Comparison of QRNN and QRF Models in Forest Biomass Estimation Based on the Screening of VIs Using an Equidistant Quantile Method

Xiao Xu [1,2,3], **Xiaoli Zhang** [4], **Shouyun Shen** [1,3,*] **and Guangyu Zhu** [5]

1   College of Landscape Architecture, Central South University of Forestry and Technology,
     Changsha 410004, China; zz2012@ynufe.edu.cn
2   School of Logistics and Management Engineering, Yunnan University of Finance and Economics,
     Kunming 650221, China
3   Hunan Big Data Engineering Technology Research Center of Natural Protected Areas Landscape Resources,
     Changsha 410004, China
4   Key Laboratory of Southwest Mountain Forest Resources Conservation and Utilization,
     Ministry of Education, Southwest Forestry University, Kunming 650233, China; karojan@swfu.edu.cn
5   Forestry College, Central South University of Forest and Technology, Changsha 410004, China;
     zgy1111999@163.com
*   Correspondence: shen-shouyun@csuft.edu.cn

Abstract: The investigation of a potential correlation between the filtered-out vegetation index and forest aboveground biomass (AGB) using the conventional variables screening method is crucial for enhancing the estimation accuracy. In this study, we examined the *Pinus densata* forests in Shangri-La and utilized 31 variables to establish quantile regression models for the AGB across 19 quantiles. The key variables associated with biomass were based on their significant correlation with the AGB in different quantiles, and the QRNN and QRF models were constructed accordingly. Furthermore, the optimal quartile models yielding the minimum mean error were combined as the best QRF (QRFb) and QRNN (QRNNb). The results were as follows: (1) certain bands exhibited significant relationships with the AGB in specific quantiles, highlighting the importance of band selection. (2) The vegetation index involving the band of blue and SWIR was more suitable for estimating the *Pinus densata*. (3) Both the QRNN and QRF models demonstrated their optimal performance in the 0.5 quantiles, with respective R^2 values of 0.68 and 0.7. Moreover, the QRNNb achieved a high R^2 value of 0.93, while the QRFb attained an R^2 value of 0.86, effectively reducing the underestimation and overestimation. Overall, this research provides valuable insights into the variable screening methods that enhance estimation accuracy and mitigate underestimation and overestimation issues.

Keywords: vegetation indices; quantile regression (QR); quantile regression neural network (QRNN); quantile random forest (QRF); *Pinus densata* forests

1. Introduction

Biomass serves as an essential parameter for evaluating forest productivity and exhibits a positive relationship with biodiversity [1,2]. The large-scale and accurate estimation of the forest aboveground biomass (AGB) in forests holds significant importance for global ecological conservation and environmental preservation, particularly in the context of China's pursuit of the dual-carbon target [3,4].

The field investigation of forest biomass is time-consuming and labor-intensive, and it is limited to estimating the biomass in a small region [5]. With the rapid development of remote sensing, the utilization of remote sensing data can expedite and streamline the acquisition of parameter information for forest AGB estimation [6]. Various remote sensing methods have been employed to estimate forest AGB, including optical methods, radar, and LiDAR [2,7]. Although LiDAR and radar possess strong vegetation penetration capability,

their application in large areas remains challenging due to the high data collection costs [8,9]. In contrast, utilizing optical images as an alternative approach for the evaluation of the forest AGB in extensive regions offers advantages such as lower cost, higher temporal resolution, and broader spatial coverage [10,11]. However, the electromagnetic waves used in optical remote sensing cannot penetrate the forest canopy; thus, they only capture the radiation information from the vegetation surface. Consequently, optical remote sensing tends to underestimate the high AGB values in high-density forests while overestimating the low AGB values due to interference from other surface vegetation's light waves [12,13]. These uncertainties are influenced by factors such as forest structure variations, topographic differences, and remote sensing data sources. Therefore, the challenges lie in improving the estimation accuracy when assessing the AGB through optical remote sensing in a large area [14–16].

Selecting an appropriate model is a viable approach to improve the estimation accuracy and mitigate uncertainty [17,18]. The AGB estimation methods encompass both parametric and non-parametric approaches [19]. The parametric methods use linear, logarithmic, exponential, and other functions to describe the correlation between the remote sensing variables and forest AGB [20]. The non-parametric methods, such as random forest (RF), k-nearest neighbor (kNN), support vector machine (SVM), and geographically weighted regression (GWR), are also utilized [21–25]. In both parametric and non-parametric modeling approaches, researchers typically extract variables from remote sensing data and subsequently select the most influential factors using random forest classifiers, Pearson's and Spearman's correlation coefficients, etc., for constructing a prediction model [26,27]. During the variable screening stage, it is important to consider the following: (1) for linear relationships between variables, a higher absolute value of the correlation coefficient indicates a stronger association between them [26]. (2) In cases where a nonlinear relationship exists between the variables, the absolute value of the correlation coefficient may be either very large or very small; hence, it cannot accurately measure their correlation strength [26]. (3) The correlation coefficient is easily affected by outliers [28]. The estimation of biomass is influenced by various complex factors, including topography, vegetation structure, weather, etc. [14,29]. Therefore, if a simple linear relationship is employed to filter out the variables, it may inadvertently eliminate factors that could potentially have hidden relationships with biomass (e.g., outliers), consequently hindering the identification of the relationship between the vegetation index and biomass. This limitation would inevitably increase the uncertainty associated with biomass estimation. On the other hand, nonlinear and machine methods aim to minimize the loss according to mean metrics like MAE or employ feature selection techniques; however, they fail to capture how independent variables impact the position, distribution, and shape of dependent variables (3). The correlation coefficient is easily affected by outliers [28]. Therefore, if a simple linear relationship is used to filter out the variables, the factors that may have hidden relationships with biomass, such as outliers, will be filtered out. As a result, the relationship between the vegetation index and AGB may not be found, which would increase the uncertainty in estimating the AGB. While the nonlinear and machine learning methods choose to minimize the loss according to means such as MAE or sorting the feature selection, they cannot describe the effects of the independent variables on the position, distribution, and shape of the dependent variables.

The quantile regression (QR) model, proposed by Koenker and Bassett [30], offers a more accurate depiction of the range of changes in both dependent and independent variables. QR provides a flexible and stable value that is unaffected by data outliers and heavy-tailed distributions, assuming the basic assumptions of the conventional models, such as independence, normality, and equal variance [31,32]. Moreover, the QR-based methods are particularly advantageous as they not only reveal the mean value but also show its quantiles, especially when there is a tendency for the data to approach extreme values [33]. Consequently, QR has been widely used in economics, biology, finances, etc., with recent research observing its potential in environmental protection [34,35]. For instance, extreme quantiles (5% and 95%) were selected to demonstrate the temperature change trends during day and night [36]. Additionally, QR has been employed as a

modeling tool to comprehend climate systems specifically focusing on variations in climate variability and extremity [37,38]. By applying QR analysis to identify the temperature trends across 19 quantile levels, it was revealed that an association exists between large-scale climate patterns and extreme temperatures [39]. Moreover, QR was employed to monitor the temporal and seasonal changes in rainfall, as well as the temperature [40,41]. When multiple factors influence the response variable and these factors exhibit varying effects that cannot be identified and measured, QR proves to be a suitable regression method. The relationship between biomass and vegetation data is intricate and exhibits significant variations. The traditional linear analysis methods tend to exclude discrete data peaks or thick tail values; however, the conventional nonlinear and machine learning approaches fail to capture the impact of independent variables on the position, distribution, and shape of dependent variables. Some previously disregarded vegetation index values may possess hidden or localized relationships with biomass, which could provide crucial implications. What are the consequences of excluding these culled values on biomass? How does the biomass change in terms of the shape or range across different quantiles and various vegetation indices? These aspects have not been considered by the traditional variable selection methods when estimating the forest AGB.

Stepwise linear regression (SLR), machine learning algorithms such as RF, artificial neural network (ANN), and deep learning algorithms like convolution neural network (CNN) are commonly used in biomass estimation [25,42–44]. However, these methods still fail to accurately capture the response characteristics and shape changes of dependent variables. To address this limitation, models integrating the features of the QR model with ANNs and RF, namely QRNN and QRF, have been developed [45]. These models retain the response characteristics of the QR model while also exhibiting variable representation through shape changes. Furthermore, they preserve the advantages offered by the ANN and RF models. Previous studies have mostly demonstrated that RF outperforms ANN in terms of fitting effectiveness [3,46], but limited research has explored the fitting performance of these two models after their combination with QR, particularly when different variable screening methods are utilized.

In this study, we integrated remote sensing with inventory data to extract 25 commonly used vegetation indices and constructed the quantile regression model with 19 equidistant quantiles ranging from 0.05 to 0.95. The purposes of this study were as follows:

(1) To propose a novel variable screening method that visualizes the shape changes and significance of each factor to reduce uncertainty by visualizing the shape changes of each factor and their significance.
(2) To investigate the potential of the quantile regression neural network (QRNN) and quantile random forest (QRF) models for enhancing the accuracy of aboveground biomass (AGB) estimation.

2. Materials and Methods

As shown in Figure 1, the work consisted of the following steps: (1) analyze the inventory data and calculate the AGB; (2) download Landsat 8 OLI images and then preprocess the image data; (3) extract the vegetation indices; (4) conduct the quantile regression (QR); (5) select the variables; (6) construct the QRNN and QRF models; and (7) draw the inversion map.
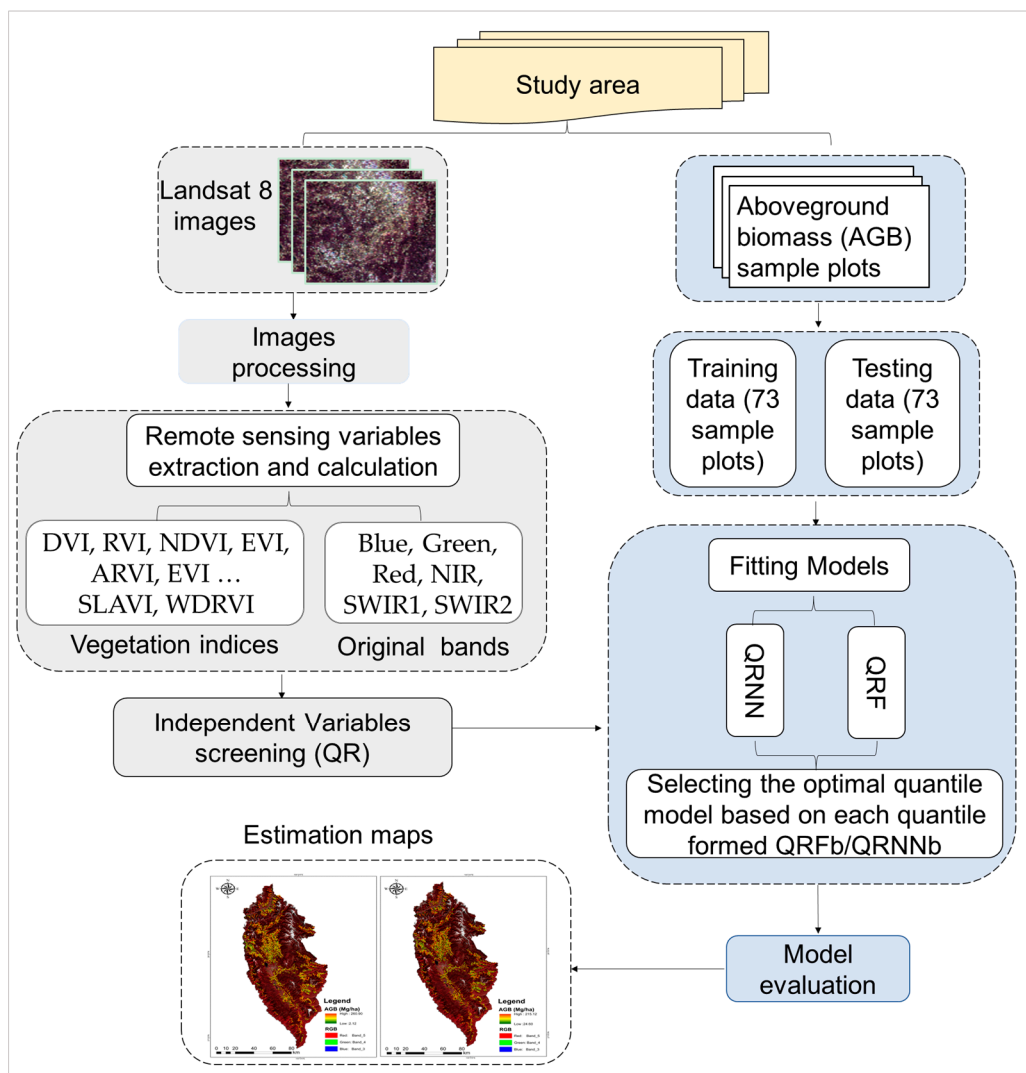
Figure 1. The workflow of aboveground biomass estimation using equidistant quantiles regression. AGB means aboveground biomass; QR means quantile regression; QRNN means quantile regression neural network; QRF means quantile random forest, QRNNb and QRFb mean the combined model that integrated the minimum mean error of each quantile.

2.1. Study Area

Shangri-La is in Diqing Prefecture, in the northwest of Yunnan Province (Figure 2); its coordinations are 26°52′~28°52′ N, 99°20′~100°19′ E. The climate changes along with the latitude. The average altitude is 3459 m, and the mean annual temperature is 5.4 °C. Under the influence of monsoons, Shangri-La has distinct dry and wet seasons. The rainy season from June to October accounts for 20%–80% of the annual precipitation, while the dry season rainfall from November to May accounts for 10%–20% of the annual precipitation. The total forest area is 75,710 km^2, and the forest cover rate is nearly 75%. The main forest type is alpine coniferous forest, and the dominant tree species are *Picea asperata*, *Abies fabri*, *Pinus densata*, *Pinus yunnanensis*, *Quercus semicarpifolia*, etc. [20].
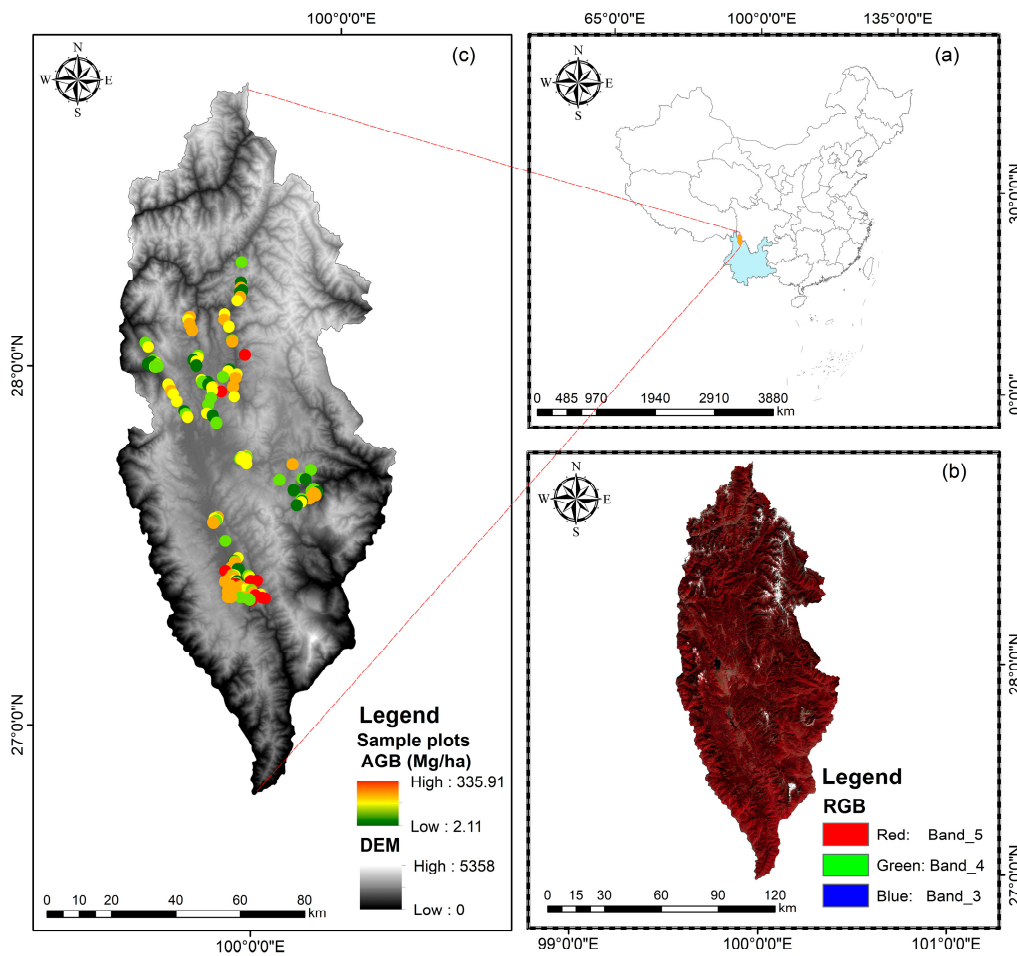
Figure 2. Location of the study area: (**a**) is the position of the study area in Yunnan Province, China; (**b**) is the imagery RGB displayed by combining bands 5, 4, and 3; (**c**) is the DEM data and the sample plots AGB distribution (from green to red indicating low to high).

2.2. Data Resources

2.2.1. The Inventory Data

The inventory data were obtained in August of the year 2016. A total of 146 random 30 m × 30 m plots were set in *Pinus densata* forest. The basic information, such as *DBH* (diameter at breast height of 1.3 m aboveground) greater than 5 cm, tree height, forest type, forest age, slope, altitude, etc., was recorded. The coordinates of the plot were recorded by GPS, with a horizontal error within 5 m. The aboveground biomass (AGB) of individual trees was calculated by Equation (1) [20], and the AGB by plot was calculated by Equation (2) [20]. The forest density was counted by sample area (Table 1).

$$AGB_i = 0.073 \times DBH^{1.739} \times H^{0.880} \tag{1}$$

where *DBH* is the diameter at breast height (1.3 m aboveground) greater than 5 cm, and *H* is tree height. AGB_i is the aboveground biomass of the sampling tree (kg).

$$AGB_S = \frac{\sum_{i=1}^{n} AGB_i}{900} \times 10,000 \div 1000 \tag{2}$$

where AGB_i is the biomass of a single tree, and AGB_s is the sum of AGB in one plot, *n* is the total tree number in the plot.

Table 1. Basic statistics of the sample plots of field surveys.

Variables	Mean *DBH* (cm)	Mean *H* (m)	Density (Stocking·hm^{-2})
Max.	41.27	24.30	8500
Min.	5.35	2.93	489
Mean	15.14	10.34	2657
Standard deviation	3.88	3.42	1326

2.2.2. Remote Sensing Data

The Landsat 8 OLI images that matched the research area were downloaded from the Geospatial Data Cloud (http://www.gscloud.cn/) on 21 December 2022. The projection of the images was UTM/WGS 84 with UTM_Zone 47. Cloud cover impacted each band of optical remote sensing, affecting the vegetation index calculation. The average cloud cover images shown in Table 2 were selected. Detailed information on the Landsat 8 OLI images is shown in Table 2.

Table 2. The image information.

Image ID	Strip No.	Average Cloud Cover (%)	Data
LC81310412016325LGN00	131	0.4	20 November 2016
LC81320402016348LGN00	132	0.73	13 December 2016
LC81320412016348LGN00	132	0.76	13 December 2016

The image data preprocessing was conducted in ENVI, including radiometric calibration, FLASSH atmospheric correction, and topographic correction; the image of the study area (Figure 1) was subsequently obtained through mosaicking and clipping.

2.2.3. Extraction of Vegetation Indices

Simple VIs, such as the RVI [47], DVI [48], and other primary VIs, have inherent limitations [49]. To overcome these limitations, advanced VIs like NDVI [50], RDVI [51], and other VIs have been developed; however, NDVI tends to overestimate low-cover vegetation and underestimate high-cover vegetation [52], while the RDVI is easily susceptible to soil background interference [53]. Composite VIs, including WDRVI, GNDVI, SAVI, OSAVI, VARI, GARI, MSR, etc., have been devised to account for the influence of both vegetation or soil on the VIs [53–58]. Furthermore, researchers have recognized that atmospheric conditions also impact the accuracy of VIs, leading to the development of ARVI [59]. With further advancements in research, it has become evident that factors influencing VIs extend beyond just soil or atmosphere; hence, considering the vegetation–atmosphere–soil interaction becomes crucial in calculating accurate vegetation indices [60]. Consequently, EVI was introduced as a solution to address the combined influence on the vegetation index [61]. In this study, 31 variables were considered, as shown in Table 3, including 6 original spectral bands along with 25 VIs related to vegetation dynamics as well as their interactions with soil and atmosphere.

Table 3. The variables' information.

Variables	Formula	Description	Reference
Single band	Band 2–7	Blue, Green, Red, NIR, SWIR1, SWIR2	[62]
NDVI	(NIR − Red)/(NIR + Red)	Normalized Difference Vegetation Index detects vegetation coverage and growth state	[50]
ARVI	(NIR − Red + Blue)/(NIR + Red + Blue)	Atmospherically resistant vegetation index is mainly used in areas with high atmospheric aerosol.	[59]

Table 3. *Cont.*

Variables	Formula	Description	Reference
VARI	(Green − Red)/(Green + Red − Blue)	Visible atmospherically resistant index. It is used to measure the amount of green vegetation	[58]
W	Blue × (0.1511) + Green × (0.1973) + Red × (0.3283) + NIR × (0.3407) + SWIR1 × (−0.7117) + SWIR2 × (−0.4559)	Tasseled cap wetness Reflects the moisture of soil and vegetation	[63]
DVI	NIR − Red	Difference vegetation index It is used to reflect the growth of vegetation	[48]
PVI	$((\text{Red}_{\text{soil}} - \text{Red}_{\text{veg}}) + (\text{IR}_{\text{soil}} - \text{IR}_{\text{veg}})^2)^{0.5}$	It is better to eliminate the influence of soil background, insensitive to the atmosphere.	[48]
EVI	2.5 × (NIR − Red)/((NIR + 6 × Red − 7.5 × Blue) + 1)	Enhanced vegetation index It increases rapidly with the increase in vegetation quantity when the vegetation coverage is 15%–25%, and it will decrease when the vegetation coverage reaches 80%.	[60]
RDVI	$(\text{NIR} - \text{Red})/(\text{NIR} - \text{Red})^{0.5}$	Renormalized difference vegetation index It can monitor plant water status effectively	[51]
G	Blue × (−0.2941) + Green × (−0.243) + Red × (−0.5424) + NIR × (0.7276) + SWIR1 × (0.0713) + SWIR × (−0.1608)	Tasseled cap greenness Reflects the greenness of the ground vegetation	[63]
MSR	$(\text{NIR}/\text{Red} - 1)/((\text{NIR}/\text{Red})^{0.5} + 1)$	Modified simple ratio. Its purpose is to linearize the relationships between the index and biophysical parameters	[53]
SLAVI	NIR/(Red + SWIR)	Specific leaf vegetation index Its links with plant ecophysiology and leaf biochemistry	[64]
MVI5	(Red + NIR − Blue)/(Red + NIR + Blue)	Moisture vegetation index Sensitivity index of soil moisture and canopy moisture	[65]
RVI	Red/NIR	Ratio vegetation index Sensitive to vegetation coverage	[47]
WDRVI	((0.1 × NIR) − Red)/((0.1 × NIR) + Red)	Wide-dynamic-range vegetation index a more robust characterization of crop physiological and phenological characteristics.	[54]
GARI	(NIR − (Green − (Blue − Red)))/(NIR − (Green + (Blue − Red)))	Green atmospherically resistant vegetation index GARI shows a much higher sensitivity to chlorophyll concentration than NDVI and a smaller sensitivity to atmospheric effects.	[55]
SARVI	(1 + L) × (NIR−Blue)/(NIR + Blue + L)	Soil-adjusted and atmospherically resistant vegetation index	[59]
MSAVI	$(2 \times \text{NIR} + 1 - ((2 \times \text{NIR} + 1)^2 - 8 \times (\text{NIR} - \text{Red}))^{0.5})/2$	Modified soil-adjusted vegetation index. It aims to address some limitations of NDVI when applied to areas with high soil surface exposure.	[66]
GNDVI	(NIR − Green)/(NIR + Green)	Green normalized vegetation index Monitor the plant with a dense canopy or in the mature stage	[55]
TVI	$(\text{NDVI} + 0.5)^{0.5}$	Transformed vegetation index. monitoring vegetation health and vigor is also useful for monitoring vegetation stress where the NDVI is saturated.	[50]
IPVI	NIR/(NIR + Red)	Infrared percentage vegetation index sensitive to the amount of green vegetation	[67]

Table 3. *Cont.*

Variables	Formula	Description	Reference
OSAVI	$(NIR - Red)/(NIR + Red + 0.16)$	Optimized soil-adjusted vegetation index monitors bare soil area of low-density vegetation through the tree canopy	[57]
NIR	$NIR/(NIR + Red + Green)$	Normalized NIR reduced the influence of soil background	[68]
MNDVI	$(NIR - SWIR2)/(NIR + SWIR2)$	Modified normalized difference vegetation index. Monitor forest health and canopy changes	[69]
ND67	$(SWIR1 - SWIR2)/(SWIR1 + SWIR2)$	Monitor the soil moisture capacity	[46]
B	$Blue \times 0.3029 + Green \times 0.2786 + Red \times 0.4733 + NIR \times 0.5599 + SWIR1 \times 0.508 + SWIR2 \times 0.1872$	Tasseled cap brightness a weighted sum of all bands and is related to the principal variation in soil reflectance	[63]

2.3. Models

We used quantile regression to identify the significant correlations with AGB in different quantiles. Quantile random forest (QRF) and quantile regression neural network (QRNN) models were applied as the fitting models. During the fitting model construction, 146 sample data were randomly split into two parts: 50% of the data were used for training and the other 50% for testing.

2.3.1. Quantile Regression

Quantile regression was developed by Koenker and Bassett [30] as a way of estimating the response variables' distribution on the conditional quantiles in a linear model. This method allows one to find the specific ratio trend of the dependent variable; it is also helpful to detect the extreme trends that are hidden in non-significant even effects or changes in median conditions. The equation form of the quantile regression is as follows:

$$Y(p \mid x) = \beta_0(p) + \beta_1(p)x + \xi \quad (3)$$

where $\beta_0(p)$ is the intercept and $\beta_1(p)$ is the slope coefficient, both of which vary depending on the value of the pth quantile being considered; ξ is the error, with the expectation of zero. The range of *p*-values is from 0 to 1. The quantile regression was conducted in R language, which was downloaded from http://www.r-project.org/on 10 February 2023. The standard deviation was calculated by the bootstrap self-sampling method, and the fitting method was the Barrodale and Roberts algorithm, where $\tau = 0.05, 0.1, 0.15, \ldots 0.95$.

2.3.2. Quantile Regression Neural Network

The quantile regression neural network nonlinear computational model, generated by Taylor [45], combines the advantages of quantile regression and artificial neural networks. This model not only reveals the conditional distribution characteristics of response variables but also elucidates the intricate nonlinear relationship between a dependent variable and an independent variable through the utilization of a nonlinear kernel function. It does not require prior knowledge about data distribution or specific relationships among variables; instead, it only necessitates setting the hidden layer and the number of nodes to control the model complexity. Weight attenuation regularization is employed to mitigate nonlinearity in the model, while 10-fold cross-validation is used to select the testing and verification datasets to prevent over-fitting. Through experimentation, it was determined that a combination of seven nodes and three hidden layers yields optimal stability. The QRNN model was implemented using R 4.3 language software with assistance from the QRNN code package.

Moreover, the minimum average error of each quantile was integrated as part of the best QRNN (QRNNb). Consequently, QRNNb exemplified a composite biomass estimation model derived from selecting the most accuracy values corresponding to the 19 quantile models.

2.3.3. Quantile Regression Forests

Quantile random forests use a quantile decision tree set based on random forests and quantiles [70], which can be used to estimate high-dimensional data and uncertainty. In this model, the complete conditional distribution of response variables can be determined by resampling the dataset. QRF provides a non-parametric and accurate method for estimating conditional quantiles. QRF was operated in R 4.3 software using the QRF language package by setting mtry, ntry, and the number of resamples. The grid search was used to determine the optimal parameter values by minimizing the root mean square error. QRF was more stable when mtry was 2, ntree was 400, and the number of bootstrappeds was 25. To ensure the quality and stability of the the QRF model, 10-fold cross-validation was used to control the accuracy of the model.

Moreover, the minimum average error of each quantile was integrated as part of the best QRF (QRFb). Therefore, the QRFb illustrated a combined biomass estimation model formed by selecting the highest accuracy values corresponding to the 19 quantile models.

2.4. Model Evaluation

In this study, we chose the coefficient of determination (R^2) and mean square root error (RMSE) to estimate the model fitting performance. The equations are as follows:

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

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

3. Results

3.1. Variable Selection

As shown in Figure 3, 19 equidistant quantiles were used to describe the shape change of each independent variable in response to the dependent variable through quantiles 0.05 to 0.95 in intervals of 0.05. It can be seen that the 31 factors could be divided into seven groups according to the estimated values of the coefficient shape graphs.
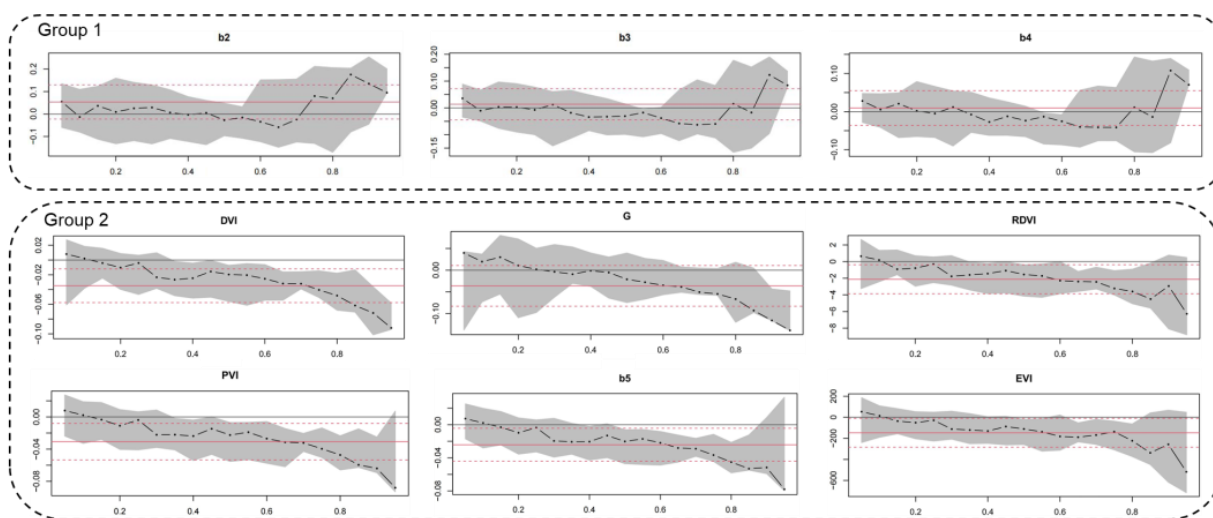
Figure 3. *Cont.*
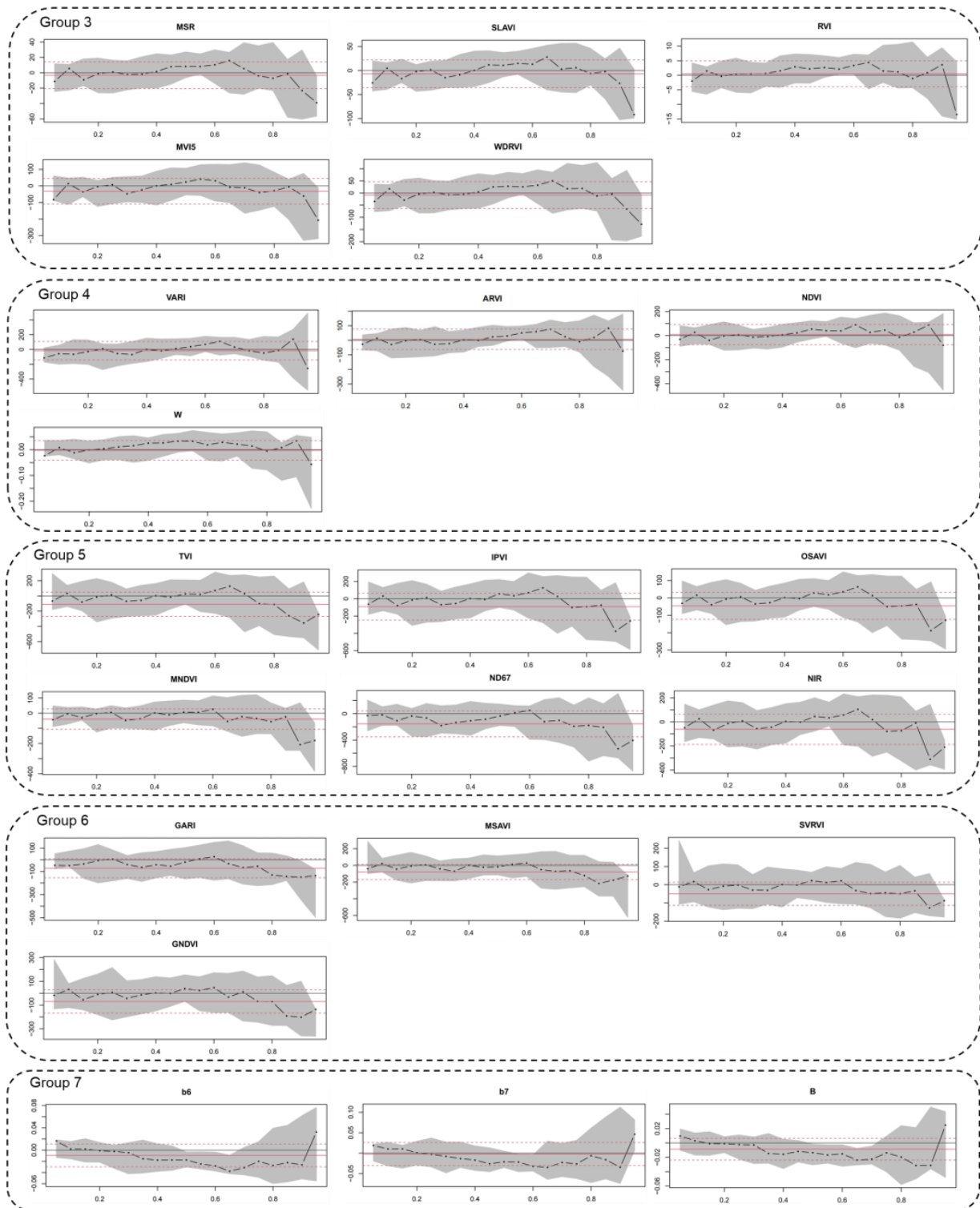
Figure 3. Shape changes of each independent variable corresponding to the dependent variables (the dark curves in the figure represent the coefficient estimates corresponding to the variables under different quantile levels; the gray area represents the 95% confidence interval of the coefficients; the solid red line represents the mean trend from a least-squares regression; the red dashed lines on both sides represent the 95% confidence interval of the coefficients in the mean regression model).

The analysis of the shape change should consider the following aspects: (1) whether the estimated coefficient value significantly influences the biomass; (2) the changing trend of the estimated coefficient values: the 31 independent variables showed an obvious

change trend in the high quantiles, while the low quantiles had relatively stable changes; (3) confidence interval: a wider confidence interval of the coefficient indicates that the standard deviation (SD) of the estimated coefficients was gradually increasing and the volatility of the estimated coefficients was increasing; (4) whether the estimated coefficient value falls within the confidence interval of the mean regression model: if it lies outside this interval, it suggests some degree of irrationality in the mean regression model; and (5) the effect of the estimated coefficient values on the biomass: a positive influence is observed for values above 0 and negative influence for those below 0.

Figure 4 shows the significance of the relationships between the biomass and the estimated coefficient values at 19 equal distance quantile points for 31 variables. In groups 1 and 7, the significance of the association between the VIs and biomass was in quantile ≤0.6. In the second group, the EVI and biomass were significant in almost all the quantiles. The relationship between b5 and the biomass was significant in quantiles 0.35–0.95. The other four VIs had a significant relationship with the biomass in quantiles 0.6–0.95. In the third group, the NDVI was significantly correlated with the biomass only in the 0.7 quantile, while the ARVI was significantly correlated with the biomass at both ends. The relationship between the VIs (VARI and W) and biomass was not significant in the higher quantiles. In the fifth group, only 2–3 points in all the quantiles showed significant associations with the AGB, and the middle four VIs showed significant correlations with the biomass in the 0.6–0.75 and 0.05 quantiles. The MNDVI was significantly correlated with the biomass in quantiles 0.15–0.45 and 0.85–0.9. In the sixth group, there was no significant relationship between the SARVI and biomass in quantiles 0.45–0.7, and the other VIs rarely showed any significant relationship with the biomass. Not all the variables had a significant association with the biomass, but we can see that there was at least one significant correlation with the biomass in one quantile. The variables showed a high correlation factor with the biomass during the normal variables screening if the significant correlation with the biomass was located around the middle quantile points. The variables with a high significant correlation with the biomass in the lower or higher quantiles were not selected as they had no strong significant relationship with the biomass in the middle quantiles. On the other hand, uncertainty could easily arise if the factors had significant associations with the biomass at nearly the middle quantiles but had no significance in the lower and higher quantiles.

A total of 19 variables that exhibited significant correlations with the AGB across at least six quantiles were selected as the independent variables (Figure 4). Subsequently, a collinearity test was carried out to validate the chosen data and address the concerns regarding the potential instability in model parameter estimates, reduced explanatory capacity, and compromised statistical reliability [62]. Finally, only those variables with significance levels below 0.01 and VIF values less than 10—namely B, EVI, MVI5, MNDVI, ND67, and SARVI—were utilized for estimating the forest AGB.
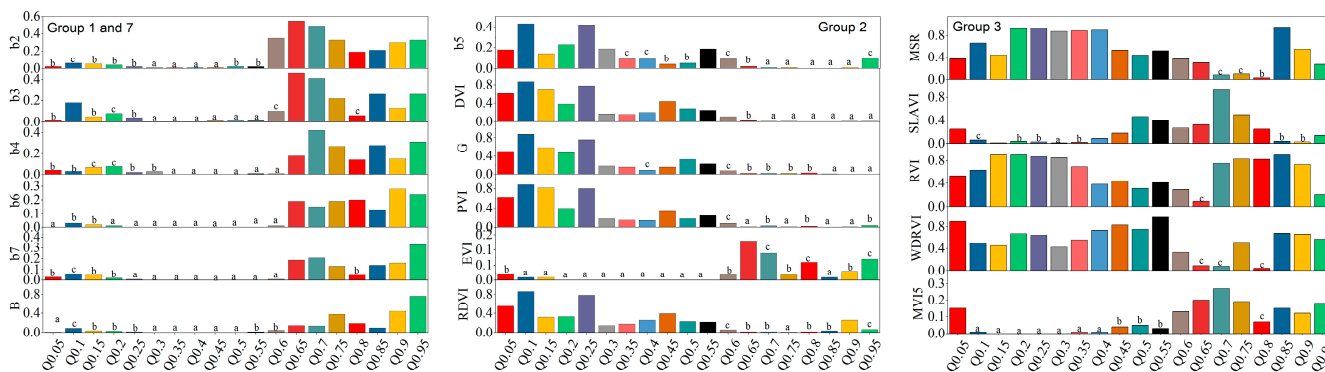
Figure 4. *Cont.*
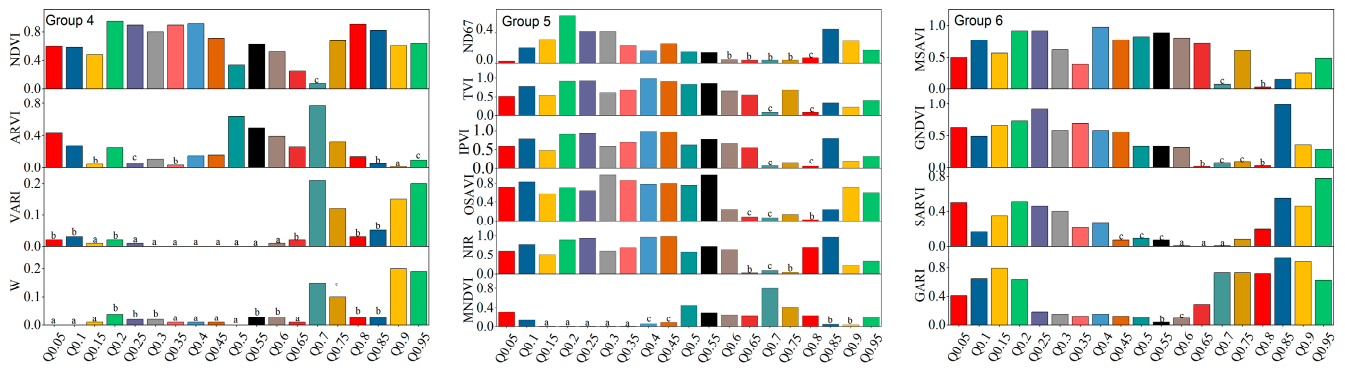
Figure 4. The significant relationship between the biomass and coefficient estimated value at 19 equal-distance quantile points for 31 variables (a represents $p < 0.01$, b represents $p < 0.05$, and c represents $p < 0.1$).

3.2. Model Performance

It can be seen from Figure 5 that the fitting performance of the model was increased and then decreased from quantiles 0.05 to 0.95. The minimum R^2 and maximum RMSE were found at both ends. The best fitting performance was in quantile 0.5, with a higher R^2 (0.68) and lower RMSE (46.89 Mg/ha). Compared with all 19 quantiles, the QRNNb had the smallest mean error in each quantile, along with the highest R^2 and lowest RMSE, with values of 0.93 and 22.59 Mg/ha, respectively.
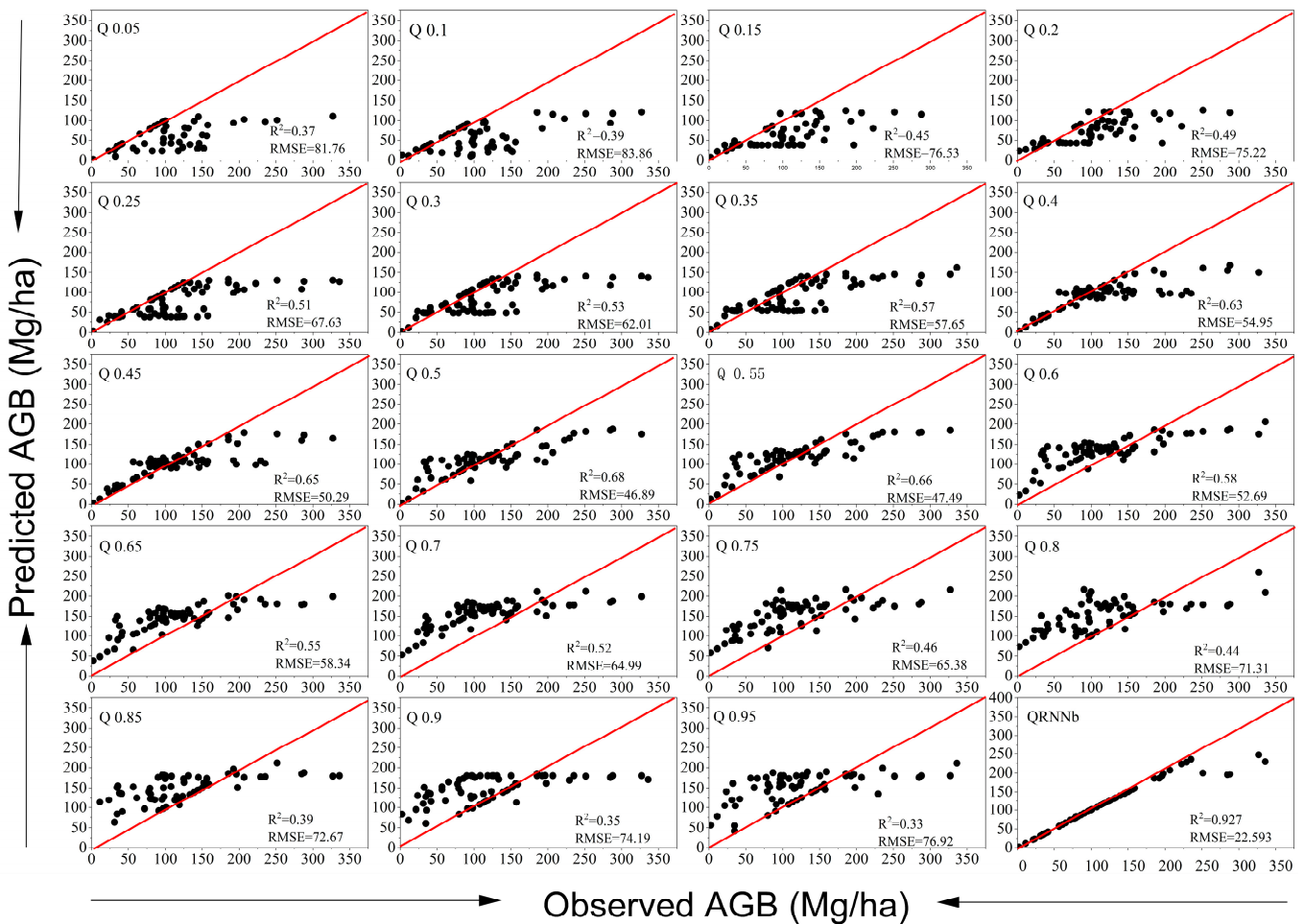
Figure 5. The scatterplot of the observed and the predicted AGB of the QRNN model. The black dots are the predicted AGB values, and the red line is the line y = x.

The performance of the model relied not only on its accuracy but also on the extent of the deviation from the y = x line. Generally, a higher slope corresponded to a smaller intercept, indicating a smaller degree of deviation from the y = x line. If the fitted line was positioned above the y = x line, it would result in an underestimation of the biomass. Figure 5 demonstrated that, for observed biomass values below 50 Mg/ha, the predicted biomass exhibited greater accuracy within the quantiles ranging from 0.05 to 0.15. Between 50 and 150 Mg/ha, only a small portion of the predicted biomass showed accurate estimation, while the other observed biomass was underestimated. In the quantiles ranging from 0.2 to 0.3, more of the predicted values were close to the red line and relatively more concentrated; however, when the biomass was less than 50 Mg/ha, overestimation occurred. The predicted values were evenly distributed along the red line in quantiles 0.4 to 0.55, where biomass values greater than 180 Mg/ha were consistently underestimated. The predicted biomass was significantly smaller than the observed biomass in the 0.6–0.95 quantiles, but the greatest predicted biomass rarely exceeded 180 Mg/ha. The deviation was mostly in the high quantiles with large biomass values, and the fitting line was below the y = x line, indicating that high biomass values were easily underestimated. We also found that the fitting line of the QRNNb coincided with y = x, indicating high fitting precision when the biomass was less than 180 Mg/ha. Thus, the phenomenon of low overvaluation and high undervaluation was effectively improved.

Figure 6 shows the observed AGB and the predicted AGB. The R^2 range of the QRF was from 0.33 (0.95 quantile) to 0.71 (0.5 quantile). Compared with the QRNN model, the mean R^2 of the QRNN and QRF was 0.5. In terms of the fitting performance in each quantile, the QRF was slightly better than the QRNN, but the QRFb, in which we integrated the predicted minimum mean error in each quantile, was inferior to the QRNNb, with an R^2 of 0.86 and an RMSE of 33.39 Mg/ha. The biomass was overestimated in all the quantiles when the AGB was less than 50 Mg/ha, and the biomass values of 50–180 Mg/ha were evenly scattered around the y = x line, while those greater than 180 Mg/ha were all underestimated in all the quantiles. The QRFb showed the best fitting performance and was able to effectively improve the underestimation and overestimation.

The boxplot was constructed to further analyze the fitting effect of the two models. As shown in Figure 7, the median fitting coefficient of the two models was nearly the same, but the median of the QRNN model deviated to the upper quartile, while that of the QRF model deviated to the lower quartile. The IQR (interquartile range) of the QRNF was wider than that of the QRNN, indicating that the prediction data of the QRF were relatively dispersed. In the same way, the QRF had the smallest error value as the estimation error dispersion was the smallest.

We used biomass inversion with the established model to calculate the biomass of the whole *Pinus densata* forest region by using the predicted biomass of the sample site (Figure 8). Generally, the heterogeneity of inversion models is used to compare the disadvantages of constructed models. Inversion maps with high heterogeneity usually show obvious spatial distributions and large color changes.

There was high heterogeneity in both models, but the QRF inversion maps showed more large values than the QRNN maps. This is also consistent with the scatterplot shown in Figure 5, where the QRF model overestimated the AGB values for all the quantiles.
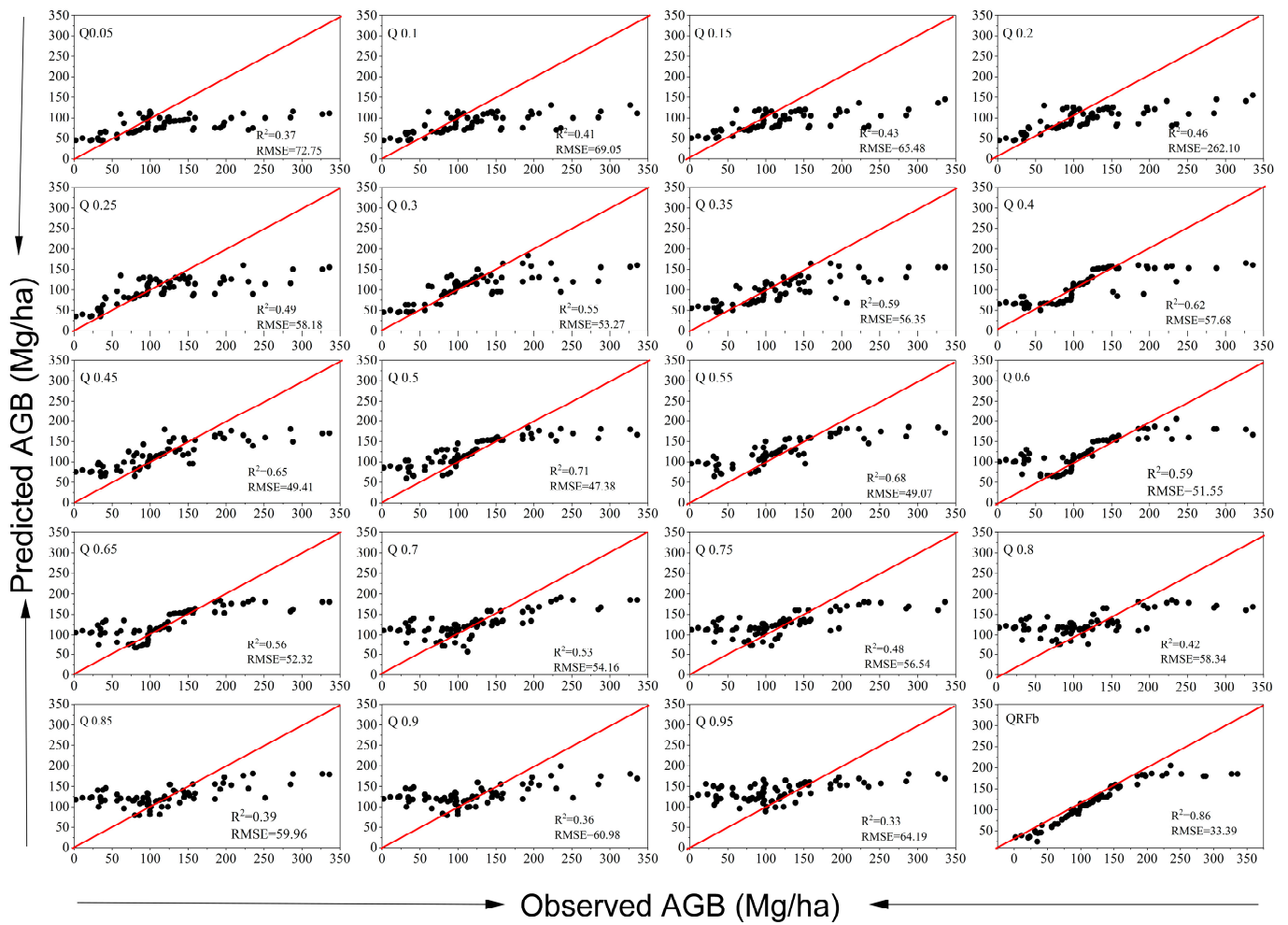
Figure 6. The scatterplot of the observed AGB and that predicted by the QRF model. The black dots are the predicted AGB values, and the red line is the line y = x.
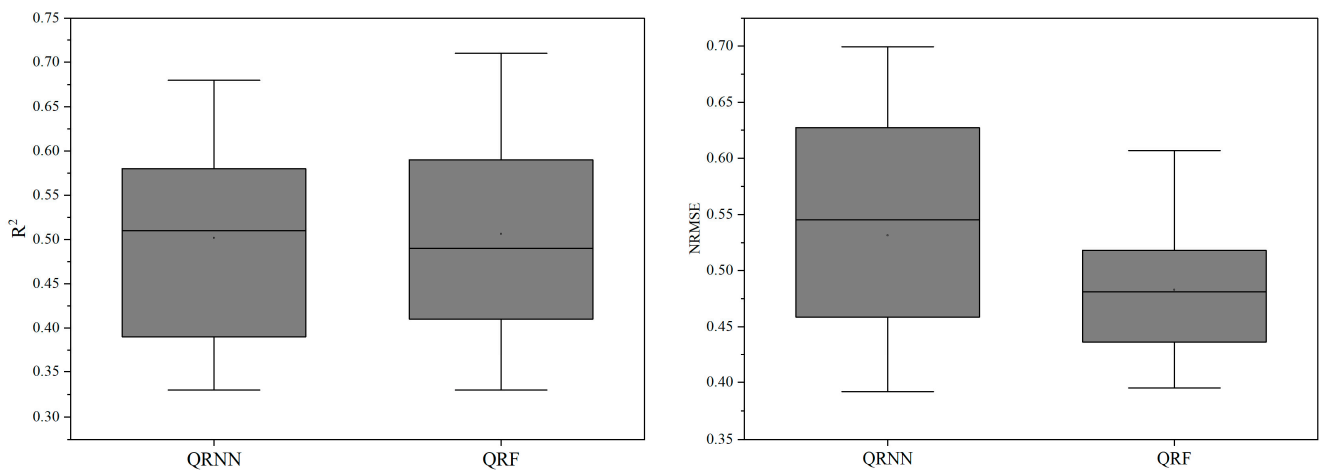
Figure 7. The boxplot of R^2 and NRMSE.
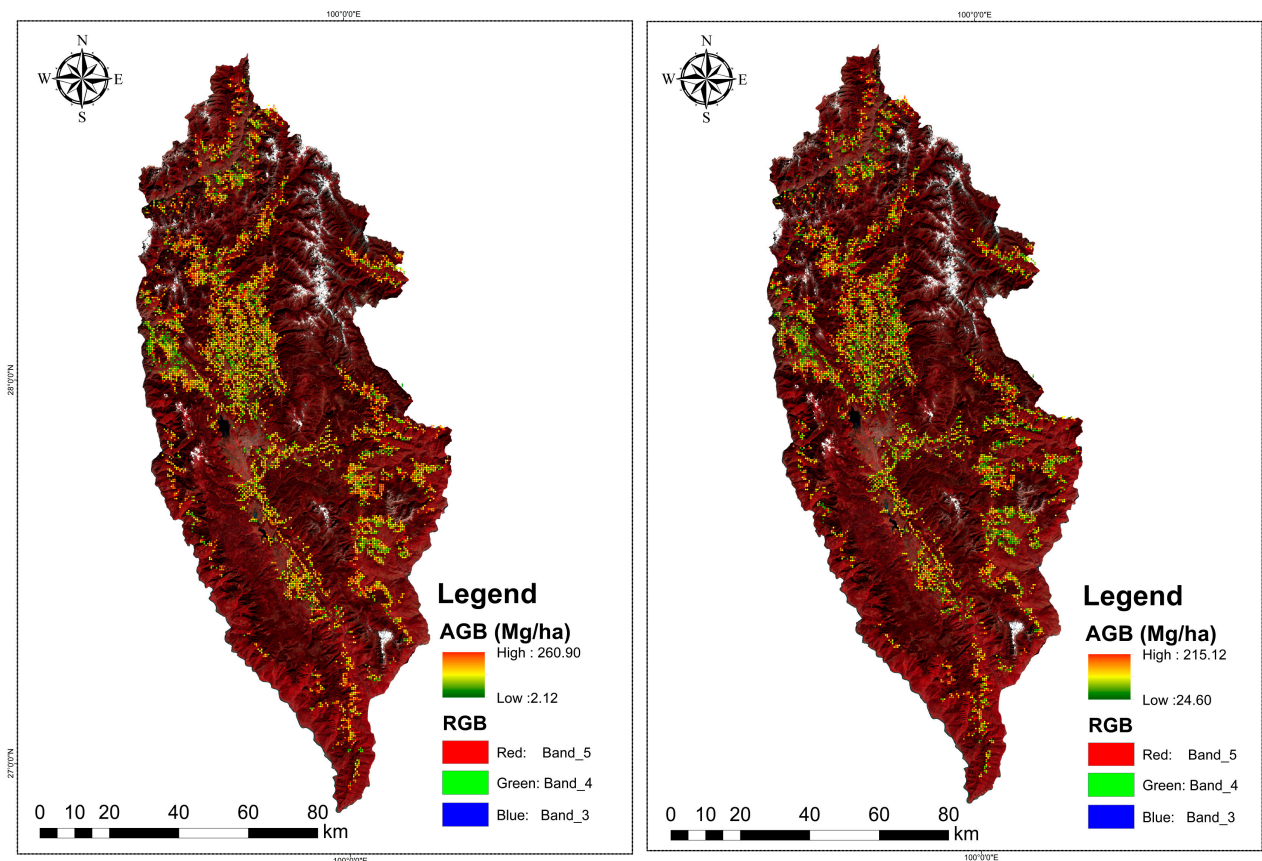
Figure 8. The inversion map of QRF and QRNN models.

4. Discussion

4.1. Variable Selection

Variable screening methods encompass correlation analysis, employing common indicators such as Pearson's correlation coefficient and Spearman's rank correlation coefficient [71], stepwise regression, where the variables are gradually introduced or removed to determine their contribution to the biomass estimation [42], and machine learning algorithms that provide feature importance assessments, such as random forests and gradient-boost regression trees [72,73]. However, due to the complex and nonlinear relationships between the biomass and various factors in highly heterogeneous forested regions with diverse topography [46], the linear analysis methods exclude discrete data peaks or thick tails, while the nonlinear and machine analysis methods fail to capture the effects of the independent variables on the position, distribution, and shape of the dependent variables. Consequently, few researchers have considered the significance of the relationships between the independent variables and dependent variables at different quantiles or evaluated the shape changes for variable selection. Conventional variable screening may lead to the exclusion of potentially important variables that play a crucial role in biomass modeling, resulting in the estimated parameter values deviating from the true values and reducing the predictive performance of the model [33]. However, if the variable selection is based on a specific sample set, it may exclude significant variables due to sample bias. Figure 4 showed the relationships between different vegetation indices and biomass varying between quantiles, with some significantly correlated in the low quantiles, some in the high quantiles, and some throughout the entire quantile range. Ignoring the significance at both ends can cause basic variable screening methods to miss variables with significant relationships. Therefore, this research method can uncover more vegetation indices related to biomass across different quantiles.

In this study, significant associations with biomass were observed for bands 2, 3, 4, 6, and 7 in the quantiles ranging from 0.05 to 0.6, while band 5 showed significance in quantiles from 0.35 to 0.95. Lu's study indicated that bands 2, 4, and 5 had the strongest relationships with forest parameters such as AGB, whereas the green band and SWIR band showed weaker correlations with these parameters in Bragantina [74]. Zhao's research illustrated that SWIR was the most effective band for estimating forest AGB [16]. Li et al. [46] applied Sentinel-2 to estimate the AGB of the *Pinus densata* forest and found that bands 2, 3, 5 (red edge band), 8 (near-infrared), and 12 (SWIR) showed strong correlations with the AGB. Interestingly, in our study, the vegetation indices calculated by bands 3 (green), 4 (red), and 5 (NIR) did not exhibit significant correlations with the AGB in the middle and low quantiles; however, they displayed significant correlations with the biomass at the higher quantiles, i.e., from 0.65 to 0.8 (DVI and RDVI were significant in the 0.65–0.95 quantiles). These results were aligned with the results of Hall et al. [75], which suggested that the NDVI is an unreliable vegetation index for boreal coniferous forests exhibiting age variation, particularly young or older forests aged over 15–20 years [76]. Additionally, it was found that complex vegetation indices were unsuitable for estimating AGB [74], especially those presented in Table 4, which showed similar variants to NDVI, essentially adding a constant term to the NDVI formula. By employing the equidistant quantile regression method, we identified varying sensitivity of the vegetation indices towards the AGB across different quantiles. Therefore, this approach aids in identifying the more sensitive vegetation indices towards the AGB at each specific quantile and subsequently enhances the estimation accuracy.

Table 4. The significant relationship between AGB and the vegetation indices calculated by band 3, band 4, and band 5.

VIs	Formula	The Significance at Quantile
NDVI	$(NIR - Red)/(NIR + Red)$	0.7
DVI	$NIR - Red$	0.65–0.95
RDVI	$(NIR - Red)/(NIR - Red)^{0.5}$	0.6–0.95
RVI	IR/Red	0.65
MSR	$(NIR/Red - 1)/((NIR/Red)^{0.5} + 1)$	0.7, 0.75, 0.8
TVI	$(NDVI + 0.5)^{0.5}$	0.7, 0.8
NIR	$NIR/(NIR + Red + Green)$	0.65, 0.7, 0.75
WDRVI	$((0.1 \times NIR) - Red)/((0.1 \times NIR) + Red)$	0.65, 0.7, 0.8
IPVI	$(NIR - Red)/(NIR + Red + 0.16)$	0.7, 0.8
OSAVI	$(NIR - Red)/(NIR + Red + 0.16)$	0.65, 0.7, 0.8
MSAVI	$(NIR - Green)/(NIR + Green)$	0.7, 0.8
GNDVI	$(NIR - Green)/(NIR + Green)$	0.65–0.8

In the VIs showing a significant relationship with the biomass between the middle and low quantiles, MVI5, B, and W represented more comprehensive soil information, while VARI represented the amount of green vegetation. Combining these indices with the initial three vegetation indices, we found that the forest coverage in the middle and low quantiles was lower, and the reflectance values included the understory vegetation and soil. The VIs in the high and middle quantiles were the PVI, RDVI, DVI, and G. Among these, the relationship between the PVI and biomass was significant, indicating that the vegetation coverage was relatively high. However, the other three VIs were not disturbed by the understory soil, and the forest's vegetation reflectance could be characterized as forest normally.

4.2. Fitting Performance

The best fitting performance of the QRNN and QRF model in this study was in quantile 0.5, with R^2 values of 0.68 and 0.71, respectively. We took several papers that all selected the same study region and same species for comparison with our results. Our result was a little higher than that of the geographically weighted regression (GWR) model with the highest fitting accuracy of R^2 0.67 in the study of Ou et al. [20], and lower than the RF model

(R^2 = 0.73) that added the habitat dataset to improve the fitting performance in the research of Tang et al. [77]. The RF model (R^2 = 0.87) employed by Zhang et al. [78] demonstrated the best fitting performance in their research, incorporating the spectral conversion of the remote sensing data and topography variables to mitigate the influence of the terrain on the AGB estimation. In this study, only the QRNN and QRF models reconstructed by exploring the relationships between the potential variables and biomass across different quantiles were considered, while other variables, such as texture, terrain, and environment, were not considered. This finding further supports the significance of mining potential variables for accurate AGB estimation. Additionally, for the QRNN model, when the data were evenly distributed, the highest estimation accuracy was observed in the middle quantile range. For right-biased data distributions, it was more appropriate to utilize middle and low quantiles, whereas, for left-biased data distributions, the high quantiles yielded better results. Although the QRF model exhibited higher fitting accuracy overall, the scatterplot analysis revealed an overestimation trend in almost all the quantiles for biomass estimation using this model, thus indicating its inferior practical applicability compared to that of the QRNN model. Combining the minimum predicted error values from the different quantiles improved the tfitting accuracy significantly for both the QRNNb and QRFb models, with respective accuracies of 0.93 and 0.86, albeit slightly lower than those reported by Li et al. and Zhang et al. [46,79]. Nevertheless, this approach effectively mitigated the issues related to overestimation or underestimation.

4.3. Limitations and Future Research

Due to their strong penetration and backscattering capabilities, synthetic-aperture radar data remain unaffected by weather conditions and light saturation, making Sentinel-1 an ideal variable for estimating the AGB with improved accuracy [80]. Moreover, Sentinel-2 is considered to be superior to Landsat due to its higher resolution (10 m compared to Landsat's 30 m) and its additional four spectral bands (three red-edge bands and one narrow near-infrared band), which make it particularly sensitive to vegetation [81–84]. Considering the superior accuracy of the Landsat images at 30 m resolution, it can be inferred that Sentinel-2 offers even higher accuracy regarding AGB estimation. The other machine learning algorithm models (such as random forests and GXboost) and deep learning algorithms (convolutional neural networks, CNN) were considered to have better fitting performance regarding AGB estimation [43,77]. Additionally, the DEM and environmental factors were also applied to enhance the AGB estimation accuracy [77]; therefore, Sentinel images, climate, DEM, and other VIs data, combined with machine learning or deep learning algorithms, could be used to compare the AGB estimation capabilities in the future. Although this study focused solely on the *Pinus densata* forest in Shangri-La, exploring the applicability of these models in other regions with complex stands would be valuable for future research.

5. Conclusions

The exploration of the relationships between the vegetation indices and biomass in different quantiles holds great significance for improving the accuracy of biomass estimation through optical remote sensing. In this study, we focused on the *Pinus densata* forest in Shangri-La as our research subject. We conducted a significant analysis of 25 vegetation indices using the quantile regression method, followed by establishing the QRNN and QRF models based on vegetation indices that exhibited significant correlations with low, high, and overall quantiles. The findings are as follows:

(1) The blue, green, red, SWIR 1, and SWIR 2 bands demonstrated a substantial association with AGB within the quantiles ranging from 0.05 to 0.6; meanwhile, the NIR displayed significance with AGB within the quantiles ranging from 0.35 to 0.95.

(2) NDVI and its analogous complex vegetation index calculated from the green, red, and NIR bands were found to be unsuitable for estimating the AGB in the regions characterized by high heterogeneity; instead, the VIs utilizing blue and SWIR bands proved to be more suitable for the *Pinus densata* estimation.

(3) The QRNN and QRF models exhibited the highest fitting accuracy for the AGB in the 0.5 quantiles, with R^2 values of 0.68 and 0.71, respectively. While both models demonstrated promising performance, the QRNN model was deemed more suitable for this study due to its ability to effectively address the overestimation of lower values and underestimation of higher values compared to the QRF model, which tended to overestimate the biomass across all the quantiles.

Notably, significant variations were observed in the relationships between the vegetation indices and AGB across different quantiles. In this study, we approached the variable selection from a novel perspective, aiming to provide insights into improving the forest AGB estimation by mitigating the underestimation and overestimation issues.

Author Contributions: X.X., investigation, data curation, formal analysis, and writing—original draft. X.Z., investigation, data curation, and writing—original draft. S.S., writing—original draft, writing—review and editing, and supervision. G.Z., writing—review and editing and supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (grant number 32271874), the Yunnan Provincial Department of Education Science Research Fund Project (grant number 2023J0652), the State Forestry Administration Key Disciplines (forest human [2016] No. 21) and the Hunan Province Double First-Class Cultivation Disciplines (Hunan education [2018] No. 469).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. 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. [CrossRef]
2. Han, H.; Wan, R.; Li, B. Estimating Forest Aboveground Biomass Using Gaofen-1 Images, Sentinel-1 Images, and Machine Learning Algorithms: A Case Study of the Dabie Mountain Region, China. *Remote Sens.* **2021**, *14*, 176. [CrossRef]
3. Wang, Y.; Wu, G.; Deng, L.; Tang, Z.; Wang, K.; Sun, W.; Shangguan, Z. Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Sci. Rep.* **2017**, *7*, 6940. [CrossRef] [PubMed]
4. Zhang, X.; Shen, H.; Huang, T.; Wu, Y.; Guo, B.; Liu, Z.; Luo, H.; Tang, J.; Zhou, H.; Wang, L.; et al. Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery. *Ecol. Indic.* **2024**, *159*, 111752. [CrossRef]
5. Feng, H.; Chen, Q.; Hu, Y.; Du, Z.; Lin, G.; Wang, C.; Huang, Y. Estimation of forest aboveground biomass by using mixed-effects model. *Int. J. Remote Sens.* **2021**, *42*, 8675–8690. [CrossRef]
6. Sun, S.; Wang, Y.; Song, Z.; Chen, C.; Zhang, Y.; Chen, X.; Chen, W.; Yuan, W.; Wu, X.; Ran, X.; et al. Modelling Aboveground Biomass Carbon Stock of the Bohai Rim Coastal Wetlands by Integrating Remote Sensing, Terrain, and Climate Data. *Remote Sens.* **2021**, *13*, 4321. [CrossRef]
7. Lu, D.; Chen, Q.; Wang, G.; Moran, E.; Batistella, M.; Zhang, M.; Laurin, G.V.; Saah, D. Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *Int. J. For. Rev.* **2012**, *2012*, 436537. [CrossRef]
8. Listopad, C.M.C.S.; Drake, J.B.; Masters, R.E.; Weishampel, J.F. Portable and Airborne Small Footprint LiDAR: Forest Canopy Structure Estimation of Fire Managed Plots. *Remote Sens.* **2011**, *3*, 1284–1307. [CrossRef]
9. Mahlangu, P.; Mathieu, R.; Wessels, K.; Naidoo, L.; Verstraete, M.; Asner, G.; Main, R. Indirect Estimation of Structural Parameters in South African Forests Using MISR-HR and LiDAR Remote Sensing Data. *Remote Sens.* **2018**, *10*, 1537. [CrossRef]
10. Xu, D.; Wang, H.; Xu, W.; Luan, Z.; Xu, X. LiDAR Applications to Estimate Forest Biomass at Individual Tree Scale: Opportunities, Challenges and Future Perspectives. *Forests* **2021**, *12*, 550. [CrossRef]
11. Zhou, Y.; Luo, J.; Feng, L.; Yang, Y.; Chen, Y.; Wu, W. Long-short-term-memory-based crop classification using high-resolution optical images and multi-temporal SAR data. *GIScience Remote Sens.* **2019**, *56*, 1170–1191. [CrossRef]
12. Li, Y.; Li, M.; Liu, Z.; Li, C. Combining Kriging Interpolation to Improve the Accuracy of Forest Aboveground Biomass Estimation Using Remote Sensing Data. *IEEE Access* **2020**, *8*, 128124–128139. [CrossRef]
13. López-Serrano, P.; Corral-Rivas, J.; Díaz-Varela, R.; Álvarez-González, J.; López-Sánchez, C. Evaluation of Radiometric and Atmospheric Correction Algorithms for Aboveground Forest Biomass Estimation Using Landsat 5 TM Data. *Remote Sens.* **2016**, *8*, 369. [CrossRef]
14. González-Jaramillo, V.; Fries, A.; Zeilinger, J.; Homeier, J.; Paladines-Benitez, J.; Bendix, J. Estimation of Above Ground Biomass in a Tropical Mountain Forest in Southern Ecuador Using Airborne LiDAR Data. *Remote Sens.* **2018**, *10*, 660. [CrossRef]
15. Sagang, L.B.T.; Ploton, P.; Sonké, B.; Poilvé, H.; Couteron, P.; Barbier, N. Airborne Lidar Sampling Pivotal for Accurate Regional AGB Predictions from Multispectral Images in Forest-Savanna Landscapes. *Remote Sens.* **2020**, *12*, 1637. [CrossRef]

16. Zhao, P.; Lu, D.; Wang, G.; Wu, C.; Huang, Y.; Yu, S. Examining Spectral Reflectance Saturation in Landsat Imagery and Corresponding Solutions to Improve Forest Aboveground Biomass Estimation. *Remote Sens.* **2016**, *8*, 469. [CrossRef]

17. Li, X.; Du, H.; Mao, F.; Zhou, G.; Chen, L.; Xing, L.; Fan, W.; Xu, X.; Liu, Y.; Cui, L.; et al. Estimating bamboo forest aboveground biomass using EnKF-assimilated MODIS LAI spatiotemporal data and machine learning algorithms. *Agric. For. Meteorol.* **2018**, *256–257*, 445–457. [CrossRef]

18. Niu, X.; Zeng, Q.; Luo, X.; Chen, L. FCAU-Net for the Semantic Segmentation of Fine-Resolution Remotely Sensed Images. *Remote Sens.* **2022**, *14*, 215. [CrossRef]

19. Lourenço, P.; Godinho, S.; Sousa, A.; Gonçalves, A.C. Estimating tree aboveground biomass using multispectral satellite-based data in Mediterranean agroforestry system using random forest algorithm. *Remote Sens. Appl. Soc. Environ.* **2021**, *23*, 100560. [CrossRef]

20. Ou, G.; Lv, Y.; Xu, H.; Wang, G. Improving Forest Aboveground Biomass Estimation of *Pinus densata* Forest in Yunnan of Southwest China by Spatial Regression using Landsat 8 Images. *Remote Sens.* **2019**, *11*, 2750. [CrossRef]

21. Alizadeh, M.; Zabihi, H.; Rezaie, F.; Asadzadeh, A.; Wolf, I.D.; Langat, P.K.; Khosravi, I.; Beiranvand Pour, A.; Mohammad Nataj, M.; Pradhan, B. Earthquake Vulnerability Assessment for Urban Areas Using an ANN and Hybrid SWOT-QSPM Model. *Remote Sens.* **2021**, *13*, 4519. [CrossRef]

22. Axelsson, C.; Skidmore, A.K.; Schlerf, M.; Fauzi, A.; Verhoef, W. Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. *Int. J. Remote Sens.* **2012**, *34*, 1724–1743. [CrossRef]

23. Beaudoin, A.; Hall, R.J.; Castilla, G.; Filiatrault, M.; Villemaire, P.; Skakun, R.; Guindon, L. Improved k-NN Mapping of Forest Attributes in Northern Canada Using Spaceborne L-Band SAR, Multispectral and LiDAR Data. *Remote Sens.* **2022**, *14*, 1181. [CrossRef]

24. Ou, G.; Li, C.; Lv, Y.; Wei, A.; Xiong, H.; Xu, H.; Wang, G. Improving Aboveground Biomass Estimation of *Pinus densata* Forests in Yunnan Using Landsat 8 Imagery by Incorporating Age Dummy Variable and Method Comparison. *Remote Sens.* **2019**, *11*, 738. [CrossRef]

25. Yadav, S.; Padalia, H.; Sinha, S.K.; Srinet, R.; Chauhan, P. Above-ground biomass estimation of Indian tropical forests using X band Pol-InSAR and Random Forest. *Remote Sens. Appl. Soc. Environ.* **2021**, *21*, 100462. [CrossRef]

26. de Almeida, C.T.; Galvão, L.S.; Ometto, J.P.H.B.; Jacon, A.D.; de Souza Pereira, F.R.; Sato, L.Y.; Lopes, A.P.; de Alencastro Graça, P.M.L.; de Jesus Silva, C.V.; Ferreira-Ferreira, J.; et al. Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian Amazon using different regression algorithms. *Remote Sens. Environ.* **2019**, *232*, 111323. [CrossRef]

27. Lu, D.; Batistella, M.; Moran, E. Satellite estimation of aboveground biomass and impacts of forest stand structure. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 967–974. [CrossRef]

28. Kim, Y.; Kim, T.-H.; Ergün, T. The instability of the Pearson correlation coefficient in the presence of coincidental outliers. *Financ. Res. Lett.* **2015**, *13*, 243–257. [CrossRef]

29. Zhao, P.; Lu, D.; Wang, G.; Liu, L.; Li, D.; Zhu, J.; Yu, S.Q. Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *53*, 1–15. [CrossRef]

30. Koenker, R.; Bassett, G. Regression Quantiles. *J. Econo. Soci.* **1978**, *46*, 33–50. [CrossRef]

31. Das, K.; Krzywinski, M.; Altman, N. Quantile regression. *Nat. Methods* **2019**, *16*, 451–452. [CrossRef] [PubMed]

32. Tian, D.; Bi, H.; Jin, X.; Li, F. Stochastic frontiers or regression quantiles for estimating the self-thinning surface in higher dimensions? *J. For. Res.* **2020**, *32*, 1515–1533. [CrossRef]

33. Cade, B.S.; Noon, B.R. A gentle introduction to quantile regression for ecologists. *Front. Ecol. Environ.* **2003**, *1*, 412–420. [CrossRef]

34. Liang, Z.; Xu, Y.; Qiu, Q.; Liu, Y.; Lu, W.; Tyler, W. A framework to develop joint nutrient criteria for lake eutrophication management in eutrophic lakes. *J. Hydrol.* **2021**, *594*, 125883. [CrossRef]

35. Simkin, S.M.; Allen, E.B.; Bowman, W.D.; Clark, C.M.; Belnap, J.; Brooks, M.L.; Cade, B.S.; Collins, S.L.; Geiser, L.H.; Gilliam, F.S.; et al. Conditional vulnerability of plant diversity to atmospheric nitrogen deposition across the United States. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 4086–4091. [CrossRef] [PubMed]

36. Barbosa, S.M.; Scotto, M.G.; Alonso, A.M. Summarising changes in air temperature over Central Europe by quantile regression and clustering. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 3227–3233. [CrossRef]

37. Jagger, T.H.; Elsner, J.B. Modeling tropical cyclone intensity with quantile regression. *Int. J. Climatol.* **2009**, *29*, 1351–1361. [CrossRef]

38. Timofeev, A.A.; Sterin, A.M. Using the quantile regression method to analyze changes in climate characteristics. *Russ. Meteorol. Hydrol.* **2010**, *35*, 310–319. [CrossRef]

39. Gao, M.; Franzke, C.L.E. Quantile Regression–Based Spatiotemporal Analysis of Extreme Temperature Change in China. *J. Clim.* **2017**, *30*, 9897–9914. [CrossRef]

40. Mazvimavi, D. Investigating changes over time of annual rainfall in Zimbabwe. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 2671–2679. [CrossRef]

41. Tan, X.; Gan, T.; Chen, S.; Liu, B. Modeling distributional changes in winter precipitation of Canada using Bayesian spatiotemporal quantile regression subjected to different teleconnections. *Clim. Dyn.* **2018**, *52*, 2105–2124. [CrossRef]

42. Chang, Y.; Bourque, C.P.A. Relating modelled habitat suitability for *Abies balsamea* to on-the-ground species structural characteristics in naturally growing forests. *Ecol. Indic.* **2020**, *111*, 105981. [CrossRef]

43. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]

44. Wang, X.; Liu, C.; Lv, G.; Xu, J.; Cui, G. Integrating Multi-Source Remote Sensing to Assess Forest Aboveground Biomass in the Khingan Mountains of North-Eastern China Using Machine-Learning Algorithms. *Remote Sens.* **2022**, *14*, 1039. [CrossRef]

45. Taylor, J.W. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *J. Forecast.* **2000**, *19*, 299–311. [CrossRef]

46. Li, L.; Zhou, B.; Liu, Y.; Wu, Y.; Tang, J.; Xu, W.; Wang, L.; Ou, G. Reduction in Uncertainty in Forest Aboveground Biomass Estimation Using Sentinel-2 Images: A Case Study of Pinus densata Forests in Shangri-La City, China. *Remote Sens.* **2023**, *15*, 559. [CrossRef]

47. Wang, F.; Huang, J.; Chen, L. Development of a Vegetation Index for Estimation of Leaf Area Index Based on Simulation Modeling. *J. Plant Nutr.* **2010**, *33*, 328–338. [CrossRef]

48. Naji, T.A.H. Study of vegetation cover distribution using DVI, PVI, WDVI indices with 2D-space plot. *J. Phys. Conf. Ser.* **2018**, *1003*, 012083. [CrossRef]

49. Jackson, R.D.; Slater, P.N.; Pinter, P.J. Discrimination of Growth and Water Stress in Wheat by Various Vegetation Indices through Clear and Turbid Atmospheres. *Remote Sens. Environ.* **1983**, *13*, 187–208. [CrossRef]

50. Rouse, J.W.; Hass, R.H.; Schell, J.A.; Deering, D.W. Monitoring vege tation systems in the great plains with ERTS. *Third Earth Resour. Technol. Satell. Symp.* **1973**, *1*, 309–317.

51. Roujean, J.L.; Breon, F.M. Estimating PAR Absorbed by Vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* **1995**, *53*, 375–384. [CrossRef]

52. Mutanga, O.; Skidmore, A.K. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int. J. Remote Sens.* **2004**, *25*, 3999–4014. [CrossRef]

53. Chen, J.M. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Can. J. Remote Sens.* **1996**, *22*, 229–242. [CrossRef]

54. Gitelson, A.A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *J. Plant Physiol.* **2004**, *161*, 165–173. [CrossRef] [PubMed]

55. Gitelson, A.A.; Yoram, J.K.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of global vegetation form EOD-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]

56. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]

57. Rondeaux, G.; Steven, M.; Baret, F. Optimization of Soil-Adjusted Vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [CrossRef]

58. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87. [CrossRef]

59. Kaufman, Y.J.; Tanre, D. Atmospherically Resistant Vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 261–270. [CrossRef]

60. Bannari, A.; Morin, D.; Bonn, F.; Huete, A.R. A review of vegetation indices. *Remote Sens. Rev.* **1995**, *13*, 95–120. [CrossRef]

61. Miura, T.; Huete, A.R.; Yoshioka, H. Evaluation of Sensor Calibration Uncertainties on vegetation indices for MODIS. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 1399–1409. [CrossRef]

62. Mondejar, J.P.; Tongco, A.F. Near infrared band of Landsat 8 as water index: A case study around Cordova and Lapu-Lapu City, Cebu, Philippines. *Sustain. Environ. Res.* **2019**, *29*, 16. [CrossRef]

63. Baig, M.H.A.; Zhang, L.; Shuai, T.; Tong, Q. Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sens. Lett.* **2014**, *5*, 423–431. [CrossRef]

64. Bognár, P.; Kern, A.; Pásztor, S.; Steinbach, P.; Lichtenberger, J. Testing the Robust Yield Estimation Method for Winter Wheat, Corn, Rapeseed, and Sunflower with Different Vegetation Indices and Meteorological Data. *Remote Sens.* **2022**, *14*, 2860. [CrossRef]

65. Freitas, S.R.; Mello, M.C.S.; Cruz, C.B.M. Relationships between forest structure and vegetation indices in Atlantic Rainforest. *For. Ecol. Manag.* **2005**, *218*, 353–362. [CrossRef]

66. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]

67. Crippen, R.E. Calculating the vegetation index faster. *Remote Sens. Environ.* **1990**, *34*, 71–73. [CrossRef]

68. Majasalmi, T.; Rautiainen, M. The potential of Sentinel-2 data for estimating biophysical variables in a boreal forest: A simulation study. *Remote Sens. Lett.* **2016**, *7*, 427–436. [CrossRef]

69. Jurgens, C. The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data. *Int. J. Remote Sens.* **1997**, *18*, 3583–3594. [CrossRef]

70. Freeman, E.A.; Moisen, G.G. An Application of Quantile Random Forests for Predictive Mapping of Forest Attributes. In Proceedings of the New Directions in Inventory Techniques & Applications Forest Inventory & Analysis (FIA) Symposium, Portland, Oregon, 8–10 December 2015; p. 362.

71. Hauke, J.; Kossowski, T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quageo* **2011**, *30*, 87–93. [CrossRef]

72. Breiman, L. Random forests. *Mach. Learn.* **2001**, *34*, 5–32. [CrossRef]

73. Pham, T.D.; Le, N.N.; Ha, N.T.; Nguyen, L.V.; Xia, J.; Yokoya, N.; To, T.T.; Trinh, H.X.; Kieu, L.Q.; Takeuchi, W. Estimating Mangrove Above-Ground Biomass Using Extreme Gradient Boosting Decision Trees Algorithm with Fused Sentinel-2 and ALOS-2 PALSAR-2 Data in Can Gio Biosphere Reserve, Vietnam. *Remote Sens.* **2020**, *12*, 777. [CrossRef]

74. Lu, D.; Mausel, P.; Brondízio, E.; Moran, E. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *For. Ecol. Manag.* **2004**, *198*, 149–167. [CrossRef]

75. Hall, F.G.; Shimabukuro, E.; Huemmrichughes, K.F. Remote sensing of forest biophysical structure using mixture decomposition and geometric reflectance models. *Ecol. Appl.* **1995**, *5*, 993–1013. [CrossRef]

76. Sader, S.A.; Waide, R.B.; Lawrence, W.T.; Joyce, A.T. Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat TM data. *Remote Sens. Environ.* **1989**, *28*, 143–156. [CrossRef]

77. Tang, J.; Liu, Y.; Li, L.; Liu, Y.; Wu, Y.; Xu, H.; Ou, G. Enhancing Aboveground Biomass Estimation for Three Pinus Forests in Yunnan, SW China, Using Landsat 8. *Remote Sens.* **2022**, *14*, 4589. [CrossRef]

78. Zhang, J.; Lu, C.; Xu, H.; Wang, G. Estimating aboveground biomass of *Pinus densata*-dominated forests using Landsat time series and permanent sample plot data. *J. For. Res.* **2018**, *30*, 1689–1706. [CrossRef]

79. Zhang, X.; Li, L.; Liu, Y.; Wu, Y.; Tang, J.; Xu, W.; Wang, L.; Ou, G. Improving the accuracy of forest aboveground biomass using Landsat 8 OLI images by quantile regression neural network for *Pinus densata* forests in southwestern China. *Front. For. Glob. Chang.* **2023**, *6*, 1162291. [CrossRef]

80. Huang, X.; Ziniti, B.; Torbick, N.; Ducey, M. Assessment of Forest above Ground Biomass Estimation Using Multi-Temporal C-band Sentinel-1 and Polarimetric L-band PALSAR-2 Data. *Remote Sens.* **2018**, *10*, 1424. [CrossRef]

81. Abdullah, H.; Skidmore, A.K.; Darvishzadeh, R.; Heurich, M.; Pettorelli, N.; Disney, M. Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (*Ips typographus*, L.) compared with Landsat-8. *Remote Sens. Ecol. Conserv.* **2018**, *5*, 87–106. [CrossRef]

82. Castillo, J.A.A.; Apan, A.A.; Maraseni, T.N.; Salmo, S.G. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 70–85. [CrossRef]

83. Chen, Y.; Guerschman, J.; Shendryk, Y.; Henry, D.; Harrison, M.T. Estimating Pasture Biomass Using Sentinel-2 Imagery and Machine Learning. *Remote Sens.* **2021**, *13*, 603. [CrossRef]

84. Cutler, M.E.J.; Boyd, D.S.; Foody, G.M.; Vetrivel, A. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 66–77. [CrossRef]