

Article

Not peer-reviewed version

A Robust Framework for Bamboo Forest AGB Estimation by Integrating Geostatistical Prediction and Ensemble Learning

[Lianjin Fu](#), [Qingtai Shu](#)^{*}, [Cuifen Xia](#), [Zeyu Li](#), Hailing He, Zhengying Li, Shaoyang Ma, ChaoGuan Qin, [Rong Wei](#), Qin Xiang, Xiao Zhang, [Yiran Zhang](#), Huashi Cai

Posted Date: 3 July 2025

doi: 10.20944/preprints202507.0301.v1

Keywords: biomass; ICESat-2; GEDI; Sentinel-2; stacking ensemble learning; interpretable machine learning; SHAP value



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

A Robust Framework for Bamboo Forest AGB Estimation by Integrating Geostatistical Prediction and Ensemble Learning

Lianjin Fu, Qingtai Shu *, Cuifen Xia, Zeyu Li, Hailing He, Zhengying Li, Shaoyang Ma, ChaoGuan Qin, Rong Wei, Qin Xiang, Xiao Zhang, Yiran Zhang and Huashi Cai

Southwest Forestry University, Kunming 650224, China

* Correspondence: shuqt@swfu.edu.cn

Abstract

Accurate above-ground biomass (AGB) quantification is confounded by signal saturation and data fusion challenges, particularly in structurally complex ecosystems like bamboo forests. To address these gaps, this study developed a two-stage framework to map the AGB of *Dendrocalamus giganteus* in a subtropical mountain environment. This study first employed Empirical Bayesian Kriging Regression Prediction (EBKRP) to spatialize sparse GEDI and ICESat-2 LiDAR metrics using Sentinel-2 and topographic covariates. Subsequently, a stacked ensemble model, integrating four machine learning algorithms, predicted AGB from the full suite of continuous variables. The stacking model achieved high predictive accuracy ($R^2 = 0.84$, $RMSE = 11.07 \text{ Mg ha}^{-1}$) and substantially mitigated the common bias of underestimating high AGB, improving the predicted-observed regression slope from a base-model average of 0.63 to 0.81. Furthermore, SHAP analysis provided mechanistic insights, identifying canopy photon rate as the dominant predictor and quantifying the ecological thresholds governing AGB distribution. The mean AGB density was $71.8 \pm 21.9 \text{ Mg ha}^{-1}$, with its spatial pattern influenced by elevation and human settlements. This research provides a robust framework for synergizing multi-source remote sensing data to improve AGB estimation, offering a refined methodological pathway for large-scale carbon stock assessments.

Keywords: biomass; ICESat-2; GEDI; Sentinel-2; stacking ensemble learning; interpretable machine learning; SHAP value

1. Introduction

Forest ecosystems cover approximately one-third of the Earth's terrestrial surface and are integral to the global carbon budget [1]. Accurate quantification of Above-Ground Biomass (AGB) at broad spatial scales is essential for advancing understanding of the global carbon cycle and for formulating effective emission reduction strategies [2]. This information also underpins sustainable forest management and ecological conservation efforts [3]. However, conventional methods for AGB estimation, which rely on field-inventory data and allometric scaling, are constrained by high costs, logistical challenges, and limited applicability in remote or inaccessible regions [4,5].

The application of remote sensing offers a powerful alternative for large-scale AGB mapping. Although optical sensors such as Sentinel-2 can provide spatially contiguous data, their signals are prone to saturation at relatively low levels of Above-Ground Biomass Density (AGBD), a phenomenon particularly acute in optical data due to limited canopy penetration. While Synthetic Aperture Radar (SAR) systems offer improved penetration, saturation still occurs, with the point of saturation being dependent on the radar wavelength used [6]. Light Detection and Ranging (LiDAR) technology overcomes this limitation. By directly quantifying vertical forest structure, LiDAR sensors maintain sensitivity in high-biomass forests and are not susceptible to signal saturation [7]. The potential of spaceborne LiDAR is highlighted by missions such as NASA's Global Ecosystem

Dynamics Investigation (GEDI) and the Ice, Cloud, and land Elevation Satellite-2 (ICESat-2). The GEDI mission uses a full-waveform 1,064 nm laser to characterize forest structure, while ICESat-2 employs photon-counting technology with its 532 nm laser to acquire high-resolution elevation data globally [8,9].

However, these spaceborne LiDAR missions possess distinct limitations. GEDI's orbital inclination limits data acquisition to latitudes between $\pm 51.6^\circ$, while the ICESat-2 signal can be attenuated in dense canopies [10]. Furthermore, both missions employ a sparse, along-track sampling strategy and thus cannot directly generate spatially contiguous maps. Conversely, while multispectral imagery from sensors like Sentinel-2 and Landsat offers high temporal resolution and rich spectral data [11], it lacks the ability to directly measure the vertical canopy structure that is highly correlated with biomass. Therefore, producing wall-to-wall AGBD maps from LiDAR data necessitates either spatial interpolation of the sparse samples [12,13] or data fusion with spatially contiguous predictor variables from other remote sensing sources [14,15].

Consequently, data fusion that integrates vertical structure information from LiDAR with spectral data from optical sensors has emerged as a primary strategy for enhancing the accuracy of AGB estimation. Recent studies demonstrate the efficacy of this fusion approach. For instance, Silva et al. (2021) developed empirical AGB models using simulated GEDI and ICESat-2 data in Sonoma County, USA. Elsewhere, Qi et al. (2019) fused GEDI LiDAR samples with Interferometric SAR (InSAR) data to evaluate the potential for regional AGB extrapolation. Innovations in modelling algorithms have also contributed to improved estimation accuracy. Seely et al. (2024), for instance, achieved high model accuracy ($R^2 = 0.77$; $RMSE = 28.38 \text{ Mg ha}^{-1}$) using deep neural networks, and the multi-source ensemble learning framework proposed by Chen et al. (2022) also yielded robust results. At a regional scale, Duncanson et al. (2020) established a multi-sensor collaborative inversion framework for major forest types across North America, demonstrating a viable pathway for large-scale AGB mapping.

Despite this progress, two critical limitations persist in the current body of research. Methodologically, many fusion techniques rely on simple feature concatenation, which may not fully exploit the complex, nonlinear complementarities inherent in multi-source datasets [16]. Stacking generalization, an ensemble learning technique, offers a promising alternative. This approach uses base-learners to process individual data sources and a meta-learner to optimally combine their outputs, which has been shown to mitigate overfitting and improve AGB retrieval accuracy in topographically complex areas. From an application perspective, the majority of spaceborne LiDAR fusion research has concentrated on arboreal forests, while bamboo ecosystems—a globally significant carbon sink—have been comparatively overlooked.

As a distinct forest type, bamboo presents unique challenges for AGB estimation. The dense, monolayered canopy structure characteristic of bamboo attenuates LiDAR returns and reduces the sensitivity of optical indices by over 40% [21,22], with signals from both sensor types becoming saturated in mature stands [23]. These challenges are amplified in the mountainous bamboo forests of Xinping County, Yunnan, China, which are characterized by rugged terrain (mean slope: 24°). Here, complex topographic effects interact with rapid bamboo phenological cycles (e.g., shooting from March to May; elongation from June to September) to further compound AGB estimation uncertainty. Steep slopes can induce a broadening of the GEDI waveform and an attenuation of ICESat-2 photon density (leading to errors $>30\%$), while topographic shadowing can obscure optical signatures corresponding to key phenological stages [24–26]. This confluence of factors makes the accurate AGB estimation of mountainous bamboo forests a pressing scientific challenge.

To address these challenges, this study introduces a multi-source remote sensing framework for AGB estimation based on stacking generalization. The framework employs Level-0 base models to fuse structural parameters from ICESat-2 and GEDI with temporal spectral features from Sentinel-2, thereby capturing the distinct vertical structure and phenological dynamics of bamboo forests. Subsequently, a Level-1 meta-model uses ridge regression to optimally weigh and integrate the predictions from the base models. By leveraging the complementary strengths of LiDAR and optical

data, this approach provides a robust and generalizable methodology for accurate AGB mapping in mountainous bamboo forests. Furthermore, to move beyond prediction and provide mechanistic insight, this study integrates the SHAP (SHapley Additive exPlanations) framework to interpret the model's behavior, identifying key drivers and their non-linear effects on AGB. This framework not only establishes a methodological foundation for future multi-source integration studies on carbon stocks but also contributes to the construction of a local-scale Digital Earth to support sustainable forest management.

2. Materials and Methods

2.1. Study Area

The study area encompasses Xinping County, Yuxi City, Yunnan Province ($23^{\circ}38'15''$ – $24^{\circ}26'05''$ N, $101^{\circ}16'30''$ – $102^{\circ}16'50''$ E), located at the eastern piedmont of the Ailao Mountain structural belt within the western Yunnan fold system (Figure 1). The region is characterized by deeply incised mid-mountain terrain (elevation range: 373–3119 m; mean: 1485 m) with considerable topographic relief and a surface fragmentation index of 0.68. Influenced by both the Indian and Pacific Ocean monsoons, the area exhibits a dry-hot valley climate, with a mean annual temperature of 19.5°C and mean annual precipitation of 838.7 mm.

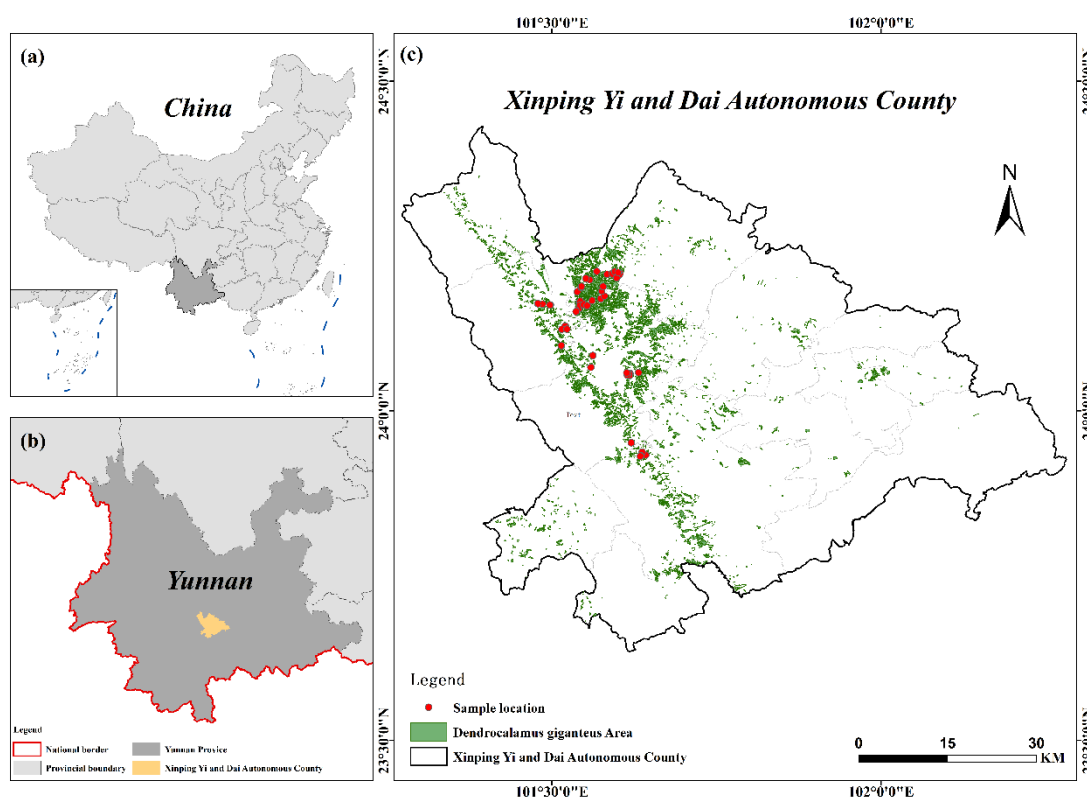


Figure 1. Location of the study area. (a) Yunnan Province within China; (b) Xinping County within Yunnan Province; (c) Distribution of *D. giganteus* forests and the locations of the 52 field plots (red dots) within Xinping County.

The county covers 4270.97 km², with a forest coverage of 64.6%. *Dendrocalamus giganteus* forests occupy 14,620 ha, representing 5.52% of the total forested area. These bamboo forests are predominantly found on south-facing steep slopes at elevations ranging from 432–1964 m. Although areas with slopes $\geq 25^{\circ}$ account for only 40% of the county's total land area, they contain 45% of its bamboo forests. The bamboo forests exhibit the dense, monolayered canopy structure previously described, with a culm density of 3200–4500 stems ha⁻¹, a mean diameter at breast height of 8–12 cm, and a canopy height of 12–18 m. The region's complex topography and unique bamboo forest

structure provide an ideal natural laboratory for evaluating multi-source remote sensing AGB retrieval models. The results of this study are therefore relevant for resource management and carbon sequestration assessments in the broader Southwest China bamboo industrial belt.

2.2. Data Collection

2.2.1. Field Data Collection and Biomass Estimation

Field data were collected in January 2024, corresponding to the dormant period for bamboo when the leaf area index is most stable. A stratified random sampling design was employed, guided by the 2023 "one map" forest resource management database from the Yunnan Provincial Forestry Survey and Planning Institute. Initially, 80 pure *D. giganteus* sub-compartments were selected across representative elevation (422–2,100 m) and slope (<15°, 15–25°, >25°) gradients. Following field reconnaissance to exclude sites with recent human disturbance (e.g., harvesting, pest damage) or accessibility constraints, a final set of 52 sample plots was established (Figure 1), satisfying minimum sample size requirements for statistical analysis.

Each sample plot was a circular area of 490.87 m² (12.5 m radius). The central coordinate of each plot was recorded using a Qianxun StarMatrix SR3 Pro receiver (WGS84 datum), with post-processing differential correction applied to ensure a horizontal positional accuracy of <0.5 m. Within each plot, the diameter at breast height (DBH, 1.3 m) was measured for all standing bamboo culms with a DBH ≥5 cm. Measurements were taken twice in perpendicular directions using a diameter tape, and the mean value was recorded. Quality control was performed by re-measuring 10% of the plots, which confirmed a high level of consistency for DBH measurements (Intraclass Correlation Coefficient, ICC = 0.98). The AGB for each individual culm was calculated using the species-specific allometric equation for *D. giganteus* developed by Fu et al. (2012). Plot-level AGB was then determined by summing the biomass of all individual culms within the plot and scaling the total to a per-hectare value (Mg ha⁻¹). A summary of the plot-level AGB statistics is provided in Table 1.

$$M = 0.8903 * D^{1.5505} (R = 0.9885, RMSE = 0.2354 \text{ kg})$$

(1)

where M is the above-ground biomass (kg) and D is the DBH (cm). Plot-scale AGB was calculated by summing the biomass of all individual culms and then converted to per-hectare values (Mg ha⁻¹).

Table 1. Statistical summary of AGB in the *D. giganteus* sample plots.

Variable	N	Minimum	Maximum	Mean	Standard Deviation (SD)
AGB (Mg ha ⁻¹)	52	9.96	132.09	66.89	27.59

2.2.2. ICESat-2 Data

This study utilized the ICESat-2 ATL08 Land and Vegetation Height product (Version 6) and the GEDI L2B Canopy Cover and Vertical Profile product (Version 2), both released in 2023 (Table 2). Data acquired between October 2023 and May 2024 were selected to ensure temporal correspondence with the field survey's phenological window.

The ICESat-2 satellite, launched in 2018, is equipped with the Advanced Topographic Laser Altimeter System (ATLAS) [28,29]. ATLAS operates using three pairs of laser beams (six beams total) arranged with ~3.3 km spacing between pairs and ~90 m spacing within each pair. Each pair consists of a strong and a weak beam, with the strong beam having approximately four times the energy of the weak beam [30]. The system produces footprints with a diameter of ~17 m and an along-track spacing of ~0.7 m [31]. The fundamental data product, ATL03, provides geolocated photon data, including height, coordinates, and signal classification for each laser pulse. The ATL08 product is a higher-level derivative of ATL03, where photons are algorithmically classified (e.g., noise, ground, canopy) and aggregated into 100 m along-track segments to produce terrain and canopy height statistics [32].

To ensure the reliability of derived land surface parameters, the ATL08 data underwent a systematic quality filtering and preprocessing workflow. Since the ICESat-2 sensor is a Photon Counting LiDAR (PCL) system sensitive to atmospheric and solar background noise, a rigorous denoising procedure is essential [9]. First, the Differential, Regressive, and Gaussian Adaptive Nearest Neighbor (DRAGANN) algorithm was applied to the photon point cloud histograms to identify and filter out noise photons [9,30,33]. Following this primary denoising, segments were further assessed for quality. Segments containing fewer than 50 resulting signal photons were discarded, as this low signal density is insufficient to accurately represent land cover characteristics. Additionally, segments with a mean canopy height exceeding 50 m or below 2 m were excluded as outliers. After this comprehensive quality control process, 19 parameters describing terrain, canopy height metrics, and laser beam characteristics were extracted from the filtered ATL08 segments for subsequent modeling. A complete list of these parameters is provided in Supplementary Table S1.

Table 2. Key parameters of the ICESat-2 and GEDI missions, which represent the state of the art in spaceborne laser altimeters.

Misson	ICESat-2	GEDI
Full name	Ice, Cloud, and land Elevation Satellite-2	Global Ecosystem Dynamics Investigation
Launch date	September 15, 2018	December 5, 2018
Detector type	Photon counting	Full waveform
Wavelength	532 nm (green)	1064 nm (near IR)
across-track spacing	90 m within pairs 3.3 km between pairs	600 m
Diameter along-track spacing	~0.7 m	~60 m
Footprint	~12 m	~25 m
Track number	6 tracks from 1 laser	8 tracks from 3 lasers
Orbit inclination and coverage	92°; coverage up to 88°N–88°S latitude	51.6°; coverage up to 51.6°N–51.6°S latitude
Laser power	120 μJ/30 μJ	15 mJ/4.5 mJ
Temporal resolution (Revisit time)	~91 days (exact repeat orbit)	~45 days (non-repeating)
Vertical accuracy	~3–5 cm for flat surfaces	~1 m (depending on waveform processing and vegetation density)

2.2.3. GEDI Data

The GEDI mission, operating from the International Space Station (ISS) since its launch in 2018, was the first spaceborne LiDAR system developed specifically for forest structure observation [8,34]. Its primary objective is the high-precision mapping of AGB and 3D structure in global temperate and tropical forests (latitudes ±51.6°). Unlike the photon-counting system of ICESat-2, GEDI employs a full-waveform LiDAR instrument that digitizes the complete vertical distribution of intercepted surfaces, enabling a more detailed characterization of the canopy profile [35]. The instrument uses three 1064 nm lasers, which are dithered to produce eight parallel ground tracks. This strategy results in a total swath width of ~4.2 km, with ~600 m spacing between tracks. Each laser produces a ~25 m footprint with an along-track sampling interval of ~60 m [36].

GEDI data are provided in a hierarchy of products from Level 1 (georeferenced waveforms) to Level 4 (gridded AGB). This study utilized the L2B product, which provides key metrics describing canopy cover, the plant area index, and the vertical foliage profile, all of which are critical for detailed structural analysis [34]. The L2B data covering the study area were acquired from NASA's Earthdata portal (<https://search.earthdata.nasa.gov/>), comprising 46 orbits. To ensure the quality of the GEDI data, a systematic filtering protocol was applied to the L2B footprints to remove low-quality returns resulting from cloud cover, atmospheric scattering, sensor noise, or complex terrain, based on established methods [37–39]. The specific quality flags and thresholds used are detailed in Table 3. From an initial dataset of 70,619 footprints acquired over the study area, this filtering process yielded 55,649 high-quality footprints suitable for analysis. For these footprints, 21 candidate variables were extracted for modeling, including terrain factors, canopy height metrics, and relative height (RH) percentiles (see Supplementary Table S2).

Table 3. GEDI L2B Data Quality Filtering Criteria.

Paramete	Retention Value	Retention Basis
lon_lowestmode	101°–103° E	Defines the longitudinal extent of the Xinping County study area.
lat_lowestmode	23°–25° N	Defines the latitudinal extent of the Xinping County study area.
algorithmrun_flag	1	Confirms successful execution of the L2B algorithm and adequate waveform fidelity.
quality_flag	1	Indicates good quality footprint data that meets multiple quality criteria and is located over a vegetated land area.
Sensitivity	≥ 0.90	Selects valid returns with high sensitivity (values approaching 1 signify high-quality signals).
degrade_flag	0	Excludes data flagged due to degraded performance of the instrument or its pointing/positioning systems.

2.2.4. Optical and Topographic Predictor Variables

To enable the spatial extrapolation of the sparse LiDAR measurements, a suite of spatially contiguous predictor variables was generated from Sentinel-2 imagery and a digital elevation model (DEM). The Sentinel-2 mission is particularly well-suited for this purpose as its multispectral instrument includes red-edge bands (e.g., B5, B6, B7), which are highly sensitive to vegetation chlorophyll content and canopy structure (Phiri et al., 2020). Level-2A surface reflectance products from May 2023 to January 2024 with cloud cover ≤5% were acquired via the Google Earth Engine platform. To mitigate the effects of seasonal phenological variation and cloud contamination, a cloud-free median composite was generated using all available images from the peak growing season (June–September 2023). Cloud masking was performed using the QA60 quality assessment band.

Topographic variables (elevation, slope, and aspect) were derived from the ALOS DEM (12.5 m resolution), which was resampled to a 25 m resolution to align with the scale of the GEDI footprints. From the processed Sentinel-2 imagery, nine vegetation indices were calculated, including the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and Soil-Adjusted Vegetation Index (SAVI) (see Table 4 for a complete list). All predictor variables were co-registered and resampled to a common 25 m grid.

Table 4. Vegetation indices and terrain variables derived from Sentinel-2 spectral data and the ALOS DEM.

Vegetation Indices / Topographic Features	Formula/Description	Citation
Difference Vegetation Index	$DVI = NIR - RED$	[40]
Enhanced Vegetation Index	$EVI = \frac{2.5 * (NIR - RED)}{1 + NIR + 6 * RED - 7.5 * BLUE}$	[41]
Green Difference Vegetation Index	$GDVI = NIR - GREEN$	[42]
Green Normalized Difference Vegetation Index	$GNDVI = \frac{NIR - GREEN}{NIR + GREEN}$	[43]
Green Ratio Vegetation Index	$GRVI = \frac{NIR}{GREEN}$	[44]
Normalized Difference Vegetation Index	$NDVI = \frac{NIR - RED}{NIR + RED}$	[45]
Normalized Pigment Chlorophyll Index	$NPCI = \frac{RED - GREEN}{RED + GREEN}$	[46]
Ratio Vegetation Index	$RVI = \frac{NIR}{RED}$	[47]
Soil Adjusted Vegetation Index	$SAVI = \frac{(1 + L) * (NIR - RED)}{NIR + RED + L}$	[48]
Elevation	Elevation	
Slope	Slope factor extracted by DEM	
Aspect	Slope aspect factor extracted by DEM	

Where RED, GREEN, BLUE, NIR are the reflectance of red band, green band, blue band and near-infrared band

2.3. Research Methods

The analytical framework for this study involved a two-stage process to upscale the field-measured plot AGB to a continuous map for the entire study area (Figure 2). The first stage involved the spatial extrapolation of key structural metrics derived from the sparse ICESat-2 and GEDI footprints. This was achieved by modeling the relationship between the LiDAR metrics and the spatially contiguous auxiliary variables to generate continuous 25 m resolution maps of these structural parameters. In the second stage, a two-level stacked ensemble model was developed to estimate AGB. This model used the field-measured plot AGB (Mg ha⁻¹) as the response variable and the full suite of continuous predictor variables (i.e., the extrapolated LiDAR metrics and the original optical and topographic layers) as inputs. The trained model was then applied to this suite of predictor variables to generate the final, wall-to-wall AGB map for the study area.

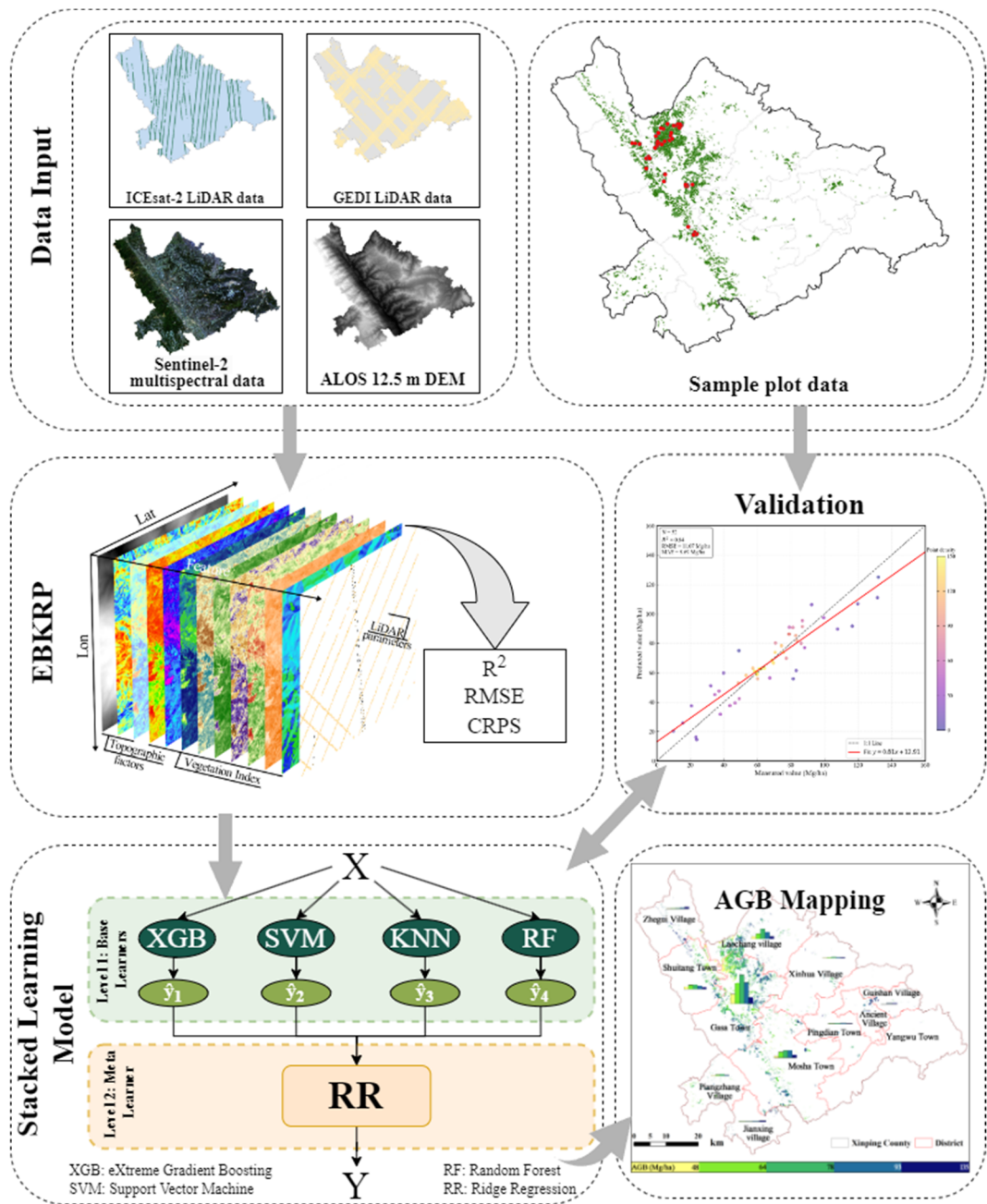


Figure 2. Method overview for generating the AGB map of *D. giganteus*.

2.3.1. Spatial Extrapolation of LiDAR Metrics

Empirical Bayesian Kriging Regression Prediction (EBKRP) was the geostatistical spatial prediction method used to generate the spatially continuous layers of LiDAR-derived structural metrics [49]. EBKRP integrates least-squares regression with kriging, and unlike conventional kriging techniques, it accounts for the uncertainty in the estimated semivariogram by simulating many possible semivariograms from the input data, resulting in more accurate predictions.

The EBKRP process was implemented in ArcGIS Pro 2.8. For each selected LiDAR metric (from GEDI and ICESat-2), a separate model was built using the LiDAR metric values at the footprint locations as the dependent variable and the co-located optical and topographic variables as independent predictors. Key parameters for the EBKRP models included an empirical data transformation, a K-Bessel variogram model, a subset size of 100, and 500 simulation iterations. To account for spatial autocorrelation anisotropy induced by the complex terrain, semivariogram ellipses were visualized to optimize the search neighborhood parameters. The performance of each model was evaluated using Leave-One-Out Cross-Validation (LOOCV). This procedure yielded

continuous 25 m resolution raster layers for each key LiDAR structural metric, which served as essential predictor variables in the final AGB model. To assess the accuracy of the EBKRP spatial prediction, several statistical metrics were calculated. The defining equations for each evaluation indicator are as follows:

$$R^2 = \frac{\sum_{i=1}^n |\hat{Z}(x_i) - \bar{Z}(x_i)|^2}{\sum_{i=1}^n |Z(x_i) - \bar{Z}(x_i)|^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n |Z(x_i) - \hat{Z}(x_i)|^2}{n}} \quad (2)$$

$$CRPS(A, F) = \int [F(x) - 1\{x \geq y\}]^2 dx \quad (3)$$

Where $\hat{Z}(x_i)$ is the predicted value at location x_i , $Z(x_i)$ is the observed value at location x_i , $\bar{Z}(x_i)$ is the mean of the observed values, and n is the total number of samples. For the Continuous Ranked Probability Score (CRPS), F represents the predicted cumulative distribution function (CDF) from the model, while y represents the true observed value. The term $1\{x \geq y\}$ is an indicator function, which equals 1 if the condition is met and 0 otherwise.

2.3.2. AGB Estimation Using Stacked Ensemble Modeling

Given the complex, nonlinear relationships between topographic and vegetation features in the study area, non-parametric machine learning models are particularly suitable for AGB estimation due to their robust fitting capabilities [50–52]. This study therefore selected four widely-used machine learning algorithms to serve as base-learners within a stacked ensemble framework to estimate bamboo forest AGB.

2.3.2.1. Base Learners Algorithms

The k-Nearest Neighbors (kNN) algorithm operates by identifying the k samples most similar to a target plot based on a distance metric and then imputes the value for the target plot using a prediction rule applied to these neighbors [53]. The algorithm's flexibility, stemming from the fact that it does not assume an underlying normal distribution, has led to its wide application in forest parameter estimation. The choice of distance metric, the value of k , and the prediction rule significantly influence the results [54]. In this study, we employed Euclidean distance and an inverse-distance weighting method. The value of k was optimized by searching a range from 1 to 50, with the final value selected via cross-validation. The prediction formula is as follows:

$$\hat{y}(x_q) = \frac{\sum_{i=1}^k w_i \cdot y_i}{\sum_{i=1}^k w_i} \quad (4)$$

where \hat{y} is the predicted value for the target plot, y_i is the observed value of the i -th nearest neighbor, and w_i is the weight of the i -th neighbor, calculated as the inverse of the distance $w_i = (1/d_i)$.

Support Vector Machine (SVM) is known to perform well even when dealing with a limited number of samples. The core principle of SVM is to map input data into a higher-dimensional feature space using a non-linear kernel function. This allows the model to address non-linear relationships, effectively reducing both model error and complexity [55,56]. We implemented the SVM model using Python's "scikit-learn" library and compared the performance of three kernel types: Radial Basis Function (RBF), polynomial, and linear. The penalty coefficient (C) and the kernel parameter (γ) were optimized via grid search (GridSearchCV) to identify the best parameter combination. The regression function is as follows:

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (5)$$

where α_i and α_i^* are the Lagrange multipliers, b is the bias term, and $K(x_i, x)$ is the kernel function.

Random Forest (RF), a classic ensemble learning method, enhances model generalization and stability by constructing a multitude of decision trees and averaging their predictions [57]. The algorithm incorporates two levels of randomness during training—bootstrapping of data samples and random selection of features at each split—which effectively mitigates the risk of overfitting. RF is also robust to outliers and provides feature importance scores, offering a valuable basis for variable selection. We developed the optimal RF model by tuning key hyperparameters, including the number of trees ($n_estimators$), maximum tree depth (max_depth), and the minimum number of samples required at a leaf node ($min_samples_leaf$). Its regression prediction formula is as follows:

$$\hat{f}(x) = \frac{1}{K} \sum_{k=1}^K k(x) \quad (6)$$

where K is the total number of trees, and $hk(x)$ represents the prediction for sample x from the k -th tree.

XGBoost (eXtreme Gradient Boosting) is an advanced ensemble learning algorithm built upon the Gradient Boosting Decision Tree (GBDT) framework. The algorithm iteratively builds a sequence of decision trees, with each new tree trained to correct the residual errors of the one before it [58]. This progressive optimization strategy significantly enhances prediction accuracy. Key advantages of XGBoost include its built-in regularization mechanism, which effectively controls overfitting, and its outstanding computational efficiency, leading to exceptional performance in a wide range of regression tasks. We implemented the model using the `xgboost` library in Python and systematically optimized its key hyperparameters—including the number of estimators ($n_estimators$), maximum tree depth (max_depth), and learning rate—via grid search (`GridSearchCV`). Its prediction model can be represented by the following formula:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (7)$$

where \hat{y}_i is the final prediction for sample i , K is the total number of trees, and f_k represents the k -th decision tree model.

2.3.2.2. Stacking Ensemble Method

The Stacking algorithm significantly enhances the accuracy of AGB estimation in complex habitats by integrating the predictive strengths of multiple base models [59]. We developed a two-level stacking architecture (Figure 4). Level 0 comprises four base-learners: XGBoost, SVM, kNN, and RF. Level 1 employs Ridge Regression as a meta-learner, which performs a weighted fusion of the base-learners' outputs to produce the final prediction.

To prevent data leakage during the meta-feature generation stage and to robustly leverage the available sample size ($N=52$), a Leave-One-Out Cross-Validation (LOOCV) strategy was applied. In this procedure, an iterative process is performed where, for each sample, the base models are trained on all other samples and then used to generate a prediction for that single held-out sample. This yields a complete set of Out-of-Fold (OOF) predictions, which were then assembled to form an $n \times 4$ meta-feature matrix (where n is the total sample size and 4 is the number of base-learners). For the meta-model training stage, the Level-1 Ridge Regression model was trained using this complete meta-feature matrix as its input and the field-measured AGB as the target variable. This model constrains regression coefficients via L2 regularization to control complexity and enhance generalization.

2.3.3. Model Performance Assessment

Model performance was evaluated using the LOOCV method. To compare predictive capabilities, each base-learner was assessed alongside the final stacking model using three quantitative metrics: the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). The formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

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

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

In these formulas, y_i is the observed value for the i -th sample, \hat{y}_i is the predicted value for the i -th sample, \bar{y} is the mean of the observed values, and n is the total number of samples.

2.3.4. SHAP-Based Model Interpretability

To move beyond assessing model performance and gain insight into the ecological mechanisms driving AGB distribution, this study employed SHAP, a method for interpreting complex machine learning model outputs [60]. Grounded in cooperative game theory, SHAP assigns a Shapley value to each feature for every prediction. This value quantifies the feature's marginal contribution to shifting the model's output from a baseline (the mean prediction across the dataset) to the final predicted value. The SHAP framework operates on the principle of additive feature attribution, wherein the explanation is a linear function of binary variables:

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \quad (11)$$

where g is the explanation model, $z' \in \{0, 1\}^M$ is the simplified input representing feature presence or absence, M is the number of input features, and $\phi_i \in \mathbb{R}$ is the Shapley value for feature i .

For this analysis, SHAP was applied to the Random Forest model, selected for its high performance among the tree-based learners and its compatibility with the efficient TreeExplainer algorithm. Two primary SHAP visualizations were generated: 1) the summary plot, which aggregates absolute Shapley values to rank global feature importance, illustrating the direction and distribution of impacts, and 2) dependence plots, which reveal the effect of a single feature on model predictions across all data points, uncovering non-linear relationships and interaction effects. This application of SHAP transformed the model from a predictive "black box" into an interpretable tool, providing a quantitative foundation for the ecological discussion and management recommendations.

3. Results

3.1. Accuracy of Spatially Extrapolated LiDAR Metrics

The spatial extrapolation of 39 structural metrics derived from GEDI and ICESat-2 using the EBKRP method revealed a clear hierarchy in prediction accuracy. Cross-validation results indicated that R^2 values ranged from 0.12 to 1.00, while RMSE and CRPS varied substantially depending on the parameter's scale, units, and intrinsic spatial variability (see Supplementary Materials).

The accuracy of the spatial prediction was stratified by parameter type. The first tier, comprising parameters related to topography and macro-canopy structure, exhibited the highest spatial continuity and thus the highest prediction accuracy. Specifically, elevation (dem_h) and seven related canopy height metrics demonstrated exceptionally strong spatial autocorrelation ($R^2 \geq 0.96$), as did the signal-to-noise ratio (snr). A second tier of parameters, representing internal canopy structure and laser return energy, yielded moderate accuracies. This group included energy-related variables such as the asr and rx_energy series ($R^2 = 0.64$ – 0.66) and other structural metrics like n_seg_ph and h_mean_canopy ($R^2 = 0.52$ – 0.60). Conversely, the third tier of parameters, which describe fine-scale canopy details, discrete photon counts, and surface reflectance properties, showed the lowest

prediction accuracies. Variables such as h_min_canopy, n_toc_photons, and the reflectance-related rv and rg series had R² values concentrated between 0.26 and 0.47. This result reflects the higher local variance and spatial randomness characteristic of these fine-scale metrics.

These findings indicate that a parameter's susceptibility to accurate spatial prediction is strongly linked to its inherent spatial structure. Parameters with high spatial continuity (e.g., topography) were predicted with the greatest accuracy, whereas those with high local variability (e.g., photon counts) were predicted with the lowest accuracy. The strong performance of the EBKRP method for variables with distinct spatial patterns underscores its ability to effectively leverage spatially explicit predictor variables, such as terrain, during the spatial prediction process.

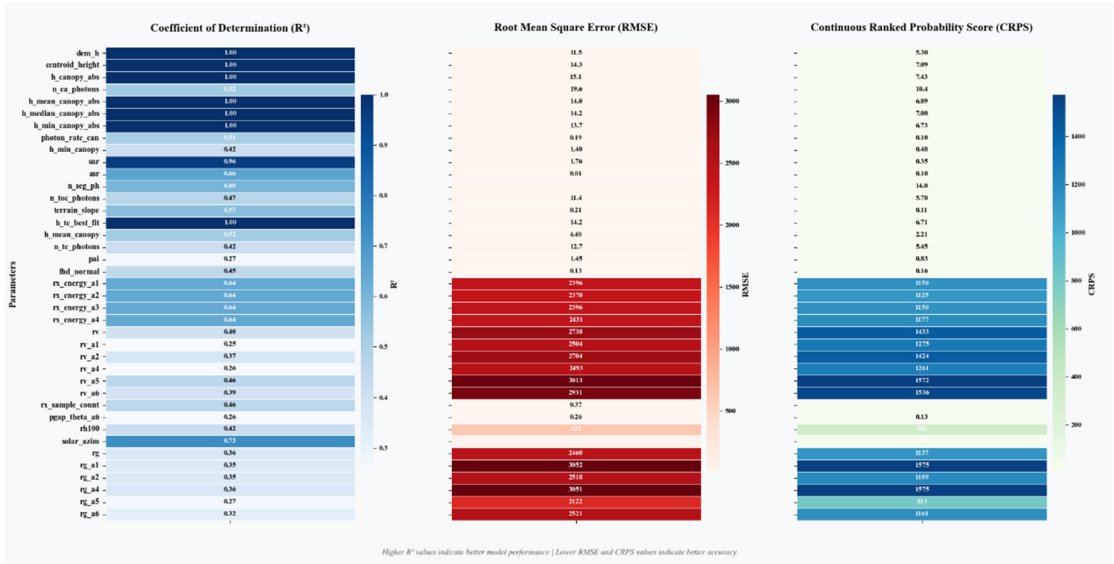


Figure 3. Performance metrics (R², RMSE, CRPS) for spatial interpolation of GEDI and ICESat-2 derived parameters using the EBKRP method.

3.2. Feature Importance and Selection for AGB Modeling

The relative importance of the 52 predictor variables in the AGB model, as determined by SHAP values, is shown in Figure 4a. The canopy photon rate (photon_rate_can) emerged as the most influential variable (SHAP value = 5.36), followed by elevation (dem_h, SHAP value = 1.86), a ground-related parameter (rg_sg, SHAP value = 1.64), and a reflectance-related parameter (rv_a4, SHAP value = 1.44). The remaining variables, including those describing vertical canopy structure, laser return energy, and waveform geometry, contributed with progressively lower SHAP values.

To identify the optimal number of variables for the final model, a recursive feature elimination (RFE) process was conducted (Figure 4b). Models were iteratively built by adding variables one by one based on their SHAP importance ranking, and the model performance was assessed at each step. This analysis revealed that model accuracy plateaued after 14 features were included.

Based on these results, the top 14 most important features were selected as the optimal set of predictor variables. This approach balances model performance and parsimony by excluding variables that contribute little explanatory power and could introduce noise or increase computational complexity.

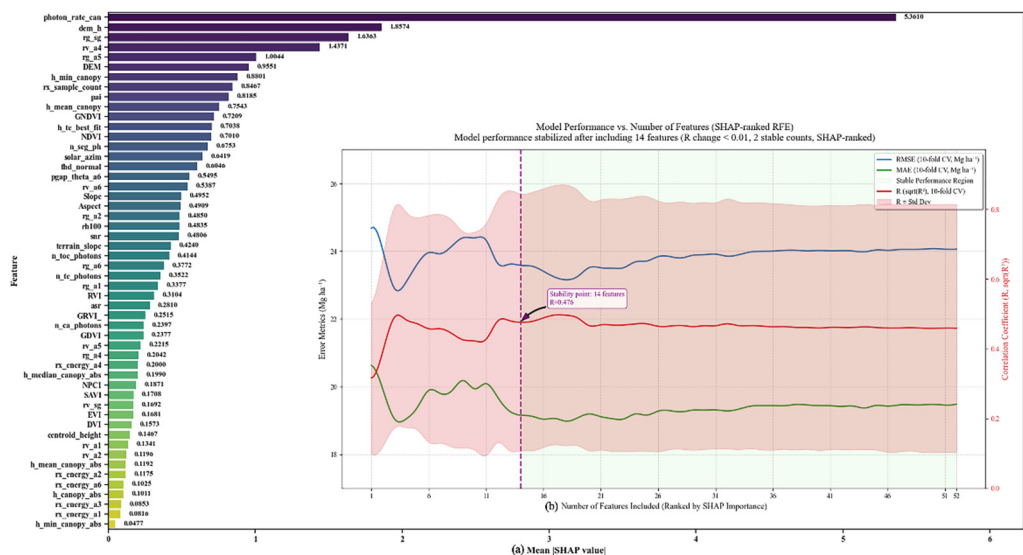


Figure 4. Feature importance results. (a) The importance ranking of explanatory variables based on SHAP values for predicting the AGB of *D. giganteus*; (b) Relationship between the number of input features and model performance metrics, including the correlation coefficient (R) (red line), RMSE (blue line), and MAE (green line).

3.3. AGB Model Performance

The validation results for the AGB models demonstrate that the stacking ensemble model significantly outperformed all individual base-learners across all evaluation metrics (Table 5, Figure 5). The stacking model achieved an R^2 of 0.84, an RMSE of 11.07 Mg ha⁻¹, and an MAE of 8.69 Mg ha⁻¹. Furthermore, the regression fit line for the stacking model’s predictions ($y = 0.81x + 12.91$) showed strong agreement with the 1:1 reference line, indicating high accuracy and low systematic bias across the full range of AGB values.

Among the four base-learners, RF yielded the best performance ($R^2=0.72$, RMSE=14.53 Mg ha⁻¹), followed by SVM ($R^2=0.69$, RMSE=15.32 Mg ha⁻¹) and XGBoost ($R^2=0.68$, RMSE=15.58 Mg ha⁻¹). The kNN model performed the poorest ($R^2=0.60$, RMSE=17.35 Mg ha⁻¹), reflecting its inherent limitations in capturing the complex, non-linear relationships between the multi-source remote sensing features and AGB in this heterogeneous landscape.

Table 5. Cross-validation results of model performance for *D. giganteus* AGB estimation

Model	R ²	RMSE (Mg/ha)	MAE (Mg/ha)	Regression Fit
Stacking	0.84	11.07	8.69	$y = 0.81x + 12.91$
Random Forest (RF)	0.72	14.53	10.53	$y = 0.53x + 31.32$
Support Vector Machine (SVM)	0.69	15.32	9.02	$y = 0.59x + 27.33$
XGBoost	0.68	15.58	11.50	$y = 0.48x + 34.48$
k-Nearest Neighbors (kNN)	0.60	17.35	14.27	$y = 0.55x + 26.62$

A notable pattern observed in the scatterplots of predicted versus observed values was that all base-learners exhibited a similar systematic bias (Figure 5a-d). The slopes of their regression lines were considerably less than 1, indicating a tendency to underestimate high AGB values and overestimate low AGB values. This bias is likely attributable to factors including the limited number of field plots in high-biomass stands and the known propensity for LiDAR signal saturation in dense bamboo canopies. The stacking model, through its two-level learning architecture, effectively integrated the predictive strengths of the diverse base-learners and, in doing so, substantially corrected for this systematic bias, which enhanced both its overall accuracy and robustness.

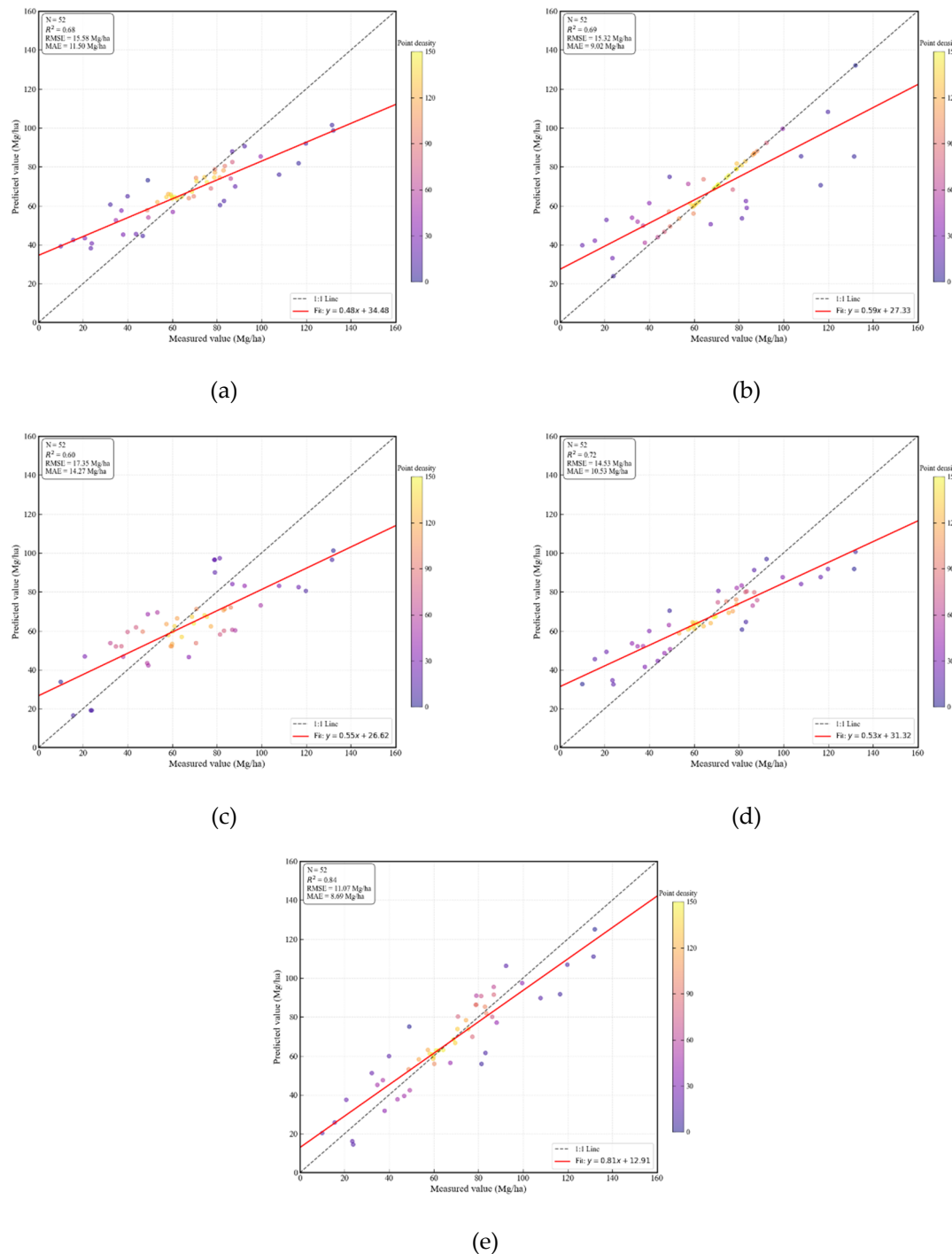


Figure 5. Scatterplots of predicted versus observed AGB for *D. giganteus* in Xinping County from different models: (a) XGBoost; (b) SVM; (c) kNN; (d) RF; and (e) Stacking. The dashed line represents the 1:1 reference line, while the solid line is the regression fit line. Model performance is considered better as the slope of the regression equation approaches 1 and the intercept approaches 0.

3.4. Mapping and Analysis of *D. giganteus* AGB

The optimized stacking model was applied to the full suite of predictor variables to generate a continuous AGB map for the 14,620 ha of *D. giganteus* forest within Xinping County (Figure 6). The mean AGB density for the bamboo forests across the study area was predicted to be 71.8 ± 21.9 Mg ha⁻¹ (mean \pm standard deviation), with densities ranging from 10.0 to 135.0 Mg ha⁻¹.

The resulting map reveals significant spatial heterogeneity and distinct patterns of geographic clustering. AGB density was systematically higher in the central-western and northern portions of

the county (e.g., Jiasa Town, Laochang Township, Shuitang Town) compared to the eastern and southern regions (e.g., Guishan Street, Pingdian Township, Jianxing Township). Analysis of the AGB frequency distribution shows that the majority of the bamboo forest area (57.4%, or 84.46 km²) falls within a medium-to-high AGB range (64.0–93.0 Mg ha⁻¹). Conversely, areas of very high AGB (>93.0 Mg ha⁻¹) and low AGB (<48.0 Mg ha⁻¹) were less common, accounting for 10.7% and 9.9% of the bamboo forest area, respectively. Spatially, the highest AGB values were concentrated in distinct patches, potentially corresponding to areas with optimal site conditions or intensive management, while the lowest AGB values were typically found in more marginal areas.

Overall, the spatial distribution of *D. giganteus* AGB in Xinping County is not random but appears to be regulated by a combination of environmental gradients and land management practices, the specific drivers of which will be explored in the discussion.

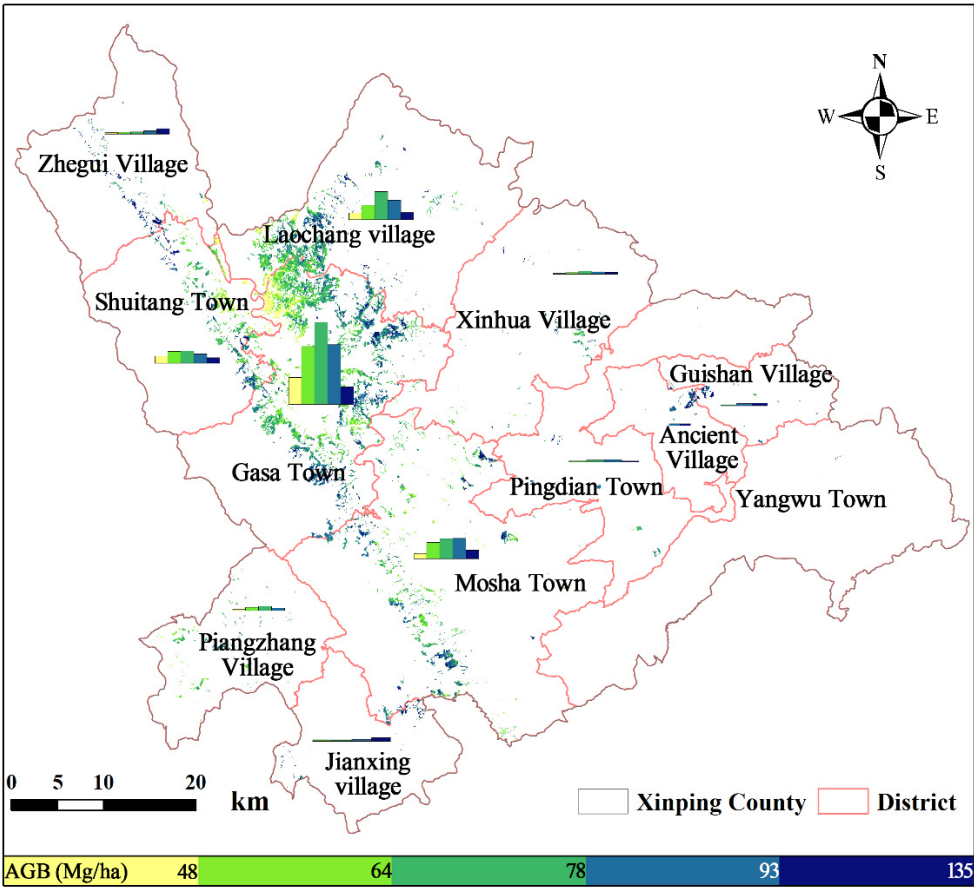


Figure 6. Spatial distribution map of *D. giganteus* AGB in the study area (bar charts show the distribution of *D. giganteus* aboveground biomass across five different density classes within each region, with classes determined using the natural breaks classification method).

4. Discussion

4.1. Analysis of Spatial Heterogeneity in EBKRP Prediction Results

The stratified prediction accuracies exhibited by the various LiDAR-derived parameters following EBKRP spatial prediction reveal the fundamental relationship between a parameter's intrinsic spatial structure and the model's underlying geostatistical assumptions. For parameters related to topography and macro-canopy height (e.g., dem_h), the model achieved an almost perfect fit ($R^2 = 1.00$). This aligns with the findings of (Borselli et al. 2008), confirming that terrain-related variables possess high spatial continuity. This high R^2 value, however, is likely less an indicator of

the model's independent predictive power and more a reflection of its ability to replicate a spatial pattern that is strongly driven by the covariates. When significant informational overlap exists between the target and predictor variables, the model effectively learns a deterministic spatial pattern. Consequently, while this result validates the efficacy of the method, the risk of overfitting due to covariate collinearity must be considered.

In contrast, parameters describing laser pulse energy (e.g., the *rx_energy* series) and certain canopy structure metrics (e.g., *n_seg_ph*) only achieved moderate prediction accuracies ($R^2 = 0.51\text{--}0.73$). This outcome reflects the limitations of EBKRP when encountering complex, non-linear spatial structures. As noted by Tian et al. (2021), LiDAR signal interaction with the canopy is governed by a non-linear combination of biophysical factors, resulting in complex spatial patterns that cannot be fully captured by the smooth semivariogram functions inherent to traditional kriging frameworks. Although EBKRP improves upon traditional methods by incorporating a Bayesian strategy and local subsets, it is still fundamentally based on Gaussian process assumptions and is therefore limited in its ability to model highly heterogeneous or complex distributions.

The poor predictive accuracy for variables with the lowest prediction accuracy, such as *h_min_canopy* and *n_toc_photons*, stems from a mismatch in spatial scale and a conflict between the data type (i.e., discrete counts) and the model's fundamental assumption of a continuous Gaussian field. The spatial variation of fine-scale variables like *h_min_canopy* is controlled by small-scale factors such as understory vegetation and microtopography [63]. This fine-scale variation occurs at a spatial grain much smaller than the resolution of the covariates, preventing the model from capturing it effectively. Furthermore, variables like *n_toc_photons* represent discrete count data, whereas EBKRP is built on the theoretical framework of a continuous Gaussian random field. As Zhao et al., (2006) emphasized, applying a continuous-field model directly to discrete count data violates the model's fundamental assumptions and can lead to systematic bias.

In summary, the performance of EBKRP for spatially predicting LiDAR metrics is highly dependent on the metric's intrinsic spatial structure, its scale relative to the predictor variables, and its data type. While the model excels at processing macro-scale variables with strong spatial continuity, its suitability for applications involving high spatial frequency, strong non-linearity, or data type mismatches requires careful assessment.

4.2. Influence of Algorithm Selection on Bamboo AGB Estimation

The systematic evaluation of multiple machine learning algorithms underscores the critical influence of model choice on AGB estimation accuracy. The superior performance of the stacking model ($R^2=0.84$, $RMSE=11.07\text{ Mg ha}^{-1}$) surpasses that of the individual base-learners and is competitive with results from recent literature that used single models for similar forest types [65,66]. The strength of this approach lies in its two-level architecture: the diverse base-learners capture different aspects of the predictor-response relationship, while the meta-learner optimally weighs their individual predictions. This process effectively balances the bias-variance trade-off, which markedly mitigated the underestimation of high AGB values ($>78\text{ Mg ha}^{-1}$) observed in the base models and improved the average regression slope from 0.54 to 0.81.

Among the individual algorithms, while RF performed best in terms of R^2 , its low regression slope (0.53) also highlighted its limitations. This underestimation is likely related to LiDAR signal saturation, a known issue in dense canopies such as the *D. giganteus* forests in our study area [67–69]. Although XGBoost and SVM yielded similar R^2 values, the lower MAE of the SVM model suggests that its principle of structural risk minimization provided greater robustness against outliers in the dataset [36]. In contrast, the poor performance of the kNN model, given the limited sample size ($n=52$), is likely attributable to the 'curse of dimensionality', where the local structure of the high-dimensional feature space becomes sparse [70].

Although the proposed framework of fusing multisource remote sensing data within a stacked ensemble model significantly improved estimation accuracy, some systematic bias was not entirely eliminated. Future work could focus on two key areas. First, incorporating a wider array of ancillary

eco-environmental factors (e.g., fine-resolution soil and climate data) may improve model performance in complex terrain and high-biomass areas. Second, while SHAP analysis provided valuable insights into feature importance and model behavior, further exploration of advanced interpretability techniques could enhance our mechanistic understanding of the relationships between remote sensing variables and biomass distribution, enabling more targeted model optimization.

4.3. Ecological Interpretation and Management Implications

The spatial distribution of AGB in *D. giganteus* stands within Xinping County, Yunnan, is governed by a complex interplay of ecological and anthropogenic factors (Figure 6). By integrating model feature importance with SHAP value analysis, this study identified primary AGB drivers and uncovered their non-linear response mechanisms and critical thresholds, providing a data-driven foundation for interpreting its distribution mechanisms.

4.3.1. Ecological Drivers and Non-Linear Mechanisms

Ecologically, the high importance of the LiDAR-derived canopy photon rate (photon_rate_can) reflects the strong relationship between LiDAR-derived metrics and the unique canopy architecture of this bamboo species. As a large sympodial bamboo, *D. giganteus* has a high LAI (6.8 ± 1.2), where its dense canopy markedly intensifies photon scattering. The SHAP summary plot (Figure 7) quantifies this relationship, showing a consistent positive correlation between photon_rate_can and predicted AGB. This confirms that LiDAR-based canopy return metrics are reliable proxies for AGB.

The SHAP dependence plot for photon_rate_can reveals a clear threshold around 0.92. Below this value, the variable predominantly contributes negatively to AGB predictions; above this threshold, its contribution to the model output shifts to strongly positive, indicating that high canopy reflectance is a salient characteristic of high-biomass stands. The significant positive effect of the vegetation return energy (rv_a4) corroborates this finding. These variables collectively underscore the primary role of canopy structure in regulating AGB.

Topographic factors also significantly influence the ecological regulation of AGB. Both elevation (dem_h) and the height above the best-fit terrain (h_te_best_fit) exhibit high SHAP contributions. GIS overlay analysis shows that approximately 73% of high-AGB areas ($>78 \text{ Mg ha}^{-1}$) are concentrated in zones with 800–1300 m elevation, $<20^\circ$ slope, and $>1200 \text{ mm}$ annual precipitation. The SHAP dependence plot provides finer-grained detail: AGB accumulation is most pronounced for dem_h values 840 m, and h_te_best_fit positively influences AGB accumulation below 1164.11 m. This indicates that topography indirectly shapes biomass distribution by influencing local microclimates and moisture availability[71,72].

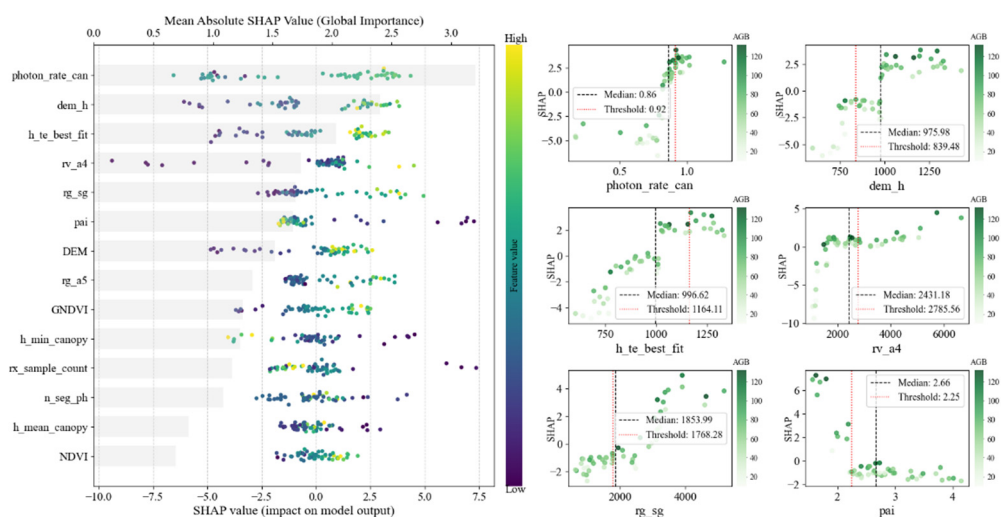


Figure 7. SHAP analysis for the Random Forest model, revealing the key drivers of AGB. The left panel is a SHAP summary plot, which ranks all features by their global importance based on the mean absolute SHAP value (gray bars). The color of each point on the plot represents the feature's value (from low to high), while its position on the x-axis indicates the impact on the prediction for an individual sample. The right panels are SHAP dependence plots for the top six most important features, detailing the non-linear relationship between each driver and its contribution to the model prediction, with key ecological thresholds indicated by red dotted lines.

4.3.2. Anthropogenic Influences and Implications for Precision Management

Anthropogenic activities markedly affect AGB patterns. The spatial association between exceptionally high-AGB areas ($>93 \text{ Mg ha}^{-1}$) and villages suggests that intensive management, including fertilization and irrigation, enhances biomass accumulation by cultivating healthier canopies and elevating *photon_rate_can* values. Our model successfully captured this coupled human-natural system; future research could integrate socioeconomic variables, such as management intensity and land-use history, for a more profound mechanistic understanding.

Based on these driving mechanisms and SHAP-derived thresholds, this study proposes differentiated management strategies. First, high-potential stands meeting critical thresholds (e.g., *photon_rate_can* > 0.92 and *h_te_best_fit* $< 1164.11 \text{ m}$) should be delineated as priority carbon sink conservation zones, where long-term carbon stock stability is ensured through strict harvesting intensity controls (e.g., annual felling rate $< 10\%$). Second, for stands with medium-to-low productivity, the primary management objective is to enhance canopy vigor to surpass the *photon_rate_can* threshold. Specific measures include density regulation ($> 3200 \text{ culms/ha}$) and soil amendment (adjusting pH to 5.5–6.5) to synergistically boost stand productivity. Last, harvesting strategies should be adaptively tailored to stand conditions. Strip thinning is advised for high-AGB zones to preserve stand structure, whereas medium-to-low AGB zones could implement selective cutting with a shorter rotation cycle (6–8 years), leveraging topographic thresholds such as *h_te_best_fit* $< 1164.11 \text{ m}$ to optimize resource turnover.

In conclusion, by utilizing the interpretable machine learning framework of SHAP, this study translates complex model outputs into ecologically significant response thresholds, providing a practical framework for advancing the data-driven adaptive management of bamboo forests. LiDAR-derived metrics, exemplified by *photon_rate_can*, can serve as critical inputs for stand health monitoring and precision interventions, thereby enhancing ecosystem service functions while helping to achieve sustainable forestry objectives.

5. Conclusion

This study successfully developed and validated a multi-stage modeling framework combining EBKRP and Stacking. This framework effectively addresses the challenge of high-accuracy spatial mapping of AGB for the large clumping bamboo species *D. giganteus* in complex subtropical mountainous environments.

The framework's core contribution is its use of the EBKRP method to fuse and spatialize discrete structural parameters from GEDI and ICESat-2 with continuous spectral and topographic factors from Sentinel-2, thereby providing high-quality predictor variables for subsequent modeling. Building on this foundation, the developed Stacking model demonstrated excellent predictive performance ($R^2=0.84$, $\text{RMSE}=11.07 \text{ Mg ha}^{-1}$). Its accuracy was significantly superior to all individual base-learners. More critically, it effectively corrected the systematic underestimation in high-biomass regions—a common issue in traditional models—by using a meta-learner, which improved the regression slope of the predicted-versus-observed fit line from a base-model average of 0.54 to 0.81, while SHAP analysis enhanced model interpretability by quantifying feature contributions and identifying critical ecological thresholds for precision silviculture.

This research demonstrates that the integrated framework provides a robust, high-accuracy technical pathway for the refined assessment of regional-scale bamboo carbon stocks. Furthermore,

it offers a methodologically generalizable reference for estimating biomass via remote sensing in other structurally complex forest types.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Table S1: Description of ICESat-2 ATL08 Parameters Used for Modeling and Analysis. Table S2: Description of GEDI L2B Parameters Used for Modeling and Analysis

Author Contributions: Conceptualization, Q.S.; Methodology, L.F. and Q.S.; Software, L.F. and Q.S.; Validation, L.F., Q.S., H.H. and S.M.; Formal Analysis, L.F. and Q.S.; Investigation, L.F., Q.S., H.H., Z.G.L., S.M., C.Q., R.W., Q.X., X.Z., Y.Z. and H.C.; Resources, Q.S.; Data Curation, Q.S., L.F. and Z.G.L.; Writing—Original Draft Preparation, L.F.; Writing—Review & Editing, L.F., Q.S., C.X., Z.L., H.H., Z.G.L., S.M., C.Q., R.W., Q.X., X.Z., Y.Z. and H.C.; Visualization, Q.S.; Supervision, Q.S., C.X., and Z.L.; Project Administration, L.F. and Q.S.; Funding Acquisition, Q.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 31860205; and the Joint Agricultural Project of Yunnan Province under Grant 202301BD070001-002.

Data Availability Statement: The GEDI and ICESat-2 data used in this study are publicly available through NASA's Earthdata portal (<https://search.earthdata.nasa.gov/>). The Sentinel-2 imagery was accessed via the Google Earth Engine platform (<https://earthengine.google.com/>), and the ALOS World 3D DEM is available from the Japan Aerospace Exploration Agency at https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm. The field plot data collected for this research are available from the corresponding author on reasonable request.

Acknowledgments: The authors would like to thank the U.S. National Aeronautics and Space Administration (NASA) for providing the GEDI and ICESat-2 data, the European Space Agency (ESA) for the Sentinel-2 imagery, and the Japan Aerospace Exploration Agency (JAXA) for the ALOS World 3D DEM data. We also acknowledge the Google Earth Engine platform for its powerful data processing capabilities that facilitated this research.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AGB	Above-Ground Biomass
AGBD	Above-Ground Biomass Density
ATLAS	Advanced Topographic Laser Altimeter System
CDF	Cumulative Distribution Function
CRPS	Continuous Ranked Probability Score
DBH	Diameter at Breast Height
DEM	Digital Elevation Model
DRAGANN	Differential, Regressive, and Gaussian Adaptive Nearest Neighbor
DVI	Difference Vegetation Index
EBKRP	Empirical Bayesian Kriging Regression Prediction
ESA	European Space Agency
EVI	Enhanced Vegetation Index
GBDT	Gradient Boosting Decision Tree
GDVI	Green Difference Vegetation Index
GEDI	Global Ecosystem Dynamics Investigation
GNDVI	Green Normalized Difference Vegetation Index
GRVI	Green Ratio Vegetation Index
ICC	Intraclass Correlation Coefficient
ICESat-2	Ice, Cloud, and land Elevation Satellite-2
InSAR	Interferometric SAR
ISS	International Space Station
kNN	k-Nearest Neighbors
LAI	Leaf Area Index
LiDAR	Light Detection and Ranging

LOOCV	Leave-One-Out Cross-Validation
MAE	Mean Absolute Error
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NPGI	Normalized Pigment Chlorophyll Index
OOF	Out-of-Fold
PCL	Photon Counting LiDAR
RBF	Radial Basis Function
RF	Random Forest
RFE	Recursive Feature Elimination
RH	Relative Height
RMSE	Root Mean Square Error
RR	Ridge Regression
RVI	Ratio Vegetation Index
SAR	Synthetic Aperture Radar
SAVI	Soil-Adjusted Vegetation Index
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine
XGBoost	eXtreme Gradient Boosting

References

1. Food and Agriculture Organization of the United Nations (FAO). *The State of the World’s Forests 2018: Forest Pathways to Sustainable Development*; FAO: Rome, Italy, 2018; ISBN 978-92-5-130561-4.
2. Chen, Q.; McRoberts, R.E.; Wang, C.; Radtke, P.J. Forest Aboveground Biomass Mapping and Estimation across Multiple Spatial Scales Using Model-Based Inference. *Remote Sens. Environ.* **2016**, *184*, 350–360, doi:10.1016/j.rse.2016.07.023.
3. Ma, T.; Zhang, C.; Ji, L.; Zuo, Z.; Beckline, M.; Hu, Y.; Li, X.; Xiao, X. Development of Forest Aboveground Biomass Estimation, Its Problems and Future Solutions: A Review. *Ecol. Indic.* **2024**, *159*, 111653, doi:10.1016/j.ecolind.2024.111653.
4. Huang, H.; Liu, C.; Wang, X.; Zhou, X.; Gong, P. Integration of Multi-Resource Remotely Sensed Data and Allometric Models for Forest Aboveground Biomass Estimation in China. *Remote Sens. Environ.* **2019**, *221*, 225–234, doi:10.1016/j.rse.2018.11.017.
5. Hummel, S.; Hudak, A.T.; Uebler, E.H.; Falkowski, M.J.; Megown, K.A. A Comparison of Accuracy and Cost of LiDAR versus Stand Exam Data for Landscape Management on the Malheur National Forest. *J. For.* **2011**, *109*, 267–273, doi:10.1093/jof/109.5.267.
6. Rodríguez-Veiga, P.; Quegan, S.; Carreiras, J.; Persson, H.J.; Fransson, J.E.S.; Hoschilo, A.; Ziolkowski, D.; Stereńczak, K.; Lohberger, S.; Stängel, M.; et al. Forest Biomass Retrieval Approaches from Earth Observation in Different Biomes. *Int. J. Appl. Earth Obs. Geoinformation* **2019**, *77*, 53–68, doi:10.1016/j.jag.2018.12.008.
7. Wulder, M.A.; White, J.C.; Nelson, R.F.; Næsset, E.; Ørka, H.O.; Coops, N.C.; Hilker, T.; Bater, C.W.; Gobakken, T. Lidar Sampling for Large-Area Forest Characterization: A Review. *Remote Sens. Environ.* **2012**, *121*, 196–209, doi:10.1016/j.rse.2012.02.001.
8. Duncanson, L.; Kellner, J.R.; Armston, J.; Dubayah, R.; Minor, D.M.; Hancock, S.; Healey, S.P.; Patterson, P.L.; Saarela, S.; Marselis, S.; et al. Aboveground Biomass Density Models for NASA’s Global Ecosystem Dynamics Investigation (GEDI) Lidar Mission. *Remote Sens. Environ.* **2022**, *270*, 112845, doi:10.1016/j.rse.2021.112845.
9. Neuenschwander, A.; Pitts, K. The ATL08 Land and Vegetation Product for the ICESat-2 Mission. *Remote Sens. Environ.* **2019**, *221*, 247–259, doi:10.1016/j.rse.2018.11.005.
10. Fu, L.; Shu, Q.; Yang, Z.; Xia, C.; Zhang, X.; Zhang, Y.; Li, Z.; Li, S. Accuracy Assessment of Topography and Forest Canopy Height in Complex Terrain Conditions of Southern China Using ICESat-2 and GEDI Data. *Front. Plant Sci.* **2025**, *16*, 1547688, doi:10.3389/fpls.2025.1547688.

11. Li, D.; Wu, B.; Chen, B.; Qin, C.; Wang, Y.; Zhang, Y.; Xue, Y. Open-Surface River Extraction Based on Sentinel-2 MSI Imagery and DEM Data: Case Study of the Upper Yellow River. *Remote Sens.* **2020**, *12*, 2737, doi:10.3390/rs12172737.
12. Hajj, M.; Baghdadi, N.; Fayad, I.; Vieilledent, G.; Bailly, J.-S.; Minh, D. Interest of Integrating Spaceborne LiDAR Data to Improve the Estimation of Biomass in High Biomass Forested Areas. *Remote Sens.* **2017**, *9*, 213, doi:10.3390/rs9030213.
13. Xia, C.; Zhou, W.; Shu, Q.; Wu, Z.; Wang, M.; Xu, L.; Yang, Z.; Yu, J.; Song, H.; Duan, D. Unlocking Vegetation Health: Optimizing GEDI Data for Accurate Chlorophyll Content Estimation. *Front. Plant Sci.* **2024**, *15*, 1492560, doi:10.3389/fpls.2024.1492560.
14. Chi, H.; Sun, G.; Huang, J.; Li, R.; Ren, X.; Ni, W.; Fu, A. Estimation of Forest Aboveground Biomass in Changbai Mountain Region Using ICESat/GLAS and Landsat/TM Data. *Remote Sens.* **2017**, *9*, 707, doi:10.3390/rs9070707.
15. Zhao, X.; Hu, W.; Han, J.; Wei, W.; Xu, J. Urban Above-Ground Biomass Estimation Using GEDI Laser Data and Optical Remote Sensing Images. *Remote Sens.* **2024**, *16*, 1229, doi:10.3390/rs16071229.
16. Silva, C.A.; Duncanson, L.; Hancock, S.; Neuenschwander, A.; Thomas, N.; Hofton, M.; Fatoyinbo, L.; Simard, M.; Marshak, C.Z.; Armston, J.; et al. Fusing Simulated GEDI, ICESat-2 and NISAR Data for Regional Aboveground Biomass Mapping. *Remote Sens. Environ.* **2021**, *253*, 112234, doi:10.1016/j.rse.2020.112234.
17. Qi, W.; Saarela, S.; Armston, J.; Ståhl, G.; Dubayah, R. Forest Biomass Estimation over Three Distinct Forest Types Using TanDEM-X InSAR Data and Simulated GEDI Lidar Data. *Remote Sens. Environ.* **2019**, *232*, 111283, doi:10.1016/j.rse.2019.111283.
18. Seely, H.; Coops, N.C.; White, J.C.; Montwé, D.; Ragab, A. Forest Biomass Estimation Using Deep Learning Data Fusion of Lidar, Multispectral, and Topographic Data Remote Sensing of Environment 2024.
19. Chen, L.; Ren, C.; Bao, G.; Zhang, B.; Wang, Z.; Liu, M.; Man, W.; Liu, J. 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 Sens.* **2022**, *14*, 2743, doi:10.3390/rs14122743.
20. Duncanson, L.; Neuenschwander, A.; Hancock, S.; Thomas, N.; Fatoyinbo, T.; Simard, M.; Silva, C.A.; Armston, J.; Luthcke, S.B.; Hofton, M.; et al. Biomass Estimation from Simulated GEDI, ICESat-2 and NISAR across Environmental Gradients in Sonoma County, California. *Remote Sens. Environ.* **2020**, *242*, 111779, doi:10.1016/j.rse.2020.111779.
21. Cao, L.; Coops, N.C.; Sun, Y.; Ruan, H.; Wang, G.; Dai, J.; She, G. Estimating Canopy Structure and Biomass in Bamboo Forests Using Airborne LiDAR Data. *ISPRS J. Photogramm. Remote Sens.* **2019**, *148*, 114–129, doi:10.1016/j.isprsjprs.2018.12.006.
22. Chaparro, D.; Duveiller, G.; Piles, M.; Cescatti, A.; Vall-llossera, M.; Camps, A.; Entekhabi, D. Sensitivity of L-Band Vegetation Optical Depth to Carbon Stocks in Tropical Forests: A Comparison to Higher Frequencies and Optical Indices. *Remote Sens. Environ.* **2019**, *232*, 111303, doi:10.1016/j.rse.2019.111303.
23. Li, R.; Xia, H.; Zhao, X.; Guo, Y. Mapping Evergreen Forests Using New Phenology Index, Time Series Sentinel-1/2 and Google Earth Engine. *Ecol. Indic.* **2023**, *149*, 110157, doi:10.1016/j.ecolind.2023.110157.
24. Alavipanah, S.K.; Karimi Firozjaei, M.; Sedighi, A.; Fatholoulumi, S.; Zare Naghadehi, S.; Saleh, S.; Naghdizadegan, M.; Gomeh, Z.; Arsanjani, J.J.; Makki, M.; et al. The Shadow Effect on Surface Biophysical Variables Derived from Remote Sensing: A Review. *Land* **2022**, *11*, 2025, doi:10.3390/land11112025.
25. Liu, A.; Cheng, X.; Chen, Z. Performance Evaluation of GEDI and ICESat-2 Laser Altimeter Data for Terrain and Canopy Height Retrievals. *Remote Sens. Environ.* **2021**, *264*, 112571, doi:10.1016/j.rse.2021.112571.
26. Zhao, R.; Ni, W.; Zhang, Z.; Dai, H.; Yang, C.; Li, Z.; Liang, Y.; Liu, Q.; Pang, Y.; Li, Z.; et al. Optimizing Ground Photons for Canopy Height Extraction from ICESat-2 Data in Mountainous Dense Forests. *Remote Sens. Environ.* **2023**, *299*, 113851, doi:10.1016/j.rse.2023.113851.
27. Fu, X.; Sun, M.; Yang, Y.; Qian, Q.; Yu, H.; Hu, X. Application of Remote Sensing in Monitoring Large Sympodial Bamboo Resources in Dehong Prefecture. *West. For. Sci.* **2012**, *41*, 88–92, doi:10.16473/j.cnki.xblykx1972.2012.04.020.
28. Fu, L.; Shu, Q.; Xia, C.; Li, Z.; Zhang, X.; Zhang, Y. ICESat-2 Performance for Terrain and Canopy Height Retrieval in Complex Mountainous Environments. *Remote Sens.* **2025**, *17*, 1897, doi:10.3390/rs17111897.

29. Markus, T.; Neumann, T.; Martino, A.; Abdalati, W.; Brunt, K.; Csatho, B.; Farrell, S.; Fricker, H.; Gardner, A.; Harding, D.; et al. The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2): Science Requirements, Concept, and Implementation. *Remote Sens. Environ.* **2017**, *190*, 260–273, doi:10.1016/j.rse.2016.12.029.
30. Neumann, T.A.; Martino, A.J.; Markus, T.; Bae, S.; Bock, M.R.; Brenner, A.C.; Brunt, K.M.; Cavanaugh, J.; Fernandes, S.T.; Hancock, D.W.; et al. The Ice, Cloud, and Land Elevation Satellite – 2 Mission: A Global Geolocated Photon Product Derived from the Advanced Topographic Laser Altimeter System. *Remote Sens. Environ.* **2019**, *233*, 111325, doi:10.1016/j.rse.2019.111325.
31. Pronk, M.; Eleveld, M.; Ledoux, H. Assessing Vertical Accuracy and Spatial Coverage of ICESat-2 and GEDI Spaceborne Lidar for Creating Global Terrain Models. *Remote Sens.* **2024**, *16*, 2259, doi:10.3390/rs16132259.
32. Magruder, L.; Brunt, K.; Neumann, T.; Klotz, B.; Alonzo, M. Passive Ground-Based Optical Techniques for Monitoring the On-Orbit ICESat-2 Altimeter Geolocation and Footprint Diameter. *Earth Space Sci.* **2021**, *8*, e2020EA001414, doi:10.1029/2020EA001414.
33. Jiang, F.; Zhao, F.; Ma, K.; Li, D.; Sun, H. Mapping the Forest Canopy Height in Northern China by Synergizing Icesat-2 with Sentinel-2 Using a Stacking Algorithm. *Remote Sens.* **2021**, *13*, 1535, doi:10.3390/rs13081535.
34. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002, doi:10.1016/j.srs.2020.100002.
35. Zhu, X.; Ren, Z.; Nie, S.; Bao, G.; Ha, G.; Bai, M.; Liang, P. DEM Generation from GF-7 Satellite Stereo Imagery Assisted by Space-Borne LiDAR and Its Application to Active Tectonics. *Remote Sens.* **2023**, *15*, 1480, doi:10.3390/rs15061480.
36. Zhou, Y.; Taylor, D.M.; Tang, H. Improved Country-Wide Estimation of above-Ground Tropical Forest Biomass Using Locally Calibrated GEDI Spaceborne LiDAR Data. *Environ. Res. Lett.* **2025**, *20*, 014017, doi:10.1088/1748-9326/ad9aba.
37. Liang, M.; González-Roglich, M.; Roehrdanz, P.; Tabor, K.; Zvoleff, A.; Leitold, V.; Silva, J.; Fatoyinbo, T.; Hansen, M.; Duncanson, L. Assessing Protected Area's Carbon Stocks and Ecological Structure at Regional-Scale Using GEDI Lidar. *Glob. Environ. Change* **2023**, *78*, 102621, doi:10.1016/j.gloenvcha.2022.102621.
38. Xia, C.; Zhou, W.; Shu, Q.; Wu, Z.; Xu, L.; Yang, H.; Qin, Z.; Wang, M.; Duan, D. Regional Scale Inversion of Chlorophyll Content of Dendrocalamus Giganteus by Multi-Source Remote Sensing. *Forests* **2024**, *15*, 1211, doi:10.3390/f15071211.
39. Xu, L.; Shu, Q.; Fu, H.; Zhou, W.; Luo, S.; Gao, Y.; Yu, J.; Guo, C.; Yang, Z.; Xiao, J.; et al. Estimation of Quercus Biomass in Shangri-La Based on GEDI Spaceborne Lidar Data. *Forests* **2023**, *14*, 876, doi:10.3390/f14050876.
40. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150, doi:10.1016/0034-4257(79)90013-0.
41. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213, doi:10.1016/S0034-4257(02)00096-2.
42. Sripada, R.P.; Heiniger, R.W.; White, J.G.; Meijer, A.D. Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. *Agron. J.* **2006**, *98*, 968–977, doi:10.2134/agronj2005.0200.
43. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298, doi:10.1016/S0034-4257(96)00072-7.
44. Verrelst, J.; Schaepman, M.E.; Koetz, B.; Kneubühler, M. Angular Sensitivity Analysis of Vegetation Indices Derived from CHRIS/PROBA Data. *Remote Sens. Environ.* **2008**, *112*, 2341–2353, doi:10.1016/j.rse.2007.11.001.
45. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A Commentary Review on the Use of Normalized Difference Vegetation Index (NDVI) in the Era of Popular Remote Sensing. *J. For. Res.* **2021**, *32*, 1–6, doi:10.1007/s11676-020-01155-1.
46. Carter, G.A. Ratios of Leaf Reflectances in Narrow Wavebands as Indicators of Plant Stress. *Int. J. Remote Sens.* **1994**, *15*, 697–703, doi:10.1080/01431169408954109.
47. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666, doi:10.2307/1936256.

48. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309, doi:10.1016/0034-4257(88)90106-X.
49. Zaresefat, M.; Derakhshani, R.; Griffioen, J. Empirical Bayesian Kriging, a Robust Method for Spatial Data Interpolation of a Large Groundwater Quality Dataset from the Western Netherlands. *Water* **2024**, *16*, 2581, doi:10.3390/w16182581.
50. Albers, A.; Collet, P.; Benoist, A.; Hélias, A. Data and Non-Linear Models for the Estimation of Biomass Growth and Carbon Fixation in Managed Forests. *Data Brief* **2019**, *23*, 103841, doi:10.1016/j.dib.2019.103841.
51. Chen, X.; Sun, Q.; Hu, J. Generation of Complete SAR Geometric Distortion Maps Based on DEM and Neighbor Gradient Algorithm. *Appl. Sci.* **2018**, *8*, 2206, doi:10.3390/app8112206.
52. Wu, C.; Tao, H.; Zhai, M.; Lin, Y.; Wang, K.; Deng, J.; Shen, A.; Gan, M.; Li, J.; Yang, H. Using Nonparametric Modeling Approaches and Remote Sensing Imagery to Estimate Ecological Welfare Forest Biomass. *J. For. Res.* **2018**, *29*, 151–161, doi:10.1007/s11676-017-0404-9.
53. Meng, Q.; Cieszewski, C.J.; Madden, M.; Borders, B.E. K Nearest Neighbor Method for Forest Inventory Using Remote Sensing Data. *GIScience Remote Sens.* **2007**, *44*, 149–165, doi:10.2747/1548-1603.44.2.149.
54. Franco-Lopez, H.; Ek, A.R.; Bauer, M.E. Estimation and Mapping of Forest Stand Density, Volume, and Cover Type Using the k-Nearest Neighbors Method. *Remote Sens. Environ.* **2001**, *77*, 251–274, doi:10.1016/S0034-4257(01)00209-7.
55. Heumann, B.W. An Object-Based Classification of Mangroves Using a Hybrid Decision Tree—Support Vector Machine Approach. *Remote Sens.* **2011**, *3*, 2440–2460, doi:10.3390/rs3112440.
56. Wu, C.-H.; Tzeng, G.-H.; Lin, R.-H. A Novel Hybrid Genetic Algorithm for Kernel Function and Parameter Optimization in Support Vector Regression. *Expert Syst. Appl.* **2009**, *36*, 4725–4735, doi:10.1016/j.eswa.2008.06.046.
57. Ali, J.; Khan, R.; Ahmad, N.; Maqsood, I. Random Forests and Decision Trees. **2012**, *9*.
58. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM: San Francisco California USA, August 13 2016; pp. 785–794.
59. Zhang, Y.; Ma, J.; Liang, S.; Li, X.; Liu, J. A Stacking Ensemble Algorithm for Improving the Biases of Forest Aboveground Biomass Estimations from Multiple Remotely Sensed Datasets. *GIScience Remote Sens.* **2022**, *59*, 234–249, doi:10.1080/15481603.2021.2023842.
60. Lamane, H.; Mouhir, L.; Moussadek, R.; Baghdad, B.; Kisi, O.; El Bilali, A. Interpreting Machine Learning Models Based on SHAP Values in Predicting Suspended Sediment Concentration. *Int. J. Sediment Res.* **2025**, *40*, 91–107, doi:10.1016/j.ijsrc.2024.10.002.
61. Borselli, L.; Cassi, P.; Torri, D. Prolegomena to Sediment and Flow Connectivity in the Landscape: A GIS and Field Numerical Assessment. *CATENA* **2008**, *75*, 268–277, doi:10.1016/j.catena.2008.07.006.
62. Tian, S.; Zheng, G.; Eitel, J.U.; Zhang, Q. A Lidar-Based 3-D Photosynthetically Active Radiation Model Reveals the Spatiotemporal Variations of Forest Sunlit and Shaded Leaves. *Remote Sens.* **2021**, *13*, 1002, doi:10.3390/rs13051002.
63. Hardiman, B.S.; Bohrer, G.; Gough, C.M.; Vogel, C.S.; Curtis, P.S. The Role of Canopy Structural Complexity in Wood Net Primary Production of a Maturing Northern Deciduous Forest. *Ecology* **2011**, *92*, 1818–1827, doi:10.1890/10-2192.1.
64. Zhao, Y.; Staudenmayer, J.; Coull, B.A.; Wand, M.P. General Design Bayesian Generalized Linear Mixed Models. *Stat. Sci.* **2006**, *21*, 35–51.
65. Li, N.; Hu, M.; Xie, J.; Wei, L.; Wu, T.; Zhang, W.; Gu, S.; Li, L. Enhancing Aboveground Biomass Estimation in Moso Bamboo Forests: The Role of on-Year and off-Year Phenomena in Remote Sensing. *Front. For. Glob. Change* **2025**, *8*, 1515767, doi:10.3389/ffgc.2025.1515767.
66. Wang, J.; Du, H.; Li, X.; Mao, F.; Zhang, M.; Liu, E.; Ji, J.; Kang, F. Remote Sensing Estimation of Bamboo Forest Aboveground Biomass Based on Geographically Weighted Regression. *Remote Sens.* **2021**, *13*, 2962, doi:10.3390/rs13152962.
67. Chen, Y.; Li, L.; Lu, D.; Li, D. Exploring Bamboo Forest Aboveground Biomass Estimation Using Sentinel-2 Data. *Remote Sens.* **2018**, *11*, 7, doi:10.3390/rs11010007.

68. Gao, Y.; Lu, D.; Li, G.; Wang, G.; Chen, Q.; Liu, L.; Li, D. Comparative Analysis of Modeling Algorithms for Forest Aboveground Biomass Estimation in a Subtropical Region. *Remote Sens.* **2018**, *10*, 627, doi:10.3390/rs10040627.
69. Zhang, L.; Zhao, Y.; Chen, C.; Li, X.; Mao, F.; Lv, L.; Yu, J.; Song, M.; Huang, L.; Chen, J.; et al. UAV-LiDAR Integration with Sentinel-2 Enhances Precision in AGB Estimation for Bamboo Forests. *Remote Sens.* **2024**, *16*, 705, doi:10.3390/rs16040705.
70. Shu, Q.; Xi, L.; Wang, K.; Xie, F.; Pang, Y.; Song, H. Optimization of Samples for Remote Sensing Estimation of Forest Aboveground Biomass at the Regional Scale. *Remote Sens.* **2022**, *14*, 4187, doi:10.3390/rs14174187.
71. Fang, Y.; Leung, L.R.; Koven, C.D.; Bisht, G.; Detto, M.; Cheng, Y.; McDowell, N.; Muller-Landau, H.; Wright, S.J.; Chambers, J.Q. Modeling the Topographic Influence on Aboveground Biomass Using a Coupled Model of Hillslope Hydrology and Ecosystem Dynamics. *Geosci. Model Dev.* **2022**, *15*, 7879–7901, doi:10.5194/gmd-15-7879-2022.
72. McNichol, B.H.; Wang, R.; Hefner, A.; Helzer, C.; McMahon, S.M.; Russo, S.E. Topography-driven Microclimate Gradients Shape Forest Structure, Diversity, and Composition in a Temperate Refugial Forest. *Plant-Environ. Interact.* **2024**, *5*, e10153, doi:10.1002/pei3.10153.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.