Aboveground biomass mapping and uncertainty estimation in tropical forests using machine learning with multispectral and SAR data

Naveen Babu Kanda, Debabrata Behera, Rahul Gour, Kurian Ayushi, Ayyappan Narayanan, Narayanaswamy Parthasarathy

▶ **To cite this version:**

Naveen Babu Kanda, Debabrata Behera, Rahul Gour, Kurian Ayushi, Ayyappan Narayanan, et al.. Aboveground biomass mapping and uncertainty estimation in tropical forests using machine learning with multispectral and SAR data. Trees, Forests and People, 2026, 25, pp.101271. ⟨10.1016/j.tfp.2026.101271⟩. ⟨hal-05606289⟩

HAL Id: hal-05606289

https://hal.science/hal-05606289v1

Submitted on 29 Apr 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Trees, Forests and People 25 (2026) 101271

Contents lists available at ScienceDirect

Trees, Forests and People

journal homepage: www.sciencedirect.com/journal/trees-forests-and-people

Aboveground biomass mapping and uncertainty estimation in tropical forests using machine learning with multispectral and SAR data

Naveen Babu Kanda [a,b,c], Debabrata Behera [b,d], Rahul Gour [e], Kurian Ayushi [b,f], Ayyappan Narayanan [b,*], Narayanaswamy Parthasarathy [a]

[a] Department of Ecology and Environmental Sciences, School of Life Sciences, Pondicherry University, Puducherry, 605 014, India
[b] Department of Ecology, Institut Français de Pondichéry / French Institute of Pondicherry, UMIFRE 21 (CNRS-MEAE) - UAR3330, 11, Saint Louis Street, Pondicherry - 605 001, India
[c] National Biobank of Thailand, National Science and Technology Development Agency, Pathum Thani, 12120, Thailand
[d] Centre for Doctoral Studies, Directorate of Research, Manipal Academy of Higher Education, Manipal, Karnataka, 576 104, India
[e] Kenneth Anderson Nature Society, Hosur, Tamil Nadu, 635 110, India
[f] Department of Agriculture, Ecotrophology, and Landscape Development, National and International Nature Conservation, Anhalt University of Applied Sciences, 06406 Bernburg, Germany

ARTICLE INFO

Keywords:
Aboveground biomass
Random forest
Quantile regression forest
Sentinel 2, texture
Uncertainty estimation
Western ghats

ABSTRACT

Tropical forests are major reservoirs of biomass and play a critical role in regulating the global climate. Accurate mapping of aboveground biomass (AGB) is therefore essential for quantifying carbon stocks, monitoring forest dynamics, and supporting climate mitigation strategies. However, tropical forests exhibit notable spatial variation, resulting in increased spatial uncertainty when estimating AGB. Therefore, estimating AGB requires integrating multiple data sources. To provide new insights into AGB modeling in the biodiverse Western Ghats region of India, this study evaluates the integration of Sentinel-1 (S1) and Sentinel-2 (S2) data, leveraging the complementary strengths of radar and optical sensors. Three machine learning techniques, Random Forest (RF), Support Vector Machine, and Extreme Gradient Boosting, were compared for their performance in accurately predicting the AGB. RF was found to be the best-performing machine learning model, with an R^2 of 0.55, a root mean square error of 75.34 Mg/ha, and a mean absolute error of 57.79 Mg/ha. Texture parameters extracted from the multispectral bands of S2 and S1's backscatter emerged as the most important variables with a relative importance score of 34.9%. The spatially predicted AGB map from the best-performing RF model exhibited a range of 16.64 to 375.73 Mg/ha with a mean value of 215.37 Mg/ha, whereas the uncertainty value ranged from 11.78 to 121.01 Mg/ha. Based on the predicted map, a high AGB was observed in semievergreen forests, followed by the moist deciduous forests and tree plantations. Low AGB was recorded in dry deciduous forests. This study recommends the integration of S1 and S2 data along with several auxiliary datasets, including topography, climate, and canopy height, to substantially improve the accuracy of AGB estimation in tropical forests. The results of this study are promising and support reliable carbon stock assessments, thereby contributing to the sustainable management of these forests within the framework of the Reducing Emissions from Deforestation and Forest Degradation (REDD+) initiatives.

1. Introduction

Tropical forests play a crucial role in global carbon dynamics, accounting for approximately 55% of the biomass stored in terrestrial vegetation and exhibiting high carbon sequestration rates per unit area (Réjou-Méchain et al., 2017; Harris et al., 2021). Aboveground biomass (AGB), defined as the total mass of living vegetation above the soil surface per unit area, is a key metric for quantifying these dynamics. Accurate estimation of AGB is essential for monitoring forest structure, productivity, and carbon storage (Rajashekar et al., 2018). It also serves as an important indicator of ecosystem function and is recognized as an Essential Biodiversity Variable (EBV; Reddy et al., 2021), supporting

* Corresponding author at: Department of Ecology, Institut Français de Pondichéry / French Institute of Pondicherry, UMIFRE 21 (CNRS-MEAE) - UAR3330, 11, Saint Louis Street, Pondicherry - 605 001, India.
E-mail address: ayyappan.n@ifpindia.org (A. Narayanan).

https://doi.org/10.1016/j.tfp.2026.101271

Available online 19 April 2026
2666-7193/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

biodiversity conservation and sustainable land management. Reliable AGB measurements further enhance our understanding of ecosystem responses to environmental change and inform strategies to mitigate climate change and reduce carbon emissions.

Traditionally, field-based estimation of AGB relies on allometric equations that incorporate stem diameter, wood density, and tree height, and are calibrated using destructive sampling. These approaches provide reliable estimates at the plot level; however, they are constrained by high costs, labor intensity, and the logistical impracticality of sampling large trees, which may introduce biases and limit their representativeness. To address these limitations, pantropical allometric models have been developed to provide a unified framework that accounts for environmental heterogeneity and structural variation across forest types (Chave et al., 2014). Despite their improved generalizability, the application of these models remains challenging at large spatial scales in tropical regions due to complex terrain, dense forest structure, and logistical constraints. Consequently, the integration of field inventory data with remote sensing technologies has emerged as an effective approach for scaling up AGB estimation (Lu et al., 2014; Réjou-Méchain et al., 2019; Tian et al., 2023). Substantial variability in AGB has been reported across forest types in the Western Ghats, India. For instance, the mean AGB in deciduous forests of Mudumalai ranges from 75.9 – 353.6 Mg/ha (Indirabai and Nilsson, 2024); in semi-evergreen forests of Kanyakumari wildlife sanctuary, from 282 – 364 Mg/ha (Kothandaraman et al., 2020); and in evergreen forests of Uppangala, Pushpagiri wildlife sanctuary, from 458.35 Mg/ha to 560.19 Mg/ha (Behera et al., 2025, Wilson et al., 2023). Capturing such pronounced variability in AGB across the landscape requires a robust, multi-faceted modeling approach supported by rigorous statistical validation.

Upscaling AGB estimation across landscapes typically follows a hierarchical modeling framework that integrates field inventory data with remote sensing observations sensitive to forest structural attributes. Remote sensing data are broadly categorized into optical, synthetic aperture radar (SAR), and light detection and ranging (LiDAR) systems. The selection of appropriate predictor variables is critical for accurate AGB estimation (Su et al., 2020; Forkuor et al., 2020), with commonly used variables including spectral bands, vegetation indices, SAR backscatter, and topographic metrics. In addition, several studies emphasize the importance of incorporating all pertinent variables into the AGB estimation process (Li et al., 2020; Zhang et al., 2023). However, each data source has inherent limitations. Optical data often suffer from signal saturation in dense tropical canopies due to their complex vertical structure, while SAR backscatter shows reduced sensitivity in high-biomass forests. Although LiDAR provides highly accurate measurements of forest structure and AGB (Jha et al., 2020), its limited accessibility and spatial coverage restrict its application for large-scale, wall-to-wall mapping.

To address these limitations, advanced approaches such as image transformation and texture analysis have been proposed. Texture metrics derived from optical and SAR data capture spatial heterogeneity and structural variation, enabling improved AGB estimation beyond the saturation limits of pixel-based approaches (Kelsey and Neff, 2014; Ghosh and Behera, 2018). Additionally, combining multiple predictors derived from passive remote sensing has been shown to enhance model performance, particularly in dense tropical forests (Behera et al., 2024). Recent studies further highlight the benefits of hierarchical modeling frameworks that integrate field data, airborne LiDAR, and spaceborne remote sensing. For example, Sagang et al. (2020) used airborne LiDAR as an intermediate link between field inventory and satellite data, while Saarela et al. (2023) developed a three-stage model-based approach that enables robust uncertainty quantification through variance-covariance estimation. Despite these advances, the availability of field inventory and airborne LiDAR data remains limited in many tropical regions (Rodda et al., 2024). In this context, the Global Ecosystem Dynamics Investigation (GEDI) mission represents a major advancement by providing spaceborne LiDAR measurements of forest vertical structure at a footprint resolution of ~25 m. However, its spatially discrete sampling design limits its applicability for continuous, wall-to-wall biomass mapping across large landscapes. Consequently, freely available optical and SAR datasets remain essential for achieving spatially continuous AGB estimation, particularly in cloud-prone tropical regions.

In recent years, machine learning (ML) algorithms and nonparametric regression approaches have become increasingly prominent in AGB estimation (Lu et al., 2014; Tian et al., 2023). These methods are particularly effective in capturing complex nonlinear relationships between AGB and predictor variables, while enabling the integration of multi-source remote sensing data (Lu et al., 2014; Ghosh and Behera, 2018). As a result, ML-based approaches have consistently demonstrated improved accuracy in biomass estimation across diverse forest ecosystems. Several studies highlight the advantages of integrating multi-sensor data within ML frameworks. For example, Laurin et al. (2018) showed that combining L-band (ALOS-2) and C-band from Sentinel-1(S1) SAR data with optical imagery significantly improved AGB predictions in Mediterranean forests. Similarly, Su et al. (2020) employed a hybrid random forest and co-kriging approach using Landsat and ALOS PALSAR data, demonstrating strong predictive performance by incorporating spectral, textural, and topographic variables.

Comparative analyses further indicate that advanced ML algorithms such as random forest (RF) and Extreme Gradient Boosting (XGBoost) outperform traditional regression methods in modeling AGB. For instance, Li et al. (2020) and Dang et al. (2019) emphasized the importance of texture variables and showed that ensemble-based models provide robust and reliable biomass estimates. In addition, approaches that quantify uncertainty, such as Monte Carlo simulations, have been successfully integrated with ML models to improve the reliability of predictions. In the Indian context, recent studies have demonstrated the effectiveness of ML-based approaches for AGB estimation using multi-source datasets. Nandy et al. (2021) combined ICESat-2, S1, and Sentinel-2 (S2) data to improve biomass and canopy height estimation, while Ghosh and Behera (2018) highlighted the synergistic use of Sentinel datasets for accurate biomass modeling in dense tropical forests. These studies collectively underscore the capability of ML algorithms to model complex forest structures and enhance large-scale AGB estimation.

A plethora of remote sensing datasets from open-access and commercial sensors have been used for AGB modeling (Lu et al., 2014; Jha et al., 2020, 2021). In recent years, Sentinel data have gained prominence due to their global coverage, high spatial and temporal resolution, and free accessibility, demonstrating strong performance in biomass estimation (Dang et al., 2019; Forkuor et al., 2020; Singh et al., 2022; Ayushi et al., 2024). S1 provides information on vegetation structure independent of weather conditions or time of day, while S2 offers high-resolution multispectral data for characterizing vegetation properties. The integration of S1 and S2 enables the complementary use of structural and spectral information, making it well-suited for large-scale AGB estimation. In addition to remote sensing data, environmental variables such as elevation, precipitation, temperature, and topographic features play a critical role in influencing biomass distribution. Previous studies have demonstrated that incorporating bioclimatic and topographic variables alongside optical and SAR datasets improves AGB estimation (Fararoda et al., 2021). These factors influence forest composition, species distribution, and biomass accumulation in tropical ecosystems (McEwan et al., 2011; Rodrigues et al., 2020; Behera et al., 2025).

In the present study, field-based AGB was estimated using the pantropical allometric equation developed by Chave et al. (2014), while species-specific equations were applied for plantation species. Although Forest Survey of India (FSI) equations are widely used as national standards, their application at local scales can introduce substantial estimation errors (Babu et al., 2023a; Behera et al., 2025). Building on these considerations, this study aims to: (1) evaluate and compare the

performance of three machine learning algorithms - random forest (RF), support vector regression (SVR), and extreme gradient boosting (XGBoost) for AGB estimation using integrated multi-source data; (2) identify key predictor variables and assess their contribution to AGB modeling; (3) generate spatially explicit AGB maps using the best-performing model; and (4) quantify and map prediction uncertainty to improve the reliability of AGB estimates.

2. Materials and methods

2.1. Study area

The study was conducted in the Shettihalli landscape, an ecologically sensitive zone located within the Western Ghats, a globally recognized biodiversity hotspot in southern India. It is a mosaic of natural and human-dominated land uses spanning over 750 km^2 and lies between 13°40′1.2″N and 14°4′58.8″N, and 75°10′1.2″E and 75°34′58.8″E. The present vegetated cover of the landscape is about 503 km^2 (Fig. 1). The forests in Shettihalli, like other regions in the Western Ghats, have experienced significant fragmentation, primarily a result of extensive logging and the conversion of forested areas for monoculture plantations like *Tectona grandis, Acacia, Eucalyptus,* and agriculture (Babu et al., 2023a). The topography of the landscape is relatively flat in the eastern regions compared to undulating terrain in the western regions, inducing variations in mean annual precipitation (1044 – 3076 mm), as well as mean annual temperature (17.5 – 29.92°C) (Kanda et al., 2021). Kanda et al. (2021) and Babu et al. (2023b) employed remote sensing and detailed field studies to classify the forest types in the study area. The identified land-use types included tropical semi-evergreen forest, moist deciduous forest, dry deciduous forest, and tree plantations including *Tectona grandis, Eucalyptus grandis, Acacia auriculiformis, A. mangium,* and *Pinus roxburghii* (Babu et al., 2023a). Shettihalli is rich in biodiversity, harbouring 263 plant species, of which 39 are endemic to the Western Ghats (Kanda et al., 2021). Some of the most prevalent native tree species in these forests include *Terminalia paniculata, Lagerstroemia microcarpa, Xylia xylocarpa, Dalbergia latifolia, Grewia tiliifolia,* and *Macaranga peltata* (Kanda et al., 2021). A recent study (Babu et al., 2023a) reported that plantation types contribute an average AGB of 139–293 Mg/ha in this landscape. In recent decades, this diversity and biomass have been threatened by high human density within the sanctuary, with over 124 villages and an extensive interior road network facilitating illegal activities (Kanda et al., 2021). Frequent forest fires are another common threat during the summer, coinciding with high temperatures (Babu et al., 2023b).

2.2. Methodology

A schematic representation of the methodology involved in estimating AGB and associated uncertainty is depicted in Fig. 2. Broadly, the approach involves 1) Earth observation and field data, 2) Variable processing and extraction, 3) Modeling, 4) Model evaluation, and 5) Spatial prediction.

2.2.1. Field inventory and AGB estimation

Between 2019 to 2022, 170 georeferenced (Garmin GPS 66SR <5 m accuracy) 0.1 ha (31.3 m × 31.3 m;) vegetation plots, oriented towards north were established using stratified random sampling across the forested region in the Shettihalli landscape (Fig. 1). Random points were generated in ArcGIS, with a 500 m minimum distance between adjacent plots to ensure maximum landscape coverage. This plot size, also employed in similar studies (Rajashekar et al., 2018; Behera et al., 2024; Ayushi et al., 2024), offers an optimal balance. Considering the topography, climate, canopy density, forest types and their extent, and the accessibility, plot distribution includes 22 plots in semi-evergreen forests (SEF), 71 in moist deciduous forests (MDF), 27 in dry deciduous forests (DDF), and 50 in plantation forests (PLN). In each plot (Fig. 1), all

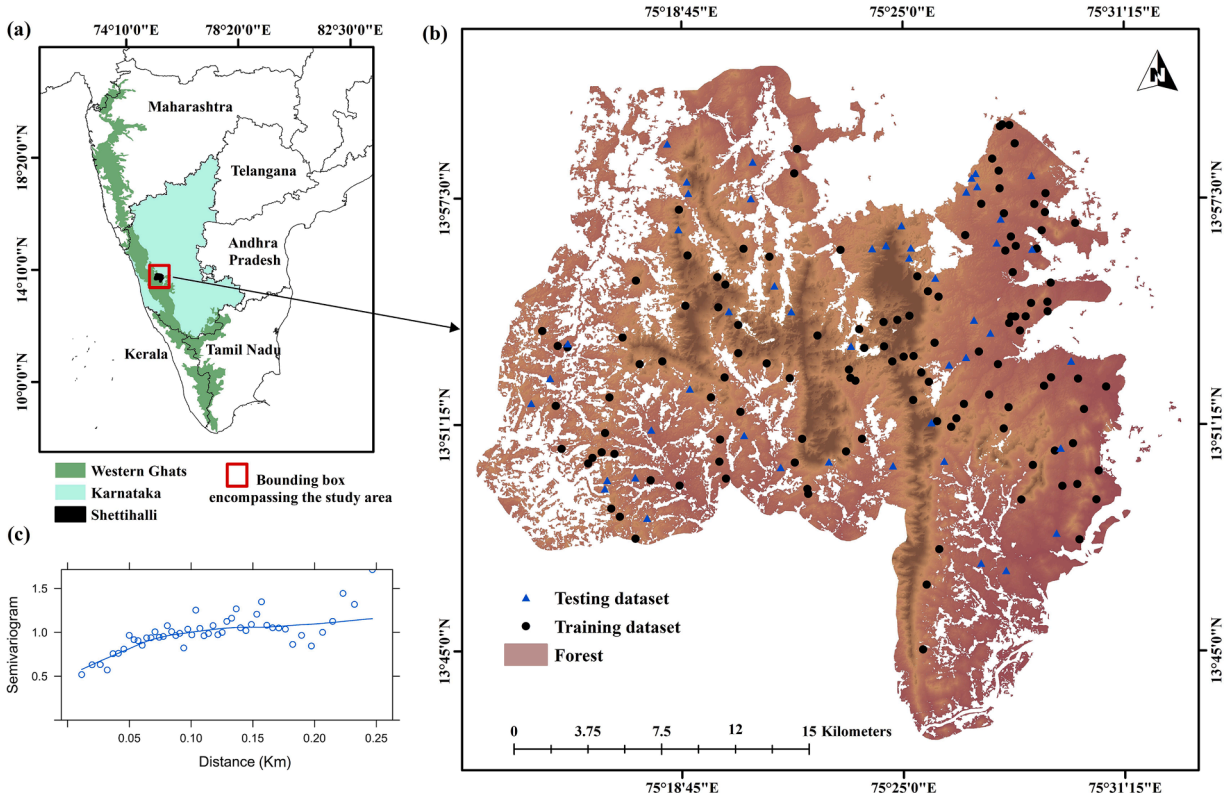

Fig. 1. (a) Location of the study area in the central Western Ghats region of peninsular India, (b) the distribution of forest inventory plots divided into training and testing sites to model aboveground biomass (AGB) and (c) semi-variogram showing the moderate spatial autocorrelation among the plots considered in the study.

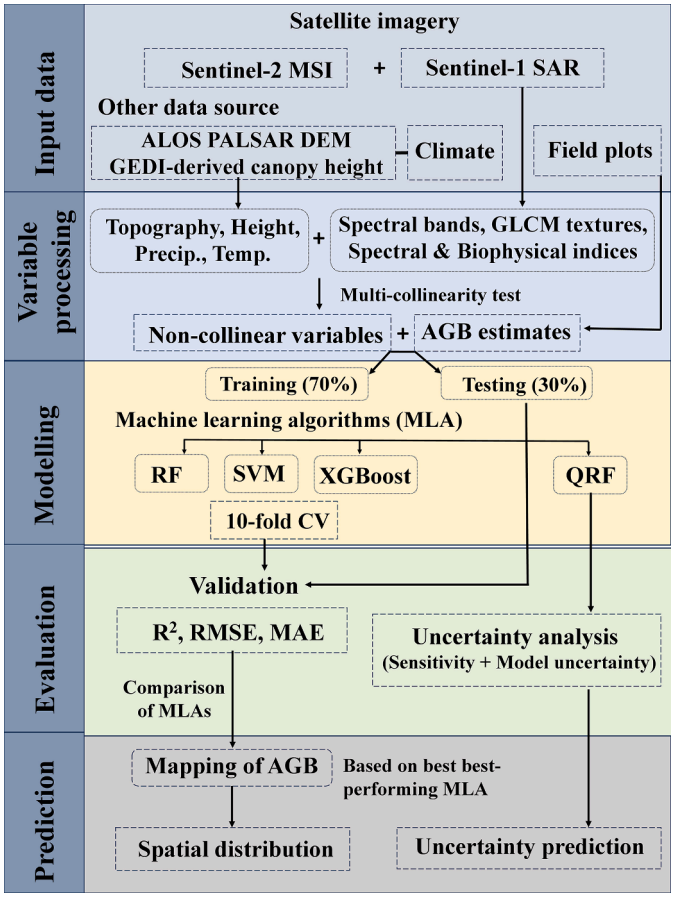

Fig. 2. The schematic representation of the methodology involved in estimating AGB and associated uncertainty in Shettihalli landscape.

trees ≥ 10 cm girth at breast height (GBH; defined at 1.3 m above-ground) were measured for their girth along with the taxonomic information. Trees with multi-stems were measured separately, and for trees with buttresses, GBH was measured above the point of measurement (i.e., 1.3 m).

The AGB of individual trees was estimated using the pantropical allometric model of Chave et al. (2014) (Eq. 1).

$$AGBest = \exp\left(-2.024 - 0.896 \cdot E + 0.920 \cdot \ln(WD) + 2.795 \cdot \ln(D) - 0.0461 \cdot \left[\ln(D)^2\right]\right)$$

Where WD is the wood density (g cm^{-3}), D represents the DBH (cm) of the individual trees, and E represents environmental stress. Wood density values were obtained from global databases (Chave et al., 2009; Zanne et al., 2009), and for species lacking specific data, genus or family-level averages were used. Individual tree AGB was calculated for each stem and summed across all stems within each plot to obtain total plot-level AGB, expressed in megagrams per hectare (Mg/ha). All computations were performed in R using the BIOMASS package (Réjou-Méchain et al., 2017).

We acknowledge that species-specific allometries provide greater accuracy compared to generalized equations. However, the high species diversity in tropical forests makes it challenging to apply species-specific regression models for AGB estimation. Although we examined India-specific models, such as the Forest Survey of India (FSI) volume equations (FSI 1996; FSI 2015), their limited coverage across the diverse species restricted their applicability. Consequently, we relied on a pantropical equation, which has been widely applied in Indian tropical forests and the Western Ghats region (Osuri et al., 2020; Ayushi et al., 2024; Dar et al., 2024; Najeeb et al., 2025). The study area contains a significant portion of tree plantations (Babu et al., 2023a), for which the Chave et al. (2014) equation may not always be suitable (Walker et al., 2016). Therefore, we used species-specific equations for the PLN type following standard literature, including the FSI equations. The choice and detailed description of allometric equations are provided in Babu et al. (2023a).

2.2.2. *Earth observing datasets*

2.2.2.1. SAR and optical data. The study used ground range detected (GRD) data from S1, a C-band SAR operating at a frequency of 5.405 GHz, obtained in dual polarisation (vertical transmit-vertical receive - VV) and (vertical transmit-horizontal receive - VH) Interferometric Wide swath mode for December 2021 directly from the Google Earth Engine platform (GEE). The images were pre-processed within GEE using S1 toolbox, which involved removing thermal noise, performing radiometric calibration, and correcting for terrain effects. The terrain-corrected backscatter coefficients were logarithmically transformed and translated to decibels (dB). Thus, we obtained VV and VH backscatter images. Similarly, GEE was used to obtain cloud-free S2 level 2A Ortho-rectified Bottom-Of-Atmosphere reflectance data for the same time period (December 2021). The S2′s multi-spectral imager (MSI) sensor captures radiation in 13 spectral bands, including bands in the vegetation red-edge zone, with a temporal resolution of around 5 days. The spatial resolutions are 10 m for four bands, 20 m for six bands, and 60 m for three bands. The spectral bands with spatial resolutions other than 10 m were resampled to 10 × 10 m using bilinear upscaling to maintain a consistent pixel size. Band 1 (Coastal aerosol), Band 9 (Water vapour), and Band 10 (SWIR-Cirrus) were removed from the analysis. Principal component analysis (PCA) was used to reduce the dimensionality of the multispectral S2 data to its two most informative components, PCA1 and PCA2, which together explained 90% of the variance. Previously, this methodology improved AGB estimation by highlighting key biomass-related features, reducing noise, and making the data easier to interpret (Behera et al., 2024). For texture analysis, both PCA1 and PCA2, along with the original VV and VH backscatter images from S1, were selected. Texture features were then derived from each of the four images using the Gray-Level Co-Occurrence Matrix (GLCM) with a 3×3 window, following Haralick et al. (1973), yielding seven texture metrics per image. The GLCM approach constructs a matrix that records how often particular pairs of pixel values appear together within a specified area of an image. As a second-order texture measure, GLCM analyzes the frequency of pixel pairs with similar brightness at defined spatial arrangements. This method is especially effective for detecting subtle, fine-scale structural patterns in dense tropical forest canopies (Csillik et al., 2019; Kupidura, P., 2019).

The spectral and biophysical indices, such as Fraction of Absorbed Photosynthetically Active Radiation, Fraction of Green Cover Vegetation, and Leaf Area Index, were calculated using the biophysical processor function in the SNAP toolbox as proxies for biophysical parameters, following the PROSAIL radiative transfer method (Jacquemoud et al., 2009). We also computed the Normalised Difference Vegetation Index (NDVI) as described by Rouse et al. (1974) and the Visible Atmospherically Resistant Index following Gitelson et al. (2002), NDVI of red-edge bands following (Forkour et al., 2020), Normalised Difference Water Index following (Gao, 1996), and Stress Related Vegetation Index 1,2,3 following Thenkabail et al. (1994) for inclusion in the analysis. Further, two SAR indices representing the discrepancy between VH and VV, as well as the sum of VH and VV, were incorporated as suggested by Laurin et al. (2018).

2.2.2.2. Climate variables. We used two bioclimatic variables, mean annual precipitation (MAP) and mean annual temperature (MAT), obtained from the CHELSA (https://chelsa-climate.org/) bioclimatic variables dataset by Karger et al. (2021).

2.2.2.3. Physical variables. The ALOS Digital Elevation Model (DEM) was procured from Alaska Satellite Facility (https://search.asf.alaska.edu/#/) for generating topographic features such as elevation and slope, while sine and cosine aspect values were transformed using "cosine" and "sine" function in ArcGIS v 10.3. We also computed topographic position index, topographic wetness index, terrain ruggedness index using SAGA GIS. GEDI-derived canopy height was obtained from the global canopy height map developed by Potapov et al. (2021) and clipped to the study area. A detailed list of predictor variables and their descriptions is provided in Table 1, and the extended metadata is provided in Table S1 (Supplementary Information).

2.2.3. Multicollinearity and variable selections

Multicollinearity among predictor variables was assessed using Pearson's correlation and the variance inflation factor (VIF) implemented in the R package usdm (Naimi, 2015). Variables with VIF $\geq$ 10 or with an absolute Pearson's correlation coefficient $\geq$ 0.7 (Fig. 3; Babu et al., 2023b) were excluded, resulting in the selection of 31 predictors out of 62 for model development (Table 1). Pearson's correlation matrix was computed using the R package ENMTools v1.0 (Warren et al., 2021).

2.2.4. Feature extraction

All predictor layers with different spatial resolutions were resampled to 10 m using bilinear interpolation. This technique preserves image quality by determining new pixel values using averages computed from adjacent pixels in both the horizontal and vertical directions (Teoh et al., 2008). The center point of each plot was geo-tagged using a Garmin 66SR GPS, and the coordinates were used to extract values using the 'extract by values' function in ArcGIS. Predictor variable values were extracted from the resampled layer stack at a 10-meter spatial resolution. To minimize geolocation errors and spatial discrepancies between field reference points and Earth observation data, an 18-meter radius buffer was applied around the center of each vegetation sample plot, following Carreiras et al. (2013). This approach has previously been shown to be effective for integrating field data with remote sensing variables (Ghosh and Behera, 2018; Behera et al., 2024).

2.2.5. Modeling - machine learning algorithms and tuning

Three ML models were implemented for AGB estimation: RF, SVR, and XGBoost. RF is a non-parametric ML technique commonly used in RS applications. It is recognized for its high predictive accuracy in identifying intricate, non-linear connections among predictor variables (Breiman, 2001). It also determines the significance of each variable by calculating the average decrease in impurity index. In order to tune the RF model, we used three hyperparameters: mtry, num.trees, and min.node.size, which were set using a grid search method in the "caret" package (Kuhn, 2017) using 10-fold cross-validation. The model with the lowest RMSE was chosen for modeling, and the resulting optimal values are mtry: 16, ntree: 300, and min.node.size: 1. To perform RF, we used the "RandomForest" package (Liaw and Wiener, 2002) in the R environment (R Core Team, 2022).

SVR transforms training samples from a low-dimensional space to a high-dimensional space for linear separation using kernel functions (Han et al., 2021). SVR minimizes the total variance of the sample points with respect to a hyperplane and separates the points by fitting a linear regression equation. Different kernel functions can be used in SVR. In our study, we used radial basis functions (Zhao et al., 2016). The SVR method was performed with the 'e1071' package (Meyer et al., 2015) in R. Using SVR, we performed a grid search approach (100 times) where the optimal parameters were tuned, and the reliable combination of C, epsilon, and gamma was detected to get the maximum R^2. The optimal values of the hyperparameters C, gamma, and epsilon were 5, 0.03, and 0.1, respectively.

Similarly, XGBoost is an ensemble learning method based on the gradient boosting framework to improve the prediction performance of

Table 1
Explanatory variables used to build aboveground biomass (AGB) model in Shettihalli, central Western Ghats, India.

Sl. no	Category	Code	Variable description	Source
1	Climate	**MAP**	Mean annual precipitation	https://chelsa-climate.org/
		MAT	Mean annual temperature	"
2	Physical	**Cos_aspect**	Cosine aspect	ArcGIS
		Elevation	Elevation	https://search.asf.alaska.edu/#/
		GEDI_height	LiDAR derived forest canopy height	Potapov et al. (2021)
		Sin_aspect	Sine aspect	ArcGIS
		Slope	Slope angle	"
		TPI	Topographic Position Index	SAGA GIS
		TRI	Terrain Ruggedness Index	"
		TWI	Topographic Wetness Index	"
3	Sentinel-1	S1_VH	Sentinel-1 Vertical-Horizontal backscatter	Open Access Hub (copernicus.eu)
		S1_VV	Sentinel-1 Vertical-Vertical backscatter	"
4	Sentinel-2	S2_B11	Sentinel-2 SWIR 1 band	"
		S2_B12	Sentinel-2 SWIR 2 band	"
		S2_B2	Sentinel-2 Blue band	"
		S2_B3	Sentinel-2 Green band	"
		S2_B4	Sentinel-2 Red band	"
		S2_B5	Sentinel-2 Vegetation Red Edge band 1	"
		S2_B6	Sentinel-2 Vegetation Red Edge band 2	"
		S2_B7	Sentinel-2 Vegetation Red Edge band 3	"
		S2_B8	Sentinel-2 NIR band	"
		S2_B8a	Sentinel-2 Narrow NIR band	"
5	Spectral & Biophysical indices	FAPAR	Fraction of Absorbed Photosynthetically Active Radiation	SNAP
		FCOVER	Fraction of Green Cover Vegetation	"
		LAI	Leaf Area Index	"
		NDWI	Normalized Difference Water Index	Gao (1996)
		NDVI	Normalized Difference Vegetation Index	Rouse et al. (1974)
		NDVI_RE	NDVI of Red Edge band	Forkuor et al. (2020)
		S1_diffVHVV	Sentinel-1 VH and VV backscatter quotient	Laurin et al. (2018)
		S1_addVHVV	Sentinel-1 VH and VV backscatter product	"
		STVI1	Stress Related Vegetation Index 1	Thenkabail et al. (1994)
		STVI2	Stress Related Vegetation Index 2	"
		STVI3	Stress Related Vegetation Index 3	"
		VARI	Visible Atmospheric Resistance Index	Gitelson et al. (2002)
6	Texture	**S1_VH_text1**	Sentinel-1 VH GLCM texture - Contrast	Haralick et al. (1973)
		S1_VH_text2	Sentinel-1 VH GLCM texture - Dissimilarity	"
		S1_VH_text3	Sentinel-1 VH GLCM texture - Entropy	"
		S1_VH_text4	Sentinel-1 VH GLCM texture - Homogeneity	"

(*continued on next page*)

Table 1 (*continued*)

Sl. no	Category	Code	Variable description	Source
		S1_VH_text5	Sentinel-1 VH GLCM texture - Mean	"
		S1_VH_text6	Sentinel-1 VH GLCM texture - Second Moment	"
		S1_VH_text7	Sentinel-1 VH GLCM texture - Variance	"
		S1_VV_text1	Sentinel-1 VV GLCM texture - Contrast	"
		S1_VV_text2	Sentinel-1 VV GLCM texture - Dissimilarity	"
		S1_VV_text3	Sentinel-1 VV GLCM texture - Entropy	"
		S1_VV_text4	Sentinel-1 VV GLCM texture - Homogeneity	"
		S1_VV_text5	Sentinel-1 VV GLCM texture - Mean	"
		S1_VV_text6	Sentinel-1 VV GLCM texture - Second Moment	"
		S1_VV_text7	Sentinel-1 VV GLCM texture -Variance	"
		S2_text1_PC1	Principal component-1 of Sentinel-2 GLCM texture - Contrast	"
		S2_text1_PC2	Principal component-2 of Sentinel-2 GLCM texture - Contrast	"
		S2_text2_PC1	PC-1 of Sentinel-2 GLCM texture - Dissimilarity	"
		S2_text2_PC2	PC-2 of Sentinel-2 GLCM texture - Dissimilarity	"
		S2_text3_PC1	PC-1 of Sentinel-2 GLCM texture - Entropy	"
		S2_text3_PC2	PC-2 of Sentinel-2 GLCM texture - Entropy	"
		S2_text4_PC1	PC-1 of Sentinel-2 GLCM texture - Homogeneity	"
		S2_text4_PC2	PC-2 of Sentinel-2 GLCM texture - Homogeneity	"
		S2_text5_PC1	PC-1 of Sentinel-2 GLCM texture - Mean	"
		S2_text5_PC2	PC-2 of Sentinel-2 GLCM texture - Mean	"
		S2_text6_PC1	PC-1 of Sentinel-2 GLCM texture - Second Moment	"
		S2_text6_PC2	PC-2 of Sentinel-2 GLCM texture - Second Moment	"
		S2_text7_PC1	PC-1 of Sentinel-2 GLCM texture - Variance	"
		S2_text7_PC2	PC-2 of Sentinel-2 GLCM texture - Variance	"

Variables with Variance inflation factor (VIF) <10 is indicated in bold letters.

regression and classification (Fan et al., 2018). XGBoost combines the advantages of both bagging and boosting. It enhances model accuracy by creating new trees that correct the residual errors of the previous trees. Here, we implemented the XGBoost regressor from the "xgboost" R package (Chen et al., 2015). The tuning grid approach was used to identify the optimal parameter combination. After the tuning grid has been refined several times and a certain number of iterations have been completed, the parameter range with the best combination for prediction on the holdout test data was selected. The chosen parameters and their optimal values for tuning XGBoost models are: (1) nrounds: 50, (2) max_depth: 2, (3) min_child_weight: 1, (4) gamma: 0.1, (5) subsample: 1, (6) colsample_bytree: 0.8, and (7) eta: 0.1.

2.2.6. Model evaluation and validation

The complete dataset, comprising 170 ground-based AGB estimates, was first randomly divided into a training set (70%) and an independent validation set (30%). Model development was conducted exclusively using the training dataset. Within this subset, a 10-fold cross-validation (CV) procedure was implemented to optimize model parameters and reduce overfitting (Kuhn and Johnson, 2013). Specifically, the training dataset was divided into ten equal subsets (folds), and the model was trained and evaluated iteratively such that in each iteration one-fold was used for validation while the remaining nine folds were used for training. This approach ensured that model tuning was performed solely on the training data, while the independent 30% validation dataset was withheld entirely from the model development process to provide an unbiased assessment of predictive performance (Gholamy et al., 2018). The final optimized model was then applied to generate spatial predictions of AGB.

Model performance was evaluated using the independent validation dataset. For each validation plot, predicted AGB was extracted by averaging all predicted pixel values within the plot boundary, ensuring spatial correspondence between field measurements and remote sensing predictors. The resulting plot-level mean predictions were compared with observed field-based AGB to compute the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). The R^2 quantifies the proportion of variance in observed AGB explained by the model, whereas RMSE and MAE measure the magnitude of prediction errors, with RMSE being more sensitive to larger deviations. All validation metrics were calculated exclusively from the independent validation dataset to ensure an unbiased evaluation. The formulae for these metrics are provided in Table 2 and are widely used in comparable biomass modeling studies. Variable importance scores were calculated using the built-in 'varImp' function implemented in the modeling framework.

2.2.7. Uncertainty assessments

RF, being a non-parametric machine learning technique, does not allow for direct quantification of uncertainty (Coulston et al., 2016). Thus, a prediction interval utilizing quantile regression forest (QRF), as outlined by Meinshausen et al. (2006), was subsequently computed to estimate model uncertainty. Model uncertainty is referred to here as the standard deviation of the QRF conditional prediction distribution. Furthermore, a model sensitivity (standard deviation of predicted maps) was computed by executing ten realizations of the AGB prediction utilizing the RF model, as outlined by Khanal et al. (2023). The sum of the standard deviation and its ratio to the mean AGB is denoted as the per-cent error in this context. We used the "quantregForest" (Meinshausen and Schiesser, 2015) package in R to perform this procedure.

2.2.8. Upscaling of AGB and associated uncertainty

We followed two upscaling pathways described by Yu et al. (2022). In pathway 1, AGB maps were first estimated at high resolution (10 m) and then upscaled to 30 m, preserving fine-scale spatial heterogeneity and mitigating mixed-pixel effects. This approach also ensures that the 10 m predictor values can be aggregated to match the field plot sizes, reducing spatial mismatch between field measurements and remote sensing data. In pathway 2, earth observation variables were upscaled to 30 m prior to AGB estimation, which can lead to the loss of spatial detail and less reliable estimates (Yu et al., 2022). Our comparative evaluation indicated that pathway 1 consistently outperformed pathway 2 in capturing variability and reducing prediction uncertainty. Overall, pathway 1 represents a balance between accuracy, spatial consistency with field data, and computational efficiency for generating 30 m AGB

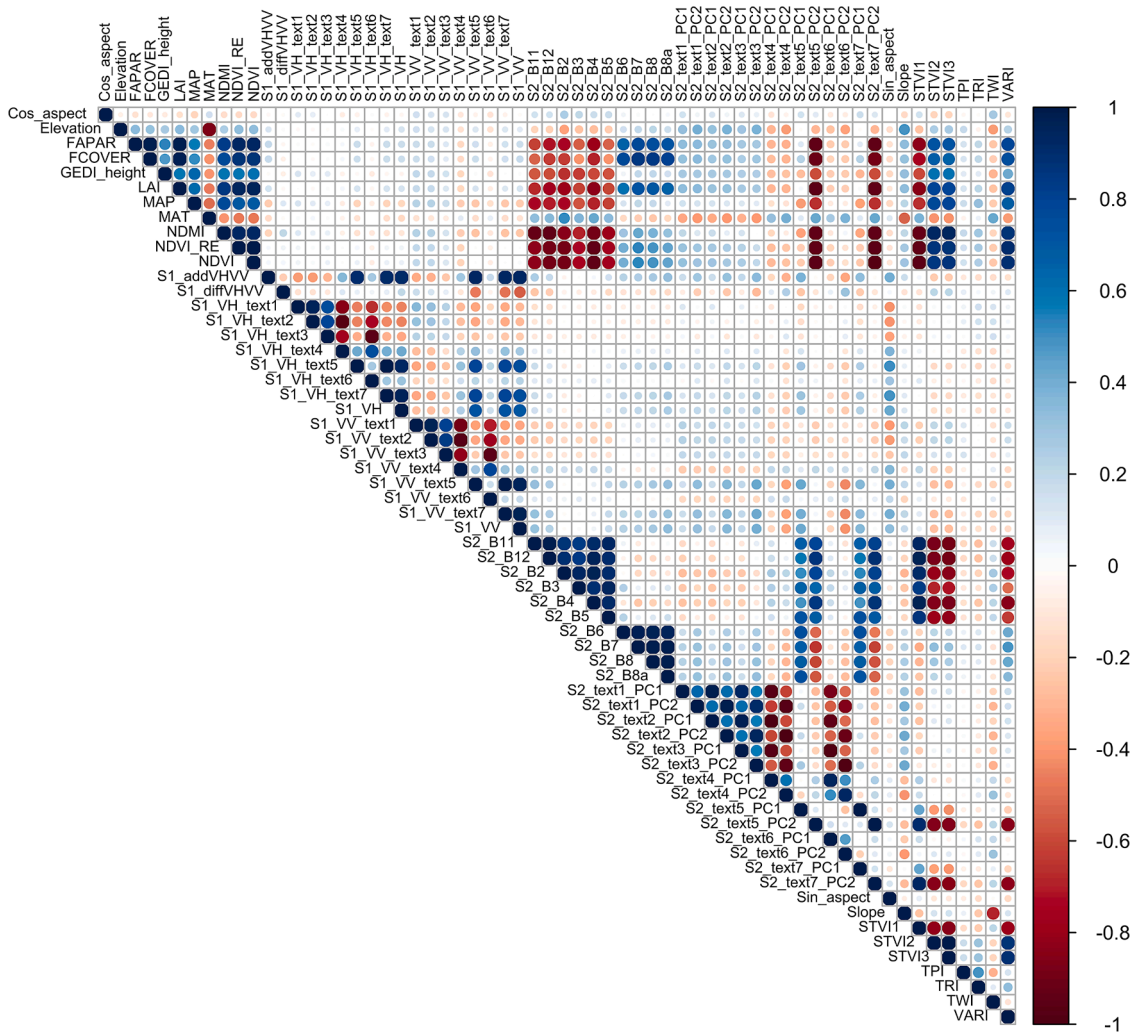

Fig. 3. Heatmap of the Pearson correlation matrix among predictor variables used for AGB modeling. Highly correlated predictors ($r > 0.70$) were filtered to minimize multicollinearity. See Table 1 for the description of predictor variables and Supplementary Table S1 for extended details on variable units, spatial resolution, and preprocessing.

Table 2
List of validation parameters used for testing the performance of machine learning models.

Validation Parameter	Formula
Coefficient of determination (R^2)	$\left(\frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2}\ \sqrt{\sum (y_i - \overline{y})^2}}\right)^2$
RMSE	$\frac{\sqrt{\sum (x_i - y_i)^2}}{n}$
MAE	$\frac{\sum (x_i - y_i)}{n}$

maps.

2.2.9. Principal component and Pearson's correlation analysis

To further interpret the relationships between the most influential predictors and AGB across forest types, PCA was performed on the top 20 variables ranked by Random Forest importance. PCA was conducted in R using the FactoMineR package (Lê et al., 2008). In addition, Pearson's correlation analysis was carried out between AGB and each of the top 20 predictors to quantify pairwise relationships. The normality of predictor variables was assessed using the Shapiro–Wilk test, and variables showing significant skewness were log-transformed to approximate normality prior to correlation analysis. These exploratory analyses were conducted to facilitate the interpretation of predictor relationships and were performed independently of the model training process.

2.2.10. Residual diagnostics and assessment of spatial robustness

To evaluate whether model relationships were stable across environmental gradients and geographic space, we conducted a comprehensive residual diagnostic analysis using the independent validation dataset. Residuals were computed as Observed − Predicted AGB. Systematic bias across the biomass range was assessed by plotting residuals against observed AGB values. To test for monotonic associations between residuals and key predictors (e.g., elevation, MAP, and GEDI canopy height), Spearman's rank correlation coefficients (ρ) were calculated, and p-values were adjusted for multiple comparisons using the Benjamini–Hochberg method (Haynes, 2013). Further, to evaluate whether model performance degraded within specific environmental ranges, each continuous predictor was divided into terciles (low, medium, high), and error metrics (RMSE, MAE, and mean bias) were computed within each group. Spatial autocorrelation of residuals was evaluated using Global Moran's I with k-nearest neighbour weights. Finally, differences in residual distributions among forest types (DDF, MDF, PLN, SEF) were evaluated using a Kruskal–Wallis test followed by Dunn's post-hoc comparisons with Benjamini–Hochberg correction.

Together, these complementary tests allowed assessment of gradient-driven bias, consistency of prediction accuracy, spatial clustering of residuals, and vegetation-type effects on model performance.

3. Results

3.1. Descriptive statistics of field-derived biomass

The field estimated AGB and vegetation characteristics for the total landscape as well as for each vegetation type are listed in Table 3. The tree species richness, density, mean DBH, minimum, and maximum AGB varied across the forest types. Field estimated AGB for the total landscape was in the range of 9.03 to 711.56 Mg/ha with a mean value (±SE) of 258.81 ± 10.76 Mg/ha. The coefficient variation of AGB was 4.15%. Out of the four major vegetation types, highest mean AGB value was scored by SEF (342.36 ± 31.56 Mg/ha) followed by MDF (322.1 ± 15.63 Mg/ha), PLN (221.21 ± 10.97 Mg/ha), and DDF (93.96 ± 11.27 Mg/ha). Of the 175 tree species recorded from the study site, the top 10 species contributed 61.32% of the total AGB. These species in the order of high to low AGB contribution include *Terminalia paniculata, Tectona grandis, Pinus roxburghii, Xylia xylocarpa, Lagerstroemia microcarpa, Terminalia elliptica, Dalbergia latifolia, Terminalia bellirica, Eucalyptus grandis,* and *Grewia tiliifolia*. Species such as *T. grandis, P. roxburghii, E. grandis* belonged to PLN, and *D. latifolia, T. elliptica,* and *G. tiliifolia* were part of MDF. Whereas *T. paniculata, T. bellirica*, L. *microcarpa*, and *X. xylocarpa* shared both MDF and SEF in the landscape.

3.2. Comparison of model performance

Table 4 presents the assessment of predictive performance for the three machine learning models. Among them, Random Forest (RF) demonstrated the best overall predictive performance, with an RMSE of 75.34 Mg/ha, MAE of 57.79 Mg/ha, and R^2 of 0.55. In comparison, SVR and XGBoost showed higher prediction errors (RMSE = 85.45 and 99.25 Mg/ha, respectively) and lower explained variance (R^2 = 0.44 and 0.40, respectively). Based on the combined evaluation of explained variance and error magnitude, RF was selected for subsequent AGB prediction and mapping. The Taylor diagram (Fig. 4) provides a complementary assessment of model behaviour by jointly comparing correlation, centered RMSE, and standard deviation relative to observed values. Although all models exhibited relatively high correlation coefficients, RF showed the most favourable balance between lower centered RMSE and variability, indicating more stable predictions across the biomass range. The scatter plot of observed versus predicted AGB for the independent validation dataset further illustrates model performance (Fig. 5). The RF model explained 55% of the variance in observed AGB (R^2 = 0.55), while the RMSE (75.34 Mg/ha) and MAE (57.79 Mg/ha) indicate the average magnitude of prediction errors. The moderate difference between RMSE and MAE suggests that although some larger deviations occur, extreme outliers are limited. Deviations from the 1:1 line were more pronounced at higher AGB values, indicating greater uncertainty. Overall, these results demonstrate moderate predictive performance, with RF providing the most consistent trade-off between explained variance and error magnitude among the evaluated models.

Table 3
Descriptive summary of major vegetation types, their characteristics and biomass estimates in Shettihalli landscape.

Vegetation type /Characteristics	DDF	MDF	SEF	PLN	Landscape
Area (ha)	2.7	7.1	2.2	5	17
No. of 0.1 ha plots	27	71	22	50	170
No. of trees	972	2689	967	3492	8120
Mean DBH (cm)	20.17 ±1.51	29.20 ±0.87	28.32 ±1.77	24.64 ±0.70	26.31 ±0.59
Mean Height (m)	9.43 ±0.36	17.99 ±0.34	17.35 ±0.49	14.19 ±0.78	15.43 ±0.37
Tree density (stems/ha)	360	378.73	439.55	698.4	477.65
Total species richness	63	110	117	38	175
Minimum AGB (Mg/ha)	9.03	95.53	126.76	41.01	9.03
Maximum AGB (Mg/ha)	284.74	686.31	711.56	382.5	711.56
Mean AGB (Mg/ha)	93.96 ±11.27	322.10 ±15.63	342.36 ±31.56	221.21 ±10.97	258.81 ±10.76

DDF – Dry deciduous forest, MDF – Mosit deciduous forest, SEF – Semi-evergreen forest, PLN – Plantation

Table 4
Performance metrics of three machine learning algorithms evaluated with independent test data for AGB estimation.

Model	Coefficient of determination (R^2)	RMSE (Mg/ha)	MAE (Mg/ha)
Random Forest (RF)	0.55	75.34	57.79
Support Vector Regression (SVR)	0.44	85.45	64.79
Extreme Gradient Boosting (XGBoost)	0.40	99.25	75.15

R^2 – Relationship between the observed and predicted; RMSE – Root mean square error, MAE – Mean absolute error

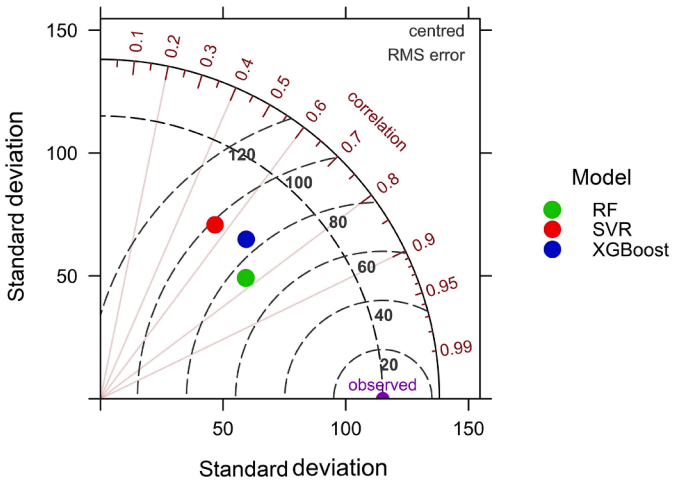

Fig. 4. Comparison of the performance of three machine learning algorithms using a Taylor diagram. Abbreviations: RF – Random Forest, SVR – Support Vector Regression and XGBoost – Extreme Gradient Boosting.

3.3. Importance of predictor variables

The analysis of relative importance of variables in predicting AGB indicated texture variables as the most important (with relative importance of 34.9%), followed by the physical (25.1%), spectral and biophysical (23.3%), and climate variables at 16.7% (Fig. 6). These findings indicate that, within the combined modeling framework, texture variables contributed more to AGB prediction than spectral, biophysical, and physical variables. Mean annual precipitation (MAP) and temperature (MAT), GEDI canopy height, and NIR band of S2 were found to be the most influential predictor variables in the analysis. The top twenty predictor variables in RF included texture indices such as the mean, second moment, and contrast of the S2 principal component, as well as the variance and mean of the S1 VV and VH, respectively. These indices demonstrate the relevance of texture analysis in capturing vegetation structure. Physical variables, such as elevation, slope, and sine aspect, contributed meaningfully to the model, highlighting the impact of topographical features on biomass distribution. Further, the analysis

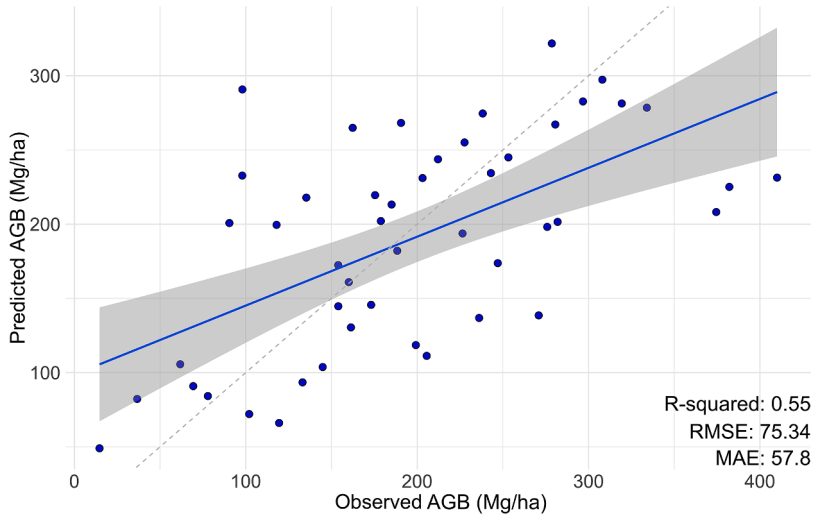

Fig. 5. Relationship between observed and predicted AGB for the independent validation dataset (30% of plots). Each point represents plot-averaged predicted AGB from 10 m pixels within the plot. Blue line represents the fitted line with 95% CI and grey dotted line represents 1:1 line.

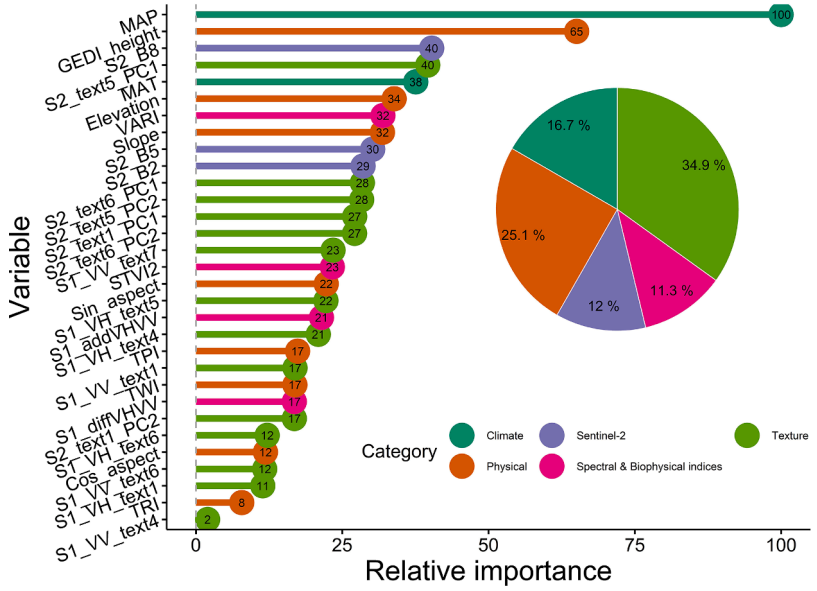

Fig. 6. Variable importance plot of explanatory variables for AGB based on the Random Forest model. The enclosed pie chart depicts the percentage contribution of variables grouped into six broad categories. Refer to Table 1 for variable description.

also revealed that the red edge and blue band of S2 and the VARI vegetation index are among the top predictor values, emphasizing the value of spectral information in AGB modeling. Overall, the results highlighted the critical role of texture variables in predicting the tropical forest biomass with substantial contributions from spectral, physical, biophysical, and climatic variables.

The principal component analysis (PCA) that included the modeled AGB revealed a distinct cluster of vegetation types determined by the top 20 covariates (Fig. 7). GEDI height, MAP, VARI, and STVI2 vegetation indices, elevation, slope, and S2 texture-contrast were found to be positively associated with AGB. Conversely, variables such as S2 blue and red-edge bands, and MAT showed a negative relationship with AGB. These biplots highlighted the multidimensional relationships between environmental variables, remote sensing indices, and AGB, underscoring the complex interplay of factors influencing biomass distribution in diverse forest types, which was evident in the clearly distinguishable patterns seen in the final predicted map. In addition, a Pearson correlation analysis was performed to examine the relationship between the top 20 predictor factors and AGB, where the GEDI canopy height, MAP, and elevation showed a strong and significant positive association, while S2 blue and red-edge bands exhibited a negative and significant correlation, affirming the results obtained by the PCA (Fig. 8).

3.4. Spatial prediction of AGB and associated uncertainty

The RF algorithm was used to construct the spatial distribution of AGB (Fig. 9) at a 10 m resolution and upscaled to 30 m. The predicted AGB ranged from 16.64 to 375.73 Mg/ha, with a mean of 215.37 Mg/ha from the RF model, which was close to the field-based mean estimates (Tables 3 and 4), reinforcing the RF model's reliability in capturing the AGB's central tendency. The map depicted a significant variation in the distribution of AGB across the forest types. The central region, dominated by SEFs, exhibited the highest AGB values, ranging from 173.87 to 375.73 Mg/ha, suggesting a dense, species-rich ecosystem with substantial biomass. Conversely, the northeastern region, characterized by DDFs, showed the lowest AGB values, mostly between 16.64 and 163.09

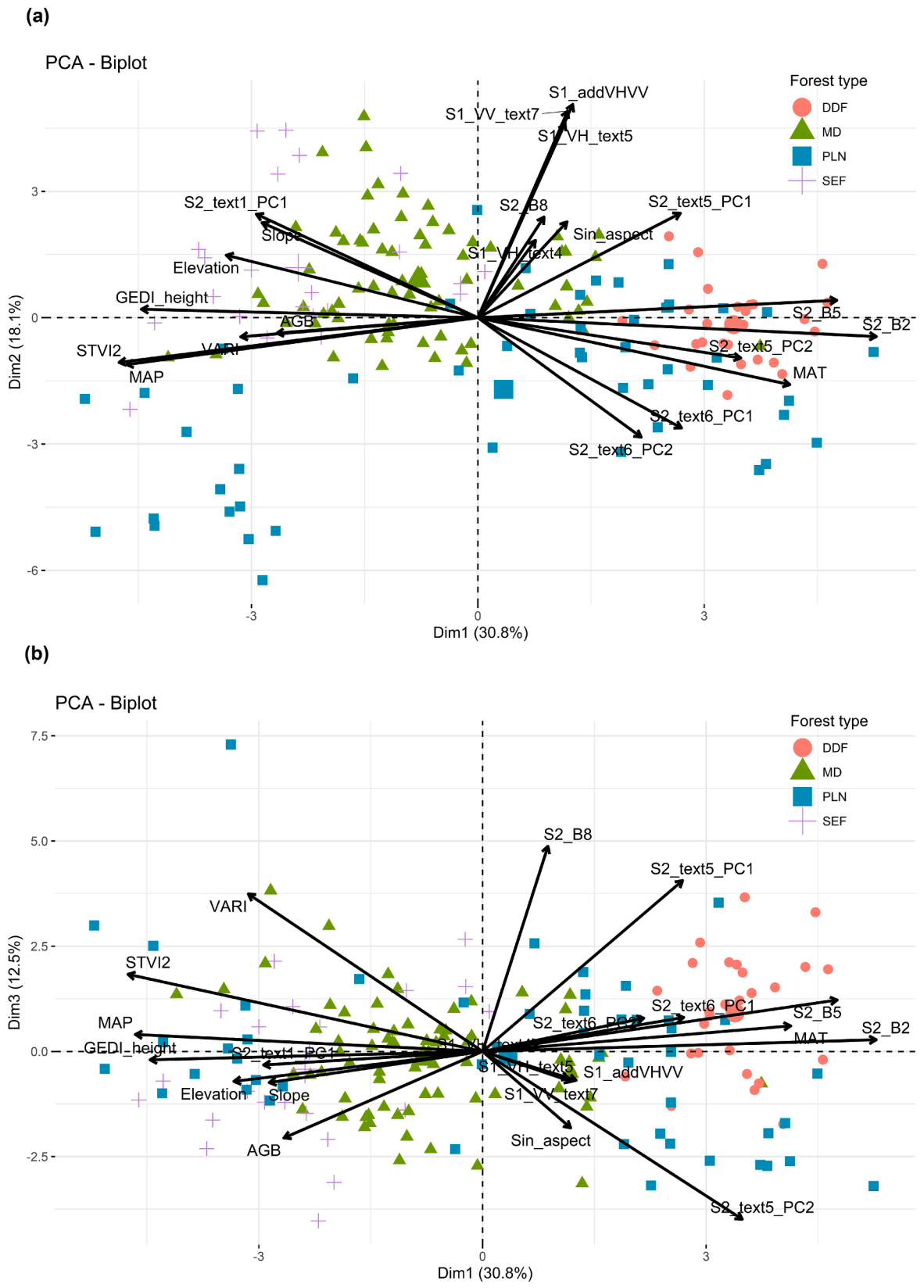

Fig. 7. Principal Component Analysis (PCA) biplots with combinations "a" PCA-1:PCA-2 and "b" PCA-1:PCA-3, with the AGB modelled and top 20 important variables representing climate, physical, spectral bands, spectral & biophysical indices and texture features.

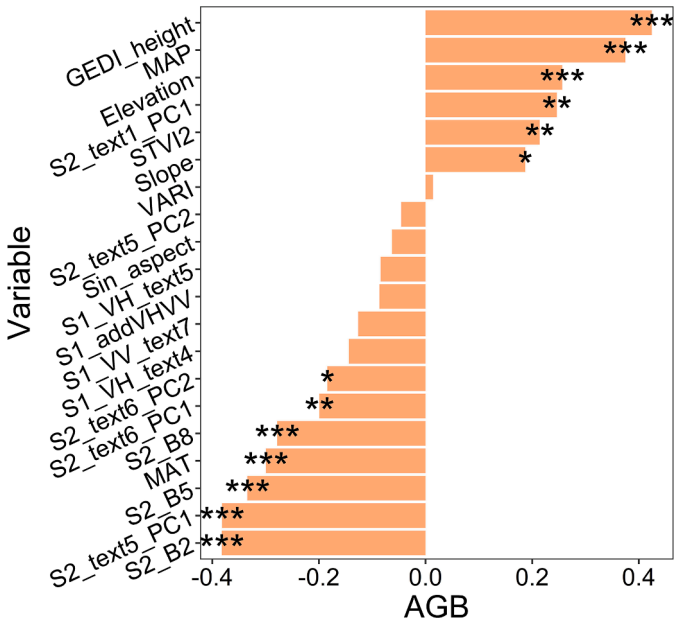

Fig. 8. Pearson correlation between the AGB and the most important explanatory variables. Significant (*p*-value) relationships are indicated as follows: * = $p < 0.05$, ** = $p < 0.01$ and *** = $p < 0.001$; while non-significant relationships are not shown in symbols.

Mg/ha, reflecting sparse vegetation and a lower biomass score. These forests are characterized by *Terminalia anogeissiana- Terminalia elliptica-Tectona grandis community* and *Eucalyptus*-dominated patches which are in the northwest region of Shettihalli. Regions that have undergone human modification, such as plantations, exhibit moderate AGB levels, typically ranging from 105.36 to 212.38 Mg/ha. This intermediate biomass value indicates that these areas have moderate vegetation cover, but they do not match the AGB of the SEFs. The variability in AGB across Shettihalli highlights the impact of vegetation type and land use on biomass distribution, with SEFs being the most biomass-rich, DDFs being the least, and MDF and PLN areas falling in between. Fig. 10 showcases the uncertainty map of estimated AGB. The uncertainty values ranged from 11.78 to 121.01 Mg/ha across the landscape, with lower uncertainty in DDF and higher in SEF. On the other hand, MDF and PLN forests had moderate to high uncertainty. The mean uncertainty in relation to the mean anticipated biomass was 35.4%.

3.5. *Residual behaviour and spatial stability of model predictions*

To evaluate whether prediction errors varied systematically across environmental gradients and geographic space, we conducted a comprehensive residual analysis using the independent validation plots (n = 48). Residuals were centered near zero and exhibited no significant monotonic association with elevation (ρ = 0.127), GEDI canopy height (ρ = -0.073), or mean annual precipitation (ρ = 0.097) after multiple comparison correction (*p*_adj = 0.62; Table S2). Residual plots showed no clear directional trends across structural, climatic, or topographic gradients (Fig. S1 a – c, Supplementary Information). To further assess

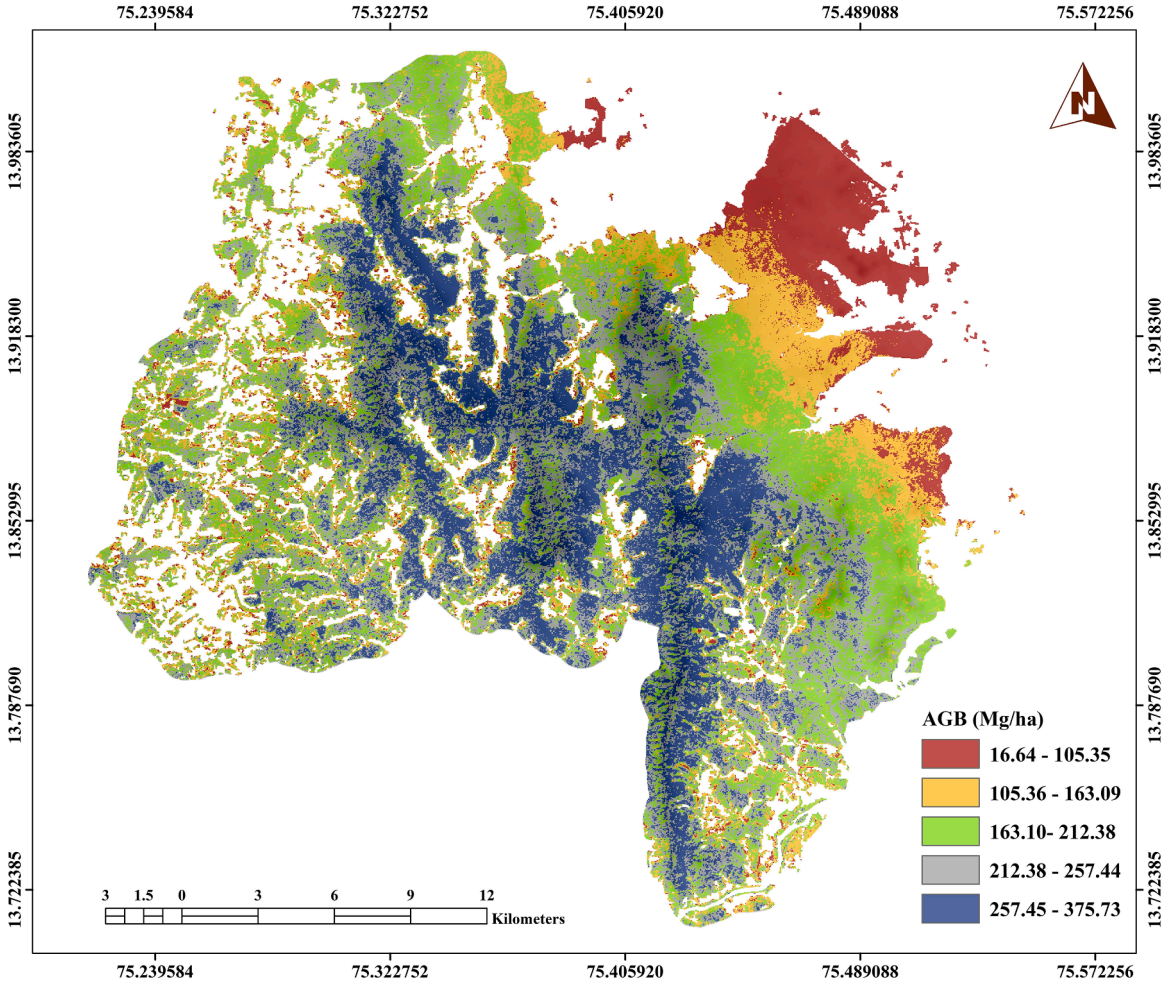

Fig. 9. Spatial prediction of AGB estimates obtained from the RF model based on the best performance.

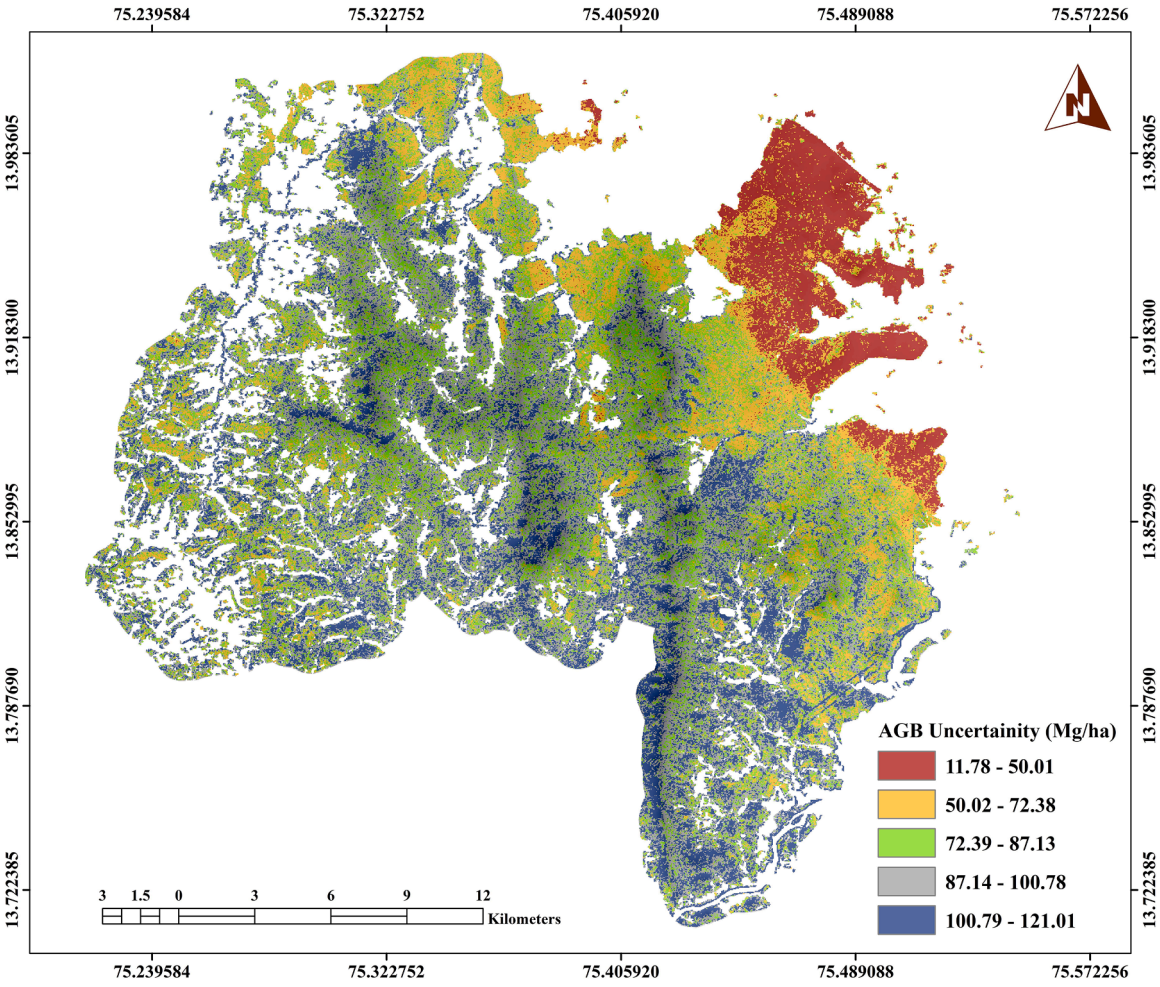

Fig. 10. The prediction uncertainty of forest AGB obtained from Quantile Regression Forest.

model stability across environmental ranges, each continuous predictor was divided into terciles (low, medium, high), and error metrics were computed within each group (Table S3). RMSE values remained broadly consistent across elevation (71–81 Mg/ha), while a noticeable variation was observed across MAP (57–85 Mg/ha), and canopy height (55–87 Mg/ha) terciles. Although a moderate positive mean bias (+30 Mg/ha) was observed in the highest elevation tercile, no consistent increase in error toward environmental extremes was evident.

Spatial analysis indicated weak and non-significant autocorrelation of residuals (Moran's I = 0.072, $p = 0.12$). The spatial distribution of residuals showed no geographic clustering of over- or underestimation (Fig. S2). Residuals were generally centered around zero but exhibited greater dispersion at mid to high predicted AGB values (Fig. S3). Both over- and under-estimation were observed at higher biomass levels, with several extreme residuals indicating increased prediction variability, which may reflect potential signal saturation effects of SAR and optical data or limited sample representation in high-biomass regions. Finally, residual distributions differed modestly among forest types (Kruskal–Wallis $\chi^2 = 9.20$, $p = 0.027$). Dunn's post-hoc test identified a significant contrast between DDF and PLN plots ($p = 0.028$), while other pairwise differences were not significant (Fig. S4). The observed difference between the two forest types is likely due to the structural differences in canopy composition rather than the spatial instability of the model.

4. Discussion

Tropical forest landscapes are characterized by significant spatial heterogeneity, leading to increased spatial uncertainty in estimating AGB. This study found moderate accuracy in estimating AGB in the heterogeneous landscape of the central Western Ghats, India. The model revealed a spatial variability in AGB across the landscape with a strong R^2 of 0.55, RMSE of 75.34 Mg/ha, MAE of 57.79 Mg/ha, and an uncertainty level of 34.6%. Our results are in line with the findings from other research works carried out in tropical forests (Kelsey and Neff, 2014; Ghosh and Behera, 2018; Fararoda et al., 2021; Li et al., 2020; Behera et al., 2024; Ayushi et al., 2024; Sainuddin et al., 2024). The study findings also endorse the use of a non-parametric and ML model for assessing AGB in tropical landscapes, notwithstanding its limitations. Parametric models, e.g., linear models (LMs) and generalized linear models (GLMs), do not consider nonlinear effects of predictors, which is often observed in ecological scenarios like AGB estimates (Ghosh and Behera, 2018; Turton et al., 2022). In contrast, ML techniques such as RF, SVM/SVR, XGBoost, and Artificial Neural Network (ANN) can estimate AGB with significant accuracy compared to linear modeling (Zhang et al., 2014; Ahmed et al., 2015; Gao et al., 2018). It suggests non-parametric methods can effectively handle the intricate interaction between AGB and remote sensing covariates. Further, the capacity to manage non-linearity and assess the significance of independent variables makes RF a powerful algorithm (Ghosh and Behera, 2018; Dang et al., 2019; Han et al., 2021; Nandy et al., 2021; Behera et al., 2024; Ayushi et al., 2024; Zhang et al., 2024), as also evidenced by our results that we estimated a better accuracy as compared to SVR (R^2 of 0.44, RMSE of 85.45 Mg/ha, MAE of 64.79 Mg/ha) and XGBoost (R^2 of 0.40, RMSE of 99.25 Mg/ha, MAE of 75.15 Mg/ha). Ghosh and Behera (2018) observed that RF outperformed SGB in predicting AGB in a tropical

landscape of India. In the Western Ghats, India, Ayushi et al. (2024) noted that RF predicted superior AGB than gradient boosting and SVM. Similarly, Han et al. (2021) reported higher prediction accuracy for the RF model than for the SVM in the Dabie Mountain region of China. In contrast, Li et al. (2020) and Sainuddin et al. (2024) found that XGBoost outperformed RF in predicting AGB. However, both models outperformed linear regression in terms of prediction accuracy.

When dealing with high-dimensional data, the abundance of irrelevant or redundant characteristics can complicate feature selection for RF. As the number of dimensions grows, it becomes more difficult to interpret the predictor variables that are genuinely important for the RF model's predictive ability. We conducted a feature selection procedure by conducting a multi-collinearity test and eliminating strongly correlated variables (Fig. 3 and Table 1), reducing the influence of irrelevant features. By using this selection approach, RF can create a more focused dataset that emphasizes the most valuable predictors, improving both the model's effectiveness and its clarity. Moreover, chosen variables were successful in distinguishing between different forest types in the landscape when subjected to a principal component analysis (Fig. 7).

Data fusion is a crucial process for enhancing the accuracy of AGB estimation. Previous research has demonstrated that combining various remote sensing sources led to substantial improvement in AGB prediction and exceeded the saturation threshold of the estimation (Hu et al., 2016; Viet Nguyen et al., 2016; Li et al., 2020; Hernández-Stefanoni et al., 2020; Nandy et al., 2021; Han et al., 2021). Each data type has unique benefits and limitations, and combining various data sources can help mitigate their deficiencies. Several studies have also attempted to determine SAR backscatter-biomass connections in forests and identified a saturation level of around 150 Mg/ha (Joshi et al., 2015; Rodríguez-Veiga et al., 2017; Kumar et al., 2019; Fararoda et al., 2021). Comparable outcomes were also observed when employing vegetation indices and spectral bands derived from optical data that become saturated at higher biomass (Zhao et al., 2016; Mutanga et al., 2023). However, high-resolution texture values derived from both SAR and optical images have exceeded the limit of biomass saturation in tropics (Ploton et al., 2017; Kelsey and Neff, 2014; Berninger et al., 2018). Thus, these variables are commonly combined to generate maps at various scales, ranging from global to local. In addition to remote sensing data, ancillary data sources such as DEMs, land use maps, climate factors, and vegetation type were also used to estimate AGB (Fararoda et al., 2021; Padalia et al., 2023; Behera et al., 2024; Ayushi et al., 2024).

The current study found that a combination of physical, climatic, SAR backscatter, spectral bands, vegetation indices, and textural characteristics produced improved accuracy in biomass estimation. Texture variables had the highest relative importance at 34.9%, followed by physical at 25.1%, spectral, and vegetation indices at 23.3% and climate at 16.7% (Fig. 6). The study's results indicate that textural factors from S2 and S1 sensors are more effective in predicting AGB in a heterogenous mixed tropical forest landscape with high canopy density and significant variability such as the SEF and MDF, aligning well with previous research (Dang et al., 2019; Behera et al., 2024; Sainuddin et al., 2024; Ayushi et al., 2024). In the similar tropical forests of Eastern Ghats, India, Behera et al. (2024) reported that textural factors from S2 data had a significant impact on estimating AGB with a relative relevance of 83.34%. Similarly, Dang et al. (2019) found that texture parameters obtained from S2 satellite imagery outperformed spectral parameters in accurately predicting biomass in Vietnam forests. Therefore, it is recommended to use high spatial resolution data, such as S2, to get better prediction as finer resolution captures spatial heterogeneity effectively compared to coarse resolution (Asner et al., 2014; Ploton et al., 2017; Jha et al., 2021). We found a valuable worldwide canopy height dataset created by Potapov et al. (2021) that greatly benefitted our modeling efforts and showed high positive correlation with AGB in the landscape (Fig. 8). Concordant with our observation, Nandy et al. (2021) concluded that the forest canopy height derived from the high-resolution data greatly enhanced the AGB prediction in tropical-sub-tropical forests of Uttarakhand, India, dominated by *Shorea roxburghii*. Additionally, research by Frank et al. (2018) emphasized the significance of incorporating tree height into models to more accurately capture variations across different regions.

We also found that terrain and climate variables account for a sizable percentage of the explanation especially MAP, MAT and elevation (Fig. 6). MAP also showed a strong positive correlation with AGB, emphasizing that higher precipitation supports greater biomass accumulation (Fig. 8). This finding aligns with prior research that underscores the influence of precipitation on AGB (Babu et al., 2023a; Bulut, 2023; Ayushi et al., 2024). Precipitation also plays a critical role in determining water availability, which in turn affects key factors such as stomatal conductance, nutrient uptake, and the overall productivity of forests (Eamus, 2003). Further, elevation positively influenced AGB, indicating that higher altitudes are associated with more biomass, possibly due to favorable climatic conditions. Moreover, elevation is a crucial predictor variable that significantly influences the distribution of AGB by impacting essential factors such as temperature, precipitation, and soil nutrient availability, which are essential for forest growth (Chen et al., 2019; Guerra-Hernández et al., 2022) and provides critical insights into forest distribution and the characteristics of different sites (Wang et al., 2021). MAT was another significant negative predictor, indicating that higher temperatures may reduce biomass, possibly due to increased evapotranspiration and stress on vegetation. Overall, at the landscape level, topographical characteristics are considered important determinants of biomass and forest structure (Salinas-Melgoza et al., 2018; Muscarella et al., 2020; Jia et al., 2022). Restricting the availability of nutrients and water in a landscape, it serves as an environmental filter, influencing the growth and survival rates of trees (Rodrigues et al., 2020). Mitchard et al. (2012) recommended including DEM as an extra layer when using SAR data to estimate AGB in order to minimize the impact of slope on radar backscatter. Furthermore, Chen et al. (2019) have demonstrated that variables derived from SRTM data were essential for estimating forest biomass and had a considerable impact on the spatial distribution of AGB.

An important practical consideration in predictive ecological modeling is whether a reduced set of predictors can maintain comparable performance while improving model interpretability and computational efficiency. In this study, we applied strict multicollinearity control ($r \geq 0.7$; VIF ≥ 10), reducing the initial predictor pool by 50% (62 to 31 variables). This procedure ensured statistical stability and minimized redundancy-driven inflation of variable importance. Nevertheless, multicollinearity filtering alone does not identify the smallest subset of predictors that maximizes predictive efficiency. Our variable-importance analysis indicates that a relatively small subset of key predictors (e.g., GEDI canopy height, MAP, Elevation, S2 bands, and texture metrics) accounts for a substantial proportion of the model's explanatory power. However, we acknowledge that the present study did not conduct a systematic feature selection analysis (e.g., recursive feature elimination) to identify the strict minimal predictor set. Such analyses would allow explicit quantification of the marginal contribution of secondary variables and evaluate how their omission influences predictive performance and the spatial distribution of model uncertainty. Therefore, while we recommend the highest-ranked predictors as priority inputs for the modeling framework, determining the optimal minimal subset and its impact on predictive uncertainty warrants systematic evaluation and represents a promising avenue for advancing a cost-effective approach for large-scale biomass modeling frameworks.

Uncertainty quantification, along with prediction mapping, offers valuable information on regional uncertainty fluctuations, enhancing decision-making in forest management (Dang et al., 2019; Padalia et al., 2023). The study utilized combinations of standard deviations derived from simulated RF predictions and QRF to assess the uncertainty of model estimations. While we estimated spatial prediction of AGB distribution in the range of 16.64 Mg/ha to 375.73 Mg/ha in the landscape (Fig. 9), the uncertainty range was observed in the range of 11.78 Mg/ha

to 121.01 Mg/ha (Fig. 10). The spatial pattern of uncertainty in AGB followed the mean predictions. In other words, High mean AGB was reported in the central region of the landscape dominated by SEF, and low mean AGB was reported in the north-eastern region dominated by DDF. A high level of uncertainty was observed for both SEF and DDF, as RF inevitably overestimates at lower AGB levels and underestimates at higher AGB levels due to its failure to incorporate spatial patterns during the modeling process (Turton et al., 2022). The other potential reasons for the increased uncertainty may include a comparatively low number of sample plots in both SEF and DDF patches. In addition, SAR interactions are disrupted in areas with intricate topography, sparse vegetation, and complex forest structure (Joshi et al., 2015; Padalia et al., 2023). These factors can affect biomass measurements and introduce more uncertainty in estimates, as observed in SEF and DDF.

This study's advantage lies in the random scattering of sample plots throughout the landscape (Fig. 1), resulting in a dataset that improves the representativeness of biomass variability. This is crucial for creating models that generalize well to the entire area of interest. Larger sampling units, such as 170 samples in this case, reduce the relative sampling error by including a wider range of variance within each unit. This leads to more precise calculations of mean biomass. We set up 0.1-hectare plots for field-based AGB estimation, a method commonly used for national-level biomass measurement (Rajashekar et al., 2018). Nevertheless, we acknowledge that a larger plot size will reduce the uncertainty of AGB estimation. Réjou-Méchain et al. (2019) demonstrated that AGB variability decreased by approximately threefold when the plot size increased from 0.1 ha to 1 ha. Similarly, Fararoda et al. (2024) reported that increasing plot size from 0.04 ha to 0.49 ha reduced relative error from approximately 22% to 5%. Consistent with these findings, several studies have shown that increasing plot size enhances explanatory power and reduces uncertainty in biomass estimation (Knapp et al., 2022; Mayamanikandan et al., 2022; Rodda et al., 2024). Additionally, the biomass validation protocol developed by the Committee on Earth Observation Satellites (CEOS) Land Product Validation Subgroup (Working Group on Calibration and Validation, 2021) recommends the establishment of large permanent plots (preferably >0.25 ha and ideally 1 ha) for field-based AGB estimation.

The allometric model and height-diameter (H–D) relationships are crucial factors in error propagation (Réjou-Méchain et al., 2017; Fararoda et al., 2024). Previous studies have shown that allometric equations alone can contribute approximately 5% uncertainty in 1-ha tropical forest plots (Chave et al., 2003). In this study, field AGB was estimated using the widely adopted pantropical allometric model proposed by Chave et al. (2014). In addition, allometric model uncertainty was quantified following the approach of Réjou-Méchain et al. (2017), which is considered a standard procedure for estimating field AGB uncertainty. Nevertheless, the use of locally calibrated H–D models is recommended when available to enhance the model accuracy and predictions. Recently, Fararoda et al. (2024) developed forest type-specific H–D models for Indian forests that outperformed several global models. Incorporating such locally developed H–D models may further reduce uncertainty in biomass estimation and is therefore recommended for future studies. Furthermore, integrating field measurements with airborne or terrestrial LiDAR has been shown to substantially improve AGB estimation and may serve as a benchmark for validating global biomass products (Santoro et al., 2021; Jha et al., 2020). Several global AGB products are now publicly available (Ploton et al., 2020a; Rodda et al., 2024), and multiple studies have attempted their validation (Ploton et al., 2020b; Bhat et al., 2024). The AGB exhibits spatial autocorrelation (Guitet et al., 2015), which is also evident in our datasets (Fig. 1). As noted, ML techniques do not account for spatial autocorrelation other than that caused by predictors, thereby introducing bias in the estimation process. Numerous modified versions of RFs are being produced, but there are inherent problems observed with model transferability and spatial predictions (Turton et al., 2022; Zhang et al., 2024). Given the situation, generalized hierarchical models offer a useful foundation for modeling and uncertainty analysis (Saarela et al., 2018) when spatial dependence is present; therefore, they are recommended for future work.

5. Conclusions

This study evaluated the performance of three ML models for estimating AGB across a heterogeneous tropical forest landscape in the central Western Ghats of India by integrating field measurements with multi-source Earth observation data. Among the tested models, RF demonstrated superior predictive performance compared to SVR and XGBoost, while QRF proved effective for quantifying prediction uncertainty. The results underscore the importance of robust variable selection and model construction in improving biomass estimation accuracy. Key predictors included S2 spectral bands and textures, elevation, GEDI-derived canopy height, and climatic variables (MAP and MAT), particularly in structurally diverse forest types such as DDF, MDF, and SEF. The integration of multi-source remote sensing data significantly enhanced model performance in this complex landscape and is therefore recommended for future studies focusing on biomass modeling frameworks. Our findings further highlight the importance of adequate field sample size for improving model reliability and reducing estimation uncertainty. The mapped AGB distribution aligned well with field-based estimates and indicated higher biomass values than national-level estimates (ISFR, 2022), emphasizing the critical role of Shettihalli's forests in carbon sequestration and climate change mitigation. Overall, this study contributes to the understanding of AGB estimation and provides valuable insights for mapping and monitoring forest biomass. The proposed approach can be extended to broader regions and supports carbon stock assessment and sustainable forest management under REDD+ initiatives. The derived AGB map may serve as a valuable tool for pinpointing areas of high conservation importance and guiding restoration efforts in low-biomass and degraded regions, thereby enhancing forest resilience and carbon offset potential.

Funding

This work was supported by the Department of Biotechnology under grant No.BT/Coord.II/10/02/2016/22.03.2018. The first author thanks the Indian Council of Social Science Research for providing a short-term doctoral fellowship (RFD/Short-Term/2022–23/ENV/ST/66).

CRediT authorship contribution statement

Naveen Babu Kanda: Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Debabrata Behera:** Writing – original draft, Validation, Methodology, Formal analysis. **Rahul Gour:** Writing – review & editing, Validation, Software, Methodology, Formal analysis. **Kurian Ayushi:** Writing – review & editing, Visualization, Validation. **Ayyappan Narayanan:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition. **Narayanaswamy Parthasarathy:** Writing – review & editing, Supervision, Investigation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ayyappan Narayanan reports financial support was provided by Department of Biotechnology, Government of India. Naveen Babu Kanda reports financial support was provided by Indian Council of Social Science Research. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was carried out as part of a project, "Biodiversity characterization at community level in India using Earth observation data". We thank the Karnataka Forest Department for granting the necessary permissions for project execution. We also thank USGS for providing the earth observation data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tfp.2026.101271.

Data availability

The data used to analyze in the study are available from the corresponding author upon request.

References

Ahmed, O.S., Franklin, S.E., Wulder, M.A., White, J.C., 2015. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. ISPRS J. Photogramm. Remote Sens. 101, 89–101.

Asner, G.P., Knapp, D.E., Martin, R.E., Tupayachi, R., Anderson, C.B., Mascaro, J., Sinca, F., Chadwick, K.D., Higgins, M., Farfan, W., Llactayo, W., Silman, M.R., 2014. Targeted carbon conservation at national scales with high-resolution monitoring. Proc. Natl. Acad. Sci. 111.

Ayushi, K., Babu, K.N., Ayyappan, N., Nair, J.R., Kakkara, A., Reddy, C.S., 2024. A comparative analysis of machine learning techniques for aboveground biomass estimation: A case study of the Western Ghats. India. Ecol. Inform. 80, 102479.

Babu, K.N., Gour, R., Ayushi, K., Ayyappan, N., Parthasarathy, N., 2023b. Environmental drivers and spatial prediction of forest fires in the Western Ghats biodiversity hotspot, India: an ensemble machine learning approach. For. Ecol. Manag. 540, 121057.

Babu, K.N., Mandyam, S., Jetty, S., Dar, A.A., Ayushi, K., Narayanan, A., Somaiah, S., Narayanaswamy, P., 2023a. Carbon stocks of tree plantations in a Western Ghats landscape, India: influencing factors and management implications. Environ. Monit. Assess. 195, 404.

Behera, D., Kumar, V.A., Rao, J.P., Padal, S.B., Ayyappan, N., Reddy, C.S., 2024. Estimating aboveground biomass of a regional forest landscape by integrating textural and spectral variables of Sentinel-2 along with ancillary data. J. Indian Soc. Remote Sens. 52, 917–929.

Behera, D., Menon, D., Wilson, V.K., Ayyappan, N., 2025. Tree diversity, community structure and aboveground biomass of a lowland dipterocarp forest of Western Ghats. India. Trees For. People 21, 101048.

Berninger, A., Lohberger, S., Stängel, M., Siegert, F., 2018. SAR-based estimation of above-ground biomass and its changes in tropical forests of Kalimantan using L-and C-band. Remote Sens 10, 831.

Bhat, Y., Kripa, M.K., Dadhwal, V.K., 2024. Validation of ESA-CCI Forest biomass products over India: methodological and data challenges and results. J. Indian Soc. Remote. Sens. 52, 931–942. https://doi.org/10.1007/s12524-023-01741-w.

Breiman, L., 2001. Random forests. Mach Learn 45, 5–32.

Bulut, S., 2023. Machine learning prediction of above-ground biomass in pure Calabrian pine (Pinus brutia Ten.) stands of the Mediterranean region. Türkiye. Ecol. Inf. 74, 101951. https://doi.org/10.1016/j.ecoinf.2022.101951.

Carreiras, J.M., 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 Sens 5, 1524–1548.

Chave, J., Condit, R., Lao, S., Caspersen, J.P., Foster, R.B., Hubbell, S.P., 2003. Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. J. Ecol. 91, 240–252.

Chave, J., Coomes, D., Jansen, S., Lewis, S.L., Swenson, N.G., Zanne, A.E., 2009. Towards a worldwide wood economics spectrum. Ecol. Lett. 12, 351–366.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Change Biol. 20, 3177–3190.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., 2015. Xgboost: extreme gradient boosting. R package version 0.4-2 1 (4), 1–4.

Chen, L., Xiang, W., Wu, H., Ouyang, S., Zhou, B., Zeng, Y., Chen, Y., Kuzyakov, Y., 2019. Tree species identity surpasses richness in affecting soil microbial richness and community composition in subtropical forests. Soil Biol. Biochem. 130, 113–121.

Coulston, J.W., Blinn, C.E., Thomas, V.A., Wynne, R.H., 2016. Approximating prediction uncertainty for random forest regression models. Photogramm. Eng. Remote. Sens. 82 (3), 189–197.

Csillik, O., Kumar, P., Mascaro, J., O'Shea, T., Asner, G.P., 2019. Monitoring tropical forest carbon stocks and emissions using Planet satellite data. Sci. Rep. 9, 17831.

Dang, A.T.N., Nandy, S., Srinet, R., Luong, N.V., Ghosh, S., Senthil Kumar, A., 2019. Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park. Vietnam. Ecol. Inform. 50, 24–32.

Dar, A.A., Babu, K.N., Sundarapandian, S., Parthasarathy, N., 2024. Disentangling the response of species diversity, forest structure, and environmental drivers to aboveground biomass in the tropical forests of Western Ghats. India. Sci. Total Env. 957, 177684. https://doi.org/10.1016/j.scitotenv.2024.177684.

Eamus, D., 2003. How does ecosystem water balance affect net primary productivity of woody ecosystems? Functional Plant Biol. 30, 187–205.

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., Xiang, Y., 2018. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. Energy Conv. Manag. 164, 102–111. https://doi.org/10.1016/j.enconman.2018.02.087.

Fararoda, R., Reddy, R.S., Rajashekar, G., Chand, T.K., Jha, C.S., Dadhwal, V.K., 2021. Improving forest above ground biomass estimates over Indian forests using multi source data sets with machine learning algorithm. Ecol. Inform. 65, 101392.

Fararoda, R., Reddy, R.S., Rajashekar, G., Mayamanikandan, T., Mutyala, P., Satish, K.V., Pasha, S.W., Jha, C.S., 2024. Improving plot-level above ground biomass estimation in tropical Indian forests. Ecol. Inform. 81, 102621.

Forkuor, G., Zoungrana, J.-B.B., Dimobe, K., Ouattara, B., Vadrevu, K.P., Tondoh, J.E., 2020. Aboveground biomass mapping in West African Dryland Forest using sentinel-1 and 2 datasets - A case study. Remote Sens., Environ. 236, 111496.

Frank, J., Castle, M.E., Westfall, J.A., Weiskittel, A.R., MacFarlane, D.W., Baral, S.K., Radtke, P.J., Pelletier, G., 2018. Variation in occurrence and extent of internal stem decay in standing trees across the eastern US and Canada: evaluation of alternative modelling approaches and influential factors. Forestry: An Int. J. Forest Res. 91, 382–399. https://doi.org/10.1093/forestry/cpx054.

FSI, 1996. Volume equations for forests of India, Nepal and Bhutan. Dehradun, Ministry of Environment and Forests. Government of India.

FSI, 2015. State of Forest Report. Ministry of Environment and Forests. Government of India.

Gao, B.C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens Environ 58.

Gao, Y., Lu, D., Li, G., Wang, G., Chen, Q., Liu, L., Li, D., 2018. Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region. Remote Sens 10, 627.

Gholamy, A., Kreinovich, V., Kosheleva, O., 2018. Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. International Journal of Intelligent Technologies and Applied Statistics 11, 105–111.

Ghosh, S.M., Behera, M.D., 2018. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. Appl. Geogr. 96, 29–40. https://doi.org/10.1016/j.apgeog.2018.05.011.

Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80, 76–87.

Guerra-Hernández, J., Narine, L.L., Pascual, A., Gonzalez-Ferreiro, E., Botequim, B., Malambo, L., Neuenschwander, A., Popescu, S.C., Godinho, S., 2022. Aboveground biomass mapping by integrating ICESat-2, SENTINEL-1, SENTINEL-2, ALOS2/PALSAR2, and topographic information in Mediterranean forests. GISci. Remote Sensing 59, 1509–1533.

Guitet, S., Hérault, B., Molto, Q., Brunaux, O., Couteron, P., 2015. Spatial structure of above-ground biomass limits accuracy of carbon mapping in rainforest but large-scale forest inventories can help to overcome. PLoS One 10, e0138456.

Han, H., Wan, R., Li, B., 2021. Estimating forest aboveground biomass using Gaofen-1 images, Sentinel-1 images, and machine learning algorithms: A case study of the Dabie Mountain region. China. Remote Sens. 14, 176. https://doi.org/10.3390/rs14010176.

Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. 610–621.

Harris, N.L., Gibbs, D.A., Baccini, A., Birdsey, R.A., De Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M.C., Herold, M., Houghton, R.A., 2021. Global maps of twenty-first century forest carbon fluxes. Nat. Clim. Change 11, 234–240.

Haynes, W., 2013. Benjamini–Hochberg Method. In: Dubitzky, W., Wolkenhauer, O., Cho, KH., Yokota, H. (Eds.), Encyclopedia of Systems Biology. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9863-7_1215.

Hernández-Stefanoni, J.L., Castillo-Santiago, M.Á., Mas, J.F., Wheeler, C.E., Andres-Mauricio, J., Tun-Dzul, F., George-Chacón, S.P., Reyes-Palomeque, G., Castellanos-Basto, B., Vaca, R., Dupuy, J.M., 2020. Improving aboveground biomass maps of tropical dry forests by integrating LiDAR, ALOS PALSAR, climate and field data. Carbon Balance Manag 15, 15.

Hu, T., Su, Y., Xue, B., Liu, J., Zhao, X., Fang, J., Guo, Q., 2016. Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data. Remote Sens 8, 565.

Indirabai, I., Nilsson, M., 2024. Estimation of above ground biomass in tropical heterogeneous forests in India using GEDI. Ecol. Inform. 82, 102712.

ISFR., 2022. Forest Survey of India. Ministry of Environment, Forest & Climate Change. Retrieved from fsi.nic.in/forest-report-2021.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+ SAIL models: A review of use for vegetation characterization. Remote Sens. Environ. 113, S56–S66.

Jha, N., Tripathi, N.K., Chanthorn, W., Brockelman, W., Nathalang, A., Pélissier, R., Pimmasarn, S., Ploton, P., Sasaki, N., Virdis, S.G., Réjou-Méchain, M., 2020. Forest aboveground biomass stock and resilience in a tropical landscape of Thailand. Biogeosciences 17 (1), 121–134.

Jha, N., Tripathi, N.K., Barbier, N., Virdis, S.G.P., Chanthorn, W., Viennois, G., Brockelman, W.Y., Nathalang, A., Tongsima, S., Sasaki, N., Pélissier, R., Réjou-

Méchain, M., 2021. The real potential of current passive satellite data to map aboveground biomass in tropical forests. Remote Sens. Ecol. Conserv. 7, 504–520.
Jia, B., Guo, W., He, J., Sun, M., Chai, L., Liu, J., Wang, X., 2022. Topography, diversity, and forest structure attributes drive aboveground carbon storage in different forest types in Northeast China. Forests 13, 455.
Joshi, N.P., Mitchard, E.T., Schumacher, J., Johannsen, V.K., Saatchi, S., Fensholt, R., 2015. L-band SAR backscatter related to forest cover, height and aboveground biomass at multiple spatial scales across Denmark. Remote Sens 7, 4442–4472.
Kanda, N.B., Ayushi, K., Wilson, V.K., Ayyappan, N., Parthasarathy, N., 2021. The woody flora of shettihalli wildlife sanctuary, central western ghats of Karnataka, India-A checklist. J. Threat. Taxa 13, 20033–20055.
Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, P., Kessler, M., 2021. Climatologies at high resolution for the Earth land surface areas. EnviDat. https://doi.org/10.16904/envidat.228.
Kelsey, K.C., Neff, J.C., 2014. Estimates of aboveground biomass from texture analysis of Landsat imagery. Remote Sens 6, 6407–6422.
Khanal, S., Nolan, R.H., Medlyn, B.E., Boer, M.M., 2023. Mapping soil organic carbon stocks in Nepal's forests. Sci. Rep. 13, 8090. https://doi.org/10.1038/s41598-023-34247-z.
Knapp, N., Attinger, S., Huth, A., 2022. A question of scale: modeling biomass, gain and mortality distributions of a tropical forest. Biogeosciences 19, 4929–4944.
Kothandaraman, S., Dar, J.A., Sundarapandian, S., Dayanandan, S., Khan, M.L., 2020. Ecosystem-level carbon storage and its links to diversity, structural and environmental drivers in tropical forests of Western Ghats. India. Sci. Rep. 10, 70313. https://doi.org/10.1038/s41598-020-70313-6.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., ..¬ Candan, C., 2017. Caret: classification and regression training (R Package version 6.0-76) [WWW Document].
Kumar, A., Kishore, B., Saikia, P., Deka, J., Bharali, S., Singha, L.B., Tripathi, O.P., Khan, M.L., 2019. Tree diversity assessment and above ground forests biomass estimation using SAR remote sensing: A case study of higher altitude vegetation of North-East Himalayas. India. Phys. Chem. Earth Parts Abc. 111, 53–64.
Kupidura, P., 2019. The comparison of different methods of texture analysis for their efficacy for land use classification in satellite imagery. Remote Sens 11, 1233.
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 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. J. Appl. Rem. Sens. 12, 1. https://doi.org/10.1117/1.JRS.12.016008.
Lê, S., Josse, J., Husson, F., 2008. FactoMineR: A package for multivariate analysis. J. Stat. Softw. 25 (1), 1–18. https://doi.org/10.18637/jss.
Li, Y., Li, M., Li, C., Liu, Z., 2020. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. Sci. Rep. 10, 9952.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2 (3), 18–22.
Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., Moran, E., 2014. A survey of remote sensing–based aboveground biomass estimation methods in forest ecosystems. Internation Journal of Earth 9, 63–105.
McEwan, R.W., Lin, Y.-C., Sun, I.-F., Hsieh, C.-F., Su, S.-H., Chang, L.-W., Song, G.-Z.M., Wang, H.-H., Hwong, J.-L., Lin, K.-C., Yang, K.-C., Chiang, J.-M., 2011. Topographic and biotic regulation of aboveground carbon storage in subtropical broad-leaved forests of Taiwan. For. Ecol. Manage. 262, 1817–1825. https://doi.org/10.1016/j.foreco.2011.07.028.
Mayamanikandan, T., Reddy, S., Fararoda, R., Thumaty, K.C., Praveen, M.S.S., Rajashekar, G., Jha, C.S., Das, I.C., Gummapu, J., 2022. Quantifying the influence of plot-level uncertainty in above ground biomass up scaling using remote sensing data in central Indian dry deciduous forest. Geocarto Int 37, 3489–3503. https://doi.org/10.1080/10106049.2020.1864029.
Meinshausen, N., Ridgeway, G., 2006. Quantile regression forests. J. Mach. Learn. Res. 7 (6), 1–12.
Meinshausen, N., Schiesser, L., 2015. Quantregforest: quantile regression forests. R package.https://cran.r-project.org.
Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.C., Lin, C.C., 2015. Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071).
Mitchard, E.T., Saatchi, S.S., White, L.J., Abernethy, K.A., Jeffery, K.J., Lewis, S.L., Collins, M., Lefsky, M.A., Leal, M.E., Woodhouse, I.H., 2012. Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud. Biogeosciences 9, 179–191.
Muscarella, R., Kolyaie, S., Morton, D.C., Zimmerman, J.K., Uriarte, M., 2020. Effects of topography on tropical forest structure depend on climate context. J. Ecol. 108, 145–159. https://doi.org/10.1111/1365-2745.13261.
Mutanga, O., Masenyama, A., Sibanda, M., 2023. Spectral saturation in the remote sensing of high-density vegetation traits: A systematic review of progress, challenges, and prospects. ISPRS J. Photogramm. Remote Sens. 198, 297–309.
Najeeb, N., Jose, K., Sreejith, K.A., Pulla, S., Suresh, H.S., Ratnam, J., Raghavendra, H.V., Chakravarthy, D., Chaturvedi, R.K., 2025. Presence of large trees and tree diversity enhances carbon storage in the Western Ghats. Biol. Conserv. 308, 111250.
Nandy, S., Srinet, R., Padalia, H., 2021. Mapping forest height and aboveground biomass by integrating ICESat-2, Sentinel-1 and Sentinel-2 data using random forest algorithm in northwest himalayan foothills of India. Geophys. Res. Lett. 48. https://doi.org/10.1029/2021GL093799 e2021GL093799.
Naimi, B., 2015. Usdm: Uncertainty analysis for species distribution models, 1, 1–12.
Osuri, A.M., Machado, S., Ratnam, J., Sankaran, M., Ayyappan, N., Muthuramkumar, S., Naeem, S., 2020. Tree diversity and carbon storage cobenefits in tropical human dominated landscapes. Conserv. Lett. 13 (2), e12699. https://doi.org/10.1111/conl.12699.
Padalia, H., Prakash, A., Watham, T., 2023. Modelling aboveground biomass of a multistage managed forest through synergistic use of Landsat-OLI, ALOS-2 L-band SAR and GEDI metrics. Ecol. Inform. 77, 102234.
Ploton, P., Barbier, N., Couteron, P., Antin, C.M., Ayyappan, N., Balachandran, N., Barathan, N., Bastin, J.-F., Chuyong, G., Dauby, G., 2017. Toward a general tropical forest biomass prediction model from very high resolution optical satellite images. Remote Sens. Environ 200, 140–153.
Ploton, P., Mortier, F., Barbier, N., Cornu, G., Réjou-Méchain, M., Rossi, V., Gourlet-Fleury, S., 2020a. A map of African humid tropical forest aboveground biomass derived from management inventories. Sci Data 7, 221. https://doi.org/10.1038/s41597-020-0561-0.
Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020b. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat Commun 11, 4540. https://doi.org/10.1038/s41467-020-18321-y.
Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sens. Environ. 253, 112165.
Rajashekar, G., Fararoda, R., Reddy, R.S., Jha, C.S., Ganeshaiah, K.N., Singh, J.S., Dadhwal, V.K., 2018. Spatial distribution of forest biomass carbon (Above and below ground) in Indian forests. Ecol. Indic. 85, 742–752. https://doi.org/10.1016/j.ecolind.2017.11.024.
R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/index.html.
Reddy, C.S., Kurian, A., Srivastava, G., Singhal, J., Varghese, A.O., Padalia, H., Ayyappan, N., Rajashekar, G., Jha, C.S., Rao, P.V.N., 2021. Remote sensing enabled essential biodiversity variables for biodiversity assessment and monitoring: technological advancement and potentials. Biodivers Conserv 30, 1–14. https://doi.org/10.1007/s10531-020-02073-8.
Réjou-Méchain, M., Barbier, N., Couteron, P., Ploton, P., Vincent, G., Herold, M., Mermoz, S., Saatchi, S., Chave, J., De Boissieu, F., Féret, J.-B., Takoudjou, S.M., Pélissier, R., 2019. Upscaling forest biomass from field to satellite measurements: sources of errors and ways to reduce them. Surv. Geophys. 40, 881–911. https://doi.org/10.1007/s10712-019-09532-0.
Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., Hérault, B., 2017. BIOMASS : an R package for estimating above-ground biomass and its uncertainty in tropical forests. Methods Ecol. Evol. 8, 1163–1167. https://doi.org/10.1111/2041-210X.12753.
Rodda, S.R., Fararoda, R., Gopalakrishnan, R., Jha, N., Réjou-Méchain, M., Couteron, P., Barbier, N., Alfonso, A., Bako, O., Bassama, P., 2024a. LiDAR-based reference aboveground biomass maps for tropical forests of South Asia and Central Africa. Sci. Data 11, 334.
Rodda, S.R., Fararoda, R., Gopalakrishnan, R., Jha, N., Réjou-Méchain, M., Couteron, P., et al., 2024b. LiDAR-based reference aboveground biomass maps for tropical forests of. South Asia and Central Africa. Sci. Data 11, 334. https://doi.org/10.1038/s41597-024-03162-x.
Rodrigues, A.C., Villa, P.M., Ali, A., Ferreira-Júnior, W., Neri, A.V., 2020. Fine-scale habitat differentiation shapes the composition, structure and aboveground biomass but not species richness of a tropical Atlantic forest. J. For. Res. 31, 1599–1611. https://doi.org/10.1007/s11676-019-00994-x.
Rodríguez-Veiga, P., Wheeler, J., Louis, V., Tansey, K., Balzter, H., 2017. Quantifying forest biomass carbon stocks from space. Curr. For. Rep. 3, 1–18. https://doi.org/10.1007/s40725-017-0052-5.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec Publ 351, 309.
Saarela, S., Holm, S., Healey, S.P., Andersen, H.-E., Petersson, H., Prentius, W., Patterson, P.L., Næsset, E., Gregoire, T.G., Ståhl, G., 2018. Generalized hierarchical model-based estimation for aboveground biomass assessment using GEDI and landsat data. Remote Sens 10, 1832.
Saarela, S., Varvia, P., Korhonen, L., Yang, Z., Patterson, P.L., Gobakken, T., et al., 2023. Three-phase hierarchical model-based and hybrid inference. MethodsX 11, 102321.
Sagang, L.B.T., Ploton, P., Sonké, B., Poilvé, H., Couteron, P., Barbier, N., 2020. Airborne LiDAR sampling pivotal for accurate regional AGB predictions from multispectral images in forest-savanna landscapes. Remote Sens 12, 1637.
Sainuddin, F.V., Malek, G., Rajwadi, A., Nagar, P.S., Asok, S.V., Reddy, C.S., 2024. Estimating above-ground biomass of the regional forest landscape of Northern western ghats using machine learning algorithms and multi-sensor remote sensing data. J. Indian Soc. Remote Sens. 52, 885–902. https://doi.org/10.1007/s12524-024-01836-y.
Salinas-Melgoza, M.A., Skutsch, M., Lovett, J.C., 2018. Predicting aboveground forest biomass with topographic variables in human-impacted tropical dry forest landscapes. Ecosphere 9, e02063. https://doi.org/10.1002/ecs2.2063.
Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D.M.A., Avitabile, V., Araza, A., Willcock, S., 2021. The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. Earth Syst. Sci. Data 13, 3927–3950. https://doi.org/10.5194/essd-13-3927-2021.
Singh, C., Karan, S.K., Sardar, P., Samadder, S.R., 2022. Remote sensing-based biomass estimation of dry deciduous tropical forest using machine learning and ensemble analysis. J. Environ. Manage. 308, 114639.
Su, H., Shen, W., Wang, J., Ali, A., Li, M., 2020. Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. For. Ecosyst. 7, 64. https://doi.org/10.1186/s40663-020-00276-7.

Teoh, K.K., Ibrahim, H., Bejo, S.K., 2008. Investigation on several basic interpolation methods for the use in remote sensing application. In: 2008 IEEE Conference on Innovative Technologies in Intelligent Systems and Industrial Applications. IEEE, pp. 60–65.

Thenkabail, P.S., Ward, A.D., Lyon, J.G., Merry, C.J., 1994. Thematic Mapper vegetation indices for determining soybean and corn growth parameters. Photogramm. Eng. Remote Sens. 60, 437–442.

Tian, L., Wu, X., Tao, Y., Li, M., Qian, C., Liao, L., Fu, W., 2023. Review of remote sensing-based methods for forest aboveground biomass estimation: progress, challenges, and prospects. Forests 14, 1086. https://doi.org/10.3390/f14061086.

Turton, A.E., Augustin, N.H., Mitchard, E.T., 2022. Improving estimates and change detection of forest above-ground biomass using statistical methods. Remote Sens 14, 4911.

Viet Nguyen, L., Tateishi, R., Kondoh, A., Sharma, R.C., Thanh Nguyen, H., Trong To, T., Ho Tong Minh, D., 2016. Mapping tropical forest biomass by combining ALOS-2, Landsat 8, and field plots data. Land (Basel) 5, 31.

Walker, S.M., Murray, L., Tepe, T., 2016. Allometric equation evaluation guidance document – Guidance Document for Lao PDR. Developed by Winrock International, on behalf Lao PDR. Funded by JICA.

Wang, J., Du, H., Li, X., Mao, F., Zhang, M., Liu, E., Ji, J., Kang, F., 2021. Remote sensing estimation of bamboo forest aboveground biomass based on geographically weighted regression. Remote Sens (Basel) 13, 2962. https://doi.org/10.3390/rs13152962.

Warren, D.L., Matzke, N.J., Cardillo, M., Baumgartner, J.B., Beaumont, L., Dinnage, R., 2021. ENMTools 1.0: an R package for comparative ecological biogeography. Ecography 44, 504–511.

Wilson, V.K., Ayyappan, N., Parthasarathy, N., 2023. Decadal forest dynamics in logged and unlogged sites at Uppangala, Western Ghats, India. Environ. Monit. Assess. 195, 66. https://doi.org/10.1007/s10661-022-10706-4.

Working Group on Calibration and Validation, C. on E.O.S., 2021. Aboveground Woody biomass product validation Good practices protocol. https://doi.org/10.5067/DOC/CEOSWGCV/LPV/AGB.001.

Yu, Y., Pan, Y., Yang, X., Fan, W., 2022. Spatial scale effect and correction of forest aboveground biomass estimation using remote sensing. Remote Sens (Basel) 14, 2828.

Zanne, A.E., Lopez-Gonzalez, G., Coomes, D.A., Ilic, J., Jansen, S., Lewis, S.L., Miller, R. B., Swenson, N.G., Wiemann, M.C., Chave, J., 2009. Global wood density database. Data From: Towards a Worldwide Wood Economics Spectrum.. In: Dryad Digital Repository. https://doi.org/10.5061/dryad.234.

Zhang, J., Huang, S., Hogg, E.H., Lieffers, V., Qin, Y., He, F., 2014. Estimating spatial variation in Alberta forest biomass from a combination of forest inventory and remote sensing data. Biogeosciences 11, 2793–2808.

Zhang, L., Zhang, X., Shao, Z., Jiang, W., Gao, H., 2023. Integrating Sentinel-1 and 2 with LiDAR data to estimate aboveground biomass of subtropical forests in northeast Guangdong, China. Int. J. Digital Earth 16, 158–182.

Zhang, X., Shen, H., Huang, T., Wu, Y., Guo, B., Liu, Z., Luo, H., Tang, J., Zhou, H., Wang, L., 2024. Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery. Ecol. Indic. 159, 111752.

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