

## Article

# Comparative Analysis of Modeling Algorithms for Forest Aboveground Biomass Estimation in a Subtropical Region

Yukun Gao <sup>1</sup>, Dengsheng Lu <sup>1,2,\*</sup> , Guiying Li <sup>1</sup> , Guangxing Wang <sup>3</sup>, Qi Chen <sup>4</sup> ,  
Lijuan Liu <sup>1</sup> and Dengqiu Li <sup>1</sup>

<sup>1</sup> State Key Laboratory of Subtropical Silviculture, Key Laboratory of Carbon Cycling in Forest Ecosystems and Carbon Sequestration of Zhejiang Province, School of Environmental & Resource Sciences, Zhejiang Agriculture and Forestry University, Hangzhou 311300, China; gykex123@163.com (Y.G.); gyli@zafu.edu.cn (G.L.); llj7885@163.com (L.L.); lidengqiu001@163.com (D.L.)

<sup>2</sup> Center for Global Change and Earth Observations, Michigan State University, East Lansing, MI 48823, USA

<sup>3</sup> Department of Geography, Southern Illinois University at Carbondale, IL 62901, USA; gxwang@siu.edu

<sup>4</sup> Department of Geography, University of Hawaii at Manoa, Honolulu, HI 96822, USA; qichen@hawaii.edu

\* Correspondence: luds@zafu.edu.cn; Tel./Fax: +86-571-6374-6366

Received: 9 February 2018; Accepted: 17 April 2018; Published: 18 April 2018



**Abstract:** Remote sensing-based forest aboveground biomass (AGB) estimation has been extensively explored in the past three decades, but how to effectively combine different sensor data and modeling algorithms is still poorly understood. This research conducted a comparative analysis of different datasets (e.g., Landsat Thematic Mapper (TM), ALOS PALSAR L-band data, and their combinations) and modeling algorithms (e.g., artificial neural network (ANN), support vector regression (SVR), Random Forest (RF), k-nearest neighbor (kNN), and linear regression (LR)) for AGB estimation in a subtropical region under non-stratification and stratification of forest types. The results show the following: (1) Landsat TM imagery provides more accurate AGB estimates (root mean squared error (RMSE) values in 27.7–29.3 Mg/ha) than ALOS PALSAR (RMSE values in 30.3–33.7 Mg/ha). The combination of TM and PALSAR data has similar performance for ANN and SVR, worse performance for RF and KNN, and slightly improved performance for LR. (2) Overestimation for small AGB values and underestimation for large AGB values are major problems when using the optical (e.g., Landsat) or radar (e.g., ALOS PALSAR) data. (3) LR is still an important tool for AGB modeling, especially for the AGB range of 40–120 Mg/ha. Machine learning algorithms have limited effects on improving AGB estimation overall, but ANN can improve AGB modeling when AGB values are greater than 120 Mg/ha. (4) Forest type and AGB range are important factors that influence AGB modeling performance. (5) Stratification based on forest types improved AGB estimation, especially when AGB was greater than 160 Mg/ha, using the LR approach. This research provides new insight for remote sensing-based AGB modeling for the subtropical forest ecosystem through a comprehensive analysis of different source data, modeling algorithms, and forest types. It is critical to develop an optimal AGB modeling procedure, including the collection of a sufficient number of sample plots, extraction of suitable variables and modeling algorithms, and evaluation of the AGB estimates.

**Keywords:** aboveground biomass; linear regression; machine learning algorithms; remote sensing imagery; stratification; subtropical forests

## 1. Introduction

Forest biomass is one of the important variables needed to quantify the structure and function of forest ecosystems [1]. A large number of studies using different technologies such as field measurements, remote sensing, and process-based ecosystem models have been conducted to estimate forest biomass at local, regional, and global scales [2]. Remotely sensed imagery can characterize land surface features over large areas and it has been extensively used for estimating forest biomass, especially aboveground biomass (AGB). However, the data limitations (e.g., data saturation for optical and radar data, and the limitations in spectral and spatial resolutions) and the complex relationships between AGB and spectral variables make AGB estimation inaccurate, especially when AGB values are higher than 150 Mg/ha or lower than 40 Mg/ha [3].

In remote sensing-based biomass modeling, the collection of a sufficient number of high-quality sample plots, the selection of suitable variables, and the selection of modeling algorithms are three critical steps [2]. Sample plots provide fundamental data for accurately estimating AGB. Once sample plot data are collected, research mainly focuses on the identification of suitable variables and modeling algorithms. Many previous studies have examined the importance of selecting suitable variables (e.g., spectral bands, vegetation indices, texture measures, and subpixel features) in improving AGB estimation [4–6]. Stepwise regression analysis is often used for the identification of variables for AGB modeling [2]. Previous research has indicated that the combination of spectral responses and texture variables can improve AGB estimation compared with using a single kind of imagery alone, especially in the tropical and subtropical regions with complex forest stand structure and composition of tree species [3,4]. Texture measures are especially valuable for sites with complex forest stand structures [4]. Another common approach to identify variables for AGB modeling is Random Forest (RF) because it can provide the ranks of importance of variables [7–10]. The stepwise regression is simple and easy to use but the identified variables are only those having a linear relationship with AGB, while the selected variables using RF can have nonlinear relationships with AGB.

In AGB modeling, in addition to the selection of variables, another critical step is to identify an appropriate algorithm to establish AGB estimation models. Lu et al. [2] summarized the major AGB modeling algorithms, covering regression and machine learning algorithms. Regression has been widely used for AGB estimation [3,4,6]. Because of its simplification and relatively good performance, this method is often used in AGB estimation for different ecosystems such as tropical, subtropical, and temperate forests [3,4,6,9]. However, linear regression (LR) methods are based on linear relationships of AGB with predictors or independent variables, thus, may not provide satisfactory results due to the complex relationships between remote sensing variables and AGB [1,3,11]. In this case, nonparametric and machine learning algorithms, such as artificial neural network (ANN), support vector regression (SVR), k-nearest neighbor (kNN), and RF, can deal with nonlinear relationships. Thus, these algorithms have gained increasing attention in the past decade [2,9,12–14] and have been widely employed in forest AGB estimation [15–17]. For example, when AGB is nonlinearly related to remote sensing variables, LR models cannot provide reliable AGB estimation; instead, machine learning algorithms such as ANN can establish nonlinear relationships and provide more accurate AGB estimates [18,19], especially when ancillary variables such as topographic factors (e.g., elevation, slope, aspect) are used [20]. ANN has the advantages of distributed parallel processing and nonlinear and adaptive learning and may outperform LR models in AGB estimation [13], but its characteristics as a black box that result in difficult variable explanation [2] and inability in training convergence and function approximation [19] are often criticized. Compared to traditional LR, SVR has become an important approach for AGB modeling in the past decade [21,22] because of its capability in dealing with a relatively small number of sample plots. Through the selection of a kernel function to realize the data space projection from a low dimension to a high dimension, SVR solves the problems of linear inseparability in sample plots, reduces the structural risk, and leads to a greater ability for processing nonlinear data. In addition, SVR can be highly generalized for dealing with high dimensional data [23,24].

The RF algorithm is an extension of the traditional decision tree method through combining multiple decision trees [8], having the advantage of dealing with a large amount of data with high speed and efficiency, and the ability to rank the importance of variables [7,25,26]. Previous research has shown that RF can lead to higher estimation accuracies and smaller errors than LR [21,24,27,28]. Another commonly used method for AGB estimation is kNN [14,29–35]. This method calculates the weighted average of forest AGB values of  $k$  nearest sample plots based on spectral distances between the plot locations and the estimated pixel using Euclidean or Mahalanobis distance. The more similar the sample plots are to the estimated pixel, the greater the weights. Without the assumptions of normal distribution and linear relationship, kNN can be utilized to estimate both continuous and discrete forest variables. Moreover, the predictors used in kNN can be any remote sensing and environmental variables, including spectral bands, vegetation indices, soil properties, and topographic features [33,34]. Previous research showed that the kNN method may produce poor AGB estimates at the pixel scale but can improve estimation at a coarser scale [14,36,37]. Considerable research for the use of kNN has been conducted to improve the accuracy of estimating forest variables by optimizing the selection of kNNs and using different distances and weighting methods [32–35].

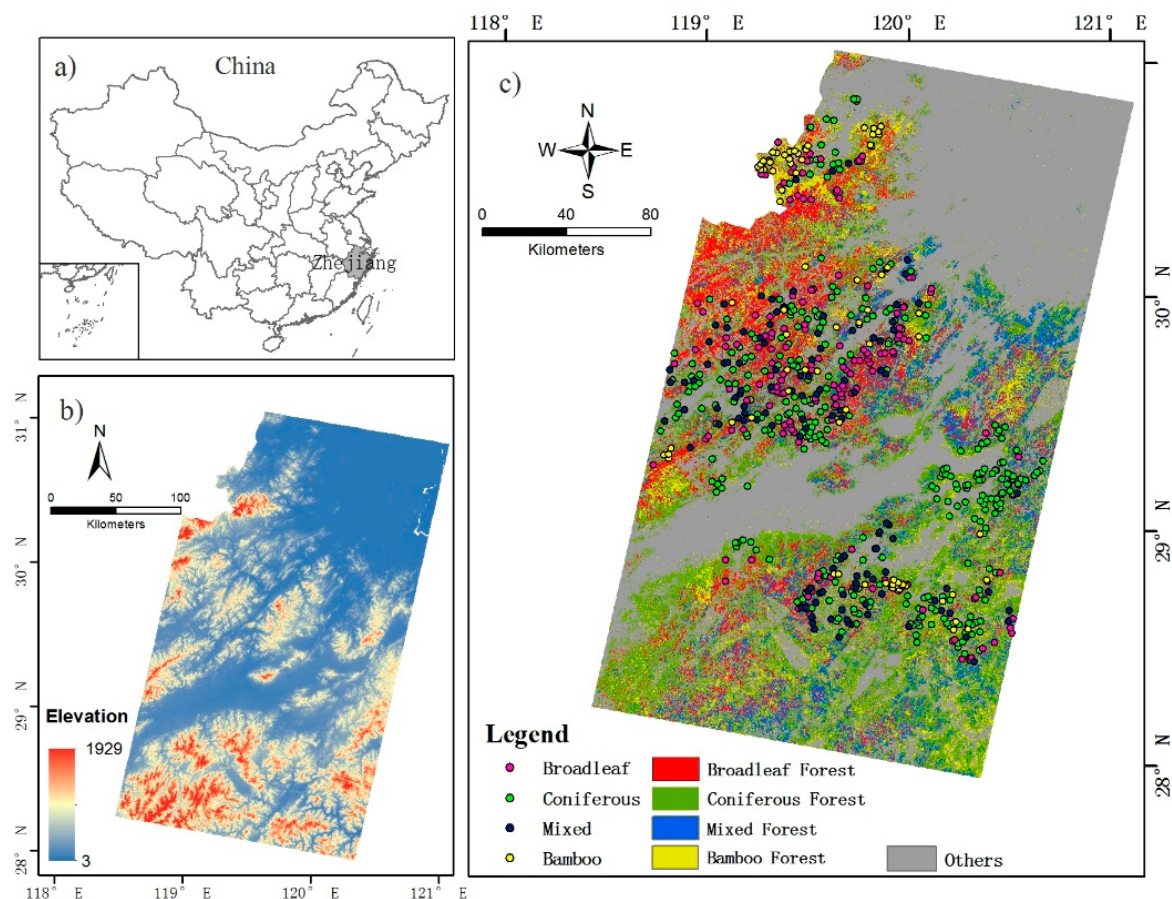
The advantages of using nonparametric and machine learning algorithms for AGB estimation have been recognized, but their estimation accuracy highly depends on the optimization of the parameters used in the relevant algorithms and representativeness of the samples. For example, ANN requires a large number of sample plots, and a small number of sample plots may lead to poor predictions [19]. SVR is found to have poor performance due to the lack of a standard approach to optimize the model parameters [38]. RF algorithm might have poor generalization and large estimation errors when the number of variables with low correlation to AGB in the model increases [25]. For a specific study area and a set of remote sensing data, it is unclear which algorithm may produce the most accurate results. It is necessary to conduct a comparative analysis of different modeling algorithms for AGB estimation.

Zhao et al. [6] used LR and explored the role of using Landsat TM images, ALOS PALSAR data, and their combination, plus data fusion to improve AGB estimation performance with and without classification of forest types. The authors found that the combination of TM and PALSAR data as extra bands increased the estimation accuracy of AGB, and the stratification of vegetation types also improved AGB estimation performance, but LR led to significant overestimation and underestimation for the small and large AGB values (e.g., less than 40 Mg/ha and greater than 160 Mg/ha), respectively. It is unknown whether the conclusions are upheld when other AGB modeling methods are used. We also have to answer the following questions: (1) Which modeling algorithm should be used for different datasets, that is, can machine learning algorithms provide better AGB estimation performance than LR? (2) Can stratification of forest types improve AGB estimation performance in the case of using machine learning and nonparametric algorithms? There are rare reports on comprehensively comparative analysis of modeling algorithms based on different datasets under non-stratification and stratification of forest types. Therefore, the objective of this research is to understand the AGB modeling algorithms that are appropriate for the subtropical region by using Landsat TM and ALOS PALSAR L-band data and by conducting a comparative analysis of different algorithms (e.g., RF, ANN, SVR, kNN, LR) under non-stratification and stratification of forest types. In addition, we attempt to examine which modeling method can improve the overestimation and underestimation that often takes place for small and large AGB values, respectively, when LR is used. Through this comprehensive comparison, we can better understand the AGB modeling mechanisms by employing suitable remote sensing variables and modeling algorithms under non-stratification and stratification of forest types.

## 2. Study Area

Zhejiang province has complex topographic conditions with mountainous and hilly areas in the southwest and plain and basin areas in the northeast. Mountainous and hilly areas account for approximately 70%, the plain and basin areas for 23%, and water for 7%. The mountains having elevation of greater than 1000 m are mainly located in southwestern Zhejiang, the hilly areas in the

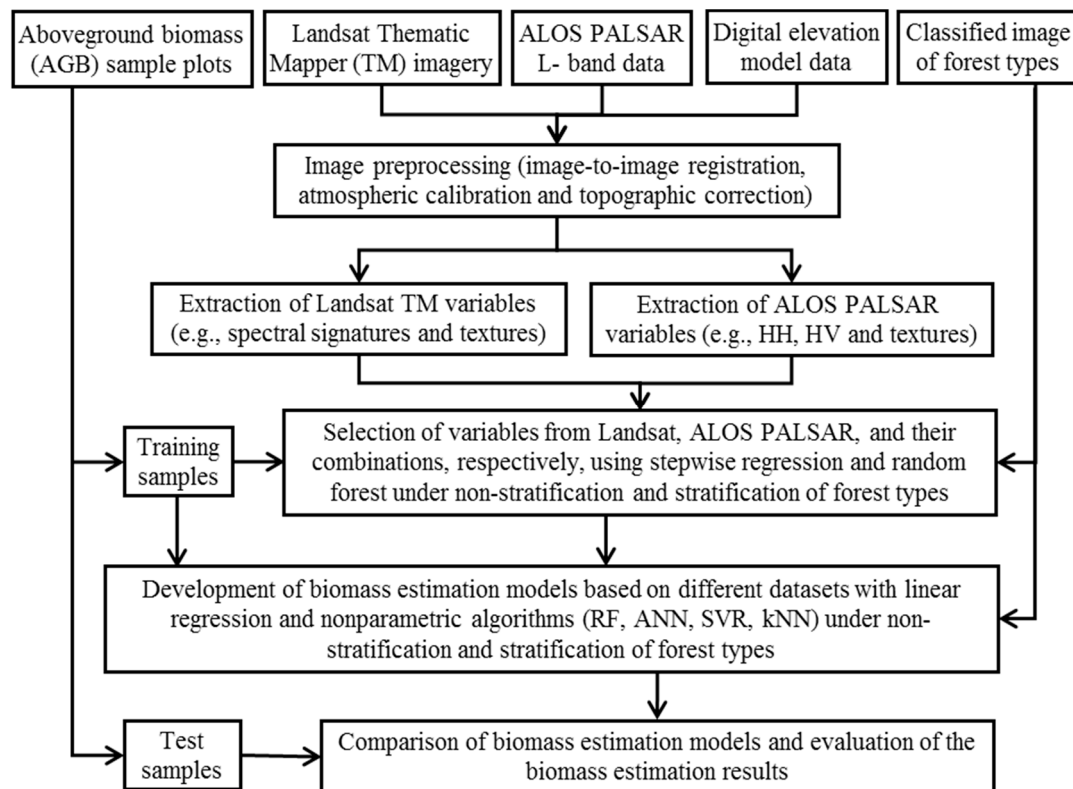
central part, and the differently sized basins are dispersed in different regions (Figure 1). This province has a subtropical moist monsoon climate with average annual temperatures of 15–18 °C and average annual precipitation of 980–2000 mm. This province has forested areas of 6680 Kha with forest coverage of approximately 60%. The dominant forest types include coniferous forests (pines and firs as the dominant trees with nearly uniform forest stand structure), broadleaf forests (dominated by evergreen broadleaf trees with complex forest canopy layers), mixed needle and broadleaf forests (two or more dominant tree species, usually with a pine overstory, broadleaf middle layer, and shrub lower layer), and bamboo forests (usually pure bamboo species).



**Figure 1.** (a) Zhejiang Province in Eastern China; (b) elevations of the study area; and (c) forest classification map for the study area based on the 2010 Landsat 5 Thematic Mapper and spatial locations of sample plots.

### 3. Materials and Methods

The framework of conducting a comparative analysis of AGB modeling approaches is illustrated in Figure 2. The major steps include (1) data preparation of different sources (i.e., geometric registration between Landsat, ALOS PALSAR, and digital elevation model (DEM); atmospheric and topographic corrections); (2) selection of the variables from Landsat, ALOS PALSAR, and DEM data; (3) development of AGB estimation models using different algorithms (i.e., LR, RF, ANN, SVR, and kNN) based on stratification and non-stratification; and (4) comparison and evaluation of the AGB modeling results.



**Figure 2.** Strategy of identifying suitable biomass modeling approaches through a comparative analysis of modeling algorithms using different source data under non-stratification and stratification of forest types (Note: HH, horizontally transmitted and received; HV, horizontally transmitted and vertically received; RF, Random Forest; ANN, artificial neural networks; SVR, support vector regression; kNN, k-nearest neighbor).

### 3.1. Data Collection and Preprocessing

The datasets used in this research are summarized in Table 1. All these data were registered in the Universal Transverse Mercator coordinate system (zone 50 north). For Landsat TM imagery, the dark object subtraction approach was used to convert digital numbers of pixels to surface reflectance [39,40] and the C-correction approach was used to conduct topographic correction [41]. The global digital elevation model (GDEM) data with 30 m spatial resolution from the United States Geological Survey were used in the topographic correction. For ALOS PALSAR data, we downloaded the 2010 global mosaic data with cell size of 25 m. The PALSAR data were co-registered to Landsat imagery. During the image-to-image registration, the PALSAR data were resampled from a 25 m to 30 m cell size using the nearest neighbor algorithm so that both TM and PALSAR data had the same cell size. Because of the speckle problem, the enhanced Lee filtering approach with a window size of  $3 \times 3$  pixels was used to reduce the speckles inherent in the PALSAR data [42]. The forest (coniferous, broadleaf, mixed, and bamboo) classification image, which was developed from the 2010 Landsat image using a hybrid approach [3], was directly used in this research. A detailed description of this classified image is provided in [3].

**Table 1.** Datasets used in this research.

| Dataset                     | Description                                                                                                                                                                                                                                                                                                                                                                                                  |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Landsat 5 TM imagery        | Two scenes of TM images (Path/row: 119/39 and 119/40) with a 30-m spatial resolution were acquired on 24 May 2010. Six spectral bands were used to develop the forest classification map [3] and for AGB estimation in this research                                                                                                                                                                         |
| ALOS PALSAR L-band          | The FBD (fine beam double polarization, HH/HV) L-band 1.5 product with 25-m cell size was downloaded from the global mosaic data with a time interval of 1 year. This downloaded image was produced using the 2010 PALSAR images. The cell size of 25-m was resampled to a 30-m cell size during the PALSAR-to-TM image registration                                                                         |
| ASTER GDEM data             | Global digital elevation model (GDEM) data with a 30-m spatial resolution were downloaded from the United States Geological Survey website                                                                                                                                                                                                                                                                   |
| Forest classification image | The forest types in this study area included pine and fir (coniferous forest), broadleaf forest, mixed forest, and bamboo. The forest distribution map was developed from a Landsat TM image using a hybrid approach [3] with an overall classification accuracy of 78%. Forest classes had user's accuracy between 71 and 87% and producer's accuracy between 72 and 87%. More details are provided in [3]. |
| Field measurements          | A total of 664 sample plots covering coniferous, broadleaf, mixed, and bamboo forests were inventoried in 2010 and 2011 [3]                                                                                                                                                                                                                                                                                  |

Note: HH, horizontally transmitted and received; HV, horizontally transmitted and vertically received.

The field inventory work was conducted at county level. Based on the forest distribution map at the subcompartment scale, sample plots were systematically allocated on this map according to the number of subcompartments, sampling interval, and proportion of samples. The sampling ratio depends on the number of subcompartments; for example, the ratios were designed as 3%, 2.5%, 2%, 1.5%, and 1% for the number of compartments of less than 1000, 1000–2000, 2000–4000, 4000–6000, and greater than 6000. Taking the number of subcompartments of 901 as an example, the number of sample plots is 30; thus, the coding of sampling subcompartment is 1, 31, 61, . . . , and 901. Figure 3 illustrates the concept of the sampling approach used in this research. After a subcompartment was selected, one plot with 20 m by 20 m and three nested subplots with 2 m by 2 m were allocated near the central area of this subcompartment, representing the forest type. Sample plots were inventoried in 2010 and 2011. Within each plot, diameters at breast height (DBH) of all trees greater than 5 cm were measured [43]. Three nested subplots were used to measure trees and shrubs with DBH less than 5 cm and grass cover [43]. Tree AGB was calculated using the allometric equations provided by [44]. In remote sensing-based AGB modeling research, the representativeness of sample plots is critical for accurate AGB estimation. In order to make sure that all plots for analysis were representative of the forests, we overlaid all sample plots on the Landsat color composite to visually examine the geometric accuracy. A total of 664 sample plots were collected and used in this research. The statistics of the sample plot data based on different forest types are summarized in Table 2. The AGB values of all sample plots ranged from 25.7 Mg/ha to 180.7 Mg/ha with an average AGB of 95.9 Mg/ha. The sample plots were randomly divided into two groups: a dataset of 498 plots (75%) for model development and a dataset of 166 plots (25%) for validation of AGB estimates under non-stratification. Because we would examine the role of stratification of forest types in improving AGB modeling performance, we needed to have a sufficient number of validation samples for each forest type. Thus, 30% of the total samples for each forest type was adopted. The numbers of modeling and validation samples are also summarized in Table 2.



algorithms [2]. AGB may have linear, nonlinear, or no relationships with remote sensing variables. Because of the high correlations among some explanatory variables, it is critical to identify the variables that have a high correlation with AGB but weak correlations among themselves [2]. Thus, correlation analysis can be used to examine the relationships between AGB and remote sensing variables to remove the variables without significant correlation. Meanwhile, the correlation analysis was used to analyze the correlation coefficients between the remote sensing variables to remove the variables that have very high coefficients (e.g., >0.9) but relatively low standard deviations. The stepwise regression method was then used to identify the variables for AGB modeling because it introduces the variables into the model one by one and tests their significance. When the original explanatory variable is no longer significant due to the introduction of a new explanatory variable, it is deleted to ensure that only the significant variables are included in the regression equation before each new variable is added. This process is repeated until no significant variables can be added to and no insignificant variables can be removed from the model. Thus, the final variables are guaranteed to be optimal and not have serious multicollinearity. During the stepwise regression analysis, the F-statistical test was used to decide whether one variable was included or not, based on the F test level of 0.1 and significance test level of 0.05.

The LR approach assumes linear relationships between explanatory variables and a response variable, specifically remote sensing-derived predictors and AGB. However, in reality, this assumption is not always met, especially when multisource data include different kinds of remote sensing variables (e.g., spectral and spatial features, radar) and DEM-derived data are used. The nonparametric and machine learning algorithms such as RF, ANN, SVR, and kNN can handle the complex nonlinear relationships and were especially useful when multisource data were used [2,9,22,27,46]. However, except for RF, most nonparametric algorithms cannot provide a variable selection method and are not able to identify the optimal variables for AGB modeling. Thus, the variables selected by RF were also used in SVR, ANN, and kNN for AGB modeling.

The principle of using RF is that the  $k$  samples are extracted from the training set using the bootstrap sampling method and the  $k$  decision tree models are set up for the extracted  $k$  sample sets, and then each sample is used to obtain the predicted values of the  $k$  group. Finally, the predicted values are averaged to obtain the final prediction value for each sample [7,15,25]. The RF algorithm has three parameters: the number of decision trees (Ntree), the minimum number of observations per tree leaf (mtry), and the number of repetitions in the calculation of importance (nperm). The nperm is often assigned as a default value of 5. In this research, the mtry and ntree were modified by repeating the setting, and the root mean square errors (RMSEs) of test samples were compared. In the RF approach, a backward feature elimination method was often used to eliminate relatively less important variables from all variables and to keep the most important variables after many iterations. A variable was removed at each iteration, and the parameters of RF were optimized. When the RMSE reached the minimum, the most appropriate mtry and ntree were obtained. This process led to an importance rank of the independent variables. Based on the rank, the least number of the variables for producing the most accurate estimates of AGB was determined.

### 3.3. Biomass Modeling Algorithms

Different algorithms including LR, RF, ANN, SVR, and kNN were used for AGB modeling in this study. LR can be expressed as model (1):

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + \varepsilon \quad (1)$$

where  $a_0$  is a constant,  $x_1, x_2, \dots, x_n$  are the explanatory variables,  $a_1, a_2, \dots, a_n$  are the regression coefficients associated with the corresponding variables,  $y$  is the value of the plot's AGB,  $n$  is the number of explanatory variables, and  $\varepsilon$  is the error term.

The advantages of using discrete or continuous datasets—less influenced by noise, high efficiency in using large datasets, no need for a priori probability distribution, flexibility, and robustness—have made RF an important tool in land cover classification and AGB estimation in the past decade [47–52]. As described in the variable selection using the RF approach, the optimal RF model was used to map AGB for the entire study area.

SVR is based on the Vapnik–Chervonenkis (VC) dimension theory and structural risk minimization, seeking the best compromise between the complexity of the model (the learning accuracy of specific training samples) and learning ability (the ability to identify any sample without error) in order to achieve the best promotion ability based on a limited number of samples. In the high dimension feature space, the “curse of dimensionality” phenomenon is prone to occur; therefore, it is necessary to transform the calculation in the high dimension feature space, which requires the kernel function to replace the inner product [21–23]. The choice of the kernel function is a core problem in SVR research. At present, there is no way to construct a suitable kernel function for a specific problem. The commonly used kernel functions are the linear, polynomial, and radial basis function and the sigmoid kernel function. In SVR, one critical step is to optimize three parameters: the kernel function, SVR type, and penalty parameters [38]. In this research, we used a grid-search approach to determine the best penalty parameters and modified the kernel function and SVR type by repeating the setting. Similar to RF, the RMSEs of test samples were compared and the optimal parameters were determined when reaching the minimum.

ANN (back-propagation neural network here) is also used for AGB estimation [18]. ANN is divided into two phases: the learning stage and the prediction phase. The learning stage is the process of finding the rule between the input variable and the output variable. Through training to modify the weight matrix, the output value is kept close to the target value. In this process, the weights and thresholds of the network are deterministic. From the structural point of view, the network weights and thresholds are consistent with the coefficients and constants in the model and the learning process is consistent with the process of solving the coefficients and constants of the model. Two important parameters—the number of neurons and the transfer function—can greatly affect the accuracy of modeling prediction. In this research, numbers from 3 to 50 neurons and 13 transfer functions were examined. The optimal parameters were determined when the RMSE reached the minimum.

The kNN algorithm is a typical nonparametric algorithm, which is widely used in forestry survey and forest parameter estimation [14,30,32–35,46,53]. kNN is based on the similarity between the observation plots and predicted pixels using univariate or multiple variables. In terms of AGB estimation, the Mahalanobis distances between the estimated pixels and the sample plots were calculated using the factors that had significant effects on AGB:

$$D_{(x,y)} = \sqrt{(x - y)^T \Sigma^{-1} (x - y)} \quad (2)$$

where  $x = (x_1, x_2, \dots, x_p)$  and  $y = (y_1, y_2, \dots, y_p)$ ,  $x$  and  $y$  represent  $p$ -dimension variable vectors in the spectral space between two AGB sample plots. In addition,

$$\Sigma = \begin{bmatrix} cov(x_1, y_1) & \cdots & cov(x_1, y_p) \\ \vdots & \ddots & \vdots \\ cov(x_p, y_1) & \cdots & cov(x_p, y_p) \end{bmatrix} \text{ and } cov(x_i, y_i) = \frac{\sum_{i=1}^p (x_i - \bar{x})(y_i - \bar{y})}{p - 1}, \quad (3)$$

where  $i = 1, 2, \dots, p$ ;  $\bar{x}$  and  $\bar{y}$  represent mean values of the  $x$  and  $y$  variables.  $T$  represents matrix transpose.  $D_{(x,y)}$  is the Mahalanobis distances between two sample plots at  $p$ -dimension variable vectors  $x$  and  $y$ . The closer the sample plot is to the target pixel, the greater the weight of the sample plot. The estimate of AGB for each pixel was created by calculating a weighted mean of AGB values from  $k$  nearest plots based on their inverse distances. In this research, we examined different values of

$k$  from 1 to 30, depending on different scenarios. When the prediction accuracy reached the maximum, the optimal parameter was selected.

### 3.4. AGB Modeling Based on Different Scenarios

Various scenarios were designed based on three datasets, five modeling algorithms, and stratification and non-stratification of forest types: (1) AGB models were developed based on three data sources with five modeling algorithms without stratification of forest type, which led to a total of 15 scenarios being examined; (2) AGB models were developed for each of four forest types based on three datasets and five modeling algorithms, which resulted in a total of 60 scenarios being examined.

### 3.5. Evaluation of AGB Estimates

In AGB modeling research, the RMSE and relative RMSE (RMSEr) are often used to assess the prediction performance [2,9]:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \text{ and } \text{RMSEr} = \frac{\text{RMSE}}{\bar{y}} \times 100, \quad (4)$$

where  $\hat{y}_i$  and  $y_i$  are the predicted AGB and corresponding AGB at the sample plot  $i$ ;  $\bar{y}$  is the mean AGB of the test sample plots (total number of  $n$ ), depending on the non-stratification (all test samples) or stratification of forest types (the samples for each forest type, respectively). In general, the smaller RMSE or RMSEr values indicate better model estimation performance. The scatterplots showing the relationships between AGB estimates and reference data were also used to evaluate the model performance. Considering the number of sample plots for evaluation, 25% of the sample plots were randomly selected as test samples for the non-stratification scenarios. However, for the stratification of forest types, we needed to make sure that a sufficient number of sample plots within each forest type were available for the evaluation of each scenario. Therefore, 30% of the sample plots were randomly selected from the sample population corresponding to each forest type for evaluation of AGB estimates. The number of test samples under non-stratification and stratification of forest types is provided in Table 2.

## 4. Results

### 4.1. Comparative Analysis of AGB Modeling Results under Non-Stratification Scenarios

The established models based on non-stratification scenarios are summarized in Tables 3 and 4, indicating the important role of textural images. For Landsat TM imagery, the spectral band 7 (SWIR2) played the most important role. Overall, the models based on ALOS PALSAR data contained more variables than Landsat data alone but Landsat imagery had better performance than ALOS PALSAR data, using either LR or RF. The combination of Landsat TM and ALOS PALSAR data improved the modeling performance for LR but not for RF. The selected variables (Table 4) using RF were also used for the AGB model using ANN, SVR, and kNN. The statistics (e.g., maximum, minimum, mean, and standard deviation) of the AGB estimates from the AGB prediction maps (Table 5) clearly indicate that PALSAR produced higher mean values than Landsat or the combination of Landsat and PALSAR data using the modeling algorithms, except RF. However, the inverse was observed for the standard deviation values, implying that PALSAR data alone could not effectively predict the AGB distribution when the AGB value was very high or very small. Table 5 indicates that the predicted mean values using all these models were smaller than the mean from the sample plots, implying the overall underestimation using these models. The much larger minimum values and smaller maximum values from the RF and kNN models, compared to those from ANN, SVR, and LR, imply that the RF and kNN modeling algorithms were not able to properly predict AGB when the values of AGB were small or large.

**Table 3.** The regression models for AGB estimation based on the non-stratification scenario.

| Datasets    | Models                                                                                                                                                                                                              | R <sup>2</sup> |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| Landsat TM  | $Y_{TM} = 111.058 - 24.186T_{b7w3ME} + 4.549T_{b7w3CON} + 23.233T_{b4w5EN} + 9.641T_{b7w3COR} + 9.132T_{b3w3ME}$                                                                                                    | 0.36           |
| ALOS PALSAR | $Y_{SAR} = 44.761 + 13.553T_{HVw5ME} + 43.273T_{HVw7DI} - 44.889T_{HVw3VA} + 15.513T_{HHw3COR} - 35.278T_{HHw7COR} + 17.115T_{HVw3CON} - 4.315T_{HHw9CON} + 23.215T_{HVw7VA}$                                       | 0.19           |
| Combination | $Y_{comb} = 179.760 - 23.772T_{b7w3ME} + 12.772T_{b7w3VA} + 51.012T_{HVw7DI} - 33.133T_{HHw9CON} + 12.233T_{HHw3COR} - 31.541T_{b4w5HO} + 8.998T_{b2w5ME} - 5.177T_{HHw9CON} + 2.989T_{HHw5CON} - 16.234T_{HVw3VA}$ | 0.41           |

**Table 4.** The identified variables for AGB estimation modeling using the RF approach based on the non-stratification scenario.

| Datasets    | The Identified Variables from Different Remote Sensing Data                                                                                                 | R <sup>2</sup> |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| Landsat TM  | $S_{b7}$ , $T_{b7w3ME}$ , $T_{b3w5SM}$ , $T_{b4w5HO}$ , $T_{b7w5DI}$ , $T_{b3w7VA}$ , $T_{b7w7EN}$ , $T_{b5w9ME}$                                           | 0.87           |
| ALOS PALSAR | $T_{HHw3ME}$ , $T_{HHw7ME}$ , $T_{HHw5HO}$ , $T_{HHw9SM}$ , $T_{HHw9VA}$ , $HV$ , $T_{HVw7ME}$ , $T_{HVw5HO}$ , $T_{HVw9CON}$ , $T_{HVw9EN}$ , $T_{HVw9VA}$ | 0.62           |
| Combination | $S_{b7}$ , $HH$ , $T_{HHw5ME}$ , $HV$ , $T_{HVw7ME}$ , $T_{HVw7CON}$ , $T_{HVw9EN}$                                                                         | 0.60           |

Note:  $S_{bi}$ , spectral band  $i$ ;  $T_{biwjxx}$ , a texture image developed using the texture measure  $xx$  ( $xx$  can be such texture measures as ME (mean), VA (variance), HO (homogeneity), CON (contrast), DI (dissimilarity), EN (entropy), SM (second moment), COR (correlation)) on spectral band  $i$  ( $bi$ ) with a window size of  $j \times j$  pixel ( $wj$ ); HH (horizontally transmitted and received) and HV (horizontally transmitted and vertically received) represent two polarization options of the PALSAR image.

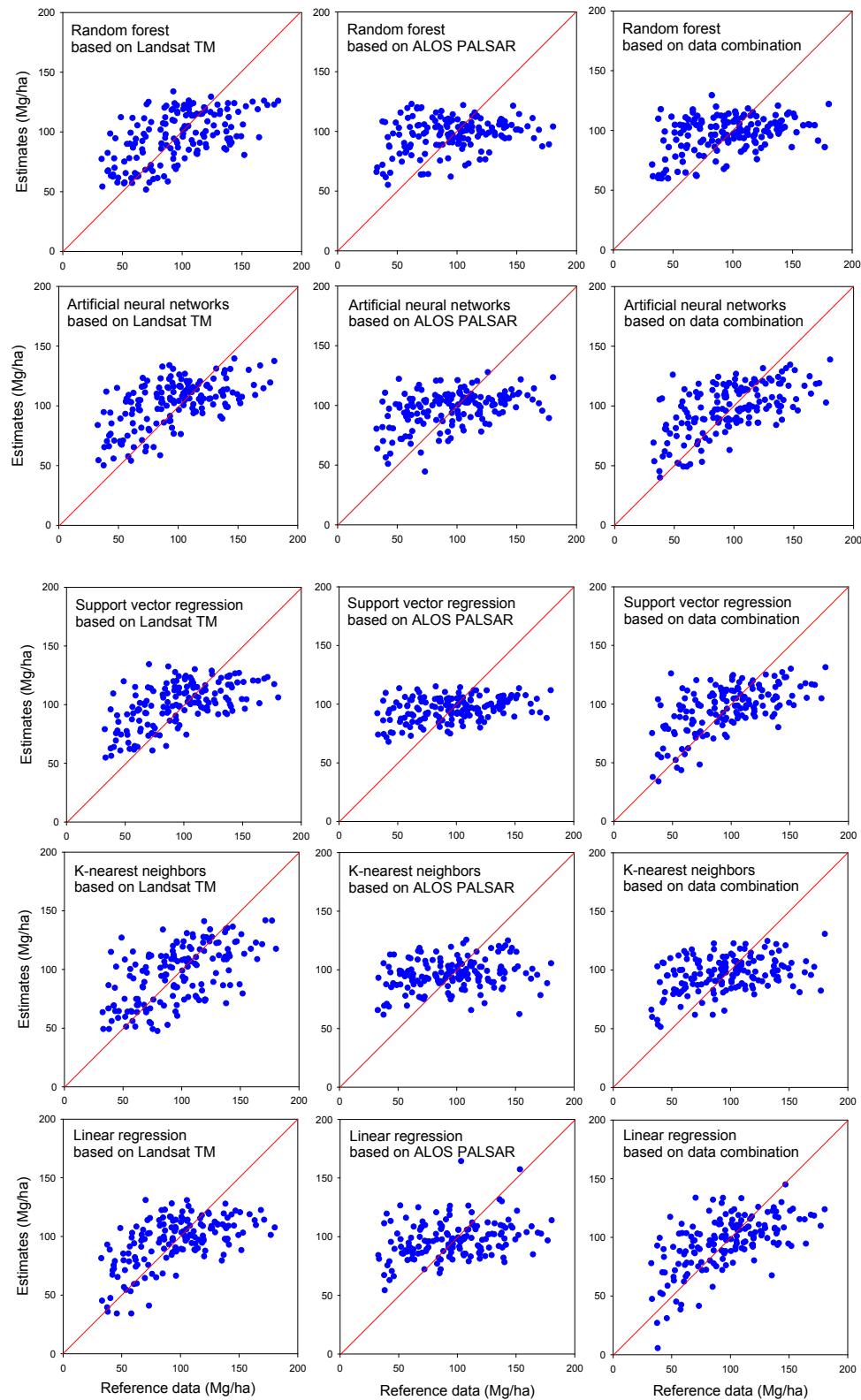
**Table 5.** Summary of statistical results of the predicted AGB images.

| Data         | Models     | Mean | Std Dev | Minimum | Maximum | Data Range |
|--------------|------------|------|---------|---------|---------|------------|
| Landsat TM   | RF         | 93.0 | 22.9    | 42.0    | 148.7   | 106.7      |
|              | ANN        | 88.4 | 26.6    | 19.1    | 187.1   | 168.0      |
|              | SVR        | 88.3 | 22.2    | 19.3    | 149.0   | 129.7      |
|              | kNN        | 88.8 | 26.9    | 32.9    | 164.7   | 131.8      |
|              | LR         | 90.6 | 26.2    | 1.7     | 180.0   | 178.3      |
| ALOS PALSAR  | RF         | 92.4 | 15.2    | 39.8    | 153.8   | 114.0      |
|              | ANN        | 91.8 | 17.0    | 1.6     | 150.5   | 148.9      |
|              | SVR        | 91.5 | 11.3    | 19.6    | 122.1   | 102.5      |
|              | kNN        | 93.7 | 14.7    | 55.8    | 127.1   | 71.3       |
|              | LR         | 92.1 | 18.4    | 1.5     | 135.5   | 134.0      |
| Combination  | RF         | 93.0 | 22.9    | 41.9    | 148.8   | 106.9      |
|              | ANN        | 90.2 | 26.3    | 0.7     | 181.3   | 180.6      |
|              | SVR        | 86.6 | 24.5    | 0.7     | 176.1   | 175.4      |
|              | kNN        | 92.0 | 16.6    | 43.2    | 137.7   | 94.5       |
|              | LR         | 89.8 | 29.7    | 1.6     | 177.8   | 176.2      |
| Sample plots | Statistics | 95.9 | 33.9    | 25.7    | 180.7   | 155.0      |

Note: RF, Random Forest; ANN, artificial neural networks; SVR, support vector regression; kNN, k-nearest neighbor; LR, linear regression.

The performance of the predictions can be explained with the scatterplots showing the relationships between the AGB estimates and reference data (Figure 4). It indicates that the overestimation and underestimation problems were obvious for all the prediction results, no matter which datasets and modeling algorithms were used. This situation especially became worse for the PALSAR-based predictions. For Landsat imagery, when AGB was within 50–130 Mg/ha, the residuals were relatively small, but for ALOS PALSAR data, the AGB ranges became narrower, only about 90–120 Mg/ha. The combination of Landsat TM and ALOS PALSAR did not improve the residuals (Figure 4). The RMSE and RMSEr results (Table 6) quantitatively confirmed this situation; that is, Landsat TM produced more accurate AGB estimates than ALOS PALSAR data no matter which modeling algorithms were used. The combination of Landsat and ALOS PALSAR data provided

slightly better performance for LR, similar performance for SVR and ANN, and worse performance for RF and kNN. Overall, Landsat TM imagery had smaller RMSE values of 27.7–29.3 Mg/ha compared with 30.3–33.7 Mg/ha for ALOS PALSAR.



**Figure 4.** The relationships between AGB estimates from different models against plot reference values.

**Table 6.** Summary of accuracy assessment results based on different scenarios.

| Algorithms | Datasets | Landsat TM |                   | ALOS PALSAR |                   | Combination |                   |
|------------|----------|------------|-------------------|-------------|-------------------|-------------|-------------------|
|            |          | RMSE       | RMSE <sub>r</sub> | RMSE        | RMSE <sub>r</sub> | RMSE        | RMSE <sub>r</sub> |
| RF         |          | 28.4       | 29.5              | 33.2        | 34.5              | 30.3        | 31.5              |
| ANN        |          | 27.7       | 28.8              | 30.3        | 31.5              | 27.6        | 28.7              |
| SVR        |          | 28.2       | 29.3              | 32.1        | 33.4              | 28.2        | 29.3              |
| kNN        |          | 28.3       | 29.4              | 33.7        | 35.0              | 30.8        | 32.0              |
| LR         |          | 29.3       | 30.5              | 32.9        | 34.2              | 27.7        | 28.8              |

Note: RF, Random Forest; ANN, artificial neural networks; SVR, support vector regression; kNN, k-nearest neighbor; LR, linear regression; RMSE, root mean squared error (Mg/ha); RMSE<sub>r</sub>, relative root mean squared error (%).

The above analysis was based on overall performance of different data sources and modeling algorithms but cannot provide detailed information on how different forest types and AGB ranges affected the AGB estimation under non-stratification scenarios. Table 7 summarizes the RMSE and RMSE<sub>r</sub> results for different scenarios. These results indicate that in terms of RMSE, mixed forests and the forests with an AGB range greater than 160 Mg/ha had the highest RMSE values for different datasets. However, because of the different AGB average values for the forest types and AGB ranges, bamboo forests and the forests with an AGB range less than 40 Mg/ha had the largest values of RMSE<sub>r</sub>. Although ANN provided the best overall estimation results for different datasets, LR provided the best estimation for broadleaf and bamboo forests and the forests with an AGB range of 40–120 Mg/ha when Landsat data were used. For the ALOS PALSAR data, SVR provided the best estimation when the AGB range was 40–160 Mg/ha. Table 7 indicates that no datasets and modeling algorithms can be optimal for all forest types or for different AGB ranges, implying the necessity to develop AGB estimation models according to specific forest types.

**Table 7.** Summary of RMSE (Mg/ha) and RMSE<sub>r</sub> (%) results from different scenarios under non-stratification conditions.

| Data        | Model | RMSE (Mg/ha) |             |      |      |      |                    |        |         |      |
|-------------|-------|--------------|-------------|------|------|------|--------------------|--------|---------|------|
|             |       | Overall      | Forest Type |      |      |      | AGB Ranges (Mg/ha) |        |         |      |
|             |       |              | MXF         | BLF  | CFF  | BBF  | <40                | 40–120 | 120–160 | >160 |
| Landsat TM  | RF    | 28.4         | 32.0        | 26.2 | 28.6 | 24.4 | 34.6               | 23.2   | 35.9    | 53.6 |
|             | ANN   | 27.7         | 30.0        | 26.2 | 28.6 | 25.0 | 37.4               | 25.5   | 29.1    | 50.2 |
|             | SVR   | 28.2         | 31.9        | 25.4 | 28.7 | 23.9 | 36.7               | 24.5   | 30.5    | 57.5 |
|             | kNN   | 28.3         | 32.6        | 26.5 | 28.5 | 22.0 | 29.7               | 24.7   | 34.7    | 61.0 |
|             | LR    | 29.3         | 32.6        | 24.4 | 28.7 | 21.3 | 35.0               | 22.4   | 34.2    | 61.0 |
| ALOS PALSAR | RF    | 33.2         | 36.8        | 29.5 | 33.8 | 28.5 | 42.5               | 27.4   | 36.9    | 74.6 |
|             | ANN   | 30.3         | 36.0        | 26.4 | 30.9 | 31.1 | 45.2               | 25.2   | 35.4    | 67.8 |
|             | SVR   | 32.1         | 35.4        | 28.2 | 31.9 | 34.7 | 50.6               | 24.7   | 32.2    | 73.6 |
|             | kNN   | 33.7         | 39.2        | 31.6 | 32.9 | 32.8 | 46.0               | 26.3   | 40.6    | 78.6 |
|             | LR    | 32.9         | 37.0        | 31.0 | 32.4 | 32.0 | 47.5               | 27.3   | 37.1    | 70.3 |
| Comb.       | RF    | 30.3         | 35.6        | 26.1 | 30.4 | 28.5 | 39.3               | 23.5   | 37.8    | 68.3 |
|             | ANN   | 27.6         | 31.7        | 24.0 | 28.2 | 23.4 | 35.5               | 23.7   | 31.6    | 53.6 |
|             | SVR   | 28.2         | 33.3        | 23.9 | 29.0 | 21.6 | 34.0               | 23.9   | 32.7    | 56.3 |
|             | kNN   | 30.8         | 35.2        | 28.7 | 30.6 | 28.1 | 35.9               | 25.3   | 36.1    | 68.3 |
|             | LR    | 27.7         | 34.4        | 24.9 | 28.2 | 20.9 | 35.7               | 22.9   | 32.4    | 58.6 |

Table 7. Cont.

| Data        | Model | RMSEr (%) |             |      |      |      |                    |        |         |      |
|-------------|-------|-----------|-------------|------|------|------|--------------------|--------|---------|------|
|             |       | Overall   | Forest Type |      |      |      | AGB Ranges (Mg/ha) |        |         |      |
|             |       |           | MXF         | BLF  | CFF  | BBF  | <40                | 40–120 | 120–160 | >160 |
| Landsat TM  | RF    | 29.5      | 30.4        | 29.3 | 27.4 | 47.7 | 95.7               | 27.8   | 26.3    | 31.3 |
|             | ANN   | 28.8      | 28.5        | 29.3 | 27.4 | 48.9 | 103.4              | 30.5   | 21.3    | 29.3 |
|             | SVR   | 29.3      | 30.3        | 28.4 | 27.5 | 46.8 | 101.5              | 29.3   | 22.3    | 33.6 |
|             | kNN   | 29.4      | 31.0        | 29.7 | 27.3 | 43.0 | 82.2               | 29.6   | 25.4    | 35.6 |
|             | LR    | 30.5      | 31.0        | 27.3 | 27.5 | 41.7 | 96.8               | 26.8   | 25.0    | 35.6 |
| ALOS PALSAR | RF    | 34.5      | 34.9        | 33.0 | 32.4 | 55.8 | 117.6              | 32.8   | 27.0    | 43.6 |
|             | ANN   | 31.5      | 34.2        | 29.6 | 29.6 | 60.8 | 125.0              | 30.2   | 25.9    | 39.6 |
|             | SVR   | 33.4      | 33.6        | 31.6 | 30.6 | 67.9 | 140.0              | 29.6   | 23.6    | 43.0 |
|             | kNN   | 35.0      | 37.2        | 35.4 | 31.5 | 64.2 | 127.2              | 31.5   | 29.7    | 45.9 |
|             | LR    | 34.2      | 35.1        | 34.7 | 31.1 | 62.6 | 131.4              | 32.7   | 27.2    | 41.1 |
| Comb.       | RF    | 31.5      | 33.8        | 29.2 | 29.1 | 55.8 | 108.7              | 28.1   | 27.7    | 39.9 |
|             | ANN   | 28.7      | 30.1        | 26.9 | 27.0 | 45.8 | 98.2               | 28.4   | 23.1    | 31.3 |
|             | SVR   | 29.3      | 31.6        | 26.8 | 27.8 | 42.3 | 94.0               | 28.6   | 23.9    | 32.9 |
|             | kNN   | 32.0      | 33.4        | 32.1 | 29.3 | 55.0 | 99.3               | 30.3   | 26.4    | 39.9 |
|             | LR    | 28.8      | 32.7        | 27.9 | 27.0 | 40.9 | 98.7               | 27.4   | 23.7    | 34.2 |

Note: RMSE, root mean squared error; RMSEr, relative root mean squared error; RF, Random Forest; ANN, artificial neural networks; SVR, support vector regression; kNN, k-nearest neighbor; LR, linear regression; MXF, mixed forest; BLF, broadleaf forest; CFF, coniferous forest; BBF, bamboo forest; AGB, aboveground biomass.

#### 4.2. Comparative Analysis of AGB Modeling Results Based on Stratification of Forest Types

The AGB models derived using five algorithms and three datasets for four forest types were compared (Tables 8 and 9). The texture measures mean, correlation, or both were involved in almost all AGB models, implying that the textures had significant contributions in improving the predictions of AGB. When Landsat TM images alone or their combination with PALSAR data were utilized, the TM spectral bands 5 and 7 and textures were frequently included in the AGB models, implying the significant roles of spectral variables in AGB estimation modeling. In the coniferous forest AGB models, the relevant Landsat TM spectral variables were often selected, meaning that the variables had a potential influence on the predictions of AGB mainly because in this study area the coniferous forests consisted mainly of pine and Chinese fir plantations that were characterized by simple forest canopy structures. This conclusion is similar to those in tropical forest regions of the Amazon where spectral responses were more important than textures for the forest sites with relatively simple forest stand structure [4,54]. On the other hand, the texture measures were involved in the models of other forests types, including broadleaf forests and mixed forests that had multiple canopy layers and complex canopy structures. This implies that the textures could account for the complicated forest structures [4]. If PALSAR data were used alone, the PALSAR-derived textures were added into almost all the models, indicating they had great potential to improve the predictions of AGB models. In the AGB models of broadleaf forests built using the PALSAR data and LR, only one texture was selected, which indicates that the PALSAR data were not appropriate for use to develop AGB modeling of the broadleaf forests.

The AGB predictions of the aforementioned models were assessed based on the values of RMSE and RMSEr according to forest types and AGB ranges (Table 10). Compared with those from the models without stratification of forest types (Table 7), most of the RMSE and RMSEr values from the models with stratification of forest types were considerably reduced (Table 10) for three datasets and five methods, and this was especially true for the combination of Landsat TM and PALSAR data. For example, the overall RMSE and RMSEr values decreased from 30.3 Mg/ha and 31.5% to 26.1 Mg/ha and 27.1%, respectively, for the RF method. For the AGB models of mixed forests, the values of RMSE decreased from a range of 31.7–35.6 Mg/ha to a range of 24.7–26.5 Mg/ha, implying a statistically significant reduction of RMSE due to the stratification of forest types. For the models of bamboo forests, all the methods except LR generated significantly smaller values of RMSE and RMSEr after the stratification of forest types than those without the stratification. The reduction of errors due

to the stratification was also noticed in the AGB models of coniferous forests but not in the models of broadleaf forests.

Comparison of the results between Tables 7 and 10 indicates that stratification of forest types was effective in reducing RMSE or RMSEr in different AGB ranges using these modeling algorithms. For both non-stratification and stratification situations, a similar conclusion is that AGB estimation had the lowest RMSE or RMSEr when AGB was in the range of 40–120 Mg/ha, had the highest RMSE values when AGB was greater than 160 Mg/ha, and had the highest RMSEr values when AGB was less than 40 Mg/ha. The stratification was especially valuable when AGB was greater than 160 Mg/ha. For example, for the Landsat TM image, RMSE values were reduced from 61 Mg/ha (see Table 7) to 53.4–53.9 mg/ha (see Table 10) when kNN or LR was used; and for ALOS PALSAR data, the RMSE values were reduced from 67.8–78.6 Mg/ha to 58.5–64.3 Mg/ha for different modeling algorithms. For the combined Landsat TM and ALOS PALSAR data, the stratification reduced RMSE from 58.6 Mg/ha (Table 7) to 41.5 Mg/ha (Table 10) using LR when AGB was greater than 160 Mg/ha. This research implies that stratification is especially valuable for forest sites having high AGB amounts.

**Table 8.** Linear regression models for AGB estimation based on different data sources under stratification of forest type.

| Data        | Linear Regression Models for Different Forest Types                                                                                                                                                                   |                                                                                                          |                                                                                                                                                                                                                 |                                                                                                                                                                                                |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|             | MXF                                                                                                                                                                                                                   | BLF                                                                                                      | CFF                                                                                                                                                                                                             | BBF                                                                                                                                                                                            |
| Landsat TM  | $Y_{TM} = 71.008 + 40.317T_{b3w9COR} - 10.625T_{b7w9ME} + 52.797T_{b2w3CON} - 206.413T_{b4w5HO} + 109.316T_{b5w5HO} + 260.741T_{b4w3SM} + 104.65T_{b3w9HO} - 6.329T_{b5w3ME}$                                         | $Y_{TM} = 177.747 - 0.165S_{b7} + 0.151S_{b3} - 28.648T_{b5w5COR} - 20.974T_{b2w9CON} + 1.282T_{b4w3VA}$ | $Y_{TM} = 183.858 - 0.021S_{b5} - 118.949T_{b5w5SM} - 9.12T_{b7w3ME}$                                                                                                                                           | $Y_{TM} = 96.657 + 23.498T_{b2w7COR} - 8.597T_{b7w9ME}$                                                                                                                                        |
| ALOS PALSAR | $Y_{SAR} = 174.082 - 11.082T_{HHw9ME}$                                                                                                                                                                                | $Y_{SAR} = 151.113 - 70.931T_{HHw7COR} - 144.26T_{HHw7SM} - 55.692T_{HVw9CON}$                           | $Y_{SAR} = 175.648 + 20.049T_{HHw3COR} - 41.359T_{HHw7COR} - 150.38T_{HVw7HO} + 13.14T_{HHw5ME} - 38.094T_{HHw9DI} + 19.933T_{HVw5COR}$                                                                         | $Y_{SAR} = 120.636 + 63.802T_{HVw7CON} - 38.269T_{HHw9EN} + 19.958T_{HHw5COR} - 39.861T_{HVw3VA} + 0.009HV - 164.441T_{HHw9SM}$                                                                |
| Comb.       | $Y_{comb} = 176.174 + 43.325T_{b3w9COR} - 11.337T_{b7w9ME} + 68.823T_{b2w3CON} - 205.811T_{b4w5HO} + 116.538T_{b5w5HO} + 315.581T_{b4w3SM} - 8.284T_{HHw3ME} - 52.339T_{b3w7VA} - 2.713T_{b4w3ME} + 36.675T_{HHw3HO}$ | $Y_{comb} = 245.956 - 0.155S_{b7} - 74.642T_{HHw7HO} - 50.144T_{HHw9COR} + 0.123S_{b3}$                  | $Y_{comb} = 179.245 - 0.044S_{b5} + 23.818T_{HHw3COR} - 52.009T_{b5w5HO} + 11.293T_{HVw5ME} - 42.93T_{HVw7COR} + 25.813T_{HVw5COR} + 18.357T_{b7w7COR} + 7.282T_{HHw3CON} - 31.22T_{HHw5EN} + 19.111T_{HHw3EN}$ | $Y_{comb} = 199.125 + 123.557T_{HVw7CON} - 41.769T_{HHw9EN} + 25.32T_{b2w7COR} + 27.018T_{HHw5COR} - 40.484T_{b4w3HO} - 125.836T_{HHw9HO} + 35.594T_{b7w3SM} + 0.006S_{b4} - 105.66T_{HVw7DI}$ |

**Table 9.** The identified variables using the RF approach based on different data sources under stratification of forest types.

| Data           | The Selected Variables for Different Forest Types Using the RF Approach                                                                                                              |                                                                                                                                                   |                                                                                                                                                                                                                                                  |                                                                                                                                                                      |
|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                | MXF                                                                                                                                                                                  | BLF                                                                                                                                               | CFF                                                                                                                                                                                                                                              | BBF                                                                                                                                                                  |
| Landsat TM     | T <sub>b3w7COR</sub> , T <sub>b3w9COR</sub> ,<br>T <sub>b4w9ME</sub> , T <sub>b5w3ME</sub> ,                                                                                         | T <sub>b4w9COR</sub> , T <sub>b5w5VA</sub> ,<br>T <sub>b5w5COR</sub> , T <sub>b5w9ME</sub> ,<br>T <sub>b7w3ME</sub>                               | S <sub>b3</sub> , S <sub>b5</sub> , T <sub>b3w9VA</sub> , T <sub>b4w7DI</sub> ,<br>T <sub>b5w5ME</sub> , T <sub>b5w5DI</sub> , T <sub>b5w7EN</sub> ,<br>T <sub>b5w7VA</sub> , T <sub>b7w3ME</sub> ,<br>T <sub>b5w7ME</sub> , T <sub>b7w9SM</sub> | T <sub>b2w7VA</sub> , T <sub>b3w7COR</sub> ,<br>T <sub>b3w9COR</sub> , T <sub>b4w3ME</sub> ,<br>T <sub>b4w9ME</sub> , T <sub>b5w3ME</sub> ,<br>T <sub>b5w9ME</sub>   |
| ALOS<br>PALSAR | T <sub>HHw9CON</sub> , T <sub>HHw7HO</sub> ,<br>T <sub>HVw3SM</sub> , T <sub>HVw5SM</sub> ,<br>T <sub>HVw7VA</sub> , T <sub>HVw7HO</sub> ,<br>T <sub>HVw9SM</sub>                    | HH, T <sub>HHw5ME</sub> ,<br>T <sub>HHw7CON</sub> , T <sub>HHw7COR</sub> ,<br>T <sub>HHw9SM</sub> , T <sub>HVw7CON</sub> ,<br>T <sub>HVw9EN</sub> | HH, T <sub>HHw3COR</sub> , T <sub>HHw3VA</sub> ,<br>T <sub>HHw5DI</sub> , T <sub>HHw7ME</sub> , T <sub>HHw9CON</sub> ,<br>T <sub>HHw9VA</sub> , T <sub>HHw9HO</sub> , T <sub>HHw9SM</sub> ,<br>T <sub>HVw3SM</sub> , T <sub>HVw5ME</sub>         | HV, T <sub>HHw3VA</sub> ,<br>T <sub>HHw5ME</sub> , T <sub>HHw5EN</sub> ,<br>T <sub>HHw9EN</sub> , T <sub>HVw7HO</sub> ,<br>T <sub>HVw9VA</sub> , T <sub>HVw9SM</sub> |
| Comb.          | S <sub>b5</sub> , T <sub>b2w7VA</sub> , T <sub>b3w7COR</sub> ,<br>T <sub>b3w9COR</sub> , T <sub>b4w3ME</sub> ,<br>T <sub>b4w9ME</sub> , T <sub>b5w5ME</sub> ,<br>T <sub>b6w7ME</sub> | S <sub>b5</sub> , T <sub>b5w9ME</sub> , T <sub>b5w3ME</sub> ,<br>T <sub>HHw7COR</sub> , T <sub>HVw7HO</sub>                                       | S <sub>b3</sub> , S <sub>b6</sub> , T <sub>b5w3ME</sub> , T <sub>b4w7DI</sub> ,<br>T <sub>b6w5ME</sub> , T <sub>b6w5HO</sub> , T <sub>b6w5COR</sub> ,<br>T <sub>HHw3COR</sub> , T <sub>HHw3VA</sub> , T <sub>HHw7ME</sub>                        | T <sub>b4w3ME</sub> , T <sub>b4w7VA</sub> ,<br>T <sub>b5w5ME</sub> , HV,<br>T <sub>HHw3ME</sub> , T <sub>HVw7HO</sub>                                                |

Note: MXF, mixed forest; BLF, broadleaf forest; CFF, coniferous forest; BBF, bamboo forest;  $S_{bi}$ , spectral band  $i$ ;  $T_{biwjxx}$ , a texture image that was developed using the texture measure  $xx$  ( $xx$  can be such texture measure as ME (mean), VA (variance), HO (homogeneity), CON (contrast), DI (dissimilarity), EN (entropy), SM (second moment), COR (correlation)) on spectral band  $i$  ( $bi$ ) with a window size of  $j \times j$  pixel ( $wj$ ); HH (horizontally transmitted and received) and HV (horizontally transmitted and vertically received) represent two polarization options of the PALSAR image.

**Table 10.** Summary of RMSE (Mg/ha) and RMSEr (%) results from different scenarios under stratification of forest types.

| Data        | Model | RMSE (Mg/ha) |             |      |      |      |                   |        |         |      |
|-------------|-------|--------------|-------------|------|------|------|-------------------|--------|---------|------|
|             |       | Overall      | Forest Type |      |      |      | AGB Range (Mg/ha) |        |         |      |
|             |       |              | MXF         | BLF  | CFF  | BBF  | <40               | 40–120 | 120–160 | >160 |
| Landsat TM  | RF    | 26.8         | 28.8        | 24.5 | 27.3 | 20.4 | 34.5              | 23.0   | 28.8    | 50.4 |
|             | ANN   | 25.5         | 28.8        | 24.2 | 26.4 | 19.9 | 31.6              | 21.6   | 29.1    | 50.4 |
|             | SVR   | 25.8         | 28.4        | 25.1 | 26.9 | 20.8 | 35.4              | 24.0   | 28.7    | 51.5 |
|             | kNN   | 28.0         | 28.9        | 26.5 | 29.1 | 21.9 | 28.1              | 24.5   | 33.7    | 53.4 |
|             | LR    | 27.4         | 28.7        | 25.4 | 26.7 | 21.4 | 35.4              | 22.6   | 33.6    | 53.9 |
| ALOS PALSAR | RF    | 30.2         | 30.7        | 29.5 | 30.8 | 24.6 | 39.3              | 25.1   | 36.5    | 63.8 |
|             | ANN   | 28.0         | 28.1        | 29.3 | 27.6 | 21.0 | 41.1              | 23.6   | 34.1    | 60.0 |
|             | SVR   | 29.1         | 29.8        | 30.3 | 28.8 | 22.5 | 46.9              | 23.7   | 31.8    | 60.1 |
|             | kNN   | 29.9         | 31.9        | 30.6 | 29.0 | 23.7 | 47.1              | 24.6   | 37.5    | 64.3 |
|             | LR    | 29.8         | 31.7        | 31.1 | 29.5 | 22.1 | 48.5              | 25.4   | 36.6    | 58.5 |
| Comb.       | RF    | 26.1         | 27.1        | 24.8 | 28.5 | 23.4 | 35.3              | 23.0   | 33.3    | 62.8 |
|             | ANN   | 24.7         | 26.8        | 23.5 | 25.9 | 19.1 | 30.4              | 21.6   | 25.1    | 52.0 |
|             | SVR   | 25.7         | 27.8        | 24.4 | 27.1 | 20.3 | 32.1              | 23.2   | 31.9    | 49.6 |
|             | kNN   | 26.5         | 27.3        | 25.2 | 27.7 | 20.7 | 33.9              | 34.1   | 33.5    | 59.4 |
|             | LR    | 26.4         | 27.9        | 25.4 | 26.1 | 20.6 | 34.3              | 22.1   | 31.0    | 41.5 |
| Data        | Model | RMSEr (%)    |             |      |      |      |                   |        |         |      |
|             |       | Overall      | Forest Type |      |      |      | AGB Range (Mg/ha) |        |         |      |
|             |       |              | MXF         | BLF  | CFF  | BBF  | <40               | 40–120 | 120–160 | >160 |
| Landsat TM  | RF    | 27.9         | 27.4        | 27.4 | 26.2 | 39.9 | 97.1              | 27.7   | 21.2    | 29.9 |
|             | ANN   | 26.5         | 27.4        | 27.1 | 25.3 | 38.9 | 88.9              | 26.0   | 21.3    | 29.9 |
|             | SVR   | 26.8         | 27.0        | 28.1 | 25.8 | 40.7 | 99.6              | 28.9   | 21.1    | 30.6 |
|             | kNN   | 29.1         | 27.4        | 29.7 | 27.9 | 42.8 | 79.1              | 29.5   | 24.8    | 31.8 |
|             | LR    | 28.4         | 27.3        | 28.4 | 25.6 | 41.9 | 99.6              | 27.2   | 24.7    | 32.1 |
| ALOS PALSAR | RF    | 31.4         | 29.2        | 33.0 | 29.5 | 48.1 | 110.5             | 30.2   | 26.8    | 37.9 |
|             | ANN   | 29.1         | 26.7        | 32.8 | 26.5 | 41.1 | 115.6             | 28.4   | 25.1    | 35.7 |
|             | SVR   | 30.2         | 28.3        | 33.9 | 27.6 | 44.0 | 131.9             | 28.5   | 23.4    | 35.7 |
|             | kNN   | 31.1         | 30.3        | 34.3 | 27.8 | 46.4 | 132.5             | 29.6   | 27.6    | 38.2 |
|             | LR    | 31.0         | 30.1        | 34.8 | 28.3 | 43.2 | 136.4             | 30.6   | 26.9    | 34.8 |
| Comb.       | RF    | 27.1         | 25.7        | 27.8 | 27.3 | 45.8 | 99.3              | 27.7   | 24.5    | 37.3 |
|             | ANN   | 25.7         | 25.5        | 26.3 | 24.8 | 37.4 | 85.5              | 26.0   | 18.4    | 30.9 |
|             | SVR   | 26.7         | 26.4        | 27.3 | 26.0 | 39.7 | 90.3              | 27.9   | 23.4    | 29.4 |
|             | kNN   | 27.5         | 25.9        | 28.2 | 26.5 | 40.5 | 95.4              | 29.0   | 24.7    | 35.3 |
|             | LR    | 27.4         | 26.5        | 28.4 | 25.0 | 40.3 | 96.5              | 26.6   | 22.8    | 24.7 |

Note: RMSE, root mean squared error; RMSEr, relative root mean squared error; RF, Random Forest; ANN, artificial neural networks; SVR, support vector regression; kNN, k-nearest neighbor; LR, linear regression; MXF, mixed forest; BLF, broadleaf forest; CFF, coniferous forest; BBF, bamboo forest.

## 5. Discussion

### 5.1. Selection of Suitable Variables for AGB Modeling

Many remote sensing variables, such as spectral bands, vegetation indices, textures, and image transformations (e.g., principal component analysis, tasseled cap, minimum noise fraction) [2] can be used as potential predictors for AGB modeling. The variables can be extracted at pixel, subpixel, and object levels. The pixel-level variables are commonly used. However, only a limited number of the remote sensing variables are useful because of their high correlations. For example, near infrared, normalized difference vegetation index and greenness from tasseled cap transform have similar vegetation information. The inclusion of such variables cannot improve AGB modeling performance [45]. In reality, the roles of spectral bands and textures in AGB modeling are dependent on the complexity of forest structures [2,4]. For example, this research indicated that the SWIR bands (e.g., Landsat TM spectral bands 5 and 7) play a more important role than visible and near-infrared bands. This conclusion is similar to that from previous studies on the moist tropical forests in the Amazon [4,45] and the Mediterranean forests [55]. The more important role of SWIR than near-infrared and visible spectral bands in AGB modeling may be due to the fact that SWIR is more sensitive to moisture and shade components inherent in the forest stand structure and that atmospheric conditions have less impact on spectral signatures than other shorter wavelength (e.g., near-infrared and visible) spectral bands. In particular, for the complex forest types in tropical and subtropical regions, SWIR is more valuable in AGB modeling than shorter wavelength spectral bands due to the wide spectral variation in the near-infrared band and being less sensitive to forest spectral signatures in the visible bands [3,4]. The higher data saturation values of AGB in SWIR than visible and near-infrared spectral bands further confirm the more important role of SWIR in AGB modeling (3). This research also implied that textures are another group of important variables for AGB modeling; in particular, textures may play more important roles than spectral bands in forest sites with complex forest stand structures. This was also confirmed in the Brazilian tropical forests [54] and subtropical forests [3,6]. However, it is critical to identify an optimal combination of textural images [5]. Because of the different features between textural images and pixel-level spectral responses, the combinations of these variables have proven helpful in improving predictions of AGB [3,4], and this research confirmed this conclusion no matter which modeling algorithm (LR or machine learning) was used. This research also indicates that AGB modeling requires multiple variables from remote sensing data, no matter which sensor data and which modeling algorithms were used. This implies that a single variable alone cannot effectively capture the complexity of forest stand structure and, thus, cannot provide satisfactory AGB estimation performance.

This research indicated that ALOS PALSAR data had poorer performance in AGB estimation than Landsat TM data no matter which modeling algorithm was used and the combination of TM and PALSAR could not improve or had limited effects in AGB estimation. A similar conclusion was also obtained in tropical forests [56–58]; for example, Hame et al. found that ALOS AVNIR data had better performance in AGB estimation than ALOS PALSAR data, and a combination of both data could not improve AGB estimation in the tropical forests in Laos [56]. This conclusion seems against our initial hypothesis that synthetic aperture radar (SAR) L-band data can provide the vertical structure of the forest and thus should have better performance than the optical sensor data. This is because SAR data represent the roughness of forest canopy and AGB is not directly related to the forest surface roughness, resulting in poor AGB modeling performance [6]. The poorer performance in SAR data than in optical sensor data was also confirmed in forest classification in the moist tropical regions in the Amazon [57,58]. Because Landsat TM and ALOS PALSAR data characterize forest structure differently, we expected that the combination of both data might improve AGB modeling performance. However, this research indicated that the combination, as either extra bands or data fusion, has limited effects in improving AGB modeling. The possible reasons may be that (1) the relatively coarse spatial resolution in ALOS-1PALSAR data (25-m spatial resolution in this research) cannot effectively capture

the forest stand structures of different forest types and (2) the data saturation in both Landsat and PALSAR data impedes the improvement of estimation [3,6]. Compared to the ALOS-1 PALSAR data used in this research, the current ALOS-2 PALSAR data with higher spatial resolution may provide rich forest stand structure information and may improve AGB modeling performance.

### 5.2. Selection of Modeling Algorithms

LR is commonly used for AGB modeling [2] and this research indicated its good performance, especially when AGB falls within 40–120 Mg/ha. This research also showed that not all machine learning and nonparametric algorithms led to more accurate AGB estimates than LR, although they have some advantages in data selection. The time for optimizing the parameters used in the relevant models was much longer than that required by LR. In addition, the requirement for sample plots (e.g., number of sample plots and the data range) was much stricter than LR; this is especially true for RF and kNN, which limited their extensions to the predictions of very low or very high AGB values, as shown in Table 5. However, when AGB values are very large or very small, machine learning and nonparametric algorithms can indeed improve AGB modeling performance. For example, this research indicated that when AGB values were greater than 120 Mg/ha, especially greater than 160 Mg/ha, ANN based on Landsat TM imagery improved AGB estimation by 3.7–6.3% of RMSEr compared with LR. When AGB values were less than 40 Mg/ha, RMSEr was reduced from 96.8% using LR to 82.2% using kNN. This implies that proper selection of modeling algorithms is also valuable in improving AGB modeling, especially for the AGB values that are very low or very high.

### 5.3. Potential Solutions to Improve AGB Estimation

This research indicated that no single algorithm was optimal for all forest types, and that bamboo forests were an important contribution resulting in poor estimation accuracy. Table 7 implies that different remote sensing data and modeling algorithms should be used for different forest types to improve AGB estimation. That is, it is valuable to develop specific AGB models for various forest types. However, this situation produces a big challenge in AGB modeling because of the requirement in number of sample plots and the transfer of AGB models in a large area. Table 9 confirms the value in improving AGB modeling using the stratification of forest types. This conclusion is similar to our previous finding that stratification of forest types [3] was an effective approach to improve AGB estimation. This research also indicated that AGB range was another factor influencing the AGB estimation. Stratification based on AGB range was especially valuable for improving the predictions of AGB with very high or very low values [9]. However, the challenge of using stratification is the requirement of a sufficient number of sample plots, which is often difficult considering the cost and intensive labor, especially when the AGB estimation is based on historical data [2].

Previous research and this case study indicate that small or large AGB values are the major factors influencing AGB modeling performance. For the forest sites with small AGB values, the complex composition of land cover type, such as bare soils, grass, and/or shrubs, is the major reason. Spectral unmixing may be one solution [54]. For the forest sites with large values of AGB, the data saturation in optical sensors such as Landsat imagery and in radar such as ALOS PALSAR [6] is the major reason for considerable underestimation of AGB. To date, there are no effective approaches to solve the data saturation problem, except the improvement of radiometric resolution and the use of stereo images and lidar data [2]. In recent years, airborne lidar and space-borne stereo images have become easily available, as well as other high spatial resolution images such as Worldview, Pleiades, and Quickbird. Therefore, more research should be directed toward integrating multiple remote sensing data for improving AGB estimation.

The representativeness of sample plots is fundamental for AGB modeling. The uncertainty of sample plot data could be from the sampling approach, plot size, the allometric equations for AGB calculation, and the measurements of tree attributes during fieldwork. Removal of the sample plots that do not have good representativeness is needed. Compared to previous work by Zhao et al. [3,6],

this research reduced the RMSEr for the forest sites with less than 40 Mg/ha, from 137 to 96.8%, by the refinement of sample plots based on LR. The overall RMSEr was also reduced from 32% in Zhao et al. [6] to 30.5% in this research. If a sufficient number of sample plots are collected for the forest sites with high AGB or low AGB values, development of AGB estimation models based on AGB ranges may considerably improve AGB estimation accuracy.

## 6. Conclusions

This research selected a subtropical forest region in Zhejiang Province, China, as a case study to explore the AGB estimation through a comparative analysis of different modeling algorithms (i.e., RF, ANN, SVR, kNN, and LR) based on Landsat TM, ALOS PALSAR, and their combination under stratification and non-stratification of forest types. The results indicate the following: (1) Landsat TM imagery provided more accurate estimates of AGB than ALOS PALSAR, and the combination of TM and PALSAR had limited effects on improving AGB estimation. (2) Overestimation for small AGB values and underestimation for large AGB values were major problems when using the optical (e.g., Landsat) and radar (e.g., ALOS PALSAR) data. (3) LR was still an important tool for AGB modeling, especially for the AGB range of 40–120 Mg/ha; machine learning and nonparametric algorithms had limited effects on improving AGB estimation; however, ANN was relatively the best model in this study. (4) Overall, RF and kNN were not suitable for AGB prediction for the forest sites with relatively low (e.g., less than 40 Mg/ha) or high (e.g., greater than 160 Mg/ha) AGB values. (5) Forest types and AGB ranges were important factors influencing AGB modeling performance; stratification based on both forest types and AGB ranges might provide great potential for improving AGB estimation if a sufficient number of sample plots were available. (6) More research should focus on improving AGB estimation for forest sites with relatively low (e.g., less than 40 Mg/ha) and high (e.g., greater than 120 Mg/ha) AGB values. However, purely Landsat or ALOS PALSAR data cannot solve this problem. As multiple source data such as lidar, stereo images, SAR, optical sensors, and ancillary data (e.g., DEM, climate, soil) become more easily available, more research should focus on the effective integration of different source data for developing AGB estimation models under the stratification based on both forest type and AGB range.

**Acknowledgments:** This study was financially supported by the National Natural Science Foundation of China (grant #41571411) and the Zhejiang Agriculture and Forestry University's Research and Development Fund (grant #2013FR052). The authors would like to thank Panpan Zhao for providing support in data collection, organization, and land cover classification.

**Author Contributions:** Dengsheng Lu conceived and designed the experiments; Yukun Gao performed the experiments; Dengsheng Lu, Guiying Li, Guangxing Wang, and Qi Chen analyzed the data; Yukun Gao, Lijuan Liu, and Dengqiu Li contributed to data collection, organization, and biomass estimation modeling; Dengsheng Lu wrote the manuscript and each co-author participated in the editing.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Zhu, J.; Huang, Z.; Sun, H.; Wang, G. Mapping forest ecosystem biomass density for Xiangjiang River Basin by combining plot and remote sensing data and comparing spatial extrapolation methods. *Remote Sens.* **2017**, *9*, 241. [[CrossRef](#)]
2. Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Li, G.; Moran, E. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *Int. J. Digit. Earth* **2016**, *9*, 63–105. [[CrossRef](#)]
3. 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](#)]
4. Lu, D. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. *Int. J. Remote Sens.* **2005**, *26*, 2509–2525. [[CrossRef](#)]

5. Lu, D.; Batistella, M. Exploring TM image texture and its relationships with biomass estimation in Rondônia, Brazilian Amazon. *Acta Amaz.* **2005**, *35*, 249–257. [[CrossRef](#)]
6. Zhao, P.; Lu, D.; Wang, G.; Liu, L.; Li, D.; Zhu, J.; Yu, S. 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](#)]
7. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
8. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning: Methods and Applications*; Springer: New York, NY, USA, 2012; pp. 157–175, ISBN 9781441993267.
9. Feng, Y.; Lu, D.; Chen, Q.; Keller, M.; Moran, E.; dos-Santos, M.N.; Bolfe, E.L.; Batistella, M. Examining effective use of data sources and modeling algorithms for improving biomass estimation in a moist tropical forest of the Brazilian Amazon. *Int. J. Digit. Earth* **2017**, *10*, 996–1016. [[CrossRef](#)]
10. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. [[CrossRef](#)]
11. Fleming, A.L.; Wang, G.; McRoberts, R.E. Comparison of methods toward multi-scale forest carbon mapping and spatial uncertainty analysis: Combining national forest inventory plot data and landsat TM images. *Eur. J. For. Res.* **2015**, *134*, 125–137. [[CrossRef](#)]
12. Chen, Q. Modeling aboveground tree woody biomass using national-scale allometric methods and airborne lidar. *ISPRS J. Photogramm. Remote Sens.* **2015**, *106*, 95–106. [[CrossRef](#)]
13. Foody, G.M.; Cutler, M.E.; McMorrow, J.; Pelz, D.; Tangki, H.; Boyd, D.S.; Douglas, I. Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Glob. Ecol. Biogeogr.* **2001**, *10*, 379–387. [[CrossRef](#)]
14. Reese, H.; Nilsson, M.; Sandström, P.; Olsson, H. Applications using estimates of forest parameters derived from satellite and forest inventory data. *Comput. Electron. Agric.* **2002**, *37*, 37–55. [[CrossRef](#)]
15. Breidenbach, J.; Næsset, E.; Gobakken, T. Improving k-nearest neighbor predictions in forest inventories by combining high and low density airborne laser scanning data. *Remote Sens. Environ.* **2012**, *117*, 358–365. [[CrossRef](#)]
16. McRoberts, R.E.; Chen, Q.; Domke, G.M.; Næsset, E.; Gobakken, T.; Chirici, G.; Mura, M. Optimizing nearest neighbour configurations for airborne laser scanning-assisted estimation of forest volume and biomass. *Forestry* **2017**, *90*, 99–111. [[CrossRef](#)]
17. Saatchi, S.; Malhi, Y.; Zutta, B.; Buermann, W.; Anderson, L.O.; Araujo, A.M.; Phillips, O.L.; Peacock, J.; ter Steege, H.; Lopez Gonzalez, G.; et al. Mapping landscape scale variations of forest structure, biomass, and productivity in Amazonia. *Biogeosci. Discuss.* **2009**, *6*, 5461–5505. [[CrossRef](#)]
18. Lek, S.; Guégan, J.F. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol. Model.* **1999**, *120*, 65–73. [[CrossRef](#)]
19. Wang, S.; Guan, D. Remote sensing method of forest biomass estimation by artificial neural network models. *Ecol. Environ.* **2007**, *16*, 108–111. [[CrossRef](#)]
20. Hame, T.; Salli, A.; Andersson, K.; Lohi, A. A new methodology for the estimation of biomass of coniferdominated boreal forest using NOAA AVHRR data. *Int. J. Remote Sens.* **1997**, *18*, 3211–3243. [[CrossRef](#)]
21. Gleason, C.J.; Im, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91. [[CrossRef](#)]
22. Li, M.; Im, J.; Quackenbush, L.J.; Liu, T. Forest biomass and carbon stock quantification using airborne LiDAR data: A case study over Huntington Wildlife Forest in the Adirondack Park. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3143–3156. [[CrossRef](#)]
23. Kim, Y.H.; Im, J.; Ha, H.K.; Choi, J.K.; Ha, S. Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *Gisci. Remote Sens.* **2014**, *51*, 158–174. [[CrossRef](#)]
24. Zhang, J.; Huang, S.; Hogg, E.H.; Liefers, V.; Qin, Y.; He, F. Estimating spatial variation in Alberta forest biomass from a combination of forest inventory and remote sensing data. *Biogeosciences* **2014**, *11*, 2793–2808. [[CrossRef](#)]
25. Belgiu, M.; Drăgu, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
26. Houghton, R.A.; Butman, D.; Bunn, A.G.; Krankina, O.N.; Schlesinger, P.; Stone, T.A. Mapping Russian forest biomass with data from satellites and forest inventories. *Environ. Res. Lett.* **2007**, *2*, 45032–45037. [[CrossRef](#)]

27. Ahmed, O.S.; Franklin, S.E.; Wulder, M.A.; White, J.C. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS J. Photogramm. Remote Sens.* **2015**, *101*, 89–101. [\[CrossRef\]](#)
28. Baccini, A.; Friedl, M.A.; Woodcock, C.E.; Warbington, R. Forest biomass estimation over regional scales using multisource data. *Geophys. Res. Lett.* **2004**, *31*, 399–420. [\[CrossRef\]](#)
29. Chirici, G.; Barbati, A.; Corona, P.; Marchetti, M.; Travaglini, D.; Maselli, F.; Bertini, R. Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. *Remote Sens. Environ.* **2008**, *112*, 2686–2700. [\[CrossRef\]](#)
30. Holmström, H.; Fransson, J.E.S. Combining remotely sensed optical and radar data in kNN-estimation of forest variables. *For. Sci.* **2003**, *49*, 409–418. [\[CrossRef\]](#)
31. Katila, M.; Tomppo, E. Stratification by ancillary data in multisource forest inventories employing k-nearest-neighbour estimation. *Can. J. For. Res.* **2002**, *32*, 1548–1561. [\[CrossRef\]](#)
32. McRoberts, R.E.; Magnussen, S.; Tomppo, E.O.; Chirici, G. Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data. *Remote Sens. Environ.* **2011**, *115*, 3165–3174. [\[CrossRef\]](#)
33. McRoberts, R.E.; Næsset, E.; Gobakken, T. Optimizing the k-Nearest neighbors technique for estimating forest aboveground biomass using airborne laser scanning data. *Remote Sens. Environ.* **2015**, *163*, 13–22. [\[CrossRef\]](#)
34. Tomppo, E.; Halme, M. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: A genetic algorithm approach. *Remote Sens. Environ.* **2004**, *92*, 1–20. [\[CrossRef\]](#)
35. Tomppo, E.; Olsson, H.; Ståhl, G.; Nilsson, M.; Hagner, O.; Katila, M. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sens. Environ.* **2008**, *112*, 1982–1999. [\[CrossRef\]](#)
36. Ahern, F.J.; Erdle, T.; Maclean, D.A.; Knepeck, I.D. A quantitative relationship between forest growth rates and thematic mapper reflectance measurements. *Int. J. Remote Sens.* **1991**, *12*, 387–400. [\[CrossRef\]](#)
37. McRoberts, R.E. Estimating forest attribute parameters for small areas using nearest neighbors techniques. *For. Ecol. Manag.* **2012**, *272*, 3–12. [\[CrossRef\]](#)
38. Cherkassky, V.; Ma, Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw.* **2004**, *17*, 113–126. [\[CrossRef\]](#)
39. Chander, G.; Markham, B.L.; Helder, D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* **2009**, *113*, 893–903. [\[CrossRef\]](#)
40. Chávez, P.S.J. Image-based atmospheric corrections—Revisited and improved. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 1025–1036.
41. Reese, H.; Olsson, H. C-correction of optical satellite data over alpine vegetation areas: A comparison of sampling strategies for determining the empirical c-parameter. *Remote Sens. Environ.* **2011**, *115*, 1387–1400. [\[CrossRef\]](#)
42. Lee, J.S.; Jurkevich, I.; Dewaele, P.; Wambacq, P.; Oosterlinck, A. Speckle filtering of synthetic aperture radar images: A review. *Remote Sens. Rev.* **1994**, *8*, 313–340. [\[CrossRef\]](#)
43. Qian, Y.; Yi, L.; Zhang, C.; Yu, S.; Shen, L.; Peng, D.; Zheng, C. Biomass and Carbon Storage of Public Service Forests in the Central Area of Zhejiang Province. *Sci. Silvae Sin.* **2013**, *49*, 17–23. [\[CrossRef\]](#)
44. Yuan, W.; Jiang, B.; Ge, Y.; Zhu, J.; Shen, A. Study on biomass models of important commonwealth forests for Zhejiang Province. *J. Zhejiang For. Sci. Technol.* **2009**, *29*, 1–5.
45. 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. [\[CrossRef\]](#)
46. Koukal, T.; Suppan, F.; Schneider, W. The impact of relative radiometric calibration on the accuracy of kNN-predictions of forest attributes. *Remote Sens. Environ.* **2007**, *110*, 431–437. [\[CrossRef\]](#)
47. Avitabile, V.; Baccini, A.; Friedl, M.A.; Schmullius, C. Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sens. Environ.* **2012**, *117*, 366–380. [\[CrossRef\]](#)
48. Baccini, A.; Laporte, N.T.; Goetz, S.J.; Sun, M.; Dong, H. A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environ. Res. Lett.* **2008**, *3*, 45011. [\[CrossRef\]](#)

49. Tanase, M.A.; Panciera, R.; Lowell, K.; Tian, S.; Hacker, J.M.; Walker, J.P. Airborne multi-temporal L-band polarimetric SAR data for biomass estimation in semi-arid forests. *Remote Sens. Environ.* **2014**, *145*, 93–104. [[CrossRef](#)]
50. Vincenzi, S.; Zucchetto, M.; Franzoi, P.; Pellizzato, M.; Pranovi, F.; De Leo, G.A.; Torricelli, P. Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecol. Model.* **2011**, *222*, 1471–1478. [[CrossRef](#)]
51. Pflugmacher, D.; Cohen, W.B.; Kennedy, R.E.; Yang, Z. Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics. *Remote Sens. Environ.* **2014**, *151*, 124–137. [[CrossRef](#)]
52. Ismail, R.; Mutanga, O.; Kumar, L. Modeling the Potential Distribution of Pine Forests Susceptible to *Sirex Noctilio* Infestations in Mpumalanga, South Africa. *Trans. GIS* **2010**, *14*, 709–726. [[CrossRef](#)]
53. McInerney, D.O.; Suarez-Minguez, J.; Valbuena, R.; Nieuwenhuis, M. Forest canopy height retrieval using LiDAR data, medium-resolution satellite imagery and kNN estimation in Aberfoyle, Scotland. *Forestry* **2010**, *83*, 195–206. [[CrossRef](#)]
54. 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](#)]
55. Chrysafis, I.; Mallinis, G.; Gitas, I.; Tsakiri-Strati, M. Estimating Mediterranean forest parameters using multi seasonal Landsat 8 OLI imagery and an ensemble learning method. *Remote Sens. Environ.* **2017**, *199*, 154–166. [[CrossRef](#)]
56. Hame, T.; Rauste, Y.; Antropov, O.; Ahola, H.A.; Kilpi, J. Improved mapping of tropical forests with optical and SAR imagery, part II: Above ground biomass estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 92–101. [[CrossRef](#)]
57. Li, G.; Lu, D.; Moran, E.; Dutra, L.; Batistella, M. A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 26–38. [[CrossRef](#)]
58. Li, G.; Lu, D.; Moran, E.; Sant'Anna, S.J.S. Comparative analysis of classification algorithms and multiple sensor data for land use/land cover classification in the Brazilian Amazon. *J. Appl. Remote Sens.* **2012**, *6*, 61706. [[CrossRef](#)]



© 2018 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 (<http://creativecommons.org/licenses/by/4.0/>).