

## Article

# Above-Ground Biomass Estimation of Plantation with Complex Forest Stand Structure Using Multiple Features from Airborne Laser Scanning Point Cloud Data

Linghan Gao<sup>1,2,3,4</sup> and Xiaoli Zhang<sup>1,2,\*</sup> 

<sup>1</sup> Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, China; gaolinghan6868@163.com

<sup>2</sup> The Key Laboratory for Silviculture and Conservation of Ministry of Education, Beijing Forestry University, Beijing 100083, China

<sup>3</sup> School of Remote Sensing and Information Engineering, North China Institute of Aerospace Engineering, Langfang 065000, China

<sup>4</sup> Hebei Collaborative Innovation Center for Aerospace Remote Sensing Information Processing and Application, Langfang 065000, China

\* Correspondence: Zhang-xl@263.net

**Abstract:** Accurate forest above-ground biomass (AGB) estimation is important for dynamic monitoring of forest resources and evaluation of forest carbon sequestration capacity. However, it is difficult to depict the forest's vertical structure and its heterogeneity using optical remote sensing when estimating forest AGB, for the reason that electromagnetic waves cannot penetrate the canopy's surface to obtain low vegetation information, especially in subtropical and tropical forests with complex layer structure and tree species composition. As an active remote sensing technology, an airborne laser scanner (ALS) can penetrate the canopy surface to obtain three-dimensional structure information related to AGB. This paper takes the Jiepai sub-forest farm and the Gaofeng state-owned forest farm in southern China as the experimental area and explores the optimal features from the ALS point cloud data and AGB inversion model in the subtropical forest with complex tree species composition and structure. Firstly, considering tree canopy structure, terrain features, point cloud structure and density features, 63 point cloud features were extracted. In view of the biomass distribution differences of different tree species, the random forest (RF) method was used to select the optimal features of each tree species. Secondly, four modeling methods were used to establish the AGB estimation models of each tree species and verify their accuracy. The results showed that the features related to tree height had a great impact on forest AGB. The top features of *Cunninghamia Lanceolata* (Chinese fir) and *Eucalyptus* are all related to height, *Pinus* (pine tree) is also related to terrain features and other broadleaved trees are also related to point cloud density features. The accuracy of the stepwise regression model is best with the AGB estimation accuracy of 0.19, 0.76, 0.71 and 0.40, respectively, for the Chinese fir, pine tree, eucalyptus and other broadleaved trees. In conclusion, the proposed linear regression AGB estimation model of each tree species combining different features derived from ALS point cloud data has high applicability, which can provide effective support for more accurate forest AGB and carbon stock inventory and monitoring.

**Keywords:** above-ground biomass (AGB); airborne laser scanner (ALS); feature extraction; random forest; estimation models



**Citation:** Gao, L.; Zhang, X. Above-Ground Biomass Estimation of Plantation with Complex Forest Stand Structure Using Multiple Features from Airborne Laser Scanning Point Cloud Data. *Forests* **2021**, *12*, 1713. <https://doi.org/10.3390/f12121713>

Academic Editor: Steven L. Petersen

Received: 18 October 2021

Accepted: 2 December 2021

Published: 6 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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

Forests have the highest biological storage capacity and play a key role in the global carbon cycle. As a natural and renewable resource, the forest also plays an irreplaceable role in maintaining global ecological balance and biodiversity [1,2]. Forest biomass is not only an important indicator of terrestrial ecosystem function evaluation, but also an important parameter of forest carbon sink assessment [3]. Therefore, accurate assessment of the forest

biomass is beneficial to quantify forest carbon storage, and can provide a key reference for forest resource management. However, the lack of forest biomass information is one of the uncertain factors of the global carbon budget [4]. China has a large area of plantations which has high carbon sequestration potential. Real-time and accurate monitoring of the forest biomass of China's plantations will not only promote the assessment accuracy of forest carbon storage, but it also has important significance for evaluating the sustainable development potential of China's plantations.

The traditional forest biomass survey is mainly based on the ground measurement. Its low efficiency and high cost are not conducive to large-scale forest biomass estimation. With the development of remote sensing technology, the limitations of traditional forest biomass estimation, such as small scale, destructive to vegetation and discontinuity, have been changed [5]. Using optical imaging, microwave remote sensing and LiDAR (Light Detection and Ranging) to estimate forest AGB has gradually become the main technical means for biomass research. The penetration of optical remote sensing to ground objects is weak, and it can only reflect or absorb the electromagnetic radiation of the top layer of the canopy. It is difficult to obtain the vertical structure parameters of vegetation and, inevitably, the signal is easily saturated [6]. Microwave radar has penetrability, which can obtain the three-dimensional structure information of the forest. However, it is easy to be affected by topography and saturation of the back-scattering signal. It also has some limitations for a large area of biomass estimation [7].

As an active remote technology, LiDAR can acquire data in a short period and has high range resolution, angular resolution and strong anti-jamming ability. It has incomparable advantages over optical remote sensing in obtaining forest vertical structure parameters closely related to biomass [8,9]. The laser pulse emitted by LiDAR can penetrate the canopy to obtain information on the terrain under the forest canopy [10], which can be used to extract the tree height and canopy height, forest coverage, density and the situation of trees with different ages. In addition, LiDAR has strong anti-interference ability and can obtain biophysical and structural parameters of the trees according to pulse intensity information or full waveform information. Many studies have shown that both small footprint LiDAR data and waveform LiDAR data can achieve high accuracy in forest biomass estimation [11–13].

Airborne LiDAR (ALS) can obtain the vertical structure information of trees through the canopy, so as to obtain information on tree height and laser point density, but it cannot directly obtain the information on forest biomass. Most of the research modeling AGB uses the correlation between the point cloud characteristics and measured biomass [14–18]. Kronseder et al. [13] used LiDAR point cloud data to estimate the AGB of two different tropical rainforest tree species in Central Kalimantan, Indonesia, and established a forest aboveground biomass model using a multiple stepwise regression method. The results showed that the relative root mean square error of two different tree species was less than 42%, and the relative root mean square error of the total aboveground biomass was 33.85%. He et al. [19] used low-density airborne LiDAR point cloud data to extract the characteristic variables, such as percentile height and density of point cloud data, and estimated the total AGB and its components' biomass estimation models of *Picea crassifolia* forest in Qilian Mountains using the multiple stepwise regression method. The results showed that the estimation accuracy of AGB was high, the adjusted  $R^2$  was 0.727, the root mean square error was 15.237 mg/ha and the relative root mean square error was 14.163%. The estimation accuracy of leaf biomass was relatively low, and the adjusted  $R^2$  was 0.356. Shao et al. [20] used LiDAR data to estimate the AGB of temperate broadleaved forest. The results showed that the accuracy of the model was high, with  $R^2$  being 0.81. Jinbo Lu et al. [21] used airborne and backpack LiDAR to estimate the aboveground biomass of a *Robinia pseudoacacia* plantation in the Yellow River Delta of China, extracted the point cloud height information and intensity information and constructed the aboveground biomass model using multiple linear regression and random forest methods. The results showed that the estimation accuracy of aboveground biomass of healthy forest, medium

dieback forest and severe dieback forest were 0.72, 0.77 and 0.70, respectively, using multiple linear regression. The accuracy of using random forest was 0.92, 0.95 and 0.91, respectively. Salum RB et al. [22] used LiDAR point cloud data combined with a forest map to study mangrove biomass in Guaras Island, Brazil, and the relationship between tree height and aboveground biomass was established. The  $R^2$  was between 0.85 and 0.92, which showed that the tree height obtained by LiDAR could replace the traditional method to estimate forest aboveground biomass.

In conclusion, ALS data can obtain the information of forest vertical structure and can directly obtain the information of vegetation height and canopy, which has incomparable advantages in forest AGB estimation. At the same time, ALS data contain abundant point cloud density and structure information, which makes it possible to explore much more information. At present, most of the existing biomass studies on stand scale do not distinguish tree species, study only a special tree species or only distinguish coniferous forest and broad-leaved forest. This makes the estimation of AGB produce great errors in the forest with a complex forest structure. Secondly, only canopy height and quantile tree height parameters are extracted from ALS point cloud data, and few ecological parameters are applied, such as leaf area index and porosity. These ecological parameters will also affect the tree growth in the actual forest environment and indirectly affect the quality of AGB. For the plantation stand with complex structure, the structural characteristics of different tree species will lead to the difference of point cloud characteristics, which indicates that the characteristic variables and estimation models of biomass may be different due to different tree species, which is also the necessity of extracting characteristic variables and modeling by distinguishing tree species.

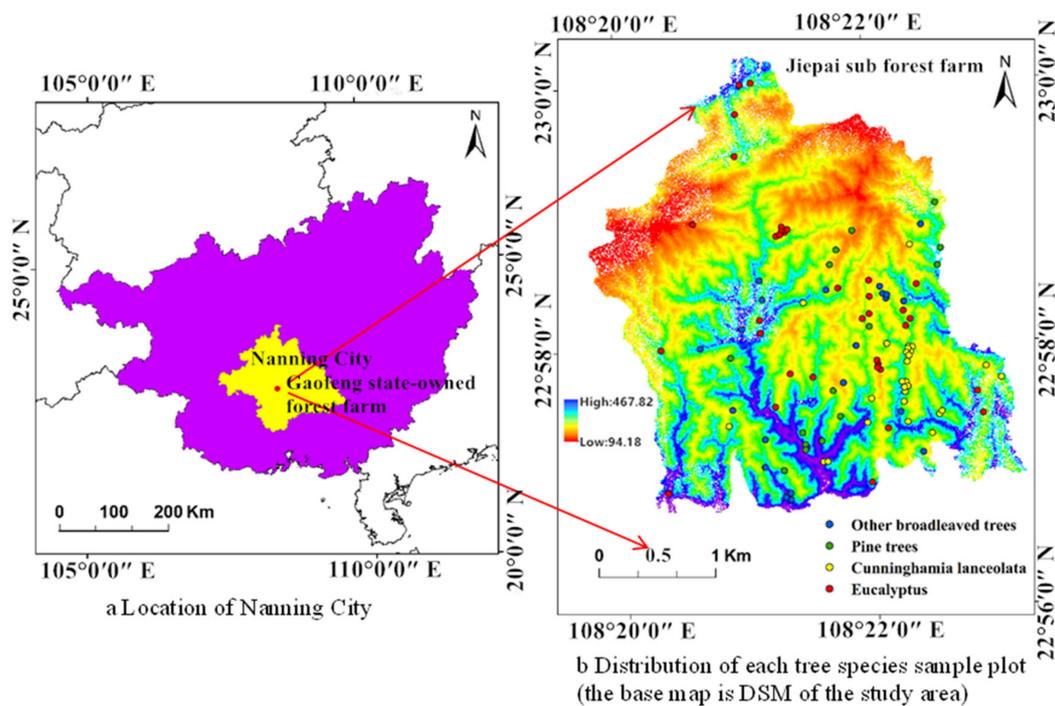
Therefore, this study takes the typical plantation in southern China as the research object, distinguishes tree species and completes the regional biomass mapping. By extracting multiple features variables from the point cloud data, including height, density, terrain and ecological parameters, more accurate biomass models for different tree species are established, and AGB mapping with the sub-compartment unit is implemented, which provide reference for forest scientific management and ecological assessment.

## 2. Materials and Methods

### 2.1. Materials

#### 2.1.1. Study Area

The study area is located in the Jiepai sub-forest farm of Gaofeng state-owned forest farm in southern China ( $22^{\circ}56'41''$ – $23^{\circ}0'21''$  N,  $108^{\circ}19'47''$ – $109^{\circ}23'16''$  E, Figure 1), covering 300 ha with 87% plantation forest. It is located in a subtropical humid monsoon climate and hilly landform with an altitude of 100–500 m and a slope of  $20^{\circ}$ – $35^{\circ}$ . The annual average temperature, precipitation and relative humidity are  $21.6^{\circ}\text{C}$ , 1304.2 mm and 80%, respectively. The proportion of non-forest land and forest land is 1:99, so the vegetation coverage rate is very high and contains a very high biomass. The main tree species in the plantation forest are *Cunninghamia Lanceolata*, *Pinus*, *Eucalyptus* and other broadleaved trees (*Castanopsis Hystrix*, *Lillicium Verum* and *Magnolia Denudata*). This forest farm is a typical representative of southern forests in China with complex forest composition and structure. The saturation point of the biomass estimated by optical remote sensing is too low. Therefore, it is very necessary to use airborne LiDAR to estimate the aboveground biomass of different tree species.



**Figure 1.** Location of the study area (a) is the location of Nanning City. (b) is the distribution of each tree species sample plot.

### 2.1.2. Field Data

The field data include the sub-compartment inventory data from 2014 and the sample plot data collected from January to February in 2018. Before the sample plot survey, Chinese GF-2 image was visually interpreted to determine the survey area and the sampling plot's location. Considering the terrain characteristics, tree species distribution and the accessibility of ground investigation, different size sample plots were set up according to the actual situation. The  $20 \times 20 \text{ m}^2$  and  $25 \times 25 \text{ m}^2$  sample plots were used for the study of single tree classification and single tree parameter estimation. The  $30 \times 30 \text{ m}^2$  sample plots were used to estimate forest stand parameters, such as canopy density and forest biomass. For  $20 \times 20 \text{ m}^2$  and  $25 \times 25 \text{ m}^2$  plots, each tree with a diameter at breast height (DBH) greater than 5 cm was measured. Real-time kinematic (RTK) Zenith 15pro was used to measure the location of the four corners and center of each plot. The total station STS-772 was used to measure the location of each tree in the plots. The laser altimeter was used to measure the height of each tree. The DBH of each tree was measured with a DBH ruler, and crown width of each tree from east to west and north to south was measured with a tape. For the plots of  $30 \times 30 \text{ m}^2$ , RTK was used to measure the locations of the northwest and southeast corners of the plots, and sample trees were randomly selected from different DBH classes and the heights and DBH were measured, then the average tree height and DBH of the plots were calculated. Detailed plot information of different tree species is shown in Table 1. The distribution of sample plots is shown in Figure 1.

The AGB of Chinese fir and Eucalyptus was calculated using the relative growth equation proposed by Wen Yuanguang et al. [23,24]. The AGB of pine tree used the relative growth equation proposed by Du Hu et al. [25]. The AGB of other broadleaved trees used the general formula of Chinese Broadleaved Trees in the handbook of main forest biomass models of China written by Luo Yunjian et al. [26]. In this study, the sample data were divided into training data and testing data. For tree species with more than 20 samples, samples were divided to the ratio of 2:1 of training and testing samples, and the training and testing samples were evenly distributed in the study area. Meanwhile, the number of testing samples should not be less than 10. For the tree species with less than 20 samples, the data were trained and tested by the leave-one method. The specific sample plots information is shown in Table 1.

Table 1. Sample plots information.

Plot No.	Species	Plot Size (m <sup>2</sup> )	Inventoried Trees (All Trees/Sample Trees)	Forest Type	Training/Testing Dataset	Number of Plots	Diameter at Breast Height (DBH cm)	Tree Height (m)	Stem Density (n·ha <sup>-1</sup> )	AGB (t·ha <sup>-1</sup> )
1–27	Chinese fir	20 × 20	All trees	Middle	Training	2	17.9 ± 11.9	13 ± 6.5	1200 ± 0	92.8 ± 5.5
					Testing	4	22.8 ± 18.1	15.6 ± 10	900 ± 300	87.8 ± 2.7
		25 × 25	All trees	Middle	Training	6	28.1 ± 20.8	19.5 ± 12.6	624 ± 160	83.8 ± 24
					Testing	4	22.1 ± 8.7	16.6 ± 4.8	656 ± 96	82.7 ± 2.8
		30 × 30	Sample trees	Middle	Training	9	15.5 ± 4.4	13.6 ± 2.6	2272 ± 1461	82.7 ± 22.2
					Testing	2	14.6 ± 1	13.4 ± 1.5	1744 ± 389	88.5 ± 2.5
28–42	Pine tree	30 × 30	Sample trees	Middle	Training/ testing	6	22 ± 3.9	16.8 ± 1.5	1211 ± 667	139.7 ± 36.1
				Young	Training/ testing	9	15.2 ± 2.5	9.4 ± 3.3	1333 ± 345	126.4 ± 73.2
43–77	Eucalyptus	20 × 20	All trees	Middle	Training	1	17.4 ± 8.4	21.3 ± 4.4	1300	185.7
				Young	Testing	2	23.1 ± 9.4	24.5 ± 10	937 ± 163	301 ± 37.8
				Young	Training	8	10.3 ± 7.3	14.2 ± 8.6	1887 ± 838	67.3 ± 54.2
				Young	Testing	7	12.3 ± 7.1	11.3 ± 10	1537 ± 188	58 ± 33.5
		25 × 25	All trees	Middle	Training	7	17 ± 11.5	20.8 ± 11.8	1232 ± 656	204.2 ± 91
				Young	Training	3	12.1 ± 7.7	13.1 ± 6.3	1608 ± 616	89.5 ± 34
		30 × 30	Sample trees	Young	Training	4	10.4 ± 0.8	15.2 ± 1.9	1855 ± 423	85.1 ± 32.8
					Testing	3	10.1 ± 2.4	14.5 ± 2.7	2177 ± 467	82.8 ± 33.3
78–98	Other broadleaved trees	25 × 25	All trees	Middle	Training	3	18.7 ± 14.4	13.1 ± 9	1192 ± 312	209.6 ± 81.2
					Testing	3	28.5 ± 21.5	15.8 ± 9.8	944 ± 240	227.7 ± 21.3
		30 × 30	Sample trees	Middle	Training	8	14.5 ± 5.3	13.9 ± 5.6	1233 ± 478	164.5 ± 112.4
					Testing	7	18.1 ± 5	14.7 ± 4.1	1372 ± 806	172.3 ± 31.8

Note:  $m \pm n$ ,  $m$  is the median of the tree parameters for each tree species,  $n$  is the maximum value by which this parameter fluctuates up or down.

### 2.1.3. ALS Data

The ALS scanning of the study area was performed in February 2018, with RIEGL LMS-Q680i laser scanning system carried by manned aircraft of fixed wings. The data format is LAS format, and the average cloud density of plantation is 3.35 points/m<sup>2</sup>. The detailed scanning parameters of ALS are shown in Table 2.

**Table 2.** Laser scanning system parameters.

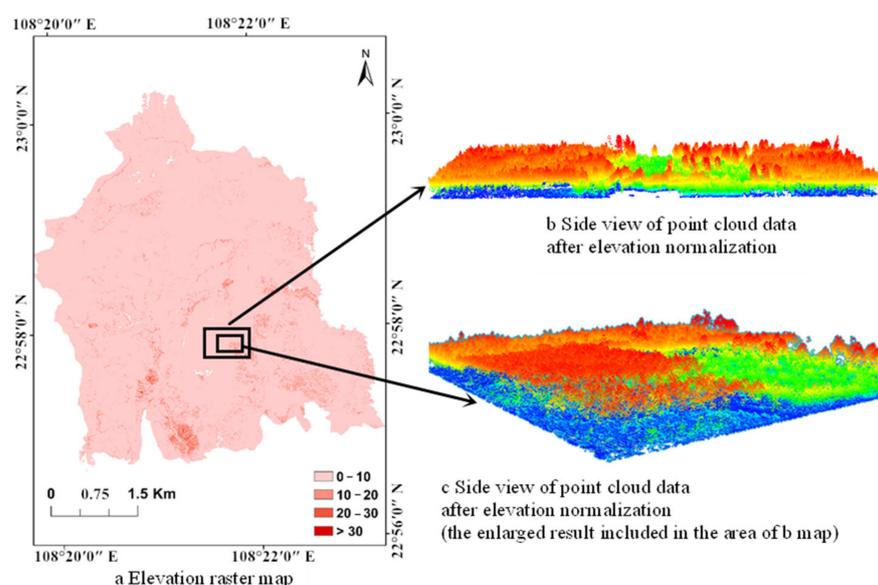
Parameters	Value
Wavelength (nm)	1550
Divergence angle (mrad)	0.5
Step length (cm)	45
Pulse repetition rate (KHz)	360
Scanning rate (Hz)	112
Width (m)	1040
Flight altitude (m)	900
Flight speed (m/s)	55
Side overlap	65%
Average point spacing (m)	0.45 × 0.45

### 2.1.4. ALS Data Processing

The original ALS point cloud data are composed of hundreds of Las format files and the data range is large, so corresponding splicing and clipping processes are needed. Secondly, due to the reflection of objects and the absorption of laser in LiDAR data acquisition, some ground points have no obvious echo signal, such that the ranging value cannot be obtained. In addition, due to the birds, wires and local terrain, abnormal distance values will also be generated in the data. Those points whose ranging values are far greater than the flight altitude or points with extremely small ranging values are called the outlier points or noise points [27,28]. This paper uses the height threshold method to remove the noise points [29]. Meanwhile, ground point and non-ground point are distinguished by the height setting. The specific parameters of using height threshold for denoising and classification filtering are shown in Table 3. Then, based on the ground points and non-ground points, the point cloud height normalization processing is carried out to eliminate the influence of terrain undulation on the elevation of surface features, and the real height information of surface features can be obtained. The triangulated irregular network (TIN) algorithm [30–32] is used to interpolate the ground points into a digital elevation model (DEM) [33]. Kriging algorithm [34] is used to interpolate non-ground points into a digital surface model (DSM). Finally, the difference between DSM and DEM is processed to generate the canopy height model (CHM), so as to complete the point cloud height normalization (Figure 2).

**Table 3.** Height threshold parameters of classification and filtering.

Category	Search Radius (m)	Height Difference (m)	Minimum Height (m)	Maximum Height (m)	Maximum Terrain Slope Angle (°)	Iteration Angle (°)	Iteration Distance (m)
High point	3	>5					
Low point	3	<0.5					
Ground point					88	8	1.5
Surface vegetation point			0	0.3			
Forest point			0.3	50			



**Figure 2.** Elevation normalization of point cloud (a) is the elevation raster map. (b) is the side view of point cloud data after elevation normalization. (c) is the enlarged result of (b).

## 2.2. Methods

### 2.2.1. Feature Variables Extraction

According to the structure features of point cloud data, considering the ecological index and spatial structure index, the four categories of feature parameters were derived from point cloud data. For the tree canopy feature, four feature parameters were extracted, such as canopy density and leaf area index. For the terrain features, four feature variables, such as slope and aspect, were extracted. For the point cloud structure features, 45 feature parameters including height percentile, height maximum and height minimum, etc., were extracted. Additionally, ten point cloud density feature parameters at different heights were extracted. In total, 63 feature parameters of point cloud were extracted (Tables 4–7).

**Table 4.** Tree canopy features.

Feature Number	Feature	Meaning	Abbreviation
1	Canopy density [35]	The ratio of vegetation points to total points in a unit grid	C.density
2	Spacing rate [36]	The ratio of ground points to total points in a unit grid	Gap
3	Leaf area index [36]	Half of the leaf surface in a unit grid	LAI
4	Canopy fluctuation rate [35]	$\frac{\text{mean} - \text{min}}{\text{max} - \text{min}}$	H.clr

**Table 5.** Terrain features.

Feature Number	Feature	Meaning	Abbreviation
5	Roughness	The ratio of the surface area to its projected area on a horizontal plane	RS
6	Slope	The steepness of terrain surface based on the DEM	Slope
7	Aspect	The projection direction of slope normal line on a horizontal plane based on the DEM	Aspect
8	Mountain shadow	The brightness of each pixel based on the DEM	Shadow

**Table 6.** Features describing the point cloud vertical distribution using heights.

Feature Number	Feature	Meaning	Abbreviation
9	Mean absolute deviation	$\frac{\sum_{i=1}^n ( Z_i - \text{mean} )}{n}$	H.mean.AD
10	Median absolute deviation	Median of $\sum_{i=1}^n ( Z_i - \text{mean} )$	H.AD.median
11–25	Cumulative Height Percentile	The height sum of X% points in a unit grid	Hc1-Hc99
26	Inter-quartile Range of Cumulative Height	Difference of Hc75 and Hc25	Hc.S
27–41	Height Percentile	The height of X% points in a unit grid	H1-H99
42	Inter-quartile Range	Difference of H75 and H25	H.S
43	Kurtosis	Kurtosis of all point height in a unit grid	H.K
44	Variation Coefficient	$\frac{\text{std}}{\text{mean}}$	H.cv
45	Mean Quadratic Power	$\sqrt{\frac{\sum_{i=1}^n Z_i^2}{n}}$	H.sq.mean
46	Mean Cubic Power	$\sqrt[3]{\frac{\sum_{i=1}^n Z_i^3}{n}}$	H.c.mean
47	Maximum	Max height of all points in a unit grid	Hmax
48	Minimum	Min height of all points in a unit grid	Hmin
49	Mean	Mean height of all points in a unit grid	Hmean
50	Median	Median height of all points in a unit grid	Hmedian
51	Skewness	$\frac{\sum_{i=1}^n (Z_i - \text{mean})^2}{n}$	H.skewness
52	Standard Deviation	Std height of all points in a unit grid	H.std
53	Variance	Var height of all points in a unit grid	H.variance

**Table 7.** Point cloud density features.

Feature Number	Feature	Meaning	Abbreviation
54–63	Point density in each horizontal layer	The point cloud was sliced to ten horizontal layers with the same heights. D1 to D10 corresponded to the point density from the lowest layer to the highest (D1 is the lowest layer)	D1–D10

2.2.2. Feature Variables Selection

The four extracted groups of features may be highly related or redundant, and a large number of feature variables increases the complexity of the model, which will lead to overfitting when the number of the sample plot is small. The Random Forest (RF) algorithm is a popular feature selection method which can realize data reduction and optimization [37,38]. The decline of target prediction accuracy after removing variables is indicated by %IncMSE, which is the growth of root mean square error rate. When the value is larger, the contribution of the variable is greater (Equation (1)). In this study, R software was used to realize RF feature selection.

$$\%IncMSE = \frac{1}{N_{tree}} \sum (err_{OOB} - err_{OOB'}) \tag{1}$$

where: *Ntree* is the number of RF decision trees (set to 1000), *OOB* (Out of Bag) is a randomly selected sample set, *errOOB* is the error of *OOB* when the sample set is not changed, and *errOOB'* is the error of *OOB* when the sample set is changed.

2.2.3. Regression Modeling of AGB

In this study, stepwise regression, ridge regression, principal component regression and nonlinear regression are used for AGB model establishing. Stepwise regression (SR) considers the variance contribution value of all variables when introducing variables and sorts them into a regression equation according to their importance. The final equation does not contain unnecessary independent variables [39,40]. Ridge regression belongs to a kind of biased estimation, which is a supplement to the improved least square method. It is considered to be a better method to deal with collinearity problems [41–44]. Principal component regression (PCR) is another biased estimated method to deal with severe

collinearity [45–47]. The number of principal component factors in a PCR input model depends on the contribution rate and cumulative contribution rate of each principal component to dependent variables. The multiple linear regression model is established using the principal components whose cumulative contribution rate reaches 80% [48,49]. Nonlinear regression refers to the form where the dependent variables and independent variable of the model are not a one-step function, but are represented graphically as curves of various forms. In this study, the nonlinear equation set is shown in Equations (2)–(5) for different tree species:

$$\text{Chinese fir } Y = b_0 + b_1x_1^2 + b_2x_2^2 + b_3x_3^2 + b_4x_4^2 + b_5x_1x_2 + b_6x_1x_3 + b_7x_1x_4 + b_8x_2x_3 + b_9x_2x_4 + b_{10}x_3x_4 + b_{11}x_1x_2x_3 + b_{12}x_1x_2x_4 + b_{13}x_1x_3x_4 + b_{14}x_2x_3x_4 + b_{15}x_1x_2x_3x_4 \quad (2)$$

$$\text{Pine tree } Y = b_0 + b_1x_1^2 + b_2x_2^2 + b_3x_3^2 + b_4x_4^2 + b_5x_1x_2 + b_6x_1x_3 + b_7x_1x_4 + b_8x_2x_3 + b_9x_2x_4 + b_{10}x_3x_4 + b_{11}x_1x_2x_3 + b_{12}x_1x_2x_4 + b_{13}x_1x_3x_4 + b_{14}x_2x_3x_4 + b_{15}x_1x_2x_3x_4 \quad (3)$$

$$\text{Eucalyptus } Y = b_0 + b_1X_1^{b_2} + \dots + b_{12}X_6^{b_{13}} \quad (4)$$

$$\text{Other broadleaved trees } Y = b_0 + b_1x_1^2 + b_2x_2^2 + b_3x_3^2 + b_4x_1x_2 + b_5x_1x_3 + b_6x_2x_3 + b_7x_1x_2x_3 \quad (5)$$

In the above formula,  $b_i$  is constant, and  $x_i$  is the  $i$ th independent variable.

#### 2.2.4. Accuracy Evaluation

The coefficient of determination  $R^2$ , accuracy  $A$ , root mean square error (RMSE) and mean absolute error (MAE) are used in the accuracy evaluation of this study. The training accuracy is expressed by coefficient of determination  $R^2$  using training data. The test accuracy is expressed by  $R^2$ ,  $A$ , RMSE and MAE using testing data. The specific formula is shown in Equations (6)–(9):

$$R^2 = 1 - \frac{\text{mean}(X_{\text{model},i} - X_{\text{obs},i})^2}{\text{mean}(\text{mean}(X_{\text{obs},i}) - X_{\text{obs},i})^2} \quad (6)$$

where:  $R^2$  is the coefficient of determination,  $X_{\text{obs},i}$  is the measured value,  $X_{\text{model},i}$  is the estimated value, and  $\text{mean}$  is the average value.

$$A = 1 - \frac{|X_{\text{obs},i} - X_{\text{model},i}|}{X_{\text{obs},i}} \quad (7)$$

$$\text{RMSE} = \left( \text{mean}(X_{\text{model},i} - X_{\text{obs},i})^2 \right)^{0.5} \quad (8)$$

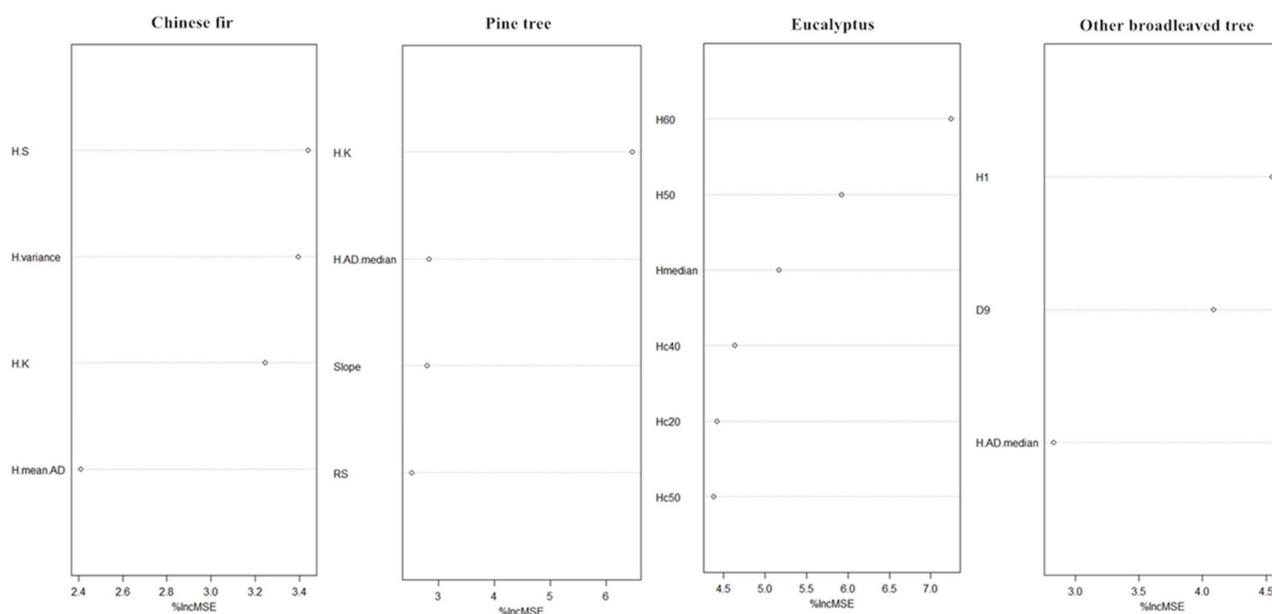
$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |X_{\text{model},i} - X_{\text{obs},i}| \quad (9)$$

where:  $\text{RMSE}$  is the root mean square error,  $\text{MAE}$  is the mean absolute error, and  $N$  is the number of samples.

### 3. Results

#### 3.1. Feature Selection

According to the number of training samples and the principle of moderate proportion, the proportion of training samples and independent variables is set as 4:1. Among the 63 feature variables extracted from the point cloud data, the top four feature variables of importance ranking were selected for modeling for Chinese fir, while the top four, top six and top three feature variables of importance ranking were selected for pine tree, eucalyptus and other broadleaved trees. The importance ranking maps of feature variables for each tree species are shown in Figure 3.



**Figure 3.** Map of feature importance ranking.

From Figure 3, the top four feature variables of Chinese fir are H.S, H.variance, H.K and H.mean.AD. The top four feature variables of pine tree are H.K, H.AD.median, Slope and RS. The top six feature variables of eucalyptus are H60, H50, Hmedian, Hc40, Hc20 and Hc50. The top three feature variables of other broadleaved trees are H1, D9 and H.AD.median. The selected top features of the four tree species are mostly related to the height, which is the key factor to determine AGB. In addition to the correlation with height features, the AGB of pine trees also has a correlation with slope and roughness, which could be explained from the sample plot situations of the pine tree. Among 15 sample plots, 9 plots are young forest and 6 plots are middle age forests, and the terrain is complex with a large slope. Therefore, the estimation results of the pine tree's AGB can be affected by the terrain. The important features of other broadleaved trees include a point cloud density feature which is consistent with various species' composition and the tree crown structure characteristics.

### 3.2. Correlation Analysis

The correlation coefficient matrix was constructed to analyze the correlation between AGB and the importance features. The heat map can directly display the correlation between each variable and AGB. Different colors are used to distinguish the positive and negative of the correlation. The depth of the color reflects the strength of the correlation. The results are shown in Figure 4.

Figure 4 shows that H.S, H.variance and H.mean.AD are moderately correlated with AGB ( $0.40 < |r| < 0.70$ ) among the top four features of Chinese fir. H.K and AGB are minimally correlated ( $0.20 < |r| < 0.40$ ). Slope and RS have relatively low correlation with AGB ( $|r| < 0.20$ ) for pine trees, and H.K has a high correlation with AGB ( $0.70 < |r| < 1.00$ ), while other features are less correlated with the AGB of pine trees. The AGB of eucalyptus is significantly correlated with the six features ( $p < 0.01$ ). Among the three importance features of other broadleaved trees, H.AD.median has a very low correlation with AGB, H1 has a low correlation with AGB and D9 has a moderate correlation with AGB. From the feature correlation of each tree species, it can be seen that the height features of Chinese fir, pine tree and eucalyptus are all moderately and highly correlated with AGB, except that the height features of other broadleaved trees are minimally correlated with AGB. This is because the canopy biomass of other broadleaved trees occupies the main body of the aboveground biomass, while the trunk biomass of the Chinese fir, pine tree and eucalyptus

accounts for a larger proportion, which also highlights the impact of height information on the AGB of different species [50,51].

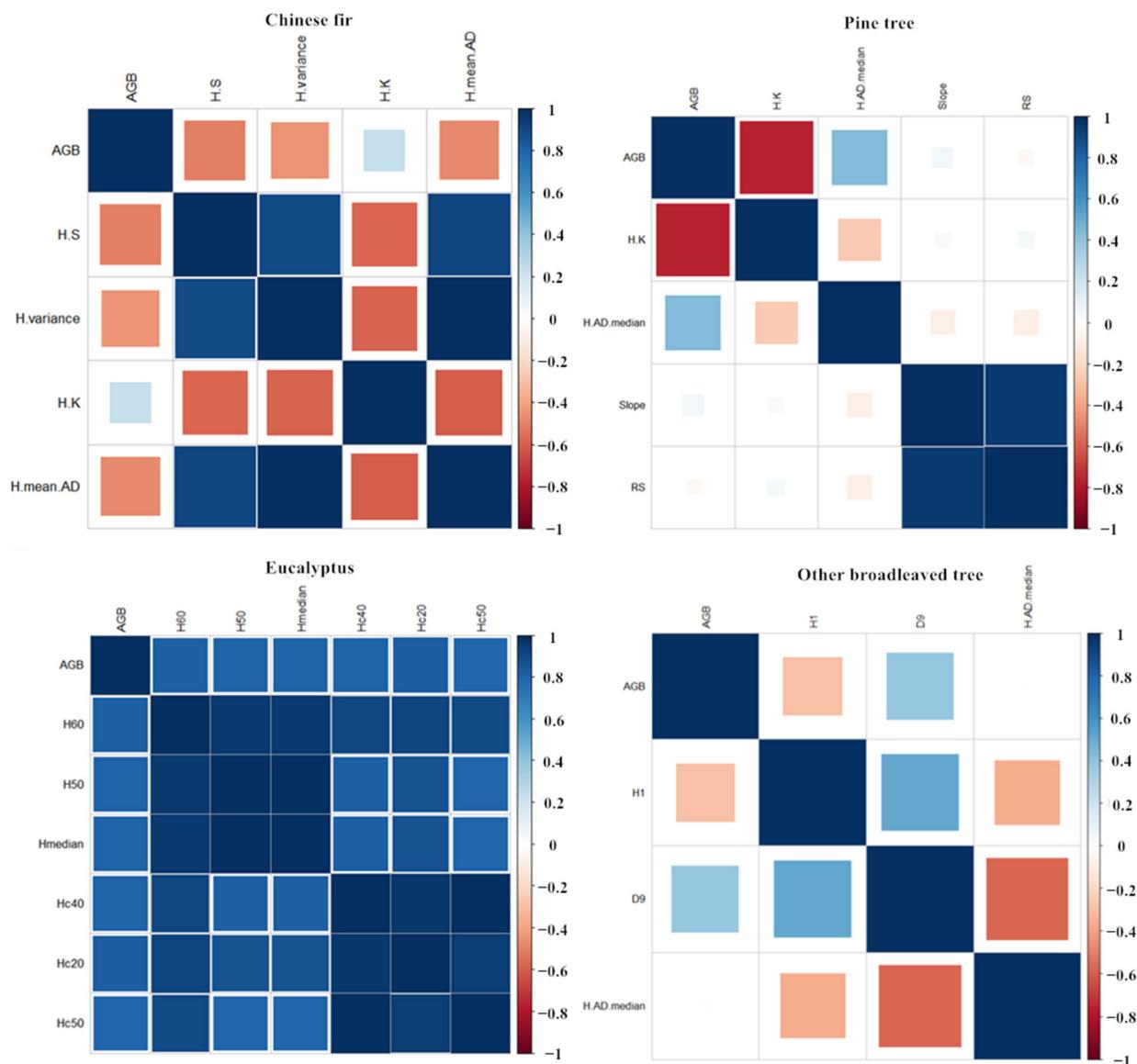


Figure 4. Features correlation of each tree species.

### 3.3. AGB Estimation Models

The aboveground biomass model of each tree species was constructed by using the selected important variables based on the four regression methods. The model formula and accuracy are shown in Table 8.

Table 8 shows that, except for eucalyptus, the training accuracy of the nonlinear model of other species is higher than that of the linear model. The training accuracy of nonlinear model of eucalyptus is 0.56, of other broadleaved trees is 0.79, and of Chinese fir and pine trees is higher than 0.9. The results showed that the nonlinear model is better for AGB estimation of coniferous forest. Among the three linear models, the training accuracy of the stepwise regression model is the highest with the training accuracy of 0.72 for eucalyptus, which is higher than the other two linear models. The ridge regression model of other broadleaved trees has the highest training accuracy (0.52), followed by the stepwise regression model. The training accuracy of the three linear models of Chinese fir is basically the same. The training accuracy of the stepwise regression model of pine trees

is 0.72, which is better than the other two linear models. To sum up, the training accuracy of the nonlinear model is relatively high, followed by the stepwise regression model [52].

**Table 8.** Different regression models and training accuracy.

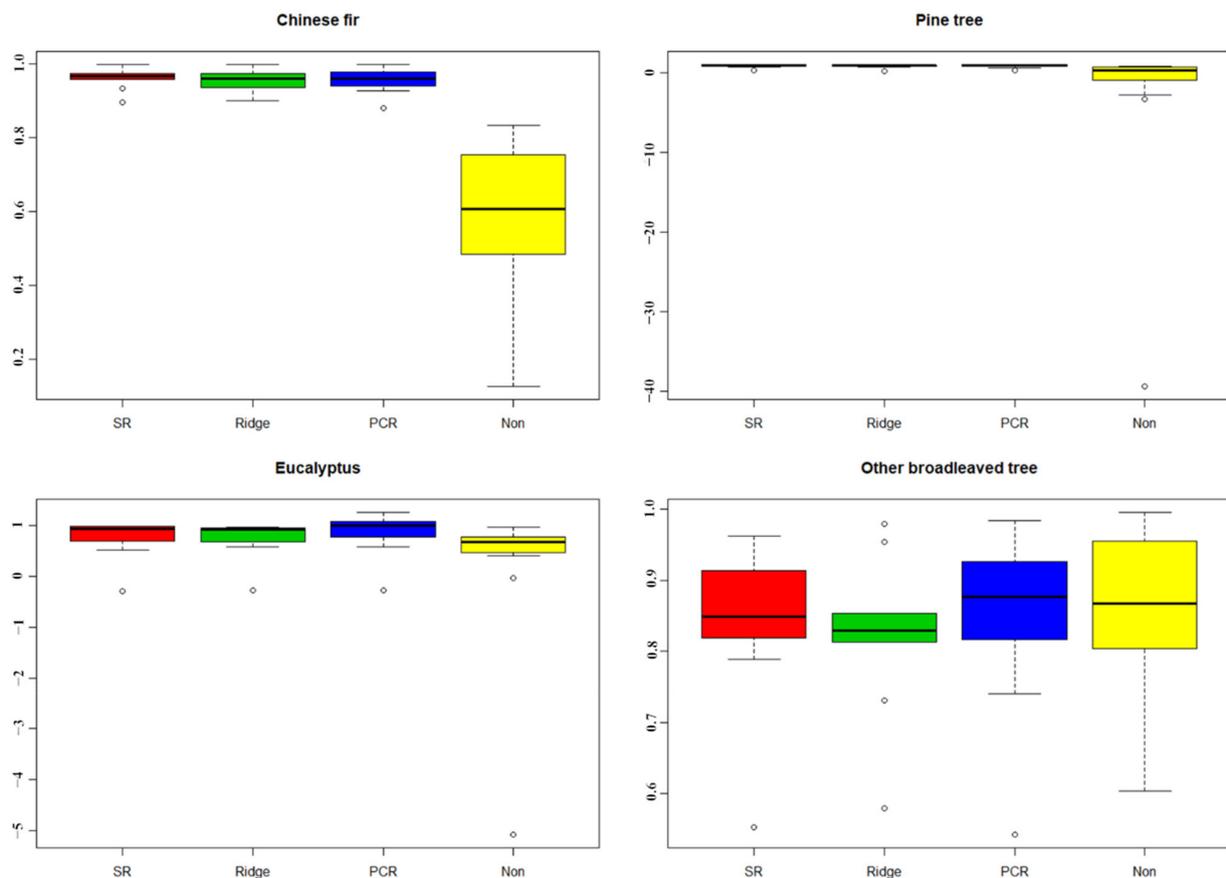
Tree Species	Method	Model	Training Accuracy $R^2$
Chinese fir	SR	$Y = 108.5 - 4.7 \times H.\text{mean.AD}$	0.23
	Ridge	$Y = 101.9 - 0.7 \times H.S - 0.1 \times H.\text{variance} - 0.01 \times H.K - 1.3 \times H.\text{mean.AD}$	0.25
	PCR	$Y = 109.5 - 0.8 \times H.S - 0.2 \times H.\text{variance} - 0.6 \times H.K - 1.8 \times H.\text{mean.AD}$	0.24
	Non	$Y = 1048.5 + 5.4 \times H.S^2 - 3.0 \times H.\text{variance}^2 + 5.5 \times H.K^2 - 284.5 \times H.\text{mean.AD}^2 + 24.9 \times H.S \times H.\text{variance} - 92.4 \times H.S \times H.K - 170.6 \times H.S \times H.\text{mean.AD} + 51.6 \times H.\text{variance} \times H.K + 62.5 \times H.\text{variance} \times H.\text{mean.AD} - 262.2 \times H.K \times H.\text{mean.AD} - 19.9 \times H.S \times H.\text{variance} \times H.K - 1.5 \times H.S \times H.\text{variance} \times H.\text{mean.AD} + 137.1 \times H.S \times H.K \times H.\text{mean.AD} - 5.8 \times H.\text{variance} \times H.K \times H.\text{mean.AD} + 1.1 \times H.S \times H.\text{variance} \times H.K \times H.\text{mean.AD}$	0.9
Pine tree	SR	$Y = 120.7 - 9.5 \times H.K + 6.6 \times H.AD.\text{median} + 4.7 \times \text{Slope} - 32.5 \times RS$	0.72
	Ridge	$Y = 135.0 - 8.4 \times H.K + 6.2 \times H.AD.\text{median} + 1.3 \times \text{Slope} - 7.5 \times RS$	0.69
	PCR	$Y = 116.1 - 6.7 \times H.K + 12.2 \times H.AD.\text{median} + 0.2 \times \text{Slope} + 1.5 \times RS$	0.58
	Non	$Y = -12790 - 20.7 \times H.K^2 - 420.9 \times H.AD.\text{median}^2 + 78.8 \times \text{Slope}^2 + 3019.7 \times RS^2 + 3502.1 \times H.K \times H.AD.\text{median} - 240.0 \times H.K \times \text{Slope} + 4364.5 \times H.K \times RS - 487.3 \times H.AD.\text{median} \times \text{Slope} + 9880.1 \times H.AD.\text{median} \times RS - 929.8 \times \text{Slope} \times RS + 80.4 \times H.K \times H.AD.\text{median} \times \text{Slope} - 4487.1 \times H.K \times H.AD.\text{median} \times RS - 76.6 \times H.K \times \text{Slope} \times RS - 169.6 \times H.AD.\text{median} \times \text{Slope} \times RS + 93.1 \times H.K \times H.AD.\text{median} \times \text{Slope} \times RS$	0.98
Eucalyptus	SR	$Y = -28.6 + 3.6 \times H50 + 5.0 \times Hc40$	0.72
	Ridge	$Y = -26.6 + 0.4 \times H60 + 1.4 \times H50 + 1.4 \times Hmedian + 1.7 \times Hc40 + 2.3 \times Hc20 + 1.3 \times Hc50$	0.72
	PCR	$Y = -27.2 + 1.2 \times H60 + 1.2 \times H50 + 1.2 \times Hmedian + 1.6 \times Hc40 + 1.7 \times Hc20 + 1.6 \times Hc50$	0.71
	Non	$Y = 2914 - 178.3 \times H60^{-0.02} - 3066.9 \times Hc50^{-0.05}$	0.56
Other broadleaved trees	SR	$Y = 114.6 - 11.4 \times H1 + 302.3 \times D9$	0.48
	Ridge	$Y = 96.7 - 8.4 \times H1 + 264.9 \times D9 + 7.7 \times H.AD.\text{median}$	0.52
	PCR	$Y = 201.2 - 5.3 \times H1 + 28.9 \times D9 - 14.1 \times H.AD.\text{median}$	0.12
	Non	$Y = 142.4 + 16.7 \times H1^2 + 583.8 \times D9^2 + 1.6 \times H.AD.\text{median}^2 - 1989.8 \times H1 \times D9 - 99.9 \times H1 \times H.AD.\text{median} - 18.5 \times D9 \times H.AD.\text{median} + 1227.6 \times H1 \times D9 \times H.AD.\text{median}$	0.79

Note: SR is the stepwise regression. Ridge is the ridge regression. PCR is the principal component regression. Non is the nonlinear regression.

### 3.4. Accuracy Evaluation

In order to preliminarily judge the distribution situation of the estimated and measured values corresponding to the testing data, the accuracy information A of different models of each tree species was calculated (Figure 5, Table 9).

As can be seen from Figure 5, except for eucalyptus, the distance between the upper quartile and the lower quartile of the accuracy A of the three linear models is smaller than that of the nonlinear model. The median value of accuracy A of the nonlinear model is 0.6 for Chinese fir. Meanwhile, there is no significant difference between the upper and lower quartiles of the accuracy A of the four models for pine trees, but the accuracy difference of the nonlinear model is still greater than that of the linear model. In addition, the maximum value of the four models has great differences for other broadleaved trees. From the preliminary accuracy results, it can be seen that the accuracy A of the three linear models is relatively concentrated, the median value is basically stable and the nonlinear model is unstable [4].



**Figure 5.** Box diagram of four models' accuracy A for each tree species.

**Table 9.** Accuracy A comparison of four models for each tree species.

Tree Species	Method	Maximum	Minimum	Average	Proportion Above 0.8
Chinese fir	SR	0.997	0.895	0.961	1
	Ridge	0.997	0.900	0.952	1
	PCR	0.997	0.880	0.953	1
	Non	0.834	0.128	0.567	0.1
Pine tree	SR	0.991	0.286	0.842	0.73
	Ridge	0.998	0.144	0.835	0.8
	PCR	0.995	0.303	0.845	0.73
	Non	0.801	−39.343	−2.886	0.07
Eucalyptus	SR	0.998	−0.289	0.780	0.67
	Ridge	0.981	−0.267	0.771	0.67
	PCR	1.267	−0.273	0.876	0.75
	Non	0.967	−5.085	0.163	0.17
Other broadleaved trees	SR	0.962	0.553	0.839	0.8
	Ridge	0.980	0.580	0.823	0.8
	PCR	0.985	0.542	0.846	0.8
	Non	0.996	0.603	0.859	0.8

Table 9 shows the numerical statistical results of accuracy A of four models for each tree species. From the proportion above 0.8, the result shows that there is a great difference between the nonlinear model and linear model. The proportion above 0.8 of the accuracy A of nonlinear model for Chinese fir, pine tree and eucalyptus is less than 0.2 for all, the proportion above 0.8 for other broadleaved trees is 0.8. The proportion above 0.8 of the three linear models has little difference. To sum up, the accuracy A of the nonlinear model is low compared with the other models, and the testing error of the model is large [4].

### 3.5. Testing Accuracy

In order to determine the optimal models of each tree species, the testing accuracy is analyzed based on measured and estimated values of testing data, and the results are shown in Table 10.

**Table 10.** Testing accuracy.

Tree Species	Method	Testing Accuracy $R^2$	RMSE (t/hm <sup>2</sup> )	MAE (t/hm <sup>2</sup> )
Chinese fir	SR	0.19	4.25	3.40
	Ridge	0.07	4.78	4.11
	PCR	0.11	5.02	4.10
	Non	0.09	42.71	37.44
Pine tree	SR	0.76	21.18	17.28
	Ridge	0.73	23.19	18.56
	PCR	0.64	25.69	17.90
	Non	0.13	1118.00	303.96
Eucalyptus	SR	0.71	50.75	25.48
	Ridge	0.68	53.33	27.84
	PCR	0.69	52.41	27.83
	Non	0.11	168.79	79.94
Other broadleaved trees	SR	0.40	46.28	33.80
	Ridge	0.51	46.78	36.05
	PCR	0.01	47.24	33.27
	Non	0.32	44.85	31.62

Table 10 shows that the testing accuracy of the nonlinear model is generally low, with about 0.1 for Chinese fir, pine tree and eucalyptus, except for other broadleaved trees at 0.32. In the linear models, the testing accuracy of the stepwise regression model is higher, and the testing accuracies of Chinese fir, pine tree, eucalyptus and other broadleaved trees are 0.19, 0.76, 0.71 and 0.40, respectively. The RMSEs are 4.25 t/hm<sup>2</sup>, 21.18 t/hm<sup>2</sup>, 50.75 t/hm<sup>2</sup> and 46.28 t/hm<sup>2</sup>, respectively. According to the testing accuracy, RMSE and MAE of the four models, the nonlinear model is unstable and has the problem of over-fitting. The stepwise regression model is relatively stable and has the highest testing accuracy [52].

### 3.6. Forest Above-Ground Biomass Mapping

The AGB estimation results of each tree species showed that the stepwise regression model has the relatively highest testing accuracy. Therefore, according to the sub-compartment data of the forest resources inventory in 2014 in Guangxi Province, the distribution area of each tree species was extracted, and the corresponding point cloud feature variables in each stand area were calculated. The stepwise regression model is used to estimate the AGB of each tree species. The thematic map of AGB in study area is shown as Figure 6.

It can be seen from Figure 6 that Chinese fir is mainly distributed in the southern and central of the study area, and the AGB is mainly concentrated in the range of 80–98 t/hm<sup>2</sup>. The distribution range of pine tree is small, mainly in the northwest and southeast of the study area, and the AGB is mainly concentrated in the range of 120–160 t/hm<sup>2</sup>. Eucalyptus is mainly distributed in the eastern and western of the study area, and the AGB is mainly concentrated in the range of 50–120 t/hm<sup>2</sup>. The other broadleaved trees are distributed evenly in the middle and among the four sides of the study area, and the AGB is mainly concentrated in the range of 130–190 t/hm<sup>2</sup>. This result proves that under the same or similar geographical environment, climate and forest age, the aboveground biomass of other broadleaved trees is higher, which plays a very important role in carbon sequestration. In the future, in the afforestation planning of artificial forest, we can increase the proportion of broadleaved trees or the proportion of pine trees in coniferous forest to improve the ecological environment's construction and carbon sequestration capacity [51].

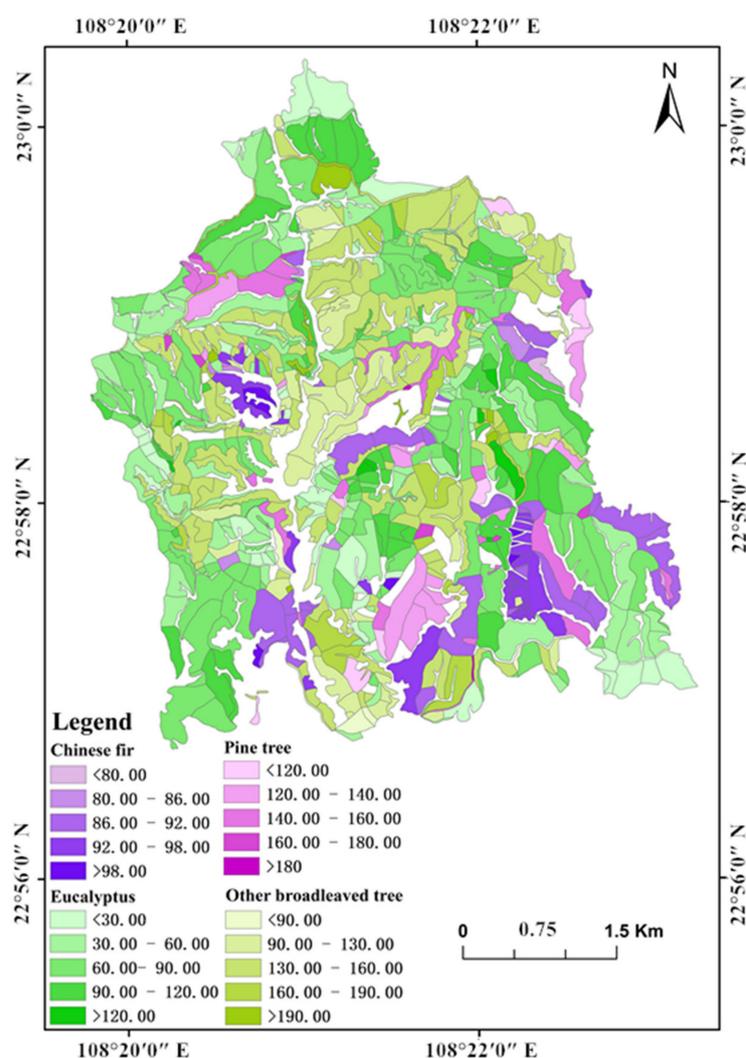


Figure 6. Forest above-ground biomass distribution in the study area.

In order to more quantitatively represent the aboveground biomass of each tree species in the study area, the obtained aboveground biomass values at the sub-compartment scale are counted, as shown in Table 11.

Table 11. Statistics table of forest AGB in the study area.

Tree Species	The Number of Sub-Compartment	The Area of Sub-Compartment (ha)	AGB of Sub-Compartment ( $t \cdot ha^{-1}$ )
Chinese fir	80	$8.495 \pm 8.475$	$89.24 \pm 14.89$
Pine tree	63	$7.835 \pm 7.655$	$146.11 \pm 54.83$
Eucalyptus	391	$8.595 \pm 8.575$	$81.17 \pm 68.82$
Other broadleaved trees	234	$6.795 \pm 6.745$	$149.795 \pm 125.925$

Note:  $m \pm n$ ,  $m$  is the median of the sub-compartment parameters for each tree species,  $n$  is the maximum value by which this parameter fluctuates up or down.

It can be seen from Table 11 that the number of sub-compartments of eucalyptus and other broadleaved trees is large and the number of sub-compartments of Chinese fir and pine tree is small in the study area, which is in line with the actual distribution law of tree species. The maximum area of sub-compartment is no more than 17.17 ha and the minimum is 0.02 ha from the statistical results, which are related to the actual planting area of tree species. From the statistical results of AGB of the sub-compartment, the AGB of other broadleaved tree fluctuates greatly, which is related to the structure of tree species [13].

#### 4. Discussion

In this study, the sensitive and optimal feature variables of aboveground biomass of different species were analyzed using airborne LiDAR point cloud data for a plantation with complex structure in South China, and the estimation model of biomass was constructed. The main research focuses are as follows:

- (a) There are differences in the optimal feature variables of different tree species. Compared with most previous studies, most of the point cloud feature variables related to aboveground biomass are height features [19,21]. In this study, the optimal features of pine trees include terrain features, and for other broadleaved trees they include point cloud density features. It also shows that the optimal features of different tree species are different, and the height feature alone cannot depict the aboveground biomass content of all tree species. Under complex terrain conditions, terrain variables should be added, and for broadleaved trees, point cloud density features must be considered as modeling variables.
- (b) It is a great advantage to distinguish tree species for estimating regional forest aboveground biomass. Compared with the forest AGB estimation in the same area, Zhang LQ [53] used Landsat TM data to estimate the forest AGB in the Gaofeng forest farm and constructed a multivariate linear equation. The accuracy was only 0.571. Compared with the forest AGB estimation using LiDAR point cloud data, Fu et al. [54] used airborne LiDAR data to estimate forest AGB in Central Yunnan Province and distinguished three forest types, including coniferous forest, broadleaved forest and mixed forest. The results showed that the AGB estimation accuracy of coniferous forest was 0.68, and that of broadleaved forests was 0.43. From the above comparisons, it can be concluded that it is necessary to distinguish tree species for estimating regional forest AGB.
- (c) The accuracy of tree species classification and distribution will affect the accuracy of regional forest AGB distribution. The regional data used in this study are the sub-compartment data of forest resources inventory, and the statistical unit is the sub-compartment. The information of tree species in the sub-compartment pertains to the dominant tree species, not the exact distribution of each tree species. Therefore, how to improve the accuracy of tree species classification and map to fine patches rather than the sub-compartment is the main direction of follow-up research.
- (d) In this study, only forest AGB in a specific area was estimated by tree species. Whether these models can be applied in other regions of same tree species has not been compared and analyzed, which will be the focus of further research.

#### 5. Conclusions

Based on LiDAR point cloud data, the features of point cloud data were extracted from different aspects and the optimal feature combination of each tree species was obtained through feature screening. The applicability of different forest AGB models for each tree species was discussed, and the regional AGB mapping was completed. The results are as follows:

- (a) 63 features of point cloud data were extracted, including tree canopy feature, terrain features, point cloud vertical distribution and point cloud density features. The top features are mostly related to the height. Since pine trees are affected by the actual sample plot environment, the tree structure is also related to the terrain factors. Other broadleaved trees have different tree species composition, so the tree shape is also related to the point cloud density features. It can be concluded that the AGB determinants of different tree species are different, which are affected by various external conditions such as environment, tree species composition and forest age.
- (b) Considering the training accuracy, testing accuracy and complexity of stepwise regression, ridge regression, principal component regression and nonlinear regression models, the accuracy of the stepwise regression model was higher than that of nonlinear model, and the model was the simplest. Therefore, the stepwise regression

method could be used to estimate forest AGB. The estimation accuracy of pine tree and eucalyptus AGB was more than 0.7, while Chinese fir and other broadleaved tree AGB was low, and that of Chinese fir was only 0.19. In conclusion, the AGB models of pine tree and eucalyptus can be used in practice, and the AGB models of Chinese fir and other broadleaved trees need to be optimized and verified.

**Author Contributions:** L.G.: Investigation, Methodology, Software, Validation, Visualization, Writing—original draft. X.Z.: Funding acquisition, Project administration, Resources. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Ministry of Science and Technology [grant number 2017YFD0600900].

**Data Availability Statement:** Not Applicable.

**Acknowledgments:** We would like to thank the Gaofeng Forest Farm in Nanning City, Guangxi Province for their aid during the field survey. We also would like to thank Erxue Chen and Lei Zhao from the Institute of Forest Resource Information Techniques CAF and Yueting Wang, Kaili Cao, Zhengqi Guo and Xuemei Zhou from Beijing Forestry University for their help in the field work.

**Conflicts of Interest:** The authors declare that they have no known competing financial interest or personal relationship that could have appeared to influence the work reported in this paper.

## References

- Shen, Y.; Cheng, R.; Xiao, W.; Yang, S.; Guo, Y.; Wang, N.; Zeng, L.; Wang, X. Labile organic carbon pools and enzyme activities of *Pinus massoniana* plantation soil as affected by understory vegetation removal and thinning. *Sci. Rep.* **2018**, *8*, 573. [[CrossRef](#)] [[PubMed](#)]
- Sun, P.; Jia, H.; Zhang, Y.; Li, J.; Lu, M.; Hu, J. Deciphering Genetic Architecture of Adventitious Root and Related Shoot Traits in *Populus* Using QTL Mapping and RNA-Seq Data. *Int. J. Mol. Sci.* **2019**, *20*, 6114. [[CrossRef](#)] [[PubMed](#)]
- Schwaab, J.; Davin, E.L.; Bebi, P.; Duguay-Tetzlaff, A.; Waser, L.T.; Haeni, M.; Meier, R. Increasing the broad-leaved tree fraction in European forests mitigates hot temperature extremes. *Sci. Rep.* **2020**, *10*, 1–9. [[CrossRef](#)] [[PubMed](#)]
- Vorster, A.G.; Evangelista, P.H.; Stovall, A.E.L.; Ex, S. Variability and uncertainty in forest biomass estimates from the tree to landscape scale: The role of allometric equations. *Carbon Balance Manag.* **2020**, *15*, 1–20. [[CrossRef](#)]
- Beaudoin, G.; Rafanoharana, S.; Boissière, M.; Wijaya, A.; Wardhana, W. Completing the Picture: Importance of Considering Participatory Mapping for REDD+ Measurement, Reporting and Verification (MRV). *PLoS ONE* **2016**, *11*, e0166592. [[CrossRef](#)]
- Sarker, L.R.; Nichol, J.E. Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sens. Environ.* **2011**, *115*, 968–977. [[CrossRef](#)]
- Lay, U.S.; Pradhan, B.; Yusoff, Z.B.M.; Bin Abdallah, A.F.; Aryal, J.; Park, H.-J. Data Mining and Statistical Approaches in Debris-Flow Susceptibility Modelling Using Airborne LiDAR Data. *Sensors* **2019**, *19*, 3451. [[CrossRef](#)]
- Bouvier, M.; Durrieu, S.; Fournier, R.A.; Renaud, J.-P. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. *Remote Sens. Environ.* **2015**, *156*, 322–334. [[CrossRef](#)]
- Dash, J.P.; Watt, M.S.; Bhandari, S.; Watt, P. Characterising forest structure using combinations of airborne laser scanning data, RapidEye satellite imagery and environmental variables. *Forests* **2015**, *89*, 159–169. [[CrossRef](#)]
- Che, E.; Jung, J.; Olsen, M.J. Object Recognition, Segmentation, and Classification of Mobile Laser Scanning Point Clouds: A State of the Art Review. *Sensors* **2019**, *19*, 810. [[CrossRef](#)]
- Zhao, K.; Popescu, S.; Nelson, R. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sens. Environ.* **2009**, *113*, 182–196. [[CrossRef](#)]
- Gleason, C.J.; Im, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91. [[CrossRef](#)]
- Kronstedter, K.; Ballhorn, U.; Böhm, V.; Siegert, F. Above ground biomass estimation across forest types at different degradation levels in Central Kalimantan using LiDAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 37–48. [[CrossRef](#)]
- Fassnacht, F.E.; Latifi, H.; Hartig, F. Using synthetic data to evaluate the benefits of large field plots for forest biomass estimation with LiDAR. *Remote Sens. Environ.* **2018**, *213*, 115–128. [[CrossRef](#)]
- Cao, L.; Coops, N.C.; Sun, Y.; Ruan, H.; Wang, G.; Dai, J.; She, G. Estimating canopy structure and biomass in bamboo forests using airborne LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2019**, *148*, 114–129. [[CrossRef](#)]
- Zolkos, S.; Goetz, S.; Dubayah, R. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sens. Environ.* **2013**, *128*, 289–298. [[CrossRef](#)]
- Takagi, K.; Yone, Y.; Takahashi, H.; Sakai, R.; Hojyo, H.; Kamiura, T.; Nomura, M.; Liang, N.; Fukazawa, T.; Miya, H.; et al. Forest biomass and volume estimation using airborne LiDAR in a cool-temperate forest of northern Hokkaido, Japan. *Ecol. Inform.* **2015**, *26*, 54–60. [[CrossRef](#)]

18. Stovall, A.E.; Vorster, A.G.; Anderson, R.S.; Evangelista, P.H.; Shugart, H.H. Non-destructive aboveground biomass estimation of coniferous trees using terrestrial LiDAR. *Remote Sens. Environ.* **2017**, *200*, 31–42. [[CrossRef](#)]
19. He, Q.; Chen, E.; An, R.; Li, Y. Above-Ground Biomass and Biomass Components Estimation Using LiDAR Data in a Coniferous Forest. *Forests* **2013**, *4*, 984–1002. [[CrossRef](#)]
20. Shao, G.; Shao, G.; Gallion, J.; Saunders, M.R.; Frankenberger, J.R.; Fei, S. Improving Lidar-based aboveground biomass estimation of temperate hardwood forests with varying site productivity. *Remote Sens. Environ.* **2018**, *204*, 872–882. [[CrossRef](#)]
21. Lu, J.; Wang, H.; Qin, S.; Cao, L.; Pu, R.; Li, G.; Sun, J. Estimation of aboveground biomass of Robinia pseudoacacia forest in the Yellow River Delta based on UAV and Backpack LiDAR point clouds. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *86*, 102014. [[CrossRef](#)]
22. Salum, R.B.; Souza-Filho, P.W.M.; Simard, M.; Silva, C.A.; Fernandes, M.E.; Cougo, M.F.; Nascimento, W.D.; Rogers, K. Improving mangrove above-ground biomass estimates using LiDAR. *Estuarine Coast. Shelf Sci.* **2020**, *236*, 106585. [[CrossRef](#)]
23. Yuanguang, W. Study on biomass and distribution of Cunninghamia lanceolata Plantation in Guangxi. *J. Guangxi Agric. Univ.* **1995**, *14*, 55–64.
24. Yuanguang, W. Study on biomass and productivity of Eucalyptus urophylla plantation. *J. Trop. Subtrop. Bot.* **2000**, *8*, 123–127.
25. Dou, L.; Deng, Q.; Li, M.; Wang, W. Biomass and distribution characteristics of *Pinus massoniana* plantations with different ages in Eastern Guangxi. *Acta Bot. Boreali-Occident. Sin.* **2013**, *33*, 394–400.
26. Luo, Y.; Chen, C.G.; Zhu, J.F. *A Handbook of Biomass Models for Major Forest Trees in China*; China Forestry Publishing House: Beijing, China, 2015.
27. Smigiel, E.; Alby, E.; Grussenmeyer, P. TLS data denoising by range image processing. *Photogramm. Rec.* **2011**, *26*, 171–189. [[CrossRef](#)]
28. Qing, S.; Tao, X.; Tatsuo, Y.; Nan, S.; Hang, Z. Classified denoising method for laser point cloud data of stored grain bulk surface based on discrete wavelet threshold. *Int. J. Agric. Biol. Eng.* **2016**, *9*, 123–131.
29. Gorgens, E.B.; Valbuena, R.; Rodriguez, L.C.E. A Method for Optimizing Height Threshold When Computing Airborne Laser Scanning Metrics. *Photogramm. Eng. Remote Sens.* **2017**, *83*, 343–350. [[CrossRef](#)]
30. Lin, X.; Zhang, J. Segmentation-Based Filtering of Airborne LiDAR Point Clouds by Progressive Densification of Terrain Segments. *Remote Sens.* **2014**, *6*, 1294–1326. [[CrossRef](#)]
31. Quan, Y.; Song, J.; Guo, X.; Miao, Q.; Yang, Y. Filtering LiDAR data based on adjacent triangle of triangulated irregular network. *Multimed. Tools Appl.* **2016**, *76*, 11051–11063. [[CrossRef](#)]
32. Liu, H.; Wu, C. Developing a Scene-Based Triangulated Irregular Network (TIN) Technique for Individual Tree Crown Reconstruction with LiDAR Data. *Forests* **2019**, *11*, 28. [[CrossRef](#)]
33. Yang, X. Cover: Use of LIDAR elevation data to construct a high-resolution digital terrain model for an estuarine marsh area. *Int. J. Remote Sens.* **2005**, *26*, 5163–5166. [[CrossRef](#)]
34. Polat, N.; Uysal, M.; Toprak, A. An investigation of DEM generation process based on LiDAR data filtering, decimation, and interpolation methods for an urban area. *Measurement* **2015**, *75*, 50–56. [[CrossRef](#)]
35. Ma, Q.; Su, Y.; Guo, Q. Comparison of Canopy Cover Estimations From Airborne LiDAR, Aerial Imagery, and Satellite Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4225–4236. [[CrossRef](#)]
36. Richardson, J.J.; Moskal, L.M.; Kim, S.-H. Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR. *Agric. For. Meteorol.* **2009**, *149*, 1152–1160. [[CrossRef](#)]
37. Cao, D.-S.; Liang, Y.-Z.; Xu, Q.-S.; Zhang, L.-X.; Hu, Q.-N.; Li, H.-D. Feature importance sampling-based adaptive random forest as a useful tool to screen underlying lead compounds. *J. Chemom.* **2011**, *25*, 201–207. [[CrossRef](#)]
38. Wang, G.; Fu, G.; Corcoran, C. A forest-based feature screening approach for large-scale genome data with complex structures. *BMC Genet.* **2015**, *16*, 1–11. [[CrossRef](#)]
39. Huang, C.; Townshend, J.R.G. A stepwise regression tree for nonlinear approximation: Applications to estimating subpixel land cover. *Int. J. Remote Sens.* **2003**, *24*, 75–90. [[CrossRef](#)]
40. Wang, M.; Wright, J.; Brownlee, A.; Buswell, R. A comparison of approaches to stepwise regression on variables sensitivities in building simulation and analysis. *Energy Build.* **2016**, *127*, 313–326. [[CrossRef](#)]
41. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
42. Park, M.; Yang, M. Ridge Regression Estimation for Survey Samples. *Commun. Stat. Theory Methods* **2008**, *37*, 532–543. [[CrossRef](#)]
43. Kaçiranlar, S.; Sakalioğlu, S.; Özkale, M.R.; Güler, H. More on the restricted ridge regression estimation. *J. Stat. Comput. Simul.* **2011**, *81*, 1433–1448. [[CrossRef](#)]
44. Liu, X.-Q.; Gao, F.; Xu, J.-W. Linearized Restricted Ridge Regression Estimator in Linear Regression. *Commun. Stat. Theory Methods* **2012**, *41*, 4503–4514. [[CrossRef](#)]
45. Massy, W.F. Principal Components Regression in Exploratory Statistical Research. *J. Am. Stat. Assoc.* **1965**, *60*, 234. [[CrossRef](#)]
46. Jeong, I.I.; Lou, I.; Ung, W.K.; Mok, K.M. Using principle component regression, artificial neural network, and hybrid models for predicting phytoplankton abundance in Macau storage reservoir. *Environ. Model. Assess.* **2015**, *20*, 355–365. [[CrossRef](#)]
47. Xiong, W.; Shi, X. Soft sensor modeling with a selective updating strategy for Gaussian process regression based on probabilistic principle component analysis. *J. Frankl. Inst.* **2018**, *355*, 5336–5349. [[CrossRef](#)]
48. Zhang, W.; Lou, I.C.; Kong, Y.; Ung, W.K.; Mok, K.M. Eutrophication analyses and principle component regression for two subtropical storage reservoirs in Macau. *Desalination Water Treat.* **2013**, *51*, 7331–7340. [[CrossRef](#)]

49. Tran, H.; Kim, J.; Kim, D.; Choi, M.; Choi, M. Impact of air pollution on cause-specific mortality in Korea: Results from Bayesian Model Averaging and Principle Component Regression approaches. *Sci. Total Environ.* **2018**, *636*, 1020–1031. [[CrossRef](#)]
50. Farhadur Rahman, M.; Onoda, Y.; Kitajima, K. Forest canopy height variation in relation to topography and forest types in central Japan with LiDAR. *For. Ecol. Manag.* **2022**, *503*, 119792. [[CrossRef](#)]
51. Knapp, N.; Fischer, R.; Cazcarra-Bes, V.; Huth, A. Structure metrics to generalize biomass estimation from lidar across forest types from different continents. *Remote Sens. Environ.* **2020**, *237*, 111597. [[CrossRef](#)]
52. Jin, S.; Su, Y.; Song, S.; Xu, K.; Hu, T.; Yang, Q.; Wu, F.; Xu, G.; Ma, Q.; Guan, H.; et al. Non-destructive estimation of field maize biomass using terrestrial lidar: An evaluation from plot level to individual leaf level. *Plant Methods* **2020**, *16*, 1–19. [[CrossRef](#)]
53. Zhang, L.Q. *Research on Remote sensing Biomass Estimate of Eucalyptus Plantation*; Guangxi University: Nanning, China, 2012.
54. Fu, T.; Pang, Y.; Huang, Q.; Liu, Q.; Xu, G. Prediction of subtropical forest parameters using airborne laser scanner. *J. Remote Sens.* **2011**, *15*, 1092–1104.