Article

Estimation of Forest Stock Volume Using Sentinel-2 MSI, Landsat 8 OLI Imagery and Forest Inventory Data

Yangyang Zhou [ID] and Zhongke Feng *[ID]

Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, China; yangyangzhou1029@163.com
* Correspondence: zhongkefeng@bjfu.edu.cn; Tel.: +86-138-1030-5579

Abstract: Forest stock volume (FSV) is a key indicator for measuring forest quality, evaluating forest management capabilities, and the main factor for evaluating forest carbon sequestration levels. In this study, to achieve an accurate estimation of FSV, we used Ninth Beijing Forest Inventory data (FID), and Landsat 8 OLI and Sentinel-2 MSI imagery to establish FSV models. The performance of Landsat 8 and Sentinel-2 imagery data in estimating forest volume in Huairou District, Beijing, China was compared. The combination of Landsat 8 and Sentinel-2 satellite data was employed to create a new data source. Two variable selection methods, linear stepwise regression (LSR) and recursive feature elimination (RFE), were used to extract feature variables. The multiple linear regression(MLR) models, Back Propagation (BP) neural network models, and Random Forest (RF) models were employed to estimate forest volume in the study area based on the feature variables obtained from both data sources. The research results indicate (1) the Sentinel-2-based model achieved higher accuracy compared to the same model based on the Landsat 8 factor set. The correlation between the red-edge band of Sentinel-2 imagery and FSV is more significant than that of other characteristic variables used. Variables derived from the red-edge band have the potential to reduce model errors; (2) the estimation accuracy of the model can be significantly improved by using the RFE (Recursive Feature Elimination) method to select remote sensing feature variables. RFE is based on the importance ranking of all feature variables and selects the feature variables that contribute the most to the model. In the variable group selected by RFE, the texture features and the derived features from the red-edge band, such as Sen_{B5}, Sen_{RVI}, $Sen_{mNDVIre}$, and Sen_{B5Mean}, contribute the most to the improvement of model accuracy. Furthermore, in the optimal Landsat 8–Sentinel-2 RFE-RF model, where texture features are involved, the rRMSE is greatly reduced by 3.7% compared to the joint remote sensing RFE-RF model without texture features; (3) the MLR, BP, and RF models based on the modeling factor set established on Sentinel-2 have accuracy superior to the model accuracy established based on the modeling factor set of Landsat 8. Among them, the Random Forest (RF) method inverted by the recursive feature elimination (RFE) method using Sentinel-2A image has the best inversion accuracy effect ($R^2 = 0.831$, $RMSE = 12.604$ m^3 ha^{-1}, $rRMSE = 36.411\%$, $MAE = 9.366$ m^3 ha^{-1}). Comparing the performance of the models on the test set, the ranking is as follows, Random Forest (RF) model > Back Propagation (BP) neural network model > multiple linear regression (MLR) model. The feature variable screening based on the Random Forest's recursive feature elimination (RFE) method is better than the linear stepwise regression (LSR). Therefore, the RFE-RF method based on the joint variables from Landsat 8 and Sentinel-2 satellite data to establish a new remote sensing data source provides the possibility to improve the estimation accuracy of FSV and provides reference for forest dynamic monitoring.

Keywords: forest stock volume (FSV); Landsat 8 OLI; Sentinel-2 MSI; forest inventory data (FID); random forest (RF); back propagation (BP) neural network; multiple linear regression (MLR)

1. Introduction

Forest stock volume (FSV, m^3 ha^{-1}) is a key indicator for measuring forest quality, evaluating forest management capability, and assessing carbon sequestration ability. Therefore, effective monitoring of FSV is a crucial task [1]. FSV is measured in cubic meters per mu (1 mu = 0.06667 ha). Traditional methods for determining FSV involve ground surveys, which use measured parameters of individual trees combined with local volume tables to estimate the wood volume of each tree, then extrapolate to the total forest volume [2]. However, this method is time-consuming and inefficient. With the advancement and application of remote sensing technology, optical remote sensing images, combined with ground sample data, are used to establish inversion models for large-scale estimation of FSV, becoming the primary method for FSV estimation. Estimating forest biomass and carbon sequestration heavily relies on forest stock volume (FSV), which serves as a fundamental data source. FSV plays a critical role in reflecting the quality of forest resources and the level of forest management [3–5]. Accurate and scientific forest management, as well as the optimization of forest ecosystems' functionality and carbon sequestration potential, depend on the dynamic estimation of FSV's spatial distribution [6–8].

Although FSV information can be easily obtained from many large-scale global products, its specific regional use is often limited by temporal and spatial resolution constraints or other unknown local errors [9–11]. Due to the strong spatial heterogeneity of forest ecosystems, the response relationship between remote sensing factors and FSV is typically complex [12,13]. There are still many issues to be addressed in the research and application of FSV estimation based on remote sensing technology. Furthermore, Light Detection and Ranging (LiDAR), as an emerging active remote sensing technology, has limited availability of satellite data and high acquisition costs for airborne data, which restricts its widespread application in FSV remote sensing estimation [14,15]. On the other hand, optical remote sensing is a passive remote sensing technology that often provides higher spatial resolution and more detailed surface information. The spatial resolution of optical remote sensing data, such as Landsat 8 and Sentinel-2, can reach below 30 m, while LiDAR, typically, has a spatial resolution ranging from a few meters to a dozen meters [16,17]. Additionally, data from satellites like Landsat 8 and Sentinel-2 are freely available, whereas LiDAR data acquisition and processing costs are higher. The successful applications of these technological tools have paved the way for estimating forest variables, such as FSV, through remote sensing technology. Launched by the European Space Agency (ESA) through its Copernicus program in 2015 and 2017, respectively, Sentinel-2A and Sentinel-2B satellite series offer imagery with a nominal five-day revisit time span worldwide [18].

Sentinel-2 imagery consists of 13 spectral bands that offer spatial resolutions ranging from 10 to 60 m. The European Copernicus program provides free access to Sentinel-2 satellite images, which are provided by the operational environment monitoring system. One of the key applications of the Sentinel-2 satellite series is vegetation analysis [9]. Sentinel-2 images are a preferred source of remote sensing data for forestry research due to their short revisit periods, numerous spectral attributes, and diverse spatial resolution bands. For example, Persson et al. [19] leveraged the Sentinel-2 data to identify common tree species in central Sweden, obtaining an overall accuracy of approximately 88.2% by incorporating all of the imagery bands in their final classification model. A study carried out by Hościło et al. [20] in southern Poland has confirmed that the Sentinel-2 series of images can accurately delineate five tree species, including beech, oak, birch, alder, and larch, with an overall accuracy higher than 85%. Similarly, Pandit et al. [21] investigated the potential of Sentinel-2 images to estimate forest biomass in Nepal. They developed a biomass estimation model that achieved an R^2 value of 0.81 and an *RMSE* value of 25.57 t ha^{-1}. Zarco-Tejada et al. [22] showed that Sentinel-2A data can estimate the chlorophyll content in conifer forests with an open canopy. The R^2 value was higher than 0.7 for June and higher than 0.4 for December. In the Brazilian Amazon, Lima et al. [23] conducted a comparative study to monitor selective logging using Sentinel-2 and Landsat 8 imagery. The effectiveness of Sentinel-2 data in detecting logging concessions exceeds Landsat

8 data. Specifically, the study revealed that 43.2% of logging concessions were detected by Sentinel-2 data, compared to only 35.5% detected by Landsat 8 data. Using the Sentinel-2 time series, Grabska et al. [24] conducted a study on mapping forest stand species in the Carpathian Mountains of Poland. The study noted that a higher accuracy level was achieved, specifically, an improvement between 5% and 10% in overall accuracy compared to relying solely on single-date imagery. The studies discussed above vividly illustrate the enormous potential of the Sentinel-2 for effective forest vegetation monitoring.

In the field of forest stock volume (FSV) estimation, researchers have investigated the use of remote sensing to assess FSV. For example, Condés et al. [25] found that combining satellite images and field data improved the prediction of plot-level growing stock volume, resulting in a significant increase in the adjusted R-squared value from 0.19 to 0.42. Another study by Chrysafis et al. [26] compared the use of Sentinel-2 and Landsat 8 images for estimating FSV using the Random Forest (RF) regression algorithm. The results showed that FSV estimates based on Sentinel-2 images had better accuracy (R^2 = 0.63, $RMSE$ = 63.11 m^3 ha^{-1}) than those based on Landsat 8 images (R^2 = 0.62, $RMSE$ = 64.40 m^3 ha^{-1}). However, some studies were conducted that combined optical images and microwave data to estimate forest variables [27–29]. Mauya et al. [30] conducted a remarkable study in which they evaluated the Sentinel-1, Sentinel-2, and ALOS PALSAR-2 images-based multiple linear regression (MLR) models for predicting FSV, and reported that the Sentinel-2 images performed much better with an $RMSE$ of 42.03% and a pseudo R^2 of 0.63. For predicting forest variables, Pham et al. showed that machine learning algorithms were likely to become more attractive in remote sensing [31]. Mura, Matteo et al. [32] discovered that the usage of eight k-nearest neighbors (kNN) methods with Sentinel-2A imagery proved successful in estimating the forest's GSV.According to the research conducted by Liu et al. [33] the combination of Sentinel-2A satellite image with kNN produced 97.0%, 93.2%, and 83.6% accuracy in three scales—forestry bureau, forest farm, and subclass, correspondingly. Li et al. [34] utilized the stepwise regression-based multiple linear regression models in their research, and with the inclusion of Sentinel-2A imagery data, forest inventory data, and digital elevation model (DEM) data of the study area, attempted to estimate the FSV of Linhai City and Chun'an County. Their study also employed the Stacking model with Least Absolute Shrinkage and Selection Operator (LASSO), and achieved the minimum Mean Absolute Percentage Error (MAPE) of 20.24% for FSV estimation. Jiang et al. [35] utilized a Random Forest regression (RFR) model with a spatial resolution of 30 m to predict the forest's growing stock volume (GSV) in Georgia. They suggested that ecophisiological variations in each forest could be accurately captured by variables derived from the Landsat time series. The authors recommend that future studies use a greater variety of methods, larger study areas, and Sentinel-2 data to predict the FSV with further accuracy.

However, ground plot survey data are still indispensable for remote sensing modeling [36]. The costs of ground plot surveys have always been high, which presents obstacles to estimating provincial FSV using remote sensing. Additionally, traditional sample location survey technology often produces serious positional deviations that may impact the modeling accuracy of plot data-based remote sensing estimations and predictions [37]. This is a reason for the biased estimations resulting from inaccurate matching of sample plots and pixels. Furthermore, limited research has compared the results of RF, BP, and MLR using Sentinel-2 and Landsat 8 images to predict FSVs. In addition, few studies have explored which variable selection methods can make the selected dataset have better explanatory and robustness to predict FSV. Hence, the novelty of this study primarily included the following aspects. (1) The field plot data used in this study were obtained from the 9th Beijing Forest Inventory Data (FID). The FID has comprehensive coverage and provides detailed information on forest resources. The measurement accuracy for forest parameters in the FID exceeds that of ordinary field measurements. By combining the forest inventory data with remote sensing imagery, more accurate actual values of forest stock volume can be provided, which is conducive to improving the accuracy of stock volume estimation.

(2) Variable selection was performed using Pearson correlation coefficient combined with linear stepwise regression (LSR) and the recursive feature elimination (RFE) method of Random Forest in the process of establishing the forest stock volume model. Six combinations of variables were formed, and multiple linear regression (MLR), back propagation (BP) neural network model, and Random Forest (RF) model were constructed to estimate the forest stock volume (FSV). The prediction accuracy of the models under different scenarios was compared and analyzed. (3) Based on Landsat 8 OLI and Sentinel-2 optical remote sensing data, the accuracy of forest stock volume estimation was compared between the two remote sensing data sources. The two remote sensing datasets were combined, and the forest stock volume was estimated using the combined data in conjunction with ground plot data. Additionally, a map of forest stock volume (FSV) in Huairou District, Beijing, was generated.

2. Materials and Methods

2.1. Study Site

The study site is situated in the northeast part of Beijing, in Huairou District, in the coordinates 116°17′–116°53′ E, and 40°41′–41°4′ N. Beijing has a total area of 2122.6 square kilometers, characterized by low-altitude terrain in the south and high-altitude terrain in the north, with elevations ranging from 34 to 1661 m. The plains in the south form part of the North China Plain, while the mountainous areas in the north belong to the Yan Mountain Range [38]. The climate is semi-humid and warm temperate, exhibiting dry and windy springs, high precipitation levels in summers, hot and humid weather patterns, drastic temperature drops and early frost in autumn, and cold and less snowy winters. The dominant tree species found in the artificial forests of Huairou District, Beijing, are *Pinus tabulaeformis Carr.* and *Populus tomentosa Carrière*. In addition, the natural forests primarily consist of *Quercus mongolica Fisch. ex Ledeb.*, *Platycladus orientalis* (L.) *Franco*, and *Betula platyphylla Suk*, among others. For clarification, the location of the study area and the distribution of the sample plots can be found in (Figure 1).
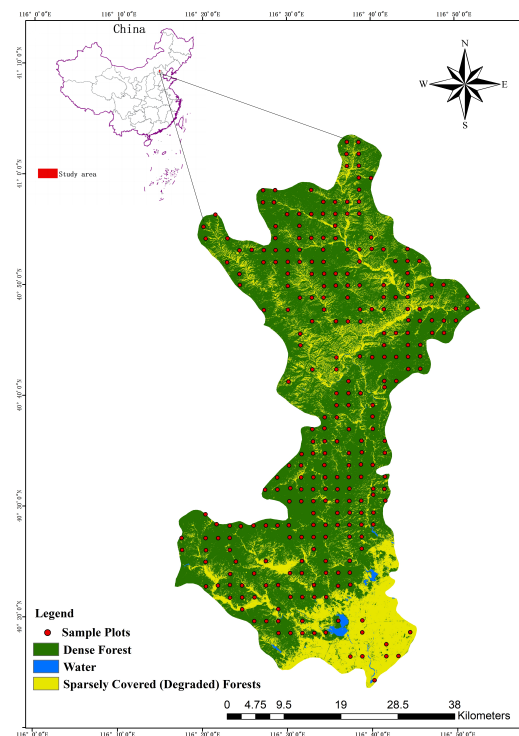

Figure 1. The study area is located in Huairou District, Beijing City, China, and the distribution of the sampling plots.

2.2. Forest Inventory Data

The data used for this study originated from the 2016 continuous inventory of forest resources in Beijing, known as "Forest Inventory Data (FID)" [39]. The systematic sampling method was used to establish fixed square sample plots, each measuring 0.067 hm² (25.8 m × 25.8 m), with a spacing of 2 km × 2 km between them. The recorded data from each sample plot pertained to latitude, longitude, soil thickness, average age, average diameter at breast height, and canopy density. The standing volume data were obtained using MySQL5.6 software by matching the plot and tree databases with the plot and tree numbers as matching criteria within the continuous forest resource inventory data. All Huairou district sample plots in Beijing were assessed, and the plots that did not meet the three-sigma rule criteria were excluded. The confidence level was set at 99.7%, with outliers identified as any plots with values that were three times the standard deviation away from the mean. A total of 273 plots remained after the abnormal data were removed. The plots were partitioned into training and testing sets in a 7:3 ratio with 70 plots each. A statistical table of plot volumes was generated.

2.3. Digital Elevation Model

The digital elevation model (DEM) primarily serves as raw data for topographical analyses [40]. It is capable of extracting terrain data, including slope, aspect, and elevation. These terrain features are used in various fields, including surveying, hydrology, and soil science. This study procured the necessary digital elevation model (DEM) data with a spatial resolution of 30 m from Geographic Space Data Cloud (www.gscloud.com, accessed on 2 December 2022). The DEM data was primarily utilized to extract and investigate the corresponding elevation, slope, and aspect values of the study area. Elevation, slope, and aspect were extracted from the DEM imagery as terrain factors. Terrain factors are important in distinguishing different landforms and their characteristics often have an impact on the growth rate of forests. Therefore, there is a theoretical relationship between terrain-based feature variables and forest stock volume (FSV) values. Thus, in this study, texture feature variables related to the terrain were extracted from elevation, slope, and aspect images of the study area.

2.4. Remote Sensing Data and Preprocessing

We used Landsat 8 OLI and Sentinel-2 MSI remote sensing sources. The Landsat 8 OLI data, along with its OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) sensors, was launched in 2013 [23]. The OLI comprises 9 bands (Table 1) of 30 m spatial resolution and one panchromatic band of 15 m resolution [41]. The TIRS contains two bands with a spatial resolution of 100 m. Studies show that the forest vegetation is lush during the summer and autumn seasons, and the overall change in forest characteristics is minimal. For this reason, this article primarily uses satellite imagery from the summer and autumn as its data source. Two Landsat 8 OIL images, taken on 11 October 2016 with a cloud cover of less than 10% and a resolution of 30 m, were chosen, taking into account image quality and cloud coverage. The chosen bands range from band 2 to band 7. (The B1 and B9 spectral bands are excluded from the study since B1 is primarily utilized for coastline observation, and B9 is predominantly used for cirrus cloud detection.) The image was downloaded from the official website of the United States Geological Survey (https://earthexplorer.usgs.gov, accessed on 23 December 2022). The study employed ENVI 5.3 to radiometrically calibrate the obtained Landsat 8 OLI image, apply FLAASH atmospheric correction, fuse images, and crop the image. The topographic correction expansion tool (Topographic Correction) in ENVI5.3 software was used to correct the image for terrain, which produced a multispectral image from Landsat 8 OLI of the study area.

Sentinel-2 is a mission by the European Space Agency (ESA) as part of the larger Copernicus program, which aims to provide high-quality Earth observation data for a range of applications [42]. Sentinel-2 is a pair of identical satellites, Sentinel-2A and Sentinel-2B, launched in 2015 and 2017, respectively [43]. One of the key features of

Sentinel-2A is its multi-spectral instrument (MSI), which captures images in 13 spectral bands (Table 1), ranging from the visible to the shortwave infrared. This enables the satellite to provide detailed information on land vegetation growth, forestry management, crop yields, and soil moisture content. The Sentinel-2 data provided by the European Space Agency (ESA) is L1C level multi-spectral instrument (MSI) data [44]. L1C data are georeferenced orthorectified imagery that needs radiometric and atmospheric calibration. Hence, it is crucial to transform Sentinel-2 data to level 2A after undergoing radiometric calibration and atmospheric correction. The Sentinel-2 imagery in this study were obtained from the website of the United States Geological Survey (https://earthexplorer.usgs.gov, accessed on 23 December 2022), collected in September 2016, and were in L1C format. The Sentinel-2 L1C level images were converted to L2A level using the ESA's Sen2Cor 2.5.5 module prior to analysis. Due to the fixed plot size of 25.8 m × 25.8 m, the spectral bands of Sentinel-2 with varying resolutions may not align. Potential image misalignments due to unmatched pixel and plot sizes could lead to significant errors in the study. Therefore, the 11 bands of Sentinel-2 imagery, including B2–B12, were resampled in SNAP8.0 software, and these 11 bands were uniformly resampled to a resolution of 25.8 m to match the sample sites.

Table 1. Spectral bands of Sentinel-2 MSI and Landsat 8 OLI images used in this study.

Band	Description	Wavelenghts		Satellite Instrument	Resolution
		min	**max**		
1	Coastal aerosol	430	457	Sentinel-2 MSI	60
2	Blue	448	546		10
3	Green	538	583		10
4	Red	646	684		10
5	Vegetation Red Edge (RE1)	694	713		20
6	Vegetation Red Edge (RE2)	731	749		20
7	Vegetation Red Edge (RE3)	769	797		20
8	Near-Infrared (NIR)	763	908		10
8a	Narrow NIR (nNir)	848	881		20
9	Water vapor	932	958		60
10	Shortwave infrared - Cirrus	1336	1411		60
11	Shortwave infrared (SWIR1)	1542	1685		20
12	Shortwave infrared (SWIR2)	2081	2323		20
1	Violet-deep Blue (V-D Blue)	433	453	Landsat 8 OLI	30
2	Blue	450	515		30
3	Green	525	600		30
4	Red	630	680		30
5	Near-Infrared (NIR)	845	885		30
6	Pan-Chromatic	1560	1660		30
7	SWIR—Cirrus	2100	2300		30
8	Shortwave infrared (SWIR1)	500	680		30
9	Shortwave infrared (SWIR2)	1360	1390		30

2.5. Characteristic Variable Extraction

Remote sensing texture feature refers to the frequency of tonal changes in an image, which reflects the surface characteristics of an object and is critical for extracting parameters and interpreting images [45]. This article utilizes the Gray-Level Co-occurrence Matrix (GLCM) method to extract texture features for the Landsat 8 image (B2–B7) and Sentinel-2A image (B2–B8A), respectively. The window size is set to 3 × 3. The extracted texture features using GLCM include mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation [46]. To differentiate the texture feature variables extracted from Landsat 8 and Sentinel-2, this article employs distinct expressions. For instance, the mean extracted from the blue band of Landsat 8 is denoted as Lan_{B2Mean}, while the correlation extracted from the blue band of Sentinel-2 is denoted as Sen_{B2Cor}. The

vegetation index (VI) is composed of a formulaic combination of diverse remote sensing spectral bands [47]. VI is of informative significance for determining vegetation characteristics and can be used to evaluate growth in both qualitative and quantitative aspects [48]. Based on Landsat 8 and Sentinel-2A image data, band calculations and extractions were carried out to obtain single-band reflectance. Vegetation indices and texture factors calculated from the single-band reflectance showed that the inclusion of single-band reflectance, vegetation indices, and texture factors can improve the accuracy of forest parameters estimation to varying degrees. Relevant literature was searched, and a total of 159 remote sensing variables were extracted in this article, including 69 remote sensing variables extracted from Landsat 8 and 84 variables extracted from Sentinel-2, and three topographic factors (Aspect, Slope, and DEM) extracted from DEM. Three forest inventory data included mean of tree age (Mage), mean diameter at breast height (DBH), and canopy density. (To avoid confusion, the expressions for Landsat 8 remote sensing variables and Sentinel-2 variables are distinguished in this article), where Landsat 8 blue band is expressed as Lan_{B2} and Sentinel-2 blue band is expressed as Sen_{B2}. In total, 6 Landsat 8 single-band factors, namely Lan_{B2}, Lan_{B3}, Lan_{B4}, Lan_{B5}, Lan_{B6}, and Lan_{B7}, were selected in this experiment, and a total of 11 Sentinel-2 single-band factors were selected, namely Sen_{B2}, Sen_{B3}, Sen_{B4}, Sen_{B5}, Sen_{B6}, Sen_{B7}, Sen_{B8}, Sen_{B8A}, Sen_{B10}, Sen_{B11}, and Sen_{B12} (Table 2).

Table 2. Description of predictor variables for the FSV estimation.

Variable	Description	Reference
Band Reflectivity	$Lan_{B2}, Lan_{B3}, Lan_{B4}, Lan_{B5}, Lan_{B6}, Lan_{B7}$	[23]
	$Sen_{B2}, Sen_{B3}, Sen_{B4}, Sen_{B5}, Sen_{B6}, Sen_{B7},$ $Sen_{B8}, Sen_{B8A}, Sen_{B10}, Sen_{B11}, Sen_{B12}$	[23]
Vegetation Index	$Lan_{NDVI} = (Lan_{B5} - Lan_{B4})/(Lan_{B5} + Lan_{B4})$	[49]
	$Lan_{ND43} = (Lan_{B4} - Lan_{B3})/(Lan_{B4} + Lan_{B3})$	[49]
	$Lan_{ND25} = (Lan_{B2} - Lan_{B5})/(Lan_{B2} + Lan_{B5})$	[49]
	$Lan_{ND563} = (Lan_{B5} + Lan_{B6} - Lan_{B3})/(Lan_{B5} + Lan_{B6} + Lan_{B3})$	[49]
	$Lan_{DVI} = Lan_{B5} - Lan_{B4}$	[49]
	$Lan_{SAVI} = (1 + L)(Lan_{B5} - Lan_{B4})/(Lan_{B5} + Lan_{B4} + L)$	[49]
	$Lan_{RVI} = Lan_{B5}/Lan_{B4}$	[49]
	$Lan_{PVI} = 0.939 Lan_{B5} - 0.344 Lan_{B4} + 0.09$	[49]
	$Sen_{RVI} = Sen_{B8}/Sen_{B4}$	[49]
	$Sen_{RVIre} = Sen_{B8}/Sen_{B5}$	[49]
	$Sen_{DVI} = Sen_{B8} - Sen_{B4}$	[49]
	$Sen_{NDVI} = (Sen_{B8} - Sen_{B4})/(Sen_{B8} + Sen_{B4})$	[49]
	$Sen_{NDVIre1} = (Sen_{B8} - Sen_{B5})/(Sen_{B8} + Sen_{B5})$	[49]
	$Sen_{NDVIre2} = (Sen_{B8} - Sen_{B6})/(Sen_{B8} + Sen_{B6})$	[49]
	$Sen_{mNDVI} = (Sen_{B8} - Sen_{B4})/(Sen_{B8} + Sen_{B4} - 2Sen_{B2})$	[49]
	$Sen_{mNDVIre} = (Sen_{B8} - Sen_{B5})/(Sen_{B8} + Sen_{B5} - 2Sen_{B2})$	[49]
	$Sen_{ND11} = (Sen_{B8} - Sen_{B11})/(Sen_{B8} + Sen_{B11})$	[49]
Topographic Factor	(Digital Elevation Model) Dem	[33]
	Aspect extracted from DEM (Aspect)	[33]
	Slope extracted from DEM (Slope)	[33]
Forest Inventory Data	Mean of Tree age (Mage)	[35]
	Mean Diameter at Breast Height (DBH)	[35]
	Canopy density	[35]
Texture analysis	Mean, Variance, Homogeneity, Contrast, Dissimilarity, Entropy, Second Moment, Correlation	[39]

2.6. Data Preprocessing

After extracting these modeling factors, it is necessary to standardize the factors with different dimensions and data levels in order to eliminate the influence of dimensional and scale differences on the results. This standardization will ensure that the numerical values of different types of data are consistent to the same scale.

The normalization formula for spectral bands [50], vegetation indices, texture features, and elevation is as follows:

$$x_i = (x - x_{min})/(x_{max} - x_{min})$$ (1)

The normalization formula for slope is as follows:

$$x_\alpha = \sin \alpha$$ (2)

The normalization formula for aspect is as follows:

$$x_\beta = \frac{\cos \beta + 1}{2}$$ (3)

2.7. Model Development

2.7.1. Correlation Analysis

The correlation relationship is a type of non-deterministic relationship that represents the degree of association between two variables. The correlation relationship between variables can be represented by correlation coefficients. Common correlation coefficients include Kendall correlation coefficient, Spearman correlation coefficient, and Pearson correlation coefficient [51]. This study uses the Person correlation coefficient to calculate the correlation between the modeling feature variables and forest stock. The Pearson correlation coefficient can be calculated using SPSS26 software. Its sign indicates the direction of the correlation, while the value represents the strength of the relationship.

The formula is as follows

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$ (4)

where x_i and y_i, respectively, indicate the values of the two variables of the ith sample, $\bar{x}$ and $\bar{y}$ represent the mean values of the two variables, and n represents the sample size. The value of r_{xy} ranges from -1 to 1. When r_{xy} is equal to 1, it indicates a completely positive linear relationship between the two variables. When r_{xy} is equal to -1, it indicates a completely negative linear relationship between the two variables. When r_{xy} is equal to 0, it indicates that there is no linear relationship between the two variables.

2.7.2. Linear Stepwise Regression

The process of linear stepwise regression (LSR) involves introducing explanatory variables into the regression equation in descending order of their correlation with the dependent variable [52]. After each addition of an explanatory variable, it undergoes an F-test followed by a T-test for each selected variable. If a previously introduced explanatory variable becomes insignificant as a result of introducing later ones, it is eliminated to ensure that the regression equation comprises only significant variables before adding another variable. This iterative process is effective in reducing the problem of multicollinearity and often yields a relatively low variance inflation factor (VIF) of the screened independent variables. Consequently, we avoid the issue of high correlation between independent variables and improve the accuracy of the model.

The formula is as follows:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p$$ (5)

b_0 represents the intercept, $x_1, x_2, \ldots, x_p$ represent the selected independent variables, and $b_1, \ldots, b_p$ represent the regression coefficients of each independent variable.

2.7.3. Recursive Feature Elimination

Random Forest recursive feature elimination is an important feature selection method in machine learning. Recursive feature elimination (RFE) repeatedly constructs models, selects the best or worst features, and repeats the process on the remaining features until all features are evaluated [53]. This method ranks features based on the order in which they are selected. This article utilizes the Random Forest algorithm for backward recursive feature elimination. The main process involves training the model using the feature selection RFE method in Python 3.6. First, the model is trained using all the feature variables, and the importance score for each feature variable is calculated. Next, based on a predetermined number of features to select, the feature variables with the highest scores are chosen and retained. Finally, the remaining feature variables are trained as a new feature set, and the above process is repeated until the desired number of features are obtained.

2.7.4. Multiple Linear Regression

Multiple linear regression (MLR) is a method of establishing the optimal fitting model by screening variables [30,54–56]. It reduces the multicollinearity of the model by removing variables with low contribution and high correlation with other variables from the candidate variables. In this study, the forest volume of the 207 sample plots in the training set was taken as the dependent variable, and the selected modeling factors after LSR and RFE method screening were taken as the independent variables. Based on different modeling factor sets, multiple linear regression models were established. The variable selection method for the independent variables in SPSS version 26.0 should be set to "Stepwise" with a confidence interval of 95%, utilizing the probability of the F statistic (sig) as the stepping method criterion. Any variable that satisfies the condition $F \leq 0.05$ will be included in the regression model, while variables with $F \geq 0.10$ will be excluded. The formula for multiple linear regression (MLR) is provided below [57].

The formula for the MLR is:

$$\widehat{GSV} = \beta + \sum_{i=1}^{n} \alpha_i \times x_i \tag{6}$$

$\widehat{GSV}$ represents the predicted FSV value of the multiple linear regression model; β is the constant intercept of the linear regression model; x_i represents the feature variable selected by the LSR or RFE method; n represents the total number of feature variables selected by the LSR or RFE method; and α_i represents the linear regression equation coefficient corresponding to the feature variable x_i.

2.7.5. Back Propagation Neural Network

The Back Propagation (BP) neural network is a widely used neural network model based on the error Back Propagation algorithm for multi-layer feedforward neural networks. It was first proposed by Rumelhart and McClelland in 1986 and is currently the most commonly used neural network model. The BP neural network is a common branch of ANN, which utilizes the back propagation of the error between the predicted and real values to update the weights and biases in the BP neural network [58]. The fundamental principle of the BP neural network algorithm is to minimize the mean squared error between the output and expected values by utilizing gradient descent, a technique which involves searching for the gradient. The BP neural network is widely applied in numerous fields due to its advantages such as self-learning, adaptivity, and parallel processing. The neural network model can be accessed directly in MATLAB through programming or the built-in toolbox. As such, this paper utilizes MATLAB R2019a(9.6.0) software to model and test the BP neural network.

The relationship between the input parameters x and output values y can be expressed as

$$y = f\left(\sum wx + b\right) \tag{7}$$

where w and b represent the weights matrix and the biases vector from the adjacent layers, respectively, f represents the activation function [59].

2.7.6. Random Forest Model

The Random Forest (RF) algorithm is a non-parametric ensemble learning algorithm that is less sensitive to noisy data and has strong noise resistance capabilities [60–65]. It does not face the risk of overfitting and does not require assumptions about the distribution of input data. Additionally, it can process input samples with high-dimensional features without the need for dimensionality reduction and can effectively handle large datasets. The RF algorithm utilizes the bootstrap method to randomly extract multiple samples from the original population and generate a group of regression trees (ntree). For each tree, the RF algorithm randomly selects a subset of predictors at each splitting node (mtry) to build a tree without requiring tree pruning. In building each tree, a procedure called the "out-of-bag" (OOB) error is used to independently construct each tree based on the training data. The RF algorithm calculates variable importance (VI) and OOB error.

The OOB error can be estimated as follows

$$OOB_{\text{error}} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{8}$$

where y_i is the measured FSV, $\hat{y}_i$ is the predicted FSV, and n is the total number of OOB samples.

2.7.7. Model Evaluation

We assessed the fitting and predictive performance of the model on a set of independent test data using various statistical metrics, including the coefficient of determination (R^2), absolute mean relative error ($RMAE$), root mean square error ($RMSE$), and mean absolute error (MAE).

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \right] \tag{9}$$

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

$$MAE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{n} \right| \tag{11}$$

$$RMAE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \tag{12}$$

where y_i is the observed value, $\hat{y}_i$ is the predicted value, n is the number of samples, and $\bar{y}$ is the average of all the observed values

3. Results

For this study, 203 out of the 273 total sample plots (70%) were used as the training data, and 70 out of the 273 total sample plots (30%) were used as the test data. Table 3 lists the summarized statistics of the plot-level training data and test data. For the training data, the minimum FSV was only 0.09 m^3 ha^{-1}, and the maximum FSV was 187.725 m^3 ha^{-1}. For the test data, the FSV ranged from 0.075 m^3 ha^{-1} to 140.190 m^3 ha^{-1}. In accordance with the three-sigma rule and using a 99.7% confidence interval, we retained the plot volume sampling points within three times the standard deviation range and removed outlier sample points that were less than $\mu - 3\sigma$ or greater than $\mu + 3\sigma$ (where μ is the sample mean and σ is the sample standard deviation). MATLAB R2019a(9.6.0) a software was used to remove data outliers based on the three-sigma rule. Finally, 203 and 70 sampling sites were, respectively, selected as the training and testing sets (Table 3).

Table 3. Summary of sample plot data processing results.

Type	Number of Sample Plots	Min m³ ha⁻¹	Max m³ ha⁻¹	Mean m³ ha⁻¹	SD * m³ ha⁻¹
Training Data	203	0.09	187.725	32.176	35.229
Test Data	70	0.075	140.19	28.897	29.885
Total sample plots	273	0.075	187.725	31.279	33.838

* SD represents standard deviation of variables.

3.1. Correlation Analysis and Variable Selection

After preprocessing of remote sensing images, this study selected feature variables extracted from Landsat 8 and Sentinel-2 images as potential predictive variables related to FSV. Modeling factors were screened for the obtained feature variables through Pearson correlation coefficient testing in SPSS26 software. The Pearson correlation coefficient can be used to describe the correlation between feature variables and ground truth data. The study calculated the Pearson correlation coefficient between the survey FSV of the sample plot and all extracted Landsat 8 feature variables and Sentinel-2 feature variables separately. Among the feature variables of Landsat 8, there are 23 feature variables that are significantly correlated at the 0.05 significance level (bilateral), and 13 feature variables that are significantly correlated with FSV at the 0.01 level of significance. Among the feature variables extracted from Landsat 8 images, except for the near-infrared band (Lan_{B5}), the reflectance of the other five bands is significantly negatively correlated with FSV. Vegetation indices (Lan_{NDVI}, Lan_{RVI}, Lan_{SAVI}) are all positively correlated with FSV and Aspect, DEM are positively correlated with FSV, while texture factors (Lan_{B2Mean}, Lan_{B3Mean}, Lan_{B4Mean}, Lan_{B6Mean}, Lan_{B7Mean}) are negatively correlated with FSV. Among the feature variables of Sentinel-2, there are 24 feature variables that are significantly correlated at the 0.01 level, among which Sen_{B2}, Sen_{B3}, Sen_{B4}, Sen_{B5} are negatively correlated with FSV, and vegetation indices ($Sen_{mNDVIre}$, Sen_{RVI}, Sen_{DVI}) are all positively correlated with FSV. Texture factors (Sen_{B2Cor}, Sen_{B2Hom}, Sen_{B2Sec}, Sen_{B4Sec}) are all positively correlated with FSV, while the remaining texture factors (Sen_{B2Mean}, Sen_{B2Ent}, Sen_{B2Diss}, Sen_{B3Mean}, Sen_{B4Ent}, Sen_{B4Mean}, Sen_{B5Ent}, Sen_{B5Mean}) are negatively correlated with FSV. The correlation between DEM and FSV in terrain factors is 0.340.

As shown in Table 4, this study used Landsat 8 and Sentinel-2 images, and selected Landsat 8 and Sentinel-2, as well as feature variable combinations using linear stepwise regression (LSR) and recursive feature elimination (RFE) methods, respectively. With Landsat 8 as the image source, nine feature variables were selected using LSR, including multi-spectral bands; (Figure 2) shows that the maximum coefficient of determination R^2 and the minimum root mean squared error ($RMSE$) were achieved when the number of variables was three, as selected by RFE. Nine feature variables were ultimately selected. Using Sentinel-2 as the image source, LSR selected five feature variables; by using RFE selection, when the number of variables was three (Figure 2), the coefficient of determination R^2 of the model reached its maximum and both R^2 and root mean squared error ($RMSE$) reached its minimum. As a result, nine feature variables were ultimately screened. To study the joint ability of Landsat 8 data and Sentinel-2 data to invert the volume, in this experiment, the resolution of the two satellite images was both set to 25.8 m during resampling, which was used to match the sample sites. By combining Sentinel-2 and Landsat 8 remote sensing variables, nine feature variables were selected after LSR screening; using RFE selection, when the number of variables is five (Figure 2), the determination coefficient of the model reaches its maximum, and at the same time, R^2 and $RMSE$ reach their minimum. Finally, 10 feature variables are selected.

Table 4. The precision of FSV estimation values was compared for three models, MLR, BP, and RF, under different combinations of variable selection methodologies of LSR and RFE.

Data Source	Variable Selection Method	Variables Combination	Methods	R^2	RMSE (m³ ha⁻¹)	rRMSE (%)	MAE (m³ ha⁻¹)
		Lan_{NDVI}, Lan_{RVI},	MLR	0.595	19.491	56.309	15.947
	LSR	Lan_{B2}, Lan_{B3}, Lan_{B4}	BP	0.704	16.666	48.147	12.239
Landsat 8 OLI		Lan_{B5}, Lan_{B6}, Dem, $Aspect$	RF	0.776	14.488	41.855	9.449
		Lan_{RVI}, Lan_{B2}, Lan_{B4},	MLR	0.602	19.323	55.824	15.699
	RFE	Lan_{B5}, Lan_{B6}, Lan_{B7},	BP	0.779	14.413	41.639	11.383
		Lan_{B2Mean}, Dem, $Aspect$	RF	0.782	14.317	41.36	9.273
		Sen_{B2}, Sen_{B5},	MLR	0.605	19.257	55.633	15.673
	LSR	$Sen_{mNDVIre}$, Sen_{RVI},	BP	0.707	16.583	47.907	12.487
Sentinel-2		Dem	RF	0.796	13.826	39.942	9.081
		Sen_{B3}, Sen_{B4}, Sen_{B5},	MLR	0.684	17.215	49.732	12.869
	RFE	Sen_{B2Cor}, Sen_{B4Sec}, Sen_{B5Ent},	BP	0.713	16.385	47.335	12.571
		Sen_{B5Mean}, Sen_{RVI}, Dem	RF	0.799	13.743	39.702	8.899
		Lan_{B2}, Lan_{B4}, Lan_{B5},	MLR	0.613	19.043	55.015	15.591
	LSR	$Sen_{mNDVIre}$, Sen_{RVI}, Sen_{B3},	BP	0.737	15.713	45.394	12.303
Landsat 8 OLI		Sen_{B5}, Dem, $Aspect$	RF	0.83	12.616	36.448	9.276
and Sentinel-2		Sen_{B3}, Sen_{B4}, Sen_{B5}, Lan_{B2},	MLR	0.627	18.719	54.077	15.252
	RFE	Sen_{RVI}, $Sen_{mNDVIre}$, Sen_{B5Mean},	BP	0.719	16.228	46.881	13.088
		Sen_{B2Cor}, $Aspect$, Dem	RF	0.831	12.604	36.411	9.366

Optimal hyperparameters are shown in Figure 3a–c for the Random Forest models. The number of estimators was varied to observe the effect on the out-of-bag (OOB) error. It is notable that the Landsat 8-based RF model had the lowest out-of-bag error when 310 estimators were used. For the Sentinel-2 based RF model, 280 estimators resulted in the lowest out-of-bag error. The joint variable data of Landsat 8 and Sentinel-2 resulted in the lowest out-of-bag error when 265 estimators were used.
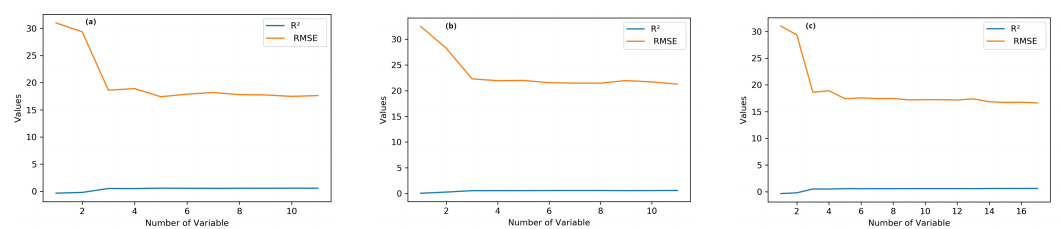

Figure 2. The precision change of recursive feature elimination (RFE) under different numbers of feature variables: The figures from left to right are, respectively, labeled as (**a**) Landsat 8 (**b**) Sentinel-2 (**c**) Landsat 8 data and Sentinel-2 combined.
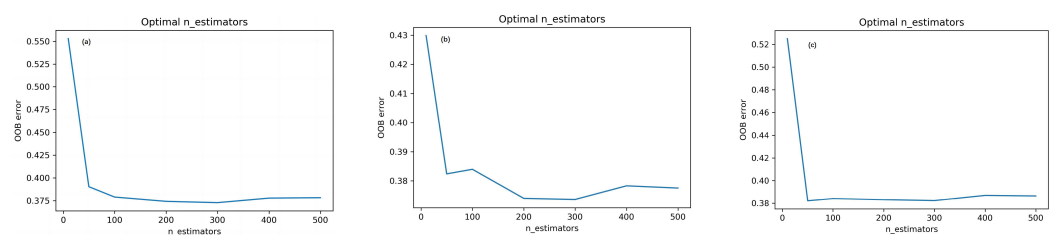

Figure 3. Landsat 8 data and Sentinel-2 combined; graph of the OOB error for the RF model with different hyperparameter combinations: (**a**) Landsat 8 (**b**) Sentinel-2 (**c**) Landsat 8 data and Sentinel-2 combined.

3.2. Forest Volume Estimation Results Based on Landsat 8 and Sentinel-2 Imagery

Table 4 displays the results of the models that selected feature variable combinations using the LSR method based on Landsat 8 images. The RF model exhibited the highest inversion accuracy, and both $rRMSE$ and MAE values were minimized (R^2 = 0.776, $RMSE$ = 14.488 m^3 ha^{-1}, $rRMSE$ = 41.855%, MAE = 9.449 m^3 ha^{-1}). Conversely, the BP model had lower precision with an R^2 value of 0.704, while the MLR model had the poorest performance with an R^2 value of 0.595. The RF model continued to have the highest inversion accuracy among those constructed with the feature variable combinations selected by the RFE method, and both $rRMSE$ and MAE values were also the lowest (R^2 = 0.782, $RMSE$ = 14.317 m^3 ha^{-1}, $rRMSE$ = 41.360%, MAE = 9.273 m^3 ha^{-1}). The BP model had slightly lower precision with an R^2 value of 0.779, while the MLR model had the poorest performance with an R^2 value of 0.602. When based on Sentinel-2 images and constructed using the LSR method, the RF model again had the highest inversion accuracy among the MLR, BP, and RF models, with minimized $RMSE$, $rRMSE$, and MAE values (R^2 = 0.796, $RMSE$ = 13.826 m^3 ha^{-1}, $rRMSE$ = 39.942%, MAE = 9.081 m^3 ha^{-1}). The BP model, again, presented lower precision with an R^2 value of 0.706, while the MLR model had the poorest performance with an R^2 value of 0.605. The RF model continued to provide the highest inversion accuracy when constructed with the feature variable combinations chosen by the RFE method (R^2 = 0.799, $RMSE$ = 13.743 m^3 ha^{-1}, $rRMSE$ = 39.702%, MAE = 8.899 m^3 ha^{-1}). In contrast, the BP model exhibited lower precision (R^2 = 0.714, $RMSE$ = 16.385 m^3 ha^{-1}, $rRMSE$ = 47.335%, MAE = 12.571 m^3 ha^{-1}), while the MLR model's performance was again the weakest with an R^2 value of 0.684.

To study the ability to jointly retrieve the storage capacity using Landsat 8 and Sentinel-2 remote sensing data, three models (MLR, BP, RF) were constructed using modeling factors based on a screening of all the feature variables of Sentinel-2 and Landsat 8 remote sensing data. The model constructed using the feature variables selected by the LSR method had the highest accuracy, and the RF model had an R^2 coefficient of 0.830. The model also had the lowest $RMSE$, $rRMSE$, and Mean Absolute Error (MAE) values, which were 12.616 m^3 ha^{-1}, 36.448%, and 9.276 m^3 ha^{-1}, respectively. The BP model was slightly less accurate with an R^2 coefficient of 0.737, while the MLR model's relatively least accurate with an R^2 coefficient of 0.613. Among the three models constructed based on the feature variables combination selected by the RFE method, the RF model had the highest accuracy, with an R^2 coefficient of 0.831, $RMSE$ of 12.604 m^3 ha^{-1}, $rRMSE$ of 36.411%, and MAE of 9.366 m^3 ha^{-1}. The BP model had a lower accuracy with an R^2 coefficient of 0.719, $RMSE$ of 16.228 m^3 ha^{-1}, $rRMSE$ of 46.881%, and MAE of 13.088 m^3 ha^{-1}. The MLR model had the least accuracy with an R^2 coefficient of 0.627. In both variable selection methods, the RF model performed better, and the RF regression accuracy based on RFE was more superior to using LSR. The Random Forest method is based on sorting the importance of all feature variables and can choose the feature variables that contribute the most to the model, thereby significantly improving the model's accuracy.

3.3. Comparison of the Predicted FSV Estimates among the MLR, BP and RF Methods

The accuracy of the model is validated using 70 sample data from the test set. The estimated values of FSV are compared with measured values in (Figures 4–6). For the same set of modeling factors, RF has the best prediction accuracy for FSV, followed by the BP model, and finally the MLR model. This indicates that machine learning methods have an advantage over traditional regression methods for FSV. Comparing the two sets of modeling factors, the model accuracy based on the Sentinel-2 (Figure 5) modeling factor set is better than that of the same type of model based on the Landsat 8 (Figure 4) modeling factor set. This suggests that Sentinel-2 is more advantageous than Landsat 8, possibly due to the high resolution and the red edge band that can reflect vegetation growth levels. The accuracy of the forest accumulation estimation model established by using both Sentinel-2 and Landsat 8 data (Figure 6) is better than that based on single optical remote sensing data. The determination coefficient R^2 of the three forest accumulation estimation models

is greater than 0.595, and the overall accuracy exceeds that of the forest accumulation estimation model constructed based on single optical remote sensing data. Further analysis of the model's fitting effect is carried out through the scatter diagram of the estimated and measured values of the three forest accumulation estimation models.
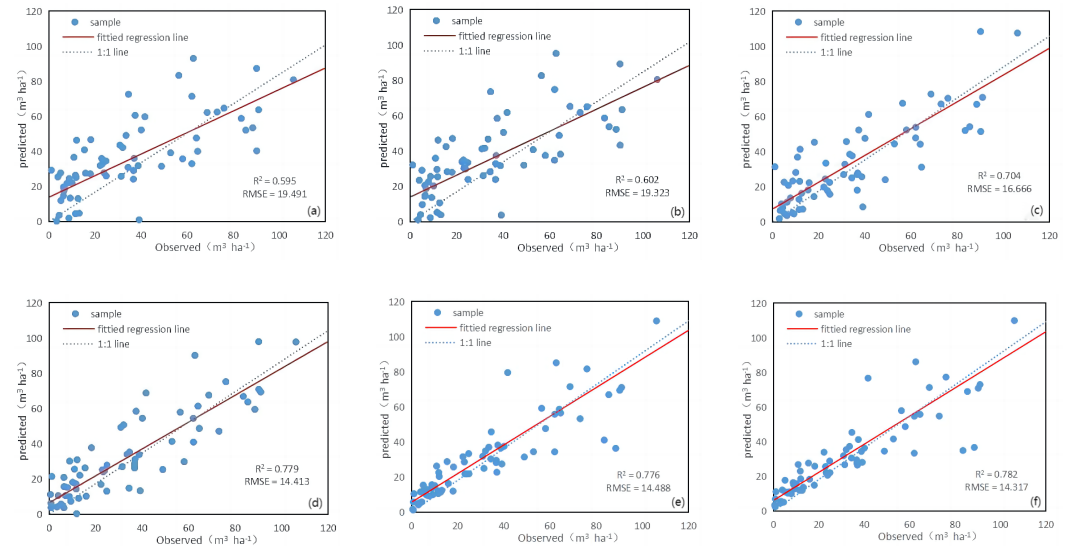

Figure 4. Scatter plots of the observed and predicted FSV values by: (**a**) Landsat 8 LSR-MLR (**b**) Landsat 8 RFE-MLR (**c**) Landsat 8 LSR-BP (**d**) Landsat 8 RFE-BP (**e**) Landsat 8 LSR-RF (**f**) Landsat 8 RFE-RF.
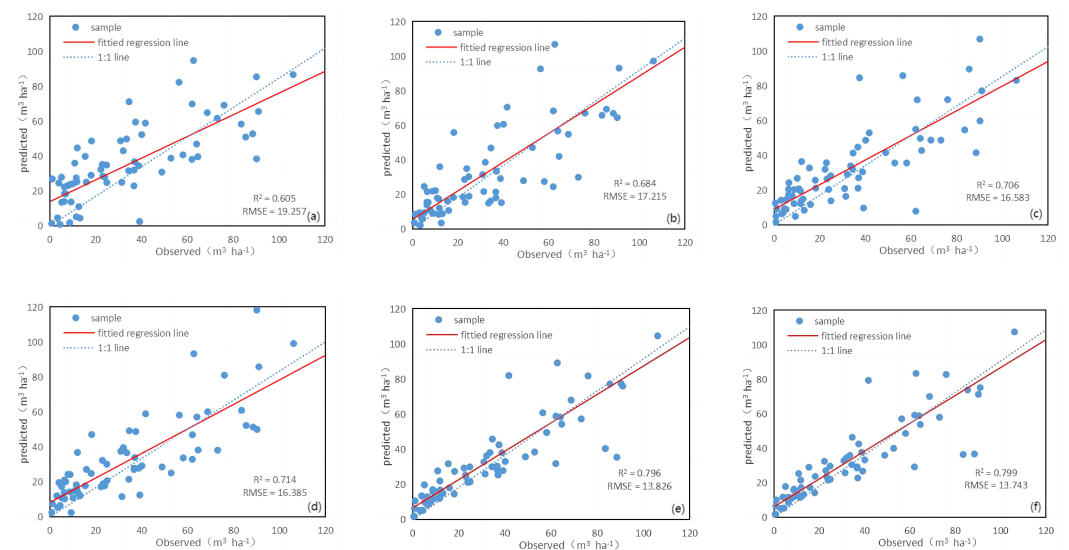
Figure 5. Scatter plots of the observed and predicted FSV values by: (**a**) Sentinel-2 LSR-MLR (**b**) Sentinel-2 RFE-MLR (**c**) Sentinel-2 LSR-BP (**d**) Sentinel-2 RFE-BP (**e**) Sentinel-2 LSR-RF (**f**) Sentinel-2 RFE-RF.
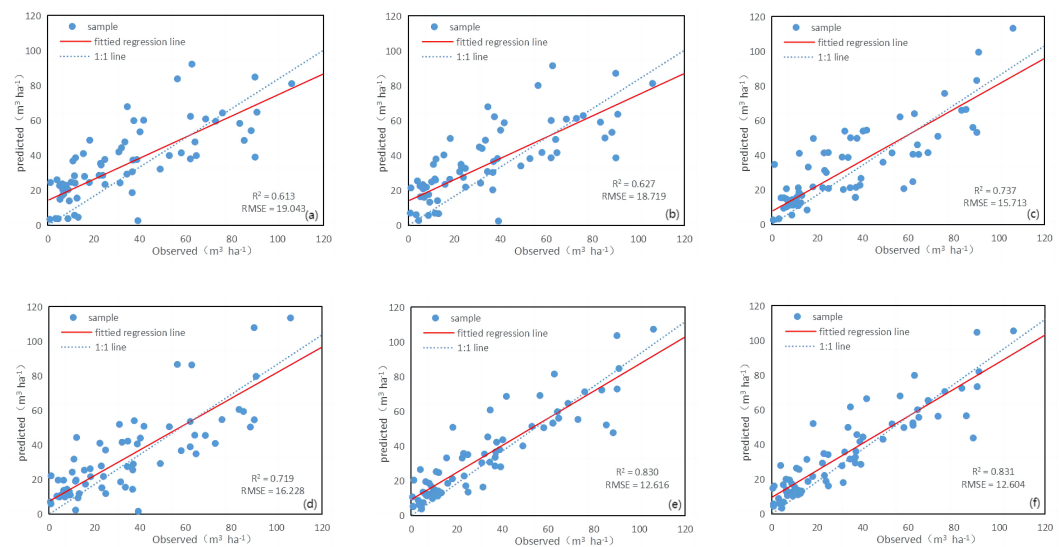
Figure 6. Scatter plots of the observed and predicted FSV values by: (**a**) Landsat 8–Sentinel-2 LSR-MLR (**b**) Landsat 8–Sentinel-2 RFE-MLR (**c**) Landsat 8–Sentinel-2 LSR-BP (**d**) Landsat 8–Sentinel-2 RFE-BP (**e**) Landsat 8–Sentinel-2 LSR-RF (**f**) Landsat 8–Sentinel-2 RFE-RF.

3.4. Map of the FSV Estimation

Firstly, using Google Earth Engine (GEE) to download the 10 m land use classification data, the forest area of Huairou District, Beijing was extracted. Then, based on the variables selected from the combined Landsat 8 and Sentinel-2 data and the best model RFE-RF ($R^2 = 0.831$), the FSV in the study area was estimated and the spatial distribution map of FSV in the study area was obtained, as shown in the (Figure 7). According to (Figure 7), areas of high forest stock volume concentration are predominantly in the northern, central, and northwestern parts of the research area, with the estimated values mostly ranging from 150 to 200 m^3 ha^{-1}. Conversely, regions with lower forest stock volumes are concentrated in the southern and southeastern parts with estimated values mostly ranging from 0 to 50 m^3 ha^{-1}. The variation in forest stock volume resulted from the relatively high terrain with dense vegetation and fewer anthropogenic activities in the northern, central, and northwestern parts and the relatively flat terrain with more human activities in the southern and southeastern parts. The forest stock volume inversion map of the study area corresponds to actual survey results, accurately reflecting the spatial distribution of the actual forest stock volume in the researched area (Figure 7).
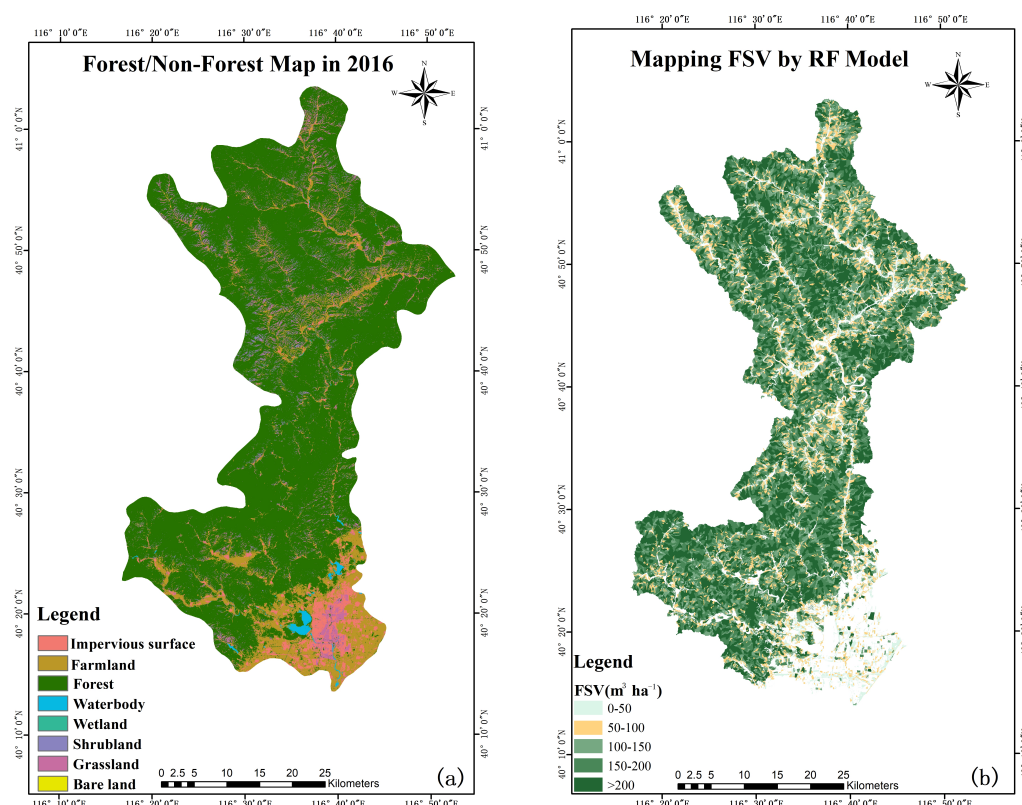
Figure 7. Land use classification map and predicted FSV map by: (**a**) the GEE land use classification forest/non-forest map in 2016; (**b**) predicted FSV map of the study area in 2016.

4. Discussion

In this study, we used the Forest Inventory Data (FID) to obtain actual survey data. We combined Landsat 8 and Sentinel-2 derived variables with remote sensing data before and after the joint variable, then studied FSV estimation by combining these data with ground sample plot data. We analyzed three data sources using Pearson correlation coefficients, combined with LSR and RFE methods to screen variables and construct three FSV estimation models (MLR, BP, and RF). The results of the regression models showed that the single data sources have strong forest stock volume estimation ability. However, the joint performance of the two data sources demonstrated a stronger capability of FSV estimation. Through precision analysis and evaluation, we determined that the RFE-RF model of the joint Landsat 8 and Sentinel-2 image was the optimal model for estimating forest stock volume. We then performed remote sensing inversion of forest stock volume in the study area and analyzed its spatial distribution characteristics.

The results show that among all the modeling factors, the one with the highest correlation with FSV is Sentinel-2's red-edge band 1 (B5). The correlation between Sentinel-2 and FSV is generally higher than that of Landsat 8, and the correlation between red-edge band 1 (B5) and FSV is generally higher. The correlation between Sentinel-2A's red-edge band 1 (B5) and FSV is more significant than the other variables used, and features derived from the red-edge band are prioritized in reducing model error. The most efficient predictive band was the red-edge 1 (B5), performing exceptionally well in both the machine learning methods and the MLR method, as validated by recent studies on forest prediction [66] and tree species classification [67]. In the domain of gross primary productivity (GPP), Lin et al. discovered the usefulness of the red-edge band in evaluating GPP. They also pointed out that the red-edge reflectance was sensitive to the chlorophyll content in leaves [68]. Additionally, the leaf chlorophyll content was a crucial variable to consider in forest analysis. While the machine learning models and the MLR did not select the same variables,

B5 stood out as a critical variable for estimating FSV. Despite the differences in modeling variables, the accuracy results from the model verification were not significantly dissimilar, emphasizing B5's advantage. Chrysafis et al. found a different result on using the RF algorithm and Sentinel-2 images for estimating FSV in a Mediterranean forest ecosystem. They highlighted B11 (SWIR 1) as the most critical variable [26]. Our study was congruent with Astola et al., where they identified B5 as the most crucial variable for predicting FSV [9].

Different combinations of feature variables can result in different modeling accuracies, and selecting appropriate variable selection methods can significantly improve the estimation accuracy of forest stock volume (FSV). The original spectral bands, vegetation indices, texture features, and terrain factors are commonly used feature variables for estimating FSV [10,69]. In this study, variables that are significantly correlated with FSV were extracted, and six feature variable combinations were formed using two variable selection methods, linear stepwise regression (LSR) and recursive feature elimination (RFE). The retained feature variables were used with multiple linear regression(MLR) models , back propagation neural(BP) network models , and Random Forest(RF) models for estimating forest stock volume in the study area. The study found that the LSR method provides the optimal linear combination of feature variables but requires the variables to be non-collinear. However, due to the complexity of forest ecosystems, the relationship between FSV and feature variables may not be linear, which limits the estimation accuracy of the model [34]. The RFE method can assess the importance of feature variables based on non-linear relationships, enabling the selection of highly important variables to be used for modeling. Unlike the LSR method, which is based on linear correlations, the RFE method selects variables based on their importance ranking, improving the selection process. However, the importance evaluation is relative, and the importance of individual variables can also vary across different feature variable combinations [70].

Therefore, we propose the RFE method for selecting appropriate combinations of feature variables. In this method, we first use the importance ranking determined with the entire dataset as a reference, and then, as the number of feature variables increases, we select suitable combinations based on changes in the error. Finally, we find that the variable combinations determined by the RFE method have the highest estimation accuracy and can effectively reduce estimation errors. Variable selection was performed using LSR and RFE. Ultimately, six variable combinations were selected, and MLR, BP, and RF models were constructed to estimate FSV. The models consisted of six variable combinations, including those with texture features and those without texture features. As shown in Table 4, overall, the models with texture features exhibited higher accuracy than the models without texture features. However, regardless of the presence of texture features, the accuracy of the forest stock volume estimation models constructed using Landsat 8 and Sentinel-2 data in combination was superior to that of using a single optical remote sensing data. Among them, the Landsat 8-Sentinel-2 RFE-RF model achieved the best estimation performance, with the highest coefficient of determination (R^2 = 0.831) and the lowest RMSE, rRMSE, and MAE. The results indicate that the RFE variable selection method achieved the best estimation performance in both single optical remote sensing data and combined remote sensing data selection processes. Additionally, after incorporating texture features in the modeling of the optimal Landsat 8-Sentinel-2 RFE-RF model, the rRMSE significantly decreased by 3.7% compared to the combined remote sensing RFE-RF model without texture features. The variable selection results of each model suggest that the principal components of texture features are more sensitive to the forest stock volume estimation.

As shown in Figure 4, both Landsat 8 and Sentinel-2 data have the potential to estimate FSV. When using a single data source to estimate FSV, the Landsat 8 data predicted an R^2 of 0.782 and the Sentinel-2 predicted an R^2 of 0.796. This demonstrates that using both remote sensing data sources for the inversion of forest stock volume is feasible. This conclusion is consistent with the findings of Astola H, Häme T, Sirro L, et al. [9]. The innovation of this article is that, by combining the person correlation coefficient method with the RFE method

to select feature variables, the dimensionality of the data can be reduced without losing the original data, and the accuracy of the stockpile estimation model can be improved. MLR, BP, and RF models established through RFE feature variable selection method are all superior to LSR method. The RFE method is based on the importance ranking of all feature variables, which can select the feature variables with the greatest contribution to the model and effectively improve the model accuracy. This conclusion is consistent with Attarzadeh R et al. [71]. This article shows that the average accuracy of models constructed using combined data from Landsat 8 and Sentinel-2 variables is improved to varying degrees compared to models constructed using single remote sensing data. This conclusion is consistent with the study by Poortinga, Ate, Tenneson, and Karis [72]. The RF model constructed by using the Pearson correlation coefficient method and the recursive feature elimination (RFE) method of Random Forest to screen the feature variables achieved the best results, with R^2 = 0.831, $RMSE$ = 12.604 m^3 ha^{-1}, $rRMSE$ = 36.411%, MAE = 9.366 m^3 ha^{-1}. Compared to the optimal model constructed using a single data source, the root mean square error of the best model constructed using combined data has decreased by 2–4 m^3 ha^{-1}.

5. Conclusions

Machine learning algorithms are currently widely applied in forest biomass estimation, but there is relatively less research on estimating FSV. In this study, Landsat 8 and Sentinel-2 data were selected as remote sensing data sources. A new data source was created by combining the two satellite datasets as variables. Three FSV estimation models (MLR, BP, RF) were constructed using two variable screening methods LSR and RFE, respectively. Overall, both machine learning models outperformed the traditional MLR model. Among them, the RF model performed the best, and the BP model also achieved good estimation results. In this study, the method of using Person correlation coefficient combined with RFE as a variable screening method was found to be more effective than the LSR variable screening method. However, this study only systematically compared the model accuracy of the Person-combined RFE method and the LSR variable screening method constructed in the same research area. In the future, this variable screening method can be applied to different research areas to better serve operational forest management practices. The current general method of combining multiple sources of remote sensing data is to use active remote sensing data combined with single optical remote sensing data. However, active remote sensing data are relatively expensive compared to optical remote sensing data. The limitation is that this study only combined remote sensing variables, and in future experiments, combining different bands of the two data sources can be attempted for image fusion, which may yield better FSV estimation results.

References

1. Gjertsen, A.K. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sens. Environ.* **2007**, *110*, 420–430. [CrossRef]
2. Katila, M.; Tomppo, E. Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sens. Environ.* **2001**, *76*, 16–32. [CrossRef]
3. Yan, W.; Wang, W.; Peng, Y.; Chen, X. Evaluation of Biomass and Carbon Stocks in Three Pine Forest Types in Karst Area of Southwestern China. *J. Sustain. For.* **2022**, *41*, 18–32. [CrossRef]
4. Sasaki, N.; Myint, Y.Y.; Abe, I.; Venkatappa, M. Predicting carbon emissions, emissions reductions, and carbon removal due to deforestation and plantation forests in Southeast Asia. *J. Clean. Prod.* **2021**, *312*, 127728. [CrossRef]
5. Zhao, J.; Zhao, L.; Chen, E.; Li, Z.; Xu, K.; Ding, X. An Improved Generalized Hierarchical Estimation Framework with Geostatistics for Mapping Forest Parameters and Its Uncertainty: A Case Study of Forest Canopy Height. *Remote Sens.* **2022**, *14*, 568. [CrossRef]
6. Gschwantner, T.; Alberdi, I.; Bauwens, S.; Bender, S.; Borota, D.; Bosela, M.; Bouriaud, O.; Breidenbach, J.; Donis, J.; Fischer, C.; et al. Growing stock monitoring by European National Forest Inventories: Historical origins, current methods and harmonisation. *For. Ecol. Manag.* **2022**, *505*, 119868. [CrossRef]
7. Rees, W.G.; Tomaney, J.; Tutubalina, O.; Zharko, V.; Bartalev, S. Estimation of boreal forest growing stock volume in russia from sentinel-2 msi and land cover classification. *Remote Sens.* **2021**, *13*, 4483. [CrossRef]
8. Persson, H.J. Estimation of boreal forest attributes from very high resolution Pléiades data. *Remote Sens.* **2016**, *8*, 736. [CrossRef]
9. Astola, H.; Häme, 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]
10. Li, L.; Zhou, X.; Chen, L.; Chen, L.; Zhang, Y.; Liu, Y. Estimating urban vegetation biomass from Sentinel-2A image data. *Forests* **2020**, *11*, 125. [CrossRef]
11. Fassnacht, F.; Hartig, F.; Latifi, H.; Berger, C.; Hernández, J.; Corvalán, P.; Koch, B. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sens. Environ.* **2014**, *154*, 102–114. [CrossRef]
12. Lindberg, E.; Hollaus, M. Comparison of methods for estimation of stem volume, stem number and basal area from airborne laser scanning data in a hemi-boreal forest. *Remote Sens.* **2012**, *4*, 1004–1023. [CrossRef]
13. Wijaya, A.; Kusnadi, S.; Gloaguen, R.; Heilmeier, H. Improved strategy for estimating stem volume and forest biomass using moderate resolution remote sensing data and GIS. *J. For. Res.* **2010**, *21*, 1–12. [CrossRef]
14. Chen, Q.; Laurin, G.V.; Battles, J.J.; Saah, D. Integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground live biomass. *Remote Sens. Environ.* **2012**, *121*, 108–117. [CrossRef]
15. García-Gutiérrez, J.; Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing* **2015**, *167*, 24–31. [CrossRef]
16. Zhang, Y.; Shao, Z. Assessing of urban vegetation biomass in combination with LiDAR and high-resolution remote sensing images. *Int. J. Remote Sens.* **2021**, *42*, 964–985. [CrossRef]
17. Fu, L.; Liu, Q.; Sun, H.; Wang, Q.; Li, Z.; Chen, E.; Pang, Y.; Song, X.; Wang, G. Development of a system of compatible individual tree diameter and aboveground biomass prediction models using error-in-variable regression and airborne LiDAR data. *Remote Sens.* **2018**, *10*, 325. [CrossRef]
18. Li, J.; Roy, D.P. A global analysis of Sentinel-2A, Sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sens.* **2017**, *9*, 902. [CrossRef]
19. Persson, M.; Lindberg, E.; Reese, H. Tree species classification with multi-temporal Sentinel-2 data. *Remote Sens.* **2018**, *10*, 1794. [CrossRef]
20. Hościło, A.; Lewandowska, A. Mapping forest type and tree species on a regional scale using multi-temporal Sentinel-2 data. *Remote Sens.* **2019**, *11*, 929. [CrossRef]
21. Pandit, S.; Tsuyuki, S.; Dube, T. Estimating above-ground biomass in sub-tropical buffer zone community forests, Nepal, using Sentinel 2 data. *Remote Sens.* **2018**, *10*, 601. [CrossRef]
22. Zarco-Tejada, P.J.; Hornero, A.; Beck, P.; Kattenborn, T.; Kempeneers, P.; Hernández-Clemente, R. Chlorophyll content estimation in an open-canopy conifer forest with Sentinel-2A and hyperspectral imagery in the context of forest decline. *Remote Sens. Environ.* **2019**, *223*, 320–335. [CrossRef]
23. Lima, T.A.; Beuchle, R.; Langner, A.; Grecchi, R.C.; Griess, V.C.; Achard, F. Comparing Sentinel-2 MSI and Landsat 8 OLI imagery for monitoring selective logging in the Brazilian Amazon. *Remote Sens.* **2019**, *11*, 961. [CrossRef]
24. Grabska, E.; Hostert, P.; Pflugmacher, D.; Ostapowicz, K. Forest stand species mapping using the Sentinel-2 time series. *Remote Sens.* **2019**, *11*, 1197. [CrossRef]
25. Condés, S.; McRoberts, R.E. Updating national forest inventory estimates of growing stock volume using hybrid inference. *For. Ecol. Manag.* **2017**, *400*, 48–57. [CrossRef]
26. Chrysafis, I.; Mallinis, G.; Siachalou, S.; Patias, P. Assessing the relationships between growing stock volume and Sentinel-2 imagery in a Mediterranean forest ecosystem. *Remote Sens. Lett.* **2017**, *8*, 508–517. [CrossRef]

27. Laurin, G.V.; Balling, J.; Corona, P.; Mattioli, W.; Papale, D.; Puletti, N.; Rizzo, M.; Truckenbrodt, J.; Urban, M. Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. *J. Appl. Remote Sens.* **2018**, *12*, 016008. [CrossRef]

28. Torbick, N.; Ledoux, L.; Salas, W.; Zhao, M. Regional mapping of plantation extent using multisensor imagery. *Remote Sens.* **2016**, *8*, 236. [CrossRef]

29. Vafaei, S.; Soosani, J.; Adeli, K.; Fadaei, H.; Naghavi, H.; Pham, T.D.; Tien Bui, D. Improving accuracy estimation of Forest Aboveground Biomass based on incorporation of ALOS-2 PALSAR-2 and Sentinel-2A imagery and machine learning: A case study of the Hyrcanian forest area (Iran). *Remote Sens.* **2018**, *10*, 172. [CrossRef]

30. Mauya, E.W.; Koskinen, J.; Tegel, K.; Hämäläinen, J.; Kauranne, T.; Käyhkö, N. Modelling and predicting the growing stock volume in small-scale plantation forests of Tanzania using multi-sensor image synergy. *Forests* **2019**, *10*, 279. [CrossRef]

31. Pham, T.D.; Yokoya, N.; Bui, D.T.; Yoshino, K.; Friess, D.A. Remote sensing approaches for monitoring mangrove species, structure, and biomass: Opportunities and challenges. *Remote Sens.* **2019**, *11*, 230. [CrossRef]

32. Mura, M.; Bottalico, F.; Giannetti, F.; Bertani, R.; Giannini, R.; Mancini, M.; Orlandini, S.; Travaglini, D.; Chirici, G. Exploiting the capabilities of the Sentinel-2 multi spectral instrument for predicting growing stock volume in forest ecosystems. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *66*, 126–134. [CrossRef]

33. Liu, Y.; Gong, W.; Xing, Y.; Hu, X.; Gong, J. Estimation of the forest stand mean height and aboveground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery. *ISPRS J. Photogramm. Remote Sens.* **2019**, *151*, 277–289. [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. Jiang, F.; Kutia, M.; Sarkissian, A.J.; Lin, H.; Long, J.; Sun, H.; Wang, G. Estimating the growing stem volume of coniferous plantations based on random forest using an optimized variable selection method. *Sensors* **2020**, *20*, 7248. [CrossRef]

36. Schepaschenko, D.; Chave, J.; Phillips, O.L.; Lewis, S.L.; Davies, S.J.; Réjou-Méchain, M.; Sist, P.; Scipal, K.; Perger, C.; Herault, B.; et al. The Forest Observation System, building a global reference dataset for remote sensing of forest biomass. *Sci. Data* **2019**, *6*, 198. [CrossRef]

37. Abdi, E.; Mariv, H.S.; Deljouei, A.; Sohrabi, H. Accuracy and precision of consumer-grade GPS positioning in an urban green space environment. *For. Sci. Technol.* **2014**, *10*, 141–147. [CrossRef]

38. Huang, H.; Chen, Y.; Clinton, N.; Wang, J.; Wang, X.; Liu, C.; Gong, P.; Yang, J.; Bai, Y.; Zheng, Y.; et al. Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. *Remote Sens. Environ.* **2017**, *202*, 166–176. [CrossRef]

39. Zhu, Y.; Feng, Z.; Lu, J.; Liu, J. Estimation of forest biomass in Beijing (China) using multisource remote sensing and forest inventory data. *Forests* **2020**, *11*, 163. [CrossRef]

40. Mukherjee, S.; Joshi, P.K.; Mukherjee, S.; Ghosh, A.; Garg, R.; Mukhopadhyay, A. Evaluation of vertical accuracy of open source Digital Elevation Model (DEM). *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 205–217. [CrossRef]

41. Quintano, C.; Fernández-Manso, A.; Fernández-Manso, O. Combination of Landsat and Sentinel-2 MSI data for initial assessing of burn severity. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 221–225. [CrossRef]

42. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]

43. Zheng, H.; Du, P.; Chen, J.; Xia, J.; Li, E.; Xu, Z.; Li, X.; Yokoya, N. Performance evaluation of downscaling Sentinel-2 imagery for land use and land cover classification by spectral-spatial features. *Remote Sens.* **2017**, *9*, 1274. [CrossRef]

44. Mallinis, G.; Mitsopoulos, I.; Chrysafi, I. Evaluating and comparing Sentinel 2A and Landsat-8 Operational Land Imager (OLI) spectral indices for estimating fire severity in a Mediterranean pine ecosystem of Greece. *GISci. Remote Sens.* **2018**, *55*, 1–18. [CrossRef]

45. Asner, G.P.; Keller, M.; Pereira, R., Jr.; Zweede, J.C. Remote sensing of selective logging in Amazonia: Assessing limitations based on detailed field observations, Landsat ETM+, and textural analysis. *Remote Sens. Environ.* **2002**, *80*, 483–496. [CrossRef]

46. Iqbal, N.; Mumtaz, R.; Shafi, U.; Zaidi, S.M.H. Gray level co-occurrence matrix (GLCM) texture based crop classification using low altitude remote sensing platforms. *PeerJ Comput. Sci.* **2021**, *7*, e536. [CrossRef] [PubMed]

47. She, X.; Zhang, L.; Cen, Y.; Wu, T.; Huang, C.; Baig, M.H.A. Comparison of the continuity of vegetation indices derived from Landsat 8 OLI and Landsat 7 ETM+ data among different vegetation types. *Remote Sens.* **2015**, *7*, 13485–13506. [CrossRef]

48. Muhsoni, F.F.; Sambah, A.B.; Mahmudi, M.; Wiadnya, D. Comparison of different vegetation indices for assessing mangrove density using sentinel-2 imagery. *GEOMATE J.* **2018**, *14*, 42–51.

49. Hu, Y.; Xu, X.; Wu, F.; Sun, Z.; Xia, H.; Meng, Q.; Huang, W.; Zhou, H.; Gao, J.; Li, W.; et al. Estimating forest stock volume in Hunan Province, China, by integrating in situ plot data, Sentinel-2 images, and linear and machine learning regression models. *Remote Sens.* **2020**, *12*, 186. [CrossRef]

50. Singh, D.; Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **2020**, *97*, 105524. [CrossRef]

51. Cohen, I.; Huang, Y.; Chen, J.; Benesty, J.; Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009, pp. 1–4.

52. Zheng, S.; Cao, C.; Dang, Y.; Xiang, H.; Zhao, J.; Zhang, Y.; Wang, X.; Guo, H. Retrieval of forest growing stock volume by two different methods using Landsat TM images. *Int. J. Remote Sens.* **2014**, *35*, 29–43. [CrossRef]

53. Senan, E.M.; Al-Adhaileh, M.H.; Alsaade, F.W.; Aldhyani, T.H.; Alqarni, A.A.; Alsharif, N.; Uddin, M.I.; Alahmadi, A.H.; Jadhav, M.E.; Alzahrani, M.Y. Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques. *J. Healthc. Eng.* **2021**, *2021*, 1004767. [CrossRef]

54. Adame-Campos, R.L.; Ghilardi, A.; Gao, Y.; Paneque-Gálvez, J.; Mas, J.F. Variables selection for aboveground biomass estimations using satellite data: A comparison between relative importance approach and stepwise Akaike's information criterion. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 245. [CrossRef]

55. Niu, W.J.; Feng, Z.K.; Feng, B.F.; Min, Y.W.; Cheng, C.T.; Zhou, J.Z. Comparison of multiple linear regression, artificial neural network, extreme learning machine, and support vector machine in deriving operation rule of hydropower reservoir. *Water* **2019**, *11*, 88. [CrossRef]

56. Wicki, A.; Parlow, E. Multiple regression analysis for unmixing of surface temperature data in an urban environment. *Remote Sens.* **2017**, *9*, 684. [CrossRef]

57. Uyanık, G.K.; Güler, N. A study on multiple linear regression analysis. *Procedia-Soc. Behav. Sci.* **2013**, *106*, 234–240. [CrossRef]

58. Li, J.; Cheng, J.H.; Shi, J.Y.; Huang, F. Brief introduction of back propagation (BP) neural network algorithm and its improvement. In *Advances in Computer Science and Information Engineering: Volume 2*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 553–558.

59. Yang, F.; Cho, H.; Zhang, H.; Zhang, J.; Wu, Y. Artificial neural network (ANN) based prediction and optimization of an organic Rankine cycle (ORC) for diesel engine waste heat recovery. *Energy Convers. Manag.* **2018**, *164*, 15–26. [CrossRef]

60. Chen, L.; Wang, Y.; Ren, C.; Zhang, B.; Wang, Z. Optimal combination of predictors and algorithms for forest above-ground biomass mapping from Sentinel and SRTM data. *Remote Sens.* **2019**, *11*, 414. [CrossRef]

61. Dube, T.; Mutanga, O.; Adam, E.; Ismail, R. Intra-and-inter species biomass prediction in a plantation forest: Testing the utility of high spatial resolution spaceborne multispectral rapideye sensor and advanced machine learning algorithms. *Sensors* **2014**, *14*, 15348–15370. [CrossRef]

62. Kilham, P.; Hartebrodt, C.; Kändler, G. Generating tree-level harvest predictions from forest inventories with random forests. *Forests* **2018**, *10*, 20. [CrossRef]

63. Ou, Q.; Lei, X.; Shen, C. Individual tree diameter growth models of larch–spruce–fir mixed forests based on machine learning algorithms. *Forests* **2019**, *10*, 187. [CrossRef]

64. Pullanagari, R.R.; Kereszturi, G.; Yule, I. Integrating airborne hyperspectral, topographic, and soil data for estimating pasture quality using recursive feature elimination with random forest regression. *Remote Sens.* **2018**, *10*, 1117. [CrossRef]

65. Soriano-Luna, M.d.l.Á.; Ángeles-Pérez, G.; Guevara, M.; Birdsey, R.; Pan, Y.; Vaquera-Huerta, H.; Valdez-Lazalde, J.R.; Johnson, K.D.; Vargas, R. Determinants of above-ground biomass and its spatial variability in a temperate forest managed for timber production. *Forests* **2018**, *9*, 490. [CrossRef]

66. Korhonen, L.; 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]

67. Immitzer, M.; Vuolo, F.; Atzberger, C. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]

68. Lin, S.; Li, J.; Liu, Q.; Li, L.; Zhao, J.; Yu, W. Evaluating the effectiveness of using vegetation indices based on red-edge reflectance from Sentinel-2 to estimate gross primary productivity. *Remote Sens.* **2019**, *11*, 1303. [CrossRef]

69. Haapanen, R.; Tuominen, S. Data combination and feature selection for multi-source forest inventory. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 869–880. [CrossRef]

70. Darst, B.F.; Malecki, K.C.; Engelman, C.D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet.* **2018**, *19*, 1–6. [CrossRef]

71. Attarzadeh, R.; Amini, J. Towards an object-based multi-scale soil moisture product using coupled Sentinel-1 and Sentinel-2 data. *Remote Sens. Lett.* **2019**, *10*, 619–628. [CrossRef]

72. Poortinga, A.; Tenneson, K.; Shapiro, A.; Nquyen, Q.; San Aung, K.; Chishtie, F.; Saah, D. Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification. *Remote Sens.* **2019**, *11*, 831. [CrossRef]