR} + R_{Green})$	(Gitelson & Merzlyak, 1996)
RDVI	$(R_{NIR} - R_{Red})/\sqrt{(R_{NIR} + R_{Red})}$	(Roujean & Breon, 1995)
SAVI	$(1 + L)(R_{NIR} - R_{Red})/(R_{NIR} + R_{Red} + L)$; $L = 0.5$	(Huete, 1988)

Table 3
Statistics of the measured above ground biomass (Mg/ha).

Species	Total Samples	Minimum	Maximum	Mean	Standard deviation
S. robusta	36	38.52	625.3	238.7	191.52
T. grandis	34	146.77	660.5	324.2	125.57

The main sources for this error are multipath propagation and tree canopies blocking satellite signals (Zheng, Wang, & Nihan, 2005). To minimize the effect of these errors, we followed the same steps as proposed by Carreiras et al. A 30 m radius buffer was created around each plot centre, as the first step. The buffer radius was restricted to 30 m as a larger radius will encroach on areas which are not homogeneous with the tree canopy cover of the plot area. Pixel values for backscatter, texture and vegetation indices images were then extracted for the buffer zones around each plot centre. Minimum, maximum, mean and standard deviation of the pixel values inside each buffer were computed to use it for establishing the relationship with AGB.

Primarily the predictor variables were four SAR images, three texture images for each polarization i.e. total six and four vegetation indices. A total of 14 variables. Now, to minimize the effect of location inaccuracy we have calculated four neighbourhood statistics for each of the variable. So we have 1-pixel value for each field plot centre and four neighbourhood statistics value for each plot, i.e. total five values for each variable. The total number of predictor variable available for the modelling was 70. SAR, texture and vegetation indices matrices were used individually as predictor variables for RF and SGB models. After that, they are used in a combination of 2 set of matrices as predictor variables. Finally, all the matrices are used together as the predictor variable. In RF models the variables, which are having a negative influence on the model, were removed and a new model was made using rest of the variables.

2.7. Data analysis

Statistical regression methods in combination with parametric (linear, logarithmic or polynomial) functions are often adopted for deriving a relationship between remote sensor measurements and field measured sample (Ali et al., 2015). The identified relationship is extended to estimate the field variable for the whole area covered by remotely sensed image (Heiskanen, 2006). However, these techniques are not helpful to establish the complex relationship between biophysical parameters and remote sensing matrices. Machine learning techniques are highly powerful non-linear regression methods which can be used as an alternative to traditional ways to handle complex and non-linear problems (Breiman, 2001). Machine learning algorithms do not depend on data distribution. As a result, it can integrate data from different sources seamlessly (Ali et al., 2015).

In last few years decision tree based machine learning algorithms got considerable attention in biophysical parameter retrieval problems. Baccini et al. mapped aboveground biomass of tropical Africa using random forest algorithm and satellite imageries (Baccini, Laporte, Goetz, Sun, & Dong, 2008). Güneralp et al. derived floodplain aboveground biomass using remote sensing data and concluded that decision tree based stochastic gradient boosting model prediction is better than several other models (Güneralp, Filippi, & Randall, 2014). Freeman et al. showed the usefulness of both random forest and stochastic gradient boosting in estimating tree canopy cover (Freeman, Moisen, Coulston, & Wilson, 2016). Suitability of these techniques in the context of Indian tropical forests has not been completely explored yet.

2.7.1. Random forest model

The Random Forest (RF) is a non-parametric machine learning algorithm. In this algorithm, a part of the training data is chosen randomly to build a decision tree. Remaining part of the training data, which haven't been used for building the tree, is used for estimating out-of-bag (OOB) error for each tree. At each node of the tree, a set of predictor variables are chosen randomly to determine the split. Hundreds of trees are built similarly and finally, new data is predicted

by aggregating the predictions of all the trees (Liaw & Wiener, 2002). RF is insensitive to noise present in the data, and it also does not overfit (Breiman, 2001). RF is also capable of ranking important variables and to produce an independent measure of prediction error (Adam, Mutanga, Abdel-Rahman, & Ismail, 2014). In the remote-sensing domain, RF has predominately been applied as a classification algorithm. However, researchers are now using the algorithm for regression-based applications. In this study R package, CARET was used to implement RF regression. Important variables were first identified using variable importance feature of random forest. A variable is considered to be important in random forest model if omitting it from the predictor variable list increases the OOB error (Liaw & Wiener, 2002). The final model was built using only important variables for 500 trees. The number of predictor variable randomly selected at each split were varied from 2 to the total number of variables with an increment of 2.

Generally, for a large dataset performance of any predictive model is measured on a validation set, which is set aside from the main dataset before training the model. The number of observations for this study was not large enough to evaluate model performance with an independent subset, so a 5-fold cross-validation approach was followed, as suggested by Hastie et al. (Hastie, Tibshirani, & Friedman, 2009). To select the best model out of the various candidates, we selected the model that had the lowest mean squared error.

2.7.2. Stochastic gradient boosting modelling

Stochastic gradient boosting is another machine learning algorithm whose usage is increasing rapidly for classification and regression applications. SGB combines both regression tree and boosted algorithms to predict the response variable. SGB typically uses a base learner and constructs additive regression models by sequentially fitting the chosen base learner to current "pseudo"-residuals by least squares at each iteration using a random fraction of the training data without replacement (Friedman, 2002). The process is continued until total residual deviance stop decreasing (Dube, Mutanga, Elhadi, & Ismail, 2014).

SGB can handle interactions among predictor variables and fit complex nonlinear relationship between them. It also remains unaffected by outliers in the dataset. The algorithm also reduces the chance of overfitting by introducing an element of stochasticity. Multiple researchers (De'ath, 2007; Dube et al., 2014; Elith, Leathwick, & Hastie, 2008; Güneralp et al., 2014) have recommended SGB for modelling with ecological data due to its flexibility and high predictive performance.

Model fitting in SGB can be fine-tuned by using user-defined values for several parameters like the learning rate, tree complexity, the number of regression trees in the ensemble. In the final model contribution of each tree is determined by learning rate. Tree complexity can be defined as the number of variables selected to determine each split. For this study, SGB models were fitted, with varying values for the number of regression trees (50–5000), tree complexity of 1, 3 and 5. The value for learning rate was fixed at 0.01. The caret package for the R statistical software was utilized to implement SGB. Here also the best model is selected by the least mean square error criteria for 5-fold cross-validation.

3. Results

Figs. 3 and 4 show the typical results of an SGB and an RF model. In those models, vegetation indices matrices were used as predictor variables.

Fig. 3 shows in SGB model, how the RMSE changes depending on shrinkage ratio, interaction depth and the number of boosting iteration, i.e., the number of regression trees. In this case, we found that the lowest RMSE can be found for interaction depth of 5, shrinkage ratio of 0.05 and 3850 number of iteration. In RF model only the number of the predictor variable is varied (Fig. 4). The final RF model shows that it gives lowest RMSE when the number of the predictor variable is 2. The detailed results of all those approaches are shown in Tables 4 and 5.

Due to the scarcity of data, we cross-validated our result instead of creating a separate train and test dataset. The k-fold cross-validation method involves splitting the dataset into k-subsets. In each fold, one subset is held out for checking the model performance, while the model is trained on all other subsets. This process is completed after RMSE and r-squared, i.e., the coefficient of determination is determined for each fold, and an overall RMSE and r-squared is provided by taking the mean of all folds. We did five-fold cross-validation for our study, results of those are given in Tables 6 and 7. Apart from providing these model derived statistics we calculated another set of RMSE and r-squared values. We compiled the prediction of all folds to a single predicted versus observed table and calculated the RMSE and r-squared values. Table 8 shows results of that method.

Results show that *S. robusta* forest biomass can be better predicted by vegetation indices matrices than SAR and texture variables when they are used individually (Tables 6–8). Irrespective of the method of derivation, i.e., either model derived or calculated after combining the prediction of all folds, SAR matrices has the highest RMSE with a value of more than 172 Mg/ha. RMSE values vary from 114 Mg/ha to 133 Mg/ha for vegetation indices matrices. In overall, the best result for *S. robusta* can be found for SAR and vegetation indices combination. RMSE values for *S. robusta* ranges from 99.3 Mg/ha, derived from RF model, to 118.92 Mg/ha, calculated using compiled results of all folds from SGB model. *T. grandis* results show a similar pattern. Here also we found the best result for SAR and vegetation indices with values varying from 72.25 Mg/ha to 80.46 Mg/ha. Other matrices, whether used individually or in combination, give less accurate results for all approaches.

The predicted and observed biomass values for all folds of cross-validation are compiled and plotted (Figs. 5 and 6) to found the correlation between them. SAR backscatter shows worst correlation with a coefficient of determination value of 0.20 for individual predictor matrices for RF model. The best correlation can be found for vegetation indices with a coefficient of determination value of 0.59. In overall, SAR and vegetation indices combination shows the best correlation with a value of 0.71. Predicted and observed biomass relationship for *S. robusta* forest shows similar trends for SGB models. Here best correlation value is 0.60 for texture and vegetation indices combination. Among RF models established for *T. grandis* forest, best correlation can be found for SAR and vegetation indices combination with an r-squared value of 0.60 (Fig. 4). Individually vegetation indices show a good correlation with r-squared values of 0.56 for both SGB and RF models. Texture and vegetation indices combination also shows a good correlation between predicted and observed values. So for both *S. robusta* forest and *T. grandis* plantation RF model predictions show a better correlation between observed and predicted values than SGB model predictions.

Fig. 7 shows the final biomass map of the whole study area for the *S. robusta* forest and *T. grandis* plantation. The maps are generated using raster package of R software. The RF/SGB model was established for each species using field plots and associated remote sensing parameters. The pixel-wise biomass values for each species was calculated using the established RF/SGB model with satellite dataset as input. This file was exported as a raster file using raster package of R software. In *S. robusta* forest estimated biomass values varies from 195.7 Mg/ha to 484.26 Mg/ha with a mean value of 228.58 Mg/ha. The range of biomass values for *T. grandis* is from 204.92 Mg/ha to 403.65 Mg/ha with a mean value of 287.22 Mg/ha. *S. robusta* forest in Katerniaghat area is a natural forest, so normally its density will vary widely throughout the forest. That is why the final map is showing a wide range of biomass values. *T. grandis* vegetation in Katerniaghat area is a plantation, not a natural forest so tree density, as well as biomass density, will not vary widely over the forested region. *T. grandis* biomass range in the final map also proves the point.

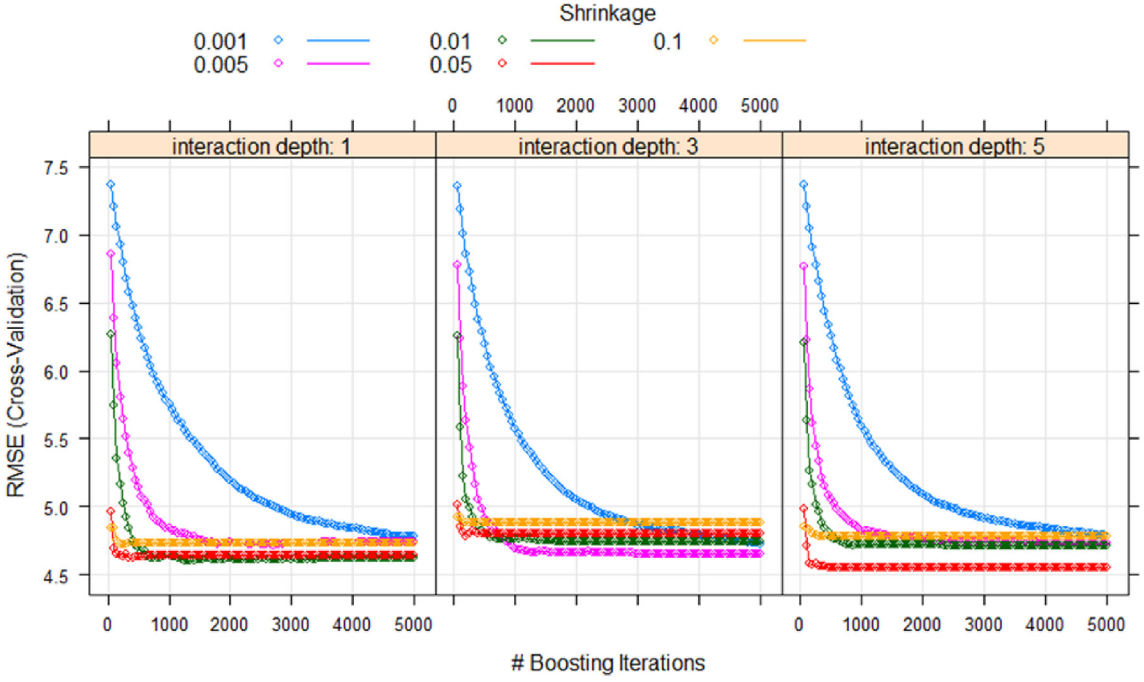

Fig. 3. The result of SGB model for *S. robusta* using vegetation indices.

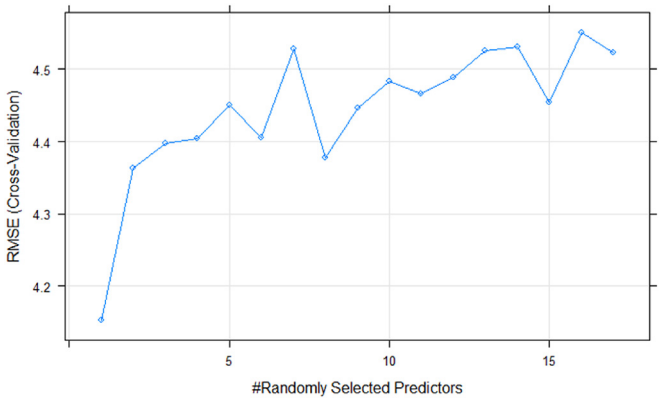

Fig. 4. The result of RF model for *S. robusta* using vegetation indices.

Table 4
Results of SGB method.

Forest Type	Variables	Number of Trees	Shrinkage Ratio	Interaction Depth
S. robusta	SAR	150	0.1	1
	Texture	100	0.05	3
	Vegetation Indices	3850	0.1	5
	SAR and Texture	2050	0.005	3
	SAR and Vegetation Indices	200	0.05	1
	Texture and Vegetation Indices	50	0.1	3
	All	200	0.1	3
T. grandis	SAR	50	0.01	3
	Texture	100	0.05	1
	Vegetation Indices	1100	0.005	5
	SAR and Texture	200	0.005	3
	SAR and Vegetation Indices	1150	0.01	1
	Texture and Vegetation Indices	100	0.05	5
	All	150	0.05	5

Table 5
Results of RF method.

Forest Type	Variables	Number of variables selected at each split
S. robusta	SAR	1
	Texture	2
	Vegetation Indices	6
	SAR and Texture	2
	SAR and Vegetation Indices	5
	Texture and Vegetation Indices	4
	All	1
T. grandis	SAR	1
	Texture	5
	Vegetation Indices	11
	SAR and Texture	2
	SAR and Vegetation Indices	7
	Texture and Vegetation Indices	9
	All	13

4. Discussion

4.1. Performance of the study compared to previous works

SAR backscatter, SAR image texture, and vegetation indices values depend on different properties of the tree canopy. Vegetation indices show the greenness or healthiness of the plant (Dong et al., 2003). SAR backscatter from vegetation depends on water content, size, density and orientation of leaves (Castel, Guerra, Caraglio, & Houllier, 2002). SAR image texture measures the canopy unevenness which is proved to be positively correlated to tropical forest biomass beyond the saturation point of the backscatter–biomass relationship (Kuplich, Curran, & Atkinson, 2005). Different combinations of these predictor variables should improve the predictive performance of biomass estimation models rather than when we use them alone. The efficiency of the current work was checked by comparing the result with two previous studies done in the same region. Behera et al. did a study on Katerniaghat Wildlife Sanctuary using L-band SAR data and field data collected in 2010. They found, *T. grandis* is better correlated to L-band

Table 6
Model derived cross-validation statistics of SGB model.

Forest Type	Variables	Cross validation RMSE (Mg/ha)	Coefficient of determination
S. robusta	SAR	173.93	0.27
	Texture	150.15	0.47
	Vegetation Indices	129.13	0.67
	SAR and Texture	161.56	0.31
	SAR and Vegetation Indices	111.39	0.75
	Texture and Vegetation Indices	121.79	0.73
	All	118.83	0.74
T. grandis	SAR	124.96	0.18
	Texture	120.89	0.31
	Vegetation Indices	83.37	0.62
	SAR and Texture	121.55	0.21
	SAR and Vegetation Indices	78.91	0.71
	Texture and Vegetation Indices	88.18	0.57
	All	83.22	0.68

Table 7
Model derived cross-validation statistics of RF model.

Forest Type	Variables	Cross validation RMSE (Mg/ha)	Coefficient of determination
S. robusta	SAR	172.49	0.25
	Texture	165.88	0.39
	Vegetation Indices	114.72	0.64
	SAR and Texture	161.23	0.42
	SAR and Vegetation Indices	99.33	0.75
	Texture and Vegetation Indices	112.05	0.66
	All	102.07	0.77
T. grandis	SAR	119.85	0.14
	Texture	115.31	0.34
	Vegetation Indices	75.14	0.73
	SAR and Texture	113.82	0.24
	SAR and Vegetation Indices	72.25	0.76
	Texture and Vegetation Indices	73.49	0.74
	All	72.66	0.74

Table 8
Statistics calculated using compiled results of all folds for SGB and RF model.

Forest Type	Variables	SGB RMSE (Mg/ha)	RF RMSE (Mg/ha)
S. robusta	SAR	174.34	172.1
	Texture	160.36	166.44
	Vegetation Indices	133.64	123.48
	SAR and Texture	163.83	154.08
	SAR and Vegetation Indices	118.92	105.03
	Texture and Vegetation Indices	122.94	119.41
	All	119.82	107.2
T. grandis	SAR	125.6	124.54
	Texture	121.83	117.4
	Vegetation Indices	83.82	84.58
	SAR and Texture	122	118.03
	SAR and Vegetation Indices	79.45	80.46
	Texture and Vegetation Indices	88.56	82.69
	All	83.21	81.99

backscatter with a coefficient of determination value of 0.71 than *S. robusta*, which has a coefficient of determination value of 0.47. *S. robusta* shows a lower RMSE value of 104 Mg/ha while *T. grandis* has an RMSE of 150 Mg/ha. Use of L-band microwave radiation for biomass estimation of dense forest is advantageous as it interacts directly with trunks which contain most of the biomass. Results of the current study show that Sentinel-1 and 2 data used in synergy has a similar capability in predicting dense biomass. In the second study, Pandey et al. (Pandey, Kushwaha, Kachhwaha, Kunwar, & Dadhwal, 2010) examined the potential of Envisat Advanced Synthetic Aperture Radar (ASAR) C-band SAR data for woody biomass assessment for a stretch of Dudhwa National Park, which is adjacent to Katerniaghat Wildlife Sanctuary. They found that for ASAR, C-band *S. robusta* and *T. grandis* forest biomass is very poorly correlated with HH and HV backscatter. For *S. robusta* the maximum coefficient of determination between observed and predicted values is 0.02 at HV polarization, and for *T. grandis* it is 0.23 at HV polarization. In contrast, using data synergy and machine learning algorithms, we are getting a better correlation between biomass and remote sensing variable.

4.2. Proficiency of Sentinel data

Sentinel-1 data does not work as a good predictor variable for the study area, as C-band SAR backscatter saturates at a biomass density of 100 Mg/ha (Ghasemi, Sahebi, & Mohammadzadeh, 2011) due to its relatively low penetrating power. Texture parameters can go beyond the normal SAR saturation level as it measures the local variation in SAR backscatter, not the backscatter itself. Image texture of tropical forested areas hugely depends on image resolution (M. L. R. Sarker, Nichol, Iz, Ahmad, & BinRahman, 2013). With the high spatial resolution, neighbouring pixels covers different forest classes' even sometimes different land cover altogether. This spatial variability increases the local variance information. Earlier studies (Cutler, Boyd, Foody, & Vetrivel, 2012; Eckert, 2012; Nichol & Sarker, 2011; Sarker & Nichol, 2011) have found that high-resolution image-textural features of both SAR and optical remote sensing data are highly correlated to forest biomass. Vegetation indices obtained from optical data generally saturates at lower biomass value (Reddersen, Fricke, & Wachendorf, 2014; Tilly, Aasen, & Bareth, 2015). However, Sentinel-2 data is different from other available optical dataset. It has much better spatial, spectral, and temporal resolution. Sentinel-2 is the only available optical data with three red-edge bands (Han et al., 2017), which is very effective for monitoring vegetation health information. Vegetation indices derived from Sentinel-2 have also shows a good correlation with tropical forest biomass with a maximum r^2 value of 0.66 (Adan, 2017).

In our result also we can see that SAR image texture performs as the better predictor than SAR backscatter. Still, the relationship between biomass and texture is not very good. The reason behind that can be the homogeneity of the chosen vegetation types. Texture parameters are most applicable under conditions of high local variance (Woodcock & Strahler, 1987). In a mixed forest, variation would be more so texture parameters would be more informative. In our study, both *S. robusta* and *T. grandis* are homogeneous patches. So, very little information can be extracted using texture parameters. Results of this study show that Sentinel-2 vegetation indices act as a better predictor variable for biomass than SAR backscatter and texture parameters. In *S. robusta* forest, a combination of SAR and vegetation indices make a better prediction than the individual variables; same goes for *T. grandis* also. Texture and vegetation indices combination yields the best correlation between observed and predicted biomass for both *S. robusta* and *T. grandis*.

4.3. Effectiveness of machine learning algorithms

Most of the earlier biomass estimation studies have done regression assuming a linear relationship between predictor and predicted variables. This assumption suppresses the inherent non-linear relationship between them. The machine learning methods can establish a complex non-linear relationship between the predictor and predicted variables. Machine learning also helps in combining data from different sources to

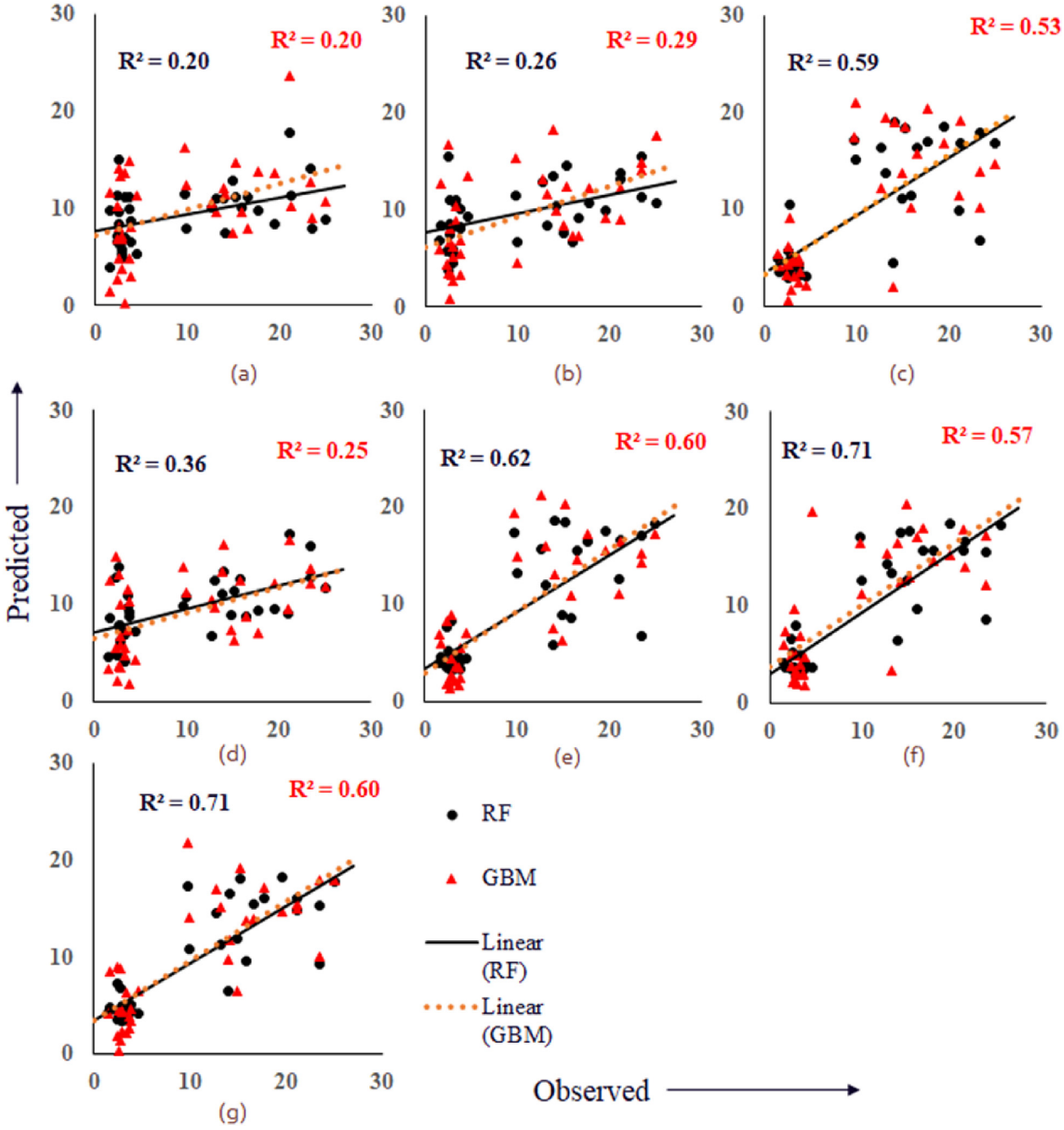

Fig. 5. Coefficient of determination (R2) values of SGB and RF models for *S. robusta* forest using (a) SAR matrices (b) Texture matrices (c) Vegetation indices (VI) matrices (d) SAR and texture combination (e) SAR and VI combination (f) Texture and VI combination (g) All matrices.

improve the result (Ali et al., 2015). In our study, we found that using machine learning and data synergy it's possible to overcome the saturation problem. We also observed that RF algorithm works better than SGB algorithm for most of the cases. Freeman et al. (Freeman et al., 2016) in their study gave the possible explanation of this kind of result. Each tree in RF is independent. So, importance tends to be divided between the predictor variables if they are highly correlated. One variable important to some of the trees and the other variables important in other trees. In SGB, each successive tree builds on the previous tree. If variables are correlated, the first variable that is randomly selected became most important. Even if other correlated variables are chosen in later trees, there is less information that they can contribute. This contributes to the better performance of RF model than SGB.

The classification and regression algorithms work better with a more extensive dataset (Beleites, Neugebauer, Bocklitz, Krafft, & Popp, 2013; Brain & Webb, 1999). In these algorithms, an increase in some predictor variable escalate the chance of overfitting (Hawkins, 2004; Subramanian & Simon, 2013). In our study, for *S. robusta* and *T. grandis* we had a sample size of 36 and 34. We also have 70 predictor variables for regression. With a larger dataset, we can divide it into separate parts for training, validation, and testing. If we divide a small dataset into multiple parts, they may not indeed represent the properties of the whole dataset, and in that case, k-fold cross validation is the most suitable procedure to remove overfitting (Lever, Krzywinski, & Altman, 2016). Although we start the regression process with 70 predictor variables, the number of important variables reduces to less than 20 after initial step which is used to build final model. Biau (2010) showed that RF model depends on active variables, not data dimensionality. He

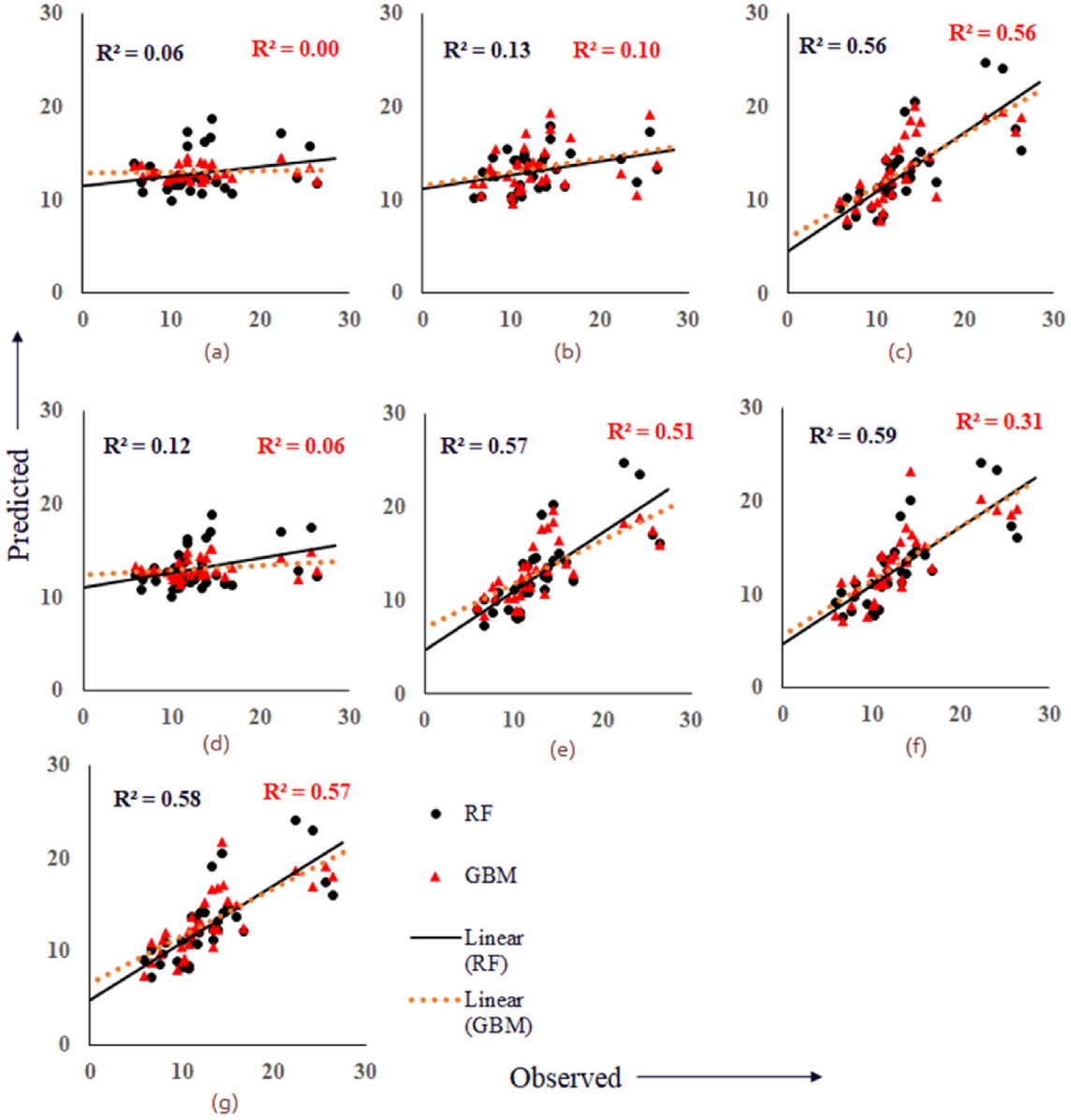

Fig. 6. Coefficient of determination (R2) values of SGB and RF models for *T. grandis* plantation using (a) SAR matrices (b) Texture matrices (c) Vegetation indices (VI) matrices (d) SAR and texture combination (e) SAR and VI combination (f) Texture and VI combination (g) All matrices.

also concluded that this is the reason why RF model can handle a large number of input variable even when it is larger than the sample size (Biau, 2010). Scornet et al. showed that RF could adapt easily when the number of influencing variable is less even though dimensionality of the data is large (Scornet, Biau, & Vert, 2015).

4.4. Possible future contribution of Sentinel data

At the beginning of next decade, NASA-ISRO Synthetic Aperture Radar (NISAR) will become world's first dual frequency (L and S-band) spaceborne SAR system. NISAR will be suitable for forest biomass measurement due to its high operating frequency. NISAR's global, detailed map of aboveground biomass density will lower the ambiguity of estimated carbon emissions from land use change. Sentinel-1 and Sentinel-2 both have long enough lifespan (Drusch et al., 2012; Torres et al., 2012) so that it will overlap with NISAR's lifespan. ESAs' P band BIOMASS mission will also come into operation in next decade as a major addition to current global efforts to build a carbon data monitoring system (Scipal et al., 2010). Biomass map generated using Sentinel-1, and 2 data can be used to validate the aboveground biomass maps generated from these future missions.

All the abilities of the state-of-the-art Sentinel-1 and Sentinel-2 data have not been fully explored yet. Especially their performance in estimating biomass of tropical forests. In this paper, we examined that prospect to a proper extent. Results show that vegetation indices from Sentinel-2 performs better than Sentinel-1 backscatter images and derived texture images. This outcome has some valuable implications. Most of the tropical forests are situated in lower and middle-income countries. Consequently, regular monitoring of these forests with costly high spatial resolution data has not been done up until now. With the free availability of Sentinel data, that observation can be possible now. Recent biomass studies mostly depend on expensive longer wavelength SAR data while ignoring the optical dataset. Sentinel-2 data has the potential to alter that trend.

5. Conclusions

This is not the first study which has used Sentinel data or RF

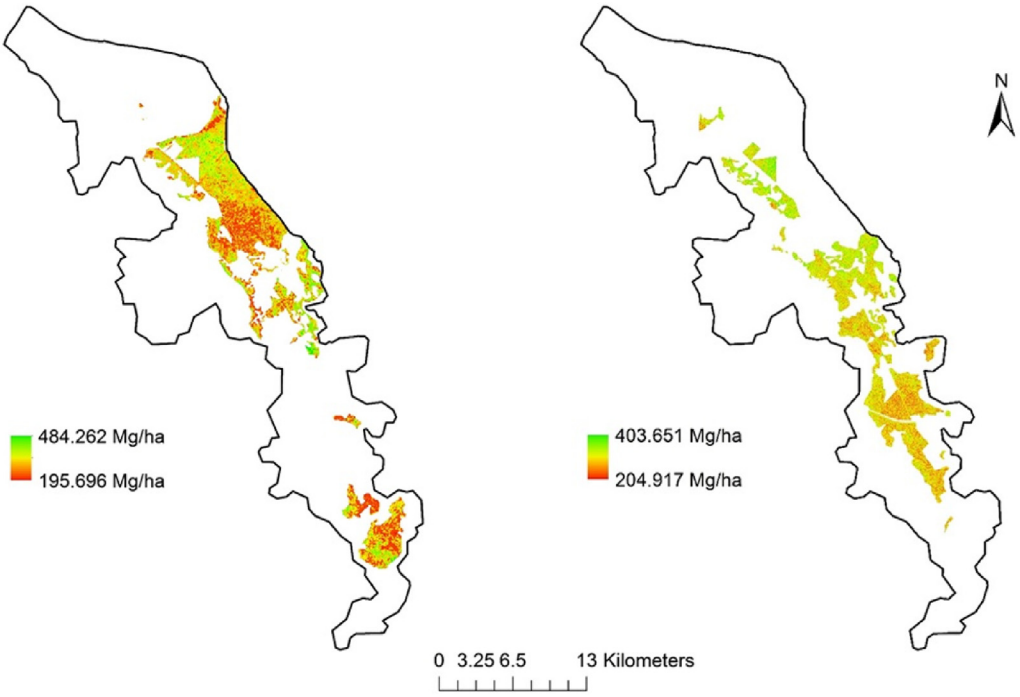

Fig. 7. Biomass (Mg/ha) maps of (a) *S. robusta* (b) *T. grandis* for the best combination of variables.

algorithm to estimate forest biomass. However, it is the first one where the effectiveness of Sentinel-1 and 2 data synergy for Indian tropical forests has been evaluated. As it was mentioned earlier, Indian tropical forests are under threat from climate change. To know their position in the global carbon cycle, whether they are acting as a source or as a sink, we have to monitor them continuously. Country-level biomass mapping at a regular interval using commercial satellite data is costly. Due to financial constraints, most of the Indian researchers try to use freely available data for this purpose. Up until now only Landsat (NASA) or LISS-3 (ISRO) data can be accessed at free of cost. Sentinel data has much better spatial, spectral, and temporal resolution than both of them. With the availability of free Sentinel data, there is scope for lots of improvement in biomass estimation studies, especially for India. In this paper, the methodology of our work is explained in detail. Image processing software and input parameters are mentioned at each step. The dataset and software used in this work are available at free of cost. So, this methodology can be replicated easily by any researcher irrespective of location and financial conditions. The results of our study show a combination of Sentinel-1 and Sentinel-2 can estimate with biomass with high accuracy. Hopefully, this results will encourage other researchers to use these data to build a country level biomass monitoring system.

References

Adam, E., Mutanga, O., Abdel-Rahman, E. M., & Ismail, R. (2014). Estimating standing biomass in papyrus (Cyperus papyrus L.) swamp: Exploratory of in situ hyperspectral indices and random forest regression. *International Journal of Remote Sensing, 35*(2), 693–714. http://doi.org/10.1080/01431161.2013.870676.

Adan, M. S. (2017). *Intergrating sentinel-2 derived vegetation indices and terrestrial laser scanner to estimate above- ground biomass/carbon in ayer hitam tropical forest Malaysia.* Msc Thesis. Retrieved from http://www.itc.nl/library/papers_2017/msc/nrm/adan.pdf.

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 Sensing, 7*(12), 16398–16421. http://doi.org/10.3390/rs71215841.

Antropov, O., Rauste, Y., Häme, T., & Praks, J. (2017). Polarimetric ALOS PALSAR time series in mapping biomass of boreal forests. *Remote Sensing, 9*(10), 999. http://doi.org/10.3390/rs9100999.

Babcock, C., Finley, A. O., Bradford, J. B., Kolka, R., Birdsey, R., & Ryan, M. G. (2015). LiDAR based prediction of forest biomass using hierarchical models with spatially varying coefficients. *Remote Sensing of Environment, 169*, 113–127. http://doi.org/10.1016/j.rse.2015.07.028.

Baccini, A., Laporte, N., Goetz, S. J., Sun, M., & Dong, H. (2008). A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environmental Research Letters, 3*(4), 045011. http://doi.org/10.1088/1748-9326/3/4/045011.

Behera, M., Tripathi, P., Mishra, B., Kumar, S., Chitale, V., & Behera, S. K. (2015). Above-ground biomass and carbon estimates of Shorea robusta and Tectona grandis forests using QuadPOL ALOS PALSAR data. *Advances in Space Research.* http://doi.org/10.1016/j.asr.2015.11.010.

Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., & Popp, J. (2013). Sample size planning for classification models. *Analytica Chimica Acta, 760*(June 2012), 25–33. http://doi.org/10.1016/j.aca.2012.11.007.

Biau, G. (2010). Analysis of a random forests model. *Journal of Machine Learning Research, 13*, 1063–1095. Retrieved from http://arxiv.org/abs/1005.0208.

Brain, D., & Webb, G. I. (1999). On the effect of data set size on bias and variance in classification learning. *Proceedings of the fourth Australian knowledge acquisition workshop (AKAW '99)* (pp. 117–128). .

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. http://doi.org/10.1023/A:1010933404324.

Campbell, B. M. (2009). Beyond Copenhagen: REDD + , agriculture, adaptation strategies and poverty. *Global Environmental Change, 19*(4), 397–399. http://doi.org/10.1016/j.gloenvcha.2009.07.010.

Carreiras, J. M. B., Melo, J. B., & Vasconcelos, M. J. (2013). Estimating the above-ground biomass in miombo savanna woodlands (Mozambique, East Africa) using L-band synthetic aperture radar data. *Remote Sensing, 5*(4), 1524–1548. http://doi.org/10.3390/rs5041524.

Carreiras, J. M. B., Vasconcelos, M. J., & Lucas, R. M. (2012). Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sensing of Environment, 121*, 426–442. http://doi.org/10.1016/j.rse.2012.02.012.

Cartus, O., Santoro, M., & Kellndorfer, J. (2012). Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band. *Remote Sensing of Environment, 124*, 466–478. http://doi.org/10.1016/j.rse.2012.05.029.

Castel, T., Guerra, F., Caraglio, Y., & Houllier, F. (2002). Retrieval biomass of a large Venezuelan pine plantation using JERS-1 SAR data. Analysis of forest structure impact on radar signature. *Remote Sensing of Environment, 79*(1), 30–41. http://doi.org/10.1016/S0034-4257(01)00236-X.

Castillo, J. A. A., Apan, A. A., Maraseni, T. N., & Salmo, S. G. (2017). Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 134*, 70–85. http://doi.org/10.1016/j.isprsjprs.2017.10.016.

Chang, J., & Shoshany, M. (2016). Mediterranean shrublands biomass estimation using Sentinel-1 and Sentinel-2. *International geoscience and remote sensing symposium (IGARSS), 2016–Novem* (pp. 5300–5303). . http://doi.org/10.1109/IGARSS.2016.7730380.

Chazdon, R. L., Broadbent, E. N., Rozendaal, D. M. A., Bongers, F., Zambrano, A. M. A., Aide, T. M., et al. (2016). Carbon sequestration potential of second-growth forest regeneration in the Latin American tropics. *Science Advances, 2*(May), e1501639 http://doi.org/10.1126/sciadv.1501639.

Chen, Q. (2015). Modeling aboveground tree woody biomass using national-scale

allometric methods and airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing, 106*, 95–106. http://doi.org/10.1016/j.isprsjprs.2015.05.007.
Chitale, V. S., & Behera, M. D. (2014). Analysing land and vegetation cover dynamics during last three decades in Katerniaghat wildlife sanctuary, India. *Journal of Earth System Science, 123*(7), 1467–1479.
Cutler, M. E. J., Boyd, D. S., Foody, G. M., & Vetrivel, A. (2012). Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS Journal of Photogrammetry and Remote Sensing, 70*, 66–77. http://doi.org/10.1016/j.isprsjprs.2012.03.011.
Das, S., & Singh, T. P. (2012). Correlation analysis between biomass and spectral vegetation indices of forest ecosystem. *International Journal of Engineering Research & Technology, 1*(5), 1–13. Retrieved from http://www.ijert.org/view-pdf/532/correlation-analysis-between-biomass-and-spectral-vegetation-indices-of-forest-ecosystem.
Deng, S., Katoh, M., Guan, Q., Yin, N., & Li, M. (2014). Estimating forest aboveground biomass by combining ALOS PALSAR and WorldView-2 data: A case study at purple mountain national Park, Nanjing, China. *Remote Sensing, 6*(9), 7878–7910. http://doi.org/10.3390/rs6097878.
De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology, 88*(1), 243–251 http://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2.
Dong, J., Kaufmann, R. K., Myneni, R. B., Tucker, C. J., Kauppi, P. E., Liski, J., et al. (2003). Remote sensing estimates of boreal and temperate forest woody biomass: Carbon pools, sources, and sinks. *Remote Sensing of Environment, 84*(3), 393–410. http://doi.org/10.1016/S0034-4257(02)00130-X.
Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment, 120*, 25–36. http://doi.org/10.1016/j.rse.2011.11.026.
Dube, T., Mutanga, O., Elhadi, A., & Ismail, R. (2014). Intra-and-inter species biomass prediction in a plantation forest: Testing the utility of high spatial resolution spaceborne multispectral RapidEye sensor and advanced machine learning algorithms. *Sensors (Basel, Switzerland), 14*(8), 15348–15370. http://doi.org/10.3390/s140815348.
Eckert, S. (2012). Improved forest biomass and carbon estimations using texture measures from worldView-2 satellite data. *Remote Sensing, 4*(4), 810–829. http://doi.org/10.3390/rs4040810.
Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802–813. http://doi.org/10.1111/j.1365-2656.2008.01390.x.
Erten, E., Lopez-Sanchez, J. M., Yuzugullu, O., & Hajnsek, I. (2016). Retrieval of agricultural crop height from space: A comparison of SAR techniques. *Remote Sensing of Environment, 187*, 130–144. http://doi.org/10.1016/j.rse.2016.10.007.
Forest Survey of India (FSI) (1996). *Volume equations for forests of India, Nepal and Bhutan.* Dehradun: Ministry of Environment and Forests (MOEF), Govt. of India.
Freeman, E. a., Moisen, G. G., Coulston, J. W., & Wilson, B. T. (2016). Random forests and stochastic gradient boosting for predicting tree canopy cover: Comparing tuning processes and model performance[1]. *Canadian Journal of Forest Research, 46*(3), 323–339. http://doi.org/10.1139/cjfr-2014-0562.
Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis, 38*(4), 367–378. http://doi.org/10.1016/S0167-9473(01)00065-2.
Gasparri, N. I., Parmuchi, M. G., Bono, J., Karszenbaum, H., & Montenegro, C. L. (2010). Assessing multi-temporal Landsat 7 ETM+ images for estimating above-ground biomass in subtropical dry forests of Argentina. *Journal of Arid Environments, 74*(10), 1262–1270. http://doi.org/10.1016/j.jaridenv.2010.04.007.
Ghasemi, N., Sahebi, M. R., & Mohammadzadeh, A. (2011). A review on biomass estimation methods using synthetic aperture radar. *International Journal of Geomatics and Geosciences, 1*(4), 776–788.
Ghosh, S. M., & Behera, M. D. (2017). Forest canopy height estimation using satellite laser altimetry: A case study in the Western Ghats, India. *Applied Geomatics, 9*(3), 159–166. http://doi.org/10.1007/s12518-017-0190-2.
Gitelson, A. a., & Merzlyak, M. N. (1996). Signature analysis of leaf reflectance Spectra: Algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology, 148*(3–4), 494–500. http://doi.org/10.1016/S0176-1617(96)80284-7.
Goodale, C. L., Apps, M. J., Birdsey, R. A., Field, C. B., Heath, L. S., Houghton, R. A., et al. (2002). Forest carbon sinks in the northern hemisphere. *Ecological Applications, 12*(3), 891–899 http://doi.org/10.1890/1051-0761(2002)012[0891:FCSITN]2.0.CO;2.
Güneralp, I., Filippi, A. M., & Randall, J. (2014). Estimation of floodplain aboveground biomass using multispectral remote sensing and nonparametric modeling. *International Journal of Applied Earth Observation and Geoinformation, 33*(1), 119–126. http://doi.org/10.1016/j.jag.2014.05.004.
Günlü, A., Ercanli, Başkent, E. Z., & Çakır, G. (2014). Estimating aboveground biomass using landsat TM imagery: A case study of anatolian crimean pine forests in Turkey. *Annals of Forest Research, 57*(2), 289–298. http://doi.org/10.15287/afr.2014.278.
Han, J., Wei, C., Chen, Y., Liu, W., Song, P., Zhang, D., et al. (2017). Mapping above-ground biomass of winter oilseed rape using high spatial resolution satellite data at parcel scale under waterlogging conditions. *Remote Sensing, 9*(3)http://doi.org/10.3390/rs9030238.
Haralick, R., Shanmugan, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*. http://doi.org/10.1109/TSMC.1973.4309314.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning the elements of statistical learning data mining, inference, and prediction* (2nd ed.). Springer series in statisticshttp://doi.org/10.1007/978-0-387-84858-7.
Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences, 44*(1), 1–12. http://doi.org/10.1021/ci0342472.
Heiskanen, J. (2006). Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *International Journal of Remote Sensing, 27*(6), 1135–1158. http://dx.doi.org/10.1080/01431160500353858.
Ho Tong Minh, D., Le Toan, T., Rocca, F., Tebaldini, S., Villard, L., Réjou-Méchain, M., et al. (2016). SAR tomography for the retrieval of forest biomass and height: Cross-validation at two tropical forest sites in French Guiana. *Remote Sensing of Environment, 175*, 138–147. http://doi.org/10.1016/j.rse.2015.12.037.
Houghton, R. A., Byers, B., & Nassikas, A. A. (2015). A role for tropical forests in stabilizing atmospheric CO2. *Nature Climate Change, 5*(12), 1022–1023. http://doi.org/10.1038/nclimate2869.
Houghton, R. A., Hall, F., & Goetz, S. J. (2009). Importance of biomass in the global carbon cycle. *Journal of Geophysical Research: Biogeosciences, 114*(3), 1–13. http://doi.org/10.1029/2009JG000935.
Huang, Y., & Genderen van, J. L. (1996). Evaluation of several speckle filtering techniques for ERS-1 & 2 imagery. *International Archives of Photogrammetry and Remote Sensing: vol. XXXI*, (pp. 164–169).
Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*(3), 295–309. http://doi.org/10.1016/0034-4257(88)90106-X.
Hunter, M. O., Keller, M., Victoria, D., & Morton, D. C. (2013). Tree height and tropical forest biomass estimation. *Biogeosciences, 10*, 8385–8399. http://doi.org/10.5194/bg-10-8385-2013.
Joshi, N., Mitchard, E. T. A., Brolly, M., Schumacher, J., Fernández-Landa, A., Johannsen, V. K., et al. (2017). Understanding "saturation" of radar signals over forests. *Scientific Reports, 7*(1), 1–11. http://doi.org/10.1038/s41598-017-03469-3.
Kaasalainen, S., Holopainen, M., Karjalainen, M., Vastaranta, M., Kankare, V., Karila, K., et al. (2015). Combining lidar and synthetic aperture radar data to estimate forest biomass: Status and prospects. *Forests, 6*(1), 252–270. http://doi.org/10.3390/f6010252.
Kachamba, D. J., Ørka, H. O., Næsset, E., Eid, T., & Gobakken, T. (2017). Influence of plot size on efficiency of biomass estimates in inventories of dry tropical forests assisted by photogrammetric data from an unmanned aircraft system. *Remote Sensing, 9*(6), 1–15. http://doi.org/10.3390/rs9060610.
Kuplich, T. M., Curran, P. J., & Atkinson, P. M. (2003). Relating SAR image texture and backscatter to tropical forest biomass. *IEEE Transactions on Geoscience and Remote Sensing (IGARSS 2003), 1*(1), 2872–2874. http://doi.org/10.1109/IGARSS.2003.1294615.
Kuplich, T. M., Curran, P. J., & Atkinson, P. M. (2005). Relating SAR image texture to the biomass of regenerating tropical forests. *International Journal of Remote Sensing, 26*(21), 4829–4854. http://doi.org/10.1080/01431160500239107.
Lever, J., Krzywinski, M., & Altman, N. (2016). Points of Significance: Model selection and overfitting. *Nature Methods, 13*(9), 703–704. http://doi.org/10.1038/nmeth.3968.
Liaw, a, & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(December), 18–22. http://doi.org/10.1177/154405910408300516.
Li, A., Glenn, N. F., Olsoy, P. J., Mitchell, J. J., & Shrestha, R. (2015). Aboveground biomass estimates of sagebrush using terrestrial and airborne LiDAR data in a dryland ecosystem. *Agricultural and Forest Meteorology, 213*, 138–147. http://doi.org/10.1016/j.agrformet.2015.06.005.
Lopes, A., Nezry, E., Touzi, R., & Laur, H. (1990). Maximum a posteriori speckle filtering and first order texture models in sar images. *10th annual international symposium on geoscience and remote sensing* (pp. 2409–2412). . http://doi.org/10.1109/IGARSS.1990.689026.
Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing, 27*(7), 1297–1328. http://doi.org/10.1080/01431160500486732.
Mauya, E. W., Hansen, E. H., Gobakken, T., Bollandsås, O. M., Malimbwi, R. E., & Næsset, E. (2015). Effects of field plot size on prediction accuracy of aboveground biomass in airborne laser scanning-assisted inventories in tropical rain forests of Tanzania. *Carbon Balance and Management, 10*(1), 10. http://doi.org/10.1186/s13021-015-0021-x.
Maynard, C., Lawrence, R., Nielsen, G., & Decker, G. (2007). Modeling vegetation amount using bandwise regression and ecological site descriptions as an alternative to vegetation indices. *GIScience & Remote Sensing, 44*(1), 68–81. http://doi.org/10.2747/1548-1603.44.1.68.
Nichol, J. E., & Sarker, M. L. R. (2011). Improved biomass estimation using the texture parameters of two high-resolution optical sensors. *IEEE Transactions on Geoscience and Remote Sensing, 49*(3), 930–948. http://doi.org/10.1109/TGRS.2010.2068574.
Ningthoujam, R., Balzter, H., Tansey, K., Feldpausch, T., Mitchard, E., Wani, A., et al. (2017). Relationships of S-Band radar backscatter and forest aboveground biomass in different forest types. *Remote Sensing, 9*(11), 1116. http://doi.org/10.3390/rs9111116.
Næsset, E., Bollandsås, O. M., Gobakken, T., Solberg, S., & McRoberts, R. E. (2015). The effects of field plot size on model-assisted estimation of aboveground biomass change using multitemporal interferometric SAR and airborne laser scanning data. *Remote Sensing of Environment, 168*, 252–264. http://doi.org/10.1016/j.rse.2015.07.002.
Omar, H., Misman, M. A., & Kassim, A. R. (2017). Synergetic of PALSAR-2 and Sentinel-1A SAR polarimetry for retrieving aboveground biomass in dipterocarp forest of Malaysia. *Applied Sciences, 7*(7), 675. http://doi.org/10.3390/app7070675.
Ouchi, K. (2013). Recent trend and advance of synthetic aperture radar with selected topics. *Remote Sensing*. http://doi.org/10.3390/rs5020716.
Pandey, U., Kushwaha, S. P. S., Kachhwaha, T. S., Kunwar, P., & Dadhwal, V. K. (2010). Potential of Envisat ASAR data for woody biomass assessment. *Tropical Ecology, 51*(1), 117–124.
Puliti, S., Solberg, S., Næsset, E., Gobakken, T., Zahabu, E., Mauya, E., et al. (2017). Modelling above ground biomass in Tanzanian miombo woodlands using TanDEM-XWorldDEM and field data. *Remote Sensing, 9*(10), 1–13. http://doi.org/10.3390/rs9100984.

Ravindranath, N. H., Joshi, N. V., Sukumar, R., & Saxena, A. (2006). Impact of climate change on forests in India. *Current Science, 90*(3), 354–361.

Reddersen, B., Fricke, T., & Wachendorf, M. (2014). A multi-sensor approach for predicting biomass of extensively managed grassland. *Computers and Electronics in Agriculture, 109*, 247–260. http://doi.org/10.1016/j.compag.2014.10.011.

Roujean, J. L., & Breon, F. M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment, 51*(3), 375–384. http://doi.org/10.1016/0034-4257(94)00114-3.

Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resources Technology Satellite (ERTS) Symposium, 1*, 309–317. http://doi.org/citeulike-article-id:12009708.

Salunkhe, O., Khare, P. K., Sahu, T. R., & Singh, S. (2016). Estimation of tree biomass reserves in tropical deciduous forests of Central India by non- destructive approach. *Tropical Ecology, 57*(2), 153–161.

Sarker, L. R., & Nichol, J. E. (2011). Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sensing of Environment, 115*(4), 968–977. http://doi.org/10.1016/j.rse.2010.11.010.

Sarker, M. L. R., Nichol, J., Ahmad, B., Busu, I., & Rahman, A. A. (2012). Potential of texture measurements of two-date dual polarization PALSAR data for the improvement of forest biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing, 69*, 146–166. http://doi.org/10.1016/j.isprsjprs.2012.03.002.

Sarker, M. L. R., Nichol, J., Iz, H. B., Ahmad, B., Bin, & Rahman, A. A. (2013). Forest biomass estimation using texture measurements of high-resolution dual-polarization C-band SAR data. *IEEE Transactions on Geoscience and Remote Sensing, 51*(6), 3371–3384. http://doi.org/10.1109/TGRS.2012.2219872.

Scipal, K., Arcioni, M., Chave, J., Dall, J., Fois, F., LeToan, T., et al. (2010). The BIOMASS Mission - an ESA Earth Explorer candidate to measure the BIOMASS of the Earth's forests. *International Geoscience and Remote Sensing Symposium (IGARSS),* 52–55. http://doi.org/10.1109/IGARSS.2010.5648979.

Scornet, E., Biau, G., & Vert, J. P. (2015). Consistency of random forests. *Annals of Statistics, 43*(4), 1716–1741. http://doi.org/10.1214/15-AOS1321.

Shen, W., Li, M., Huang, C., & Wei, A. (2016). Quantifying live aboveground biomass and forest disturbance of mountainous natural and plantation forests in northern Guangdong, China, based on multi-temporal landsat, PALSAR and field plot data. *Remote Sensing, 8*(7), 595. http://doi.org/10.3390/rs8070595.

Sinha, S., Jeganathan, C., Sharma, L. K., Nathawat, M. S., Das, A. K., & Mohan, S. (2016). Developing synergy regression models with space-borne ALOS PALSAR and landsat TM sensors for retrieving tropical forest biomass. *Journal of Earth System Science, 125*(4), 725–735. http://doi.org/10.1007/s12040-016-0692-z.

Solberg, S., Hansen, E. H., Gobakken, T., Næssset, E., & Zahabu, E. (2017). Biomass and InSAR height relationship in a dense tropical forest. *Remote Sensing of Environment, 192*, 166–175. http://doi.org/10.1016/j.rse.2017.02.010.

Subramanian, J., & Simon, R. (2013). Overfitting in prediction models - is it a problem only in high dimensions? *Contemporary Clinical Trials, 36*(2), 636–641. http://doi.org/10.1016/j.cct.2013.06.011.

Su, Y., Guo, Q., Xue, B., Hu, T., Alvarez, O., Tao, S., et al. (2016). Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne lidar, optical imagery, and forest inventory data. *Remote Sensing of Environment, 173*, 187–199. http://doi.org/10.1016/j.rse.2015.12.002.

Takagi, K., Yone, Y., Takahashi, H., Sakai, R., Hojyo, H., Kamiura, T., et al. (2015). Forest biomass and volume estimation using airborne LiDAR in a cool-temperate forest of northern Hokkaido, Japan. *Ecological Informatics, 26*, 54–60. http://doi.org/10.1016/j.ecoinf.2015.01.005.

Thumaty, K. C., Fararoda, R., Middinti, S., Gopalakrishnan, R., Jha, C. S., & Dadhwal, V. K. (2016). Estimation of above ground biomass for central Indian deciduous forests using ALOS PALSAR L-band data. *Journal of the Indian Society of Remote Sensing, 44*(1), 31–39. http://doi.org/10.1007/s12524-015-0462-4.

Tilly, N., Aasen, H., & Bareth, G. (2015). Fusion of plant height and vegetation indices for the estimation of barley biomass. *Remote Sensing, 7*(9), 11449–11480. http://doi.org/10.3390/rs70911449.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., et al. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment, 120*, 9–24. http://doi.org/10.1016/j.rse.2011.05.028.

Vaglio, G. L., Pirotti, F., Callegari, M., Chen, Q., Cuozzo, G., Lingua, E., et al. (2017). Potential of ALOS2 and NDVI to estimate forest above-ground biomass, and comparison with lidar-derived estimates. *Remote Sensing, 9*(1)http://doi.org/10.3390/rs9010018.

Wang, Y., Li, G., Ding, J., Guo, Z., Tang, S., Wang, C., et al. (2016). A combined GLAS and MODIS estimation of the global distribution of mean forest canopy height. *Remote Sensing of Environment, 174*, 24–43. http://doi.org/10.1016/j.rse.2015.12.005.

Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment, 21*, 311–332. http://doi.org/10.1016/0034-4257(87)90015-0.

Xu, L., Saatchi, S. S., Shapiro, A., Meyer, V., Ferraz, A., Yang, Y., et al. (2017). Spatial distribution of carbon stored in forests of the democratic republic of Congo. *Scientific Reports, 7*(1), 1–12. http://doi.org/10.1038/s41598-017-15050-z.

Zhao, P., Lu, D., Wang, G., Liu, L., Li, D., Zhu, J., et al. (2016). Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. *International Journal of Applied Earth Observation and Geoinformation, 53*, 1–15. http://doi.org/10.1016/j.jag.2016.08.007.

Zheng, J., Wang, Y., & Nihan, N. L. (2005). Quantitative evaluation of GPS performance under forest canopies. *2005 IEEE networking, sensing and control, ICNSC2005-proceedings, 2005* (pp. 777–782). . http://doi.org/10.1109/ICNSC.2005.1461289.