DENDROBIOLOGY

2026, vol. 95, 30–50

https://doi.org/10.12657/denbio.095.003

*Yao Dong, Le Yang, Mymuna Islam Moon, Yang Liu, Hongke Hao**

Estimation of above-ground biomass in forests of North and South mountains in Xining City using multi-source remote sensing data

Received: 19 January 2026; Accepted: 11 February 2026

Abstract: This study investigates a hierarchical multi-source remote sensing framework for estimating forest aboveground biomass (AGB) in the ecologically fragile North and South Mountains of Xining City, northeastern Qinghai-Tibet Plateau. The framework integrates Sentinel-2 optical imagery, spaceborne LiDAR (GEDI, ICESat-2), airborne LiDAR, and ground surveys to establish a "plot-local-regional" estimation system. Machine learning, particularly Random Forest, was used to model multi-source data relationships. The framework demonstrates strong applicability across scales: (1) at the local scale, the random forest model based on airborne LiDAR canopy structural features achieved high precision (R^2 = 0.90, RMSE = 2.14 t · ha^{-1}); (2) at the regional scale, the fusion of GEDI canopy profile metrics and Sentinel-2 spectral indices significantly enhanced estimation accuracy (R^2 = 0.82, RMSE = 3.90 t · ha^{-1}), markedly outperforming single-data-source models; (3) machine learning algorithms, particularly Random Forest, proved effective in handling multi-source data and capturing complex nonlinear relationships; (4) the generated 25-meter resolution AGB distribution map reveals clear spatial patterns, with biomass increasing with elevation and being significantly higher on shaded slopes than on sunlit slopes. The results confirm that the proposed framework is applicable for high-precision AGB estimation in high-altitude arid regions and provides a scalable technical pathway for forest carbon monitoring and spatialized management support.

Keywords: forest aboveground biomass, multi-source remote sensing data, machine learning, collaborative inversion

Addresses: Y. Dong, L. Yang, M. I. Moon, H. Hao, College of Forestry, Northwest A&F University, Yangling 712100, China; e-mail: hhk2018@126.com;
Y. Liu, College of Landscape Architecture & Arts, Northwest Agriculture and Forestry University, Yangling 712100, China
* corresponding author

Introduction

Forest ecosystems are the planet's largest carbon pool on land, holding roughly 76–98% of the world's terrestrial organic carbon and making up nearly 90% of all land-based plant biomass (Wang et al., 2013). They play an irreplaceable role in regulating the global carbon cycle, mitigating climate change, and maintaining ecological balance. As a key part of the world's forest resources, planted forests are growing in importance for timber production, environmental enhancement, landscape development, and climate

change mitigation. With more focus on sustainable forest management and improving overall quality and efficiency, plantations aren't just helping maintain ecosystem services and wood supply – they also ease pressure on natural forests and add to wood resources – plus, they're crucial in ecological restoration, landscape rehabilitation, and environmental improvement as part of broader eco-civilization efforts (Messier et al., 2022) . Furthermore, they provide essential support for sustainable modern forestry and regional carbon balance maintenance (Jepma et al., 2020; Ameray et al., 2021). Forest aboveground biomass (AGB) acts as both a major carbon stock and a telling indicator of forest ecosystem health. Getting AGB estimates right – and doing it efficiently – is the starting point for assessing how much carbon forests can soak up, and it's just as vital for evaluating regional carbon budgets scientifically and backing up sustainable forest management.

Traditional methods for estimating forest AGB mostly depend on field surveys using sample plots. These include approaches like destructive harvesting (where trees are cut and weighed), the sample tree method, or estimating through measurements such as diameter at breast height and tree height, combined with allometric equations (Nam et al., 2016; Possu et al., 2016). These techniques are heavily reliant on manual plot surveys, which are time-consuming and labor-intensive. They simply can't deliver comprehensive or rapid assessments across vast forest areas spanning millions of hectares. Long survey cycles also mean they often miss timely capture of dynamic processes like forest growth, degradation, or disturbance impacts, which can delay management decisions.

On the other hand, multi-source remote sensing technologies such as LiDAR, multispectral imaging, and satellite remote sensing offer distinct advantages: broad coverage, rapid data acquisition, high precision, and penetration capability. They've become go-to tools for precise ecological engineering management. Optical remote sensing (for example, the Landsat and Sentinel-2 satellite series) delivers continuous land cover and multispectral data. By calculating various vegetation indices (like NDVI and EVI) and pulling out textural features, optical data can effectively describe forest horizontal structure and leaf area dynamics, and they're already widely used in biomass estimation (Pham et al., 2019; Li et al., 2019). That said, optical remote sensing signals are sensitive to atmospheric conditions and often run into noticeable "signal saturation." This happens when forest canopy cover or biomass exceeds certain thresholds ($>150\ t \cdot ha^{-1}$), making vegetation indices less responsive to further biomass increases (Eckert, 2012; Zhang et al., 2024). As a result, high-biomass areas get underestimated while low-biomass areas are overestimated – a real limitation for accurate estimation in dense forests (Wang et al., 2022; Zhu et al., 2017).

LiDAR, as an active remote sensing technology, offers a more direct and accurate way to capture the three-dimensional vertical structure of forests – details like tree height, canopy height models, and vertical profile energy distribution. These parameters have strong physical ties to forest biomass and are much less prone to signal saturation compared to optical data, making LiDAR a more reliable option for biomass estimation (Kacic et al., 2023; Hu et al., 2020). Spaceborne LiDAR systems, such as the Global Ecosystem Dynamics Investigation (GEDI) and ICESat-2/ATLAS, have now taken this capability to a global scale (Xiong et al., 2025). That said, LiDAR data – especially from spaceborne platforms – come as discretely distributed sampling points, which can't directly produce spatially continuous biomass distribution maps. Also, LiDAR inherently doesn't provide continuous spectral information, which limits its usefulness for tasks like forest type classification and species identification (Pu, 2021).

Given the inherent limitations of any single remote sensing source for forest biomass estimation, blending multiple data sources – multi-source remote sensing fusion – has become a central research focus and a clear trend (Campbell et al., 2021). Collaborative approaches using statistical models, machine learning, and geostatistical spatial interpolation are now widely applied in forest AGB estimation. Methods based on statistical models work by building mathematical relationships between remote sensing features and ground-measured biomass. Among these, nonparametric machine learning models like Random Forest (RF) and Support Vector Machines (SVM) have become go-to choices for fusing multi-source data (e.g., optical, LiDAR, SAR), thanks to their strong ability to handle nonlinear patterns and high-dimensional data, which has noticeably boosted estimation accuracy (Zhao et al., 2023; Vafaei et al., 2018).

Geostatistical methods take advantage of the natural spatial autocorrelation in biomass to extend information from discrete high-precision sampling points (like spaceborne LiDAR footprints) into continuous spatial fields. Ordinary Kriging is a basic spatial interpolation method that works well in areas with spatially consistent attributes (Wang et al., 2025). Cokriging goes a step further by including auxiliary variables – such as topography and spectral data – as covariates, using spatial cross-correlations between variables to improve biomass mapping accuracy, especially in complex terrain (Wai et al., 2022). Together, these approaches provide a solid mathematical foundation for turning point-based biomass estimates into full spatial distributions. The North

and South Mountains of Xining City, located on the northeastern edge of the Qinghai-Tibet Plateau, serve as an important ecological security barrier and forest carbon sink. The forests in this region consist primarily of planted forests and natural shrublands, forming fragile ecosystems highly sensitive to climate change. Current research on forest biomass in this area remains limited. Existing studies predominantly rely on single optical remote sensing data sources (e.g., Landsat-8) and traditional regression models, leaving room for improvement in estimation accuracy and model robustness (Yang et al., 2016). Therefore, developing high-precision biomass estimation methods suitable for this region is urgently needed.

This study aims to establish a hierarchical "plot-local-regional" estimation system for forest aboveground biomass (AGB). Focusing on the North and South Mountains of Xining, we integrate Sentinel-2 optical imagery, spaceborne LiDAR, and ground plot survey data to develop a multi-scale collaborative inversion framework. The main objectives are: (1) to develop a cross-scale AGB estimation methodology based on multi-source remote sensing features; (2) to systematically compare the scale adaptability of different statistical models and identify the optimal estimation approach; (3) to generate high-resolution regional AGB spatial distribution maps and analyze their spatial pattern characteristics.

The findings are expected to provide an efficient technical solution for forest carbon stock assessment in ecologically fragile areas along the northeastern margin of the Tibetan Plateau, supporting dynamic monitoring and precision management decisions through an integrated approach from field measurements to regional mapping.

Materials and methods

Study Area

The study area is centered on Xining City in eastern Qinghai Province, situated along the northeastern edge of the Qinghai-Tibet Plateau. We focus specifically on the North–South Mountains Afforestation Project Zone surrounding the city, which lies between 36°28'–37°01'N and 101°06'–101°55'E (Fig. 1).

The terrain here is mostly made up of medium-low mountains, loess hills, and river valley terraces. Overall, the land is higher in the southwest and lower in the northeast, averaging around 2,300 meters in elevation with a relief of nearly 700 meters. The area is characterized by extensively eroded hills and dry sunny slopes covered with loess, featuring crisscrossing gullies and fragmented topography. In terms of climate, this region exhibits a typical inland semi-arid plateau climate characterized by dry conditions, with annual precipitation averaging only 350–400 millimeters. Evaporation significantly exceeds rainfall. Rainfall is concentrated primarily between July and September, predominantly occurring as short, intense downpours.

Since large-scale afforestation projects kicked off in 1989, forest coverage in the North–South Mountains has jumped from just 7.2%, turning the area into something of a model for ecological restoration in high-altitude arid zones. The study area is dominated by planted forests, with main tree species including *Populus spp., Picea crassifolia, Pinus tabuliformis*, and *Juniperus przewalskii*.

Data Sources and Pre-processing

Ground-based measurement data

Between August and October 2020, we conducted plot surveys within the study area, establishing a total of 90 plots, each measuring 25 × 25 m. Using high-precision surveying equipment, we accurately recorded the geographic coordinates of each plot's four corners and center point. Within each plot, all trees with a diameter at breast height (DBH) of 5.0 cm or greater were inventoried, and we measured their DBH, tree height, and crown width. To calculate the aboveground biomass of individual trees, we applied species-specific allometric equations derived from the Chinese National Standard (National Technical Committee for Standardization of Forest Resources, 2024) „Biomass Models and Carbon Accounting Parameters for Major Tree Species" (GB/T 43648-2024). The specific calculation formulas corresponding to dominant tree species are provided in Table 1.

Using the formulas described above, we first calculated the aboveground biomass for each tree within a plot. These individual tree values were then summed to get the total aboveground biomass per plot, which was finally converted to a per-hectare forest AGB value. This gave us plot-level aboveground biomass per unit area. Across the 90 plots, calculated AGB ranged from 6.82 to 44.49 $t \cdot ha^{-1}$, with a mean of 27.62 $t \cdot ha^{-1}$. This dataset serves as the baseline for developing and validating subsequent remote sensing models.

Table 1. Heterogeneous growth equation table for various tree species. (In the equations, M represents the above-ground biomass per tree (kilograms), D denotes the diameter at breast height (DBH, centimeters), and H indicates the tree height (meters))

Tree species	Calculation formula
Populus spp.	$0.04607 \times D^{2.14892} \times H^{0.59163}$
Picea crassifolia	$0.12890 \times D^{2.09828} \times H^{0.25663}$
Pinus tabuliformis	$0.067765 \times D^{2.18050} \times H^{0.43610}$
Juniperus przewalskii	$0.08947 \times D^{1.91489} \times H^{0.61516}$

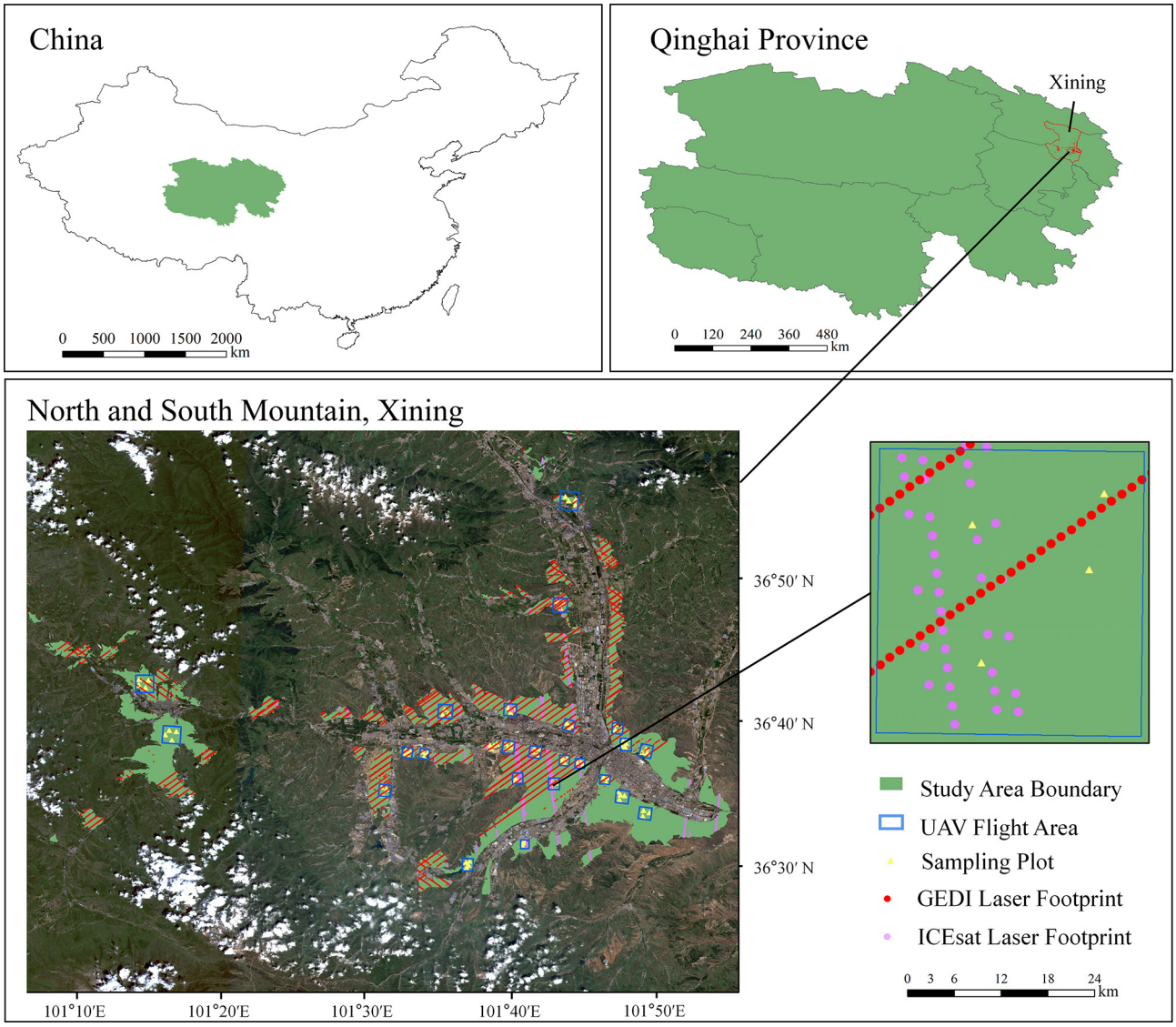

Fig. 1. Schematic diagram of study area location, remote sensing data, and plot locations

UAV-LiDAR data

Between August and October 2020, we conducted UAV-LiDAR surveys across 24 flight zones covering all 90 ground sample plots. The system consisted of a DJI M600 Pro multi-rotor UAV equipped with a Li-Air 220 laser scanner (GreenValley, China). The Li-Air 220 integrates a 40-line rotating laser scanner, an inertial navigation system, and a visible-light camera, with a detection range of 0.3–220 m and a measurement frequency of 720 kHz. The acquired point cloud data were used to generate digital elevation models (DEMs) and digital surface models (DSMs) for subsequent analysis.

We preprocessed the raw point cloud data in LiDAR360 software (GreenValley, China). First, we applied a height threshold to remove noise while preserving vegetation and ground returns. We then used an irregular triangulation network (TIN) filtering algorithm to classify the point cloud into ground and non-ground points. Based on this classification, we generated a DEM and a DSM. To reduce terrain effects, we performed elevation normalization using the DEM and derived a canopy height model (CHM) by subtracting the DEM from the DSM.

GEDI data

We obtained the Level-2B (L2B) data product from the Global Ecosystem Dynamics Investigation (GEDI) mission covering the period from June to December 2020. The data were downloaded from the NASA Earthdata website (https://search.earthdata.nasa.gov/search/). This product offers footprint-level estimates of aboveground biomass, canopy cover, and vertical profile metrics. Using the Python library h5py, we read the HDF5-formatted orbital data and extracted the unique shot identifier (shot_number) along with its geographic coordinates (lon_lowestmode, lat_lowestmode).

To ensure data reliability for subsequent modeling, we implemented strict quality control filtering

based on recommendations from the official data product handbook (Dubayah et al., 2020) and related literature (Moudrý et al., 2024). We kept only those shots that met the following criteria: quality_flag = 1 (indicating the shot is usable according to the standard algorithm), degrade_flag = 0 (signifying no significant degradation in instrument performance), and sensitivity ≥ 0.95 (ensuring a high signal-to-noise ratio and reliable detection of ground and canopy returns). This combination of thresholds is widely adopted to select high-quality, full-power laser shots with minimal noise and optimal geolocation accuracy for vegetation parameter retrieval.. This process resulted in a total of 6637 high-quality GEDI shots distributed across the study area, which were then used for further analysis.

ICESat-2/ATLAS data

To maintain temporal consistency with the GEDI data, we also acquired the Level 3A Land and Vegetation Height product (ATL08) from the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) mission for the same period (June to December 2020, available at https://nsidc.org/data/atl08)). The ATL08 product includes key structural parameters such as vegetation canopy height percentiles and apparent surface reflectance (ASR). Using the h5py library in Python, we extracted the geographic coordinates (latitude, longitude) of photon events and corresponding vegetation height metrics, like h_canopy.

To ensure data quality, we carried out an initial screening based on the ATL08 quality description flag (quality_flag). We then applied the three-sigma rule to remove outliers from major parameters, including h_canopy (Zhang et al., 2022). After this processing, we ended up with a total of 1115 valid ICESat-2 vegetation photon points within the study area. Their spatial distribution is illustrated in Figure 1. This dataset serves as an independent source of forest vertical structure observations and is spatiotemporally aligned with the GEDI data for subsequent collaborative inversion.

Sentinel-2 data

Sentinel-2 multispectral imagery was obtained from the Copernicus Open Access Hub of the European Space Agency (https://scihub.copernicus.eu). To align with the phenology of our field surveys, we acquired two Level-2A (L2A) scenes from August 2020, close to the investigation period. The Sentinel-2 satellite operates at an altitude of 786 km with a swath width of 290 km. Its MultiSpectral Instrument (MSI) provides data across 13 spectral bands at spatial resolutions of 10 m, 20 m, and 60 m. The 20 m resolution red-edge bands (Bands 5, 6, 7) and shortwave infrared bands (Bands 11, 12) are particularly sensitive to chlorophyll content and water stress in vegetation, making them highly useful for monitoring forest condition and biomass.

The Level-2A product provides atmospherically corrected surface reflectance. We performed additional processing in ENVI 5.6, including resampling bands with 20 m and 60 m resolutions to a uniform 25 m pixel size, followed by masking and clipping using the study area boundary vector. The resulting image was used for feature extraction.

All remote sensing data were resampled to a uniform 25-meter resolution to align with both our field plot size (25 × 25 m) and the nominal footprint diameter of GEDI data (~25 m; Dubayah et al., 2020). This scale consistency supports direct plot-to-pixel comparison and enables seamless fusion of multi-source data – including Sentinel-2 spectral imagery, GEDI/ICESat-2 structural metrics, and SRTM topography – while mitigating biases from mixed resolutions during modeling and prediction.

Digital elevation data

The Digital Elevation Model (DEM), a raster dataset representing ground surface elevation, is widely utilized in natural resource management. We acquired the global DEM data from the Shuttle Radar Topography Mission (SRTM) through the Google Earth Engine (GEE) platform as the topographic data source for the study area, with an original spatial resolution of 30 m (data accessed via Google Earth Engine, https://earthengine.google.com).

To maintain spatial consistency with the multi-source remote sensing data, we resampled the DEM to a 25 m resolution in ArcGIS 10.8. The resampled elevation data were then mosaicked and clipped to generate a seamless topographic raster covering the entire study area. Based on this processed DEM, key topographic variables including slope and aspect were derived using GIS spatial analysis tools (Banerjee et al., 2017).

Research methodology

Rationale for the Hierarchical Framework

The spatial discontinuity of spaceborne LiDAR footprints (e.g., GEDI, ICESat-2) poses a significant challenge for direct calibration with ground plots, which are rarely spatially coincident. Unlike optical imagery, which provides continuous coverage, spaceborne LiDAR samples the landscape discretely, making matched plot collection logistically demanding and often unfeasible at regional scales. To overcome this limitation, we designed a two-stage scaling framework that introduces UAV-LiDAR as an intermediate layer. This approach allows us to (1) generate a continuous, high-resolution AGB reference surface from UAV-LiDAR and plot data, and (2) use this surface to calibrate spaceborne remote sensing data without

relying on direct plot–satellite matching, thereby enhancing scalability and reducing spatial uncertainty.

Feature Extraction

Based on the preprocessed UAV-LiDAR vegetation point cloud, we extracted 64 feature variables (Yan et al., 2023), including height, density, intensity, and canopy-related metrics derived from the point cloud data (Table 2).

Multi-modal remote sensing feature variables were then extracted as summarized in Table 3. The extracted features can be grouped into three main categories: (1) canopy-related metrics from the GEDI L2B product, such as canopy cover, leaf area index, and foliage height diversity; (2) percentile heights and canopy structural parameters from the ICESat-2/ATL08 product; and (3) spectral bands and multiple vegetation indices derived from Sentinel-2 imagery.

Additionally, topographic factors extracted from SRTM data were included to better characterize the environmental drivers influencing the spatial distribution of AGB.

Feature Selection

We examined the correlation between each feature variable and AGB using Pearson's correlation coefficient (with a significance level of $P < 0.05$). The correlation coefficient (r) ranges from -1 to 1, where values closer to |1| indicate stronger relationships. To reduce the impact of multicollinearity on model stability, we excluded feature pairs with $|r| > 0.8$.

Using Python, we carried out feature selection based on the Random Forest (RF) importance ranking method (Li et al., 2021). In the context of multi-source remote sensing data synergy, we ranked all features according to their importance scores derived from RF to identify the optimal feature subset. Features were then added to the RF model sequentially in descending order of importance, and the root mean square error (RMSE) of the model output was recorded at each step. The feature combination corresponding to the minimum RMSE was considered the optimal subset for nonparametric modeling and was subsequently used to build the final model.

Forest AGB Inversion Model

This study developed multiple models to evaluate how different algorithms and feature combinations perform in estimating forest aboveground biomass (AGB) across the study area. The models we built included multivariable linear regression (MLR), support vector machine (SVM), and random forest (RF).

MLR assumes a linear relationship between the dependent variable and multiple independent

Table 2. Feature variables of UAV LiDAR

Variable type	Variable name	Description
Elevation metrics	h_n	Height percentile, n=1,5,10,20,25,30,40,50,60,70,75,80,90,95,99
	h_max	Maximum
	h_min	Minimum
	h_mean	Mean
	h_median	Median
	h_kur	Kurtosis
	h_ske	Skewness
	h_sqrt	Quadratic mean
	h_std	Standard deviation
	h_var	Variance
	h_cur	Cubic mean
	h_cv	Coefficient of variation
Intensity metrics	i_n	Intensity percentile, n=1,5,10,20,25,30,40,50,60,70,75,80,90,95,99
	i_max	Maximum
	i_mean	Mean
	i_median	Median
	i_min	Minimum
	i_ske	Skewness
	i_var	Variance
	i_std	Standard deviation
	i_kur	Kurtosis
	i_cv	Coefficient of variation
Density metrics	d_n	Density variable, n=0,1,2,3,4,5,6,7,8,9
Canopy feature	CC	Canopy cover
	GF	Gap fraction
	LAI	Leaf area index
	CRR	Canopy relief ratio

variables. It estimates AGB by fitting a linear regression equation – offering clear interpretability and computational efficiency when the underlying relationships are actually linear (Zhang et al., 2024); Support vector regression (SVR) extends support vector machine theory to regression tasks. Using kernel functions, SVR projects the original data into a higher-dimensional space, which helps capture complex nonlinear patterns between predictor variables and AGB. Among available kernel options, the Gaussian radial basis function (RBF) is commonly chosen for its flexibility in modeling smooth, nonlinear responses (Georgopoulos et al., 2023); RF is an ensemble learning method made up of multiple decision trees. It aggregates predictions from individual trees to improve accuracy and stability. RF shows strong robustness, handles high-dimensional data well, and naturally provides variable importance estimates, making it especially suitable for modeling the intricate ecological relationships found in multisource remote sensing data (Adugna et al., 2022).

Table 3. Feature variables of spaceborne LiDAR

Data source	Product	Variable name	Description
GEDI	L2B	pai	Pai, Plant area index
		fhd	fhd_norma, Leaf-height diversity index
		cover	Cover, Total canopy cover percentage
		rg	Rg, Integral of the ground component
		rv	Rv, Integral of the vegetation component
		modis_treecover	Tree cover fraction from MODIS
		modis_nonvegetated	Non-Vegetated percentage in MODIS
		pta_N	pgap_theta_aN, N = 1,2,3,4,5,6,Forest canopy gap probability
		rea_N	rx_energy_aN, N = 1,2,3,4,5,6, Received waveform energy between toploc and botloc with noise removed
		hl	height_lastbin, Height of the last bin of the pgap_theta_z, relative to the ground.
ICEsat-2	ATL08	asr	Apparent surface reflectance
		dem_H	Terrain height best fit
		H_min	Minimum height
		H_mean	Mean height
		H_max	Maximum height
		H_median	Median height
		H_N	N=10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,95
Sentinel-2	L2A	B2,B3,B4,B5,B6,B7,B8,B8A,B11,B12	Single-band reflectance
		NDVI (Normalized Difference Vegetation Index)	(B8 − B4) / (B8 + B4)
		NDVI45 (Narrow Red-Edge NDVI)	(B5 − B4) / (B5 + B4)
		GNDVI (Green normalized difference vegetation index)	(B8 − B3) / (B8 + B3)
		EVI (Enhanced vegetation index)	2.5 × (B8 − B4) / (B8 + 6 × B4 − 7.5 × B2 + 1)
		EVI2 (Two-Band Enhanced Vegetation Index)	2.5 × (B8 − B4) / (B8 + 2.4 × B4 + 1)
		RVI (Ratio Vegetation Index)	B8 / B4
		DVI (Difference Vegetation Index)	B8 − B4
		TVI (Triangular Vegetation Index)	0.5 × (120 × (B6 − B3) − 200×(B4 − B3))
		SAVI (Soil-Adjusted Vegetation Index)	(1 + L) × (B8 − B4) / (B8 + B4 + L)
		MSAVI (Modified Soil-Adjusted Vegetation Index)	(2×B8 + 1 − sqrt((2 × B8 + 1)2 − 8 × (B8 − B4))) / 2
		GRVI (Green ratio vegetation index)	(B3 − B4) / (B3 + B4)
		S2REP (Sentinel-2 Red-Edge Position Index)	705 + 35 × ((B7 + B4)/2 − B5) / (B6 − B5)
		CI (Chlorophyll Index)	(B8 − B5) / (B8 + B5)
		CIRE (Red-Edge Chlorophyll Index)	(B8 − B6) / (B8 + B6)
		IRECI (Inverted Red-Edge Chlorophyll Index)	(B7 − B4) / (B5 / B6)
SRTM DEM		Elevation, Slope, Aspect	

Note: Key vegetation indices and their primary relevance: NDVI – sensitivity to chlorophyll content and canopy greenness; EVI/EVI2 – reduced sensitivity to atmospheric and soil background effects, mitigating saturation in dense vegetation; Red-edge indices (NDVI45, CIRE, S2REP, IRECI) – sensitivity to variations in leaf chlorophyll, nitrogen, and canopy structure; SAVI/MSAVI – correction for soil background reflectance; GNDVI – sensitivity to canopy nitrogen and photosynthetic activity; CI – sensitivity to chlorophyll content in high-biomass vegetation. The selection aimed to capture complementary aspects of vegetation physiology and structure related to AGB.

Through comparative analysis of these models, the study aims to identify the most reliable modeling framework for accurate and generalizable forest AGB estimation in the region.

Spatial Fusion and Interpolation Methods for Multi-Source Remote Sensing Data

The spaceborne LiDAR structural parameters acquired in this study are discrete point data, requiring spatial autocorrelation testing prior to spatialization. Global Moran's I was employed to analyze the spatial autocorrelation of parameters, calculated as follows:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(y_j - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (1)$$

where n is the number of sample points, x_i and x_j are the observed values at locations i and j respectively, $\bar{x}$ is the mean of all observed values, and w_{ij} are elements of the spatial weight matrix constructed based on the inverse distance weighting method. Significance testing was performed using Monte Carlo simulation (999 permutations), with $p < 0.05$ indicating significant spatial autocorrelation.

For parameters passing the spatial autocorrelation test, Ordinary Kriging was employed for spatial interpolation. Kriging is an optimal unbiased spatial prediction method based on the theory of regionalized variables and variogram models. The predicted value $\hat{Z}(s_0)$ and prediction variance $\sigma^2(s_0)$ are calculated as:

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) \quad (2)$$

$$\sigma 2(s_0) = C(0) - \sum_{i=1}^{\eta} \lambda_i C(s_i - s_0) - \mu \quad (3)$$

where λ_i are Kriging weights obtained by solving the Kriging system of equations; $C(0)$ is the covariance function; and μ is the Lagrange multiplier. Semivariogram models including spherical, exponential, and Gaussian were fitted, with the optimal model selected based on root mean square error (RMSE) and coefficient of determination (R^2) from leave-one-out cross-validation.

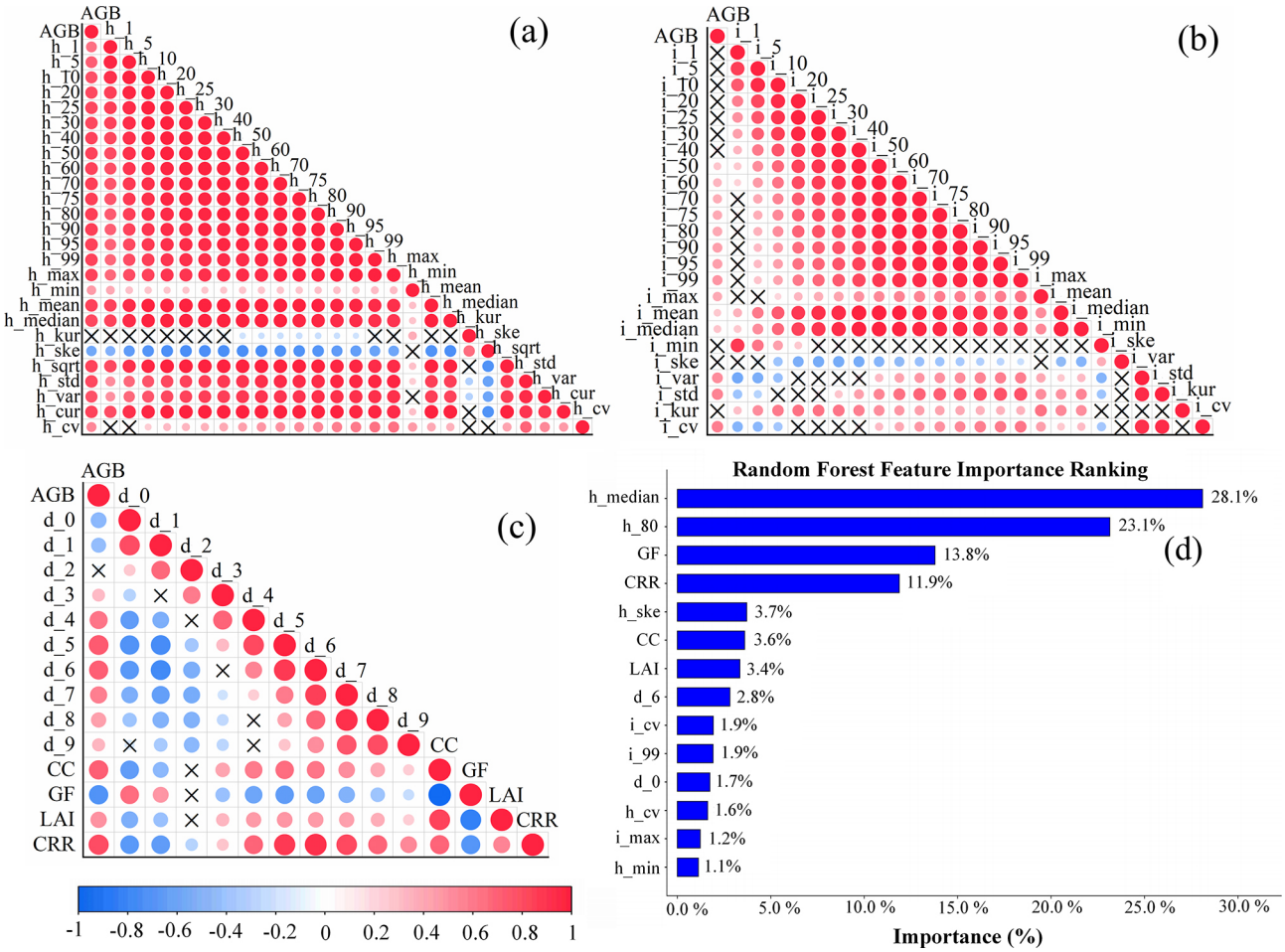

Fig. 2. Correlation and importance ranking of UAV-LiDAR feature variables relative to field-measured AGB (a. Elevation metrics; b. Intensity metrics; c. Density metrics; d. Feature importance ranking of initially screened variables using random forest)

Accuracy Evaluation Metrics

To fully utilize the sample data and enhance model reliability, different cross-validation strategies were employed according to the sample size at each stage of the hierarchical framework.

For the plot-to-airborne AGB inversion (Stage 1), which is based on a relatively small number of field plots (n = 90), we employed leave-one-out cross-validation (LOOCV). LOOCV treats each individual sample as the test set in turn, maximizing the use of limited data and providing a robust assessment of model performance with small samples. For the airborne-to-regional AGB inversion (Stage 2), which utilizes the larger set of AGB reference points derived from UAV-LiDAR within the airborne coverage area, we applied 5-fold cross-validation. This method randomly partitions the dataset into five subsets, iteratively using four for training and one for testing. It provides a stable and computationally efficient performance estimate suitable for larger sample sizes, while still maintaining a rigorous validation standard.The accuracy of the models was assessed using the following evaluation metrics: Coefficient of determination (R^2), Root mean square error (RMSE) , Mean absolute error (MAE) and Relative root mean square error (rRMSE). R^2 measures the goodness of fit of the model, with values closer to 1 indicating better fit. RMSE quantifies the prediction error of the model, while MAE represents the average absolute deviation between predicted and observed values. Lower RMSE and MAE values indicate better model performance. rRMSE is a normalized form of RMSE that reflects the error proportion relative to the observed values. The formulas for these evaluation metrics are as follows:

$$r = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{6}$$

$$rRMSE = \frac{RMSE}{\bar{y}_i} \tag{7}$$

$$EA = 1 - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{y_i}\right) \times 100\% \tag{8}$$

In the formula, y_i denotes the measured value, $\hat{y}_i$ denotes the predicted value, $\bar{y}_i$ denotes the average of the measured values, and n denotes the sample size.

Results and analysis

Inversion Results of Forest AGB from UAV Data and Sample Plots

UAV-LiDAR Point Cloud Feature Selection

Fig. 2 shows the Pearson correlation coefficient matrix between UAV-LiDAR feature variables and field-measured AGB. Overall, height-related features generally showed strong correlations with AGB, while density and intensity features had relatively weaker relationships. This suggests that forest vertical structure information – particularly canopy height distribution – serves as a key physical basis for AGB estimation.

After removing redundant features, the initial feature set included: h_median, h_80, h_cv, h_min, h_ske, GF, CRR, CC, LAI, d_6, d_0, i_99, and i_max. We then performed feature importance ranking using the Random Forest algorithm, which identified h_median, h_80, GF, and CRR as the key variables. This outcome indicates that features representing overall canopy height (h_median), upper canopy structure (h_80), and canopy geometry (GF, CRR) hold the highest explanatory power in the model, while some statistical features (e.g., skewness h_ske) and lower percentile heights (h_min) were filtered out due to their lower sensitivity to AGB variation.

It is worth noting that although intensity features (i_99, i_max) and certain density features (d_6, d_0) were kept in the initial screening, they were not included in the final key variable set during importance ranking. This is likely because intensity metrics are more susceptible to variations in sensor calibration, surface moisture conditions, and scan geometry, which can introduce noise unrelated to biomass. In contrast, structural height and canopy geometry metrics (e.g., h_median, CRR) provide a more direct and stable physical relationship with AGB. This result implies that for UAV-LiDAR data in this study area, structural features contribute more significantly and robustly to AGB estimation than echo intensity and point cloud density features.

Estimation of Forest AGB Based on UAV-LiDAR

Using the selected UAV-LiDAR feature set as independent variables and field-measured AGB data from the study area as the dependent variable, we constructed three models – MLR, SVR, and RF – for forest AGB estimation. The AGB fitting results from these inversion models are presented in Fig. 3.

The results show that among the three models used in this study, the Random Forest model exhibited the best predictive performance, achieving an R^2 of 0.90 against field-measured AGB, with an RMSE of 2.14 $t \cdot ha^{-1}$ and an rRMSE of 7.75%. The MLR and SVM models showed similar prediction

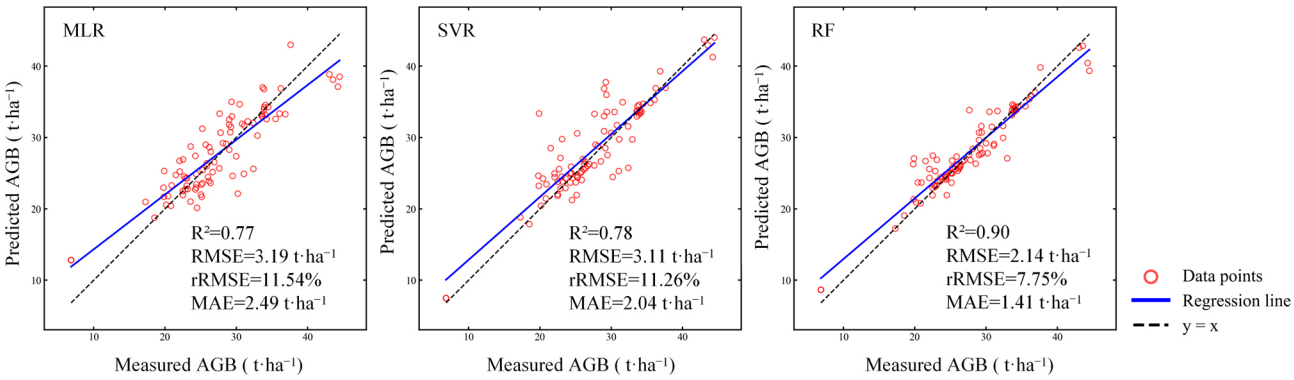

Fig. 3. Scatter plots of UAV-LiDAR-derived AGB predicted by MLR, SVR, and RF against field-measured plot AGB

performance. Specifically, MLR achieved an R^2 of 0.77, with an RMSE of 3.19 t · ha^{-1} and an rRMSE of 11.54%, while SVM reached an R^2 of 0.78, with an RMSE of 3.11 t · ha^{-1} and an rRMSE of 11.25%.

For AGB values in the range of 20–35 t · ha^{-1}, the RF model demonstrated superior fitting performance, with overall prediction accuracy significantly outperforming both MLR and SVR. It is worth noting that the RF model maintained high accuracy even within the moderate AGB range (20–35 t · ha^{-1}), indicating that the selected height and canopy structure features (h_median, h_80, GF, CRR) provide robust explanatory power across varying biomass levels. This further supports the advantage of UAV-LiDAR in characterizing forest vertical heterogeneity.

Therefore, the variables selected from the airborne LiDAR data were used to develop the RF model, which was then applied to predict AGB reference values corresponding to the laser footprints within the masked airborne coverage area. This approach generated a spatially distributed forest AGB map consistent with the resolution of spaceborne LiDAR data. This method not only validates the reliability of UAV-LiDAR in regional-scale AGB mapping but also provides a high-precision ground validation benchmark for subsequent multi-source remote sensing collaborative inversion.

Estimation of Forest AGB Using Satellite Remote Sensing Data

Spaceborne Remote Sensing Data Feature Selection

Following the similar methodological approach to that used for UAV-LiDAR data, we performed feature selection on the spaceborne remote sensing data – Sentinel-2, ICESat-2, and GEDI – to identify the best predictors for forest aboveground biomass estimation. The process mainly involved two steps: correlation analysis and machine learning-based importance ranking using Random Forest.

The selected spaceborne features included vegetation indices from Sentinel-2 (CI, GRVI, CIRE, GNDVI, NDVI45, NDVI705, EVI), spectral bands (B2, B11), canopy structural parameters from GEDI (rv, fhd, pta_4, pta_1, cover, pai), and height percentiles from ICESat-2 (H_60, H_95, H_65). We also incorporated elevation data derived from the SRTM DEM as a key topographic variable to account for terrain effects on biomass distribution.

This multi-sensor feature set captures complementary information related to forest structure, chlorophyll content, and moisture status, supporting robust AGB estimation across the study area. Building on the analysis of single data sources, we adopted two variable combination schemes: ICESat-2 + Sentinel-2 and GEDI + Sentinel-2. The top-ranked features were added sequentially to a Random Forest regression model, and the root mean square error was recorded at each step. The feature subset corresponding to the minimum RMSE was ultimately selected as the final modeling feature set (Fig. 5). These combination schemes further demonstrate the effectiveness of multi-source data synergy. The ICESat-2 + Sentinel-2 combination emphasizes the integration of „vertical structure + spectral features," making it suitable for areas where canopy height and spectral information are highly complementary. The GEDI + Sentinel-2 combination integrates „canopy profile + spectral reflectance" information, which is particularly useful for stands with high canopy closure and complex vertical structures. By selecting the feature subset based on RMSE minimization (Fig. 5), this study not only avoids overfitting but also ensures the model's generalizability with limited samples, laying a reliable foundation for subsequent spatial mapping of AGB.

Comparison and Analysis of RF Model Accuracy with Different Data Source Combinations

Based on different data sources and using the spaceborne laser spot AGB reference data within the airborne coverage area, we constructed models with the Random Forest algorithm – given its superior

fitting accuracy shown in the UAV-LiDAR-based inversion (Section 4.1.2). Five distinct combinations were tested: Sentinel-2 only, ICESat-2 only, GEDI only, ICESat-2+Sentinel-2, and GEDI+Sentinel-2. Scatter plots of model accuracy are shown in Figure 6, and the optimal variables along with corresponding results are summarized in Fig. 6 and Table 4.

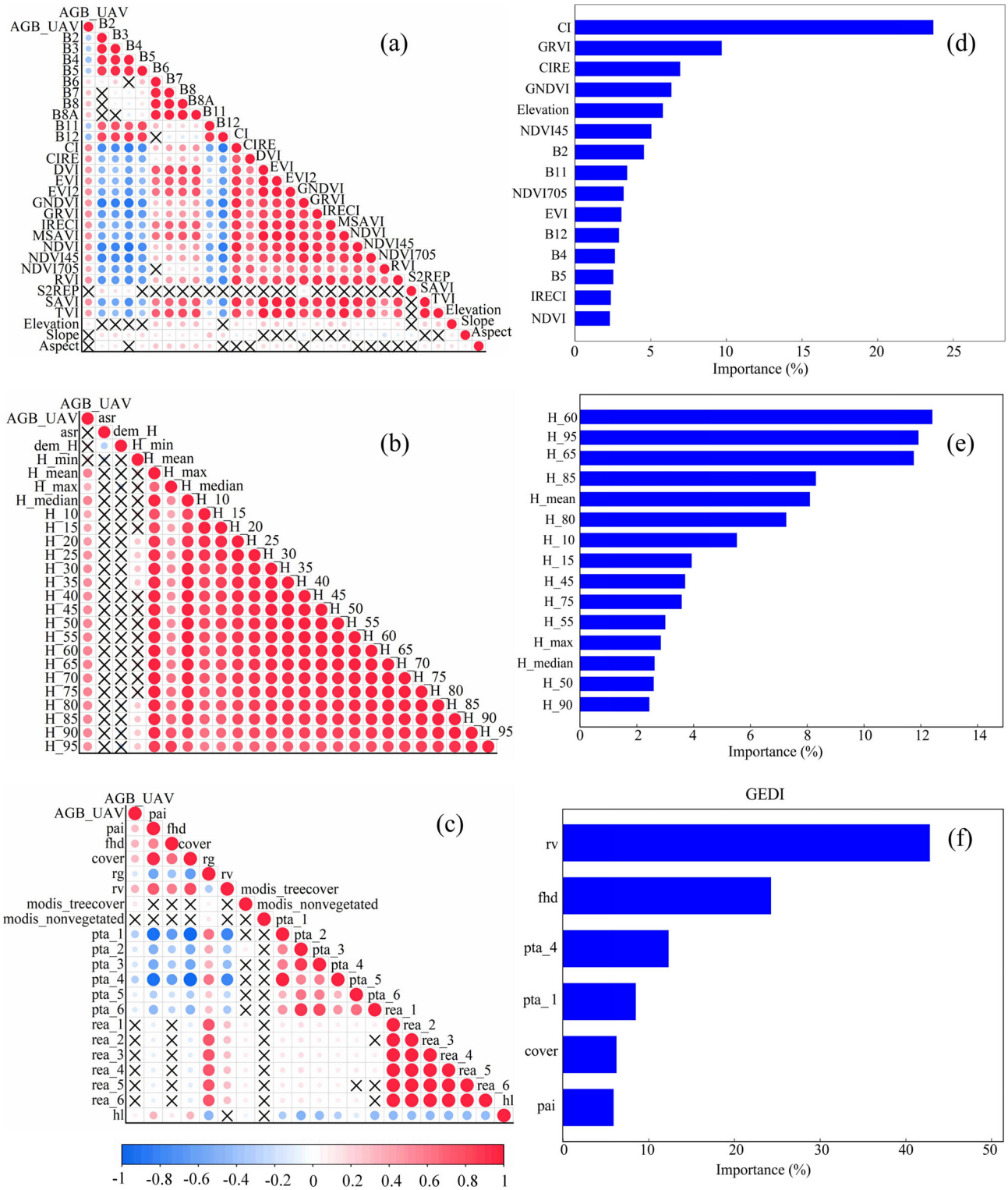

Fig. 4. Correlation and importance ranking between spaceborne remote sensing feature variables and UAV-LiDAR regional AGB (a. Pearson correlation matrix of Sentinel-2 feature variables; b. Pearson correlation matrix of ICEsat-2 feature variables; c. Pearson correlation matrix of GEDI feature variables; d. Random Forest importance ranking of Sentinel-2 feature variables; e. Random Forest importance ranking of ICEsat-2 feature variables; f. Random Forest importance ranking of GEDI feature variables)

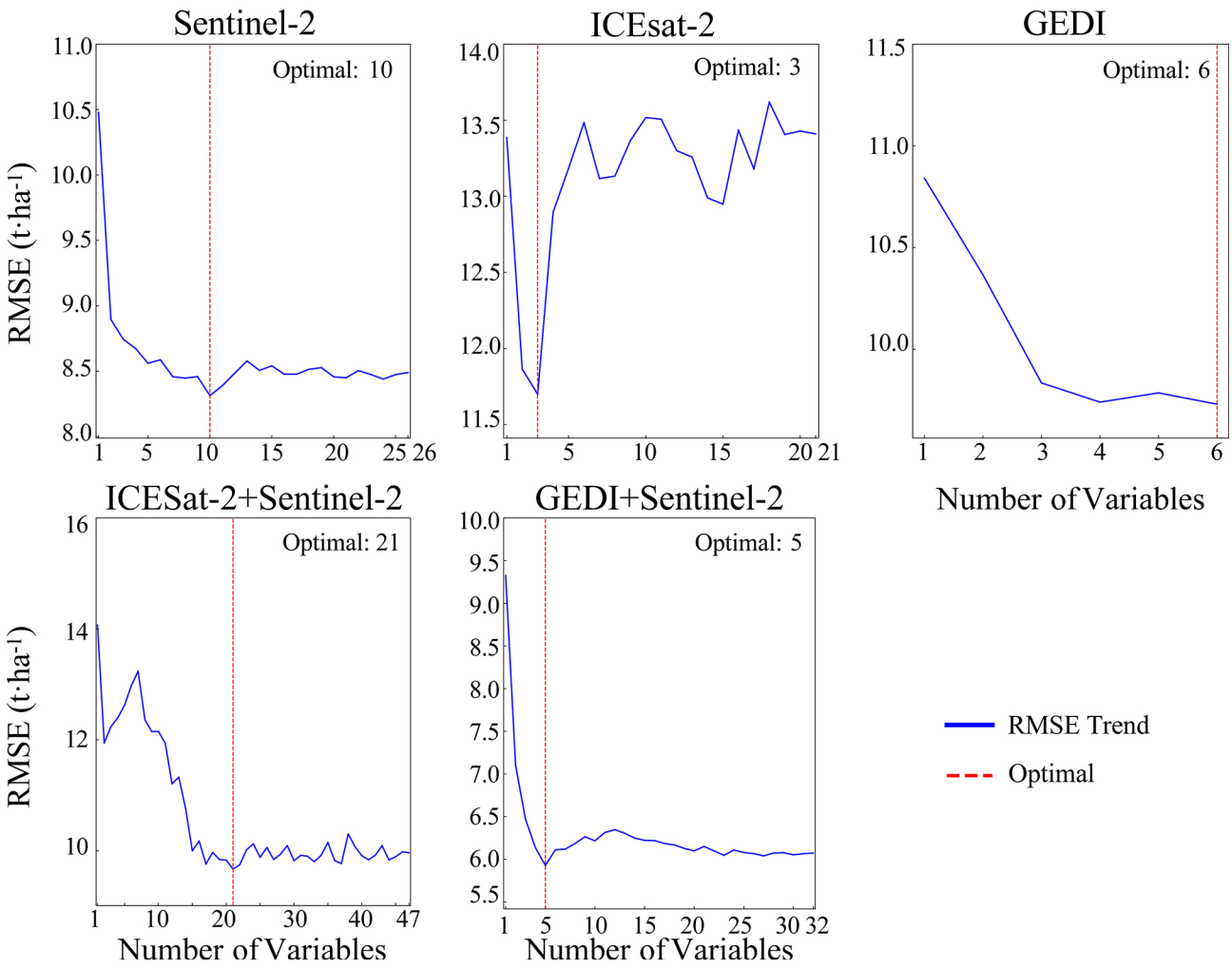

Fig. 5. Different variable combinations RMSE variation chart of RF model

Table 4. Summary table of random forest model accuracy validation results for different data sources and combinations

Data sources	Optimal Variables	R^2	RMSE (t · ha^{-1})	rRMSE(%)	MAE (t · ha^{-1})
Sentinel-2	CI, GRVI, CIRE, GNDVI, Elevation, NDVI45, B2, B11, NDVI705, EVI	0.70	5.27	19.87	3.53
ICESat-2	H_60, H_95, H_65	0.65	7.34	26.50	4.79
GEDI	rv, fhd, pta_4, pta_1, cover, pai	0.57	6.06	22.94	4.15
ICEsat-2+Sentinel-2	H_95, H_60, H_mean, H_75, H_45, H_80, H_65, CIRE, H_85, H_median, H_10, B2, H_90, NDVI45, B8, H_15, B12, H_30, H_50, B8A, B7	0.76	6.01	21.71	4.05
GEDI+Sentinel-2	CI, rv, fhd, NDVI45, CIRE	0.82	3.90	14.79	2.70

According to the accuracy validation results, when using feature parameters from single data sources – Sentinel-2, ICESat-2, and GEDI – as predictors in the Random Forest models, notable performance differences were observed. Sentinel-2 achieved the best fitting performance with an R^2 of 0.70, largely thanks to its rich multispectral features, which effectively capture the relationship between vegetation physiological status and biomass. ICESat-2 (R^2 = 0.65) and GEDI (R^2 = 0.57), limited by factors such as data resolution, coverage continuity, or feature dimensionality, did not fully leverage their structural information advantages when used alone.

When ICESat-2 and GEDI spaceborne LiDAR data were synergistically combined with Sentinel-2 spectral data, multi-source synergy led to significant improvements in fitting accuracy for both cases. Among the combined models, GEDI+Sentinel-2 performed best, achieving an R^2 of 0.82 and an RMSE of 3.90 t · ha^{-1}. This combination reached optimal performance with only five highly informative features, demonstrating efficient complementarity between GEDI canopy profile parameters (rv, fhd) and Sentinel-2 chlorophyll-sensitive indices (CI, CIRE, NDVI45) in characterizing forest structure and function. In comparison, although the ICESat-2+Sentinel-2 combination

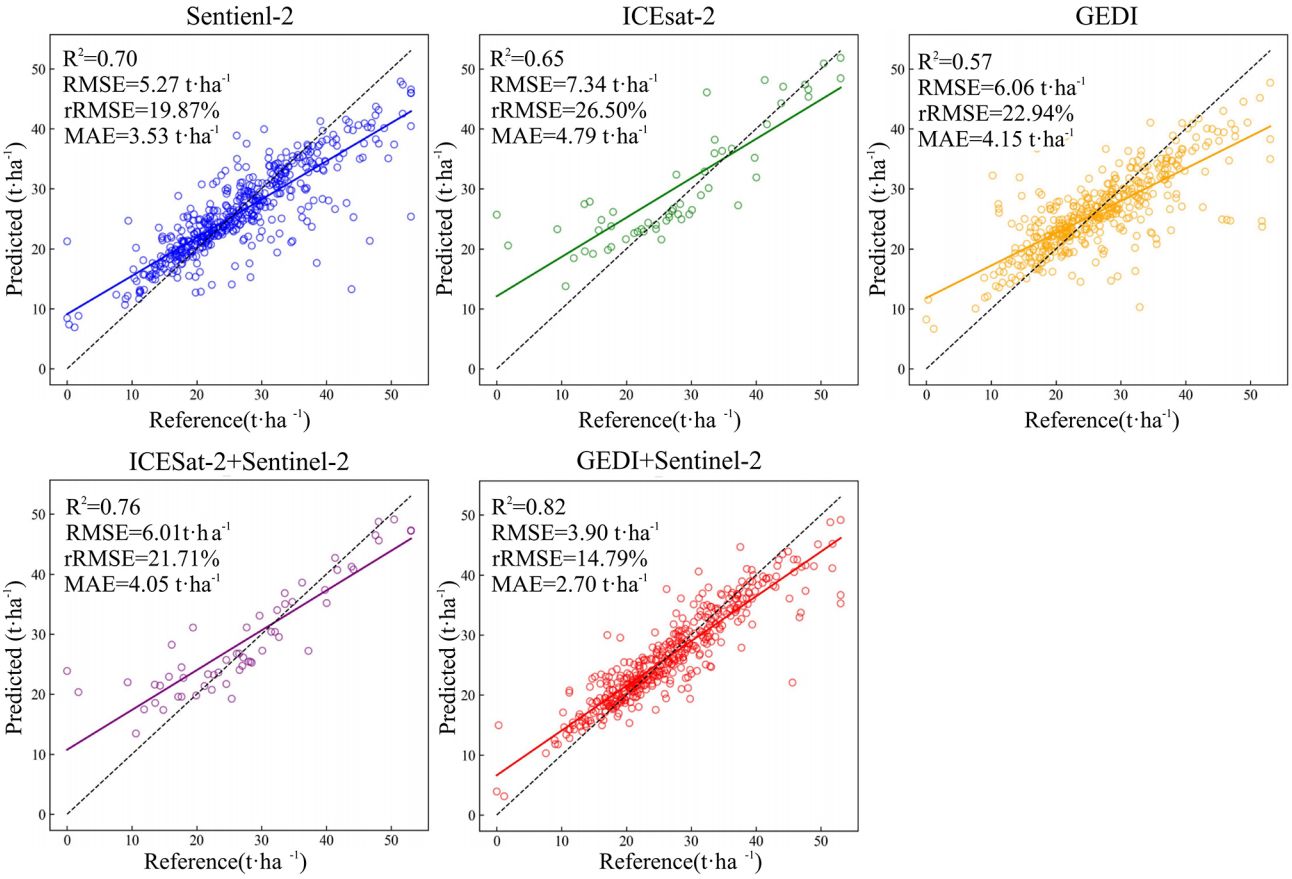

Fig. 6. Fitting plots of forest AGB based on different data source combinations using random forest models

also showed higher accuracy than the single-source models ($R^2 = 0.76$, RMSE = 6.01 t · ha^{-1}), it required more variables (21) and exhibited relatively lower accuracy, suggesting higher information redundancy in ICESat-2 height metrics and limited synergy efficiency with spectral features.

Forest AGB Spatial Distribution

To generate a continuous high-spatial-resolution map of forest aboveground biomass (AGB) across the study area, spatial extrapolation and mapping were performed using the optimal model identified in previous analyses – the Random Forest model integrating GEDI and Sentinel-2 data. It should be noted that key canopy structural parameters from GEDI, such as relative vertical energy distribution (rv) and foliage height diversity (fhd), are discrete point-based observations and do not have the continuous spatial coverage characteristic of Sentinel-2 spectral indices. To ensure the reasonableness and reliability of spatial interpolation, we first examined the spatial autocorrelation of the GEDI structural parameters. Global Moran's I analysis confirmed significant positive spatial autocorrelation for both rv ($I = 0.111715$, $p < 0.001$) and fhd ($I = 0.110375$, $p < 0.001$), meeting the basic requirements for spatial interpolation.

Based on this, spatialization was carried out using the Ordinary Kriging interpolation method. Through semivariogram analysis, the spherical model was selected as the optimal fitting model. Leave-one-out cross-validation yielded root-mean-square standardized error (RMSSE) values of 1.015 for rv and 0.9565 for fhd on the 25-meter resolution interpolated surfaces. The RMSSE measures the accuracy of the kriging variance estimates; a value close to 1 indicates that the prediction errors are well-calibrated and the interpolation model is reliable, whereas values significantly greater or less than 1 suggest over- or under-estimation of the prediction uncertainty, respectively (Wang et al., 2025). The obtained RMSSE values near 1 indicate reliable interpolation accuracy for both parameters.

The spatialized structural parameters were registered and fused with continuous data layers including Sentinel-2 spectral indices and topographic factors to form a complete predictive variable dataset. The trained Random Forest model was applied for pixel-level prediction across the entire study area, resulting in a preliminary AGB distribution map. After forest masking and statistical outlier removal, the final spatial distribution map of forest AGB for the study area was obtained (Fig. 7a).

Based on the final AGB distribution map, we further analyzed the statistical characteristics of AGB under different classifications of topographic factors (Table 5). The statistical results show that the mean AGB exhibits a clear increasing trend with rising elevation, from 22.80 t · ha^{-1} in areas below 2400 m to 29.92 t · ha^{-1} in the 2800–3000 m range. Across the slope gradient, AGB distribution displays

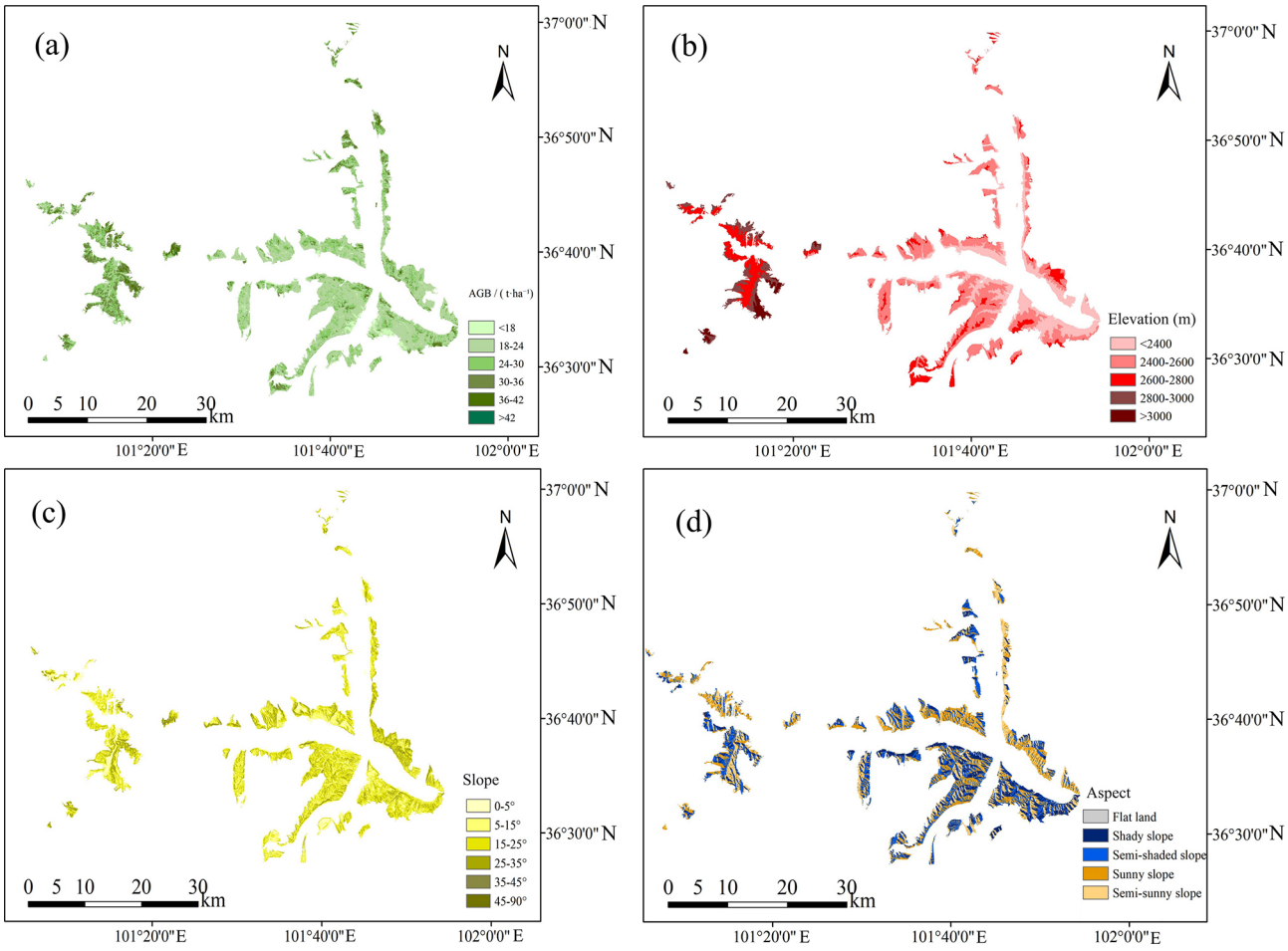

Fig. 7. Spatial distribution of forest AGB in the study area (a. Forest AGB Spatial Distribution; b. Elevation Spatial Distribution; c. Slope spatial distribution; d. Aspect spatial distribution)

Table 5. Statistical table of AGB by different grades of topographic factors

Terrain factors	Classification levels	Min (t · ha^{-1})	Max (t · ha^{-1})	Mean (t · ha^{-1})	Proportion
Elevation	<2400m	12.02	42.76	22.80	24.50%
	2400–2600m	13.42	42.20	23.92	41.71%
	2600–2800m	14.40	43.90	26.53	20.72%
	2800–3000m	14.04	42.40	29.92	9.24%
	>3000m	12.65	42.07	29.19	3.84%
Slope	0–5°	13.19	43.55	25.25	7.88%
	5–15°	12.02	43.02	25.26	32.20%
	15–25°	12.47	43.65	24.64	40.66%
	25–35°	12.98	43.90	23.86	16.61%
	35–45°	14.33	43.46	24.13	2.49%
	45–90°	16.29	42.20	25.79	0.17%
Aspect	Flat land	15.09	41.86	25.49	0.44%
	Shady slope (0–45°)	12.15	42.20	26.05	12.30%
	Semi–shaded slope (45–135°)	12.65	43.55	24.32	24.79%
	Sunny slope (135–225°)	12.91	43.02	23.61	23.53%
	Semi–sunny slope (225–315°)	12.41	43.90	25.06	28.01%
	Shady slope (315–360°)	12.02	42.99	26.07	10.92%

nonlinear characteristics, with gentle slopes (0–5°) and steep slopes (35–45°) showing relatively higher mean AGB, while areas with moderate slopes (15–25°) have the lowest mean AGB. Aspect significantly influences AGB distribution, with shady slopes (0–45° and 315–360°) exhibiting the highest mean AGB (26.05 t · ha^{-1} and 26.07 t · ha^{-1}, respectively), while sunny slopes show the lowest mean AGB (23.61 t · ha^{-1}).

Discussion

This study established a collaborative inversion framework for forest aboveground biomass (AGB) in the North and South Mountains of Xining City by integrating multi-source remote sensing data (Sentinel-2, GEDI, ICESat-2) with ground plot surveys, resulting in a high-spatial-resolution map of AGB distribution. Compared to existing regional studies that mainly rely on single optical data sources (e.g., Landsat) and traditional regression models (Yang et al., 2016), this research shows clear progress in data fusion strategy, model algorithm application, and interpretation of spatial results. Specifically, Yang et al. (2016) reported an R^2 of 0.554 and an average relative error of 13.509% for their Landsat-8-based biomass model in the same study area. In contrast, our multi-source fusion model (GEDI + Sentinel-2) achieved an R^2 of 0.82, representing a 48.0% relative improvement in explained variance. Moreover, our model's RMSE (3.90 t · ha^{-1}) is substantially lower than the error range implied by their model (e.g., mean biomass error around 1.60 t · ha^{-1} in validation). This marked enhancement in accuracy underscores the effectiveness of integrating spaceborne LiDAR structural metrics with optical spectral information, overcoming the saturation limitations of purely optical approaches and providing a more robust estimation framework for forest AGB in high-altitude arid regions. Our results also compare favorably with studies employing more complex data combinations. For instance, research in boreal forests of Northern China found the most effective model combined Sentinel-2, L-band SAR (ALOS/PALSAR), and topographic data (Li et al., 2024). Another study on the Google Earth Engine platform integrating Sentinel-2 spectral bands, vegetation indices, topography, and GEDI canopy height metrics reported a best-performing model R^2 of 0.77 (Zurqani, 2025). The accuracy of our GEDI+Sentinel-2 model (R^2 = 0.82) is comparable or superior, underscoring the high complementary value and efficiency of fusing GEDI's vertical structure metrics with Sentinel-2's spectral information, particularly red-edge indices.

Advantages of Multi-Source Data Fusion over Single Data Sources

This study demonstrates that integrating optical and LiDAR features can more comprehensively capture forest structural information, thereby significantly improving AGB inversion accuracy. While models using single data sources (Sentinel-2, ICESat-2, GEDI) provided baseline estimates for the region (R^2 = 0.70, 0.65, and 0.57, respectively), each had its limitations. The Sentinel-2 model, though performing best among single sources, may be affected by signal saturation in dense forest areas due to its reliance on spectral information (Wang et al., 2022; Pandit et al., 2018). Conversely, single spaceborne LiDAR datasets (ICESat-2, GEDI), constrained by discrete point sampling and limited spatial coverage continuity, did not fully realize the explanatory potential of their structural parameters when used alone (Mandl et al., 2023; Rajab Pourrahmati et al., 2023; Kombate et al., 2024).

Model performance improved systematically through data synergy. The combination of GEDI and Sentinel-2 data yielded the best results (R^2 = 0.82). This outcome is largely due to the complementary information from the two data types: parameters like the relative vertical energy distribution (rv) and foliage height diversity index (fhd) from GEDI directly characterize the three-dimensional structural complexity of forests, which is a key physical basis for AGB accumulation and less prone to saturation in high-biomass areas. In contrast, chlorophyll-related indices (CI, CIRE, NDVI45) derived from Sentinel-2 reflect the physiological status and photosynthetic capacity of vegetation. Their integration couples „structure–function" information, enabling the model to capture both physical structural and physio-ecological drivers of AGB simultaneously, thereby exceeding the explanatory power of models based solely on spectral or structural data (Chen et al., 2022; Qin et al., 2025; Fu et al., 2025).

The synergistic efficiency of ICESat-2 and Sentinel-2 was relatively lower. This may be because ICESat-2 primarily provides a series of height percentiles, whose information dimensions and representation differ from GEDI's canopy profile parameters. ICESat-2 metrics may be less refined than GEDI's in capturing vertical heterogeneity within the canopy, resulting in slightly weaker complementarity with spectral information and the introduction of more redundant variables.

Suitability of Machine Learning Models for Handling Multi-Source Data

The results suggest that machine learning models, particularly Random Forest (RF), show better fitting and generalization capabilities compared to traditional multiple linear regression (MLR) when handling multi-source, high-dimensional remote sensing data. In modeling based on UAV-LiDAR features, the accuracy of the RF model ($R^2 = 0.895$) was notably higher than that of MLR ($R^2 = 0.768$) and Support Vector Regression (SVR) ($R^2 = 0.780$). This difference comes down to algorithmic nature: MLR assumes linear relationships between variables, making it hard to capture the inherent nonlinear, high-order interactions between AGB and complex forest structural and spectral features. As an ensemble learning algorithm, RF can approximate complex nonlinear relationships by building many decision trees, shows robustness to high-dimensional data and multicollinearity, and its built-in feature importance assessment helps understand variable contributions (Cao et al., 2018; Liu et al., 2017; Li et al., 2024). In multi-source data fusion scenarios, where variable interactions are more complex, these advantages make RF a more suitable choice.

Spatial Fusion Based on Geostatistics and Ecological Interpretation of AGB Spatial Patterns

To generate continuous surfaces from the discrete GEDI metrics (rv and fhd), we employed Ordinary Kriging (OK) interpolation (Section 3.3). This method assumes stationarity (constant mean and variance) and spatial continuity of the interpolated variable. In our mountainous study area, these assumptions were preliminarily validated by the significant spatial autocorrelation observed in the GEDI parameters (Global Moran's I, $p < 0.001$). The spherical variogram model provided the best fit, and leave-one-out cross-validation yielded reliable accuracy (RMSSE ≈ 1).

However, the application of OK in complex terrain requires careful consideration. While it performed well within the sampled domain, the method may smooth local variations in areas of extreme heterogeneity or sparse data coverage. Its assumption of a uniform spatial process might not fully capture non-stationary trends driven by sharp topographic gradients. Future studies could explore more adaptive methods, such as regression kriging that explicitly incorporates elevation and aspect as covariates, to potentially enhance spatialization in complex landscapes.

By combining spatial autocorrelation analysis with Ordinary Kriging interpolation, this study successfully spatialized discrete point-based structural parameters from GEDI into continuous surfaces, subsequently generating a continuous AGB distribution map for the entire study area (Fig. 7a). The spatialized AGB distribution clearly shows its strong relationship with topographic factors (Table 5), offering useful insights into the spatial heterogeneity of regional forest carbon sinks.

The increasing trend of AGB with elevation may be related to reduced human disturbance at higher altitudes (Salinas-Melgoza et al., 2018). The influence of aspect stands out: mean AGB was highest on shaded slopes (0–45°, 315–360°) and lowest on sunny slopes. This is mainly due to the intense solar radiation and water stress characteristic of the semi-arid region on the northeastern margin of the Tibetan Plateau. Shaded slopes receive less solar radiation, have lower evapotranspiration water loss, and generally better soil moisture conditions, which favor tree growth and biomass accumulation (Klinge et al., 2015; Warren II, 2010).

The effect of slope showed a nonlinear pattern: relatively higher AGB on gentle (0–5°) and steep slopes (35–45°), and the lowest on moderate slopes (15–25°). This pattern may be explained by a combination of edaphic and anthropogenic factors: gentle slopes typically support deeper, more stable soils conducive to root development and biomass accumulation (Salinas-Melgoza et al., 2018), whereas steeper slopes are often less accessible and may experience reduced anthropogenic pressure, allowing forest stands to develop with less disturbance. Moderate slope areas might have been more affected historically by erosion or human activities.

These spatial pattern analyses indicate that in this ecologically fragile study area, topography – through regulating water and heat redistribution – is a key environmental factor driving the spatial variation of forest AGB. This provides a direct scientific basis for future refined and differentiated forest management and carbon sink strategies, such as developing tending or supplemental planting plans tailored to different aspects and elevations.

Methodological Justification for the Hierarchical Framework

The hierarchical "plot–UAV–spaceborne" framework adopted in this study was motivated by the spatial and structural gaps between field observations and satellite data. Spaceborne LiDAR data are inherently discontinuous, making direct plot–satellite matching inefficient and error-prone. By introducing UAV-LiDAR as an intermediary, we effectively created

a continuous AGB reference layer that is both spatially extensive and structurally detailed. This layer serves as a "bridge" to calibrate GEDI and Sentinel-2 data, mitigating the need for coincident ground plots and reducing the impact of geolocation errors. While this approach may introduce additional uncertainty from UAV data processing, our rigorous validation at each stage indicates that the framework remains robust for regional AGB estimation in complex terrain. We first established a high-fidelity model (R^2 = 0.90) at the plot scale using UAV-LiDAR, which then served as a bridge to generate high-confidence training data for the regional satellite-based model. This approach effectively addresses common challenges of scarce ground truth data and uncertainty propagation in large-scale AGB mapping. Similar hierarchical strategies, using LiDAR to "enhance" plot data, have been successfully applied in other regions, such as Northern Chinese forests and tropical areas, leading to significant accuracy improvements (R^2 increases from 0.65–0.75 to around 0.82) (Liu et al., 2025; Tian et al., 2024). Our work validates the applicability of this framework in high-altitude arid regions.

Future Perspectives

Future research could explore several directions. First, temporal analysis using multi-temporal remote sensing data could investigate dynamic changes in AGB since afforestation projects began and their response to climate change (Lucas et al., 2021). Second, integrating synthetic aperture radar (SAR) data, such as from Sentinel-1, could provide complementary advantages. SAR signals, operating in the microwave spectrum, can penetrate cloud cover and are less affected by atmospheric conditions, ensuring data availability in persistently cloudy regions. Furthermore, the backscatter intensity and interferometric coherence from SAR are sensitive to forest vertical structure and moisture content, offering a potential means to mitigate the saturation effects commonly encountered with optical vegetation indices in high-biomass forests (Doblas et al., 2020; Mauya et al., 2019). To effectively integrate SAR data into the existing multi-source framework, we propose a feature-level fusion strategy combining SAR-derived metrics – such as dual-polarization backscatter coefficients (VV, VH), interferometric coherence, and texture measures – with optical spectral indices and LiDAR structural profiles. A possible workflow includes: extracting SAR features over the study area and resampling them to match the spatial resolution of Sentinel-2 (e.g., 25 m); performing correlation and importance analysis (e.g., using Random Forest) to identify SAR features that are complementary to existing optical and LiDAR variables; incorporating selected SAR features into the existing machine learning model (e.g., Random Forest or deep learning fusion networks) to evaluate their contribution to AGB estimation accuracy, particularly in high-biomass or topographically complex areas. Preliminary experiments could focus on a subset of the study area with concurrent Sentinel-1, Sentinel-2, and GEDI acquisitions to validate the added value of SAR in reducing model uncertainty. Finally, applying the model framework from this study to other similar ecological zones on the Tibetan Plateau could test its generalizability and robustness, supporting the development of a regional or broader forest carbon monitoring network.

Conclusions

This study established a multi-source remote sensing collaborative inversion framework for forest aboveground biomass (AGB) in the North and South Mountains of Xining City, located along the northeastern edge of the Qinghai-Tibet Plateau. The framework integrates Sentinel-2 multispectral imagery, GEDI and ICESat-2 spaceborne LiDAR data, and ground-based plot measurements. The main findings are as follows:

At the local scale, the Random Forest model utilizing airborne LiDAR-derived canopy structural features achieved the highest accuracy (R^2 = 0.90, RMSE = 2.14 t · ha^{-1}). This demonstrates the capability of high-resolution UAV-LiDAR in accurately characterizing forest vertical structure and providing a reliable reference for AGB estimation.

At the regional scale, data fusion significantly improved estimation performance. The combination of GEDI canopy profile metrics (e.g., rv, fhd) and Sentinel-2 spectral indices (e.g., CI, CIRE) yielded the optimal result (R^2 = 0.82, RMSE = 3.90 t · ha^{-1}), substantially outperforming models based on any single data source. This confirms the effectiveness of synergistically integrating vertical structure information from spaceborne LiDAR with spectral information from optical imagery.

Machine learning algorithms, particularly Random Forest, outperformed traditional multiple linear regression in handling the multi-source, high-dimensional feature space and capturing the complex nonlinear relationships between AGB and predictor variables.

By integrating spatial autocorrelation analysis with Ordinary Kriging interpolation, discrete GEDI structural parameters were successfully spatialized. The resulting 25-meter resolution AGB distribution map revealed clear spatial patterns: biomass generally increases with elevation and is significantly higher on shaded slopes (mean: 26.06 t · ha^{-1}) than on sunlit slopes (mean: 23.61 t · ha^{-1}).

In summary, this study confirms that a hierarchical estimation framework – using airborne LiDAR-derived AGB as a high-precision intermediary to link plot measurements with satellite observations – is effective. The synergistic integration of Sentinel-2 and GEDI data through machine learning algorithms provides a reliable and efficient technical solution for high-precision AGB estimation in ecologically fragile, high-altitude arid regions. The generated AGB distribution map and the analysis of its relationship with topography offer a valuable data basis and scientific reference for regional forest carbon stock assessment and informed forest management.

Author Contributions

Conceptualization, Hongke Hao and Yao Dong; investigation, Le Yang, Mymuna Islam Moon and Yang Liu; methodology, Hongke Hao and Yao Dong; resources, Hongke Hao; software, Yao Dong, Le Yang, Mymuna Islam Moon and Yang Liu; validation, Hongke Hao; formal analysis, Yao Dong and Hongke Hao; data curation, Hongke Hao and Yao Dong; writing – original draft preparation, Hongke Hao, Yao Dong, Le Yang, Mymuna Islam Moon and Yang Liu; visualization, Le Yang and Yang Liu. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We sincerely thank the National Snow and Ice Data Center (NSIDC) for the ICESat-2/ATLAS data, the NASA Goddard Space Flight Center and GEDI Science Team for providing and maintaining GEDI data products, the European Space Agency (ESA) for Sentinel-2 imagery via the Copernicus Open Access Hub, and NASA's Shuttle Radar Topography Mission for SRTM DEM data. We are grateful to the National Technical Committee for Standardization of Forest Resources for the allometric equations in GB/T 43648-2024, the field and laboratory team members for their diligent work, and the editor and anonymous reviewers for their constructive comments. This work was supported by the Tree Fall Risk Assessment Algorithm Technical Service Contract (2026006030200213) and the Forest, Grassland, and Wetland Survey Software Development Contract (TG20241146).

Funding

Tree Fall Risk Assessment Algorithm Technical Service Contract (2026006030200213); Forest, Grassland, and Wetland Survey Software Development Contract (TG20241146).

Institutional review board statement
Not applicable.

Informed consent statement
Not applicable.

Data availability statement
Not applicable.

Conflicts of interest
The authors declare no conflict of interest.

References

Adugna T, Xu W & Fan J (2022) Comparison of random forest and support vector machine classifiers for regional land cover mapping using coarse resolution FY-3C images. Remote Sensing 14: 574. doi:10.3390/RS14030574.

Ameray A, Bergeron Y, Valeria O, Montoro Girona M & Cavard X (2021) Forest carbon management: a review of silvicultural practices and management strategies across boreal, temperate and tropical forests. Current Forestry Reports 7: 245–266. doi:10.1007/S40725-021-00151-W.

Banerjee A, Singh P & Pratap K (2017) Morphometric evaluation of Swarnrekha watershed, Madhya Pradesh, India: an integrated GIS-based approach. Applied Water Science 7: 1807–1815. doi:10.1007/s13201-015-0354-3.

Campbell MJ, Dennison PE, Kerr KL, Brewer SC & Anderegg WR (2021) Scaled biomass estimation in woodland ecosystems: testing the individual and combined capacities of satellite multispectral and lidar data. Remote Sensing of Environment 262: 112511. doi:10.1016/J.RSE.2021.112511.

Cao L, Pan J, Li R, Li J & Li Z (2018) Integrating airborne LiDAR and optical data to estimate forest aboveground biomass in arid and semi-arid regions of China. Remote Sensing 10: 532. doi:10.3390/rs10040532.

Chen L, Ren C, Bao G, Zhang B, Wang Z, Liu M, Man W & Liu J (2022) Improved object-based estimation of forest aboveground biomass by integrating LiDAR data from GEDI and ICESat-2 with multi-sensor images in a heterogeneous mountainous region. Remote Sensing 14: 2743. doi:10.3390/RS14122743.

Doblas J, Shimabukuro Y, Sant'Anna S, Carneiro A, Aragão L & Almeida C (2020) Optimizing near real-time detection of deforestation on tropical rainforests using sentinel-1 data. Remote Sensing 12: 3922. doi:10.3390/rs12233922.

Dubayah R, Blair JB, Goetz S, Fatoyinbo L, Hansen M, Healey S, Hofton M, Hurtt G, Kellner J, Luthcke S, Armston J, Tang H, Duncanson L, Han-

cock S, Jantz P, Marselis S, Patterson PL, Qi W & Silva C (2020) The Global Ecosystem Dynamics Investigation: high-resolution laser ranging of the Earth's forests and topography. Science of Remote Sensing 1: 100002. doi:10.1016/j.srs.2020.100002.

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

Fu L, Shu Q, Xia C, Li Z, He H, Li Z, Ma S, Qin C, Wei R, Xiang Q, Zhang X, Zhang Y & Cai H (2025) A robust framework for bamboo forest AGB estimation by integrating geostatistical prediction and ensemble learning. Remote Sensing 17: 2682. doi:10.3390/RS17152682.

Georgopoulos N, Gitas IZ, Korhonen L, Antoniadis K & Stefanidou A (2023) Estimating crown biomass in a multilayered fir forest using airborne LiDAR data. Remote Sensing 15: 2919. doi:10.3390/RS15112919.

Hu T, Zhang Y, Su Y, Zheng Y, Lin G & Guo Q (2020) Mapping the global mangrove forest aboveground biomass using multisource remote sensing data. Remote Sensing 12: 1690. doi:10.3390/rs12101690.

Jepma CJ, Nilsson S, Amano M, Bonduki Y, Lönnstedt L, Sathaye J & Wilson T (2020) Carbon sequestration and sustainable forest management: common aspects and assessment procedures: Economics of Carbon Sequestration in Forestry. CRC Press, Boca Raton, FL, USA, pp. S83–S96.

Kacic P, Thonfeld F, Gessner U & Kuenzer C (2023) Forest structure characterization in Germany: novel products and analysis based on GEDI, Sentinel-1 and Sentinel-2 data. Remote Sensing 15: 1969. doi:10.3390/RS15081969.

Klinge M, Böhner J & Erasmi S (2015) Modeling forest lines and forest distribution patterns with remote-sensing data in a mountainous region of semiarid central Asia. Biogeosciences 12: 2893–2905. doi:10.5194/bg-12-2893-2015.

Kombate A, Fotso Kamga GA & Goïta K (2024) Modeling canopy height of forest–savanna mosaics in Togo using ICESat-2 and GEDI spaceborne LiDAR and multisource satellite data. Remote Sensing 17: 85. doi:10.3390/RS17010085.

Li C, Li Y & Li M (2019) Improving forest aboveground biomass (AGB) estimation by incorporating crown density and using landsat 8 OLI images of a subtropical forest in Western Hunan in Central China. Forests 10: 104. doi:10.3390/f10020104.

Li H, Tang X, Cui L, Zhai X, Wang J, Zhao X, Li J, Lei Y, Wang J, Wang R & Li W (2024) Estimating aboveground biomass of wetland plant communities from hyperspectral data based on fractional-order derivatives and machine learning. Remote Sensing 16: 3011. doi:10.3390/RS16163011.

Li J, Bao W, Wang X, Song Y, Liao T, Xu X & Guo M (2024) Estimating aboveground biomass of boreal forests in Northern China using multiple datasets. IEEE Transactions on Geoscience and Remote Sensing 62: 1–10. doi:10.1109/TGRS.2024.3408316.

Li X, Zhang M, Long J & Lin H (2021) A novel method for estimating spatial distribution of forest aboveground biomass based on multispectral fusion data and ensemble learning algorithm. Remote Sensing 13: 3910. doi:10.3390/RS13193910.

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

Liu P, Ren C, Wang Z & Jia M (2025) Improving aboveground biomass estimation by integrating ICESat-2 LiDAR sampling and tree species stratification strategies using multi-source datasets. Geo-spatial Information Science. doi:10.1080/10095020.2025.2516609.

Lucas R, Otero V, Van De Kerchove R, Lagomasino D, Satyanarayana B, Fatoyinbo T & Dahdouh-Guebas F (2021) Monitoring Matang's mangroves in Peninsular Malaysia through Earth observations: a globally relevant approach. Land Degradation & Development 32: 354–373. doi:10.1002/ldr.3652.

Mandl L, Stritih A, Seidl R, Ginzler C & Senf C (2023) Spaceborne LiDAR for characterizing forest structure across scales in the European Alps. Remote Sensing in Ecology and Conservation 9: 599–614. doi:10.1002/RSE2.330.

Mauya EW, Koskinen J, Tegel K, Hämäläinen J, Kauranne T & Käyhkö N (2019) Modelling and predicting the growing stock volume in small-scale plantation forests of Tanzania using multi-sensor image synergy. Forests 10: 279. doi:10.3390/f10030279.

Messier C, Bauhus J, Sousa-Silva R, Auge H, Baeten L, Barsoum N, Bruelheide H, Caldwell B, Cavender-Bares J, Dhiedt E, Eisenhauer N, Ganade G, Gravel D, Guillemot J, Hall JS, Hector A, Hérault B, Jactel H, Koricheva J, Kreft H, Mereu S, Muys B, Nock CA, Paquette A, Parker JD, Perring MP, Ponette Q, Potvin C, Reich PB, Scherer-Lorenzen M, Schnabel F, Verheyen K, Weih M, Wollni M & Zemp DC (2022) For the sake of resilience and multifunctionality, let's diversify planted forests! Conservation Letters 15: e12829. doi:10.1111/CONL.12829.

Moudrý V, Prošek J, Marselis S, Marešová J, Šárovcová E, Gdulová K, Kozhoridze G, Torresani M, Rocchini D, Eltner A, Liu X, Potůčková M, Šedová A, Crespo-Peremarch P, Torralba J, Ruiz LA,

Perrone M, Špatenková O & Wild J (2024) How to find accurate terrain and canopy height GEDI footprints in temperate forests and grasslands? Earth and Space Science 11: e2024EA003709. doi:10.1029/2024EA003709.

Nam VT, Van Kuijk M & Anten NP (2016) Allometric equations for aboveground and belowground biomass estimations in an evergreen forest in Vietnam. PLOS ONE 11: e0156827. doi:10.1371/journal.pone.0156827.

Pandit S, Tsuyuki S & Dube T (2018) Estimating above-ground biomass in sub-tropical buffer zone community forests, Nepal, using Sentinel 2 data. Remote Sensing 10: 601. doi:10.3390/rs10040601.

Pham TD, Yokoya N, Bui DT, Yoshino K & Friess DA (2019) Remote sensing approaches for monitoring mangrove species, structure, and biomass: opportunities and challenges. Remote Sensing 11: 230. doi:10.3390/rs11030230.

Possu WB, Brandle JR, Domke GM, Schoeneberger M & Blankenship E (2016) Estimating carbon storage in windbreak trees on US agricultural lands. Agroforestry Systems 90: 889–904. doi:10.1007/s10457-016-9896-0.

Pu R (2021) Mapping tree species using advanced remote sensing technologies: a state-of-the-art review and perspective. Journal of Remote Sensing 2021: 9812624. doi:10.34133/2021/9812624.

Qin Z, Yang H, Shu Q, Yu J, Yang Z, Ma X & Duan D (2025) Estimation of *Dendrocalamus giganteus* leaf area index by combining multi-source remote sensing data and machine learning optimization model. Frontiers in Plant Science 15: 1505414. doi:10.3389/FPLS.2024.1505414.

Rajab Pourrahmati M, Baghdadi N & Fayad I (2023) Comparison of GEDI LiDAR data capability for forest canopy height estimation over broadleaf and needleleaf forests. Remote Sensing 15: 1522. doi:10.3390/RS15061522.

Salinas-Melgoza MA, Skutsch M & Lovett JC (2018) Predicting aboveground forest biomass with topographic variables in human-impacted tropical dry forest landscapes. Ecosphere 9: e02063. doi:10.1002/ecs2.2063.

Tian X, Li J, Zhang F, Zhang H & Jiang M (2024) Forest aboveground biomass estimation using multisource remote sensing data and deep learning algorithms: a case study over Hangzhou Area in China. Remote Sensing 16: 1074. doi:10.3390/rs16061074.

Vafaei S, Soosani J, Adeli K, Fadaei H, Naghavi H, Pham TD & 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 Sensing 10: 172. doi:10.3390/rs10020172.

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 Sensing 14: 2146. doi:10.3390/RS14092146.

Wang G, Guan D, Peart MR, Chen Y & Peng Y (2013) Ecosystem carbon stocks of mangrove forest in Yingluo Bay, Guangdong Province of South China. Forest Ecology and Management 310: 539–546. doi:10.1016/j.foreco.2013.08.045.

Wang Q, Putri NA, Gan Y & Song G (2022) Combining both spectral and textural indices for alleviating saturation problem in forest LAI estimation using Sentinel-2 data. Geocarto International 37: 10511–10531. doi:10.1080/10106049.2022.2037730.

Wang X, Liu C, Lv G, Xu J & Cui G (2022) Integrating multi-source remote sensing to assess forest aboveground biomass in the Khingan mountains of north-eastern China using machine-learning algorithms. Remote Sensing 14: 1039. doi:10.3390/rs14041039.

Wang Y, Yuan F, Cammarano D, Liu X, Tian Y, Zhu Y, Cao W & Cao Q (2025) Integrating machine learning with spatial analysis for enhanced soil interpolation: balancing accuracy and visualization. Smart Agricultural Technology 10: 101032. doi:10.1016/J.ATECH.2025.101032.

Warren II RJ (2010) An experimental test of well-described vegetation patterns across slope aspects using woodland herb transplants and manipulated abiotic drivers. New Phytologist 185: 1038–1049. doi:10.1111/j.1469-8137.2009.03147.x.

Xiong H, Chang B, Lan X, Zhou H, Chen Y & Zhang W (2025) Up-to-date high-resolution understory terrain extraction based on satellite stereo photogrammetry and spaceborne LiDAR. Forest Ecosystems 14: 100372. doi:10.1016/J.FECS.2025.100372.

Yan M, Xia Y, Yang X, Wu X, Yang M, Wang C, Hou Y & Wang D (2023) Biomass estimation of subtropical arboreal forest at single tree scale based on feature fusion of airborne liDAR data and aerial images. Sustainability 15: 1676. doi:10.3390/SU15021676.

Yang W, Zhao P, Xue D, Hou Y, Zhang X & Wang Z (2016) Estimation of forest biomass model of North and South Mountains based on Landsat-8 remote sensing image data in Xining. Journal of Northwest Forestry University 31: 33–37. doi:10.3969/j.issn.1001-7461.2016.02.06.

Zhang L, Zhao Y, Chen C, Li X, Mao F, Lv L, Yu J, Song M, Huang L, Chen J, Zheng Z & Du H (2024) UAV-LiDAR integration with Sentinel-2 enhances precision in AGB estimation for bamboo

forests. Remote Sensing 16: 705. doi:10.3390/RS16040705.

Zhang Z, Chen G, Bo Y, Guo X & Bao J (2022) Performance evaluation of combining ICESat-2 and GEDI laser altimetry missions for inland lake level retrievals. Geoscience Letters 9: 35. doi:10.1186/S40562-022-00243-W.

Zhao Z, Chuah JH, Lai KW, Chow CO, Gochoo M, Dhanalakshmi S, Wang N, Bao W & Wu X (2023) Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: a review. Frontiers in Computational Neuroscience 17: 1038636. doi:10.3389/FNCOM.2023.1038636.

Zhu J, Huang Z, Sun H & Wang G (2017) Mapping forest ecosystem biomass density for Xiangjiang river basin by combining plot and remote sensing data and comparing spatial extrapolation methods. Remote Sensing 9: 241. doi:10.3390/rs9030241.

Zurqani HA (2025) A multi-source approach combining GEDI LiDAR, satellite data, and machine learning algorithms for estimating forest aboveground biomass on Google Earth Engine platform. Ecological Informatics 86: 103052. doi:10.1016/J.ECOINF.2025.103052.