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

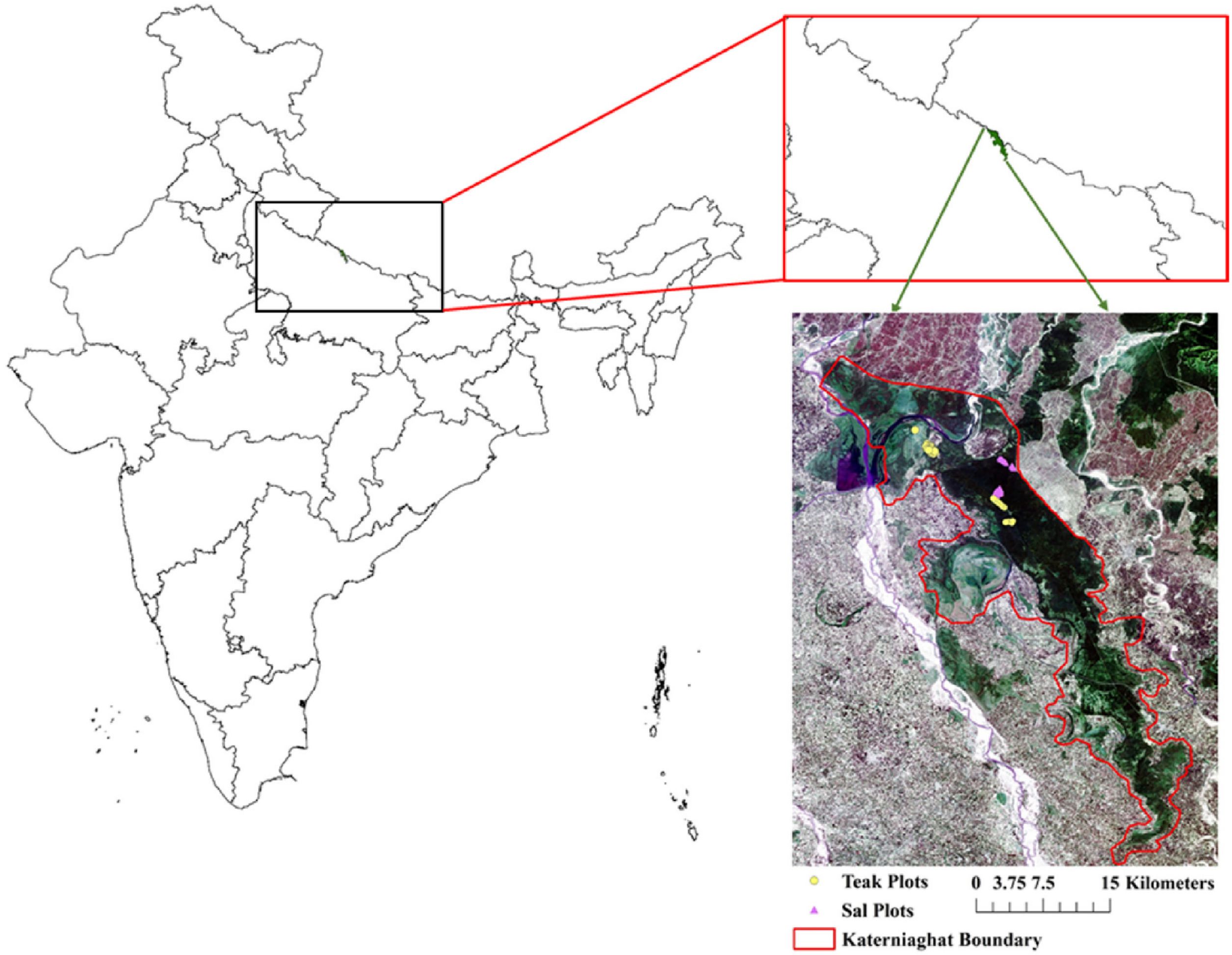

Fig. 1. The geographical location of study area.

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