Ecological Modelling 508 (2025) 111242

Contents lists available at ScienceDirect

Ecological Modelling

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

Enhanced aboveground biomass density estimation in Central Vietnamese forests

Viet Hoang Ho [a,d,*], Hidenori Morita [b], Felix Bachofer [c], Thanh Ha Ho [d]

[a] Graduate School of Environmental and Life Science, Okayama University, 1 Chome-1-1 Tsushimanaka, Kita Ward, Okayama 700-8530, Japan
[b] Graduate School of Environmental, Life, Natural Science and Technology, Okayama University, 1 Chome-1-1 Tsushimanaka, Kita Ward, Okayama 700-8530, Japan
[c] German Aerospace Center (DLR), Earth Observation Center, 82234 Wessling, Germany
[d] University of Agriculture and Forestry, Hue University, 102 Phung Hung Str, Hue City, Thua Thien Hue 53000, Vietnam

ARTICLE INFO

Keywords:
Forest aboveground biomass density
Random forest
Ordinary kriging
Co-kriging
Multispectral
Multi-frequency synthetic aperture radar

ABSTRACT

Accurate estimation of spatially explicit forest aboveground biomass density (AGBD) is essential for supporting climate change mitigation strategies. Recent studies have demonstrated the predictive effectiveness of the random forest (RF) algorithm in forest AGBD estimation utilizing multi-source remote sensing (RS) data. However, the RF-based estimates may be further enhanced by integrating RF with kriging techniques that account for spatial autocorrelation in model residuals. Therefore, we investigated the performance of random forest ordinary kriging (RFOK) and random forest co-kriging (RFCK) for estimating AGBD in Central Vietnamese forests using Advanced Land Observing Satellite-2 Phased Array L-band Synthetic Aperture Radar-2 (ALOS-2 PALSAR-2), Sentinel-1 (S1), and Sentinel-2 (S2) imageries. 277 predictors, including spectral bands, radar backscatter coefficients, vegetation indices, biophysical variables, and texture metrics, were derived from these RS datasets and statistically linked to field measurements from 104 geo-referenced forest inventory plots. The results showed that textures, modified chlorophyll absorption ratio index (MCARI), and radar backscatters were key contributors to AGBD variability. The fusion of ALOS-2 PALSAR-2 and S2 data yielded the highest RF performance, with coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) achieving 0.75, 39.15 t.ha^{-1}, and 32.20 t.ha^{-1}, respectively. Incorporating interpolated residuals by ordinary kriging and co-kriging into RF predictions enhanced estimation accuracy, with relative improvements of 5.74–7.04 % in R^2, 8.73–10.91 % in RMSE, and 13.62–15.27 % in MAE, yet these gains remained limited. Although RFOK achieved marginally better accuracy (R^2 = 0.80, RMSE = 34.88 t.ha^{-1}, MAE = 27.28 t.ha^{-1}) compared to RFCK (R^2 = 0.79, RMSE = 35.73 t.ha^{-1}, MAE = 27.81 t.ha^{-1}), the latter reduced estimation bias more effectively, likely due to the inclusion of elevation as a covariate in the co-kriging process. These findings underscore the potential of the hybrid RF-kriging frameworks for improving spatial AGBD estimation, offering a robust approach for carbon accounting in tropical ecosystems.

1. Introduction

Forest ecosystems play a crucial role by effectively reducing atmospheric carbon levels (Canadell and Raupach, 2008; Ji et al., 2024). Prior research has demonstrated that forest ecosystems on a worldwide scale absorb over 30 % of human-caused carbon dioxide emissions by extracting nearly 2 Pg carbon from the atmosphere per year (Bellassen and Luyssaert, 2014). Among forest ecosystems, tropical forests serve as vital contributors to carbon sequestration and storage, surpassing other types of forests (Nesha et al., 2020). Specifically, on average, tropical forests store 303 t carbon ha^{-1} (Lü et al., 2010), while temperate and boreal forests store 66 and 44 t carbon ha^{-1}, respectively (Thurner et al., 2014). However, in most tropical nations, deforestation and forest degradation substantially affect the carbon cycle globally (Mwambala et al., 2023; Pearson et al., 2017) since the aboveground biomass (AGB), which includes the trunks, branches, and leaves of forest trees, holds the vast majority of the forest carbon reservoir (Pan et al., 2011). In addition, the United Nations Framework Convention on Climate Change (UNFCCC) has acknowledged AGB as a crucial climate parameter for monitoring carbon storage in forest landscapes (Rodríguez-Veiga et al., 2016). Thus, spatially explicit estimates of aboveground biomass density (AGBD) are a critical quantitative approach to estimating terrestrial

* Corresponding author.
E-mail addresses: paoe9ws1@s.okayama-u.ac.jp, hoviethoang@hueuni.edu.vn (V.H. Ho).

https://doi.org/10.1016/j.ecolmodel.2025.111242
Received 18 November 2024; Received in revised form 29 May 2025; Accepted 17 June 2025
Available online 23 June 2025
0304-3800/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

carbon stocks and enhancing our comprehension of the mechanisms that control carbon storage and release in tropical forests where uncertainty remains high (Rodríguez-Veiga et al., 2017; Vorster et al., 2020).

To estimate AGBD, conventional methods using field measurements, such as destructive sampling and allometric models, are the most accurate, yet pose challenges when used in broad regions due to their negative impacts on the environment, high costs, time requirements, and labor intensiveness (Sinha et al., 2015; Timothy et al., 2016). Meanwhile, remote sensing (RS) is widely acknowledged as a promising approach for estimating AGBD, as it enables low-cost, rapid data acquisition and provides coverage of forested areas that are otherwise difficult or hazardous to access through ground-based methods (Barbosa et al., 2014; Koch, 2010; Lu, 2006; Yu et al., 2022). The primary types of remotely sensed data for extracting information on AGBD are optical RS data, synthetic aperture radar (SAR) data, and Light Detection and Ranging (LiDAR) data (Sinha et al., 2015). Although optical RS data have been widely employed for estimating forest AGBD because of their accessibility and cost-effectiveness (Cutler et al., 2012; Foody et al., 2003; Timothy et al., 2016), they typically have constraints when distinguishing ecosystems that exhibit moderate to high amounts of AGBD (GFOI, 2013). Whereas, SAR data can penetrate the forest canopy, a capability completely lacking in optical data (Sinha et al., 2015). Most AGBD estimation studies frequently use the C-, L-, and P-band SAR data (Ghasemi et al., 2011). Among these, the L-band is regarded as the most appropriate frequency for estimating forest biomass because it can penetrate deeper into vegetation canopies and interact with the limbs and trunks of trees, with minimal sensitivity to environmental conditions (Berninger et al., 2018; Hamdan et al., 2015; Sun et al., 2002). In addition to its benefits, SAR data presents drawbacks related to speckles and image distortions in uneven terrain areas (Sinha et al., 2015). The LiDAR technology is deemed more precise than optical or radar sensors for estimating AGBD because of its robust correlation with biomass level and better resistance to the saturation effect (Lu et al., 2012; Yavaşlı, 2012; Zolkos et al., 2013). However, the exorbitant expenses related to the collecting and processing of LiDAR data continue to be a significant obstacle, especially for large-scale forest monitoring programs (Su et al., 2020; Xu et al., 2021). To improve the prediction accuracy in estimating forest AGBD, the multi-frequency SAR data (Berninger et al., 2018; Huang et al., 2018; Sivasankar et al., 2018; Zeng et al., 2022) and the integration of SAR and optical sensor data (Li et al., 2020a; Z. Li et al., 2022; Monsalve-Tellez et al., 2022; Naik et al., 2022; Pham et al., 2020; Vafaei et al., 2018) are considered potential ways.

As opposed to direct forest biomass estimations, RS techniques generally estimate forest AGBD by establishing relationships between ground measurements and RS-derived parameters, such as spectral indices, canopy cover and height, texture, shaded fraction, leaf and basal area, and timber volume, through the use of predictive models (Goetz et al., 2009; Li et al., 2020a; Tian et al., 2023). Over the past decade, machine learning (ML) algorithms, including random forest (RF), support vector regression (SVR), artificial neural networks (ANN), k-nearest neighbors (kNN), stochastic gradient boosting (SGB), and extreme gradient boosting (XGBoost), have been widely employed as predictive models for estimating AGBD in forest ecosystems (Li et al., 2020a; Su et al., 2020; Zhang et al., 2020). Their advantages over traditional statistical regression methods include the ability to handle nonlinear data and require no assumptions about the underlying data distribution or the relationships between AGBD and RS features (Wu et al., 2016; Zhang et al., 2019). Existing research indicates that RF frequently achieved improved accuracy in forest AGBD estimation compared to other ML algorithms and traditional statistical regressions due to its less sensitivity to noise in the training data and multicollinearity issue (Li et al., 2024; Su et al., 2020). For example, Wu et al. (2016) compared stepwise linear regression, kNN, SVR, RF, and SGB approaches for estimating forest AGBD from Landsat imagery in northwestern Zhejiang Province, China, and found that RF yielded the highest accuracy. Ghosh and Behera (Ghosh and Behera, 2018) evaluated RF and SGB using multiple data sources in a dense tropical forest and concluded that RF achieved greater accuracy. Similarly, Ramachandran and Dikshit (2022) assessed the performance of RF, ANN, and XGBoost using airborne L-band SAR data to predict AGBD across forested landscapes and reported that RF outperformed the other methods. Kumari and Kumar (2023) also confirmed the superior performance of RF over SVR in estimating forest AGBD from MODIS data in Uttarakhand, India. However, one of the limitations of RF is its inability to account for the influence of neighboring observed data, known as spatial autocorrelation, when predicting the spatial distribution (Chen et al., 2019a), making it difficult to completely mine the relationship between the regionalized variable and its predictors. Consequently, the relationship that is not captured by the trend component is transferred to the residuals, causing them to still retain spatial autocorrelation structure (Zhu et al., 2022). To overcome the disadvantage, the kriging methods can be used to capture the spatial dependence of the residual components obtained from RF (Islam et al., 2017). Utilizing variogram theory and structural analysis, the kriging methods can not only characterize the underlying spatial structure but also achieve the best linear unbiased estimation of regionalized variables (Jiang et al., 2022). Integrating the predictive power of RF for generating deterministic trends with the capability of ordinary kriging (OK) or co-kriging (CK) for capturing spatial autocorrelation in residuals is likely to provide an initiative to yield more accurate estimates compared to using standalone RF in various environmental fields, such as soil attributes and pollutant concentrations (Su et al., 2020). However, these hybrid methods do not consistently yield improved accuracy in estimating forest AGBD, as spatial autocorrelation is not uniformly present across all forest ecosystems because of its strong dependence on the heterogeneity and structural complexity of local forest landscapes (Viana et al., 2012).

Tropical forests present significant challenges for field access due to their characteristic dense vegetation and complex topography, leading to a scarcity of in situ data (Saatchi et al., 2011; Sinha et al., 2015). This data limitation often results in AGBD prediction models, those employing RF trained over extensive areas, being susceptible to overfitting and failing to capture local variability adequately (Jha et al., 2021; Su et al., 2020). In such cases, the synergistic integration of optical and multi-frequency SAR data, combined with hybrid approaches that fuse RF and kriging techniques, offers promising avenues. However, the effectiveness of these methodologies in tropical forest ecosystems remains underexplored and warrants further investigation. Taking these issues above into account, we conducted this study to (1) identify the most influential RS-derived predictors, (2) evaluate the performance of different RS data fusion strategies with RF, and (3) assess the extent of improvement of the random forest ordinary kriging (RFOK) model and the random forest co-kriging (RFCK) model compared to RF for estimating AGBD in tropical forests. The findings of this study are expected to contribute to the strategic development of carbon sequestration forests and to inform evidence-based approaches for sustainable forest management.

2. Materials and methods

2.1. Study area

Danang is one of six centrally-controlled cities in Vietnam, roughly in the nation's center. The study area is located between 15.92^0 and 16.22°N latitude and 107.81^0 and 108.34°E longitude (Fig. 1), covering approximately 960 km^2 (Thi An et al., 2022). The city comprises two distinct primary topographic settings: mountains and plains. Over half of the city's territory is covered by mountainous areas along the northern and northwestern borders, with altitudes varying from 700 m to 1,500 m. The coastal lowlands are dominated by plains to the east of the city. Danang experiences a tropical monsoon environment characterized by rainy and dry seasons. The rainy season typically runs from August to December, with heavy rains and typhoons, while the dry season occurs

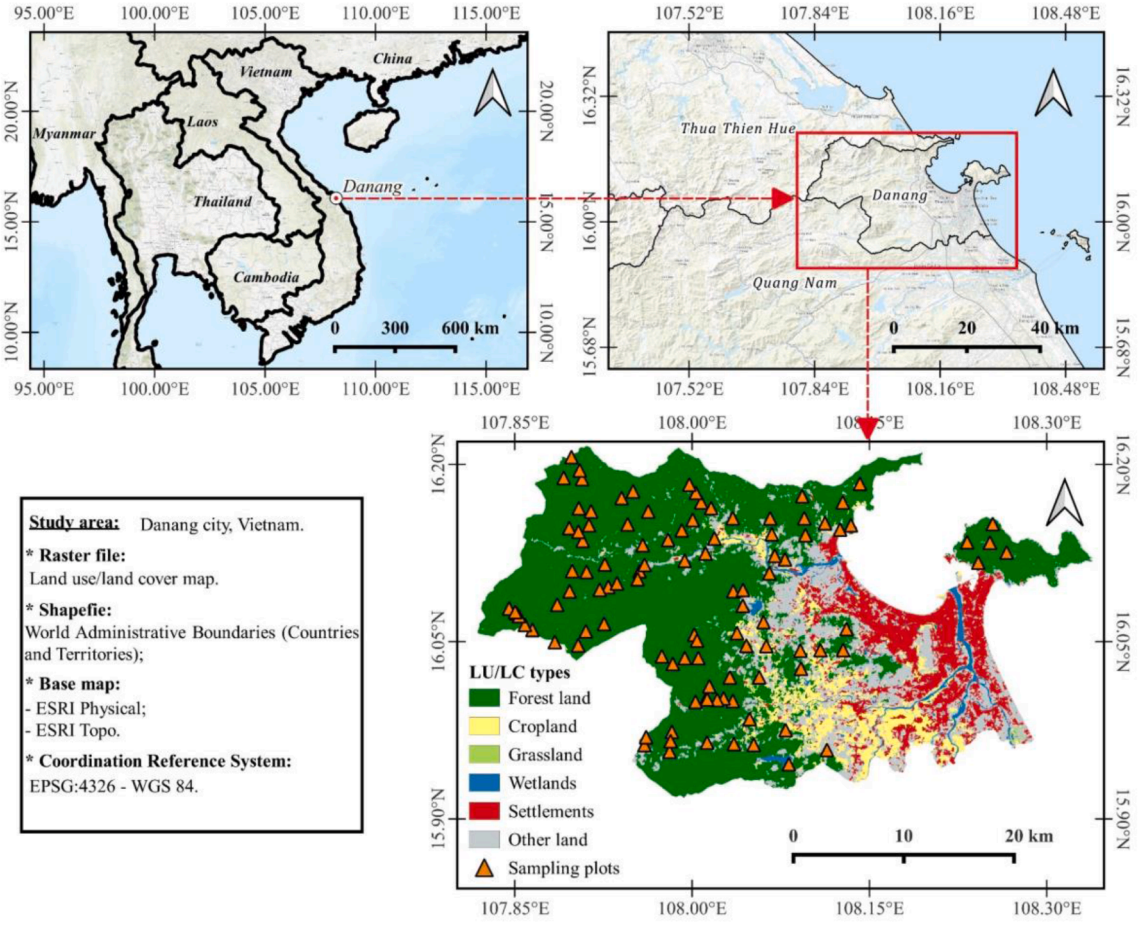

Fig. 1. The location of the study site (top); the LULC map of Danang for 2023 and the location of sampling plots (bottom right).

during the remainder of the year in hot and humid conditions (Hoang Khanh Linh and Van Chuong, 2015). Besides, Danang lies in a tropical rainforest zone (Iremonger and Gerrand, 2011), leading to the domination of evergreen broadleaf vegetation (e.g., *Lithocarpus annamensis, Polyalthia nemoralis, Scaphium lychnophorum, Syzygium levinei, Shorea farinosa*, and *Aglaia roxburghiana*) in the forest ecosystems (Huy et al., 2016).

Since the initiation of the *Doi Moi* (renovation) policy in 1986, Danang and the broader central region of Vietnam have undergone rapid land use/land cover (LULC) transformations. These changes, driven by accelerated economic development and globalization, have led to the conversion of tropical forests into built-up and agricultural land (Liang et al., 2022). Consequently, there has been a substantial decline in ecosystem carbon storage, particularly in forested zones (Shi et al., 2024). In response to growing environmental concerns, the Vietnamese government has committed to achieving national carbon neutrality by 2050 (Hoang Ha and Da Hanh, 2024). To support this objective, the development of appropriate LULC strategies aimed at minimizing carbon loss in Central Vietnam is crucial, highlighting the need for accurate quantification of both the magnitude and spatial distribution of forest carbon stocks, with particular emphasis on AGBD which typically accounts for nearly half of the total carbon stored in tropical forest ecosystems (Ghosh and Behera, 2018).

2.2. Data

2.2.1. In situ data

The field data, including living trees, shrubs, and herbs, were collected from 104 nested sampling plots between 13th July and 21st September 2023. The sampling plots were established within the forest land layer using a systematic unaligned sampling design, whereby one plot was randomly located within each 2.5 × 2.5km grid. The coordinates of the generated sampling plots were transferred to portable global positioning system (GPS) receivers (Garmin GPSMAP 64x) for navigation during field data collection, and to ensure consistency across measurements, all plots were oriented with their northern edge aligned to geographic north.

The 900-m^2 (30 × 30 m) plots were designed for living trees with a diameter at breast height (DBH) of above 5 cm. In each plot, DBH was measured by using D-tapes. For herbaceous and shrub vegetation, three 1-m^2 subplots were systematically placed along a 14.14-meter interval transect, extending from the southwest to the northeast corner of a 900-m^2 plot (Fig. 2). All herbs and shrubs in these subplots were cut into pieces and subsequently weighed directly in the field. Mixed subsamples were stored in sample plastic bags and brought to the laboratory to determine moisture content (Pearson et al., 2007).

Tree AGB is usually quantified by allometric equations that relate essential plot-level variables to AGB (Roy et al., 2021; L. Zhang et al., 2023). The tree AGB in our study area was calculated based on an allometric equation with a single variable (DBH only) created for the south-central coastal regions of Vietnam (Eq. (1)). The suggested equation demonstrated more excellent reliability than the single variable equations by Brown (1997) and IPCC (2003), as well as generic equations with up to three variables proposed by Chave et al. (2014) for tropical forests in the study site (Huy et al., 2016).

$$\boldsymbol{AGB_{tree}} = 0.104189 \times \boldsymbol{DBH}^{2.491453} \quad (1)$$

Where: AGB_{tree} is tree aboveground biomass (kg), and *DBH* is the

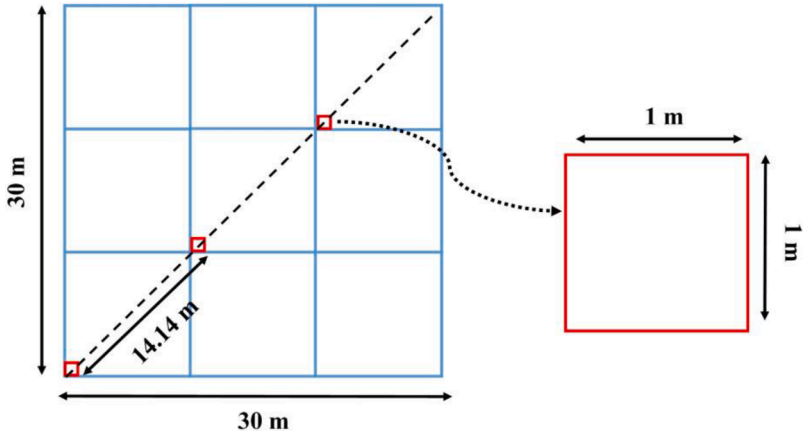

Fig. 2. Plot design for field survey.

diameter at breast height (cm).

Tree aboveground biomass density was then calculated as follows:

$$AGBD_{tree} = \frac{\sum AGB_{tree}}{A_{plot}} \times 10 \quad (2)$$

Where $AGBD_{tree}$ is tree aboveground biomass density (t.ha^{-1}), A_{plot} is the area of the 30 × 30m plot (m^2).

The estimation of herb/shrub biomass density was conducted with the formula:

$$HBD \Big/ SBD = \frac{W_{herb/shrub\_field}}{A_{subplots}} \times \left[\frac{W_{herb/shrub-subsample(dry)}}{W_{herb/shrub-subsample(fresh)}}\right] \times 10 \quad (3)$$

Where HBD/SBD is biomass density of herb/shrub layer (t.ha^{-1}), $W_{herb/shrub\_field}$ is the fresh weight of herb/shrub sample (kg), $A_{subplots}$ is the total area of 3 × 1 × 1m subplot (m^2), $W_{herb/shrub-subsample(dry)}$ is the oven-dried weight of herb/shrub subsample (g), and $W_{herb/shrub-subsample(fresh)}$ is the fresh weight of herb/shrub subsample (g).

AGBD was calculated as the sum of $AGBD_{tree}$, HBD, and SBD in each sampling plot.

2.2.2. Remote sensing data and processing

The Advanced Land Observing Satellite-2 Phased Array L-band SAR-2 (ALOS-2 PALSAR-2) is a SAR sensor developed by the Japan Aerospace Exploration Agency (JAXA). The sensor was released on May 24, 2014, and can be used to monitor tropical rainforests and detect carbon sinks (Baig et al., 2017). In this study, we obtained ALOS-2 PALSAR-2 fine beam dual–polarized (HV, HH) mode level 2.1 CEOS format data from the https://www.eorc.jaxa.jp website.

Sentinel-1 (S1), operated by the European Space Agency (ESA), performs C-band SAR imaging of the Earth (Yang et al., 2021). The S1 SAR sensor is capable of operating in both single and dual polarizations with a temporal resolution of 12 days (Abdikan et al., 2016). Sentinel-2 (S2), also launched by the ESA, is equipped with a multispectral (MSI) instrument, an optical sensor capable of capturing 13 spectral bands with an average revisit time of 5 days (Dong and Fu, 2023). S1 and S2 images were acquired from the https://dataspace.copernicus.eu website. The details of the Sentinel and ALOS-2 PALSAR-2 scenes are shown in Table 1. The preprocessing and use of RS data products are outlined in Fig. 3.

RS data was processed by the ESA SNAP toolbox in this study. The ALOS-2 PALSAR-2 images were transformed into normalized radar backscattering coefficients by applying the equation provided below:

$$\sigma^0 = 10.log_{10}(DN)^2 + CF \quad (4)$$

$$\gamma^0 = \frac{\sigma^0}{cos\varphi} \quad (5)$$

Table 1
Acquired remote sensing data.

Sensors	Quantity of imageries	Acquisition date (yyyy/mm/dd)	Processing level	Spatial resolution (m)
ALOS-2 PALSAR-2	02	2023/06/04	2.1	6.25
S1	01	2023/08/15	1	10
S2	09	2023/05/07 - 2023/08/15	1C	10 - 20

Where σ^0 is the backscattering coefficient in decibels (dB), γ^0 is the normalized radar backscattering coefficient in dB, φ is the incidence angle, DN is the digital number of the amplitude image, and CF is the calibration factor. For ALOS-2 PALSAR-2 imageries, CF was set to -83dB (Shimada et al., 2009).

The ALOS-2 PALSAR-2 images continued to undergo a Lee speckle filter with a 3 × 3-pixel kernel to decrease speckle noise. For the S1 image, an array of processing procedures was implemented, encompassing orbit file application, removal of border and thermal noise, radiometric calibration, terrain correction, and conversion to the normalized radar γ^0 in dB (Filipponi, 2019). For S2, considering cloud cover, we used 09 level 1C products exclusively with cloud cover below 10 % to create a cloud-free mosaic. These S2 level 1C orthorectified images were converted into orthoimage bottom-of-atmosphere corrected reflectance level 2A data by the Sen2Cor plug-in, while cloud covers were also automatically created by the Idepix-assembly plug-in. The S2 images without cloud were subsequently created, mosaicked together, and processed by averaging using a 3 × 3-pixel mean filter to minimize noise or random variations in raster data. All RS data were resampled to 10 m spatial resolution using bilinear interpolation and co-registered to the pixel grid of the reference layer (band 2 from S2). To ensure spatial correspondence, the center of each sampling plot was aligned with the nearest pixel center of the reference layer for the extraction of corresponding satellite-derived values.

To study the potential of ALOS-2 PALSAR-2, S1, and S2 for forest AGBD estimation, we selected spectral bands, SAR backscatter coefficients, vegetation indices (VIs), biophysical variables (BIs), and textures as input variables. Specifically, the S2 spectral bands included band 2 (B2), band 3 (B3), band 4 (B4), band 5 (B5), band 6 (B6), band 7 (B7), band 8 (B8), band 8A (B8A), band 11 (B11), and band 12 (B12), while S1 provided VV and VH polarizations, and ALOS-2 PALSAR-2 contributed HH and HV polarizations. A total of 20 widely used VIs associated with forest AGBD estimation were calculated from the S2 mosaic (as shown in Table 2). Additionally, 03 BIs, including leaf area index (LAI), fraction of absorbed photosynthetically active radiation

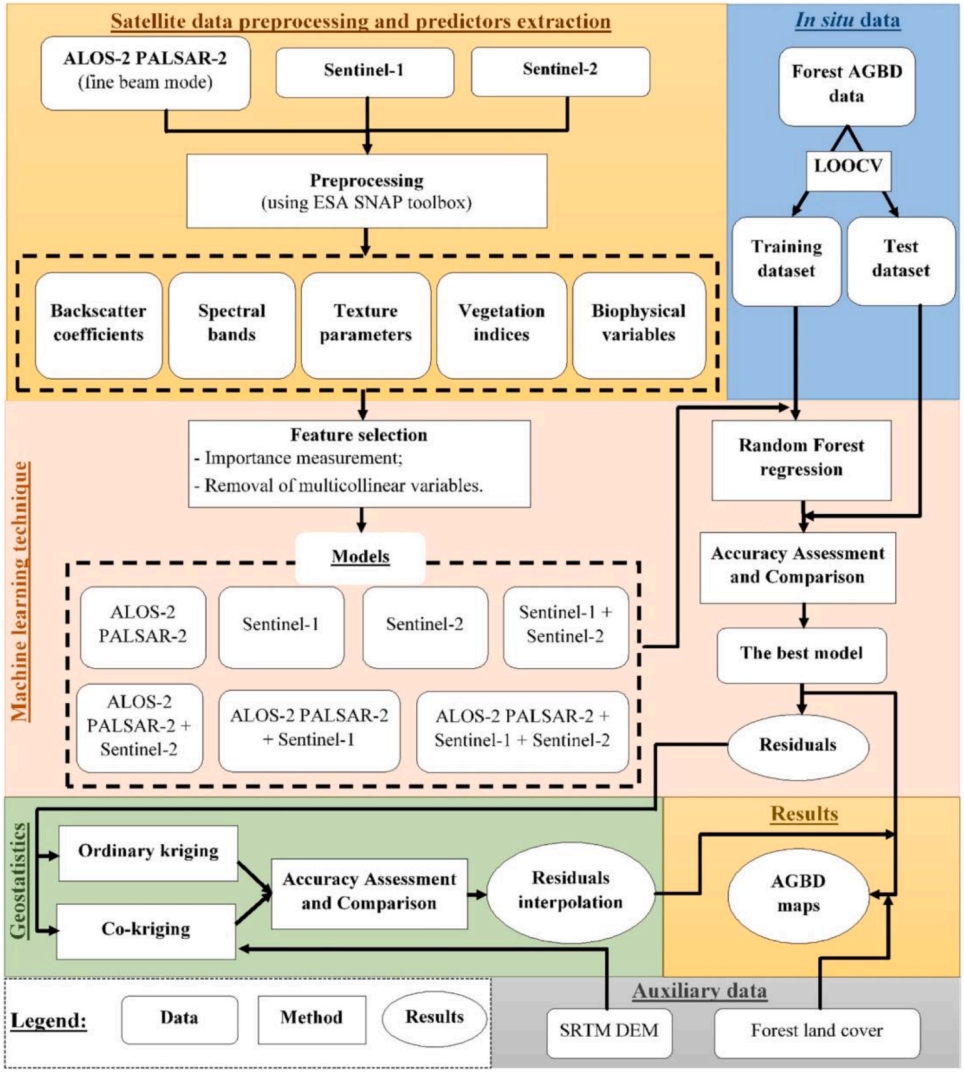

Fig. 3. The methodology workflow of the research.

(FAPAR), and fractional vegetation cover (Fcover), were also derived from S2 using SNAP's biophysical processor. These VIs and BVs have proven to correlate with forest AGBD in previous studies (Baloloy et al., 2018; Chen et al., 2019b, 2018; Jha et al., 2021). Besides, the gray-level co-occurrence matrix (GLCM) -derived texture features were utilized as predictors to enhance the accuracy of AGBD estimation because they can broaden the saturation (Berninger et al., 2018; H. Li et al., 2022; F. Zhang et al., 2023). In addition, testing various window sizes and evaluating their effect on model performance is necessary to optimize capturing relevant information and avoid noise or computational burden (Kelsey and Neff, 2014; Li et al., 2021; Sarker and Nichol, 2011). Thus, we calculated textures for ALOS-2 PALSAR-2 (HV, HH), S1 (VH, VV), and S2 (10m resolution bands: B2, B3, B4, and B8) utilizing GLCM analysis with three different window sizes of 5 × 5, 7 × 7, and 9 × 9 pixels. For each spectral band and SAR backscatter coefficient, ten GLCM texture metrics, including contrast (CONT), dissimilarity (DIS), homogeneity (HO), angular second moment (ASM), entropy (ENT), mean (ME), variance (VA), correlation (COR), energy (ENER) and maximum probability (MAX), were computed, following the definitions established by Haralick et al. (1973).

2.2.3. Topographic data

A 30-meter resolution digital elevation model (DEM) encompassing the entire study area was acquired from the shuttle radar topography mission (SRTM) dataset via the Google Earth Engine (GEE, https://code.earthengine.google.co.in/) platform. The elevation variable extracted from the DEM data was also resampled to a 10-meter spatial resolution and co-registered with the reference layer to ensure alignment with the other input datasets.

2.2.4. Land use/land cover data

A 10-m resolution LULC map for Danang city in 2023 was produced by integrating simple non-iterative clustering (SNIC) segmentation, time-series analysis of S1 and S2 imagery, and kernel principal component analysis (k-PCA) to enhance classification performance using the RF algorithm. The resulting LULC map classified the study area into six IPCC (2006) categories: forest land, cropland, grassland, wetlands, settlements, and other land. Classification accuracy was assessed using 427 ground truth points collected between July and September 2023, with 80 % of the samples used for model calibration and 20 % for validation. The RF classifier achieved an overall accuracy (OA) of 91.86 % and a Kappa coefficient (K) of 0.88, both of which exceed commonly accepted thresholds for reliable LULC classification (OA > 85 %, K > 0.80) (Nyamekye et al., 2021). These performance metrics indicate that the LULC map (Fig. 1) was sufficiently accurate for subsequent analysis in this study.

Table 2
List of Sentinel-2 vegetation indices.

Index	Description	Formula	Reference
NDVI	Normalized difference vegetation index	$(B8 - B4)/(B8 + B4)$	Tucker (Tucker, 1979)
RVI	Ratio vegetation index	$B8/B4$	Broge and Leblanc (2001)
PVI	Perpendicular vegetation index	$\left(1/\sqrt{0.149^2+1}\right) \times (B8 - 0.374 - 0.735)$	Richardsons and Wiegand (1977)
IPVI	Infrared percentage vegetation index	$B8/B8 + B4$	Crippen (1990)
WDVI	Weighted difference vegetation index	$B8 - (0.752 \times B4)$	Clevers (1989)
TNDVI	Transformed normalized difference vegetation index	$\sqrt{B8 - B4/B8 + B4} + 0.5$	Zhang et al. (2023)
GNDVI	Green normalized difference vegetation index	$B7 - B3/B7 + B3$	Moradi et al. (2022)
GEMI	Global environmental monitoring index	$w \times (1 - (0.25 \times w)) - B4 - 0.125/1 - B4$ $w = ((2 \times B8A^2 - B4^2)) + (1.5 \times B8A) + (0.5 \times B4))/(B8A + B4 + 0.5)$	Pinty and Verstraete (1992)
ARVI	Atmospherically resistant vegetation index	$B8 - B4 - 0.069(B4 - B2)/B8 + B4 - 0.069(B4 - B2)$	Kaufman and Tanre (1992)
NDI45	Normalized difference index	$B5 - B4/B5 + B4$	Kumar et al. (2021)
MTCI	Meris terrestrial chlorophyll index	$B6 - B5/B5 - B4$	Chen et al. (2019b)
MCARI	Modified chlorophyll absorption ratio index	$((B5 - B4) - 0.2(B5 - B3))(B5/B4)$	Chen et al. (2019b)
S2REP	Sentinel-2 red-edge position index	$705 + 35(B4 + B7/2 - B5/B6 - B5)$	Chen et al. (2019b)
IRECI	Inverted red-edge chlorophyll index	$(B7 - B4)/(B5/B6)$	Moradi et al. (2022)
PSSRa	Pigment specific simple ratio index	$B7/B4$	Blackburn (1998)
DVI	Difference vegetation index	$B8 - B4$	Moradi et al. (2022)
SAVI	Soil adjusted vegetation index	$(1 + 0.725) \times B8 - B4/B8 + B4 + 0.725$	Huete (Huete, 1988)
TSAVI	Transformed soil adjusted vegetation index	$(0.421 \times (B8 - (0.421 \times B4) - 0.824))/(B4 + 0.421(B8 - 0.824) + 0.114(10.421^2))$	Baret et al. (1989)
MSAVI	Modified soil adjusted vegetation index	$(1 + L)(B8 - B4)/B8 + B4 + L$ Where: L = 1 - 2 × s × NDVI × WDVI and s is the soil line slope	Qi et al. (1994)
MSAVI2	Modified soil adjusted vegetation index 2	$2B8 + 1 - \sqrt{(2B8+1)^2 - 8(B8 - B4)}/2$	Qi et al. (1994)

2.3. Modelling techniques and assessment

2.3.1. Random forest regression model

RF is a tree-based ML technique categorized under ensemble methods (Rincy and Gupta, 2020). The approach involves creating decision trees by randomly pulling several bootstrap samples with replacements from the initial training dataset. The unused portion of the training data can be used to compute the out-of-bag error. While implementing RF, each node in a tree randomly selects a subset of input predictors from the overall set to perform binary partitioning. This procedure is iterated across multiple trees, resulting in the creation of a substantial ensemble, and the predicted observation value is ascertained by calculating the average of the predictions from all trees within the ensemble (Mitchell, 2011; Wang et al., 2016).

Many predictor variables contribute little to model performance, leading to redundancy and increased model complexity, thereby selecting the most informative variables is essential to enhance model efficiency and interpretability (H. Li et al., 2022; Lu et al., 2016). RF can also be used to determine the significance of predictors through Gini importance values, which are the overall reduction in the impurity of all nodes averaged across all ensemble trees (Fan et al., 2022). A higher Gini importance suggests that the variable is more crucial in influencing the model's predicted accuracy. In our study, RF and attribute importance analysis were implemented using the Scikit-learn package. Moreover, an absolute Pearson correlation coefficient (r) greater than 0.8 was employed to detect multicollinearity among RS variables, as high multicollinearity can adversely impact prediction accuracy in regression analysis (Chen et al., 2019a; Luo et al., 2022).

2.3.2. Random forest regression kriging models

Since RF disregards spatial autocorrelation of the field-based AGBD, RFOK and RFCK were proposed to enhance the RF estimation. Specifically, kriging techniques, including OK and CK, were used to interpolate the RF-derived AGBD residuals, which were then added to the predicted values of AGBD from RF to obtain the new predictions (Jiang et al., 2022).

The OK approach is appropriate for interpolating the RF residuals of AGBD in regions influenced by the spatial variability of topography and climate (Su et al., 2020). For OK implementation, the assumption of data normality and stationarity needs to be satisfied (Meul and Van Meirvenne, 2003). In this study, we validated normality by skewness and kurtosis values, as well as the bell-shaped curves of the probability density function (PDF), and assessed stationarity by employing the intrinsic hypothesis (Oliver and Webster, 2014). The formula for OK interpolation is detailed below:

$$Z_{u,OK}(x_0) = \sum_{i=1}^{n} \lambda_{ui} Z_u(x_i) \tag{6}$$

Where $Z_{u,OK}(x_0)$ is the interpolated residual value of AGBD at location x_0, λ_{ui} is the weighting coefficient at location x_i, $Z_u(x_i)$ is the AGBD residual at location x_i, and n denotes the quantity of sample points utilized for interpolation.

CK is a spatial estimation technique that aims to provide the best possible estimate of a target variable at a given location by incorporating information from neighboring values of one or more additional variables, referred to as co-variables (Yalçin, 2005). CK is an extension of OK (Su et al., 2020) and adheres to the data normality and stationarity assumptions. Nevertheless, CK appears to enforce stricter criteria than OK when the co-variable must exhibit feature-space correlation with the target variable (Golden et al., 2020). The interpolation equation for the CK method is outlined below:

$$Z_{uv,CK}(x_0) = \sum_{i=1}^{n} \lambda_{ui} Z_u(x_i) + \sum_{j=1}^{m} \lambda_{vj} Z_v(x_j) \tag{7}$$

Where $Z_{uv,CK}(x_0)$ represents the residual value of the AGBD to be interpolated at location x_0, λ_{vj} is the weight assigned to co-variable at location x_j, $Z_v(x_j)$ refers to the value of co-variable at location x_j, and m is the number of co-variable samples. As the study area has considerable spatial terrain variation, the elevation data extracted from the DEM data was utilized as a co-variable.

Weighting coefficients in kriging are determined by fitting a semivariogram model for OK, and two semivariogram models along with a

cross-semivariogram model for CK (Webster and Oliver, 2007). This study employed spherical, exponential, and gaussian functions to determine nugget, sill, and range via the gstat package. The nugget-to-sill (N/S) ratio was utilized to assess the degree of spatial autocorrelation, where values below 0.25 indicate strong spatial dependence, values between 0.25 and 0.75 represent moderate spatial dependence, and values exceeding 0.75 suggest low spatial dependence (Li et al., 2020b).

2.3.3. Model assessment

A leave-one-out cross-validation (LOOCV) approach was used to calculate metrics for assessing the performance of predictive models. In LOOCV, during each iteration, one data point is excluded as the validation set, while the rest are utilized as the training set (Yue et al., 2018). Three evaluation metrics were chosen for model assessment: the mean absolute error (MAE), the root mean square error (RMSE), and the coefficient of determination (R^2). A model is considered superior performance if the RMSE and MAE values are lower while the R^2 value is greater (Zhang et al., 2020).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{10}$$

Where n is the number of observed values, y_i is the observed AGBD at the i^{th} location, $\hat{y}_i$ is the predicted AGBD at the i^{th} location, and $\bar{y}$ is the mean of the observed AGBD values.

Additionally, the relative improvement (RI) indices were used to quantify the extent to which the performance of a given model was enhanced compared to a baseline model (Zhang et al., 2022).

$$RI_{MAE/RMSE} = \frac{MAE/RMSE_b - MAE/RMSE_m}{MAE/RMSE_b} \times 100 \tag{11}$$

$$RI_{R2} = \frac{{R^2}_m - {R^2}_b}{{R^2}_b} \times 100 \tag{12}$$

Where RI_{MAE}, RI_{RMSE}, and RI_{R2} are the relative improvement indices of MAE, RMSE, and R^2, respectively; b and m represent the baseline model and model m, respectively. A positive RI suggests that the model m has improved compared to the baseline model, while a negative RI implies the opposite. Besides, greater values of the RI indicate more significant improvements.

3. Results and discussion

3.1. Descriptive statistics of forest AGBD

Table 3 provides the summary of descriptive statistics for forest AGBD in the research site. Among 104 observation data, the AGBD values ranged from 54.00 to 325.68 t.ha^{-1}, with a mean of 173.63 t.ha^{-1} and a standard deviation (SD) of 78.93 t.ha^{-1}. The coefficient of variation (CV) was 45.46 %, indicating a moderate level of relative dispersion around the mean. The distribution of AGBD data was slightly positively skewed (skewness = 0.24), suggesting a longer right tail, while the kurtosis value of -1.28 implies a relatively flatter distribution compared to the normal distribution. These descriptive statistics highlight the heterogeneity of AGBD in tropical forest ecosystems with complex stand structures.

3.2. Feature selection

In the importance analysis, the variables from various RS sources were ranked separately by the Gini importance. Fig. 4 shows the top 20 variables of each RS source identified by the Gini importance. In this study, to reduce input dimensionality, we chose the variables with a Gini importance exceeding 0.02. As a result, 12 out of 153 variables from S2, 12 out of 62 variables from S1, and 10 out of 62 variables from ALOS-2 PALSAR-2 were selected. Overall, GLCM texture variables accounted for the largest proportion of the better-performing predictors in all RS sources. Besides, MCARI from S2 and backscatters from S1 and ALOS-2 PALSAR-2 were also essential for AGBD estimation. Additionally, to mitigate the effect of multicollinearity on the performance of the ML technique, predictor variable pairs with absolute r values greater than 0.8 were identified, and the variable in each pair with the lower Gini importance value, as illustrated in Fig. 4, was removed. After multicollinear variable removal processing, five variables (B4_ENER_w9 × 9, B4_CONT_w9 × 9, B4_VA_w9 × 9, MCARI, and B4_COR_w9 × 9) from S2, three variables (VH_MAX_w7 × 7, VH_COR_w9 × 9, and VV_HO_w5 × 5) from S1 and two variables (HV_VA_w9 × 9 and HV_MAX_w5 × 5) from ALOS-2 PALSAR-2 were retained for RF implementation.

Many previous studies have shown that an improvement in forest AGBD estimation can be achieved by using GLCM-based texture measures (Berninger et al., 2018; Chen et al., 2019b; Eckert, 2012; H. Li et al., 2022; Salazar Villegas et al., 2023; F. Zhang et al., 2023). Our research results confirmed the undeniable role of texture indices in predicting AGBD. Three potential reasons can explain the importance of texture features in this study. Firstly, texture measurements can enhance the differentiation of spatial information regardless of tone and expand the saturation level for AGBD estimation in complex forest stand structures (Gao et al., 2018; Lu et al., 2016). Secondly, texture measures are able to effectively capture the heterogeneity of forest canopy structure (Eckert, 2012), as in our study area. The third reason may be attributed to the diverse species of hardwood trees present in the study site. Texture metrics are particularly demonstrated to be informative for characterizing a mixed forest where there is significant variation (Ghosh and Behera, 2018). Moreover, the study results show that GLCM-based textures derived from B2 and B4 of S2 data are vital to estimating AGBD, emphasizing the importance of textures generated from traditional bands, as the findings of Pandit et al. (2020). We also found that, among the selected important variables from SAR data, GLCM texture features derived from cross-polarization channels (VH from S1 and HV from ALOS-2 PALSAR-2) accounted for a larger proportion than those derived from co-polarization channels (VV from S1 and HH from ALOS-2 PALSAR-2). It is compatible with the results of previous studies (H. Li et al., 2022; Salazar Villegas et al., 2023). This is because cross-polarized waves are supposed to be more responsive to biomass than co-polarized ones (Salazar Villegas et al., 2023; Sinha et al., 2015). According to Lu et al. (2016), in light of the complexity of stand structure in forest landscapes, the relationships between VIs and AGBD can differ.

Table 3
Descriptive analysis of forest AGBD in the research site.

Parameter	N	Min (t.ha^{-1})	Max (t.ha^{-1})	Mean (t.ha^{-1})	SD (t.ha^{-1})	CV (%)	Skewness	Kurtosis
AGBD	104	54.00	325.68	173.63	78.93	45.46	0.24	-1.28

Where N is the number of AGBD observations; Min and Max refer to the minimum and maximum values, respectively; SD is the standard deviation; and CV is the coefficient of variation.

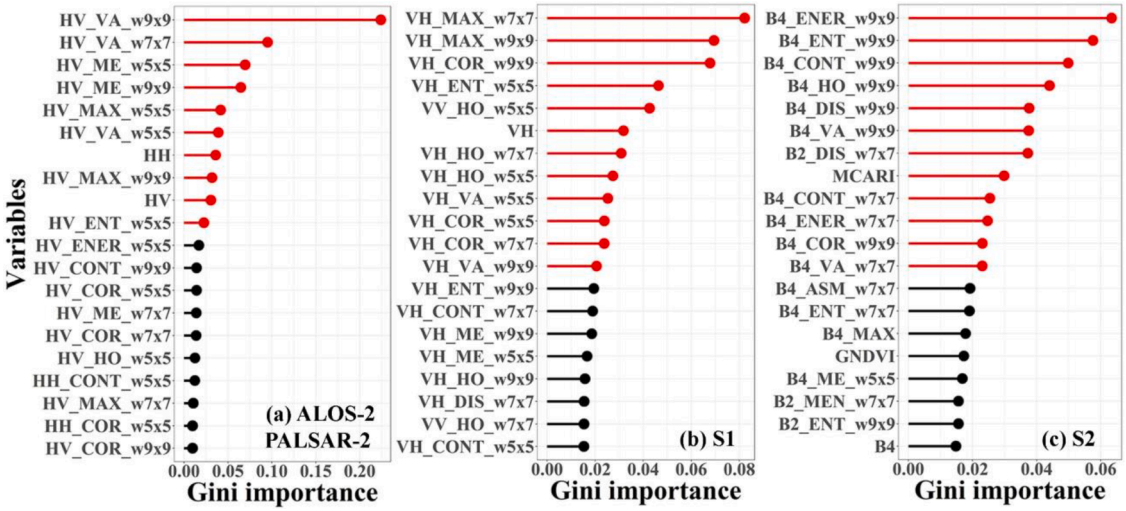

Fig. 4. Gini importance of the predictors derived from ALOS-2 PALSAR-2 (a), S1 (b), and S2 (c) for AGBD estimation. w is the window size of GLCM texture analysis, and red objects illustrate the variables retained after importance ranking.

Although many S2 VIs have been proposed for AGBD estimation in previous research (Chen et al., 2019b; Mngadi et al., 2021; F. Zhang et al., 2023), MCARI was the only vegetation index selected after the importance ranking in our study. The weak performance of VIs could have originated from the average tree height of 17.5m at the study site (Huy et al., 2016), which may have increased the shadow content due to the heterogeneous canopy, causing S2 spectral measurements to become saturated and less effective in discriminating different biomass values (Pandit et al., 2020). The importance of MCARI in our study is harmonious with the result of Chen et al. (2019b), who claimed that in RF using S2 data, MCARI was the most essential predictor variable. This is attributed to the high correlation between MCARI and chlorophyll concentration, which can be explained by the non-deciduous phenomena of the evergreen forests, facilitating maximum chlorophyll absorption (Maleki et al., 2020). Furthermore, unlike traditional broadband VIs, MCARI, based on red-edge bands, exhibited greater sensitivity to dense vegetation structures while being less influenced by spectral saturation (Mngadi et al., 2021; Wai et al., 2022). HH and HV polarizations from L-band ALOS-2 PALSAR-2 also displayed a significant contribution to AGBD estimation in the study. The results of our study align with existing literature. Mitchard et al. (2012) found a strong correlation between AGBD and L-band polarizations, with saturation thresholds of approximately 100 $t.ha^{-1}$ for HH polarizations and around 150 $t.ha^{-1}$ for HV polarizations in tropical rainforests. However, other authors have proposed higher saturation values for the L-band. Specifically, in the research about L-band saturation level for AGBD in Malaysian tropical forests, Hamdan et al. (2015) found that HV showed a good relationship with AGBD below 200 $t.ha^{-1}$ and appeared to reach saturation at 200 $t.ha^{-1}$. The saturation level for the L-band backscatter even showed a higher figure in the results of Lucas et al. (2010), in which HH and HV saturated at approximately 170 $t.ha^{-1}$ and 272 $t.ha^{-1}$ in Australian tropical rainforests. When it comes to S1 backscatter data, although being a significant variable of S1 for AGBD estimation in our study, VH had a negative impact on predictive models after saturating at 50 $t.ha^{-1}$ (Englhart et al., 2011; Huang et al., 2018).

3.3. Random forest regression modelling

After running RF for seven models of RS source combination, MAE, RMSE, and R^2 calculated from LOOCV are shown in Table 4. Among single-source models, model 1 containing S2 variables outperformed the others for AGBD prediction when it achieved the highest R^2 of 0.62 and the lowest RMSE and MAE, at 48.73 and 39.50 $t.ha^{-1}$, respectively. Thus, Model 1 was chosen as the benchmark for evaluating the relative improvement of other models.

Regarding relative improvement, the values of RI_{R2}, RI_{RMSE}, and RI_{MAE} were all negative in model 2 (S1), model 3 (ALOS-2 PALSAR-2), model 4 (ALOS-2 PALSAR-2 + S1), and model 5 (S1 + S2), indicating a decline in model performance relative to the baseline model (S2). Meanwhile, model 6 (ALOS-2 PALSAR-2 + S2) and model 7 (ALOS-2 PALSAR-2 + S1 + S2) exhibit positive patterns, reflecting improved predictive accuracy. Notably, the maximum RI values of 22.17 %, 19.65 %, and 18.49 % for R^2, RMSE, and MAE, respectively, achieved in model 6, point out that the merging of ALOS-2 PALSAR-2 and S2 was optimal for RF. Therefore, residuals from model 6 (ALOS-2 PALSAR-2 + S2) were utilized in the subsequent kriging-based spatial prediction.

In dual-source combinations, model 4 (ALOS-2 PALSAR-2 + S1) achieved an R^2 of 0.46, RMSE of 57.76 $t.ha^{-1}$, and MAE of 46.21 $t.ha^{-1}$, while model 6 (ALOS-2 PALSAR-2 + S2) obtained superior performance with an R^2 of 0.75, RMSE of 39.15 $t.ha^{-1}$, and MAE of 32.20 $t.ha^{-1}$. Both models outperformed their respective single-source counterparts, underscoring the effectiveness of RS data fusion in improving AGBD estimation accuracy. Consistent with findings by Vafaei et al. (2018), Li et al. (2022), and Salazar Villegas et al. (2023), the integration of S2 and

Table 4
Evaluation metrics of RF based on different RS data fusions.

Models	Variables	MAE ($t.ha^{-1}$)	RMSE ($t.ha^{-1}$)	R^2	RI_{MAE} (%)	RI_{RMSE} (%)	RI_{R2} (%)
Model 1	S2	39.50	48.73	0.62	-	-	-
Model 2	S1	61.14	72.93	0.14	−54.77	−49.67	−77.59
Model 3	ALOS-2 PALSAR-2	47.91	59.30	0.43	−21.29	−21.70	−30.10
Model 4	ALOS-2 PALSAR-2 + S1	46.21	57.76	0.46	−16.97	−18.54	−25.35
Model 5	S1 + S2	38.99	48.97	0.61	1.31	−0.49	−0.61
Model 6	ALOS-2 PALSAR-2 + S2	32.20	39.15	0.75	18.49	19.65	22.17
Model 7	ALOS-2 PALSAR-2 + S1 + S2	33.09	41.55	0.72	16.23	14.73	17.07

ALOS-2 PALSAR-2 proved to be particularly effective. The enhancement in RF can be ascribed to the advantages of SAR-MSI fusion in alleviating the issue of mixed pixels and the problem of data saturation (Ghasemi et al., 2011; Lu, 2006). Likewise, the fusion of S1 and ALOS-2 PALSAR-2 obtained a more acceptable performance of RF than a single one. The application of a multi-frequency SAR dataset to improve forest AGBD predictions is extensively reported in the literature. Sivasankar et al. (2018) pointed out that the combined use of L- and C-band cross-polarization backscatter can correct mistakes in lower biomass areas for the L-band and errors in larger biomass areas for the C-band. In addition, the research by Berninger et al. (2018) stated that the C-band is more sensitive to surface roughness variations, leading to better modelling in burned and grass areas, thereby incorporating the C-band alongside the L-band marginally enhances the correlation between estimated and predicted AGBD relative to utilizing the L-band in separation. In contrast, the combination of S1 and S2 data in model 5 yielded an R^2 of 0.61, RMSE of 48.97 t.ha^{-1}, and MAE of 38.99 t.ha^{-1}, which was less accurate than using S2 data alone, indicating that the inclusion of S1 did not enhance predictive performance in our study. On the one hand, this finding conflicts with Forkuor et al. (2020), David et al. (2022), Zhang et al. (2023), Zhang et al. (2023), and Salazar Villegas et al. (2023). On the other hand, the result is consistent with Stratoulias et al. (2022), who posited that the decreased accuracies could be due to either the used indices or the predictive model's performance that might not fully utilize the information acquired from S1 and S2 data.

When combining all three data sources (model 7), RF acquired a notably accurate result with R^2 of 0.72, RMSE of 41.55 t.ha^{-1}, and MAE of 33.09 t.ha^{-1}. As expected, triple-source data outweighed both single-source and dual-source data, except for model 6 (ALOS-2 PALSAR-2 + S2). This finding indicates that fusing all satellite data (optical MSI and multi-frequency SAR) within a single model did not constantly improve AGBD estimation. The findings align with previous research. The best example can be observed in the study of Kanmegne Tamga et al. (2022), who used VIs and texture variables extracted from MSI (S2) and SAR data (S1 and ALOS-2 PALSAR-2) as predictors to predict AGBD in vegetated areas. After implementing the predictive models, they suggested that the fusion of S1 and S2 data returned a better RF performance than all three satellite datasets. Another example could be the research of Laurin et al. (2018), who also estimated AGBD by S1, ALOS-2 PALSAR-2, and S2 data. However, when adding S2 predictors to multi-frequency SAR data, the ALOS-2 PALSAR-2 predictors were eliminated from the predictive model.

Besides, the study results indicate that the addition of S1 data into the RF models, based on either single S2 data or the combination of S2 and ALOS-2 PALSAR-2 data, was inclined to worse performances due to the short wavelength (5.3 cm) C-band limitation in forest AGBD estimation (Huang et al., 2018). As mentioned, C-band S1 demonstrates saturation at roughly 50 t.ha^{-1} (Huang et al., 2018), occurring at a markedly lower AGBD than ALOS-2 PALSAR-2, which attains saturation above 250 t.ha^{-1} (Lucas et al., 2010) because C-band has less penetration and six times higher attenuation coefficient than L-band (Huang et al., 2018). Furthermore, sensitivity to large AGBD values is much higher for S2 compared to S1 since S2 saturation points can reach 204 t.ha^{-1} with the inclusion of the red-edge region (Jha et al., 2021). Even using the texture features, the gap between S1 and S2 or ALOS-2 PALSAR-2 in saturation levels was still huge. In other words, C-band S1 is solely appropriate for subtropical forests with low biomass, whereas S2 and ALOS-2 PALSAR-2 are prone to be better suited in tropical forests (Laurin et al., 2018). Therefore, in our study site with a complex forest stand structure, the S1 variables soon exhibited a negative correlation with AGBD, decreasing the performance of the predictive models.

3.4. Random forest regression kriging modelling

3.4.1. Residuals of random forest-derived AGBD and semivariogram analysis

The spatial distribution of the RF-derived residuals from model 6 (ALOS-2 PALSAR-2 + S2) is shown in Fig. 5a, and their descriptive statistics are summarized in Fig. 5c. The residuals ranged from −88.90 to 116.23 t.ha^{-1}, and the mean value was −1.03 t.ha^{-1}. In addition, the skewness value was close to 0, and the PDF had a bell-shaped curve, suggesting that the residuals followed an approximately normal distribution. However, 0.20 of the absolute kurtosis value showed that the curve had a lighter tail and a flatter peak. Therefore, we subsequently transformed residual data by Box-Cox transformation with a lambda of 0.85 to obtain a closer normal distribution of data (skewness = −0.11, kurtosis = −0.12) (Fig. 5e), as well as provide the same scale with the transformed elevation variable in later CK analysis. Furthermore, the semivariogram cloud of Box-Cox transformed RF residuals (Fig. 6a) exhibited no trend or systematic variation in the semivariogram, indicating that the data satisfied the intrinsic stationarity assumption. After confirming the approximate normality and stationarity, the Box-Cox transformed RF residuals were used to fit the experimental semivariogram for OK interpolation. Table 5 illustrates fitted theoretical semivariogram models and their evaluation metrics. Among the used theoretical semivariogram models, spherical achieved the highest value of R^2 (0.20) and the lowest values of RMSE (18.13) and MAE (14.13), suggesting the most suitable for kriging. Thus, we used the spherical function for the semivariogram analysis. The high nugget value of 220.04 in the spherical model of the transformed residuals, visible in Fig. 6b, which should ideally be zero, indicated a considerable error in the data attributable to the short-scale variability. In addition, the N/S value of 0.52 indicated the presence of a moderate spatial autocorrelation structure existing in the transformed residuals.

The elevation of sampling plots extracted from SRTM DEM was used as a co-variable in CK. Nonetheless, elevation data did not follow the normal distribution pattern and skewed to the right (skewness = 0.70, kurtosis = 0.14) (Fig. 5d). Hence, we also used Box-Cox transformation to turn elevation into approximately normally distributed data (skewness = −0.09, kurtosis = −0.98) (Fig. 5f). The transformed elevation had a constant semivariance (Fig. 6d), as well as there was a feature-space linear correlation between the target variable (Box-Cox transformed RF residuals) and co-variable (Box-Cox transformed elevation) with $r = 0.21$ (p-value < 0.05) (Fig. 6c), implying the Box-Cox transformed elevation could be used for CK. We fitted the direct semivariograms and the cross-semivariogram (Fig. 6b, e, f) simultaneously by a linear model of co-regionalization, which used the spherical model and the same range of 5056.71 m, but different partial sills and nuggets to achieve positive definiteness in the CK system. Based on Table 5 and Fig. 6, the distribution of the transformed RF residuals of forest AGBD, interpolated using OK and CK, was obtained and then converted back to the initial scale. The interpolated maps of residuals from OK and CK are shown in Fig. 9b, d.

3.4.2. Forest AGBD estimation via random forest regression kriging

The spatial distribution, data distribution, and description of the residuals from RFOK and RFCK are shown in Fig. 7. The residuals from both models exhibited an approximately normal distribution, revealing that the models still fit the input data well. Table 6 shows the evaluation metrics of RF, RFOK, and RFCK. Comparing RF and the hybrid models revealed that the addition of residuals interpolated by OK or CK yielded more accurate estimations of forest AGBD than solely considering the effects of predictor covariates from RS data. The R^2 values for RFOK and RFCK reached 0.80 and 0.79, respectively, both higher than that of RF, while their RMSE values (34.88 and 35.73 t.ha^{-1}, respectively) and MAE values (27.28 and 27.81 t.ha^{-1}, respectively) were lower than those for RF, indicating their superior predictive performance.

Moreover, as illustrated in Fig. 7d, the scatter points of RFOK

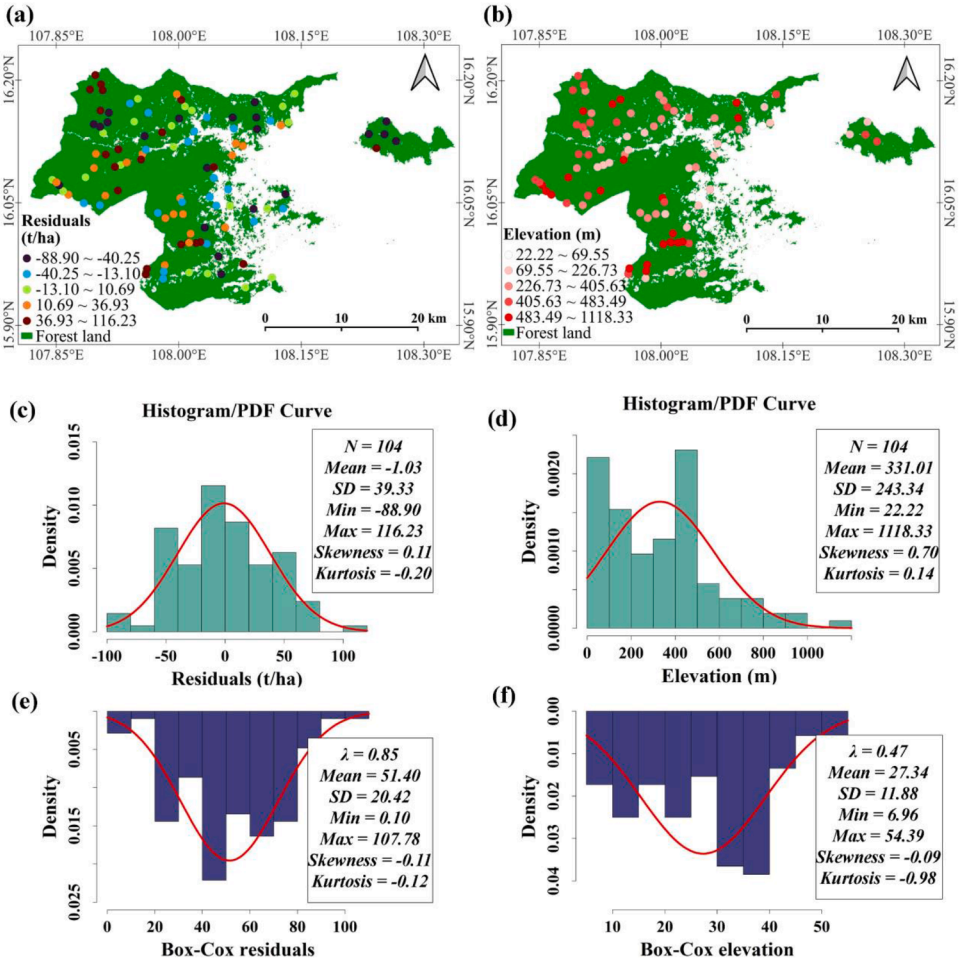

Fig. 5. Spatial distributions and histograms/PDF of residuals and elevation. (a, b) are spatial distributions of residuals and elevation, respectively; (c, d) are histograms/PDF of residuals and elevation, respectively; (e, f) are histograms/PDF of transformed residuals and elevation, respectively. λ denotes the Box-Cox transformation parameter.

(brown) are generally closer to the 1:1 line than those of RF (gray), suggesting an enhanced prediction accuracy for RFOK. However, the magnitude of this improvement was limited, as evidenced by the relatively low values of RI_{R2} (7.04 %), RI_{RMSE} (10.91 %), and RI_{MAE} (15.27 %). This finding aligns with previous studies, which similarly reported slight improvements when integrating RF with OK for estimating forest AGBD. For instance, Chen et al. (2019a) conducted a study in temperate continental forests in northeastern China and reported relatively low improvement in RI_{RMSE} values ranging from 6.90 % to 12.00 %, depending on the specific fusion of RS data sources. Similarly, Su et al. (2020), investigating subtropical forests in northern Guangdong Province, China, found RI_{RMSE} values of 17.60 % in 1992, 5.00 % in 2002, and 3.10 % in 2010. In tropical forest environments of Myanmar, Wai et al. (2022) reported RI_{RMSE} values of 2.10 % for evergreen forests and 13.40 % for deciduous forests. In contrast, Jiang et al. (2022), conducting research in subtropical forests in Chifeng City, China, observed significant improvements when applying RFOK, with RI_{RMSE} values of 55.95 % using Landsat 8-derived variables and 51.08 % using Landsat 9-derived variables. The limited improvement observed in our study is likely attributable to the moderate spatial dependence of the residuals derived from RF (Su et al., 2020), as indicated by the N/S value of 0.52, which falls within the range of 0.25 to 0.75 (Li et al., 2020b), as reported in Table 5. Two possible factors could account for the observed spatial dependence structure. Firstly, the wide spacing between sampling plots hindered the ability of kriging to effectively capture and model spatial autocorrelation (Wai et al., 2022). Theoretically, implementing denser sampling schemes could potentially enhance the spatial structure of residuals (Viana et al., 2012), thereby improving the accuracy of spatial predictions, but such efforts remain challenging in tropical forest landscapes as our study site, due to accessibility constraints attributed to terrain complexity and dense vegetation (Saatchi et al., 2011; Sinha et al., 2015). Secondly, forest parameters, including AGBD, are more strongly influenced by regionalized variables like soil characteristics and nutrient availability than by the AGBD values of neighboring sites, leading to a limited level of inherent spatial autocorrelation (Chen et al., 2019a; Viana et al., 2012).

In particular, RFOK recorded higher values of RI_{R2}, RI_{RMSE}, and RI_{MAE} than its bivariate counterpart. Specifically, the former had values of 7.04 %, 10.91 %, and 15.27 % for RI_{R2}, RI_{RMSE}, and RI_{MAE}, respectively, while the latter had values of 5.74 %, 8.73 %, and 13.62 %, respectively, suggesting that RFOK had marginally better improvement than RFCK in AGBD prediction. It was also confirmed through Fig. 7g, in which the scatter points of RFOK (brown) slightly tended to cluster nearer to the 1:1 line compared to those of RFCK (blue). The possible explanation may stem from the weak feature-space linear correlation between the target variable and the co-variable (Fig. 6c). This finding is consistent with Wai et al. (2022), who similarly estimated the AGBD of tropical evergreen forests and reported that while RFCK offered higher accuracy than initial RF, it was still lower than RFOK. However, the results contradict Su et al. (2020), who pointed out that RFCK provided a more accurate method than RFOK for AGBD mapping with complex topography. The discrepancy is ascribed to the difference in study sites

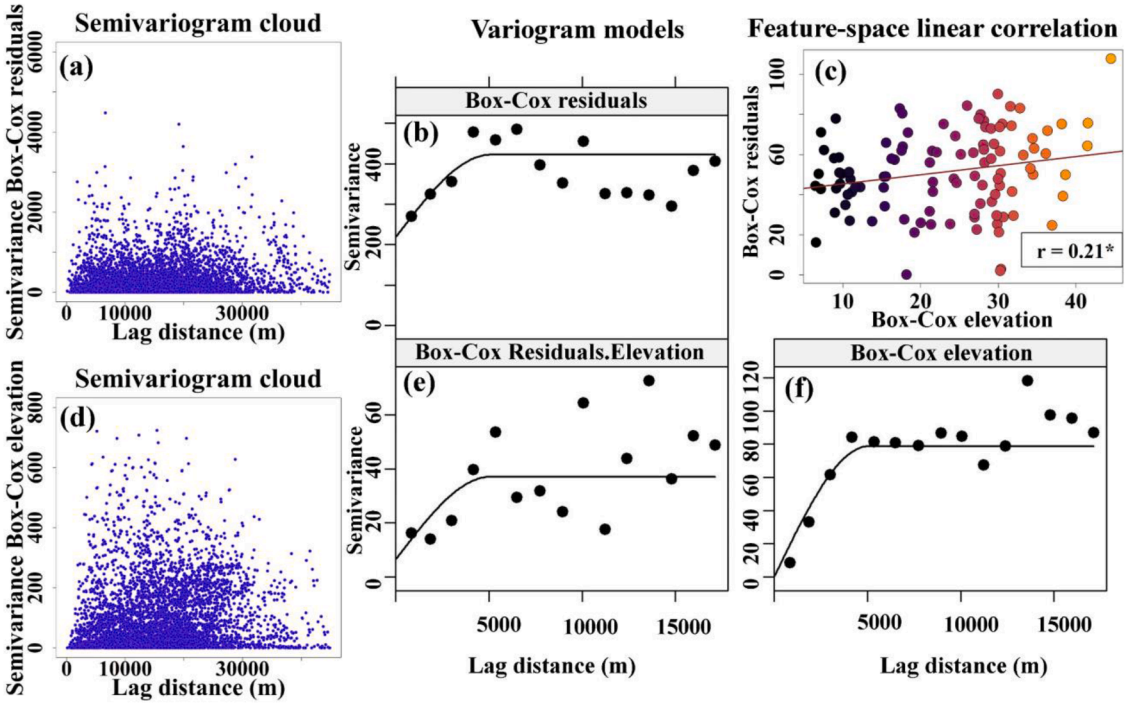

Fig. 6. Semivariogram clouds (a, d), feature-space linear correlation (c, '*' means that p-value was below 0.05, and colors transition from dark to bright as elevation increases), and semivariogram modelling of co-regionalization (b, e, f) for CK.

Table 5
Parameter estimations for semivariogram analysis.

Theoretical models	Nugget	Sill	N/S	Range (m)	MAE	RMSE	R^2
Exponential	152.08	416.25	0.37	1476.89	14.24	18.21	0.20
Gaussian	249.95	421.22	0.59	2434.92	14.17	18.13	0.20
Spherical	220.04	421.31	0.52	5056.71	14.13	18.13	0.20

and the RS data employed. Specifically, the study site of Su et al. (2020) was situated within subtropical forests, which typically exhibit simpler forest structures with lower biomass levels compared to tropical forests in our study site, and Landsat data was used instead of S2 data.

The results in Fig. 8 illustrate that, in RF, RFOK, and RFCK, negative residuals were inclined to cluster at low observed AGBD values, while positive residuals were more common at high observed AGBD values. This pattern reveals a tendency for overestimation at low AGBD levels and underestimation at large AGBD levels across all three models. The underestimation of high AGBD values was attributed to signal saturation in both optical MSI and SAR sensors under dense biomass conditions, whereas overestimation in low AGBD areas likely resulted from mixed pixels comprising shrubs, grasses, and bare soil, which influence the spectral reflectance characteristics of RS data (Li et al., 2020b). However, the hybrid approaches appeared to alleviate these issues more effectively than RF, as indicated by the lower slopes of their fitted lines (0.22 and 0.19 for RFOK and RFCK, respectively, compared to 0.29 for RF). Likewise, based on the slope values, RFCK exhibited a slightly improved ability to mitigate overestimation in low biomass regions and underestimation in high biomass areas compared to its univariate counterpart, despite yielding lower overall prediction accuracy. This improvement may be attributed to the use of elevation as a covariate in the CK interpolation process. In our study area, elevation was positively correlated with AGBD, suggesting that underestimation and overestimation were more pronounced at higher and lower elevations, respectively. Unlike RFOK, which only accounts for spatial autocorrelation, RFCK additionally captures terrain-related variability, which likely contributed to its more effective handling of estimation biases in ecologically complex areas, particularly in transitional zones between mountainous and lowland regions (Su et al., 2020).

3.5. Spatial distribution of forest AGBD

The distribution of forest AGBD using RFOK was acquired by combining Fig. 9a and 9b, whereas the distribution based on RFCK was obtained by integrating Fig. 9a and 9d. Two maps demonstrated a comparable distribution of AGBD with RF, while values varied significantly (Fig. 9a, c, e). The maps generated by RFOK and RFCK displayed a more realistic spatial distribution of forest AGBD than RF. In other words, the addition of the kriged residuals from OK and CK interpolation provided smoother spatial surfaces of forest AGBD, mitigating erratic patterns in the deterministic trend produced by RF. Aside from assessing the prediction accuracy, the model's generalization capability was also taken into account by analyzing the range of AGBD values in the predicted maps. The ranges of predicted AGBD using RF, RFOK, and RFCK were 73.15-299.19 $t.ha^{-1}$, 39.82-331.29 $t.ha^{-1}$, and 41.11-330.23 $t.ha^{-1}$, respectively. The broader range of values showed that RFOK and RFCK had enhanced generalization capabilities and greater robustness than RF. The spatial distribution of AGBD exhibited a decrease from west to east. The highest AGBD values were found in the western part of the city, predominantly covered by natural evergreen broadleaf forests in the high-altitude mountain regions protected by the Danang Department of Forest Protection. Meanwhile, the lowest AGBD values were located in the eastern region, where lower vegetation coverage and intensive

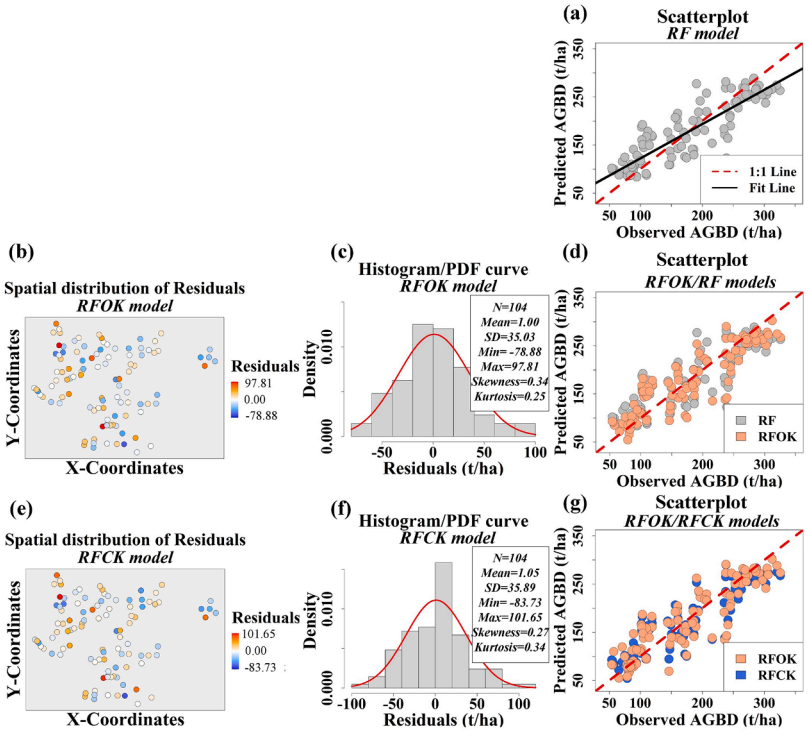

Fig. 7. Spatial distributions and histograms of residuals from RFOK (b, c) and RFCK (e, f); Scatterplots of observed and predicted AGBD for RF (a), RFOK/RF (d), and RFOK/RFCK (g).

Table 6
Evaluation metrics of RFOK and RFCK.

Models	MAE ($t.ha^{-1}$)	RMSE ($t.ha^{-1}$)	R^2	RI_{MAE} (%)	RI_{RMSE} (%)	RI_{R2} (%)
RF	32.20	39.15	0.75	-	-	-
RFOK	27.28	34.88	0.80	15.27	10.91	7.04
RFCK	27.81	35.73	0.79	13.62	8.73	5.74

human activities prevail.

4. Conclusion

This study investigates the potential of RS derivatives from ALOS-2 PALSAR-2, S1, and S2 data, in combination with field measurements, for modelling AGBD using RF in tropical evergreen broadleaf forests of Central Vietnam. In addition, RFOK and RFCK were employed to improve prediction accuracy by addressing the lack of spatial autocorrelation in RF. Through the study results, it is concluded that: (1) the predictor variables, including texture features, MCARI, and radar backscatters, were significantly influential in predicting forest AGBD; (2) RF integrating ALOS-2 PALSAR-2 and S2 data achieved the highest predictive performance (R^2 = 0.75, RMSE = 39.15 $t.ha^{-1}$, and MAE = 32.20 $t.ha^{-1}$), outperforming not only models using single-source datasets but also those employing combinations of ALOS-2 PALSAR-2 with S1 data, S1 with S2 data, and even ALOS-2 PALSAR-2 with S1 and S2 data; (3) Although RFOK and RFCK provided relative improvements over RF, these gains were still limited, as indicated by the relatively low RI values (5.74–7.04 % in RI_{R2}, 8.73–10.91 % in RI_{RMSE}, and 13.62–15.27 % in RI_{MAE}); (4) the hybrid approaches yielded more spatially realistic AGBD maps compared to RF; and (5) while RFOK produced marginally higher accuracy in AGBD estimation within tropical evergreen broadleaf forests (R^2 = 0.80, RMSE = 34.88 $t.ha^{-1}$, MAE = 27.28 $t.ha^{-1}$) compared to RFCK (R^2 = 0.79, RMSE = 35.73 $t.ha^{-1}$, MAE = 27.81 $t.ha^{-1}$), RFCK demonstrated better capacity to reduce overestimation in low biomass areas and underestimation in high biomass regions. We highlight that the hybrid approaches integrating ALOS-2 PALSAR-2 and S2 data in our study provide a reliable framework for

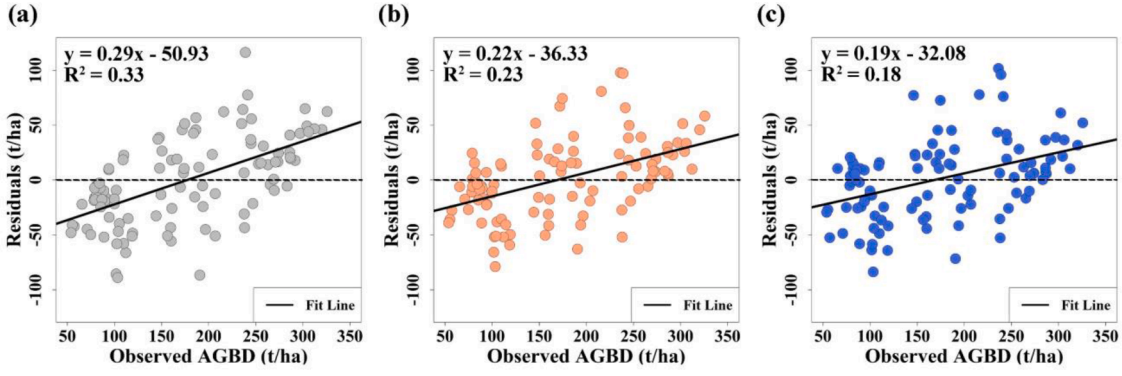

Fig. 8. Scatter plots of residuals and observed AGBD values for RF (a), RFOK (b), and RFCK (c).

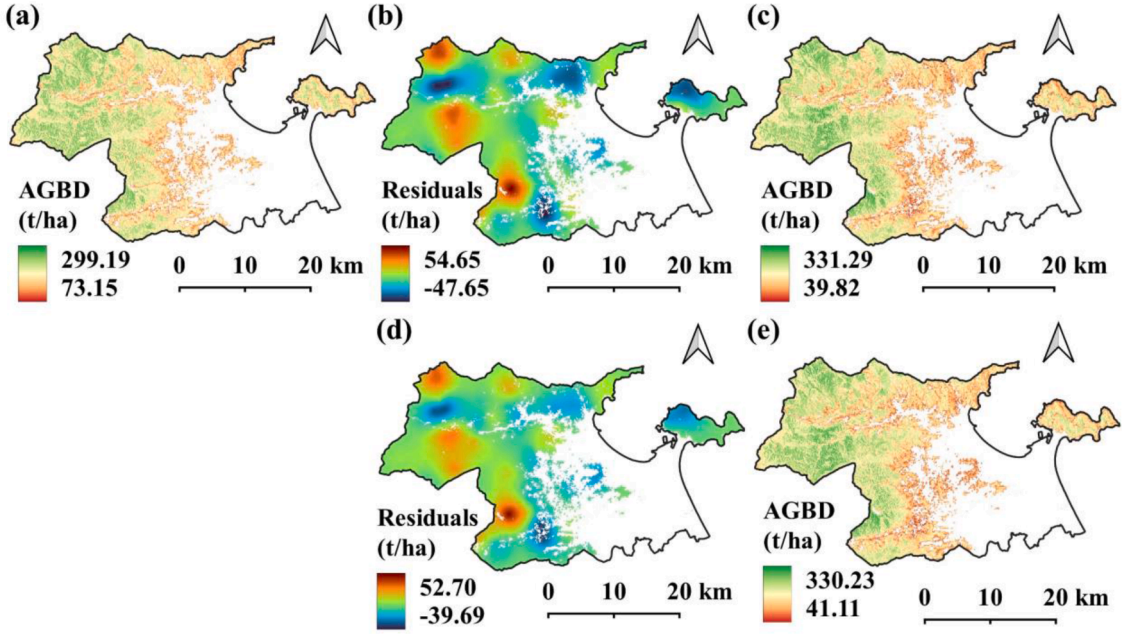

Fig. 9. Estimated AGBD maps at 10m spatial resolution using RF (a), RFOK (c), and RFCK (e); the interpolated residuals from OK (b) and CK (d).

spatial mapping of AGBD, which can support more frequent monitoring and accurate carbon accounting within tropical forests.

Future research should aim to enhance the capacity to detect spatial autocorrelation of model residuals, a critical factor in maximizing the effectiveness of kriging-based hybrid approaches in forest AGBD estimation. This may be achieved through the implementation of denser sampling schemes, which can better capture local variability and spatial structure. Additionally, to strengthen the spatial structure of residuals and improve interpolation performance in CK, it is recommended to explore alternative environmental variables, such as soil characteristics or climatic data, as covariates. The potential of upcoming SAR missions, including the BIOMASS mission (Europe, P-band, scheduled for launch in 2025), the NISAR mission (USA, L-band, scheduled for 2025), and the TanDEM-L mission (Germany, L-band, expected in 2028), should also be investigated for their capability to reduce signal saturation in dense tropical forests. Lastly, the practical validity of the hybrid modelling frameworks needs to be further verified, especially their applicability under different geological and environmental conditions.

CRediT authorship contribution statement

Viet Hoang Ho: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Hidenori Morita:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Felix Bachofer:** Writing – review & editing, Methodology. **Thanh Ha Ho:** Writing – review & editing, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors thank the staff of the Danang Forest Protection Department and the colleagues of Hue University of Agriculture and Forestry, who participated in the field surveys and data collection.

Appendix A. List of variable notations

Variable notation	Description	Unit
Field-based AGBD measurement		
AGB_{tree}	Aboveground biomass of a single tree	kg
DBH	Diameter at breast height of a tree	cm
$AGBD_{tree}$	Tree aboveground biomass density	$t.ha^{-1}$
A_{plot}	Area of the sample plot (30 × 30 m)	m^2
HBD	Herb biomass density	$t.ha^{-1}$
SBD	Shrub biomass density	$t.ha^{-1}$
$W_{herb\ field}$	Total fresh weight of herb sample measured in the field	kg
$W_{shrub\ field}$	Total fresh weight of shrub sample measured in the field	kg
$A_{subplots}$	Total area of subplots (3 × 1 × 1m) used for herb/shrub sampling	m^2
$W_{herb-subsample(dry)}$	Oven-dried weight of herb subsample	g
$W_{shrub-subsample(dry)}$	Oven-dried weight of shrub subsample	g
$W_{herb-subsample(fresh)}$	Fresh weight of herb subsample	g
$W_{shrub-subsample(fresh)}$	Fresh weight of shrub subsample	g
$AGBD$	Aboveground biomass density	$t.ha^{-1}$
SAR data transformation		

(*continued on next page*)

(*continued*)

Variable notation	Description	Unit
σ^0	Radar backscattering coefficient	dB
γ^0	Normalized radar backscattering coefficient	dB
DN	Digital number (pixel value of amplitude image)	-
CF	Calibration factor for SAR images	dB
φ	Incidence angle	Degrees (^{0})
VIs calculation		
B2	Band 2 (Blue, 490 nm) reflectance	-
B3	Band 3 (Green, 560 nm) reflectance	-
B4	Band 4 (Red, 665 nm) reflectance	-
B5	Band 5 (Red edge, 705 nm) reflectance	-
B6	Band 6 (Red edge, 749 nm) reflectance	-
B7	Band 7 (Red edge, 783 nm) reflectance	-
B8	Band 8 (Near infrared, 842 nm) reflectance	-
B8A	Band 8A (Near infrared, 865 nm) reflectance	-
B11	Band 11 (Short wave infrared, 1610 nm) reflectance	-
B12	Band 12 (Short wave infrared, 2190 nm) reflectance	-
NDVI	Normalized difference vegetation index	-
RVI	Ratio vegetation index	-
PVI	Perpendicular vegetation index	-
IPVI	Infrared percentage vegetation index	-
WDVI	Weighted difference vegetation index	-
TNDVI	Transformed normalized difference vegetation index	-
GNDVI	Green normalized difference vegetation index	-
GEMI	Global environmental monitoring index	-
ARVI	Atmospherically resistant vegetation index	-
NDI45	Normalized difference index	-
MTCI	Meris terrestrial chlorophyll index	-
MCARI	Modified chlorophyll absorption ratio index	-
S2REP	Sentinel-2 red-edge position index	-
IRECI	Inverted red-edge chlorophyll index	-
PSSRa	Pigment specific simple ratio index	-
DVI	Difference vegetation index	-
SAVI	Soil adjusted vegetation index	-
TSAVI	Transformed soil adjusted vegetation index	-
MSAVI	Modified soil adjusted vegetation index	-
MSAVI2	Modified soil adjusted vegetation index 2	-
L	Canopy background adjustment constant (in MSAVI formula)	-
s	Soil line slope (used in MSAVI formula)	-
w	Constant used in GEMI index formula	-
Kriging analysis		
n	Number of sample points used for kriging interpolation	-
m	Number of co-variable sample points used for CK	-
x_0	Target location for interpolation	m
x_i	Locations of sample points for residuals	m
x_j	Locations of sample points for co-variable data (elevation)	m
$Z_{u,OK}(x_0)$	Interpolated residual value of AGBD at location x_0 using OK	t.ha^{-1}
$Z_{uv,CK}(x_0)$	Interpolated residual value of AGBD at location x_0 using CK	t.ha^{-1}
$Z_u(x_i)$	Residual of AGBD at location x_i	t.ha^{-1}
$Z_v(x_j)$	Value of co-variable (elevation) at location x_j	m
λ_{ui}	Weighting coefficient for AGBD residual at location x_i in OK	-
λ_{vj}	Weighting coefficient for co-variable at location x_j in CK	-
Model evaluation metrics calculation		
y_i	Observed AGBD value at the i^{th} location	t.ha^{-1}
$\widehat{y}_i$	Predicted AGBD value at the i^{th} location	t.ha^{-1}
$\bar{y}$	Mean of observed AGBD values	t.ha^{-1}
MAE	Mean absolute error: average of absolute differences between predicted and observed values	t.ha^{-1}
RMSE	Root mean square error: square root of the average of squared differences	t.ha^{-1}
R^2	Coefficient of determination: proportion of variance explained by the model	-
RI_{MAE}	Relative improvement in *MAE* between the baseline and the improved model	%
RI_{RMSE}	Relative improvement in *RMSE* between the baseline and the improved model	%
RI_{R2}	Relative improvement in R^2 between the baseline and the improved model	%
MAE_b	*MAE* of the baseline model	t.ha^{-1}
MAE_m	*MAE* of the improved model	t.ha^{-1}
$RMSE_b$	*RMSE* of the baseline model	t.ha^{-1}
$RMSE_m$	*RMSE* of the improved model	t.ha^{-1}
$R^2{}_b$	R^2 of the baseline model	-
$R^2{}_m$	R^2 of the improved model	-

Data availability

Data will be made available on request.

References

Abdikan, S., Sanli, F.B., Ustuner, M., Calò, F., 2016. Land cover mapping using Sentinel-1 SAR data. In: ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLI-B7, pp. 757–761. https://doi.org/10.5194/isprsarchives-XLI-B7-757-2016.

Baig, S., Qazi, W.A., Akhtar, A.M., Waqar, M.M., Ammar, A., Gilani, H., Mehmood, S.A., 2017. Above ground biomass estimation of Dalbergia sissoo forest plantation from dual-polarized ALOS-2 PALSAR data. Can. J. Remote Sens. 43, 297–308. https://doi.org/10.1080/07038992.2017.1330143.

Baloloy, A.B., Blanco, A.C., Candido, C.G., Argamosa, R.J.L., Dumalag, J.B.L.C., Dimapilis, L.L.C., Paringit, E.C., 2018. Estimation of mangrove forest aboveground biomass using multispectral bands, vegetation indices and biophysical variables derived from optical satellite imageries: RapidEye, Planetscope and Sentinel-2. In: ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. IV–3, pp. 29–36. https://doi.org/10.5194/isprs-annals-IV-3-29-2018.

Barbosa, J.M., Broadbent, E.N., Bitencourt, M.D., 2014. Remote sensing of aboveground biomass in tropical secondary forests: a review. Int. J. For. Res. 2014, 1–14. https://doi.org/10.1155/2014/715796.

Baret, F., Guyot, G., Major, D.J., 1989. TSAVI: A vegetation index which minimizes soil brightness effects on LAI and APAR estimation. In: 12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium. IEEE, pp. 1355–1358. https://doi.org/10.1109/IGARSS.1989.576128.

Bellassen, V., Luyssaert, S., 2014. Carbon sequestration: managing forests in uncertain times. Nature 506, 153–155. https://doi.org/10.1038/506153a.

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. https://doi.org/10.3390/rs10060831.

Blackburn, G.A., 1998. Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. Int. J. Remote Sens. 19, 657–675. https://doi.org/10.1080/014311698215919.

Broge, N., Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sens. Environ. 76, 156–172. https://doi.org/10.1016/S0034-4257(00)00197-8.

Brown, S., 1997. Estimating Biomass and Biomass Change of Tropical Forests: A Primer, FAO Forestry Paper. FAO Forestry Paper.

Canadell, J.G., Raupach, M.R., 2008. Managing forests for climate change mitigation. Science (80-.) 320, 1456–1457. https://doi.org/10.1126/science.1155458.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Chang. Biol. 20, 3177–3190. https://doi.org/10.1111/gcb.12629.

Chen, L., Ren, C., Zhang, B., Wang, Z., Xi, Y., 2018. Estimation of forest above-ground biomass by geographically weighted regression and machine learning with sentinel imagery. Forests 9, 582. https://doi.org/10.3390/f9100582.

Chen, L., Wang, Y., Ren, C., Zhang, B., Wang, Z., 2019a. Assessment of multi-wavelength SAR and multispectral instrument data for forest aboveground biomass mapping using random forest kriging. For. Ecol. Manage. 447, 12–25. https://doi.org/10.1016/j.foreco.2019.05.057.

Chen, L., Wang, Y., Ren, C., Zhang, B., Wang, Z., 2019b. Optimal combination of predictors and algorithms for forest above-ground biomass mapping from Sentinel and SRTM data. Remote Sens 11, 414. https://doi.org/10.3390/rs11040414.

Clevers, J.G.P.W., 1989. Application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. Remote Sens. Environ. 29, 25–37. https://doi.org/10.1016/0034-4257(89)90076-X.

Crippen, R., 1990. Calculating the vegetation index faster. Remote Sens. Environ. 34, 71–73. https://doi.org/10.1016/0034-4257(90)90085-Z.

Cutler, M.E.J., Boyd, D.S., Foody, G.M., Vetrivel, A., 2012. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: an assessment of predictions between regions. ISPRS J. Photogramm. Remote Sens. 70, 66–77. https://doi.org/10.1016/j.isprsjprs.2012.03.011.

David, R.M., Rosser, N.J., Donoghue, D.N.M., 2022. Improving above ground biomass estimates of Southern Africa dryland forests by combining Sentinel-1 SAR and Sentinel-2 multispectral imagery. Remote Sens. Environ. 282, 113232. https://doi.org/10.1016/j.rse.2022.113232.

Dong, R., Fu, H., 2023. Fusing time-inconsistent sentinel-2 images and high-resolution remote sensing images. In: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 5994–5996. https://doi.org/10.1109/IGARSS52108.2023.10283425.

Eckert, S., 2012. Improved forest biomass and carbon estimations using texture measures from WorldView-2 satellite data. Remote Sens. 4, 810–829. https://doi.org/10.3390/rs4040810.

Englhart, S., Keuck, V., Siegert, F., 2011. Aboveground biomass retrieval in tropical forests — the potential of combined X- and L-band SAR data use. Remote Sens. Environ. 115, 1260–1271. https://doi.org/10.1016/j.rse.2011.01.008.

Fan, X., He, G., Zhang, W., Long, T., Zhang, X., Wang, G., Sun, G., Zhou, H., Shang, Z., Tian, D., Li, X., Song, X., 2022. Sentinel-2 images based modeling of grassland above-ground biomass using random forest algorithm: a case study on the Tibetan plateau. Remote Sens 14, 5321. https://doi.org/10.3390/rs14215321.

Filipponi, F., 2019. Sentinel-1 GRD preprocessing workflow. In: 3rd International Electronic Conference on Remote Sensing. MDPI, Basel Switzerland, p. 11. https://doi.org/10.3390/ECRS-3-06201.

Foody, G.M., Boyd, D.S., Cutler, M.E.J., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. Remote Sens. Environ. 85, 463–474. https://doi.org/10.1016/S0034-4257(03)00039-7.

Forkuor, G., Benewinde Zoungrana, J.-B., Dimobe, K., Ouattara, B., Vadrevu, K.P., Tondoh, J.E., 2020. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets - a case study. Remote Sens. Environ. 236, 111496. https://doi.org/10.1016/j.rse.2019.111496.

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. https://doi.org/10.3390/rs10040627.

GFOI, 2013. Integrating remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative: Pub: Group on Earth Observations. Geneva, Switzerland.

Ghasemi, N., Sahebi, M.R., Mohammadzadeh, A., 2011. A review on biomass estimation methods using synthetic aperture radar data. Int. J. OFGEOMATICS Geosci. 1, 776–788.

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.

Goetz, S.J., Baccini, A., Laporte, N.T., Johns, T., Walker, W., Kellndorfer, J., Houghton, R.A., Sun, M., 2009. Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. Carbon Balance Manag. 4, 2. https://doi.org/10.1186/1750-0680-4-2.

Golden, N., Zhang, C., Potito, A., Gibson, P.J., Bargary, N., Morrison, L., 2020. Use of ordinary cokriging with magnetic susceptibility for mapping lead concentrations in soils of an urban contaminated site. J. Soils Sediments 20, 1357–1370. https://doi.org/10.1007/s11368-019-02537-7.

Hamdan, O., Mohd Hasmadi, I., Khali Aziz, H., Norizah, K., Helmi Zulhaidi, M.S., 2015. L-band saturation level for aboveground biomass of dipterocarp forests in peninsular Malaysia. J. Trop. For. Sci. 27, 388–399.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man. Cybern SMC-3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314.

Hoang Ha, A., Da Hanh, T.M., 2024. Vietnam's energy-related carbon emissions projections for the shared socioeconomic pathways towards net zero emission: a system dynamic analysis. Int. J. Sustain. Energy Plan. Manag. 42, 72–87. https://doi.org/10.54337/ijsepm.8327.

Hoang Khanh Linh, N., Van Chuong, H., 2015. Assessing the impact of urbanization on urban climate by remote satellite perspective: a case study in Danang city, Vietnam. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XL-7/W3, 207–212. https://doi.org/10.5194/isprsarchives-XL-7-W3-207-2015.

Huang, X., Ziniti, B., Torbick, N., Ducey, M.J., 2018. Assessment of forest above ground biomass estimation using multi-temporal C-band Sentinel-1 and polarimetric L-band PALSAR-2 data. Remote Sens. 10, 1424. https://doi.org/10.3390/rs10091424.

Huete, A., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X.

Huy, B., Poudel, K.P., Temesgen, H., 2016. Aboveground biomass equations for evergreen broadleaf forests in South Central Coastal ecoregion of Viet Nam: selection of eco-regional or pantropical models. For. Ecol. Manage. 376, 276–283. https://doi.org/10.1016/j.foreco.2016.06.031.

IPCC, 2006. 2006 IPCC guidelines for national greenhouse gas inventories. In: Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K. (Eds.), Prepared by the National Greenhouse Gas Inventories Programme. Japan.

IPCC, 2003. In: Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Wagner, F. (Eds.), Good Practice Guidance for Land Use, Land-Use Change and Forestry. Kanagawa, Japan.

Iremonger, S., Gerrand, A.M., 2011. Global Ecological Zones Map for FAO FRA 2010. Rome, Italy.

Islam, A.R.M.T., Ahmed, N., Bodrud-Doza, M., Chu, R., 2017. Characterizing groundwater quality ranks for drinking purposes in Sylhet district, Bangladesh, using entropy method, spatial autocorrelation index, and geostatistics. Environ. Sci. Pollut. Res. 24, 26350–26374. https://doi.org/10.1007/s11356-017-0254-1.

Jha, N., Tripathi, N.K., Barbier, N., Virdis, S.G.P.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. https://doi.org/10.1002/rse2.203.

Ji, Y., Wang, L., Zhang, W., Marino, A., Wang, M., Ma, J., Shi, J., Jing, Q., Zhang, F., Zhao, H., Huang, G., Yang, F., Wang, G., 2024. Forest above-ground biomass estimation using X, C, L, and P band SAR polarimetric observations and different inversion models. Int. J. Digit. Earth 17. https://doi.org/10.1080/17538947.2024.2310730.

Jiang, F., Sun, H., Chen, E., Wang, T., Cao, Y., Liu, Q., 2022. Above-ground biomass estimation for coniferous forests in Northern China using regression kriging and landsat 9 images. Remote Sens. 14, 5734. https://doi.org/10.3390/rs14225734.

Kanmegne Tamga, D., Latifi, H., Ullmann, T., Baumhauer, R., Bayala, J., Thiel, M., 2022. Estimation of aboveground biomass in agroforestry systems over three climatic regions in West Africa using Sentinel-1, Sentinel-2, ALOS, and GEDI data. Sensors 23, 349. https://doi.org/10.3390/s23010349.

Kaufman, Y.J., Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Trans. Geosci. Remote Sens. 30, 261–270. https://doi.org/10.1109/36.134076.

Kelsey, K., Neff, J., 2014. Estimates of aboveground biomass from texture analysis of landsat imagery. Remote Sens 6, 6407–6422. https://doi.org/10.3390/rs6076407.

Koch, B., 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS J. Photogramm. Remote Sens. 65, 581–590. https://doi.org/10.1016/j.isprsjprs.2010.09.001.

Kumar, P., Krishna, A.P., Rasmussen, T.M., Pal, M.K., 2021. An approach for fraction of vegetation cover estimation in forest above-ground biomass assessment using

sentinel-2 images. Communications in Computer and Information Science. Springer, Singapore, pp. 1–11. https://doi.org/10.1007/978-981-16-1086-8_1.
Kumari, K., Kumar, S., 2023. Machine learning based modeling for forest aboveground biomass retrieval. In: 2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS). IEEE, pp. 1–4. https://doi.org/10.1109/MIGARS57353.2023.10064607.
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. Remote Sens. 12, 1. https://doi.org/10.1117/1.JRS.12.016008.
Li, C., Zhou, L., Xu, W., 2021. Estimating aboveground biomass using sentinel-2 MSI data and ensemble algorithms for grassland in the Shengjin Lake Wetland, China. Remote Sens. 13, 1595. https://doi.org/10.3390/rs13081595.
Li, H., Kato, T., Hayashi, M., Wu, L., 2022. Estimation of forest aboveground biomass of two major conifers in Ibaraki Prefecture, Japan, from PALSAR-2 and Sentinel-2 data. Remote Sens. 14, 468. https://doi.org/10.3390/rs14030468.
Li, R., Tan, S., Zhang, M., Zhang, S., Wang, H., Zhu, L., 2024. Geological disaster susceptibility evaluation using a random forest empowerment information quantity model. Sustainability 16, 765. https://doi.org/10.3390/su16020765.
Li, Y., Li, M., Li, C., Liu, Z., 2020a. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. Sci. Rep. 10, 9952. https://doi.org/10.1038/s41598-020-67024-3.
Li, Y., Li, M., Liu, Z., Li, C., 2020b. Combining kriging interpolation to improve the accuracy of forest aboveground biomass estimation using remote sensing data. IEEE Access 8, 128124–128139. https://doi.org/10.1109/ACCESS.2020.3008686.
Li, Z., Bi, S., Hao, S., Cui, Y., 2022. Aboveground biomass estimation in forests with random forest and Monte Carlo-based uncertainty analysis. Ecol. Indic. 142, 109246. https://doi.org/10.1016/j.ecolind.2022.109246.
Liang, Y., Zeng, J., Li, S., 2022. Examining the spatial variations of land use change and its impact factors in a coastal area in Vietnam. Land 11, 1751. https://doi.org/10.3390/land11101751.
Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. Int. J. Remote Sens. 27, 1297–1328. https://doi.org/10.1080/01431160500486732.
Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., Moran, E., 2016. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. Int. J. Digit. Earth 9, 63–105. https://doi.org/10.1080/17538947.2014.990526.
Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., Vaglio Laurin, G., Saah, D., 2012. Aboveground forest biomass estimation with landsat and LiDAR data and uncertainty analysis of the estimates. Int. J. For. Res. 2012, 1–16. https://doi.org/10.1155/2012/436537.
Lü, X.-T., Yin, J.-X., Jepsen, M.R., Tang, J.-W., 2010. Ecosystem carbon storage and partitioning in a tropical seasonal forest in Southwestern China. For. Ecol. Manage. 260, 1798–1803. https://doi.org/10.1016/j.foreco.2010.08.024.
Lucas, R., Armston, J., Fairfax, R., Fensham, R., Accad, A., Carreiras, J., Kelley, J., Bunting, P., Clewley, D., Bray, S., Metcalfe, D., Dwyer, J., Bowen, M., Eyre, T., Laidlaw, M., Shimada, M., 2010. An evaluation of the ALOS PALSAR L-band backscatter—above ground biomass relationship Queensland, Australia: impacts of surface moisture condition and vegetation structure. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3, 576–593. https://doi.org/10.1109/JSTARS.2010.2086436.
Luo, K., Wei, Y., Du, J., Liu, L., Luo, X., Shi, Y., Pei, X., Lei, N., Song, C., Li, J., Tang, X., 2022. Machine learning-based estimates of aboveground biomass of subalpine forests using Landsat 8 OLI and Sentinel-2B images in the Jiuzhaigou National Nature Reserve, Eastern Tibet Plateau. J. For. Res. 33, 1329–1340. https://doi.org/10.1007/s11676-021-01421-w.
Maleki, M., Arriga, N., Barrios, J.M., Wieneke, S., Liu, Q., Peñuelas, J., Janssens, I.A., Balzarolo, M., 2020. Estimation of gross primary productivity (GPP) phenology of a short-rotation plantation using remotely sensed indices derived from sentinel-2 images. Remote Sens. 12, 2104. https://doi.org/10.3390/rs12132104.
Meul, M., Van Meirvenne, M., 2003. Kriging soil texture under different types of nonstationarity. Geoderma 112, 217–233. https://doi.org/10.1016/S0016-7061(02)00308-7.
Mitchard, E.T.A., Saatchi, S.S., White, L.J.T., Abernethy, K.A., Jeffery, K.J., Lewis, S.L., Collins, M., Lefsky, M.A., Leal, M.E., Woodhouse, I.H., Meir, P., 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. https://doi.org/10.5194/bg-9-179-2012.
Mitchell, M.W., 2011. Bias of the random forest out-of-bag (OOB) error for certain input parameters. Open J. Stat. 01, 205–211. https://doi.org/10.4236/ojs.2011.13024.
Mngadi, M., Odindi, J., Mutanga, O., 2021. The utility of sentinel-2 spectral data in quantifying above-ground carbon stock in an urban reforested landscape. Remote Sens 13, 4281. https://doi.org/10.3390/rs13214281.
Monsalve-Tellez, J.M., Torres-León, J.L., Garcés-Gómez, Y.A., 2022. Evaluation of SAR and optical image fusion methods in oil palm crop cover classification using the random forest algorithm. Agriculture 12, 955. https://doi.org/10.3390/agriculture12070955.
Moradi, F., Darvishsefat, A.A., Pourrahmati, M.R., Deljouei, A., Borz, S.A., 2022. Estimating aboveground biomass in dense hyrcanian forests by the use of sentinel-2 data. Forests 13, 104. https://doi.org/10.3390/f13010104.
Mwambala, A.N., Nyundo, B.A., Kalumanga, E., 2023. Tree biomass, carbon stock characteristics and ground beetles (Coleoptera: Carabidae) diversity in the Uzungwa Scarp Forest Nature Reserve, Tanzania. Environ. Manage. 71, 190–200. https://doi.org/10.1007/s00267-022-01733-5.
Naik, P., Dalponte, M., Bruzzone, L., 2022. Generative feature extraction from sentinel 1 and 2 data for prediction of forest aboveground biomass in the Italian Alps. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 4755–4771. https://doi.org/10.1109/JSTARS.2022.3179027.
Nesha, M.K., Hussin, Y.A., van Leeuwen, L.M., Sulistioadi, Y.B., 2020. Modeling and mapping aboveground biomass of the restored mangroves using ALOS-2 PALSAR-2 in East Kalimantan, Indonesia. Int. J. Appl. Earth Obs. Geoinf. 91, 102158. https://doi.org/10.1016/j.jag.2020.102158.
Nyamekye, C., Kwofie, S., Agyapong, E., Ofosu, S.A., Arthur, R., Appiah, L.B., 2021. Integrating support vector machine and cellular automata for modelling land cover change in the tropical rainforest under equatorial climate in Ghana. Curr. Res. Environ. Sustain. 3, 100052. https://doi.org/10.1016/j.crsust.2021.100052.
Oliver, M.A., Webster, R., 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. CATENA 113, 56–69. https://doi.org/10.1016/j.catena.2013.09.006.
Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S.W., McGuire, A.D., Piao, S., Rautiainen, A., Sitch, S., Hayes, D., 2011. A large and persistent carbon sink in the world's forests. Science (80-.) 333, 988–993. https://doi.org/10.1126/science.1201609.
Pandit, S., Tsuyuki, S., Dube, T., 2020. Exploring the inclusion of Sentinel-2 MSI texture metrics in above-ground biomass estimation in the community forest of Nepal. Geocarto Int. 35, 1832–1849. https://doi.org/10.1080/10106049.2019.1588390.
Pearson, T.R.H., Brown, S., Murray, L., Sidman, G., 2017. Greenhouse gas emissions from tropical forest degradation: an underestimated source. Carbon Balance Manag. 12, 3. https://doi.org/10.1186/s13021-017-0072-2.
Pearson, T.R.H., Brown, S.L., Birdsey, R.A., 2007. Measurement Guidelines for the Sequestration of Forest Carbon. U.S. Department of Agriculture, Forest Service, Northern Research Station. https://doi.org/10.2737/NRS-GTR-18.
Pham, Tien Dat, Yokoya, N., Xia, J., Ha, N.T., Le, N.N., Nguyen, T.T.T., Dao, T.H., Vu, T. T.P., Pham, Tien Duc, Takeuchi, W., 2020. Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the Red River Delta Biosphere Reserve, Vietnam. Remote Sens. 12, 1334. https://doi.org/10.3390/rs12081334.
Pinty, B., Verstraete, M.M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. Vegetatio 101, 15–20. https://doi.org/10.1007/BF00031911.
Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126. https://doi.org/10.1016/0034-4257(94)90134-1.
Ramachandran, N., Dikshit, O., 2022. Forest aboveground biomass estimation from Airborne L-Band SAR data using machine learning. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 6403–6405. https://doi.org/10.1109/IGARSS46834.2022.9884839.
Richardsons, A.J., Wiegand, A., 1977. Distinguishing vegetation from soil background information. Photogramm. Eng. Remote Sens. 43, 1541–1552.
Rincy, T.N., Gupta, R., 2020. Ensemble learning techniques and its efficiency in machine learning: a survey. In: 2nd International Conference on Data, Engineering and Applications (IDEA). IEEE, pp. 1–6. https://doi.org/10.1109/IDEA49133.2020.9170675.
Rodríguez-Veiga, P., Saatchi, S., Tansey, K., Balzter, H., 2016. Magnitude, spatial distribution and uncertainty of forest biomass stocks in Mexico. Remote Sens. Environ. 183, 265–281. https://doi.org/10.1016/j.rse.2016.06.004.
Rodríguez-Veiga, P., Wheeler, J., Louis, V., Tansey, K., Balzter, H., 2017. Quantifying forest biomass carbon stocks from space. Curr. For. Reports 3, 1–18. https://doi.org/10.1007/s40725-017-0052-5.
Roy, S., Mudi, S., Das, P., Ghosh, S., Shit, P.K., Bhunia, G.S., Kim, J., 2021. Estimating above ground biomass (AGB) and tree density using sentinel-1 data. Environmental Science and Engineering, pp. 259–280. https://doi.org/10.1007/978-3-030-56542-8_11.
Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M., Morel, A., 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. Proc. Natl. Acad. Sci. 108, 9899–9904. https://doi.org/10.1073/pnas.1019576108.
Salazar Villegas, M.H., Qasim, M., Csaplovics, E., González-Martinez, R., Rodriguez-Buritica, S., Ramos Abril, L.N., Salazar Villegas, B., 2023. Examining the potential of sentinel imagery and ensemble algorithms for estimating aboveground biomass in a tropical dry forest. Remote Sens. 15, 5086. https://doi.org/10.3390/rs15215086.
Sarker, L.R., Nichol, J.E., 2011. Improved forest biomass estimates using ALOS AVNIR-2 texture indices. Remote Sens. Environ. 115, 968–977. https://doi.org/10.1016/j.rse.2010.11.010.
Shi, G., Wang, Y., Zhang, J., Xu, J., Chen, Y., Chen, W., Liu, J., 2024. Spatiotemporal pattern analysis and prediction of carbon storage based on land use and cover change: a case study of Jiangsu coastal cities in China. Land 13, 1728. https://doi.org/10.3390/land13111728.
Shimada, M., Isoguchi, O., Tadono, T., Isono, K., 2009. PALSAR radiometric and geometric calibration. IEEE Trans. Geosci. Remote Sens. 47, 3915–3932. https://doi.org/10.1109/TGRS.2009.2023909.
Sinha, S., Jeganathan, C., Sharma, L.K., Nathawat, M.S., 2015. A review of radar remote sensing for biomass estimation. Int. J. Environ. Sci. Technol. 12, 1779–1792. https://doi.org/10.1007/s13762-015-0750-0.
Sivasankar, T., Lone, J.M., S, K.K., Qadir, A., R, P.L.N., 2018. The potential of multi-frequency multipolarized ALOS-2/PALSAR-2 and Sentinel-1 SAR data for aboveground forest biomass estimation. Int. J. Eng. Technol. 10, 797–802. https://doi.org/10.21817/ijet/2018/v10i3/181003095.
Stratoulias, D., Nuthammachot, N., Suepa, T., Phoungthong, K., 2022. Assessing the spectral information of Sentinel-1 and Sentinel-2 satellites for above-ground biomass retrieval of a tropical forest. ISPRS Int. J. Geo-Information 11, 199. https://doi.org/10.3390/ijgi11030199.

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.

Sun, G., Ranson, K., Kharuk, V., 2002. Radiometric slope correction for forest biomass estimation from SAR data in the Western Sayani Mountains, Siberia. Remote Sens. Environ. 79, 279–287. https://doi.org/10.1016/S0034-4257(01)00279-6.

Thi An, T., Izuru, S., Narumasa, T., Raghavan, V., Hanh, L.N., Van An, N., Long, N.V., Thi Thuy, N., Minh, T.P., 2022. Flood vulnerability assessment at the local scale using remote sensing and GIS techniques: a case study in Da Nang City, Vietnam. J. Water Clim. Chang. 13, 3217–3238. https://doi.org/10.2166/wcc.2022.029.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S.R., Schmullius, C., 2014. Carbon stock and density of northern boreal and temperate forests. Glob. Ecol. Biogeogr. 23, 297–310. https://doi.org/10.1111/geb.12125.

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.

Timothy, D., Onisimo, M., Cletah, S., Adelabu, S., Tsitsi, B., 2016. Remote sensing of aboveground forest biomass: a review. Trop. Ecol.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.

Vafaei, S., Soosani, J., Adeli, K., Fadaei, H., Naghavi, H., Pham, T., Tien Bui, D., 2018. Improving accuracy estimation of forest aboveground biomass based on incorporation of ALOS-2 PALSAR-2 and Sentinel-2A imagery and machine learning: a case study of the hyrcanian forest area (Iran). Remote Sens. 10, 172. https://doi.org/10.3390/rs10020172.

Viana, H., Aranha, J., Lopes, D., Cohen, W.B., 2012. Estimation of crown biomass of Pinus pinaster stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models. Ecol. Modell. 226, 22–35. https://doi.org/10.1016/j.ecolmodel.2011.11.027.

Vorster, A.G., Evangelista, P.H., Stovall, A.E.L., Ex, S., 2020. Variability and uncertainty in forest biomass estimates from the tree to landscape scale: the role of allometric equations. Carbon Balance Manag. 15, 8. https://doi.org/10.1186/s13021-020-00143-6.

Wai, P., Su, H., Li, M., 2022. Estimating aboveground biomass of two different forest types in Myanmar from Sentinel-2 data with machine learning and geostatistical algorithms. Remote Sens. 14, 2146. https://doi.org/10.3390/rs14092146.

Wang, L., Zhou, X., Zhu, X., Dong, Z., Guo, W., 2016. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. Crop J. 4, 212–219. https://doi.org/10.1016/j.cj.2016.01.008.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. John Wiley & Sons.

Wu, C., Shen, H., Shen, A., Deng, J., Gan, M., Zhu, J., Xu, H., Wang, K., 2016. Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. J. Appl. Remote Sens. 10, 035010. https://doi.org/10.1117/1.JRS.10.035010.

Xu, D., Wang, H., Xu, W., Luan, Z., Xu, X., 2021. LiDAR applications to estimate forest biomass at individual tree scale: opportunities, challenges and future perspectives. Forests 12, 550. https://doi.org/10.3390/f12050550.

Yalçin, E., 2005. Cokriging and its effect on the estimation precision. J. South African Inst. Min. Metall.

Yang, L., Wang, L., Abubakar, G.A., Huang, J., 2021. High-resolution rice mapping based on SNIC segmentation and multi-source remote sensing images. Remote Sens. 13, 1148. https://doi.org/10.3390/rs13061148.

Yavaşlı, D., 2012. Recent approaches in above ground biomass estimation methods. Aegean Geogr. J. 21, 39–49.

Yu, Y., Pan, Y., Yang, X., Fan, W., 2022. Spatial scale effect and correction of forest aboveground biomass estimation using remote sensing. Remote Sens. 14, 2828. https://doi.org/10.3390/rs14122828.

Yue, J., Feng, H., Yang, G., Li, Z., 2018. A comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy. Remote Sens. 10, 66. https://doi.org/10.3390/rs10010066.

Zeng, P., Zhang, W., Li, Y., Shi, J., Wang, Z., 2022. Forest total and component above-ground biomass (AGB) estimation through C- and L-band polarimetric SAR data. Forests 13, 442. https://doi.org/10.3390/f13030442.

Zhang, F., Marino, A., Ji, Y., Zhang, W., 2023. Coniferous forests aboveground biomass inversion in typical regions of China with Joint Sentinel-1 and Sentinel-2 remote sensing data supported by different feature optimizing algorithms. Forests 15, 56. https://doi.org/10.3390/f15010056.

Zhang, J., Lu, C., Xu, H., Wang, G., 2019. Estimating aboveground biomass of Pinus densata-dominated forests using Landsat time series and permanent sample plot data. J. For. Res. 30, 1689–1706. https://doi.org/10.1007/s11676-018-0713-7.

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. Digit. Earth 16, 158–182. https://doi.org/10.1080/17538947.2023.2165180.

Zhang, Y., Ma, J., Liang, S., Li, X., Li, M., 2020. An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products. Remote Sens. 12, 4015. https://doi.org/10.3390/rs12244015.

Zhang, Y., Ma, J., Liang, S., Li, X., Liu, J., 2022. A stacking ensemble algorithm for improving the biases of forest aboveground biomass estimations from multiple remotely sensed datasets. GIScience Remote Sens. 59, 234–249. https://doi.org/10.1080/15481603.2021.2023842.

Zhu, C., Wei, Y., Zhu, F., Lu, W., Fang, Z., Li, Z., Pan, J., 2022. Digital mapping of soil organic carbon based on machine learning and regression kriging. Sensors 22, 8997. https://doi.org/10.3390/s22228997.

Zolkos, S.G., Goetz, S.J., Dubayah, R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sens. Environ. 128, 289–298. https://doi.org/10.1016/j.rse.2012.10.017.