Journal of the Indian Society of Remote Sensing (April 2024) 52(4):771–786
https://doi.org/10.1007/s12524-024-01821-5

RESEARCH ARTICLE

Forest Aboveground Biomass and Forest Height Estimation Over a Sub-tropical Forest Using Machine Learning Algorithm and Synthetic Aperture Radar Data

Noman Ali[1] · Unmesh Khati[2]

Received: 23 March 2023 / Accepted: 12 January 2024 / Published online: 3 February 2024

Abstract
Forest aboveground biomass (AGB) is a key measurement in studying terrestrial carbon storage, carbon cycle, and climate change. Machine learning based algorithms can be applied to estimate forest AGB using remote sensing-based data. Our study utilized L-band ALOS-2/PALSAR-2 Synthetic Aperture Radar (SAR) data in combination with multi-parameter linear regression (LR) and Random forest regression (RF) for forest carbon estimation. Six L-band fully polarimetric acquisitions are used in this study. The input parameters to the RF algorithm are the backscatter, decomposition powers and species information. The multi-temporal backscatter (HH1 to HH6, HV1 to HV6, VV1 to VV6) and the temporal average are used. Furthermore, average decomposi-tion parameters from G4U decomposition—Double bounce (Dbl), Odd bounce (Odd), Volume scattering (Vol), and Helix scattering (Hlx) for all six dates. In the first case (1), the model is trained to estimate only the AGB. In the second case (2), the model is trained for forest height estimation. In the third case (3), the model is trained to predict both the AGB and height of the forest. In contrast to the LR method, there is a significant improvement in AGB estimation achieved with the RF algorithms. This study shows the potential of combined retrieval of AGB and forest height using time-series L-band backscatter data.

Keywords L-Band ALOS-2/PALSAR-2 SAR data · Aboveground biomass model · Height of forest model · AGB and height of forest model

Introduction

Forests play a pivotal role within terrestrial ecosystems, making substantial contributions to their ecological dynamics and functions. These contributions are critical due to their role in contributing roughly eighty percent of aboveground carbon (Liu et al., 2017). As a result, forest aboveground biomass (AGB) is deemed an essential element in the global carbon cycle (Brown et al., 2002; Houghton et al., 2009). The precise assessment of woodland biomass assumes critical significance in the context of sustainable forest management, carbon sequestration, and the conservation of biodiversity. The fluctuations in biomass can be employed as a means of monitoring the health of forests and effectively identifying management strategies that prove to be efficacious. Accurately estimating the biomass of forests is of utmost importance as it enables informed decision-making regarding the sustainable yield and quality of forest products that can be obtained.

In recent times, the adoption of SAR remote sensing has gained momentum as an effective method for monitoring and managing forests (Achard et al., 2012; Nandy et al., 2019). This trend is attributed to its unique capability to infiltrate phases of haze, fumes, and vegetal concealment. The technology in question utilizes microwaves to perceive the electromagnetic backscatter signals reflected from the Earth's surface, rendering significant insights into the composition and characteristics of the forest canopy. Numerous scholarly investigations have evinced that SAR can furnish noteworthy insights into the architecture of

✉ Noman Ali
alinoman.itz@gmail.com

Unmesh Khati
unmesh.khati@iiti.ac.in

1 UIET Hoshiarpur, Panjab University, Chandigarh, Punjab 160014, India

2 DAASE, Indian Institute of Technology Indore, Simrol, Indore, Madhya Pradesh 453552, India

forests, biomass evaluations, and the detection of variations (Laurin et al., 2018; Lucas et al., 2010; Pham et al., 2020). One of the primary advantages of SAR is its ability to provide reliable data in areas with persistent cloud cover or frequent rainfall, where traditional optical sensors may be limited in their ability to collect data (Liu et al., 2019). Aboveground biomass (AGB) and forest height are key parameters for understanding forest structure and productivity. AGB refers to the total dry weight of living vegetation aboveground, including stems, branches, leaves, and any other biomass. Height, on the other hand, is a measure of the vertical distance between the canopy and the ground. These two indicators are essential for forest inventory and management, carbon accounting, and biodiversity conservation. AGB is a key parameter in forest carbon accounting because it is used to estimate carbon sequestration potential.

SAR backscattering has been widely used for forest AGB estimation using various SAR data collected in the P, L, S, C, and X bands frequencies (Kumar et al., 2017; Le Toan et al., 1992, 2011; Luckman et al., 1998). The correlation between AGB and SAR backscatter is dependent on the nuanced interactions between electromagnetic waves and vegetative structures. Numerous investigations have been conducted, demonstrating that SAR backscatter is associated with AGB in a nonlinear fashion, with a plateauing of the relationship observed at a particular point. (Joshi et al., 2017; Mermoz et al., 2015; Yu & Saatchi, 2016). Consequently, SAR backscatter has the potential to serve as a valuable tool for inferring AGB in various tree species within forests. The interaction of SAR signals with various components of a forest, which comprises trunks, branches, leaves, and ground, is dependent on the wavelength. (Ulaby et al., 1990; Fransson, 1999; Woodhouse, 2016). Low-frequency SAR data, specifically from the P and L bands, are commonly deemed more appropriate for biomass estimation owing to their heightened saturation level. The sensitivity of SAR backscatter with respect to AGB diminishes in the presence of a thicker canopy, as the SAR signal fails to penetrate the entire canopy. The saturation level of SAR backscatter, wherein alterations in AGB no longer evoke a discernible sensitivity, is determined by various factors such as frequency, polarization mode, angle of incidence, forest type, and humidity conditions. According to current literature, saturation values for L-band SAR data have been documented to vary between 40 and 150 Mg/ha (megagrams per hectare). (Luckman, 1997; Mermoz et al., 2015; Neumann et al., 2012). Several studies have reported L-band saturation levels above 200 Mg/h (Behera et al., 2016; Sarker et al., 2012). For the P-band, the saturation is usually between 150 and 300 Mg/h (Cartus et al., 2019; Hoekman & Quinones, 2000; Liao et al., 2019), while for the X-band backscatter, it is between 30 and 80 Mg/h.

State-of-the-art techniques for the estimation of AGB and altitude are encompassed by polarimetric SAR interferometry (Kumar et al., 2017), SAR backscatter (Lucas et al., 2010), and SAR coherence (Lee et al., 2003). The present study employs L-band ALOS 2/PALSAR-2 Synthetic Aperture Radar (SAR) backscatter in conjunction with multi-parameter linear regression (LR) and random forest regression (RF) techniques for the purpose of forest AGB and forest height evaluation within a sub-tropical forest. The SAR backscatter methodology is a widely applied means of estimating AGB and altitude measurements in forested regions.

The present study aims to explore the impact of various SAR polarizations, specifically HH (Horizontal-Horizontal), HV (Horizontal-Vertical), and VV (Vertical-Vertical), on the precision of the AGB and height of the forest and harbors the potential to benefit the broader scientific constituency through the provision of a novel and economical technique for determining the forest's aboveground biomass and height. This technique is deemed suitable for forest monitoring and management, carbon assessment, and climate change modeling. The present findings can stimulate additional scholarly investigation and advancement in the area of remote sensing, ultimately resulting in the emergence of novel and enhanced technologies for the purposes of forest monitoring and management.

Study Area and Data Sets

The present study has chosen the Haldwani forest range test site, which constitutes a managed forest inhabiting an area of 405 square kilometers, sited in the foothills of the Himalayas, located in the state of Uttarakhand, India. The central point of this study area is situated at approximately 29.16°N latitude and 79.08°E longitude. Haldwani represents a forested ecosystem with a relatively low aboveground biomass, ranging from 10 to 260 metric tons per hectare. The Haldwani Forest Range can be characterized as having a relatively level surface, with a ground slope at the plot level measuring less than 5°, and an average ground slope measuring 2.39°. The forest has been fragmented into distinct sub-zones that have been designated for the cultivation of various plant species including teak (*Tectona grandis*), *Eucalyptus* sp., poplar (*Populus* sp.), Charcoal Tree (*Trema orientalis*), kanju (*Holoptelea integrifolia*) as well as mixed plantations. The taxonomic units comprised of gutel, kanju, amaltas (*Cassia fistula*), and sisau (*Dalbergia sissoo*) are considered species in academic parlance. The classifications of tree species include deciduous types such as teak, poplar, and gutel, and the evergreen species of eucalyptus.

The present forest range has been the subject of an investigation pertaining to forest height, detection of logging activities (Khati et al., 2018), and tomographic assessments (Khati et al., 2019; Kumar et al., 2017). An extensive analysis was conducted on the impacts of the phenological cycle on the PolInSAR height (Khati et al., 2017). The Forest Department office in Uttarakhand, India maintains a comprehensive account of diverse forest management practices, encompassing timber harvesting, and land clearing operations, as well as the planting and maturation stages of each sub-unit. Figure 1 outlines an engineered color composite picture of the Haldwani Timberland Run, delivered utilizing satellite imagery on May 12, 2018. Besides, the picture complements the geological positions of the plots.

The field research conducted in Haldwani encompassed two principal objectives, namely, the measurement of forest height and aboveground biomass (AGB). Specifically, in November 2015, a comprehensive survey was undertaken to evaluate the forest's height. The term "H100" was employed to denote the average height of the tallest one hundred trees within a forest area spanning one hectare (ha), a measure commonly referred to as Lorey's height. The AGB forest mapping process entailed executing a field campaign during the months of March 2017 and November 2018. Data from 60 field plots was gathered through field inventory. Each plot encompasses an area of 0.1 ha (31.6 m × 31.6 m) and is typically established in homogenous plots comprising non-specific plantations or mixed species. In the context of the study, the selection of trees in each plot was limited to those with a diameter at breast height (dbh) that exceeded 15 cm.

The dbh was measured at 1.3 meters aboveground level as per standard protocol. The diameter at breast height (dbh), height, taxonomic classification, and estimated age of every tree within the study area were determined with the aid of technical personnel from the State Forest Service. The collective tally of individual trees observed during the two surveys amounted to 4150 in number (Table 1).

Table 1 Precis information of the numerous parameters gathered ultimately of the area campaigns (Where std. dev denotes standard deviation)

Parameter	Minimum	Maximum	Mean	Std. dev
Height (m)	3.54	28.72	15.6	6.02
Stem volume m^3	0.2	417	153.9	122.1
AGB (Mg/ha)	3.76	310	123	69.5
Tree per plot	8	247	93	52

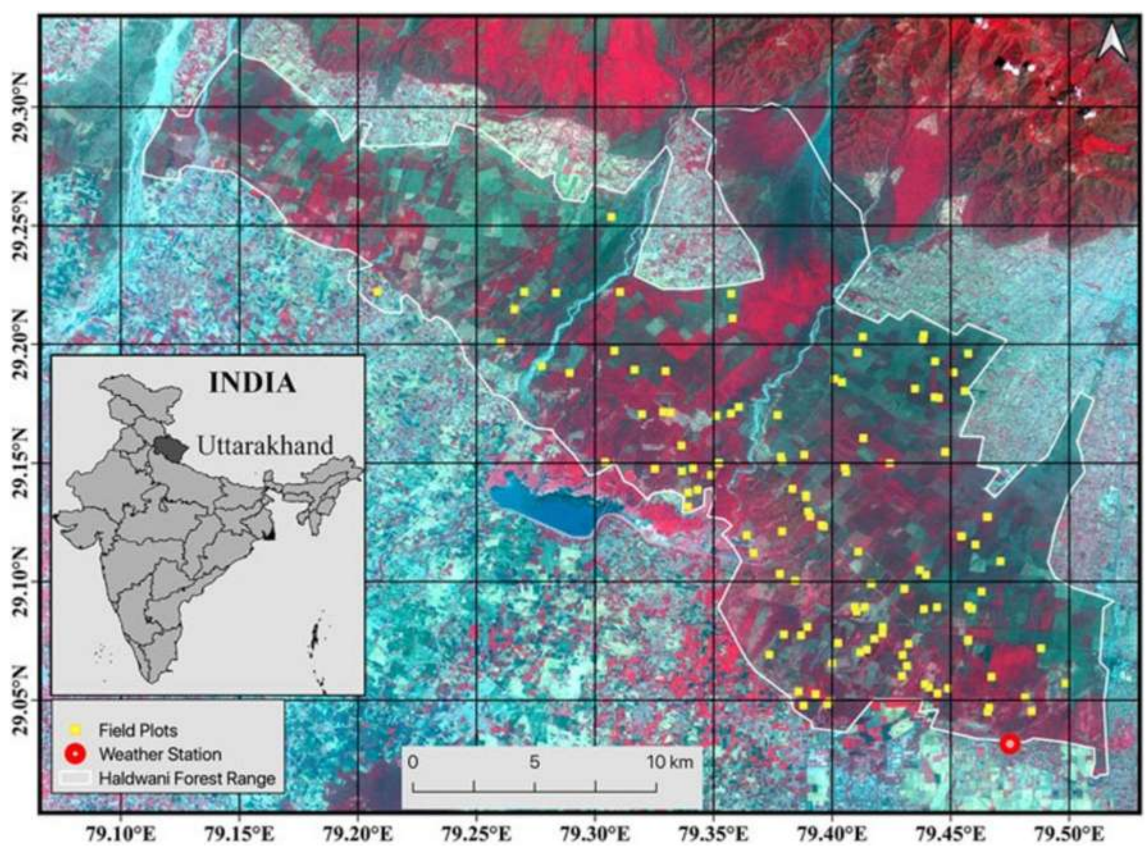

Fig. 1 The Haldwani Woodland Run as seen within the Sentinel-2 false-color composite picture taken on 12th of May, 2018. The plot area and climate station are also presented

In this paper, we also utilize the processed TanDEM-X height data derived from six distinct TanDEM-X acquisitions to complement our investigation. For a more detailed understanding of the processing methodologies, workflow, and profound analysis of the TanDEM-X Polarimetric Interferometric Synthetic Aperture Radar (PolInSAR) data, we direct interested readers to the comprehensive studies conducted by (Khati et al., 2017, 2018). The spatial distribution of Aboveground Biomass (AGB) in our designated test site is depicted in Fig. 2.

According to current scientific literature, the term AGB denotes the quantity of organic matter, either living or deceased, derived from woody vegetation in the form of trees or shrubs. A common measure of AGB is expressed as a density per unit area, typically in terms of metric units such as megagrams (Mg) of aboveground biomass per hectare.

The following steps detail the procedure used to estimate the biomass measured in the field.

- The volume of each tree was calculated using the site and species-specific volume equations given by the Forest Survey of India (1996). The input to these equations is the dbh of the tree.
- Then, using wood density information from the State Forest Department, the mass was converted to biomass for each tree.
- Finally, biomass was aggregated for the 0.1-hectare plot to obtain the AGB of the plot. This is then scaled accordingly to obtain AGB in Mg/ha.

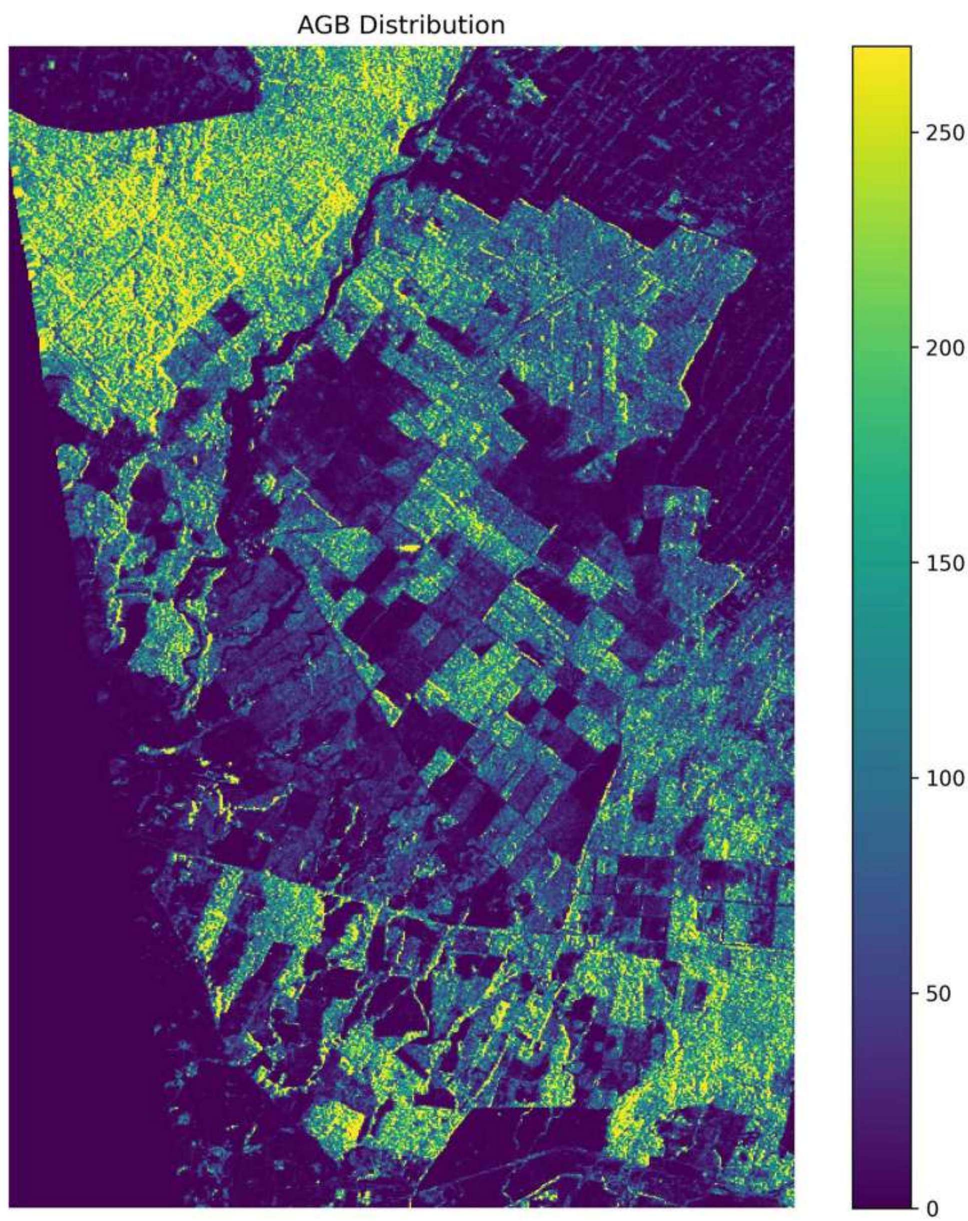

Fig. 2 Illustration of the spatial distribution of Aboveground Biomass (AGB) in the form of a geographical map

Volume Equations for Tree Species

The volume equations for major tree species are presented below:

$$Populus\ sp.: V = (-0.143393 + 3.040067 \cdot D)^2$$

$$Eucalyptus\ sp.: V = 0.02894 - (0.89284 \cdot D) + \left(8.72416 \cdot D^2\right)$$

$$Acacia\ catechu: V = 0.02384 - (0.72161 \cdot D) + \left(7.46888 \cdot D^2\right)$$

$$Tectona\ grandis: V = \left(0.08847 - 1.46936 \cdot D + 11.98979 \cdot D^2 + 1.97056 \cdot D^3\right) \cdot 1.34$$

$$Senegalia\ catechu: V = 0.02384 - (0.72161 \cdot D) + \left(7.46888 \cdot D^2\right)$$

$$Dalbergia\ sissoo: V = (-0.3238 + 3.0077 \cdot D)^2$$

Equations: Equations that show the volume of the major species.

The ensuing equations represent volumetric estimates for the principal species in the Haldwani experiment region as prescribed by the Forest Survey of India (1996). These mathematical expressions involve the variables V and D, where V represents the tree's volume and D denotes its diameter measured at its breast height. Numerous studies have emphasized the capacity of allometric equations, featuring field-measured H100 and dbh, to provide precise measurements of aboveground biomass in field settings (Feldpausch et al., 2011, 2012). Nevertheless, it is worth noting that within the Haldwani test area, numerous species lack allometric equations. Consequently, the volume equation is employed. By utilizing the prescribed equations, the volumetric measurement of the tree is computed. The present study measured the wood density of six different species, namely *Poplus* sp., *Eucalyptus sp.*, *Acacia catechu*, *Tectona grandis*, *Senegalia catechu*, and *Dalbergia sisoo*. The calculated values of wood density were 0.4, 0.697, 0.825, 0.57, 0.825, and 0.692 for *Poplus* sp., *Eucalyptus sp.*, *Acacia catechu*, *Tectona grandis*, *Senegalia catechu*, and *Dalbergia sisoo*, respectively. The field derived aboveground biomass (AGB) exhibited a range of 3.76 to 310 Mg/ha with a mean estimate of 123 Mg/ha. By utilizing the 10 most prominent trees within the 0.1 hectare site, the recorded heights of individual parcels were transformed into H100 heights (Khati et al., 2017, 2018).

ALOS-2/PALSAR-2 SAR Data

Polarimetric L-band ALOS-2/PALSAR-2 Synthetic Aperture Radar (SAR) data were acquired in a fully polarized mode over Haldwani on five distinct occasions in the year 2017, specifically on March 19, April 2, April 16, April 30, and June 11. The data in question is acquired in a step-by-step manner at the stroke of midnight, precisely at 18:39 UTC or 00:09 of the local time, exclusively within the summer season in the geographical region of India.

The acquired data possesses a range and azimuth spacing of 2.8 m and 3.2 m, respectively. Data on temperature and precipitation can be ascertained through the nearest meteorological station situated at Pantnagar Airport situated along the southern confines of Haldwani Forest. The pre-processing of ALOS-2/PALSAR-2 fully-polarimetric acquisitions is conducted to extract the (HH), (HV), and (VV) parameters.

The corpus in question comprises six distinct acquisitions, each of which is characterized by three distinct polarizations, specifically (HH), (HV), and (VV). The present study has utilized six sets of polarizations, namely (HH1) to (HH6), (HV1) to (HV6), and so forth, and as computed the mean value of each polarization over the aforementioned six dates. In a parallel manner, the (General 4 Component Decomposition) (G4U) technique has been employed to derive mean decomposition parameters, notably (Double Bounce) (Dbl), (Odd Bounce) (Odd), (Volume Scattering) (Vol), and (Helix Scattering) (Hlx), across the six dates and species data sets.

Interrelationships Among SAR Backscatter, Aboveground Biomass, and Canopy Height

We can see the connections between three important variables in Fig. 3: SAR Backscatter (particularly, HH avg, HV avg, and VV avg), Aboveground Biomass (AGB), and Canopy Height (H-100). These associations shed important light on how these crucial elements of forest ecosystems interact with one another.

SAR Backscatter and AGB: There are substantial positive connections between SAR Backscatter variables HH avg, HV avg, and VV avg and AGB. This demonstrates the usefulness of SAR data to offer insights on forest biomass by showing a correlation between greater SAR Backscatter values and increasing AGB. Particularly, HH avg and HV avg have significant positive correlations with AGB, indicating their potential as reliable measures of biomass.

SAR Backscatter and Canopy Height: Canopy Height (H-100) and SAR Backscatter variables show a little positive association. This suggests that taller forest canopies are related with higher SAR Backscatter values, suggesting a potential for SAR data to reveal information about canopy structure.

AGB and Canopy Height: An association between AGB and canopy height (H-100) is shown, highlighting the intimate connection between biomass and forest canopy structure. This relationship emphasizes how crucial it is to take into account both elements when evaluating ecosystem dynamics and forest health.

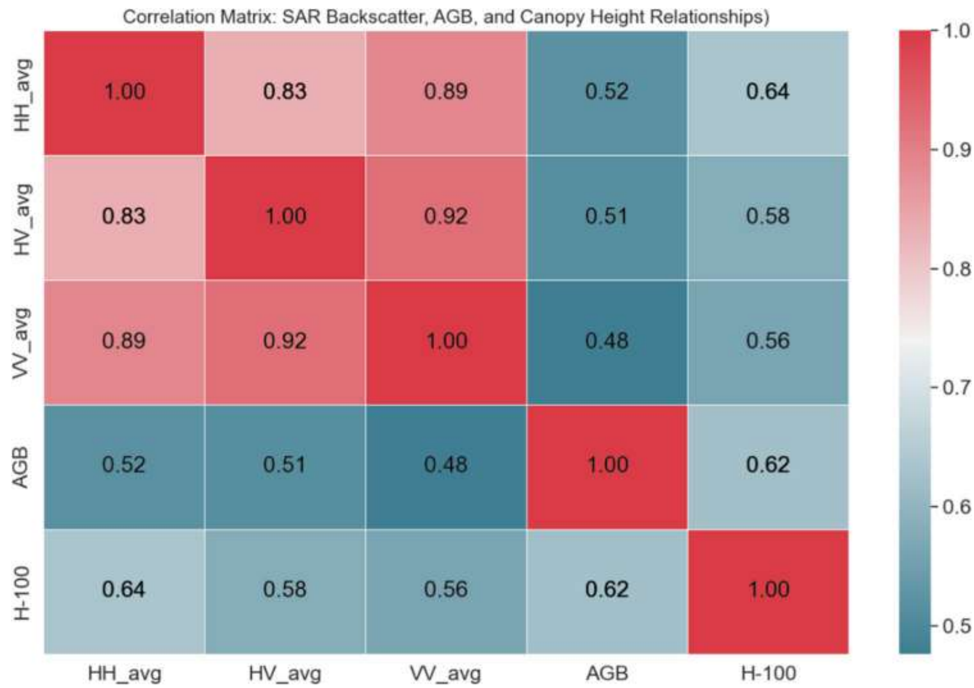

Fig. 3 Correlations Among SAR Backscatter, AGB, and Canopy Height in Forest

Model Training

Linear Regression

The linear regression model postulates a linear association between a dependent variable and a designated set of independent predictors, specifically, the Aboveground Biomass (AGB) and remote sensing predictive variables. The present investigation employed a multi-parameter linear regression approach to establish a model capable of determining forest height and Aboveground Biomass.

Random Forest

The association between forest AGB and remote sensing data is often intricate and multifaceted, which can lead to instances where conventional statistical regression techniques are insufficient in comprehensively characterizing this association. The machine learning algorithm like random forest possesses the ability to establish a multifaceted non-linear correlation amidst an uncertain distribution of data, and is capable of adroitly integrating varied data sources to ameliorate the precision of prediction. The Random Forest (RF) algorithm displays a notable resistance to noise that may impede data accuracy, and also avoids instances of overfitting (Breiman, 2001). RF has the capability to discern significant variables and produce autonomous metrics for forecasting inaccuracies (Adam et al., 2014). RF has been widely applied as a classification algorithm (Luo et al., 2016) and used for time series forecasting in large-scale regression-based applications (Adam et al., 2014; Tyralis & Papacharalampous, 2017; Tyralis et al., 2019).

Results and Discussions

Evaluation of Linear Regression

The multi-parameter linear regression is used to train a model to predict the aboveground biomass (AGB) and H100 (height of forest) variables and evaluate the model using evaluation metrics like root mean square error (RMSE), mean absolute error (MAE), and R-squared value, or the coefficient of determination (Table 2).

It was found that AGB can be predicted with an R^2 value of 0.60, RMSE of 49.04 Mg/ha, which measures the standard deviation of residuals and MAE of 39.13. And H100 (Height of forest) is predicted with an R^2 value of 0.56, RMSE of 3.97 m, and MAE of 3.29 m, which represents the average of the absolute difference between the actual and predicted values (Fig. 4).

Interpreting the multiple linear regression prediction error plot: Figs. 5 and 6 emphasize the uncertainty in estimating Aboveground Biomass (AGB), especially concerning the Height of the forest (H-100) based on the selected predictors. It indicates that our current model faces challenges in accurately capturing the relationships. This observation suggests the presence of complex and potentially non-linear interactions within our data. In light of this, we recognize the

Table 2 Evaluation metrics for linear regression

Model	RMSE	MAE	R^2
AGB prediction	49.04 (Mg/ha)	39.139 (Mg/ha)	0.60
Height prediction	3.97 (m)	3.29 (m)	0.56

importance of exploring alternative modeling approaches to improve the accuracy of AGB predictions.

Evaluation of Random Forest Regression

The random forest (RF) algorithm is applied in three different ways:

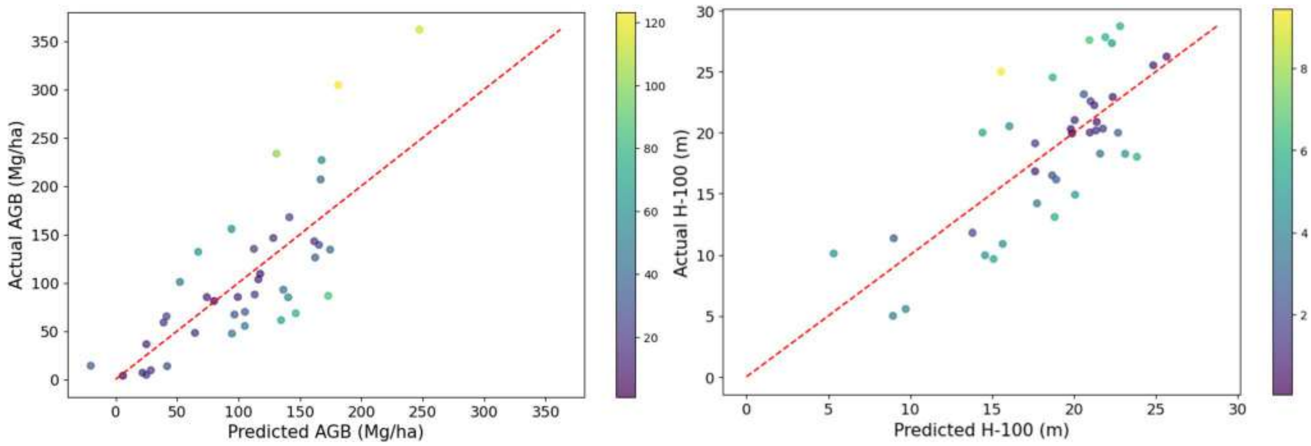

Fig. 4 Scatter Plots between actual AGB and predicted AGB, and between actual H100 (Height of forest) and predicted H100

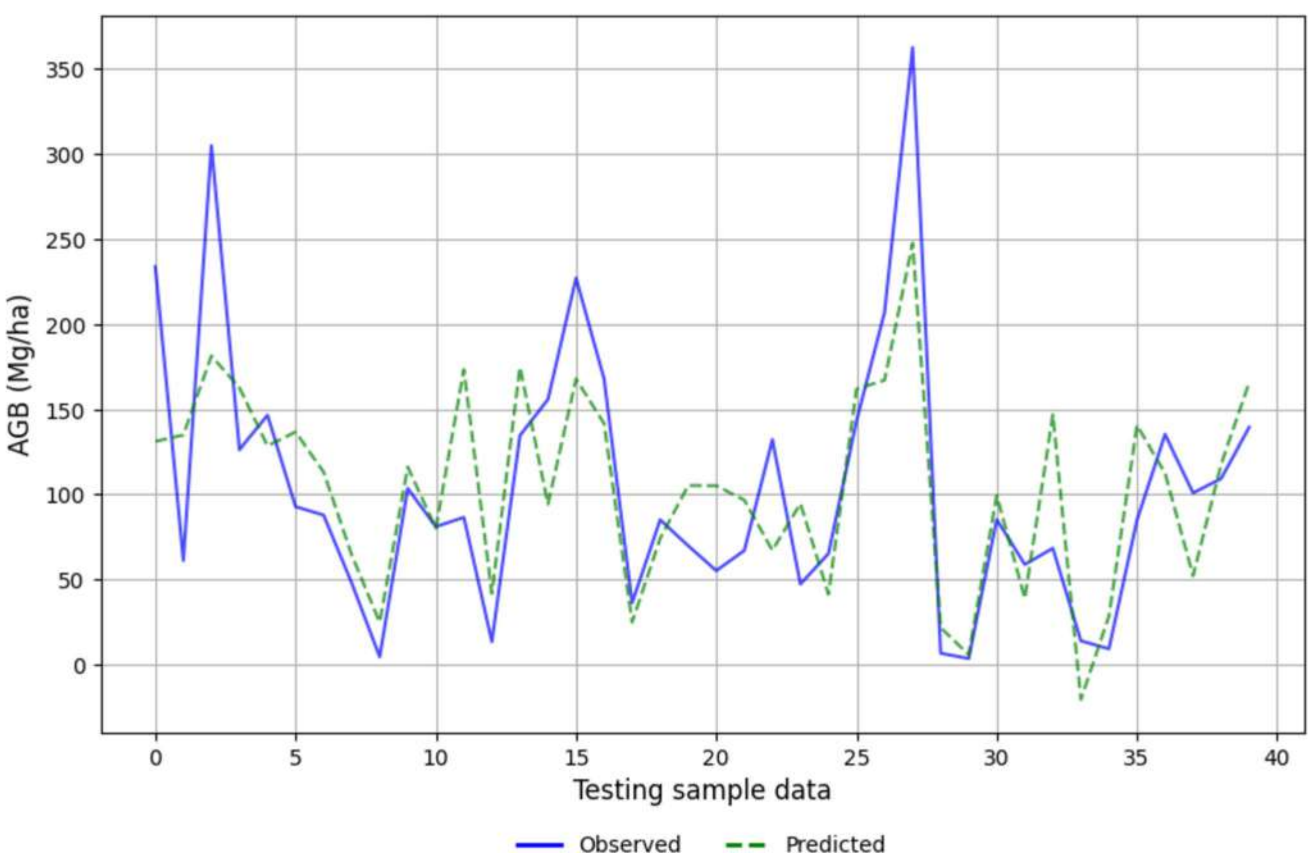

Fig. 5 Comparison of Observed and Predicted AGB (Mg/ha) Values for Testing Sample Data

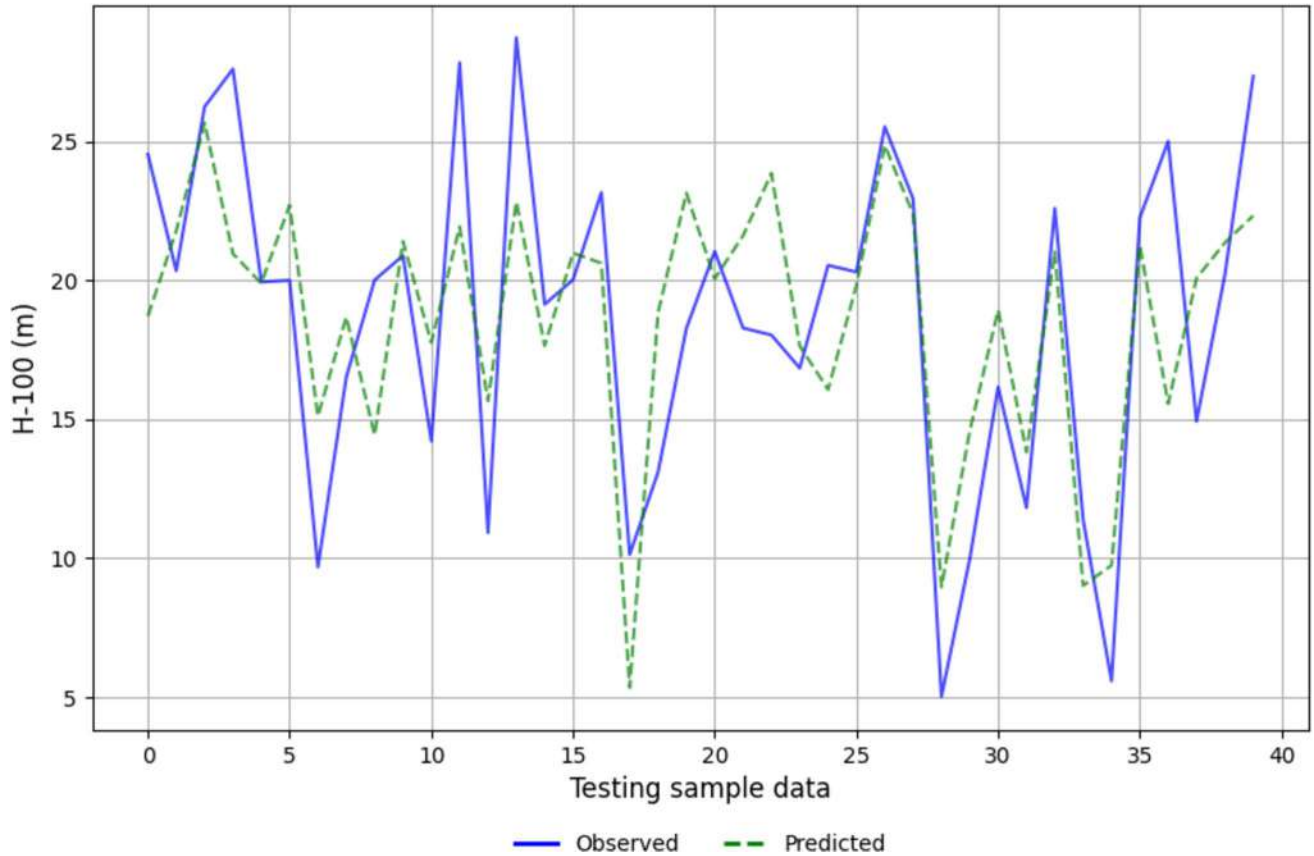

Fig. 6 Comparison of Observed and Predicted H-100 (m) Values for Testing Sample Data

1. We model only the aboveground biomass (AGB).
2. We model only the height of the forest.
3. We model both the AGB and height simultaneously.

And evaluated the model for all three cases using evaluation metrics like root mean square error (RMSE), R-squared value or the coefficient of determination, and mean absolute error (MAE) (Table 3).

It was found that in the first case, AGB can be predicted with an R^2 value of 78.04, RMSE of 36.51 Mg/ha, which measures the standard deviation of residuals and MAE of 27.89 Mg/ha, which represents the average of the absolute difference between the actual and predicted values, and in the second case, H100 (Height of forest) is predicted with an R^2 value of 79.16, RMSE of 2.74 (m), and MAE of 2.15 (m). For the combined prediction case we found the R^2 value is 70.45, RMSE of 24.54, and MAE of 17.18.

Estimation of Aboveground Biomass and Forest Canopy Height: Individual Assessments

Figure 7 displays scatter plots for actual vs. predicted values of Aboveground Biomass (AGB) and Height of Forest (H100). These plots visually assess the accuracy of our predictive models, helping us gauge how well our predictions align with real-world data.

Feature importance: In examining the graph depicting feature importance, we uncover essential insights into the factors driving our Random Forest Regression models' predictions.

Table 3 Evaluation metrics for random forest

Model fo	RMSE	MAE	R^2
AGB prediction	36.51 (Mg/ha)	27.89 (Mg/ha)	78.04
Height prediction	2.15 (m)	2.74 (m)	79.16
Combined prediction	24.54 $((\text{Mg/ha})^2+(\text{m})^2)^{1/2}$	17.18 (Mg/ha—m)	70.45

For AGB estimation, key features include acquisition ('H2' and'H4') and backscatter polarizations (HH avg, VV avg).'H2' and'H4' dominated, while HH avg, VV avg also contributed significantly (Fig. 8).

In Canopy Height (H-100) estimation, Dbl avg (Double bounce) is the most critical feature, followed by HV avg (HV polarization) and Hlx avg (helix scattering). These factors play pivotal roles in predicting canopy height.

Comparatively, 'H2' and 'H4' are crucial for AGB, while Dbl avg leads in height estimation. These findings highlight distinct predictors for accurate AGB and Canopy Height estimates.

To construct our Random Forest Regression models for estimating AGB (Aboveground Biomass) and H-100 (Canopy Height), we adopted a rigorous approach that encompassed both feature selection and hyperparameter tuning.

For the AGB estimation, we selected a forest of decision trees with n estimators set to 7 and min samples split at 2. To ensure model robustness, we set min samples leaf to 1 and max depth to 6. These hyperparameter configurations proved effective in capturing the complex relationships among predictor variables, resulting in accurate biomass estimations.

In the case of'H-100' estimation, we fine-tuned our model with n estimators set to 8 to build a forest of trees. To control overfitting, we limited max depth to 9 and specified min samples leaf at 2. We also set max features to 1.0 to further enhance predictive performance.

Estimation of Aboveground Biomass and Forest Canopy Height: Combined Assessment

Figure 9 shows scatter plots that indicate how Aboveground Biomass (AGB) and Height of Forest (H-100) data actually and anticipated values line up. These charts provide a visual evaluation of the precision of our predictive models and reveal how well our forecasts correspond to actual data. It demonstrates the combined predictive strength of our models, allowing us to simultaneously estimate AGB and H-100.

Feature importance: As illustrated in the figure above, our combined Random Forest Regression model for estimating both'AGB' (Aboveground Biomass) and'Height of the Forest' reveals crucial insights into feature importance. Dbl avg (Double bounce) takes center stage with a feature importance score of 0.3650, highlighting its pivotal role in our model. Vol avg (volume scattering) and Odd avg (odd bounce) follow suit with scores of 0.2428 and 0.1323, respectively, emphasizing the importance of scattering characteristics (Fig. 10).

Hlx avg (helix scattering), HH avg (average backscatter with HH polarization), HV avg (average backscatter with HV polarization), and VV avg (average backscatter with VV polarization) also make significant contributions with scores ranging from 0.0269 to 0.1145. Collectively, these insights underscore the robustness of our model, allowing simultaneous estimation of AGB and Canopy Height while highlighting the key features driving precise predictions.

We used a rigorous methodology that included feature selection and hyperparameter tweaking to build our Random Forest Regression models for simultaneously predicting AGB and H-100 (Canopy Height). We used a forest of decision trees with n estimators set to 9 and min samples split at 2 for the "AGB" model. We set the min samples leaf to 1 and the max depth to 8 to ensure model robustness. Accurate biomass prediction resulted from these hyperparameter configurations' success in capturing the intricate interactions between predictor factors.

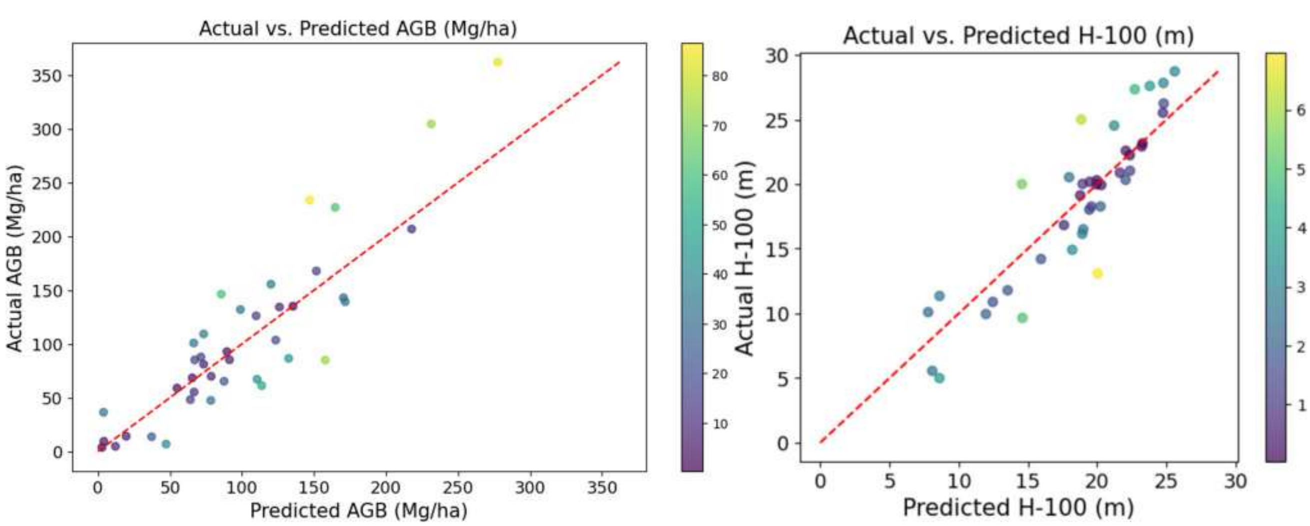

Fig. 7 Scatter Plots between actual AGB and predicted AGB, and between actual H100 (Height of forest) and predicted H100 for RF (individual)

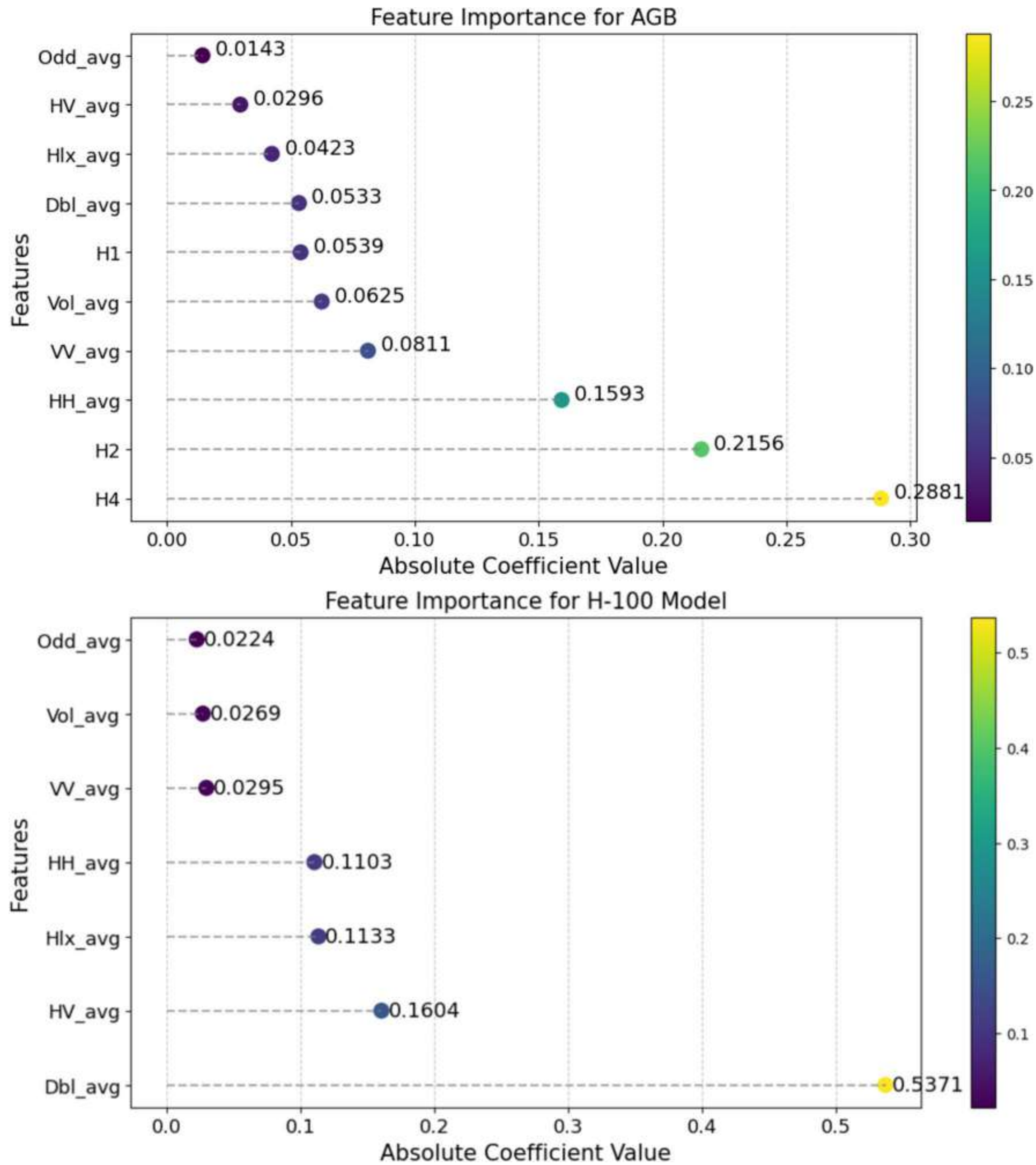

Fig. 8 Feature importance of individual assessments of AGB and H100

Interpreting the random forest regression prediction error plot: AGB and H-100 (tree height) observed and anticipated values are thoroughly compared in the Fig. 11. The solid line in this graphic depiction represents the expected values, while the dispersed data points show the actual values.

A thorough look reveals that the projected values closely match the observed ones, demonstrating a noteworthy degree of agreement. While some areas suggest the model is facing some challenges. In summary, the graph suggests that our model possesses a commendable capability to predict AGB and H-100 simultaneously with a reasonable degree of accuracy.

Figure 12 illustrates the predictive efficacy of the models in estimating AGB both independently and in conjunction with height prediction. Similarly, Fig. 13 delineates the model performance in predicting height individually and in tandem with AGB prediction. This concurrent evaluation sheds light on the synergistic effects and potential improvements achieved through the simultaneous prediction of AGB and height. The nuanced insights garnered from these comparisons contribute to a

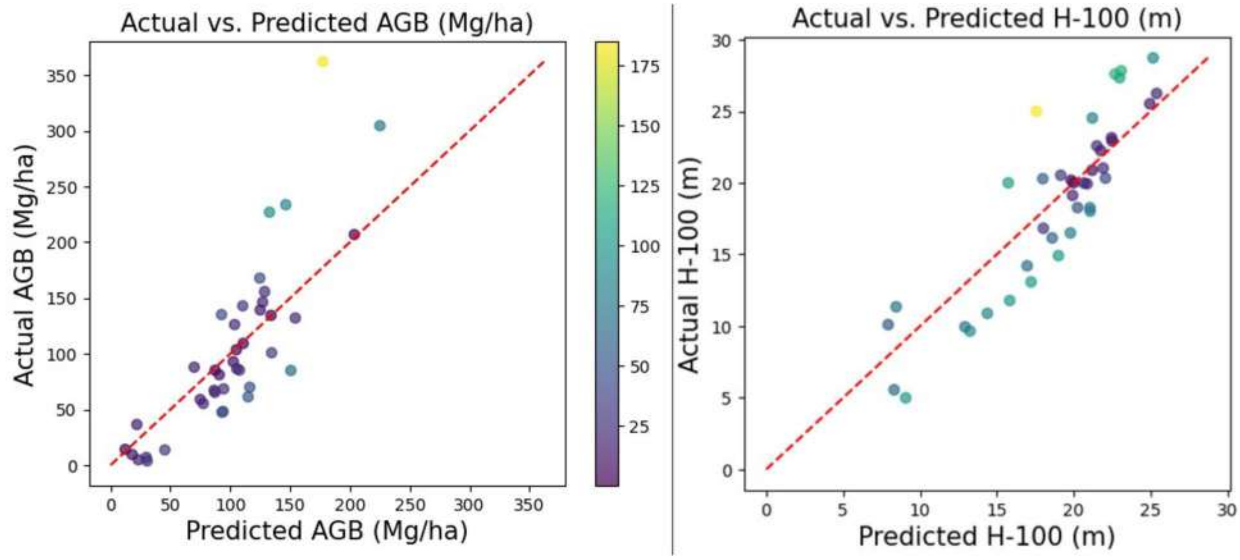

Fig. 9 Scatter Plots between actual AGB and predicted AGB, and between actual H100 (Height of forest) and predicted H100 for the combined case

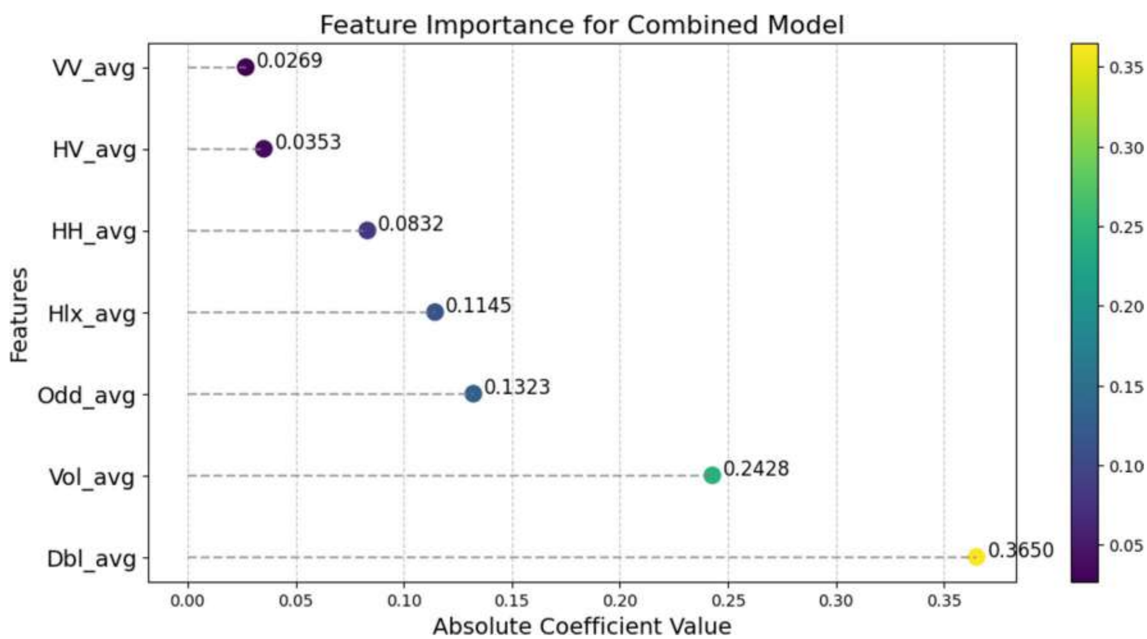

Fig. 10 Feature importance for combined assessment of AGB and H100

comprehensive understanding of the predictive capabilities of the employed models in the study area.

Moderate Correlation and Random Forest's Nonlinear Capability:

The model excels in estimating both AGB and H-100 due to the moderate positive correlation (0.62) between the two variables. This correlation simplifies predictions as changes in one variable often correspond to changes in the other (Schmid et al., 2023). In several drug response predictions, (Rahman et al., 2017) discovered that multivariate Random Forests outperform Random Forests when the outputs are closely related. Additionally, the model benefits from Random Forest's innate capacity to handle complex and nonlinear relationships, and due to the fact that backscatter and AGB are related (Joshi et al., 2015). With

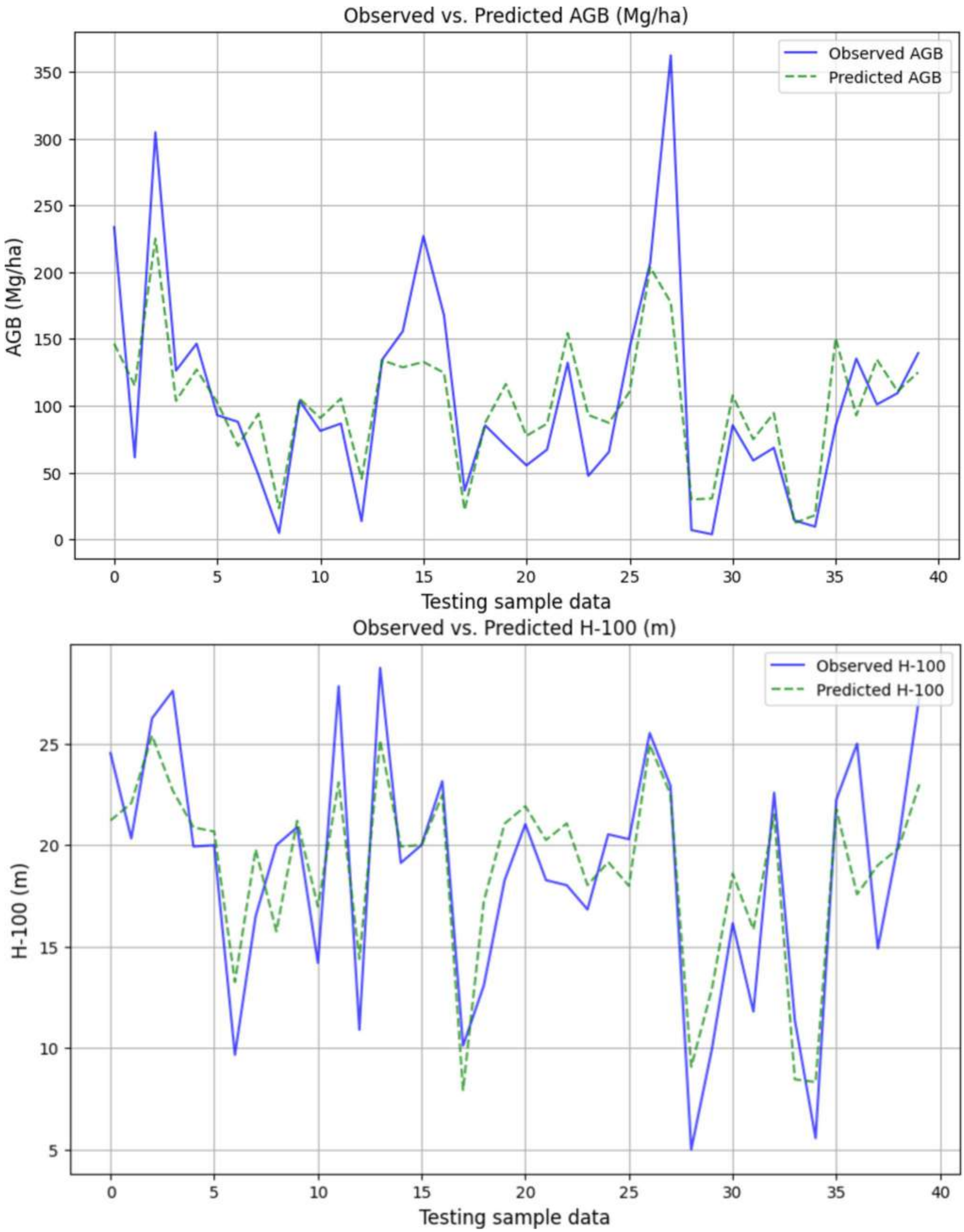

Fig. 11 Comparison of Observed and Predicted AGB and H-100 Values for Testing Sample Data for combined model

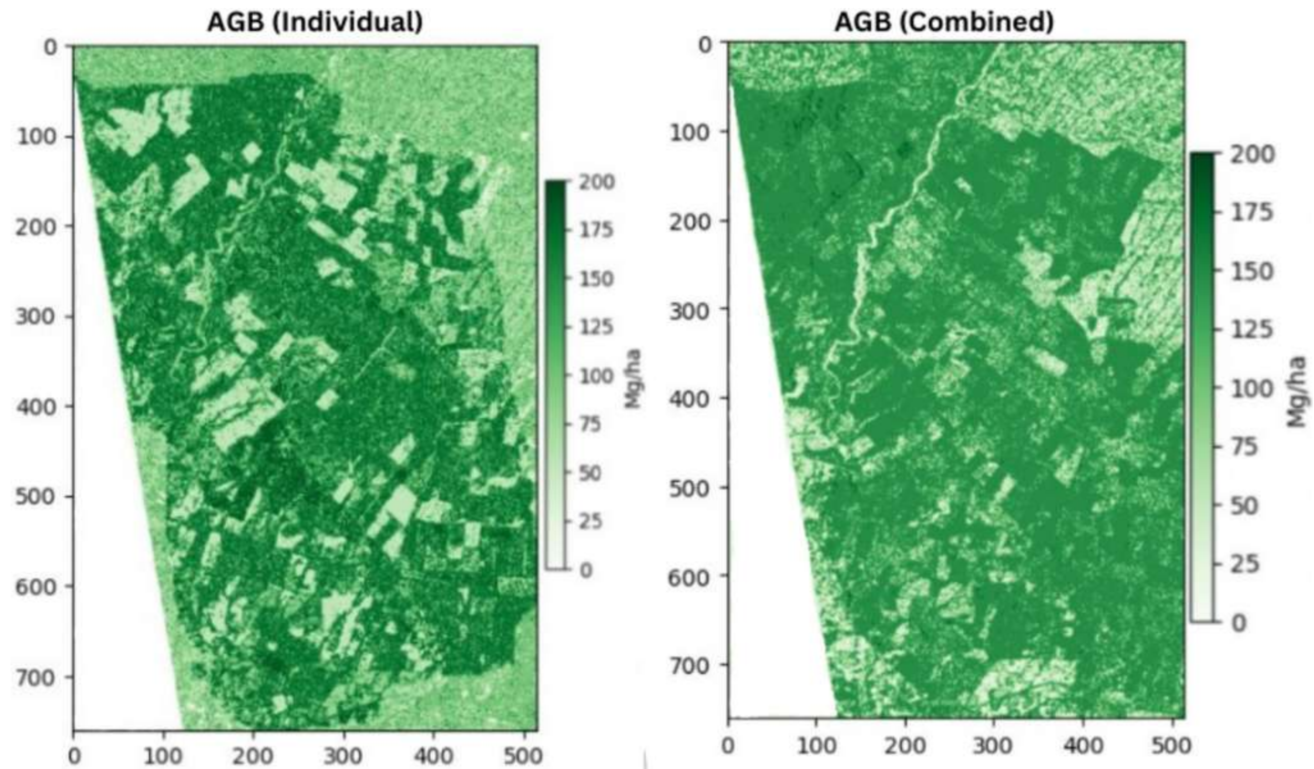

Fig. 12 Comparison of individual and combined prediction for AGB values

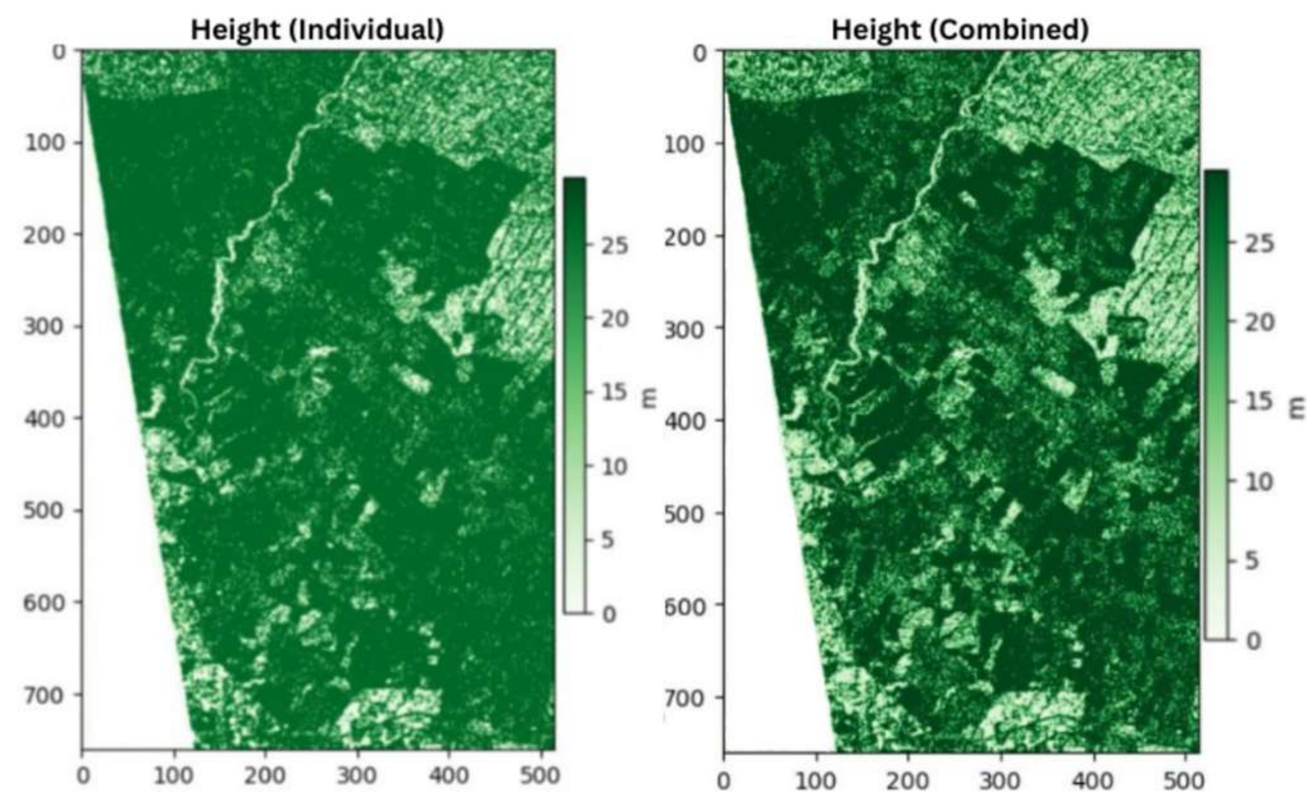

Fig. 13 Comparison of individual and combined prediction for Height values

its ensemble of decision trees, Random Forest adeptly captures intricate patterns and interactions, making it well-suited for the nuanced dependencies between AGB, H-100, and related features. This synergy between correlation and modeling capability enables the model to provide accurate and consistent estimations for both targets.

Conclusion

The work involves the integration of remote sensing, machine learning, and forest ecology. Such interdisciplinary research can lead to new insights and approaches for addressing complex environmental issues. The study proposes an innovative approach to estimating forest aboveground biomass and height using SAR data and machine learning algorithms, which combine the advantages of these technologies and can provide accurate and cost-effective estimates. The use of PALSAR data and incorporation of field measurements further enhance the accuracy of the estimates, and the comparison with other methods provide valuable insights into the performance of different approaches.

This study focused on the Haldwani timberland trial area (29.16 N and 79.08 E), situated within the Himalayan foothills in Uttarakhand, India. The site was chosen to investigate the estimation of aboveground biomass (AGB) and forest height using Synthetic Aperture Radar (SAR) data and machine learning algorithms. The findings of this research demonstrate the potential of utilizing L-Band backscatter data for the joint retrieval of AGB and forest height.

The outcomes indicate a significant enhancement in accuracy through the utilization of the Random Forest model in comparison to the linear regression model. The results affirm the viability of employing time-series L-Band backscatter data for the prediction of both aboveground biomass and forest height.

Acknowledgements We are deeply grateful to the anonymous reviewers for their valuable comments and feedback, which greatly enhanced the quality of our manuscript. Their expertise and dedication to scholarly excellence are much appreciated.

Funding This research was not funded by any external sources or organizations.

Declarations

Conflict of interest The authors declare the following regarding competing interests related to the work submitted for publication: The authors declare that they have no competing financial interests related to the work submitted for publication. This declaration confirms the absence of any financial conflict of interest and the absence of a requirement for publication consent.

Ethical approval No ethical approvals required.

Consent for Publication Not applicable.

References

Achard, F., & Hansen, M. C. (2012). *Global forest monitoring from earth observation*. CRC Press.

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. https://doi.org/10.1080/01431161.2013

Behera, M. D., Tripathi, P., Mishra, B., Kumar, S., Chitale, V. S., & Behera, S. K. (2016). Above-ground biomass and carbon estimates of *Shorea robusta* and *Tectona grandis* forests using Quad-POL ALOS PALSAR data. *Advances in Space Research, 57*(2), 552–561. https://doi.org/10.1016/j.asr.2015.11.010

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

Brown, S. (2002). Measuring carbon in forests: Current status and future challenges. *Environmental Pollution, 116*(3), 363–372. https://doi.org/10.1016/S0269-7491(01)00212-3

Cartus, O., Santoro, M., Wegmuller, U., & Rommen, B. (2019). Benchmarking the retrieval of biomass in boreal forests using p-band SAR backscatter with multi-temporal c- and l-band observations. *Remote Sensing, 11*(14), 1695. https://doi.org/10.3390/rs11141695

Feldpausch, T. R., Banin, L., Phillips, O. L., Baker, T. R., Lewis, S. L., Quesada, C. A., Affum-Baffoe, K., Arets, E. J. M. M., Berry, N. J., Bird, M., Brondizio, E. S., de Camargo, P., Chave, J., Djagbletey, G., Domingues, T. F., Drescher, M., Fearnside, P. M., Franca, M. B., Fyllas, N. M., … Lloyd, J. (2011). Height-diameter allometry of tropical forest trees. *Biogeosciences, 8*(5), 1081–1106. https://doi.org/10.5194/bg-8-1081-2011

Feldpausch, T. R., Lloyd, J., Lewis, S. L., Brienen, R. J. W., Gloor, M., Monteagudo Mendoza, A., Lopez-Gonzalez, G., Banin, L., Abu Salim, K., Affum-Baffoe, K., Alexiades, M., Almeida, S., Amaral, I., Andrade, A., Aragao, L. E. O. C., Araujo Murakami, A., Arets, E. J. M. M., Arroyo, L., Aymard, G. A., … Phillips, O. L. (2012). Tree height integrated into pantropical forest biomass estimates. *Biogeosciences, 9*(8), 3381–3403. https://doi.org/10.5194/bg-9-3381-2012

Hoekman, D. H., & Quinones, M. J. (2000). Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the colombian amazon. *IEEE Transactions on Geoscience and Remote Sensing, 38*(2), 685–696. https://doi.org/10.1109/36.841998

Houghton, R. A., Hall, F., & Goetz, S. J. (2009). Importance of biomass in the global carbon cycle: Biomass in the global carbon cycle. *Journal of Geophysical Research: Biogeosciences*. https://doi.org/10.1029/2009JG000935

Joshi, N., Mitchard, E., Schumacher, J., Johannsen, V., Saatchi, S., & Fensholt, R. (2015). L-band SAR backscatter related to forest cover, height and aboveground biomass at multiple spatial scales across Denmark. *Remote Sensing, 7*(4), 4442–4472. https://doi.org/10.3390/rs70404442

Joshi, N., Mitchard, E. T. A., Brolly, M., Schumacher, J., Fernandez-Landa, A., Johannsen, V. K., Marchamalo, M., & Fensholt, R. (2017). Understanding 'saturation' of radar signals over forests. *Scientific Reports, 7*(1), 3505. https://doi.org/10.1038/s41598-017-03469-3

Khati, U., Lavalle, M., & Singh, G. (2019). Spaceborne tomography of multi-species Indian tropical forests. *Remote Sensing of Environment, 229*, 193–212. https://doi.org/10.1016/j.rse.2019.04.017

Khati, U., Singh, G., & Ferro-Famil, L. (2017). Analysis of seasonal effects on forest parameter estimation of Indian deciduous forest using TerraSAR-x PolInSAR acquisitions. *Remote Sensing of Environment, 199*, 265–276. https://doi.org/10.1016/j.rse.2017.07.019

Khati, U., Singh, G., & Kumar, S. (2018). Potential of space-borne PolInSAR for forest canopy height estimation over india—a case study using fully polarimetric L -, C -, and X -band SAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11*(7), 2406–2416. https://doi.org/10.1109/JSTARS.2018.2835388

Kumar, S., Khati, U. G., Chandola, S., Agrawal, S., & Kushwaha, S. P. S. (2017). Polarimetric SAR interferometry based modeling for tree height and aboveground biomass retrieval in a tropical deciduous forest. *Advances in Space Research, 60*(3), 571–586. https://doi.org/10.1016/j.asr.2017.04.018

Laurin, G. V., Balling, J., Corona, P., Mattioli, W., Papale, D., Puletti, N., Rizzo, M., Truckenbrodt, J., & Urban, M. (2018). Above-ground biomass prediction by sentinel-1 multi-temporal data in central Italy with integration of ALOS2 and sentinel-2 data. *Journal of Applied Remote Sensing, 12*(1), 1. https://doi.org/10.1117/1.JRS.12.016008

Le Toan, T., Beaudoin, A., Riom, J., & Guyon, D. (1992). Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing, 30*(2), 403–411. https://doi.org/10.1109/36.134089

Le Toan, T., Quegan, S., Davidson, M. W. J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., & Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment, 115*(11), 2850–2860. https://doi.org/10.1016/j.rse.2011.03.020

Liao, Z., He, B., Quan, X., van Dijk, A. I. J. M., Qiu, S., & Yin, C. (2019). Biomass estimation in dense tropical forest using multiple information from single-baseline p-band PolInSAR data. *Remote Sensing of Environment, 221*, 489–507. https://doi.org/10.1016/j.rse.2018.11.027

Liu, C., Chen, Z., Shao, Y., Chen, J., Hasi, T., & Pan, H. (2019). Research advances of SAR remote sensing for agriculture applications: A review. *Journal of Integrative Agriculture, 18*(3), 506–525. https://doi.org/10.1016/S2095-3119(18)62016-7

Lee, J.-S., Cloude, S. R., Papathanassiou, K. P., Grunes, M. R., & Woodhouse, I. H. (2003). Speckle filtering and coherence estimation of polarimetric sar interferometry data for forest applications. *IEEE Transactions on Geoscience and Remote Sensing, 41*(10), 2254–2263. https://doi.org/10.1109/TGRS.2003.817196

Liu, K., Wang, J., Zeng, W., & Song, J. (2017). Comparison and evaluation of three methods for estimating forest above ground biomass using TM and GLAS data. *Remote Sensing, 9*(4), 341. https://doi.org/10.3390/rs9040341

Lucas, R., Armston, J., Fairfax, R., Fensham, R., Accad, A., Carreiras, J., Kelley, J., Bunting, P., Clewley, D., Bray, S., Metcalfe, D., Dwyer, J., Bowen, M., Eyre, T., Laidlaw, M., & Shimada, M. (2010). An evaluation of the ALOS PALSAR l-band backscatter—above ground biomass relationship queensland, australia: Impacts of surface moisture condition and vegetation structure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3*(4), 576–593. https://doi.org/10.1109/JSTARS.2010.2086436

Luckman, A. (1997). A study of the relationship between radar backscatter and regenerating tropical forest biomass for spaceborne SAR instruments. *Remote Sensing of Environment, 60*(1), 1–13. https://doi.org/10.1016/S0034-4257(96)00121-6

Luckman, A., Baker, J., Honzak, M., & Lucas, R. (1998). Tropical forest biomass density estimation using JERS-1 SAR: Seasonal variation, confidence limits, and application to image mosaics. *Remote Sensing of Environment, 63*(2), 126–139. https://doi.org/10.1016/S0034-4257(97)00133-8

Luo, Y.-M., Huang, D.-T., Liu, P.-Z., & Feng, H.-M. (2016). An novel random forests and its application to the classification of mangroves remote sensing image. *Multimedia Tools and Applications, 75*(16), 9707–9722. https://doi.org/10.1007/s11042-015-2906-9

Mermoz, S., Rejou-Mechain, M., Villard, L., Le Toan, T., Rossi, V., & Gourlet-Fleury, S. (2015). Decrease of l-band SAR backscatter with biomass of dense forests. *Remote Sensing of Environment, 159*, 307–317. https://doi.org/10.1016/j.rse.2014.12.019

Nandy, S., Ghosh, S., Kushwaha, S. P. S., & Senthil Kumar, A. (2019). Remote sensing-based forest biomass assessment in northwest himalayan landscape. In R. R. Navalgund, A. Senthil Kumar, & S. Nandy (Eds.), *Remote sensing of northwest Himalayan ecosystems* (pp. 285–311). Springer.

Neumann, M., Saatchi, S. S., Ulander, L. M. H., & Fransson, J. E. S. (2012). Assessing performance of l- and p-band polarimetric interfero-metric SAR data in estimating boreal forest above-ground biomass. *IEEE Transactions on Geoscience and Remote Sensing, 50*(3), 714–726. https://doi.org/10.1109/TGRS.2011.2176133

Pham, T. D., Le, N. N., Ha, N. T., Nguyen, L. V., Xia, J., Yokoya, N., To, T. T., Trinh, H. X., Kieu, L. Q., & Takeuchi, W. (2020). Estimating mangrove above-ground biomass using extreme gradient boosting decision trees algorithm with fused sentinel-2 and ALOS-2 PALSAR-2 data in can Gio biosphere reserve, Vietnam. *Remote Sensing, 12*(5), 777. https://doi.org/10.3390/rs12050777

Rahman, R., Otridge, J., & Pal, R. (2017). IntegratedMRF: Random forest-based framework for integrating prediction from different data types. *Bioinformatics, 33*(9), 1407–1410. https://doi.org/10.1093/bioinformatics/btw765

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. https://doi.org/10.1016/j.isprsjprs.2012.03.002

Schmid, L., Gerharz, A., Groll, A., & Pauly, M. (2023). Tree-based ensembles for multi-output regression: Comparing multivariate approaches with separate univariate ones. *Computational Statistics & Data Analysis, 179*, 107628. https://doi.org/10.1016/j.csda.2022.107628

Tyralis, H., & Papacharalampous, G. (2017). Variable selection in time series forecasting using random forests. *Algorithms, 10*(4), 114. https://doi.org/10.3390/a10040114

Tyralis, H., Papacharalampous, G., & Tantanee, S. (2019). How to explain and predict the shape parameter of the generalized extreme value distribution of streamflow extremes using a big dataset. *Journal of Hydrology, 574*, 628–645. https://doi.org/10.1016/j.jhydrol.2019.04.070

Ulaby, F. T., Sarabandi, K., Mcdonald, K., Whitt, M., & Dobson, M. C. (1990). Michigan microwave canopy scattering model. *International Journal of Remote Sensing, 11*(7), 1223–1253. https://doi.org/10.1080/01431169008955090

Woodhouse, I. H. (2016). Predicting backscatter-biomass and height-biomass trends using a macroecology model. *IEEE Transactions on Geoscience and Remote Sensing, 44*(4), 871–877. https://doi.org/10.1109/TGRS.2006.872356

Yu, Y., & Saatchi, S. (2016). Sensitivity of l-band SAR backscatter to aboveground biomass of global forests. *Remote Sensing, 8*(6), 522. https://doi.org/10.3390/rs8060522

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.