




Article

Improving Forest Above-Ground Biomass Estimation by Integrating Individual Machine Learning Models

Mi Luo ^{1,2}, Shoaib Ahmad Anees ³, Qiuyan Huang ^{1,2,*}, Xin Qin ⁴, Zhihao Qin ^{1,2,5}, Jianlong Fan ⁶, Guangping Han ⁷, Liguozhang ^{1,2} and Helmi Zulhaidi Mohd Shafri ⁸

¹ Key Laboratory of Remote Sensing for Subtropical Agriculture, School of Geography and Planning, Nanning Normal University, Nanning 530001, China; lm@nnnu.edu.cn (M.L.); qinzhihao@caas.cn (Z.Q.); 20191221@nnnu.edu.cn (L.Z.)

² Key Laboratory of Environment Change and Resources Use in Beibu Gulf, Nanning Normal University, Nanning 530001, China

³ Department of Forestry, The University of Agriculture, Dera Ismail Khan 29050, Pakistan; anees.shoaib@gmail.com

⁴ Guangxi Vocational Normal University, Nanning 530001, China; lindaqx@163.com

⁵ State Key Laboratory of Efficient Utilization of Arid and Semi-Arid Arable Land in Northern China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

⁶ National Satellite Meteorological Center, Beijing 100081, China; fanjl@cma.gov.cn

⁷ Guangxi Zhuang Autonomous Region Institute of Natural Resources Remote Sensing, Nanning 530028, China; hgpjczx@163.com

⁸ Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), Serdang 43400, Malaysia; helmi@upm.edu.my

* Correspondence: hqy@nnnu.edu.cn; Tel.: +86-136-2771-2102

Abstract: The accurate estimation of forest above-ground biomass (AGB) is crucial for sustainable forest management and tracking the carbon cycle of forest ecosystem. Machine learning algorithms have been proven to have great potential in forest AGB estimation with remote sensing data. Though many studies have demonstrated that a single machine learning model can produce highly accurate estimations of forest AGB in many situations, efforts are still required to explore the possible improvement in forest AGB estimation for a specific scenario under study. This study aims to investigate the performance of novel ensemble machine learning methods for forest AGB estimation and analyzes whether these methods are affected by forest types, independent variables, and spatial autocorrelation. Four well-known machine learning models (CatBoost, LightGBM, random forest (RF), and XGBoost) were compared for forest AGB estimation in the study using eight scenarios devised on the basis of two study regions, two variable types, and two validation strategies. Subsequently, a hybrid model combining the strengths of these individual models was proposed for forest AGB estimation. The findings indicated that no individual model outperforms the others in all scenarios. The RF model demonstrates superior performance in scenarios 5, 6, and 7, while the CatBoost model shows the best performance in the remaining scenarios. Moreover, the proposed hybrid model consistently has the best performance in all scenarios in spite of some uncertainties. The ensemble strategy developed in this study for the hybrid model substantially improves estimation accuracy and exhibits greater stability, effectively addressing the challenge of model selection encountered in the forest AGB forecasting process.

Keywords: above-ground biomass; ensemble model; CatBoost; machine learning



Citation: Luo, M.; Anees, S.A.; Huang, Q.; Qin, X.; Qin, Z.; Fan, J.; Han, G.; Zhang, L.; Shafri, H.Z.M. Improving Forest Above-Ground Biomass Estimation by Integrating Individual Machine Learning Models. *Forests* **2024**, *15*, 975. <https://doi.org/10.3390/f15060975>

Academic Editor: Eric Casella

Received: 23 April 2024

Revised: 28 May 2024

Accepted: 30 May 2024

Published: 1 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The forest ecosystem, as an important carbon sink and carbon source on land, plays a major role in the global carbon cycle [1]. Accurate and robust forecasts of forest above-ground biomass (AGB) can improve our understanding of the global carbon cycle to achieve

effective carbon reduction strategies, sustainable forest management, and forest ecosystem productivity monitoring [2]. Two approaches are available for forest AGB estimation on a regional scale: field measurements and remote sensing estimation. Generally speaking, regional forest AGB estimation with field measurements is extremely expensive, time-consuming, difficult, and even destructive. It is necessary to combine field survey data and remote sensing techniques to estimate AGB for carbon science and policy.

Three strategies have been utilized to improve forest AGB estimation accuracy through remote sensing [2,3]. The first is to combine multi-source remote sensing data, such as optical, hyperspectral, and Light Detection and Ranging (Lidar) [3,4]. It has been demonstrated that a combination of multiple remote sensing data sources could effectively improve the accuracy of forest AGB estimation [5,6]. However, the high cost of hyperspectral and Lidar remote sensing data restricts their extensive applications to forest AGB estimation at regional scale. The second is to select the optimal features for the estimation. Since different remote sensing data have their own features in both spectral and spatiotemporal dimensions, several features can be extracted or computed from multisource remote sensing data. The successful selection of these features can often improve forest AGB estimation accuracy. Therefore, it is important to develop suitable feature-rich prediction models for AGB estimation, which generally requires an effective way to screen out the most important and useful information from numerous variables and eliminate redundant or irrelevant variables involved in multiple remote sensing data [5]. Obvious advances have been made in reducing model over-fitting and computation risks for forest AGB estimation [7]. Variable selection using random forests (VSURF), recursive feature elimination (RFE), and least absolute shrinkage and selection operator (LASSO) as feature selection methods have been applied in forest AGB estimation [8,9]. The third is to adopt an appropriate estimation algorithm for a specific scenario [10,11]. As is well known, a good model is very important to have an accurate estimation of forest AGB in any scenario. Several models have been developed for the estimation, ranging from simple linear regression to advanced machine learning, such as random forest (RF), support vector machine (SVM), and artificial neural networks (ANN).

Machine learning algorithms have been proven to be more effective than parameterized methods that use linear models for AGB estimation [12–14]. The exploration of new machine learning algorithms represents a direction to improve forest AGB forecast accuracy. Ensemble learning is a branch of machine learning that can mainly be divided into Bagging and Boosting algorithms [15]. RF is a popular technique in the Bagging algorithm widely used in different forecasting fields [16–19]. Boosting algorithms include adaptive boosting (AdaBoost), gradient boosting decision tree (GBDT), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), and categorical boosting (CatBoost). It has been demonstrated that the boosting algorithms (XGBoost, CatBoost, LightGBM, and GBDT) can provide significant improvements in estimation accuracy; hence, they have been widely used in such studies in biology [20] and climate science [21]. Although several efforts have attempted to use these algorithms to construct forest parameter models from remote sensing data [8,22], the new ensemble learning algorithms XGBoost, CatBoost, and LightGBM have only been recently introduced to forest AGB estimation. Few studies have been reported comparing their performance in forest AGB estimation at a regional scale. The CatBoost and XGBoost algorithms have been applied in our previous forest AGB study [8], showing their great potential in application to forest AGB estimation at a regional scale. Since these algorithms have their strengths and weaknesses, none of them have been demonstrated to be the best for forest AGB estimation.

Hybrid models have been commonly recognized as an effective strategy to improve the predictive performance of individual models and have been widely used in many forecasting issues [23]. This is because hybrid models have the ability to make maximal use of the available information involved in various single-forecast results, hence avoiding effective information waste and reducing the influence of random factors on forecast results. These strengths of hybrid models give them the best accuracy in estimation and allow them

to produce the most stable forecast results in comparison with the prediction by individual models. Several studies have demonstrated that hybrid models can enhance prediction accuracy for applications in energy [24], climate [25], and soil [26]. The performance of four individual machine learning algorithms (RF, quantile regression forest, cubist, and fuzzy logic) and two hybrid methods (random forest-ordinary kriging and quantile random regression forest-ordinary kriging) were compared in Matinfar et al. [27] to predict soil organic carbon. The results demonstrated that hybrid models outperformed individual models. Zhao et al. [28] developed a hybrid model by integrating ANN and a modified Penman-Monteith equation for latent heat flux estimation. Their results revealed that the hybrid model had better performance in extrapolation than the individual machine learning model. Machine learning integration models have not been explicitly tested for forest AGB forecasting, and studies using this model have not been seen in the literature of forest AGB estimation. Therefore, on the basis of our previous study [8], we would like to explore the capability of hybrid models in improving forest AGB prediction accuracy.

Forest AGB estimation with the above models generally ignores the important issue of spatial autocorrelation (SAC), which refers to the potential interdependence in the observation data of the variables within the same group. The presence of SAC violates the assumption of independence among the variables used to establish the models for forest AGB estimation. Therefore, ignoring SAC would make the models prone to result in a biased estimation. The best approach to account for SAC is through setting the SAC as a separate (weighing) variable or removing it from the observations, for example, by selecting a non-correlated subset for the model establishment [29]. Another effective approach for the consideration of SAC in model construction is to use a spatial cross-validation strategy, i.e., to conduct spatial clustering to the evaluation partitions in the cross-validation procedure so that the underestimation of the model due to SAC can be corrected [30]. Spatial cross-validation was used in Mayr et al. [31] to solve the SAC between two parameters in their study of a burned forest area and the number of fire occurrences. The results showed that the lower but more realistic model performances were retrieved from spatial cross-validation. Meyer et al. [32] pointed out that spatial cross-validation was essential in preventing an overoptimistic model when predicting leaf area index and land coverage. Spatial cross-validation, however, is yet to be explored in forest AGB models.

The objective of this study is to develop a framework for the accurate estimation of forest AGB, including the procedures from data input and modeling to validation. The performance of four individual machine learning models (CatBoost, LightGBM, RF, and XGBoost) will be compared for eight scenarios composed of two distinct study areas, two variable types, and two validation strategies. The effects of forest types, independent variables, and spatial autocorrelation on the forest AGB estimation accuracy will be analyzed. Consequently, a hybrid model will be developed for forest AGB estimation. Spectral variables, vegetation indexes, red-edge vegetation indexes, and texture measures of Sentinel-2 will be calculated to maximize the potential information between the derived remote sensing data and AGB. Moreover, VSURF was selected to screen out effective features to participate in establishing the forest AGB model. The influence of SAC on the forest AGB model will be evaluated by comparing random cross-validation and spatial cross-validation. Therefore, the structure of the study will be as follows: Material and Methods will be presented in Section 2, followed by the Results and Analysis given in Section 3. Finally, Sections 4 and 5 will be used to present the Discussion and Conclusion of the study.

2. Materials and Methods

2.1. Methodological Framework

Theoretically, forest AGB can be viewed as the total dry materials in weight under a unit of acreage that the forest has above the ground surface. In terms of the temporal dimension, this is the summation of the forest's growth during its life cycle [1]. Many variables play an important role in the formation of forest AGB during the forest life cycle.

Difficulties are obvious for a precise estimation of forest AGB on a regional scale. Direct measurement through felling all the trees in the forest to scale their weight would be the ideal way to obtain the most accurate data on forest AGB. However, this is destructive, highly costly, and time consuming. Thus, this approach is generally not viewed as the proper method for forest AGB estimation. With the development of modern technology, especially for remote sensing and data processing, a number of novel approaches have been developed for forest AGB estimation [3,10]. The most practical one with wide applications is to obtain a sample representing the various conditions of a forest and to develop a model with the samples for forest AGB estimation. In recent decades, machine learning algorithms have been widely used in many study fields including forest AGB estimation [7,13,16]. Due to the advantages in spatial dimensions, remote sensing has also been widely applied for forest AGB estimation on a regional scale. Therefore, the combination of machine learning with remote sensing represents a new direction for the study of forest AGB estimation [23,28]. The basic hypothesis for this combination is that remote sensing could provide a best way to simultaneously obtain the required information about the ground that directly related to the forest growth conditions and hence the biomass of the forest under study.

In order to perform this combination approach for forest AGB estimation, a number of studies have been devoted to investigating the methods for model establishment, variable selection, data processing, sampling implementation, and so on that relate to the process of forest AGB estimation from data input to final result output. In this study, we intend to compare the performance of four well-known machine learning models for forest AGB estimation under eight different scenarios. Therefore, the methodological framework of this study can be represented by Figure 1, which indicates that the framework will mainly consist of four steps: (1) data preprocessing, (2) feature extraction, (3) feature selection, and (4) modeling and validation.

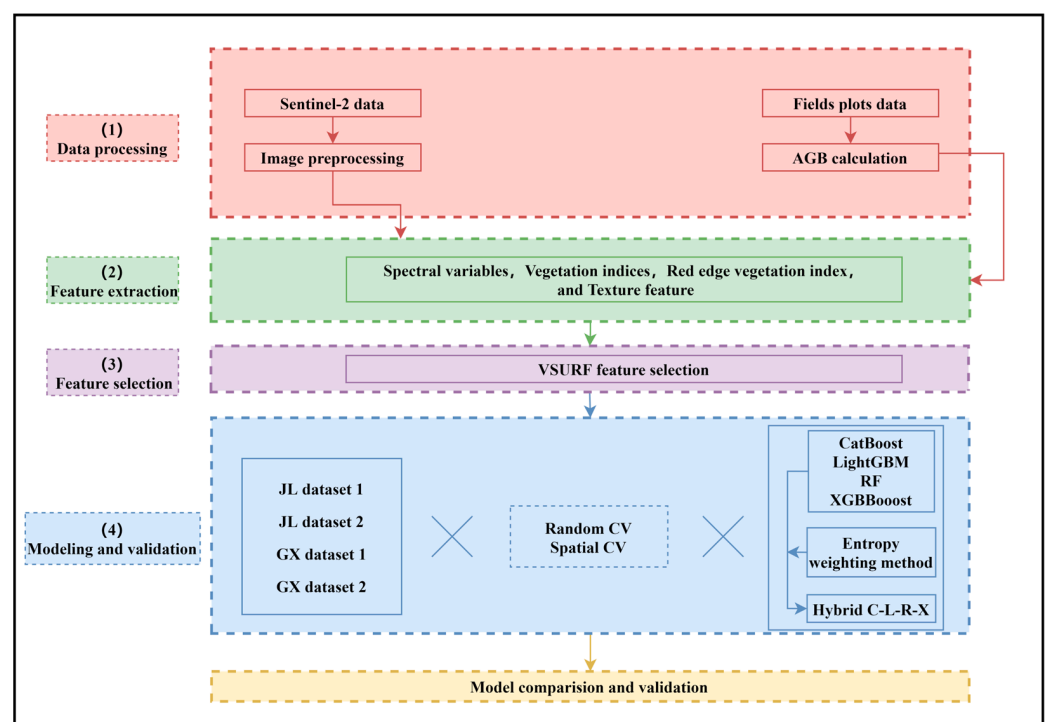


Figure 1. The methodological framework for estimating the AGB using the individual models and hybrid model under the spatial cross-validation and random cross-validation based on Sentinel-2 images and filed survey data.

2.2. Study Regions

This study was carried out in two typical forest regions of China, i.e., east Jilin (JL) and central Guangxi (GX) (Figure 2), representing different climate conditions, forest structure,

and species composition. The JL region has an area of $12.60 \times 10^4 \text{ km}^2$. The landscape of the JL region mainly features with low mounts and hills. The elevation of the region ranges from 200 m to about 1000 m. The climate of the region is the type of temperate continental monsoon, with an average air temperature below $-11 \text{ }^\circ\text{C}$ in winter and above $23 \text{ }^\circ\text{C}$ in summer. The forest under study is mainly the natural forest in the region, with a few planted forests. Mixed broad-leaved forest dominates the forest type of the region, and the main tree species in the forest are *Pinus koraiensis* Siebold et Zuccarini, *Larix olgensis* Henry, and *Juglans mandshurica* Maxim.

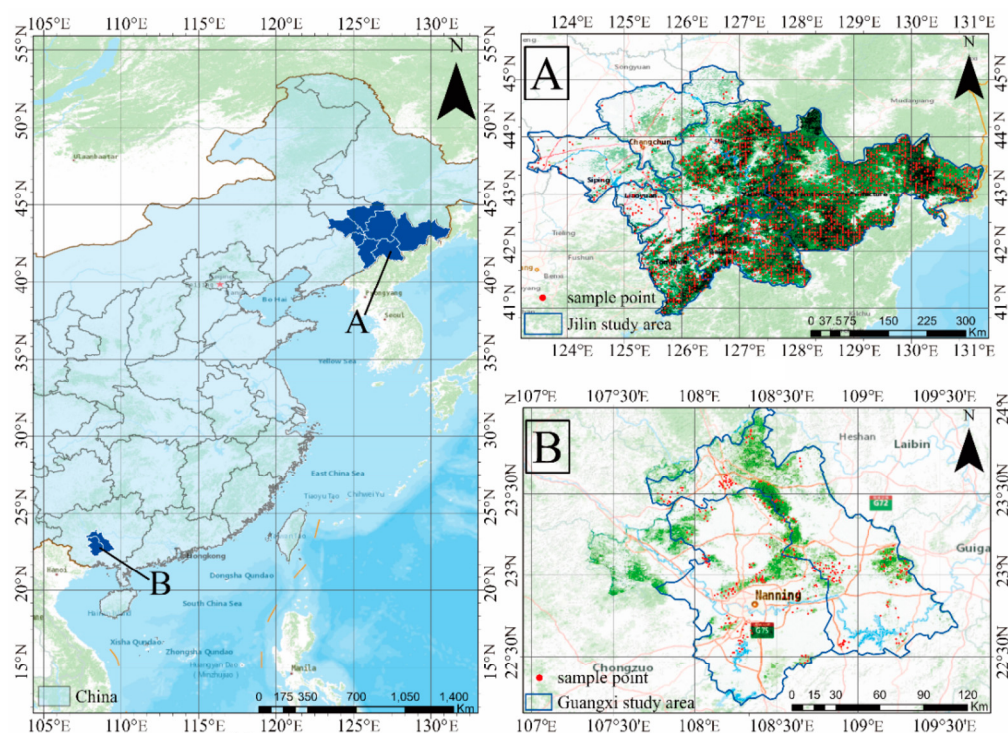


Figure 2. Geographical location of the two study regions in China: east Jilin (A) and central Guangxi (B). In blue is the study area, in red are the sample points, and in green is the forest area.

The GX region is an area of $2.21 \times 10^4 \text{ km}^2$. The landform of the region is dominated by plains and hills, with an elevation ranging from 80 m to about 400 m above sea level. The climate of the GX region is mainly a pattern of humid subtropical monsoon, with an annual average temperature of $21.6 \text{ }^\circ\text{C}$. The type of forest under study is mainly the planted forest, with the main tree species being *Pinus massoniana* Lamb., *Eucalyptus robusta* Smith, and *Cunninghamia lanceolata* (Lamb.) Hook.

2.3. Field Data and Preprocessing

In the JL region, the required field AGB data of forest plot samples were obtained from the 9th National Forest Continuous Inventory datasets of Jilin Province, which were generated from the forest inventory campaigns conducted in 2014. The plot samples had an acreage of 0.06 hectares extracted with a fixed spatial grid of $4 \text{ km} \times 8 \text{ km}$. To obtain the field AGB data in the JL region, we assumed that potential changes in forest AGB due to temporal differences between field inventories and remote sensing data acquisitions in 2018 would have very limited effects on the estimated results from the modeling in the study. In the GX region, the field campaigns for forest AGB data were carried out during 2017 and 2018 with a plot size of $20 \text{ m} \times 30 \text{ m}$. The plots were representative and randomly distributed, and non-forest areas were masked out. All living trees with a diameter at breast height (DBH) larger than 5 cm in the sampling plot were measured to obtain the information on species, tree height, canopy density, slope, aspect, slope position, and coordinates. Furthermore, the field sampling plots either in non-forested areas (cropland,

water, urban land, and bare land) or covered with clouds in the remote sensing images were excluded.

The AGB of an individual tree was calculated using the general one-variable AGB model [33,34] as follows:

$$M = a \times D^{7/3} \quad (1)$$

$$a = 0.3 \times p \quad (2)$$

where M (kg) is the AGB of a tree, D (cm) is the diameter at breast height of the tree, a is the parameter of tree species, and p (g cm^{-3}) is the basic wood density (Table S1 in Supplementary Materials). The plot AGB was converted to per hectare biomass (Mg ha^{-1}).

2.4. Sentinel-2 Image Preprocessing and Variable Extraction

Image data used in the study for forest AGB modeling were Copernicus Sentinel-2 satellite imagery downloaded from the Google Earth Engine (GEE) platform (<https://earthengine.google.com/>, accessed on 22 April 2024). Sentinel-2 MSI data contain two products, Level-1C and Level-2A. The Level-2A product has been orthographically and atmospherically corrected and geographically registered. To represent the best growing status of forest for AGB estimation, the Normalized Difference Vegetation Index (NDVI) maximum synthesis method, which uses the best observed value in adjacent time as the filling value, was used to generate the composite images of the Level-2A product in 2018. Then, the quality assessment band (QA60) was used for cloud masking based on cloudy pixels identified. In addition, image splicing and image cropping were also performed on the GEE platform.

To maximize information on forest AGB in the study regions, 10 of the 13 Sentinel-2 bands (4 visible, 4 red-edge, and 2 short-wavelength infrared (SWIR)) were extracted and resampled to 10 m for the study. Consequently, a total of 125 predictive variables were extracted (Table 1). A total of 22 widely-used vegetation indices and 13 red-edge vegetation indices were also calculated for the study, and the specific description was presented in reference [35]. Texture measures were extracted using the gray level co-occurrence matrix method (GLCM), and eight GLCM measures with 5×5 window sizes were calculated from the 10 spectral bands.

2.5. Selection of Variables Using VSURF

VSURF, a wrapper-based algorithm that uses random forests as the base classifier [36], was used in the study to select the most discriminant features for the study. The feature selection process consists of three stages, namely thresholding step, interpretation step, and prediction step. Four parameters need to be set, including “ntree”, “nfor.thres”, “nfor.interp”, and “nfor.pred”. Specific steps are as follows. First, feature variables were ranked according to their relative importance, and low-scoring features were eliminated to reduce the number of features. Second, the VSURF algorithm was used to produce two separate subsets: a primary subset of important variables including redundancy, and a small subset avoiding redundancy. Finally, the smallest subset with minimum redundancy was selected as the input variables of the forest AGB estimation model in the study. The procedures of VSURF algorithm were performed using the VSURF package in R-4.0.5 [37]. Two datasets were outputted by the performance of VSURF procedures: the first output dataset includes all the 125 variables, and the second output dataset only includes 10 spectral variables.

Table 1. Summary of predictor variables including Sentinel-2 spectral variables, vegetation indices, red edge vegetation indices, and texture measures for AGB estimation.

Variable Type	Variable Number	Variables	Description
Spectral variables	10	B2_blue, B3_green, B4_red, B5_re1, B6_re2, B7_re3, B8_NIR, B8a_nNir, B11_SWIR1, B12_SWIR2	Band2-Blue, Band3-Green, Band4-Red, Band5-Vegetation red edge, Band6-Vegetation red edge, Band7-Vegetation red edge, Band8-NIR, Band8A-Narrow NIR, Band11-SWIR, Band12-SWIR
Vegetation indices	22	DVI	$DVI = B8 - B4$
		Clg	$Clg = (B8/B3) - 1$
		NR	$NR = B4/(B8 + B4 + B3)$
		NNIR	$NNIR = B7/(B8 + B4 + B3)$
		NG	$NG = B3/(B8 + B4 + B3)$
		NLI	$NLI = (B8^2 - B4)/(B8^2 + B4)$
		RDVI	$RDVI = (B8 - B4)/(B8 + B4)^{0.5}$
		SPVI	$SPVI = 0.4 \times 3.7 \times (B8 - B4) - 1.2 \times (B3 - B4)$
		RI	$RI = (B4 - B3)/(B4 + B3)$
		EVI	$EVI = 2.5 \times [(B8 - B4)/(B7 + (6 \times B4) - (7.5 \times B2) + 1)]$
		GARI	$GARI = [B8 - (B3 - (B2 - B4))/(B7 + (B3 - (B2 - B4)))]$
		VARIg	$VARIg = (B3 - B4)/(B3 + B4 - B2)$
		NDII	$NDII = (B8a - B11)/(B8a + B11)$
		NBR	$NBR = (B8a - B12)/(B8a + B12)$
		NMDI	$NMDI = [B8a - (B11 - B12)/(B8a + (B11 - B12))]$
		NDVI	$NDVI = (B8 - B4)/(B8 + B4)$
RGR	$RGR = B4 - B3$		
SAVI	$SAVI = (B8 - B4)/(B8 + B4 + 0.5) \times 1.5$		
RVI	$RVI = B8/B4$		
RVI54	$RVI54 = B11/B4$		
RVI64	$RVI64 = B12/B4$		
WDRVI	$WDRVI = [(0.02 \times B8 - B4)/(0.02 \times B8 + B4)] + [(1 - 0.02)/(1 + 0.02)]$		
Red edge vegetation indices	13	CIre	$CIre = (B8/B5) - 1$
		PSRI	$PSRI = (B4 - B3)/B6$
		MTCI	$MTCI = (B6 - B5)/(B5 - B4)$
		MCARI	$MCARI = [(B5 - B4) - 0.2 \times (B5 - B3)] \times (B5/B4)$
		MCARI2	$MCARI2 = [1.5 \times (2.5 \times (B8 - B4) - 1.3 \times (B8 - B3))]/[(2 \times B7 + 1)^2 - (6 \times B7 - 5 \times B4^{0.5}) - 0.5]^{0.5}$
		TCARI	$TCARI = [3 \times (B5 - B4) - 0.2 \times (B5 - B3)] \times (B5/B4)$
		TCARI2	$TCARI2 = [3 \times (B6 - B5) - 0.2 \times (B5 - B3)] \times (B5/B4)$
		TVI	$TVI = 0.5 \times (120 \times (B6 - B3) - 200 \times (B4 - B3))$
		MTVI2	$MTVI2 = [1.5 \times (1.2 \times (B8 - B3) - 2.5 \times (B4 - B3))]/[(2 \times B7 + 1)^2 - (6 \times B7 - 5 \times B4^{0.5}) - 0.5]^{0.5}$
		IRECI	$IRECI = (B7 - B4)/(B5 + B6)$
		S2REP	$S2REP = 700 + 35 \times [(B8 - B4)/2 - B5]/(B6 - B5)$
MSRren	$MSRren = [(B8a/B5) - 1]/[(B8a/B5)^{0.5} + 1]$		
WDRVIre	$WDRVIre = [(0.01 \times B8 - B5)/(0.01 \times B8 + B5)] + [(1 - 0.01)/(1 + 0.01)]$		
Texture measures	80	Bi_Mea_5, Bi_Var_5, Bi_Hom_5, Bi_Con_5, Bi_Dis_5, Bi_Ent_5, Bi_ASM_5, Bi_Cor_5	10 spectral bands of Sentinel-2 texture measurement using gray-level co-occurrence matrix based on 5×5 window size

Note: NIR is near infrared; SWIR is shortwave infrared. Bi_X_5 represents a texture image developed in Sentinel-2 10 spectral bands using the texture measure X with 5×5 pixels window, where Bi is band number, X is Mean (Mean), Var is variance, Hom represents homogeneity, Con is the contrast, Dis is dissimilarity, Ent is entropy, ASM is angular second moment, and Cor is correlation.

2.6. Spatial Cross-Validation

Since spatial autocorrelation exists among the variables used to estimate forest AGB in our study, we intend to use two strategies for the spatial cross-validation of our estimation results: random k-fold cross-validation and spatial k-fold cross-validation. The basic idea of cross-validation is to repeatedly split a dataset into two subsets, the training subset and the validation subset, where the training subset was used to train the above machine learning algorithms (4 individual models and 1 hybrid model) for model establishment, while the validation subset was used to validate the accuracy of the established model for forest AGB estimation. Details on the spatial cross-validation of forest AGB estimation with machine learning algorithms could be found in Kuhn and Johnson [38]. The only difference between random cross-validation and spatial cross-validation is in their parti-

tioning strategy of a sample dataset. In random cross-validation, the k-means clusters of geographic coordinates were used to derive the validation subset from the sample dataset, while in spatial cross-validation, a bias-reduced assessment was conducted for the model's predictive performance to enhance the stability of the model.

2.7. Development of Individual Models

The following four individual models belonging to the ensemble learning method were used to develop the forest AGB estimation models in the study:

- (1) Categorical Boosting (CatBoost). This model was developed by Dorogush et al. [39] from the GBDT algorithm for gradient bias and prediction shift [40]. Building the CatBoost model using the R language consists of four steps: data preprocessing, model initialization, model training, and model evaluation. During model training, depth, learning rate, and the number of iterations are the main parameters that need to be adjusted to improve the accuracy of the CatBoost model, which is run in R-4.05 with the 'catboost' function.
- (2) Light Gradient Boosting (LightGBM). This model was proposed by Ke et al. [41] on the basis of the decision tree algorithm. An obvious advantage of the LightGBM model is its computational speed and memory consumption [42]. Implementing LightGBM involves several crucial steps, from data preparation to model evaluation and interpretation. The initial phase in employing LightGBM within R involves careful data preparation. This includes cleaning the dataset to handle missing values and outliers and ensure data consistency. Feature engineering is also pivotal, involving the transformation of categorical variables using one-hot encoding and the creation of new features that can enhance model performance. The dataset is then split into training and testing subsets to evaluate the model's performance accurately. This model is also run in R-4.0.5 using the 'lightgbm' package. The primary hyperparameters affecting the model include learning rate, the number of iterations, and min-data.
- (3) Random Forest (RF). This model has been widely used for many applications including forest AGB estimation [43]. The running environment of the RF algorithm is also the R package 'randomForest' using 'ntree' and 'mtry' parameters.
- (4) Extreme Gradient Boosting (XGBoost). This model is a boosting algorithm based on GBDT and RF approaches, which was first introduced by Chen and Guestrin [44] who demonstrated that, in comparison with GBDT, the XGBoost model could significantly improve multithreaded processing and the optimization function, which consequently overcome the over-fitting problem [45]. The 'xgboost' package in R-4.0.5 was used to run this model. Extended grid tuning was used to train the XGBoost model on the training dataset, allowing for the investigation of different hyperparameter combinations. In addition to the number of boosting iterations (nrounds = 200), the hyperparameters also included the depth of the tree (max_depth = 2), learning rate (eta = 0.3), and minimum loss reduction, namely gamma.

The 'Caret' package (version: 6.0-84) in R-4.0.5 was used to optimize the parameters of the above models in the study.

2.8. Development of Hybrid Model

A hybrid model [46] is commonly defined as a model that combines the forecasting advantages of several individual models so that the possible errors, bias, and uncertainty caused by individual models can be maximally reduced. Therefore, it is believed that a hybrid model provides one of the best ways to improve prediction accuracy and to minimize the variance error between the predicted values. Every individual model has its own uncertainty in modeling, hence unavoidably generating its own prediction errors when subjected to the new data for modeling. The hybrid model used in this study integrated the above four machine learning algorithms (Catboost, XGB, RF, and LightGBM) on the basis of the entropy weighting method [47]. Thus, our hybrid model could be termed as

the CLRX model. The prediction process of our hybrid model, i.e., the CLRX model, can be outlined as follows:

Step 1: Calculate the prediction error e_{ij} of the individual model i ($i = 1, 2, \dots, m$, where $m = 4$ in our study due to only 4 individual models being used in the study) for sample j ($j = 1, 2, \dots, n$) in the training subset.

$$e_{ij} = \begin{cases} 1, & \text{if } \frac{|y_j - \hat{y}_j(i)|}{y_j(i)} \geq 1, \\ \frac{|y_j - \hat{y}_j(i)|}{y_j}, & \text{if } 0 \leq \frac{|y_j - \hat{y}_j(i)|}{y_j} < 1 \end{cases} \quad (3)$$

where y_j is the observed forest AGB, and $\hat{y}_j(i)$ is the predicted forest AGB using the individual model i .

Step 2: Normalize the prediction errors of each individual model i for sample j .

$$p_{ij} = \frac{e_{ij}}{\sum_{j=1}^n e_{ij}} \quad (4)$$

Step 3: Calculate the entropy weighting for each individual model.

$$h_i = -k \sum_{j=1}^n p_{ij} \ln p_{ij} \quad (5)$$

where $k = 1/\ln(n)$.

Step 4: Calculate the weight coefficient for each individual model.

$$l_i = \frac{d_i}{\sum_{i=1}^m d_i} \quad (6)$$

where $d_i = 1 - h_i$. The entropy weighting of the individual models in our hybrid model is expected to decrease with the increasing variation in its prediction error.

Step 5: Compute the hybrid model (CLRX model) by summarizing the prediction value of individual models with their weight coefficients.

$$\hat{y}_j = l_1 \hat{y}_j(1) + \dots + l_i \hat{y}_j(i) + \dots + l_m \hat{y}_j(m), \quad j = 1, 2, \dots, n \quad (7)$$

2.9. Model Evaluation

The sample dataset obtained from field investigation was divided into a training subset (80%) and a validation subset (20%). In order to avoid the over-fitting of the model for forest AGB estimation in our study, two cross-validations were performed: random 5-fold cross-validation and spatial 5-fold cross-validation. The estimation accuracy of the forest AGB models was assessed with the independent validation subset to compute the following four parameters that are generally used for validation: determination coefficient (R^2), root-mean-square error (RMSE), relative RMSE (RMSE%), and bias (Bias):

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

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

$$RMSE\% = \frac{RMSE}{\bar{y}} * 100 \quad (10)$$

$$Bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{N} \quad (11)$$

where \hat{y}_i is the predicted value of forest AGB, y_i is the observed value of forest AGB from field investigation, \bar{y} is the mean of the observed values, and N is the number of observations.

3. Results and Analysis

3.1. Forest AGB Characteristics in the Two Study Regions

Table 2 shows the forest AGB characteristics from field data in the two study regions. The forest AGB value ranged from 39.50 Mg ha⁻¹ to 235.70 Mg ha⁻¹ in the JL region and from 0.15 Mg ha⁻¹ to 108.63 Mg ha⁻¹ in GX region. Thus, it could be found that forest AGB was much higher in the JL region (mean = 159.90 Mg/ha) than in GX (mean = 53.40 Mg ha⁻¹). This may be attributed to the fact that the forest in the JL region is mainly the type of natural forest that has a long period of growing as a result of being a conservation region, while the forest in the GX region was mainly the type of man-planted economic forest that has relatively short period of growing due to its frequent chopping down for economic income. Histograms of the measured forest AGB in the JL and GX regions are depicted in Figure 3, which indicates that forest AGB was mainly distributed in the range of 50–80 Mg ha⁻¹ in the GX region and 150–200 Mg ha⁻¹ in the JL region.

Table 2. Summary descriptive statistics of the field-measured forest AGB in the two study regions.

Study Region	No. of Sampling Plots	Min (Mg ha ⁻¹)	Max (Mg ha ⁻¹)	Mean (Mg ha ⁻¹)	Standard Deviation (Mg ha ⁻¹)
JL	1812	39.50	235.70	159.90	34.69
GX	344	0.15	108.63	53.40	26.55

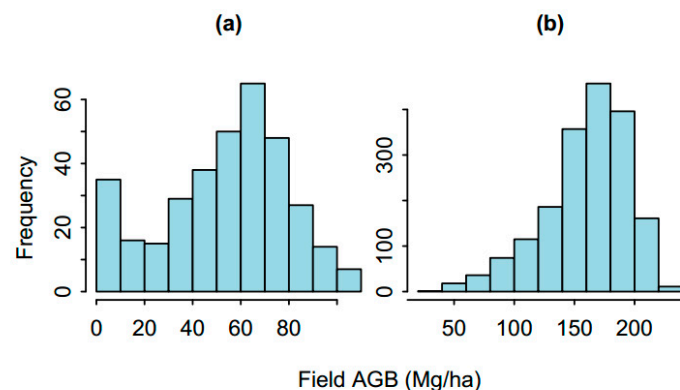


Figure 3. Histogram of the field-measured forest AGB in the (a) GX and (b) JL regions.

3.2. Optimal Number of Variables for Modelling

As described in Section 2.5, the VSURF method was used to determine the optimal number of variables for forest AGB modeling in the study. Therefore, the field datasets were divided into two subsets using the VSURF method: two for the JL region and two for the GX region. The first output subset, i.e., the primary subset, included all 125 variables, and the second subset, i.e., the small subset, only contained 10 spectral variables. The parameters are set as $n_{tree} = 500$, $n_{for.thres} = 50$, $n_{for.interp} = 25$, and $n_{for.pred} = 25$. Table 3 shows the variable selection results from using the VSURF method. GX dataset 2 is a primary subset of the VSURF outputs to the field sample dataset in the GX region. Though the dataset has 125 original variables at the threshold step (Figure 4a,b) of VSURF analysis, only 54 variables and 17 variables were selected at the interpretation and prediction steps, respectively (Figure 4c,d), after the VSURF analysis on GX dataset 2. Therefore, the dataset containing 17 final variables would be used as the input dataset to establish the prediction model for forest AGB estimation in the study of the GX region. Figure 4a shows the variable ranking results based on the mean variable importance (VI). Figure 4b presents the standard deviation change in VI (black curve), the prediction given by a CART (Classification and

Regression Trees) fitted to the standard deviations (green curve), and the CART prediction's minimum value (dotted red line). Figure 4c shows the mean out-of-bag (OOB) error for nested random forest models (black curve) and the smallest OOB error (red line) at the interpretation step. Figure 4d presents the mean OOB error rate for embedded random forest models at the prediction step. Similar results could also be obtained from the VSURF analysis on the other datasets, such as JL dataset 2 and JL dataset 1 for the JL region.

Table 3. Results of VSURF feature selection method for four datasets.

Datasets	Before Selection	After Selection
JL Dataset 1	10	7
JL Dataset 2	125	18
GX Dataset 1	10	5
GX Dataset 2	125	17

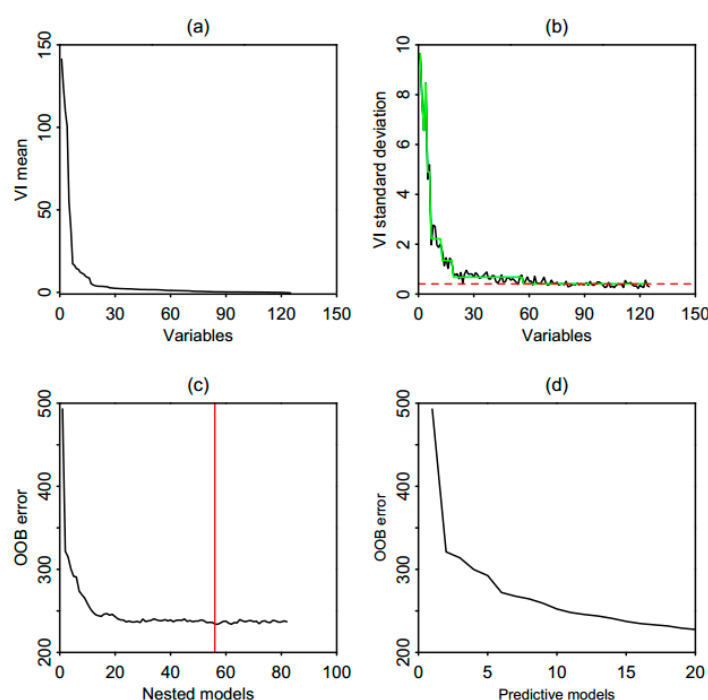


Figure 4. Results of VSURF feature selection method for GX dataset 2, including threshold step (a,b), interpretation step (c), and prediction step (d).

3.3. Variable Importance

In order to determine which variable is suitable for the final modelling of forest AGB in the two study regions, the relative importance of the selected variables in each dataset was computed for the four machine learning models (CatBoost, LightGBM, RF, and XGBoost) under the spatial cross-validation. Since different variables may have different functions in the formation of forest AGB in different regions, the selection of relatively important variables for forest AGB modelling is very necessary for different regions. Figures 5 and 6 show the top 10 variables ranked in order of importance in primary datasets, i.e., JL dataset 2 and GX dataset 2, of the two study regions. Great variation can be seen in the ranking of variable importance to the four machine learning models in the two study regions.

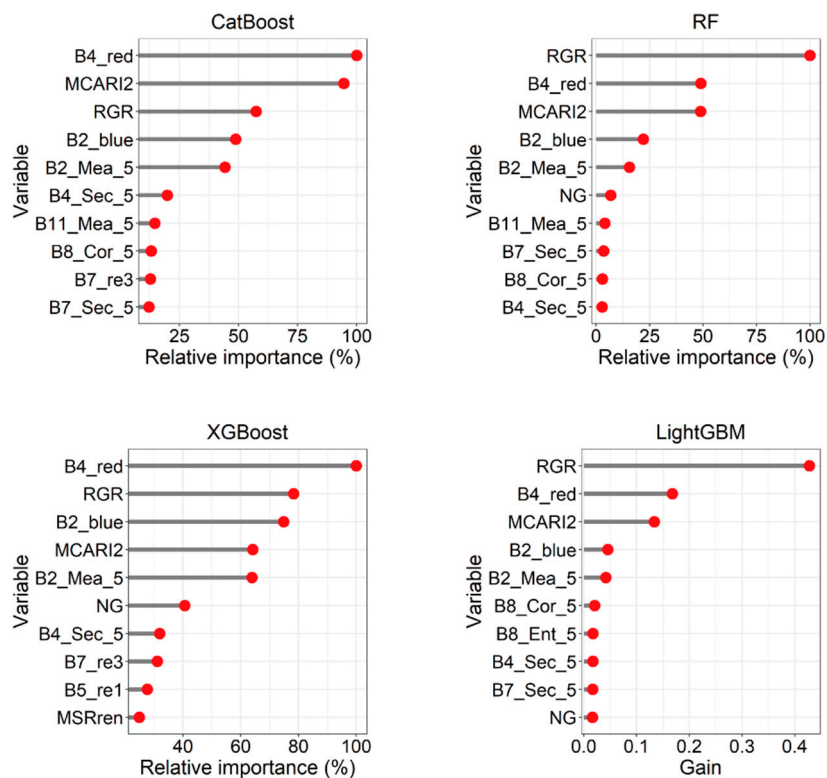


Figure 5. Variable importance ranking for JL dataset 2 using four machine learning models (CatBoost, RF, XGBoost, and LightGBM) under spatial cross-validation strategies.

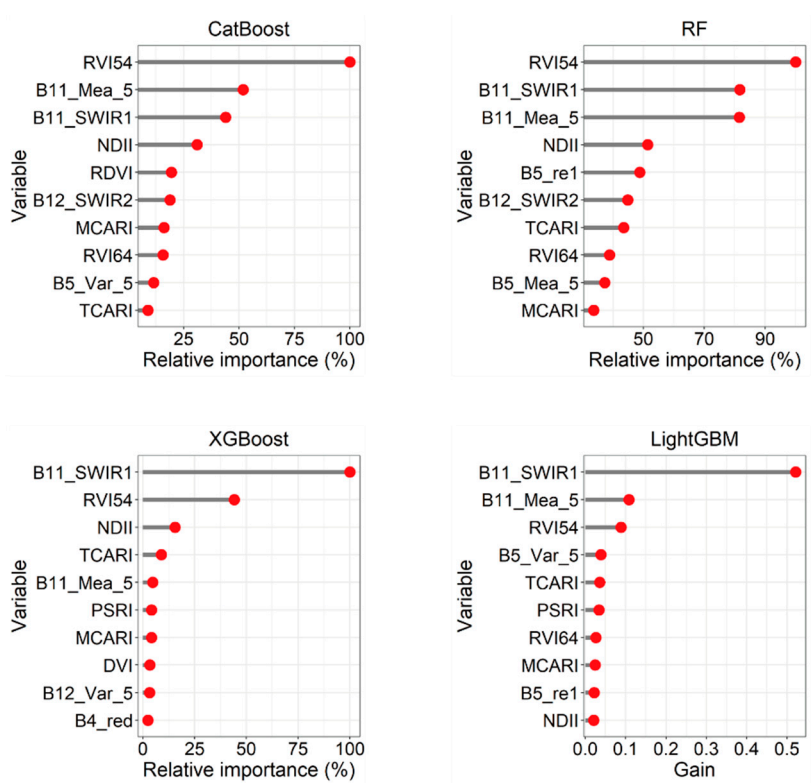


Figure 6. Variable importance ranking for GX dataset 2 using four machine learning models (CatBoost, RF, XGBoost, and LightGBM) under spatial cross-validation strategies.

In JL dataset 2, there were five variables (B4_red, MCARI2, RGR, B2_blue, and B2_Mea_5) that were listed as the top three variables in the four machine learning models.

These variables belong to two spectral variables (B4_red and B2_blue), one vegetation index (RGR), one red-edge vegetation index (MCARI2), and one texture measure (B2_Mea_5). It has been widely demonstrated that forest AGB estimation was affected by multiple features rather than a single feature [48–50]. In our study of the JL region, we found that the spectral variable B4-red was the most important factor in the CatBoost and XGBoost models, while the vegetation variable RGR was the most needed predictor in the RF and LightGBM models. The red-edge variable MCARI2 was also an important variable, indicating that the red-edge band index has excellent potential in forest AGB estimation in our study regions.

Figure 6 shows the relative importance of variables for the four models in GX dataset 2. It was found that the vegetation variable RVI54 was the most essential factor, ranking first in the CatBoost and RF models and in the top three in the XGBoost and LightGBM models. Meanwhile, the red edge vegetation index TCARI and vegetation index MCARI were also found to be important to forest AGB models in the GX region. The relative importance of TCARI in our forest AGB estimation in the GX region confirmed the superiority of the red-edge band vegetation indices in forest AGB modeling. This may be attributed to the fact that the spectral reflectance in the red edge range is highly sensitive to the vegetation growth quality, thus affecting the biomass of the forest ecosystem over the ground surface. Our study further confirms this well-known phenomenon in the remote sensing of territory ecosystems, which stated that forest AGB estimates were remarkably influenced by multiple remote sensing variables, and red-edge band vegetation indices have a significant contribution to improving the accuracy of forest AGB prediction.

3.4. Scenarios for Comparison of the Individual Models and the Hybrid Model

With the above results, eight scenarios (Table 4) were constructed on the basis of two study areas (JL and GX), two variable types (spectral and textural variables), and two validation strategies (random cross-validation and spatial cross-validation) to compare the performance of the four individual models and the hybrid model. Table 5 shows the general capabilities of these five models (four individual models and one hybrid model) for forest AGB prediction over the four scenarios of the JL region. The results of these models over the four scenarios of the GX region are presented in Table 6. More detailed results of this comparison are given in the Supplementary Materials. Specifically, the relative weights of the four individual models in the hybrid model for forest AGB estimation are given in Table S2 of the Supplementary Materials, while the key tuning hyper-parameters are given in Table S3.

Table 4. Description of the eight scenarios used to compare the four individual models and the one hybrid model for forest AGB estimation in our two study regions.

Scenario	Dataset	Validation Strategy	Variable Type
Scenario 1	JL Dataset 1	Random cross-validation	Spectral variables
Scenario 2	JL Dataset 1	Spatial cross-validation	Spectral variables
Scenario 3	JL Dataset 2	Random cross-validation	Spectral and textural variables
Scenario 4	JL Dataset 2	Spatial cross-validation	Spectral and textural variables
Scenario 5	GX Dataset 1	Random cross-validation	Spectral variables
Scenario 6	GX Dataset 1	Spatial cross-validation	Spectral variables
Scenario 7	GX Dataset 2	Random cross-validation	Spectral and textural variables
Scenario 8	GX Dataset 2	Spatial cross-validation	Spectral and textural variables

Table 5. Forest AGB estimation accuracy assessment on the five selected models in the JL region.

Scenario	Model	R ²	RMSE (Mg ha ⁻¹)	Bias (Mg ha ⁻¹)	Relative RMSE (%)
Scenario 1	CatBoost	0.62	21.48	−0.08	13.43
	LightGBM	0.61	21.80	0.04	13.63
	RF	0.60	21.83	−0.31	13.64
	XGBoost	0.60	21.91	−0.03	13.70
	CLR X	0.63	21.40	−0.09	13.39

Table 5. Cont.

Scenario	Model	R^2	RMSE (Mg ha ⁻¹)	Bias (Mg ha ⁻¹)	Relative RMSE (%)
Scenario 2	CatBoost	0.51	22.38	−0.75	14.40
	LightGBM	0.49	22.78	−0.33	14.69
	RF	0.49	22.96	−0.93	14.78
	XGBoost	0.49	22.78	−0.64	14.66
	CLR X	0.59	22.21	−0.10	13.92
Scenario 3	CatBoost	0.66	20.29	−0.12	12.69
	LightGBM	0.64	20.72	−0.03	12.96
	RF	0.66	20.36	−0.17	12.73
	XGBoost	0.65	20.55	−0.08	12.85
	CLR X	0.67	20.21	−0.10	12.63
Scenario 4	CatBoost	0.54	21.31	−0.83	13.67
	LightGBM	0.53	21.51	−0.41	13.83
	RF	0.54	21.32	−0.77	13.68
	XGBoost	0.52	21.89	−1.1	14.05
	CLR X	0.63	21.16	−0.36	13.23

Table 6. Forest AGB estimation accuracy assessment on the five selected models in the GX region.

Scenario	Model	R^2	RMSE (Mg ha ⁻¹)	Bias (Mg ha ⁻¹)	Relative RMSE (%)
Scenario 5	CatBoost	0.58	17.18	0.007	33.32
	LightGBM	0.59	17.41	−0.008	32.79
	RF	0.59	17.14	0.16	32.29
	XGBoost	0.58	17.41	0.32	32.75
	CLR X	0.60	16.80	0.12	31.61
Scenario 6	CatBoost	0.51	17.43	0.017	34.52
	LightGBM	0.46	18.20	1.42	36.08
	RF	0.51	17.42	0.13	34.46
	XGBoost	0.47	18.38	−0.06	36.60
	CLR X	0.55	17.28	−0.24	33.46
Scenario 7	CatBoost	0.64	15.89	0.05	29.86
	LightGBM	0.64	15.88	0.24	29.84
	RF	0.65	15.62	0.11	29.42
	XGBoost	0.60	16.77	0.32	31.52
	CLR X	0.66	15.59	0.18	29.34
Scenario 8	CatBoost	0.59	15.80	0.16	31.10
	LightGBM	0.54	16.60	0.22	32.87
	RF	0.58	15.95	−0.14	31.22
	XGBoost	0.55	16.51	−0.09	32.29
	CLR X	0.62	15.34	−0.18	30.72

3.5. Performance of the Individual Models and the Hybrid Model in JL Region

Table 5 compares the performance of the five models (four individual models and one hybrid model) for forest AGB estimation with the four scenarios in the JL region. It can be seen that all the five models have better performances in datasets 2 than in datasets 1, which implies that the datasets containing both spectral and textural variables are more suitable for forest AGB estimation than those with only spectral variables. In other words, forest AGB estimation with more variables will produce a more accurate forest AGB estimation than those only with a few spectral variables. Specifically, scenario 4, i.e., JL dataset 2 with the hybrid model CLR X under spatial cross-validation, has $R^2 = 0.63$, which is increased by 6.78% in comparison to that of scenario 2 ($R^2 = 0.59$), which is the JL dataset 1 in the JL region. Scenario 3, which is composed of dataset 2 with the hybrid model CLR X under

random cross-validation in the JL region, has an even greater determination coefficient ($R^2 = 0.67$) than scenario 4. Scenarios 1 and 2 are composed of JL dataset 1 under random and spatial cross-validation, and they only have determination coefficients of $R^2 = 0.63$ and $R^2 = 0.59$ for the hybrid model CLRX. Another feature of Table 5 for the JL region is that the hybrid model has the highest R^2 and lowest RMSE among the five models in comparison to all the scenarios. This surely proves that the hybrid model has the capability to combine the advantages of the four individual models constituting the hybrid model. Our finding aligns with Ploton et al. [48], which confirms the high importance of texture features for improving the precision of forest AGB estimation.

3.6. Performance of the Individual Models and the Hybrid Model in GX Region

Table 6 shows the performance of the five selected models for forest AGB estimation with the four scenarios in the GX region. Similar results can be seen in the GX region. Generally speaking, all five models have better performance in both scenarios 7 and 8 than that in scenarios 5 and 6. This is because scenarios 7 and 8 are composed of both spectral variables and textural variables, while scenarios 5 and 6 only involve the spectral variables. Therefore, it can be said that more variables are better to increase the accuracy of forest AGB estimation with the selected machine learning models in this study. Among the five selected models, the hybrid model CLRX has the best performance in all scenarios. For example, for scenario 7, the R^2 of the hybrid model CLRX is 0.66, while it is 0.64, 0.64, 0.65, and 0.60 for the individual models CatBoost, LightGBM, RF, and XGBoost, respectively. The RMSE of the hybrid model is also the lowest among the five models in all the scenarios. For example, the hybrid model CLRX has an RMSE of 15.34 Mg ha⁻¹ for scenario 8, which is composed of both spectral and textural variables under a spatial cross-validation strategy, while the RMSEs of the other four individual models are 15.80, 16.60, 15.95, and 16.51 Mg ha⁻¹ for this scenario. The comparison of the four individual models indicates that the model RF has the lowest RMSE in both scenarios 7 and 8, implying that RF would be the best individual machine learning model for forest AGB estimation in GX region. However, we also found that the R^2 of all the models is not very high, and the greatest R^2 is only 0.66 for the hybrid model in scenario 7. Since R^2 is the determination coefficient indicating the percentage of the dependent variable explained by the selected independent variables, the relatively low R^2 in the study also implies that forest AGB is determined by a number of variables, as only about two-thirds of forest AGB variation in the GX region can be predicted by the selected variables involved in the scenarios.

3.7. Comparison of the Two Cross-Validation Strategies

The forest AGB estimation accuracies of the five selected models also need to be compared under spatial cross-validation and random cross-validation. The reason for this comparison is that training and validation datasets through spatial cross-validation are not mixed in the same space. Figure 7 shows the spatial visualization of the selected validation and training observations for the random 5-fold cross-validation and spatial 5-fold cross-validation of one repetition in the JL region. As shown in Table 5 for the JL region, R^2 for individual machine learning models under spatial cross-validation has a decrease within the range of 17.74%~20.00%, while the decrease range of this coefficient is within 7.81%~22.03% in the GX region (Table 6), which is expected due to the SAC in the observations. As for the hybrid model CLRX, the coefficient R^2 was reduced by 6.35% in JL dataset 1, which is higher than that in JL dataset 2 (5.97%). A similar decrease in R^2 is observed for the hybrid model in the GX region, which shows that the decrease in R^2 was 8.33% in GX dataset 1 and 6.06% in GX dataset 2. Since dataset 1 is with random cross-validation while dataset 2 is under spatial cross-validation, it can be said that random cross-validation has better performance than spatial cross validation for the models in forest AGB estimation.

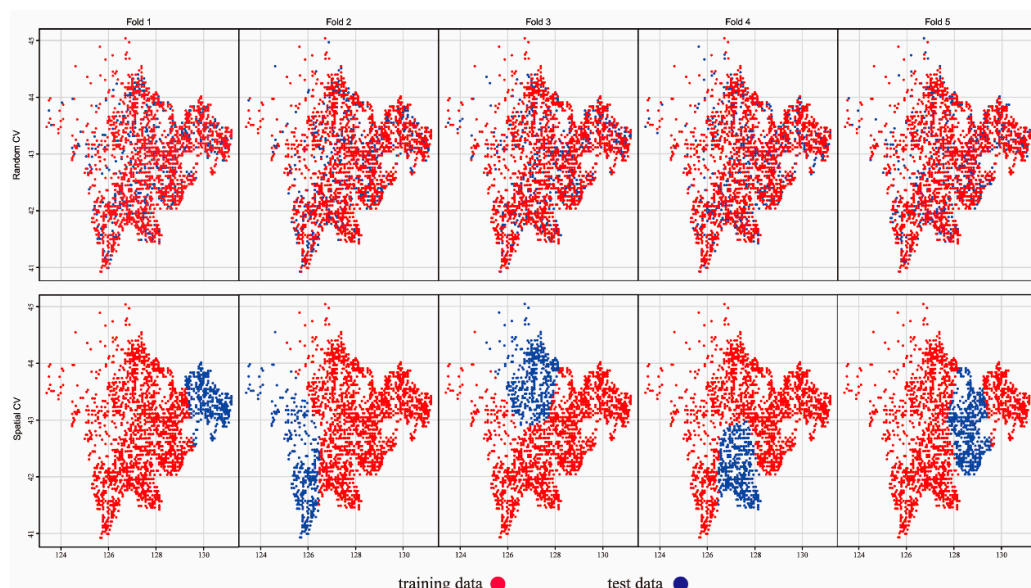


Figure 7. Spatial visualization of the selected validation and training observations for cross-validation of one repetition: random cross-validation (**upper row**) and spatial cross-validation (**lower row**).

Figures 8 and 9 present a detailed comparison of the five models for forest AGB estimation under two cross-validation strategies in JL and GX regions, respectively. In the JL region, as shown in Figure 8, though the general application of the models for forest AGB estimation was observed, the strategy of random cross-validation is commonly better than the spatial cross-validation strategy for all the models. However, this feature is not observed in the GX region, where similar accuracy can be seen for all models in both the random and spatial cross-validation strategies (Figure 9). Moreover, the performance of the models is much better in the JL region than in the GX region. This may be attributed to the difference in forest types in the two regions. Another important feature is that the overestimation at low values and underestimation at high values can be found for all the models in the two regions. In the JL region, as indicated in Figure 8, all models underestimate the forest AGB for values $> 200 \text{ Mg ha}^{-1}$ and overestimate the forest AGB for values $< 50 \text{ Mg ha}^{-1}$. In the GX region, an overestimation of forest AGB for the values $< 20 \text{ Mg ha}^{-1}$ is obviously found (Figure 9). This overestimation at low value and underestimation at high value may be a trend.

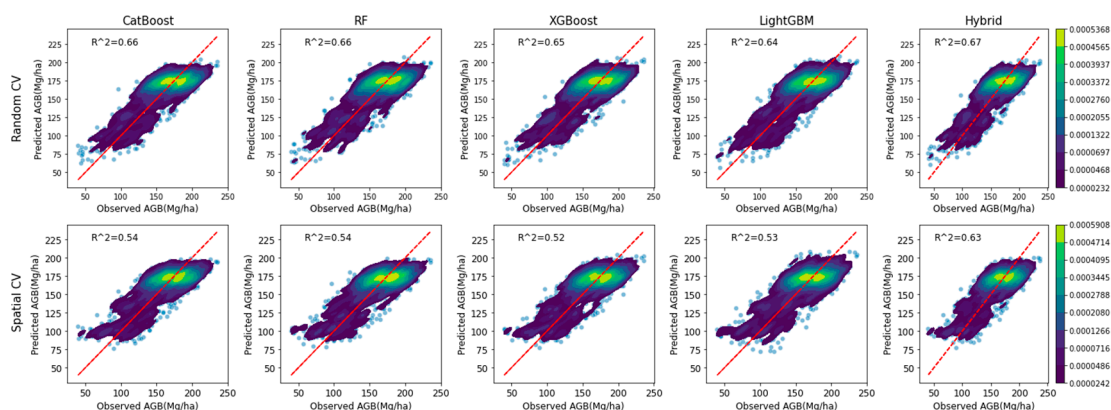


Figure 8. Scatter plot of the predicted and observed forest AGB using the selected models under the random and spatial cross-validation (CV) strategies for the JL region.

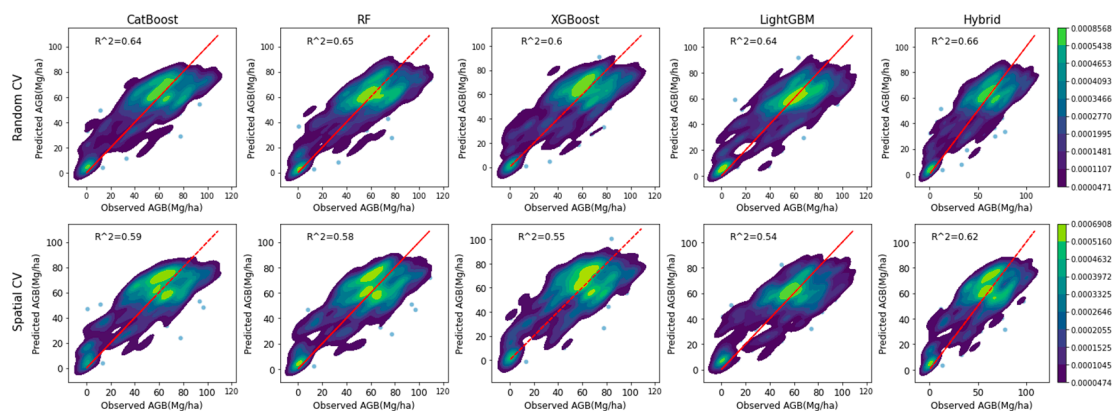


Figure 9. Scatter plot of the predicted and observed forest AGB using the selected models under the random and spatial cross-validation (CV) strategies for the GX region.

3.8. Comparison of the Individual Models with the Hybrid Model for Forest AGB Estimation

The results in Tables 5 and 6 indicate that among the four individual models, the RF model performs best in scenarios 5, 6, and 7, while the CatBoost model performs best in the remaining scenarios. The CatBoost model yielded the highest predictive performance in the JL region, with $R^2 = 0.66$ and $RMSE = 20.29 \text{ Mg ha}^{-1}$ in random cross-validation and $R^2 = 0.54$ and $RMSE = 21.31 \text{ Mg ha}^{-1}$ in spatial cross-validation. The RF model showed the best predicted performance in the GX region with $R^2 = 0.65$ and $RMSE = 15.62 \text{ Mg ha}^{-1}$ in random cross-validation. However, in spatial cross-validation, the CatBoost model ($R^2 = 0.59$, $RMSE = 15.80 \text{ Mg ha}^{-1}$) was slightly better than the RF model ($R^2 = 0.58$, $RMSE = 15.95 \text{ Mg ha}^{-1}$), followed by the XGBoost and LightGBM models. Overall, no single prediction model can always maintain excellent performance in the scenarios of the study.

It can clearly be seen in Tables 5 and 6 that, in comparison with the four individual models, the hybrid model resulted in more accurate predictions of forest AGB in our two study regions, with the highest R^2 and lowest RMSE in all the scenarios. For example, in scenario 4 (Table 5), the R^2 of the hybrid model CLRX has the ability to improve the forest AGB estimation by 16.67% in comparison with the optimal individual model (CatBoost) and by 21.15% in comparison with the worst individual model (XGBoost). One interesting feature was that the improvement in the hybrid model for forest AGB estimation under the spatial cross-validation strategy was more evident than that of the random cross-validation strategy. In comparison with the CatBoost model, the R^2 of the hybrid model is slightly improved by 1.61% in scenario 1, while it is increased by 15.68% in scenario 2. Overall, forest AGB estimation with the hybrid model in our study regions may result in a higher accuracy than with individual models. This finding can provide some useful insights into model selection for forest AGB estimation.

We used the hybrid model under the spatial cross-validation strategy to generate forest AGB maps for 2018 at 100 m resolution in the JL and GX regions (Figure 10). The results showed that high AGB values tended to be found in regions covered by dense forest. In the JL region, the average value of AGB was 140.27 Mg/ha , with a minimum of 54.31 Mg/ha and a maximum of 190.25 Mg/ha , indicating the presence of dense, carbon-rich forests in this area. The considerable range in AGB values highlights the spatial heterogeneity across the JL region. In the GX region, the average value of AGB was 55.58 Mg/ha . This difference in mean AGB between the two regions may be attributed to factors such as climate, forest type, and management practices.

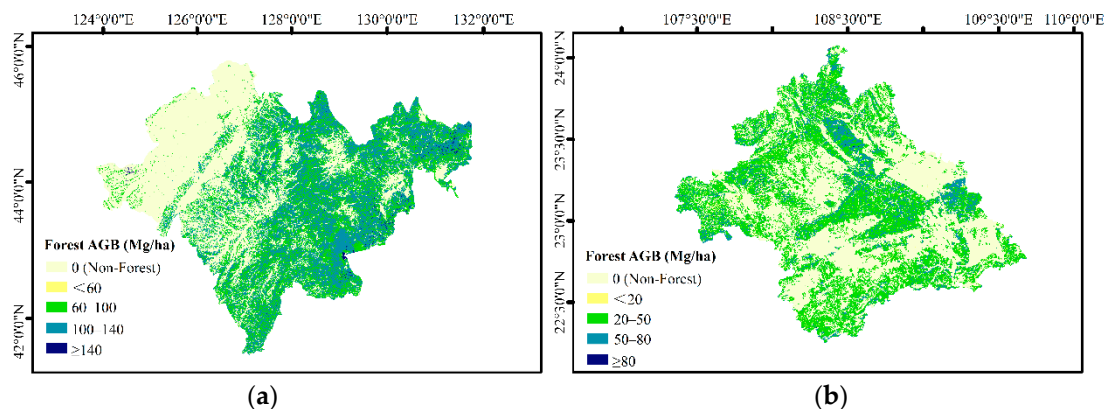


Figure 10. Forest AGB maps for 2018 at 100 m spatial resolution generated: (a) study area in JL; (b) study area in GX.

4. Discussion

In this study, the improvement in the forest AGB prediction accuracy was investigated from three dimensions: model input (feature extraction and selection), modeling algorithm (individual machine learning models and the hybrid model CLRX), and model performance (spatial cross-validation and random cross-validation).

With the availability and accessibility of multiple-source remote sensing data, the applications of machine learning algorithms in many practical issues including forest AGB estimation have experienced dramatic development in recent decades [2,10,50]. However, these applications have mainly been based on traditional machine learning models such as ANN and RF. With the development of computational sciences, several new ensemble learning algorithms such as CatBoost, XGBoost, and LightGBM have been developed and demonstrated to have high potential for practical applications in issues in earth sciences such as climate change modelling, AGB estimation, land use and land cover mapping, and crop yield prediction [8,13,22]. However, the applicability of these new ensemble learning algorithms for forest AGB estimation have not been reported. In this study, we compare the three new ensemble learning algorithms (CatBoost, XGBoost, and LightGBM) with the widely used machine learning model RF to determine their applicability for forest AGB prediction in two typical Chinese forest regions: the JL region in northeastern China and the GX region in southern China. On the basis of this comparison, we developed a hybrid model to combine the advantages of these four individual models and compare the performance of this hybrid model with the four individual models for forest AGB in the two study regions.

To obtain a complete comparison of the selected four individual models and the hybrid model for forest AGB estimation, we established eight scenarios from the available field datasets according to the two study regions, the selected variables, and the cross-validation strategies. The results showed that CatBoost was superior to other individual models except in scenarios 5, 6, and 7. RF was slightly better than CatBoost, which might be due to the smaller sample number and feature number in scenarios 5, 6, and 7. CatBoost showed more obvious advantages in the datasets with large samples and many features. Thus, it could be said that CatBoost has great potential to improve forest AGB estimates' accuracy in the two study regions. This finding is consistent with those presented in Zhang et al. [49] and Zhao et al. [24]. CatBoost solves the problems leading to gradient bias and prediction shift in modeling results and hence has the ability to improve the accuracy of our forest AGB estimation. Moreover, the model can also be easy to generalize, hence making it suitable for applications to many situations, especially in enhancing parameter estimation accuracy for forest AGB modeling.

On the basis of the above four individual models, we developed in this study the hybrid model CLRX for forest AGB prediction. Validation with the datasets in the two study regions indicates that the hybrid model can result in a better performance than any indi-

vidual model. The hybrid model CLRX is mainly affected by the choice of models and the combined weighting method. RF has an advantage in anti-noise but is insensitive to outliers. The model XGBoost can automatically use CPU multithreading for parallel computing to improve the modelling accuracy. The model CatBoost can avoid over-fitting by adding additional procedures to compute the leaf nodes when selecting the tree structures in forest AGB estimation. The model LightGBM has apparent advantages in computational speed and memory consumption. The better performance of our hybrid model on the basis of the four individual models confirms the direction of improving forest AGB estimation accuracy through the development of a hybrid model, which intends to extend the application of different machine learning algorithm combinations (such as kernel-based learners, tree-based models, and linear-based learners) or the coupling of physical knowledge with machine learning algorithms for the hybrid model. The investigation into innovative combination techniques and fusion approaches presents substantial prospects for optimizing the performance of hybrid models, facilitating their proficient utilization of diverse algorithms and strategies. Additionally, the incorporation of novel variables, such as topography, and the integration of emerging sensor technologies contribute significantly to the progress and practicality of hybrid models.

Two strategies of cross-validation were applied in our study, and we found that the spatial cross-validation was less optimistic than the random one. Our finding in the study is consistent with previous investigations into forest fires [31] showing that random cross-validation yielded higher performance than spatial cross-validation. The RMSEs of the four machine learning models increased more apparently in the GX region than in the JL region. This probably is due to the higher SAC in the GX region than in the JL region. To the best of our knowledge, this is the first time that SAC has been considered in forest AGB estimation. The lower but more realistic model performance was presented when taking spatial effects into account [50]. This is especially problematic when there is strong SAC in the sampling dataset and when training samples are clustered in space. It has been commonly understood that over-optimistic forest AGB estimation may provide wrong information when formulating guidance or policies in ecological decision-making processes relating to terrestrial ecosystems [10,18]. However, in order to solve the problem of SAC, we still recommend the general consideration of spatial cross-validation in assessing forest AGB models. It is worth mentioning that the performance improvement in the hybrid model under the spatial cross-validation strategy was more prominent than the random one. Therefore, it is highly recommended to exploit the hybrid model combined with the spatial cross-validation strategy, which may represent an interesting perspective for future research on more realistic and accurate AGB estimation.

Though obvious improvements have been made in our hybrid model for forest AGB estimation, the problem of overestimation at low values and underestimation at high values still remains, as is reported in previous studies [51]. All the algorithms underestimated the forest AGB observed values when the AGB values were greater than $\sim 200 \text{ Mg ha}^{-1}$ (Figure 8). As is well known, the factors affecting forest AGB estimation with a remote sensing approach are complex [3,46,50]. The following three aspects are generally considered for the estimation: (1) remote sensing data source and its spatial resolution; (2) differences in the forest ecosystem itself and topographic changes in its distribution area; and (3) prediction algorithms. This study mainly investigated the issues relating to the third aspect, while those related to the first and second aspects were not fully explored. On one hand, the pixel point saturation of vegetation indices exists in sampling blocks with high forest AGB values in Sentinel-2 images [52]. On the other hand, spectral variables, vegetation indices, red-edge band vegetation indices, and texture measures are considered as the variables for the estimation of forest AGB in this study.

Our results in the study indicate that the determination coefficient R^2 is not very high. The best R^2 is for the hybrid model was in scenario 3, with $R^2 = 0.67$ (Table 5), implying that the selected variables can only explain about two-thirds of the forest AGB variance in this scenario, leaving about one-third unable to be predicted by the variables. This

surely is due to the fact that many factors shape forest AGB, and our study to conduct this estimation using a remote sensing approach only involved a few factors that could be extracted from the selected remote sensing images. As is well known, many potentially important variables affecting forest AGB prediction, such as canopy coverage, vegetation type, and topography, may need to be included into the estimation [51]. The reflectance of spectral bands is affected by ground vegetation and soil for plots with low AGB values [53]. Therefore, it is necessary to continue the exploration of the methodology for an accurate estimation of forest AGB. One direction for this is the extension of combining multi-source remote sensing images with a hybrid model under a spatial cross-validation strategy.

5. Conclusions

The improvement of forest AGB estimation was investigated in the study so that an effective and adaptive framework could be developed for this estimation, including variable extraction, variable selection, model development, and model evaluation. The spectral variables, vegetation indices, red-edge vegetation indices, and texture measures were extracted from the Sentinel-2 remote sensing images, and VSURF was used for feature selection. The forecasts from four individual machine learning models (RF, XGBoost, LightGBM, and CatBoost) were combined to create the hybrid model CLRX. The influence of SAC on model evaluation was considered by comparing random cross-validation and spatial cross-validation. The results showed that VSURF could provide an effective approach for the selection of appropriate variables from a larger set of variables to participate in model construction for forest AGB estimation. No single prediction model could always maintain excellent performance in every situation. The hybrid model CLRX could overcome this deficiency and resulted in the most accurate predictions in comparison with the four individual machine learning models. Both individual and hybrid models had relatively lower performances under spatial cross-validation than under random cross-validation. This is probably due to SAC in the dataset. It should be mentioned that the hybrid model CLRX has greater potential to improve the accuracy of AGB estimation under spatial cross-validation than under random cross-validation. This study contributes to investigating the potential use of the hybrid machine learning method to estimate forest AGB. Meanwhile, we underline the necessity of spatial cross-validation in forest AGB estimation.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/f15060975/s1>, Table S1: The wood density of the tree species or groups used in the study; Table S2: Weights of the forest AGB models on the basis of entropy weighting method. Table S3: Optimal hyper-parameter with spatial 5-fold cross-validation for the individual models.

Author Contributions: Conceptualization, M.L. and Q.H.; data curation, S.A.A. and L.Z.; formal analysis, X.Q.; investigation, M.L., J.F. and G.H.; methodology, M.L. and S.A.A.; project administration, Q.H.; resources, X.Q., Z.Q. and J.F.; software, H.Z.M.S.; supervision, X.Q., J.F. and H.Z.M.S.; validation, Q.H., Z.Q. and G.H.; writing—original draft, M.L.; writing—review and editing, M.L., S.A.A. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Base and Talent Project of Guangxi (grant no. Guike-AD23026073), Guangxi Young and Middle-aged University Teachers' Scientific Research Ability Enhancement Project (Grant No.2023KY0399), Ecosystem Soil and Water Conservation Function Assessment Project in Beibu Gulf, Guangxi Province (Grant No.84-Y50-G29-9001-22/23), and the MNR-CN Key Laboratory of China-ASEAN Satellite Remote Sensing Applications (Grant No. ZDMY202310).

Data Availability Statement: Not applicable. The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

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

References

- Fang, J.Y.; Wang, Z.M. Forest biomass estimation at regional and global levels, with special reference to China's forest biomass. *Ecol. Res.* **2001**, *16*, 587–592. [\[CrossRef\]](#)
- Jacon, A.D.; Galvao, L.S.; Dalagnol, R.; dos Santos, J.R. Aboveground biomass estimates over Brazilian savannas using hyperspectral metrics and machine learning models: Experiences with Hyperion/EO-1. *GISci. Remote Sens.* **2021**, *58*, 1112–1129. [\[CrossRef\]](#)
- Oehmcke, S.; Li, L.; Trepekli, K.; Revenga, J.C.; Nord-Larsen, T.; Gieseke, F.; Igel, C. Deep point cloud regression for above-ground forest biomass estimation from airborne LiDAR. *Remote Sens. Environ.* **2024**, *302*, 113968. [\[CrossRef\]](#)
- Mohite, J.; Sawant, S.; Pandit, A.; Sakkan, M.; Pappula, S.; Parmar, A. Forest aboveground biomass estimation by GEDI and multi-source EO data fusion over Indian forest. *Int. J. Remote Sens.* **2024**, *45*, 1304–1338. [\[CrossRef\]](#)
- Puliti, S.; Hauglin, M.; Breidenbach, J.; Montesano, P.; Neigh, C.S.R.; Rahlf, J.; Solberg, S.; Klingenberg, T.F.; Astrup, R. Modelling above-ground biomass stock over Norway using national forest inventory data with ArcticDEM and Sentinel-2 data. *Remote Sens. Environ.* **2020**, *236*, 111501. [\[CrossRef\]](#)
- Domingues, G.F.; Soares, V.P.; Leite, H.G.; Ferraz, A.S.; Ribeiro, C.; Lorenzon, A.S.; Marcatti, G.E.; Teixeira, T.R.; de Castro, N.L.M.; Mota, P.H.S.; et al. Artificial neural networks on integrated multispectral and SAR data for high-performance prediction of eucalyptus biomass. *Comput. Electron. Agric.* **2020**, *168*, 105089. [\[CrossRef\]](#)
- Chen, Z.; Jia, K.; Xiao, C.; Wei, D.; Wang, L. Leaf area index estimation algorithm for GF-5 hyperspectral data based on different feature selection and machine learning methods. *Remote Sens.* **2020**, *12*, 2110. [\[CrossRef\]](#)
- Luo, M.; Wang, Y.F.; Xie, Y.H.; Zhou, L.; Qiao, J.J.; Qiu, S.Y.; Sun, Y.J. Combination of feature selection and catBoost for prediction: The first application to the estimation of aboveground biomass. *Forests* **2021**, *12*, 216. [\[CrossRef\]](#)
- Hu, Y.F.; Nie, Y.H.; Liu, Z.H.; Wu, G.M.; Fan, W.Y. Improving the Potential of Coniferous Forest Aboveground Biomass Estimation by Integrating C- and L-Band SAR Data with Feature Selection and Non-Parametric Model. *Remote Sens.* **2023**, *15*, 4194. [\[CrossRef\]](#)
- de Almeida, C.T.; Galvao, L.S.; Aragao, L.; Ometto, J.; Jacon, A.D.; Pereira, F.R.D.; Sato, L.Y.; Lopes, A.P.; Graca, P.; Silva, C.V.D.; et al. Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian Amazon using different regression algorithms. *Remote Sens. Environ.* **2019**, *232*, 111323. [\[CrossRef\]](#)
- Ghosh, S.M.; Behera, M.D.; Jagadish, B.; Das, A.K.; Mishra, D.R. A novel approach for estimation of aboveground biomass of a carbon-rich mangrove site in India. *J. Environ. Manag.* **2021**, *292*, 13. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jiang, F.G.; Kutia, M.; Ma, K.S.; Chen, S.; Long, J.P.; Sun, H. Estimating the aboveground biomass of coniferous forest in Northeast China using spectral variables, land surface temperature and soil moisture. *Sci. Total Environ.* **2021**, *785*, 15. [\[CrossRef\]](#) [\[PubMed\]](#)
- Su, H.Y.; Shen, W.J.; Wang, J.R.; Ali, A.; Li, M.S. Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. *For. Ecosyst.* **2020**, *7*, 64. [\[CrossRef\]](#)
- Pelletier, F.; Cardille, J.A.; Wulder, M.A.; White, J.C.; Hermosilla, T. Inter- and intra-year forest change detection and monitoring of aboveground biomass dynamics using Sentinel-2 and Landsat. *Remote Sens. Environ.* **2024**, *301*, 113931. [\[CrossRef\]](#)
- Abbaszadeh, P.; Moradkhani, H.; Zhan, X.W. Downscaling SMAP Radiometer Soil Moisture Over the CONUS Using an Ensemble Learning Method. *Water Resour. Res.* **2019**, *55*, 324–344. [\[CrossRef\]](#)
- Meng, B.P.; Liang, T.G.; Yi, S.H.; Yin, J.P.; Cui, X.; Ge, J.; Hou, M.J.; Lv, Y.Y.; Sun, Y. Modeling Alpine Grassland Above Ground Biomass Based on Remote Sensing Data and Machine Learning Algorithm: A Case Study in East of the Tibetan Plateau, China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2986–2995. [\[CrossRef\]](#)
- Zimbres, B.; Rodriguez-Veiga, P.; Shimbo, J.Z.; Bispo, P.D.; Balzter, H.; Bustamante, M.; Roitman, I.; Haidar, R.; Miranda, S.; Gomes, L.; et al. Mapping the stock and spatial distribution of aboveground woody biomass in the native vegetation of the Brazilian Cerrado biome. *For. Ecol. Manag.* **2021**, *499*, 15. [\[CrossRef\]](#)
- Freeman, E.A.; Moisen, G.G.; Coulston, J.W.; Wilson, B.T. Random forests and stochastic gradient boosting for predicting tree canopy cover: Comparing tuning processes and model performance. *Can. J. For. Res.* **2016**, *46*, 323–339. [\[CrossRef\]](#)
- Lou, X.W.; Weng, Y.H.; Fang, L.M.; Gao, H.L.; Grogan, J.; Hung, I.K.; Oswald, B.P. Predicting stand attributes of loblolly pine in West Gulf Coastal Plain using gradient boosting and random forests. *Can. J. For. Res.* **2021**, *51*, 807–816. [\[CrossRef\]](#)
- Li, W.; Yin, Y.B.; Quan, X.W.; Zhang, H. Gene Expression Value Prediction Based on XGBoost Algorithm. *Front. Genet.* **2019**, *10*, 484931. [\[CrossRef\]](#)
- Fan, J.; Wang, X.; Zhang, F.; Ma, X.; Wu, L. Predicting daily diffuse horizontal solar radiation in various climatic regions of China using support vector machine and tree-based soft computing models with local and extrinsic climatic data. *J. Clean. Prod.* **2020**, *248*, 119264. [\[CrossRef\]](#)
- Suwanlee, S.R.; Pinasu, D.; Som-ard, J.; Borgogno-Mondino, E.; Sarvia, F. Estimating Sugarcane Aboveground Biomass and Carbon Stock Using the Combined Time Series of Sentinel Data with Machine Learning Algorithms. *Remote Sens.* **2024**, *16*, 750. [\[CrossRef\]](#)
- Mallick, J.; Talukdar, S.; Shahfahad, Pal, S.; Rahman, A. A novel classifier for improving wetland mapping by integrating image fusion techniques and ensemble machine learning classifiers. *Ecol. Inform.* **2021**, *65*, 101426. [\[CrossRef\]](#)

24. Hu, S.; Xiang, Y.; Zhang, H.C.; Xie, S.Y.; Li, J.H.; Gu, C.H.; Sun, W.; Liu, J.Y. Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction. *Appl. Energy* **2021**, *293*, 116951. [[CrossRef](#)]
25. Ahmed, A.A.M.; Deo, R.C.; Feng, Q.; Ghahramani, A.; Raj, N.; Yin, Z.L.; Yang, L.S. Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. *J. Hydrol.* **2021**, *599*, 126350. [[CrossRef](#)]
26. Yu, J.X.; Zhang, X.; Xu, L.L.; Dong, J.; Zhangzhong, L.L. A hybrid CNN-GRU model for predicting soil moisture in maize root zone. *Agric. Water Manag.* **2021**, *245*, 106649. [[CrossRef](#)]
27. Matinfar, H.R.; Maghsodi, Z.; Mousavi, S.R.; Rahmani, A. Evaluation and prediction of topsoil organic carbon using machine learning and hybrid models at a field-scale. *Catena* **2021**, *202*, 105258. [[CrossRef](#)]
28. Zhao, W.L.; Gentine, P.; Reichstein, M.; Zhang, Y.; Zhou, S.; Wen, Y.Q.; Lin, C.J.; Li, X.; Qiu, G.Y. Physics-constrained machine learning of evapotranspiration. *Geophys. Res. Lett.* **2019**, *46*, 14496–14507. [[CrossRef](#)]
29. Scudiero, E.; Skaggs, T.H.; Corwin, D.L. Regional-scale soil salinity assessment using Landsat ETM plus canopy reflectance. *Remote Sens. Environ.* **2015**, *169*, 335–343. [[CrossRef](#)]
30. Schratz, P.; Muenchow, J.; Iturrirxa, E.; Richter, J.; Brenning, A. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecol. Model.* **2019**, *406*, 109–120. [[CrossRef](#)]
31. Mayr, M.J.; Vanselow, K.A.; Samimi, C. Fire regimes at the arid fringe: A 16-year remote sensing perspective (2000–2016) on the controls of fire activity in Namibia from spatial predictive models. *Ecol. Indic.* **2018**, *91*, 324–337. [[CrossRef](#)]
32. Meyer, H.; Reudenbach, C.; Wollauer, S.; Nauss, T. Importance of spatial predictor variable selection in machine learning applications—Moving from data reproduction to spatial prediction. *Ecol. Model.* **2019**, *411*, 108815. [[CrossRef](#)]
33. Zeng, W.S. Developing one-variable individual tree biomass models based on wood density for 34 tree species in China. *For. Resour. Manag.* **2018**, *6*, 41–46. [[CrossRef](#)]
34. Li, Y.; Li, C.; Li, M.; Liu, Z. Influence of variable selection and forest type on forest aboveground biomass estimation using machine learning algorithms. *Forests* **2019**, *10*, 1073. [[CrossRef](#)]
35. Korhonen, L.; Hadi; Packalen, P.; Rautiainen, M. Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sens. Environ.* **2017**, *195*, 259–274. [[CrossRef](#)]
36. Genuer, R.; Poggi, J.M.; Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **2010**, *31*, 2225–2236. [[CrossRef](#)]
37. Genuer, R.; Poggi, J.M.; Tuleau-Malot, C. VSURF: An R package for variable selection using random forests. *R J.* **2016**, *7*, 19–33. [[CrossRef](#)]
38. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013.
39. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363. [[CrossRef](#)]
40. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2018; Volume 11, pp. 6639–6649.
41. Ke, G.L.; Meng, Q.; Finley, T.; Wang, T.F.; Chen, W.; Ma, W.D.; Ye, Q.W.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the Thirty-First Conference on Neural Information Processing System*, Long Beach, CA, USA, 4 December 2017.
42. Zhou, K.B.; Hu, Y.X.; Pan, H.; Kong, L.; Liu, J.; Huang, Z.; Chen, T. Fast prediction of reservoir permeability based on embedded feature selection and LightGBM using direct logging data. *Meas. Sci. Technol.* **2020**, *31*, ab4a45. [[CrossRef](#)]
43. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
44. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. *Knowl. Discov. Data Min.* **2016**, *785*, 2939785. [[CrossRef](#)]
45. Ma, M.H.; Zhao, G.; He, B.S.; Li, Q.; Dong, H.Y.; Wang, S.G.; Wang, Z.L. XGBoost-based method for flash flood risk assessment. *J. Hydrol.* **2021**, *598*, 126382. [[CrossRef](#)]
46. Bates, J.M.; Granger, C.W.J. Combination of forecasts. *Oper. Res. Q.* **1969**, *20*, 451–468. [[CrossRef](#)]
47. Liu, D.J.; Li, L. Application study of comprehensive forecasting model based on entropy weighting method on trend of pm2.5 concentration in Guangzhou, China. *Int. J. Environ. Res.* **2015**, *12*, 7085–7099. [[CrossRef](#)] [[PubMed](#)]
48. Ploton, P.; Barbier, N.; Coutron, P.; Antin, C.M.; Ayyappan, N.; Balachandran, N.; Barathan, N.; Bastin, J.F.; Chuyong, G.; Dauby, G.; et al. Toward a general tropical forest biomass prediction model from very high resolution optical satellite images. *Remote Sens. Environ.* **2017**, *200*, 140–153. [[CrossRef](#)]
49. Zhang, Y.; Ma, J.; Liang, S.; Li, X.; Li, M. An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products. *Remote Sens.* **2020**, *12*, 4015. [[CrossRef](#)]
50. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guilleri-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schroder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
51. Su, Y.J.; Guo, Q.H.; Xue, B.L.; Hu, T.Y.; Alvarez, O.; Tao, S.L.; Fang, J.Y. Spatial distribution of forest aboveground biomass in China: Estimation through combination of spaceborne lidar, optical imagery, and forest inventory data. *Remote Sens. Environ.* **2016**, *173*, 187–199. [[CrossRef](#)]

-
52. Forkuor, G.; Zoungrana, J.B.B.; Dimobe, K.; Ouattara, B.; Vadrevu, K.P.; Tondoh, J.E. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets—A case study. *Remote Sens. Environ.* **2020**, *236*, 111496. [[CrossRef](#)]
 53. Astola, H.; Hame, T.; Sirro, L.; Molinier, M.; Kilpi, J. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sens. Environ.* **2019**, *223*, 257–273. [[CrossRef](#)]

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.