

Article

Mapping Forest Ecosystem Biomass Density for Xiangjiang River Basin by Combining Plot and Remote Sensing Data and Comparing Spatial Extrapolation Methods

Jia Zhu ^{1,2,3}, Zhihong Huang ^{1,2,*}, Hua Sun ^{3,4,5} and Guangxing Wang ^{3,4,5,*}

¹ National Engineering Laboratory for Applied Technology of Forestry & Ecology in South China, Changsha 410004, China; zhujia186@126.com

² Faculty of Life Science and Technology, Central South University of Forestry and Technology, Changsha 410004, China

³ Department of Geography and Environmental Resources, Southern Illinois University, Carbondale, IL 62901, USA

⁴ Research Center of Forestry Remote Sensing & Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China; sunhuayiw@vip.126.com

⁵ Key Laboratory of Forestry Remote Sensing Based Big Data & Ecological Security for Hunan Province, Changsha 410004, China

* Correspondence: huanghugh2013@yahoo.com (Z.H.); gxwang@siu.edu (G.W.);
Tel.: +86-731-8562-3483 (Z.H.); +1-618-453-6017 (G.W.)

Academic Editors: Erkki Tomppo, Huaiqing Zhang, Qi Chen, Lars T. Waser and Prasad S. Thenkabail
Received: 18 October 2016; Accepted: 1 March 2017; Published: 5 March 2017

Abstract: The distribution of forest biomass in a river basin usually has obvious spatial heterogeneity in relation to the locations of the upper and lower reaches of the basin. In the subtropical region of China, a large amount of forest biomass, comprising diverse forest types, plays an important role in maintaining the balance of the regional carbon cycle. However, accurately estimating forest ecosystem aboveground biomass density (AGB) and mapping its spatial variability at a scale of river basin remains a great challenge. In this study, we attempted to map the current AGB in the Xiangjiang River Basin in central southern China. Three approaches, including a multivariate linear regression (MLR) model, a logistic regression (LR) model, and an improved k-nearest neighbors (kNN) algorithm, were compared to generate accurate estimates and their spatial distribution of forest ecosystem AGB in the basin. Forest inventory data from 782 field plots across the basin and remote sensing images from Landsat 5 in the same period were combined. A stepwise regression method was utilized to select significant spectral variables and a leave-one-out cross-validation (LOOCV) technique was employed to compare their predictions and assess the methods. Results demonstrated the high spatial heterogeneity in the distribution of AGB across the basin. Moreover, the improved kNN algorithm with 10 nearest neighbors showed stronger ability of spatial interpolation than other two models, and provided greater potential of accurately generating population and spatially explicit predictions of forest ecosystem AGB in the complicated basin.

Keywords: multivariate linear regression; logistic regression; improved kNN algorithm; leave-one-out cross-validation; spatial distribution; Xiangjiang River Basin

1. Introduction

Forest biomass is one of the important variables for the quantitative study of structures and functions of forest ecosystems [1]. In the context of the Kyoto Protocol, forest biomass density is one

the key parameters for evaluating the potential of forest carbon sinks and studying global climate change in terms of offsetting greenhouse gas emissions [2]. However, accurately estimating and mapping forest ecosystem biomass density at large scales such as regionally, nationally and globally, is very challenging for the study of forest carbon sinks [3].

Traditional techniques based on field measurements are accurate, but time-consuming, labor-intensive and destructive to forest ecosystems [4,5]. In fact, the methods work for small scales only. Huxley [6] mathematically proposed an idea of using relative growth rates of biomass components that was then widely utilized by researchers to estimate biomass values of tree components (stem, branch and leaves) [7–13]. Subsequent approaches, such as the method proposed in Intergovernmental Panel on Climate Change (IPCC) [14], that is, using biomass conversion or expansion factors [15,16], were proposed to estimate forest ecosystem biomass density at large scales. However, these models proved to be poor in transferability [17]. Remote sensing technology has been widely applied for mapping forest ecosystem biomass density for regions, countries, and even the whole world due to its characteristics of rapidly, dynamically and repeatedly acquiring images of large areas [3,18–20]. The data used for remote sensing based prediction of forest ecosystem biomass include optical images [21,22], radar images [23,24], and LiDAR data [25–28]. However, optical remote sensing has serious obstacles such as cloud cover and saturation of reflectance due to high biomass and complexity of multi-layer forests [29–32]. A universal and accurate estimation model for forest ecosystem biomass has not been found [33].

There are three kinds of methods for forest ecosystem biomass estimation using remote sensing based empirical models (including parametric and non-parametric models), process models and simulation algorithms [33,34]. Wang et al. [34,35] proposed image-aided co-simulation algorithms in which spatial correlations of interest variables are taken into account. Both process models and simulation algorithms are advanced, but complicated and computation intensive, and thus difficult to use for generating spatially explicit estimates of forest ecosystem biomass density at large scales [34,35]. Moreover, the empirical models have been most widely used, but their prediction ability is often strictly limited by the data used. In addition, the regression models require normal distribution of variables and sometimes produce negative values [34,35]. Wang et al. [36] proposed a classification method for hierarchical multinomial logistic regression of hyperspectral images to improve the accuracy of regression modeling. In contrast, non-parametric empirical models are often more effective in mapping forest ecosystem biomass density due to the use of more flexible algorithms, no assumptions to be made about the forms of the data and distributions of variables, and no necessary inputs of complex parameters. These methods include k-nearest neighbors (kNN) method [24,37,38], artificial neural network (ANN) [39,40], random forest [41–44], support vector machine (SVM) [45,46], and maximum entropy [47–49]. Among the non-parametric models, kNN has become popular in recent years mainly due to its ability of spatial interpolation for complicated forest ecosystems in terms of topographic features and forest canopy structures [37,38,50,51].

Although various methods have been developed and used for mapping biomass density of forest ecosystems using different remote sensing data, there have been few reports that indicate which method can lead to most accurate population estimates and their spatial distributions at watershed scales [52]. The aim of this study was to demonstrate a novel and cost-effective mapping method of above-ground biomass (AGB) for forest ecosystems in the Xiangjiang River basin based on a combination of existing forest inventory sample plot data and free Landsat images. This was achieved by improving k-nearest neighbors (kNN) algorithm using correlations between spectral variables and forest ecosystem AGB to calculate weighted spectral distances of unknown pixels to each sample plot. This method was called correlation-weighted kNN (CW-kNN) and its results with different k nearest neighbors were compared with those from a widely used multivariate linear regression (MLR) model, a logistic regression (LR) model, and a general kNN algorithm (g-kNN) using a leave-one-out cross-validation (LOOCV) technique. It has to be pointed out that, in this study, the abbreviation AGB actually implied aboveground biomass density of forest ecosystems.

2. Materials and Methods

2.1. Study Area

The Xiangjiang River is the largest tributary of the Dongting Lake river system in the Yangtze River Basin. The Xiangjiang River Basin is located mainly in Hunan Province ($110^{\circ}31'–114^{\circ}15'E$, $24^{\circ}31'–29^{\circ}52'N$), China (Figure 1). The basin has complicated topographic features, and is surrounded by mountains and hills in the eastern, southern, and western parts of the basin and dominated by flat terrains in the central and northern parts. The basin has a humid mid-subtropical monsoon climate with distinct seasonality [53,54]. The mean annual temperature is approximately $17.5^{\circ}C$ with a mean annual precipitation of 1400 mm. The annual precipitation distribution is uneven with rainfall mainly in spring and summer. The annual average evaporation is about 1200 mm, with a frost-free period of 270–311 days. The main soil types include red soil, paddy soil, purple soil, yellow soil, lime (rock) soil, and yellow brown soil [55]. The typical vegetation in the basin is subtropical evergreen broad-leaved forests with a forest cover percentage of 54.4%. The basin has been subject to population increase and anthropogenic disturbances and it is necessary to know the spatial distribution and variability of the forest ecosystem AGB, and its dynamic change and trend in the basin.

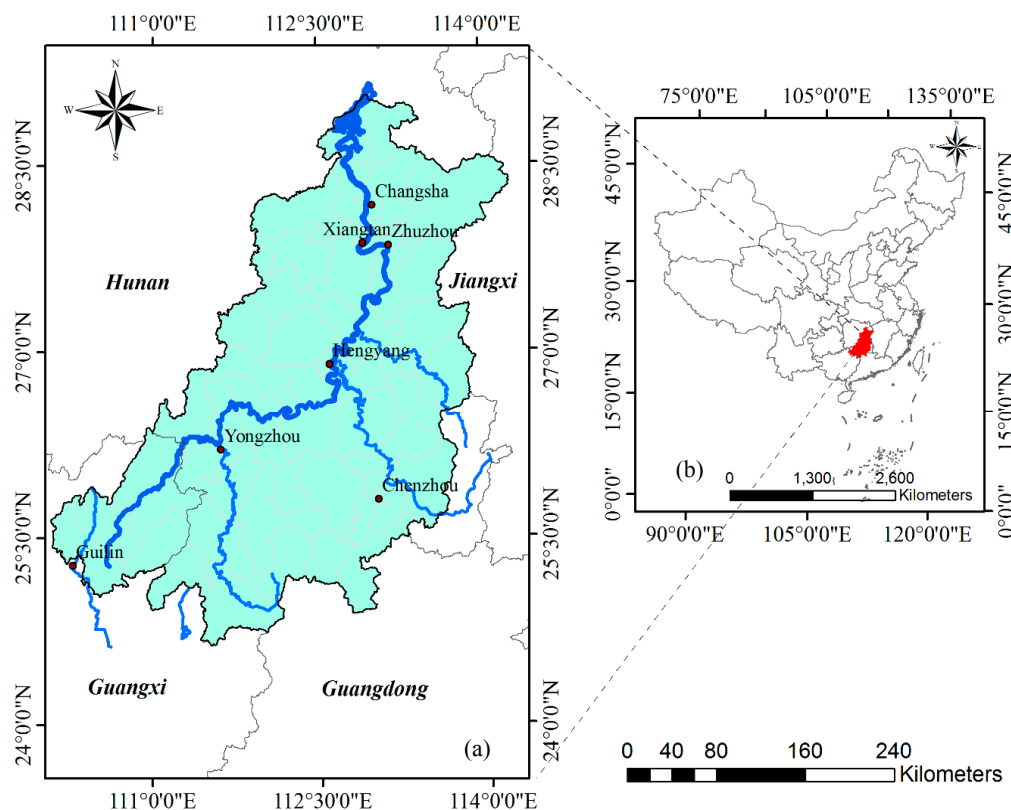


Figure 1. The study area: (a) the Xiangjiang River Basin in cyan with the river in blue; and (b) the basin location in China marked in red.

2.2. Data Sets

2.2.1. Forest Inventory Data

The forest inventory database of the Xiangjiang River Basin was obtained and provided information on the area size, spatial distribution and condition of forests. The information consists of detailed coordinates, canopy cover percentage, average shrub height and cover percentage, vegetation height, herbaceous cover percentage, stand types, dominant tree species, average diameter at breast

height, average stand height, number of non-timber forest trees, number of trees surrounding villages, houses, roads, and water bodies, number of bamboo trees, and number of miscellaneous bamboo stalks. The plot-level data were collected from a total of 782 fixed or permanent plots across the Basin in the summer of 2009, with a sampling interval of 4 km × 8 km and a square plot area of 0.067 ha (25.82 m × 25.82 m).

By using the regression models reported by Li et al. [56], plot-level biomass was estimated according to tree species groups, and the biomass of non-timber forests was estimated with average ground diameters. The biomass of shrubs and herbs was obtained by using the models proposed by Fan et al. [57] and estimating biomass of individual plants with regression equations based on height of shrubs and herbs. The total biomass of shrubs and herbs was obtained by summing the values of individual plant biomass and converting the values per unit. The values of biomass for the sample plots of mixed coniferous forests, mixed broad-leaved forests, and mixed coniferous and broad-leaved forests, were estimated by using tree species biomass and corresponding mixed percentage. The average AGB for all forest inventory field plots was 64.53 Mg/ha with a standard deviation of 46.80 Mg/ha and a confidence interval of 61.25 Mg/ha to 67.81 Mg/ha at a significant level of 0.05. The spatial distribution of the AGB estimates for the forest inventory sampled plots is shown in Figure 2. Relatively, the plot AGB biomass had larger values in northeastern, eastern, southern and southwestern mountainous and hilly areas, and smaller values in the central and northern flat areas. In this study, the plot AGB values were used as reference data.

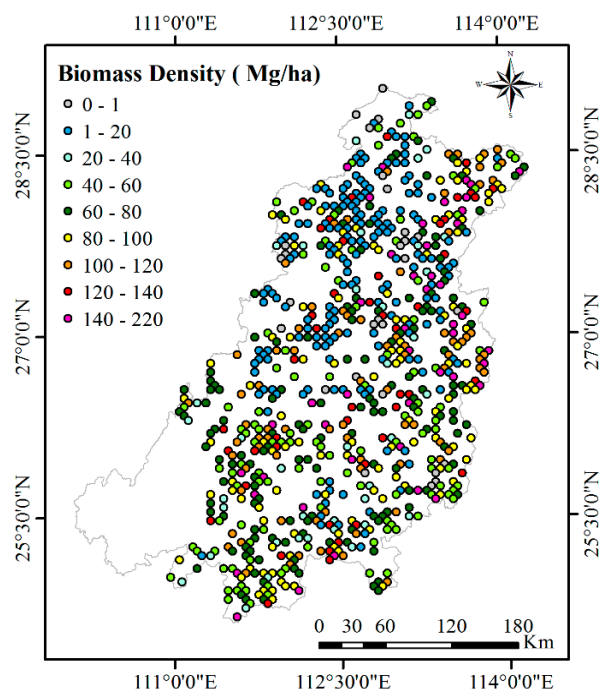


Figure 2. Spatial distribution of sampling plots and corresponding plot AGB values across the Xiangjiang River Basin.

2.2.2. Remote Sensing Data

The remote sensing images used in this study were obtained from Landsat 5 Thematic Mapper (TM) because the Landsat 5 data completely covered the study area and the acquired images were consistent with the field plot data collected in the summer of 2009. The images had six bands consisting of bands 1–5 and band 7 at a spatial resolution of 30 m. Seven adjacent and cloud-free images from the summer of 2009 were selected with path 123 and rows 40–43 and path 124 and rows 41–43. The images were susceptible to interference and distortion due to the sensor response characteristics and atmospheric absorption and scattering as well as other random factors. The images were enhanced by

radiometric, atmospheric and geometric correction, as well as ortho rectification and then mosaicked. The pixel digital numbers of the images were first converted to the values of radiance at the sensor's aperture and further to the values at-satellite reflectance using the parameters of Landsat 5 and solar zenith angles. The spectral values at-satellite reflectance were then converted to the reflectance values at ground surface. Moreover, the effects of slope, aspect and shade on the images were eliminated by conducting a topographic correction using Minnaert model. Finally, all the images were geo-referenced to the Universal Transverse Mercator (UTM) projection and coordinate system using a first-order affine transformation and a root mean square error (RMSE) of less than one pixel was yielded. Figure 3 shows a false color composite image by combining Landsat TM band 3, band 5 and band 4 as blue, green and red, respectively.

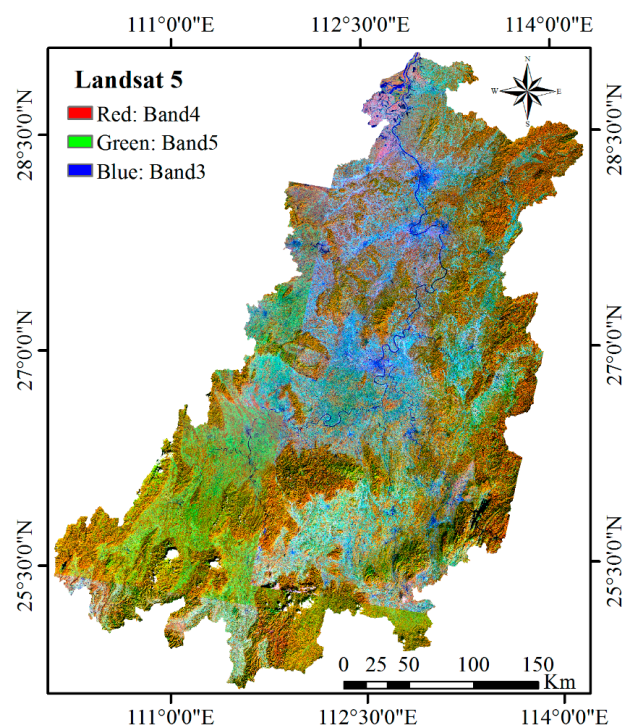


Figure 3. Landsat 5 TM composite image covering the study area after pre-processing.

2.3. Methods

2.3.1. Extraction and Selection of Spectral Variables

In this study, a total of 105 spectral variables were extracted, including normalized difference vegetation index (NDVI), infrared index II, spectral vegetation index (SVI), four soil-adjusted vegetation indices ($SAVI_l$, $l = 0.1, 0.25, 0.3$ and 0.5), modified soil-adjusted vegetation index (MSAVI), modified normalized difference vegetation index (MNDVI), difference vegetation index (DVI), transformed vegetation index (TVI), reduced simple ratio (RSR), atmospherically resistant vegetation index (ARVI), visible atmospherically resistant index (VARI), enhanced vegetation index (EVI), thirty-two-band ratio indices and sixty-three-band ratio indices. The spectral variables were selected to capture the characteristics and canopy structures of complicated forest ecosystems and to reduce the effects of slope and aspects [58]. The remote sensing variables as the independent variables of models might be significantly correlated with each other and the correlations would lead to the duplication of information and interfere the performance of the models. Thus, the spectral variables were first screened using significant coefficients of their correlations with the dependent variable—the plot AGB from the forest inventory database (FID) at the significance level of 0.05. A stepwise regression method with a variance inflation factor (VIF) ≥ 10 was utilized to diagnose the possible interference

of the correlations among the spectral variables; that is, multicollinearity. The used VIF value was determined based on a rule of thumb, literature and examination of data [59,60].

2.3.2. Multivariate Linear Regression (MLR) and Logistic Regression (LR) Model

In this study, both multivariate linear regression (MLR) model and logistic regression (LR) model were used to account for the relationship of forest ecosystem AGB with the spectral variables selected by stepwise regression analysis [61,62]. MLR is the most widely used method, but many studies have shown that the MLR has several shortcomings such as leading to negative and extremely large estimates. In this study, it was employed mainly for the purpose of comparison. LR is a method mainly used for analysis of binary dependent variables and probability prediction in which the predictions range from zero to one [63,64]. LR model can be expressed in its simplest form as proposed by Atkinson and Massari [63], and Devkota et al. [65]. In order to use the LR model to account for the relationship of the selected spectral variables with forest ecosystem AGB, range normalization of the biomass density data was implemented [60]. In the present study, the biomass data were transformed into a dimensionless quantity using the method of range normalization [59], which was in accordance with the dependent variable being probability distribution of the logistic model. The LR model was selected and compared with the MLR model, on one hand, to test whether the relationships of the forest ecosystem AGB with spectral variables were linear or non-linear, and on the other hand, to validate if the LR model performed better because of its non-linearity and positive estimates to be created than the MLR model in prediction of the forest ecosystem AGB.

2.3.3. kNN Algorithms

k-nearest neighbors (kNN) is one of non-parametric statistical methods and can be used to simultaneously estimate multiple forest ecosystem variables using the same underlying field dataset [37,38,66,67]. In this study, the estimation was conducted in a spectral space based on a spectral distance-weighted algorithm in which the estimate of AGB of pixel p was calculated [37,38,50,67] from Equation (1):

$$\hat{AGB}_p = \sum_{j=1}^k w_{pj} AGB_j \quad (1)$$

where k was the number of plots closest to pixel p in the spectral space, AGB_j was the AGB value for the j th plot, and w_{pj} was the weight for the j th nearest plot of pixel p . The plot weights w_{pj} were calculated from Equation (2):

$$w_{pj} = \frac{1}{d_{pj}} \left[\sum_{j=1}^k \frac{1}{d_{pj}} \right]^{-1} \quad (2)$$

The spectral distance metric between pixel p and the pixel corresponding to the j th plot, d_{pj} , was calculated from Equations (3) and (4):

$$d_{pj} = \sqrt{\sum_{l=1}^m v_l (x_{pl} - x_{jl})^2} \quad (3)$$

$$v_l = \frac{|r_l|}{\sum_{l=1}^m |r_l|} \quad (4)$$

where d_{pj} was the spectral distance between pixel p and the pixel corresponding to the j th plot, l was the l th remote sensing variable, m was the number of remote sensing variables, x_{pl} was the value of remote sensing variable l for pixel p , x_{jl} was the value of remote sensing variable l for pixel j , v_l was the weight for remote sensing variable l on spectral distance, and r_l was the correlation coefficient between remote sensing variable l and the AGB of the plot.

In this study, the kNN method was selected mainly because several studies have proven that it is appropriate for estimating and mapping forest ecosystem AGB at a large scale because of its non-parametric characteristics, that is, normal distributions of interest variables are not required [24,37,38,50,67]. Moreover, the kNN was improved by using the correlations between the spectral variables and plot AGB to calculate a weighted spectral distance of each unknown pixel to each sample plot. The improved kNN was called correlation-weighted kNN (CW-kNN). Compared with the general kNN algorithm (g-kNN) without the correlation-based weighting, CW-kNN takes into account the importance of spectral variables for mapping forest ecosystem AGB and provides the potential to more accurately account for the difference between one unknown pixel and one sample plot in the spectral space and thus in canopy structure and biomass density of forest ecosystems. In this study, both g-kNN and CW-kNN with four k values (3, 5, 7 and 10 nearest plots) were compared for mapping forest ecosystem AGB. Based on previous studies, using more than 10 neighbors did not produce a greater estimation accuracy [37,38,50,67,68] and, therefore, the maximum k value of 10 was selected.

2.3.4. Leave-One-Out Cross-Validation (LOOCV) and Model Evaluation

The LOOCV technique [69] was applied to evaluate the accuracies of three kinds of methods and corresponding AGB maps. The LOOCV followed the algorithm as reported by Ji et al. [70]: one single plot was withheld as a validation sample and the remaining plots were used to train the models. This step was repeated until each plot was used once as a validation sample. The AGB values of all field plots were then compared with their estimates using the “leave-one-out” training samples. The LOOCV technique has the advantage of providing an unbiased estimation of the prediction error [71].

Moreover, in this study four indices including the coefficient of determination (R^2), RMSE, the mean and variance values ($\hat{\mu}_{map}, Var_{map}$) of prediction maps, were employed to quantify the errors and to clarify which method performed most accurately in terms of plot and pixel levels. The uncertainty was quantified at the 95% confidence interval. $\hat{\mu}_{map}$ and Var_{map} were calculated using model-assisted regression estimators [51] from Equations (5) and (6):

$$\hat{\mu}_{map} = \frac{1}{N} \sum_{j=1}^N \hat{AGB}_j - \frac{1}{n} \sum_{i=1}^n \varepsilon_i \quad (5)$$

$$Var_{map} = \frac{1}{n(n-p)} \sum_{i=1}^n \varepsilon_i^2 \quad (6)$$

where $\hat{\mu}_{map}$ was the mean value of the predicted results at the pixel level in the study area, that is, a AGB map; N was the total number of pixels in the study area; n was the total number of sample plots used for the predictions; \hat{AGB}_j was the predicted value of each pixel; ε_i was the residual between the referenced and predicted value from the same plot i ; Var_{map} was the variance of the predicted results for the study area; and p was the number of variables used for modeling, including constant variables.

3. Results

3.1. Independent Variables of Models

A total of 82 spectral variables were significantly correlated with the plot AGB and the absolute values of the correlation coefficients ranged from 0.211 to 0.706. The coefficients of correlation for the five most correlated spectral variables were -0.706 , -0.698 , 0.697 , 0.683 , and 0.681 , corresponding to SR_{314} : band 3/(band 1 + band 4); SR_{324} : band 3/(band 2 + band 4); MNDVI: ((band 4 – band 3)/(band 4 + band 3))(1 – (band 5 – band 5_{min})/(band 5_{max} – band 5_{min})); SR_{436} : band 4/(band 3 + band 7); and NDVI: (band 4 – band 3)/(band 4 + band 3), respectively. It was found that many of the spectral variables were highly correlated with each other. In this study, the stepwise regression

analysis with a VIF value of 10 eliminated most of the correlated spectral variables and led to five spectral variables selected, including NDVI; SR_{23} : band 2/band 3; SR_{415} : band 4/(band 1 + band 5); SR_{546} : band 5/(band 4 + band 7); and SR_{625} : band 7/(band 2 + band 5) with the correlation coefficients of 0.618, 0.389, 0.646, -0.282 , and -0.401 , respectively. This implied that the stepwise regression analysis successfully excluded the spectral variables that contained similar information. For example, except for NDVI, other four most significant spectral variables were not selected because of their high correlations with NDVI. The five selected spectral variables were used as the independent variables in the multivariate linear regression model, logistic regression model, and kNN algorithms with plot AGB as the dependent variable.

3.2. Multivariate Linear Regression (MLR) Modeling

The five selected remote sensing variables were used as independent variables to fit a MLR model to the AGB data from the forest inventory dataset as follows:

$$AGB = 189.625 + 312.217NDVI - 38.934SR_{23} - 85.718SR_{415} - 136.343SR_{546} - 194.794SR_{625} \quad (7)$$

Based on the results of validation, there was a strong correlation between the plot predicted and referenced biomass density values (Figure 4a). The predicted average biomass density of the sample plots by the MLR method was almost the same as the average FID biomass. The mean $\hat{\mu}_{map}$ of the AGB map for the basin was slightly underestimated compared to the average biomass density value of the sample plots (Table 1). The underestimation mainly took place for the plots with AGB values greater than 100 Mg/ha and smaller 20 Mg/ha (Figure 4a). The distribution of the residuals as estimates of the error variance was in the shape of a horn (Figure 5a).

Table 1. The accuracy assessments of forest aboveground biomass density (AGB) estimates from a multivariate linear regression (MLR) model, a logistic regression (LR) model, and k-nearest neighbors (kNN) algorithms in the basin based on the leave-one-out cross-validation (LOOCV). The mean AGB means the average estimates for the forest inventory sample plots; and R^2 and RMSE are the coefficient of determination and root mean square error between the estimated and referenced values of the sample plots, respectively. $\hat{\mu}_{map}$ and Var_{map} are the mean estimate and its variance of a forest AGB map using model-assisted regression estimators from the MLR and LR models, and both the g-kNN without correlation based weighting and the CW-kNN with correlation based weighting.

Approach	Mean AGB (Mg/ha)	R^2	RMSE (Mg/ha)	$\hat{\mu}_{map}$ (Mg/ha)	Var_{map} (Mg/ha)
FID	64.53	—	—	—	—
MLR	64.51	0.54	31.55	60.31	1.30
LR	64.52	0.52	32.43	59.11	1.35
g-kNN	k = 3	0.48	34.78	60.24	1.56
	k = 5	0.51	32.90	59.74	1.39
	k = 7	0.53	32.14	59.55	1.33
	k = 10	0.54	31.87	59.32	1.31
CW-kNN	k = 3	0.48	34.66	59.74	1.55
	k = 5	0.51	32.95	59.88	1.40
	k = 7	0.53	32.36	59.69	1.35
	k = 10	0.54	31.93	59.47	1.31

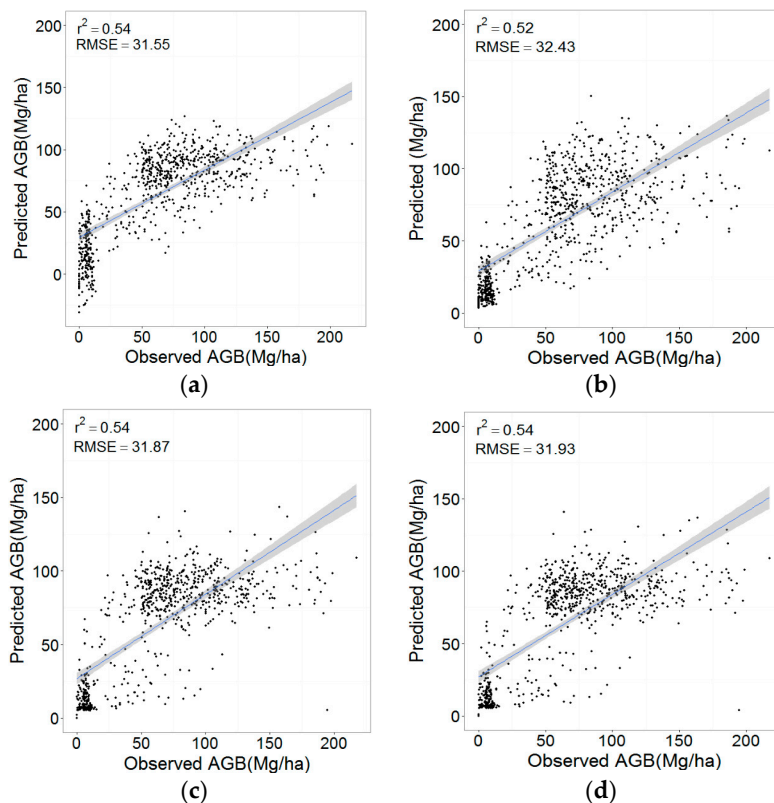


Figure 4. Predicted vs. observed (that is, referenced) AGB of sample plots using: (a) the multivariate linear regression (MLR); (b) the logistic regression (LR); (c) g-kNN with $k = 10$ nearest neighbors; and (d) CW-kNN with $k = 10$ nearest neighbors.

The percentages of the sample plot AGB values and the corresponding estimates falling in the biomass density intervals for the AGB map from the MLR are listed in Table 2. Compared to those of the sample plot AGB values, there were much smaller percentages of estimates falling in the biomass density intervals less than 20 Mg/ha and larger than 100 Mg/ha, indicating the underestimations happened in the intervals. The percentages of estimates falling in the biomass density intervals of 20–40 Mg/ha, 40–60 Mg/ha, 60–80 Mg/ha, and 80–100 Mg/ha were much greater than those of the sample plot AGB values, implying that overestimations took place in the intervals. In addition, the percentage for the biomass density interval less than 0 Mg/ha showed that the MLR method generated negative predictions at many places.

Table 2. Statistical percentages of forest aboveground biomass density (AGB) estimates falling in the biomass intervals for forest ecosystem AGB maps using MLR, LR, g-kNN, and CW-kNN modeling. The reference AGB was obtained from a forest inventory dataset (FID column).

Interval (Mg/ha)	FID	MLR	LR	g-kNN				CW-kNN			
				k = 3	k = 5	k = 7	k = 10	k = 3	k = 5	k = 7	k = 10
<0	0.00	5.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0~1	3.60	0.00	0.01	1.00	0.92	0.93	0.89	1.00	0.91	0.92	0.89
1~20	22.60	7.33	15.62	23.50	22.16	21.90	22.47	23.81	22.29	22.08	22.69
20~40	4.80	14.47	19.35	7.71	9.93	10.49	10.10	7.35	9.90	10.42	9.80
40~60	14.80	18.17	16.19	12.37	9.57	7.88	7.05	12.02	9.03	7.40	6.50
60~80	18.50	25.50	19.02	20.77	21.30	21.42	21.18	20.98	21.18	21.17	21.38
80~100	13.70	23.64	17.30	20.74	25.22	28.03	30.91	21.10	25.81	28.74	31.43
100~120	10.20	4.99	10.09	9.75	8.61	8.06	6.72	9.58	8.69	8.08	6.65
120~140	6.00	0.11	2.28	2.94	1.90	1.04	0.62	2.95	1.81	0.99	0.60
140~220	5.80	0.03	0.15	1.21	0.39	0.25	0.07	1.20	0.39	0.21	0.05

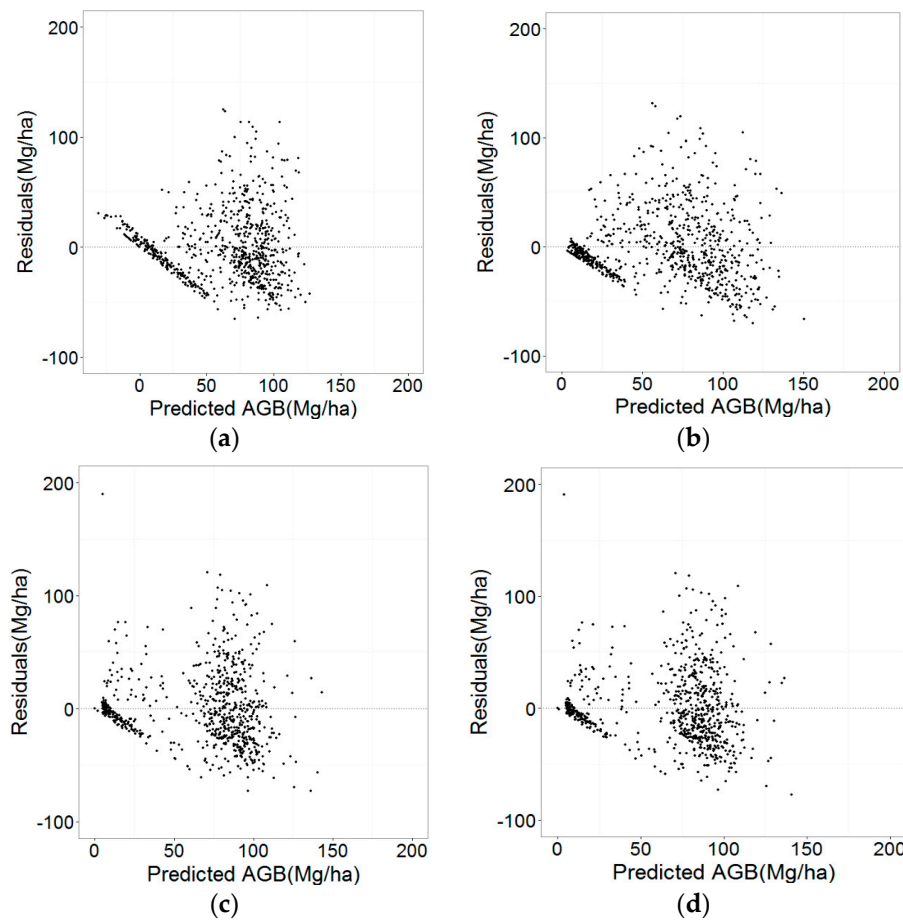


Figure 5. Distributions of residuals for predictions of sample plot AGB from: (a) the multivariate linear regression (MLR); (b) the logistic regression (LR); (c) g-kNN with $k = 10$ nearest neighbors; and (d) CW-kNN with $k = 10$ nearest neighbors.

3.3. Logistic Regression (LR) Modeling

The LR method led to following AGB estimation model:

$$AGB = 217.603 \frac{e^{(4.1681+10.9561NDVI-1.3100SR_{23}-3.9319SR_{415}-5.5985SR_{546}-5.9164SR_{625})}}{1 + e^{(4.1681+10.9561NDVI-1.3100SR_{23}-3.9319SR_{415}-5.5985SR_{546}-5.9164SR_{625})}} \quad (8)$$

The validation results showed that the coefficient of determination R^2 between the predictions and the plot biomass density values was statistically significant at the level of 0.05 (Table 1 and Figure 4b). The mean AGB prediction at the plot level was close to the sample plot average, but the mean $\hat{\mu}_{map}$ of the AGB map for the basin was smaller (Table 1). Compared with those from the MLR model (Figure 5a), the residuals of predictions from the LR model were distributed more evenly and randomly (Figure 5b). The LR model also led to underestimations for the plots with biomass density values smaller than 20 Mg/ha and larger than 120 Mg/ha (Figures 4b and 5b), but the underestimations were slightly improved compared to those from the MLR model (Figures 4a and 5a).

Unlike the MLR model, the LR method did not lead to negative predictions of biomass density (Table 2). Compared to those of the sample plot AGB values, there were smaller percentages of estimates falling in the biomass density intervals less than 20 Mg/ha and larger than 120 Mg/ha. This indicated that the LR method resulted in underestimations in the intervals, but the underestimations were improved compared to those by the MLR method. Moreover, the percentages of estimates falling in the biomass density intervals of 40–60 Mg/ha, 60–80 Mg/ha, and 80–100 Mg/ha were slightly larger

than those of the sample plot AGB values, implying that overestimations took place in the intervals, but were not very serious. The great overestimations happened mainly in the interval of 20–40 Mg/ha.

3.4. kNN Modeling

The mean predictions of kNN methods derived AGB for the FID-plots had a small range, but all fell in the confidence interval of the sample plot data (Table 1). For both kNN methods, the coefficient of determination R^2 between the plot predicted and referenced biomass density values increased gradually with the increasing number of k-nearest neighbors (Table 1 and Figure A1). Moreover, the larger the k value, the smaller the values of RMSE. For the same k values, the values of RMSE by CW-kNN were similar to those by g-kNN. Compared to those from the MLR and LR models, the RMSE values from both g-kNN and CW-kNN were slightly larger when $k = 3$ and 5 nearest neighbors and did not significantly differ when $k = 7$ and 10 nearest neighbors (Table 1 and Figure 4). The predicted $\hat{\mu}_{map}$ of the AGB maps from the kNN methods slightly fluctuated and the corresponding variance Var_{map} slightly decreased with the increasing values of k (Table 1). The map mean estimates were smaller than that of the sample plot data and fell out of the confidence interval, but very close to the lower bound, implying slight underestimations.

Similar to the LR method, the kNN algorithms did not result in negative estimates (Table 2). For both g-kNN and CW-kNN algorithms, the percentages of the map estimates falling in the biomass density intervals of 1–20 Mg/ha, 40–60 Mg/ha and 60–80 Mg/ha were very close to those of the sample plot AGB values. Compared to those of the sample plot AGB values, there were smaller percentages of the estimates in the biomass density intervals larger than 100 Mg/ha, while greater percentage existed in the interval of 80–100 Mg/ha. However, overall both g-kNN and CW-kNN algorithms greatly improved the underestimations and overestimations of AGB for the intervals compared to the MLR and LR models. Both g-kNN and CW-kNN algorithms led to similar characteristics of plot and pixel estimates.

The distributions of residuals looked horn-shaped for both kNN methods and for all k values (Figure A2). With an increase of k values, the residual distribution tended to be random, that is, the residuals fluctuating around the line of zero except for one plot with an extremely residual (Figure A2). Compared to those from the MLR and LR models, both g-kNN and CW-kNN algorithms improved the distributions of residuals (Figure 5).

3.5. Spatial Distribution of AGB

In Figure 6a,b, the spatial distributions of predicted AGB values for the basin by the MLR and LR models looked different from those by both g-kNN and CW-kNN algorithms with $k = 10$ nearest neighbors in Figure 6c,d, although the locations of the areas where large and small estimates existed were similar. The greater biomass estimates were distributed mainly in the northeastern and the upper reaches of the Xiangjiang River Basin, including the eastern, southern and southwestern parts. The smaller biomass estimates were allocated in the middle and lower reaches, that is, central, northwestern and northern parts of the basin. Among the methods, the LR model led to the highest spatial variability of predicted AGB values, especially in the northeast parts (Figure 6b), and then the MLR model (Figure 6a). Both g-kNN and CW-kNN algorithms with $k = 10$ resulted in similar spatial distributions of the estimates (Figure 6c,d) with lower spatial variability than that from the regression models. For the kNN algorithms, slightly higher spatial variability of the predicted values was derived by using smaller k values (Figure A3). In addition, the MLR model resulted in negative predictions in the middle and lower reaches (central, northwestern and northern parts) of the basin.

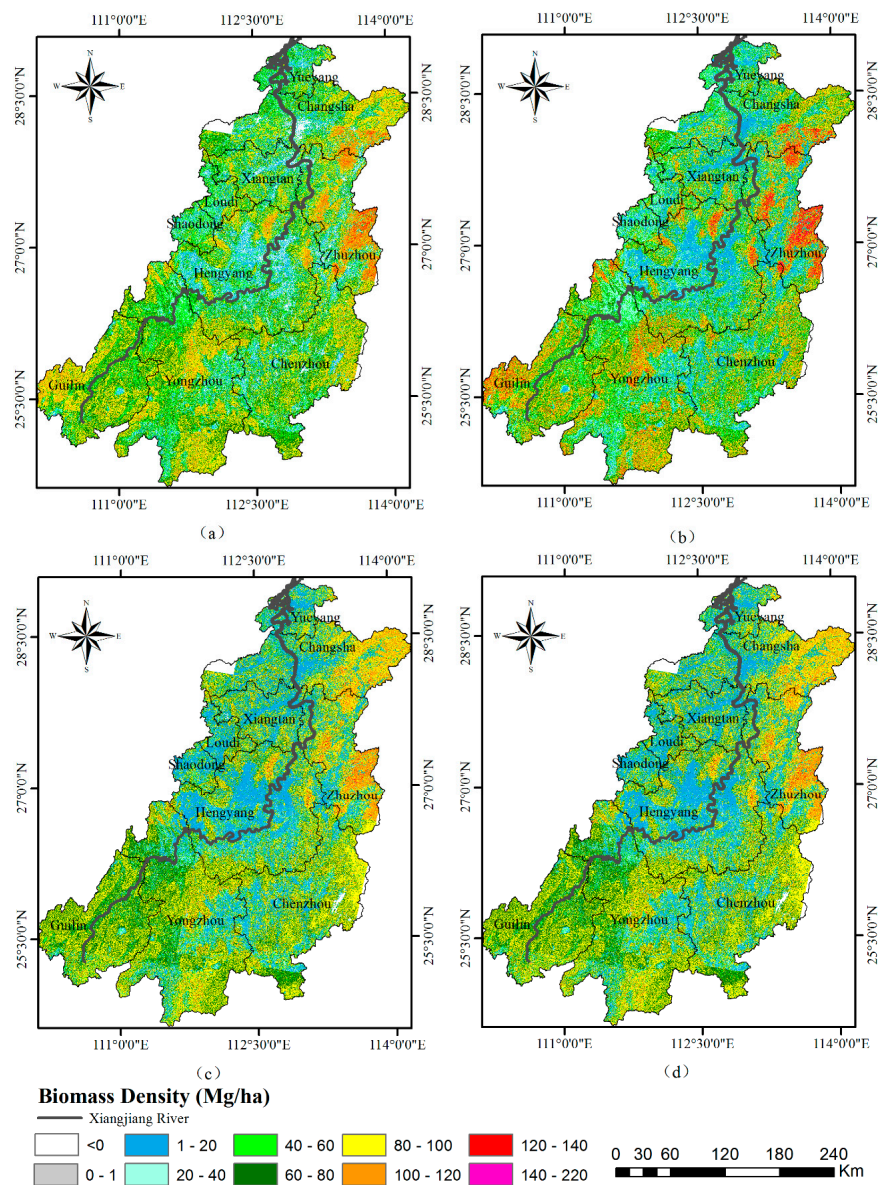


Figure 6. Spatial distributions, that is, maps of above-ground biomass density (AGB) predictions for the basin by: (a) the multivariate linear regression (MLR); (b) the logistic regression (LR); (c) the g-kNN with $k = 10$; and (d) the CW-kNN with $k = 10$.

4. Discussion

4.1. Rationality of Spectral Variable Selection

Although Landsat TM images have been widely employed for AGB estimation of forest ecosystems [4,52,72–78], extracting and selecting spectral variables to accurately derive spatial distribution of AGB is still challenging mainly due to the saturation of spectral reflectance and the presence of mixed pixels [4,58,79]. The image data and spectral bands from different sensors have their own characteristics in reflecting land surfaces [4]. Some vegetation indices such as simple ratios of spectral bands and *NDVI* obtained from Landsat data have been demonstrated to be useful predictors of biomass density in forest ecosystems [80–82]. However, not all vegetation indices are significantly correlated with AGB of forest ecosystems [4]. In this work, the independent variables that contributed to significantly improving the statistical fit of models to data and reducing the sum of squared errors were first selected from a total of 82 Landsat TM image derived spectral variables that

were significantly correlated with plot AGB. The spectral variables were utilized in all three kinds of methods and thus the differences among the estimates should be attributed to the properties of the algorithms. The selected spectral variables matching the ground survey of the sample plots in time were site-specific and could not be generalized for the use of other years or areas.

4.2. Comparison of Different Methods for Biomass Estimation

The selection of appropriate spatial extrapolation methods plays a central role in mapping biomass density of forest ecosystems [3,4,43]. In this work, three kinds of spatial modeling approaches, including MLR model, LR model, and kNN algorithms with and without spectral variable-AGB correlation based weighting (CW-kNN and g-kNN), were compared to yield the spatial estimates of AGB in the study area. The results showed that the smallest RMSE was obtained by the MLR, then the g-kNN and CW-kNN algorithms with $k = 10$. However, the differences of RMSE values were not significant. Moreover, the MLR resulted in negative estimates at many places (5.79% of the pixels). The distributions of the residuals and the scatter graph of predicted vs. referenced AGB values from the MLR showed a “horn” shape and underestimations existing for the sample plots and areas that had the AGB values smaller than 20 Mg/ha and larger than 100 Mg/ha. These implied that the relationship of the selected spectral variables with the forest ecosystem AGB was not linear [76]. We tested and verified the nonlinear relationships of the five selected spectral variables with forest ecosystem AGB. The distributions of the residuals and the underestimations were improved slightly by the non-linear LR model and greatly by the non-parametric methods CW-kNN and g-kNN with $k = 10$.

In addition, for the sample plots and areas with AGB values smaller than 20 Mg/ha, the underestimations could be caused by the impacts of spectral reflectance from soils and bare lands within young forest stands on the values of the selected spectral variables. On the other hand, the underestimations for the sample plots and areas with AGB values larger than 100 Mg/ha for the MLR model or 120 Mg/ha for the LR model were also due to reflectance saturation of multi-layer canopy and high biomass forest stands [79]. More importantly, this was mainly because both regression models were global methods that modeled and used global trends of AGB to generate estimates of local areas or pixels and in contrast, both CW-kNN and g-kNN were local methods that modeled and utilized local variability of AGB to create the estimates and thus improved the underestimations.

It had to be pointed out that both g-kNN and CW-kNN algorithms led to a great residual for one sample plot located in a shadow slope close to the south border of the basin with elevation of 1120 m and covered by a high dense and mixed forest of deciduous trees and shrub. This suggested that, although a topographic correction was conducted in this study, the effects of shadows still existed. If a more advanced method for topographic correction can be developed, the accuracy of AGB estimation can be further increased. In addition, in this study the kNN algorithms were investigated only using k values not larger than 10, mainly because based on previous reports using a k value larger than 10 would often smooth the spatial distribution of AGB estimates and not increase the accuracy of estimation [37,38,50,67,68].

4.3. Spatial Distribution of AGB in the Xiangjiang River Basin

Compared with those from the MLR and kNN algorithms, the spatial distribution of AGB estimates using the LR model had the higher spatial variability, especially in the northeast part of the basin (Figure 6b), because of its non-linear characteristics. However, all the predicted maps (Figure 6) were characterized by larger estimates of AGB distributed in the northeastern parts of the basin and the upper reaches (eastern, southern and southwestern parts) and smaller predictions of AGB allocated in the middle reaches (central areas) and the lower reaches (northwestern and northern parts). The spatial characteristics and patterns of AGB estimates were consistent with those of the sample plot biomass density values and reasonable because the northeastern (Liuyang of Changsha and Liling of Zhuzhou), eastern (Youxian and Chalin county of Zhuzhou), southern (Yongzhou) and southwest (Guilin of Guangxi) parts of the basin were mountainous and hilly areas and dominated by various

high dense forests, and the central (Hengyang), northwestern (Xiangtan) and northern (Changsha) parts were flat areas and dominated by cities and villages.

The spatial patterns of AGB predictions were supported by the results of the report from Jiao et al. [83,84], i.e., the estimates of AGB were relatively greater in the areas of Yongzhou and Chengzhou (the upper reaches of the basin) and smaller in the areas of Hengyang and Xiangtan cities (the middle and lower reaches of the basin). The similar spatial patterns of carbon storage for *Cunninghamia lanceolata* were characterized by Huang and Tong [85]. The AGB values of *Pinus massoniana* forests tended to decrease from the southwestern and southern parts to the northern parts of Hunan Province [86]. Xiangtan City had the lowest forest cover percentage and AGB value, followed by Hengyang City due to the implementation of the Hunan Agricultural Development Strategy in the mid-1990s, resulting in a decrease in the forested land in the areas of the basin [87,88]. However, in recent years net primary production (NPP) has been increasing in the middle and downstream reaches of the basin due to the programs of the “returning farmland to forests” and “forest protection”, based on the data of four forest inventories from 1983 to 2009 with the area of broad-leaved forests increased by four times [87,88].

4.4. Comparison with Previous Biomass Estimations

The average AGB estimates of the sample plots for the forest ecosystems in the Xiangjiang River Basin in 2009 varied from 63.84 Mg/ha to 64.52 Mg/ha, very close to the referenced value (64.53 Mg/ha) of the plots measured in the same year. The corresponding average estimates of the obtained maps ranged from 59.32 Mg/ha to 60.31 Mg/ha, slightly underestimated compared to the referenced value. Based on previous studies [83,89], the average AGB estimates of vegetation ecosystems (including forests, shrubs and grasslands) in the subtropical regions of China had a great range of 31.76 Mg/ha to 74.78 Mg/ha. In Hunan province, the forest average AGB values obtained from the sample plots of the 4th and 8th national forest inventories in 1990 and 2009 respectively were 31.76 Mg/ha [83] and 27.56 Mg/ha. For the Xiangjiang River Basin, the forest average AGB values obtained from the sample plots of national forest inventories in 1999, 2004 and 2009 respectively were 32.55 Mg/ha, 19.5 Mg/ha and 47.25 Mg/ha. This implied that the AGB values of forests in the Xiangjiang River Basin were larger than those of the whole Hunan in the same years mainly because the upper reaches of the Xiangjiang River was dominated by various protected forests. However, the value of 47.25 Mg/ha in 2009 did not include biomass of shrubs and grass within the forests. This study dealt with the forest ecosystems of the Xiangjiang River Basin consisting of forests, and shrubs and grass within the forests and thus, the obtained average AGB estimates in this study were greater than that of the forests from the sample plots of the 8th national forest inventory in 2009. The forest AGB of the Xiangjiang River Basin varied greatly over time from 1999 to 2004 and 2009, which was mainly caused by Human activities including city sprawling and urbanization, returning farmland to forests and forest protection.

4.5. Uncertainty Analysis of Forest Biomass

The estimates of AGB for the forest ecosystems of the Xiangjiang River basin are associated with uncertainties. The uncertainties may come from the images used and measurement errors of field observations for the tree variables involved in allometric equations, including tree height and diameter at breast height (DBH) [4,90–92]. In the present work, five remote sensing variables were selected and they were all significantly correlated with the sample plot AGB at the significance level of 0.01. Using the same five remote sensing variables three kinds of approaches were employed to produce the AGB estimates. The estimates of the sample plots from the methods were similar to the FID-based AGB values, suggesting that the five spectral variables contributed to statistically significantly improving the fit of the models to the data and reducing the sum of square errors. In this study, the forest inventory dataset collected in 2009 served as a reference, and allowed us to test and compare the remote sensing based methods. The dataset contained sampling and measurement errors of tree height and DBH that led to uncertainty of AGB estimates for the forest ecosystems of the Xiangjiang River

basin. However, this study did not analyze the effects of sampling and measurement errors on the accuracy of estimation because of lacking the corresponding information of the tree variables.

Moreover, the coefficients of allometric equations are site-specific [8] and not generalized. Applying the allometric equations to other study areas will cause uncertainties of forest ecosystem AGB estimates [11,93–95]. In this study, several allometric equations were used to estimate the values of AGB of trees by species; however, the uncertainty from the used allometric equations was unknown. In addition, Tang et al. [54] demonstrated that selecting spatial extrapolation methods is also vital for estimation and mapping of forest ecosystem AGB. In this study, three kinds of methods were compared using the same dataset and spectral variables. Based on the average estimates of the AGB maps at the pixel level, both CW-kNN and g-kNN had more robust ability of spatial extrapolation than the MLR and LR models.

5. Conclusions

The objective of this study was to develop a novel and cost-effective mapping method of forest ecosystem AGB for the Xiangjiang River Basin by combining an existing forest inventory database and free Landsat images and by improving kNN algorithm and comparing it with other three spatial extrapolation methods. It was found that: (1) spectral variables *NDVI*, *SR*₂₃, *SR*₄₁₅, *SR*₅₄₆, and *SR*₆₂₅ had significant contributions to improving the models' fit to the data and reducing the residuals of predictions; (2) all the spatial extrapolation methods led to reasonable and similar spatial patterns of forest ecosystem AGB estimates, but different spatial variability, and both the original g-kNN and the improved CW-kNN with 10 nearest plots resulted in slightly smaller RMSE than the LR model; (3) there was no significant difference between the g-kNN and the improved CW-kNN in terms of estimation accuracy at both plot and pixel levels; (4) compared to the MLR and LR models, both the g-kNN and CW-kNN algorithms improved the distributions of residuals of predictions and showed stronger ability of spatial interpolation; and (5) although the MLR model created the smallest RMSE, it led to negative estimates and was not appropriate for mapping the biomass of the forest ecosystems. Thus, this study suggested that the CW-kNN or g-kNN algorithm provided greater potential for generating accurate plot level estimates and spatially explicit predictions of AGB for the forest ecosystems in the complicated basin by combining free Landsat TM images and existing plot data, and can be applied to other areas at similar scales.

Acknowledgments: We would like to thank the faculty and staff of the Forest Ecology Laboratory, CSUFT, for their advice and assistance with this study. We appreciate Ying Wang of the International Institute for Earth System Science of Nanjing University for constructive comments on the draft of this paper. We also thank Lei Ji of Earth Resources Observation and Science (EROS) of the U.S. Geological Survey (USGS) for reviewing the manuscript and providing valuable comments. Supported by the project of Synergy Simulation of multi-resolution remote sensing data for city vegetation carbon mapping and uncertainty analysis funded by Hunan Provincial Education Department. All rights reserved.

Author Contributions: Jia Zhu, Zhihong Huang, and Hua Sun designed the study. Hua Sun collected the sample data and conducted the preprocessing of remote sensing and forest inventory data. Jia Zhu completed the calculations, map production and the accuracy assessment. Jia Zhu completed the first version of the draft. Guangxing Wang and Zhihong Huang wrote and revised the manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

AGB	aboveground biomass density
kNN	the k-nearest neighbors algorithm
LOOCV	Leave-one-out cross-validation
LR	Logistic regression model
MLR	Multivariate linear regression model
NDVI	normalized difference vegetation index

SVI	spectral vegetation index
MSAVI	modified soil-adjusted vegetation index
MNDVI	modified normalized difference vegetation index
DVI	difference vegetation index
TVI	transformed vegetation index
RSR	reduced simple ratio
ARVI	atmospherically resistant vegetation index
VARI	visible atmospherically resistant index
EVI	enhanced vegetation index

Appendix A

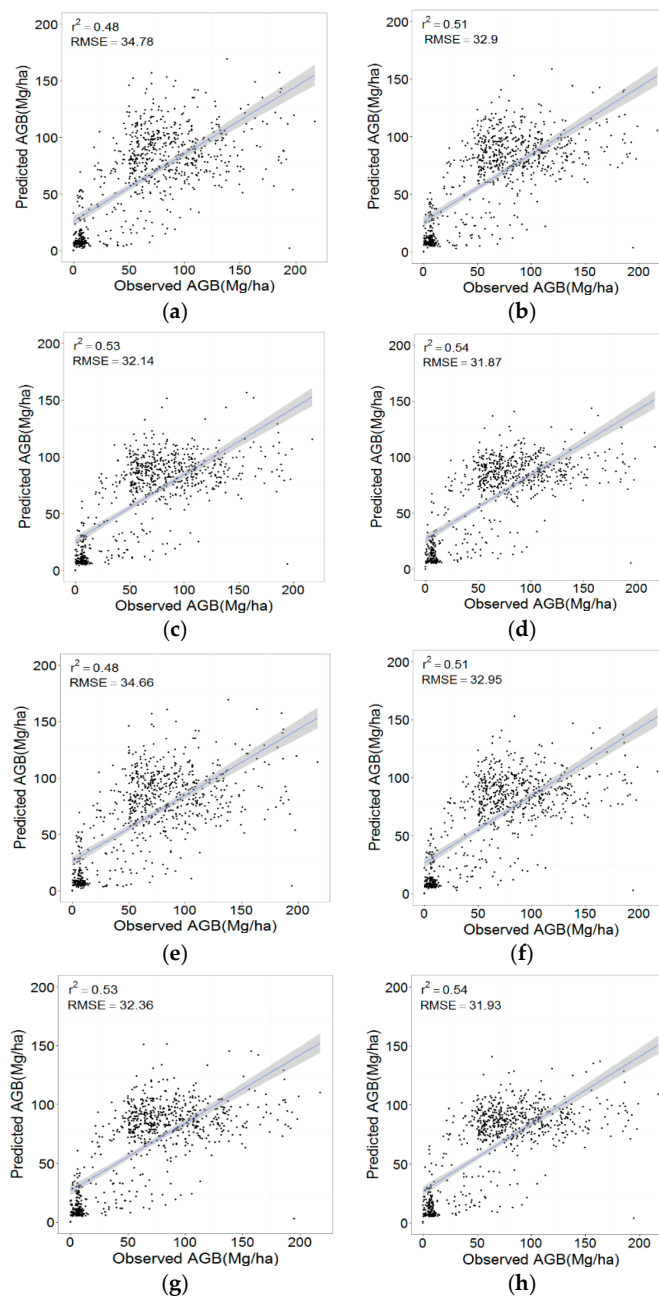


Figure A1. Predicted vs. observed (that is, referenced) AGB of sample plots using g-kNN with k value: (a) 3; (b) 5; (c) 7; and (d) 10; and CW-kNN with k value: (e) 3; (f) 5; (g) 7; and (h) 10.

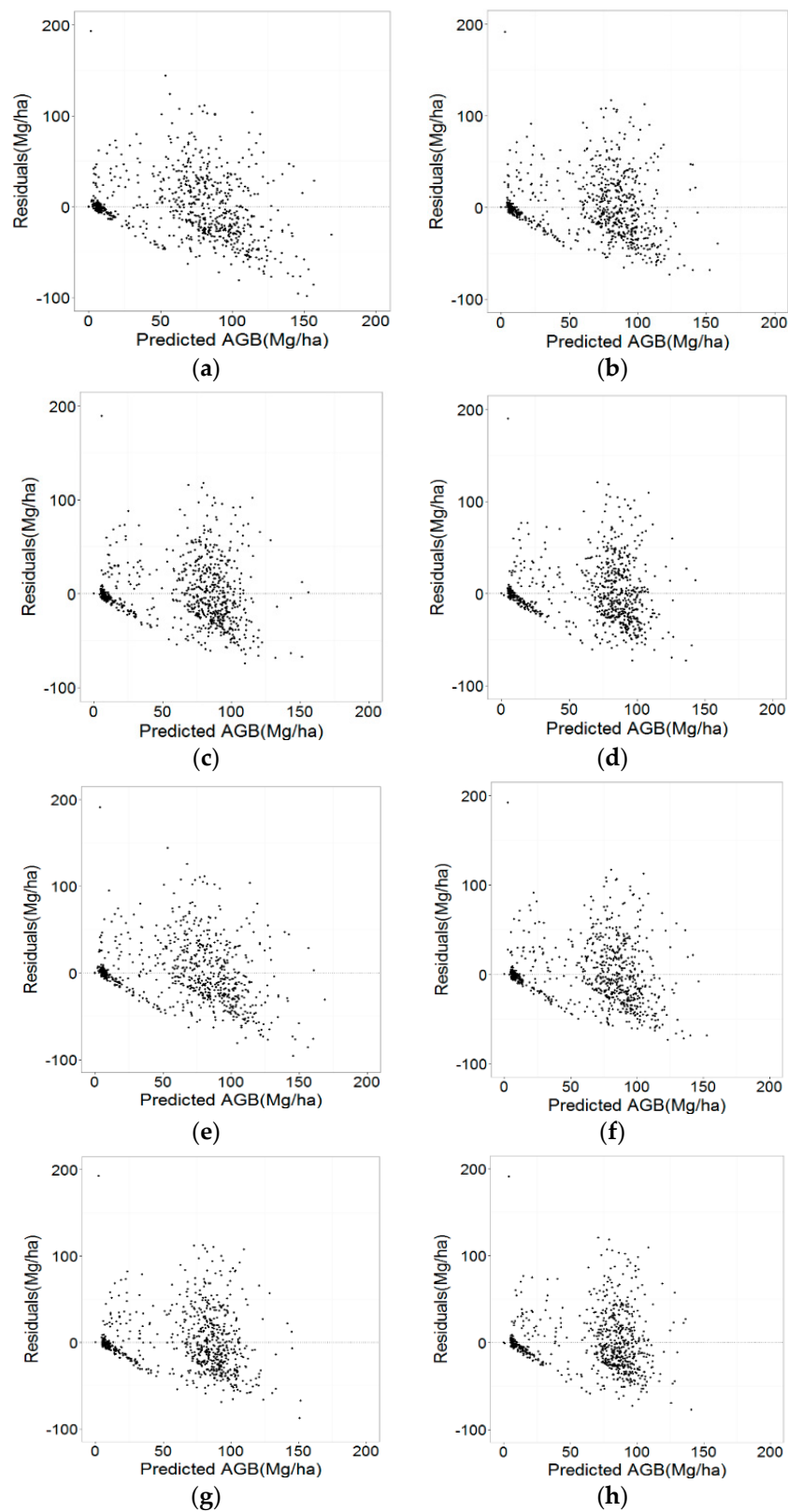


Figure A2. Distributions of residuals for predictions of sample plot AGB using g-kNN algorithm with k values: (a) 3; (b) 5; (c) 7; and (d) 10; and CW-kNN algorithm with k values: (e) 3; (f) 5; (g) 7; and (h) 10.

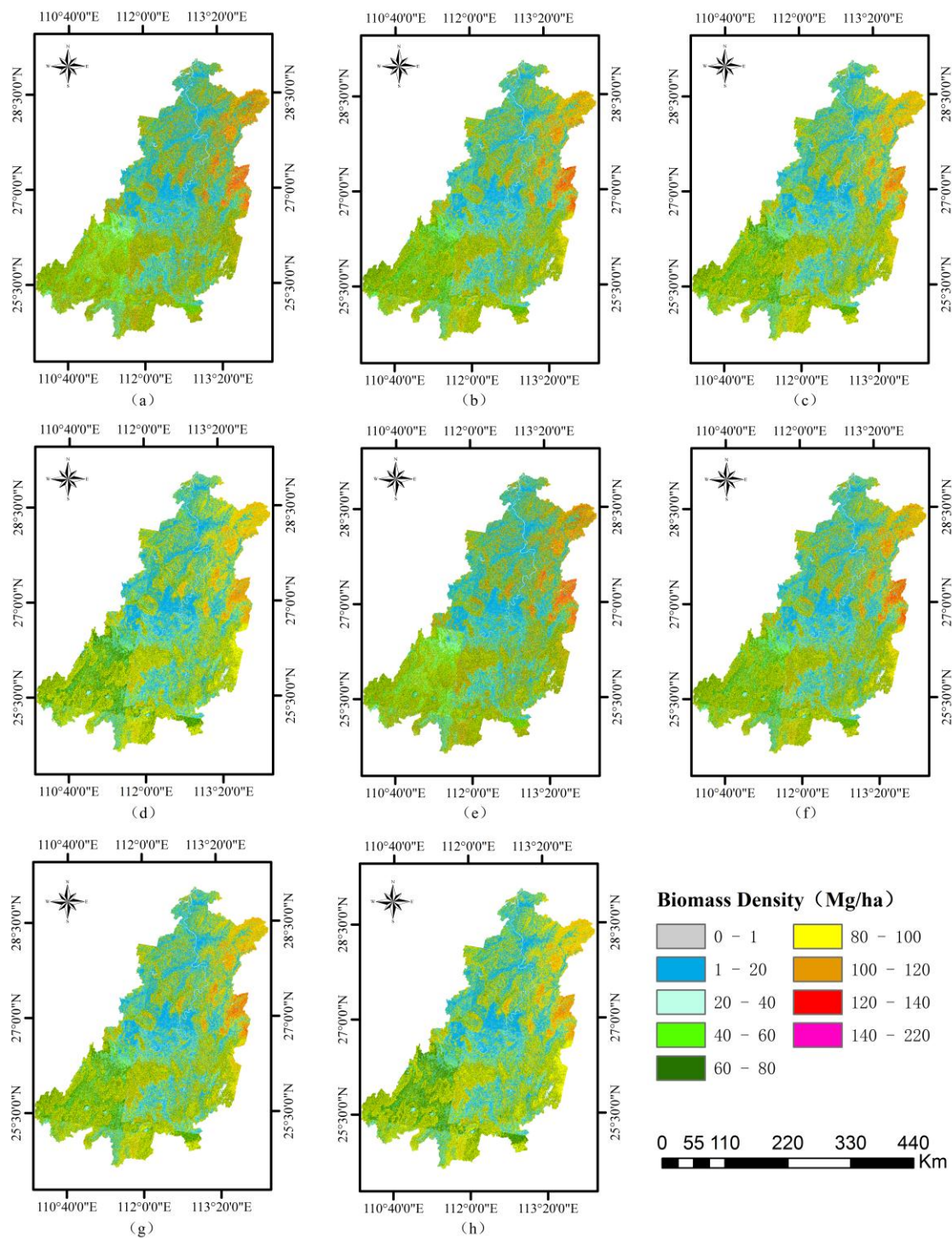


Figure A3. Spatial distributions, that is, maps of above-ground biomass density (AGB) predictions for the basin by using general k nearest neighbors (g- k NN) algorithm with numbers of nearest neighbors: (a) $k = 3$; (b) $k = 5$; (c) $k = 7$; and (d) $k = 10$; and spectral variable-AGB correlation-weighted k NN (CW- k NN) algorithm with numbers of nearest neighbors: (e) $k = 3$; (f) $k = 5$; (g) $k = 7$; and (h) $k = 10$.

References

1. Xiang, W.H.; Tian, D.L.; Yan, W.D. Review of researches on forest biomass and productivity. *J. Cent. South For. Invent. Plan.* **2003**, *22*, 57–60. (In Chinese)

2. Hu, Y.; Su, Z.; Li, W.; Li, J.; Ke, X. Influence of tree species composition and community structure on carbon density in a subtropical forest. *PLoS ONE* **2015**, *10*, e0136984. [[CrossRef](#)] [[PubMed](#)]
3. Zhang, J.; Huang, S.; Hogg, E.H.; Loeffers, 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](#)]
4. Lu, D. The potential and challenge of remote sensing-based biomass estimation. *Int. J. Remote Sens.* **2006**, *27*, 1297–1328. [[CrossRef](#)]
5. Klinge, H.; Rodrigues, W.A.; Brunig, E.; Fittkau, E.J. Biomass and structure in a Central Amazonian rain forest. In *Ecological Studies: Tropical Ecological Systems*; Golley, F.B., Medina, E., Eds.; Springer: Berlin, Germany, 1975; pp. 115–122.
6. Huxley, J.S. The variation in the width of the abdomen in immature fiddler crabs considered in relation to its relative growth-rate. *Am. Nat.* **1924**, *58*, 468–475. [[CrossRef](#)]
7. Brown, S.; Gillespie, A.; Lugo, A. Biomass estimation methods for tropical forests with applications to forest inventory data. *For. Sci.* **1989**, *35*, 881–902.
8. Ketterings, Q.M.; Coe, R.; van Noordwijk, M.; Palm, C.A. Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. *For. Ecol. Manag.* **2001**, *146*, 199–209. [[CrossRef](#)]
9. Geudens, G.; Staelens, J.; Kint, V.; Goris, R.; Lust, N. Allometric biomass equations for Scots pine (*Pinus sylvestris* L.) seedlings during the first years of establishment in dense natural regeneration. *Ann. For. Sci.* **2004**, *61*, 653–659. [[CrossRef](#)]
10. Bi, H.; Turner, J.; Lambert, M.J. Additive biomass equations for native eucalypt forest trees of temperate Australia. *Trees* **2004**, *18*, 467–479. [[CrossRef](#)]
11. Kenzo, T.; Furutani, R.; Hattori, D.; Kendawang, J.J.; Tanaka, S.; Sakurai, K.; Ninomiya, I. Allometric equations for accurate estimation of above-ground biomass in logged-over tropical rainforests in Sarawak, Malaysia. *J. For. Res.* **2009**, *14*, 365–372. [[CrossRef](#)]
12. Tinker, D.; Stakes, G.K.; Arcano, R.M. Allometric equation development, biomass, and aboveground productivity in ponderosa pine forests, Black Hills, Wyoming. *West. J. Appl. For.* **2010**, *25*, 112–119.
13. Blujdea, V.N.B.; Pilli, R.; Dutca, I.; Ciuvat, L.; Abrudan, I.V. Allometric biomass equations for young broadleaved trees in plantations in Romania. *For. Ecol. Manag.* **2012**, *264*, 172–184. [[CrossRef](#)]
14. Penman, J.; Gytarsky, M.; Hiraishi, T.; Krug, T.; Kruger, D.; Pipatti, R.; Buendia, L.; Miwa, K.; Ngara, T.; Tanabe, K.; et al. Definitions and methodological options to inventory emissions from direct human-induced degradation of forests and revegetation of other vegetation types. In *Good Practice Guidance for Land Use, Land-Use Change and Forestry*; Intergovernmental Panel on Climate Change (IPCC): Geneva, Switzerland; Institute for Global Environmental Strategies (IGES): Hayama, Japan, 2003.
15. Fang, J.; Chen, A.; Peng, C.; Zhao, S.; Ci, L. Changes in forest biomass carbon storage in China between 1949 and 1998. *Science* **2001**, *292*, 2320–2322. [[CrossRef](#)] [[PubMed](#)]
16. Fang, J.Y.; Wang, Z.M. Forest biomass estimation at regional and global levels, with special reference to China's forest biomass. *Ecol. Res.* **2001**, *16*, 587–592. [[CrossRef](#)]
17. Paustian, K.; Ravindranath, N.H.; Amstel, A.R.V. 2006 IPCC Guidelines for National Greenhouse Gas Inventories; Intergovernmental Panel on Climate Change (IPCC): Geneva, Switzerland, 2006.
18. Cohen, W.B.; Goward, S.N. Landsat's role in ecological applications of remote sensing. *BioScience* **2004**, *54*, 535–545. [[CrossRef](#)]
19. Boudreau, J.; Nelson, R.F.; Margolis, H.A.; Beaudoin, A.; Guindon, L.; Kimes, D.S. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sens. Environ.* **2008**, *112*, 3876–3890. [[CrossRef](#)]
20. Lucas, R.M.; Mitchell, A.L.; Armston, J. Measurement of forest above-ground biomass using active and passive remote sensing at large (subnational to global) scales. *Curr. For. Rep.* **2015**, *1*, 162–177. [[CrossRef](#)]
21. Kobayashi, S.; Sanga-Ngoie, K. The integrated radiometric correction of optical remote sensing imagery. *Int. J. Remote Sens.* **2008**, *29*, 5957–5985. [[CrossRef](#)]
22. Liu, J.K.; Shih, P.T. Topographic correction of wind-driven rainfall for landslide analysis in Central Taiwan with validation from aerial and satellite optical images. *Remote Sens.* **2013**, *5*, 2571–2589. [[CrossRef](#)]

23. Harrell, P.; Bourgeau-Chavez, L.L.; Kasischke, E.S.; French, N.H.F.; Christensen, N.L. Sensitivity of ERS-1 and JERS-1 radar data to biomass and stand structure in Alaskan boreal forest. *Remote Sens. Environ.* **1995**, *54*, 247–260. [[CrossRef](#)]
24. Konovalyuk, M.; Gorbunova, A.; Baev, A.; Kuznetsov, Y. Parametric reconstruction of radar image based on Multi-point Scattering Model. *Int. J. Microw. Wirel. Trans.* **2014**, *6*, 543–548. [[CrossRef](#)]
25. Zhao, K.; Popescu, S. Lidar-based mapping of leaf area index and its comparison with satellite GLOBCARBON LAI Products. *Remote Sens. Environ.* **2009**, *113*, 1628–1645. [[CrossRef](#)]
26. Koch, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS J. Photogramm.* **2010**, *65*, 581–590. [[CrossRef](#)]
27. Olsoy, P.J.; Glenn, N.F.; Clark, P.E.; Derryberry, D.R. Aboveground total and green biomass of dryland shrub derived from terrestrial laser scanning. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 166–173. [[CrossRef](#)]
28. Srinivasan, S.; Popescu, S.C.; Eriksson, M.; Sheridan, R.D.; Ku, N.W. Multi-temporal terrestrial laser scanning for modeling tree biomass change. *For. Ecol. Manag.* **2014**, *318*, 304–317. [[CrossRef](#)]
29. Sun, G.; Ranson, K.J.; Kharuk, V.I. Radiometric slope correction for forest biomass estimation from SAR data in the western Sayani Mountains, Siberia. *Remote Sens. Environ.* **2002**, *79*, 279–287. [[CrossRef](#)]
30. Neilson, E.T.; MacLean, D.A.; Meng, F.R.; Arp, P.A. Spatial distribution of carbon in natural and management stands in an industrial forest in New Brunswick, Canada. *For. Ecol. Manag.* **2007**, *253*, 148–160. [[CrossRef](#)]
31. Nafiseh, G.; Mahmood, R.S.; Mohammadzadeh, A. A review on biomass estimation methods using synthetic aperture radar data. *Int. J. Geomat. Geosci.* **2011**, *1*, 776–788.
32. Sinha, S.; Jeganathan, C.; Sharma, L.K.; Nathawat, M.S. A review of radar remote sensing for biomass estimation. *Int. J. Environ. Sci. Technol.* **2015**, *12*, 1779–1792. [[CrossRef](#)]
33. 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* **2014**, *9*, 1–43. [[CrossRef](#)]
34. Wang, G.; Oyana, T.; Zhang, M.; Adu-Prah, S.; Zeng, S.; Lin, H.; Se, J. Mapping and spatial uncertainty analysis of forest vegetation carbon by combining national forest inventory data and satellite images. *For. Ecol. Manag.* **2009**, *258*, 1275–1283. [[CrossRef](#)]
35. Wang, G.; Zhang, M.; Gertner, G.Z.; Oyana, T.; McRoberts, R.E.; Ge, H. Uncertainties of mapping forest carbon due to plot locations using national forest inventory plot and remotely sensed data. *Scand. J. For. Res.* **2011**, *26*, 360–373. [[CrossRef](#)]
36. Wang, C.; Wang, S.; Guo, Z.; Wang, L.; Ma, C. Study on classification method of hyperspectral remote sensing image based on hierarchical multinomial logistic regression algorithm. *Int. J. Earth Sci. Eng.* **2014**, *7*, 415–420.
37. 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](#)]
38. Tomppo, E.O.; Gagliano, C.; de Natale, F.; Katila, M.; McRoberts, R. Predicting categorical forest variables using an improved k-Nearest Neighbour estimator and Landsat imagery. *Remote Sens. Environ.* **2009**, *113*, 500–517. [[CrossRef](#)]
39. 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](#)]
40. Almeida, A.C.; Barros, P.L.C.; Monteiro, J.H.A.; Rocha, B.R.P. Estimation of above-ground forest biomass in Amazonia with neural networks and remote sensing. *IEEE Lat. Am. Trans.* **2009**, *7*, 27–32. [[CrossRef](#)]
41. Feng, Y.M.; Lei, X.D.; Lu, Y.C. Interpretation of pixel-missing patch of remote sensing image with Kriging interpolation of spatial statistics. *J. Remote Sens.* **2004**, *8*, 317–322.
42. Saatchi, S.; Halligan, K.; Despain, D.; Crabtree, R. Estimation of forest fuel load from radar remote sensing. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1726–1740. [[CrossRef](#)]
43. Saatchi, S.; Marlier, M.; Chazdon, R.L.; Clark, D.B.; Russell, A.E. Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. *Remote Sens. Environ.* **2011**, *115*, 2836–2849. [[CrossRef](#)]
44. Meng, Q.; Borders, B.; Madden, M. High-resolution satellite image fusion using regression Kriging. *Int. J. Remote Sens.* **2010**, *31*, 1857–1876. [[CrossRef](#)]
45. Diao, Y.; Zhang, C.; Liu, J.; Liang, Y.; Hou, X.; Gong, X. Optimization model to estimate Mount Tai forest biomass based on remote sensing. *IFIP Adv. Inf. Commun. Technol.* **2011**, *370*, 453–459.

46. Lee, J.H.; Im, J.H.; Kim, K.M.; Heo, J. Change Analysis of aboveground forest carbon stocks according to the land cover change using multi-temporal Landsat TM images and machine learning algorithms. *J. Korean Assoc. Geogr. Inf. Stud.* **2015**, *18*, 81–99. [[CrossRef](#)]
47. Jackett, C.J.; Turner, P.J.; Lovell, J.L.; Williams, R.N. Deconvolution of MODIS imagery using multiscale maximum entropy. *Remote Sens. Lett.* **2011**, *2*, 179–187. [[CrossRef](#)]
48. Li, W.K.; Guo, Q.H. A maximum entropy approach to one-class classification of remote sensing imagery. *Int. J. Remote Sens.* **2010**, *31*, 2227–2235. [[CrossRef](#)]
49. Rodríguez-Veiga, P.; Saatchi, S.; Tansey, K.; Balzter, H. Magnitude, spatial distribution and uncertainty of forest biomass stocks in Mexico. *Remote Sens. Environ.* **2016**, *183*, 265–281. [[CrossRef](#)]
50. Mcroberts, R.E.; Tomppo, E.O. Remote sensing support for national forest inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [[CrossRef](#)]
51. 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](#)]
52. Buma, B.; Krapek, J.; Edwards, R.T. Watershed-scale forest biomass distribution in a perhumid temperate rainforest as driven by topographic, soil, and disturbance variables. *Can. J. For. Res.* **2016**, *46*, 844–854. [[CrossRef](#)]
53. Zhao, L.; Xiang, W.; Li, J.; Lei, P.; Deng, X.; Fang, X.; Peng, C. Effects of topographic and soil factors on woody species assembly in a Chinese subtropical evergreen broadleaved forest. *Forests* **2015**, *6*, 650–669. [[CrossRef](#)]
54. Tang, X.; Fehrmann, L.; Guan, F.; Forrester, D.I.; Guisasola, R.; Kleinn, C. Inventory-based estimation of forest biomass in Shitai County, China: A comparison of five methods. *Ann. For. Res.* **2016**, *59*, 269–280. [[CrossRef](#)]
55. Luo, Q.; Wang, K.L.; Wang, Q.X. Using SWAT to simulate runoff under different land use scenarios in Xiangjiang River Basin. *Chin. J. Eco-Agric.* **2011**, *19*, 1431–1436. [[CrossRef](#)]
56. Li, H.K.; Lei, Y.C. *Estimation and Evaluation of Forest Biomass Carbon Storage in China*; China Forestry Press: Beijing, China, 2010. (In Chinese)
57. Fan, W.Y.; Zhang, H.Y.; Yu, Y.; Mao, X.G.; Yang, J.M. Comparison of three models of forest biomass estimation. *Chin. J. Plant Ecol.* **2011**, *35*, 402–410. (In Chinese) [[CrossRef](#)]
58. Sun, H.; Qie, G.P.; Wang, G.X.; Tan, Y.F.; Li, J.P.; Peng, Y.G.; Ma, Z.G.; Luo, C.Q. Increasing the accuracy of mapping urban forest carbon density by combining spatial modeling and spectral unmixing analysis. *Remote Sens.* **2015**, *7*, 15114–15139. [[CrossRef](#)]
59. Gao, H.X. *Applied Multivariate Statistical Analysis*, 1st ed.; Peking University Press: Beijing, China, 2005; pp. 125–220. (In Chinese)
60. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013; Available online: <http://www.R-project.org/> (accessed on 18 October 2016).
61. Vincent, G.; Sabatier, D.; Blanc, L.; Chave, J.; Weissenbacher, E.; Pélissier, R.; Fonty, E.; Molino, J.F.; Coutron, P. Accuracy of small footprint airborne LiDAR in its predictions of tropical moist forest stand structure. *Remote Sens. Environ.* **2012**, *125*, 23–33. [[CrossRef](#)]
62. He, Q.; Chen, E.; An, R.; Li, Y. Above-ground biomass and biomass components estimation using LiDAR data in a coniferous forest. *Forests* **2013**, *4*, 984–1002. [[CrossRef](#)]
63. Atkinson, P.M.; Massari, R. Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. *Comput. Geosci.* **1998**, *24*, 373–385. [[CrossRef](#)]
64. Wang, J.C.; Guo, Z.G. *Logistic Regression Models: Methods and Application*; Higher Education Press: Beijing, China, 2001; pp. 1–17. (In Chinese)
65. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat. Hazards* **2013**, *65*, 135–165. [[CrossRef](#)]
66. Franco-Lopez, H.; Ek, A.R.; Bauer, M.E. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sens. Environ.* **2001**, *77*, 251–274. [[CrossRef](#)]
67. Rajaniemi, S.; Tomppo, E.; Ruokolainen, K.; Tuomisto, H. Estimating and mapping pteridophyte and Melastomataceae species richness in western Amazonian rainforests. *Int. J. Remote Sens.* **2005**, *26*, 475–493. [[CrossRef](#)]

68. Liang, S.L.; Li, X.W.; Wang, J.D. *Quantitative Remote Sensing: Concepts and Algorithms*; Science Press: Beijing, China, 2013.
69. Moore, A.W.; Lee, M.S. Efficient algorithms for minimizing cross validation error. In Proceedings of the Eleventh International Conference on Machine Learning, New Brunswick, NJ, USA, 10–13 July 1994; pp. 190–198.
70. Ji, L.; Wylie, B.K.; Nossov, D.R.; Peterson, B.; Waldrop, M.P.; McFarland, J.W.; Rover, J.; Hollingsworth, T.N. Estimating aboveground biomass in interior Alaska with Landsat data and field measurements. *Int. J. Appl. Earth Observ. Geoinf.* **2012**, *18*, 451–461. [[CrossRef](#)]
71. Clevers, J.; van der Heijden, G.; Verzakov, S.; Schaepman, M. Estimating grassland biomass using SVM band shaving of hyperspectral data. *Photogr. Eng. Remote Sens.* **2007**, *73*, 1141–1148. [[CrossRef](#)]
72. Kajisa, T.; Murakami, T.; Mizoue, N.; Top, N.; Yoshida, S. Object-based forest biomass estimation using Landsat ETM+ in Kampong Thom Province, Cambodia. *J. For. Res.* **2009**, *14*, 203–211. [[CrossRef](#)]
73. Lu, D.; Chen, Q.; Wang, G.; Moran, E.; Batistella, M.; Zhang, M.; Laurin, G.V.; Saah, D. Aboveground forest biomass estimation with Landsat and Lidar data and uncertainty analysis of the estimates. *Int. J. For. Res.* **2012**, *2012*, 436537. [[CrossRef](#)]
74. Timothy, D.; Onesimo, M.; Cletah, S.; Adelabu, S.; Tsitsi, B. Remote sensing of aboveground forest biomass: A review. *Trop. Ecol.* **2016**, *57*, 125–132.
75. Nelson, R.F.; Kimes, D.S.; Salas, W.A.; Routhier, M. Secondary forest age and tropical forest biomass estimation using thematic mapper imagery. *BioScience* **2000**, *50*, 419–431. [[CrossRef](#)]
76. Foody, G.M.; Boyd, D.S.; Cutler, M.E. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sens. Environ.* **2003**, *85*, 463–474. [[CrossRef](#)]
77. Zheng, D.; Rademacher, J.; Chen, J.; Crow, T.; Bresee, M.; Le Moine, J.; Ryu, S.R. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sens. Environ.* **2004**, *93*, 402–411. [[CrossRef](#)]
78. Lu, D. Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon. *Int. J. Remote Sens.* **2005**, *26*, 2509–2525. [[CrossRef](#)]
79. 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**. [[CrossRef](#)]
80. Fassnacht, K.S.; Gower, S.T.; MacKenzie, M.D.; Nordheim, E.V.; Lillesand, T.M. Estimating the leaf area index of North Central Wisconsin forests using the landsat thematic mapper. *Remote Sens. Environ.* **1997**, *61*, 229–245. [[CrossRef](#)]
81. Jakubauskas, M.E. Thematic Mapper characterization of lodgepole pine seral stages in Yellowstone National Park, USA. *Remote Sens. Environ.* **1996**, *56*, 118–132. [[CrossRef](#)]
82. Steininger, M.K. Satellite estimation of tropical secondary forest above-ground biomass: Data from Brazil and Bolivia. *Int. J. Remote Sens.* **2000**, *21*, 1139–1157. [[CrossRef](#)]
83. Jiao, X.M.; Xiang, W.H.; Tian, D.L. Carbon Storage of Forest Vegetation and Its Geographical Distribution in Hunan Province. *J. Cent. South For. Univ.* **2005**, *25*, 4–8. (In Chinese)
84. Liu, C.; Liu, M.; Wang, K.L.; Chen, H.S. Evolution of landscape pattern in upper and middle reaches of Xiangjiang River. *Chin. J. Ecol.* **2007**, *26*, 1822–1827. (In Chinese)
85. Huang, X.N.; Tong, J. Dynamics of carbon storage of Chinese fir in Hunan province. *J. Cent. South Univ. For. Technol.* **2011**, *31*, 80–84. (In Chinese)
86. Xu, X.; Yang, D. Spatial distribution and dynamic changes of total biomass quantity of *Pinus massoniana* forests in Hunan province. *J. Cent. South Univ. For. Technol.* **2012**, *32*, 73–78.
87. Jiang, Y.T. Analysis of Spatio-Temporal Dynamics and Factors Influencing Vegetation NPP in Xiangjiang River Basin. Master's Thesis, Hunan University of Science and Technology, Xiangtan, China, May 2015. (In Chinese)
88. Liu, Z.D.; Li, B.; Fang, X.; Xiang, W.H.; Tian, D.L.; Yan, W.D.; Lei, P.F. Dynamic characteristics of forest carbon storage and carbon density in Hunan Province. *Acta Ecol. Sin.* **2016**, *36*, 6897–6908.
89. Wang, X.; Feng, Z.; Ouyang, Z. Vegetation carbon storage and density of forest ecosystems in China. *Chin. J. Appl. Ecol.* **2001**, *12*, 13–16. (In Chinese)
90. Chave, J.; Condit, R.; Aguilar, S.; Hernandez, A.; Lao, S.; Perez, R. Error propagation and scaling for tropical forest biomass estimates. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **2004**, *359*, 409–420. [[CrossRef](#)] [[PubMed](#)]

91. Ledo, A.; Cornulier, T.; Illian, J.B.; Iida, Y.; Kassim, A.R.; Burslem, D.F. Re-evaluation of individual diameter: Height allometric models to improve biomass estimation of tropical trees. *Ecol. Appl.* **2016**, *26*, 2374–2380. [[CrossRef](#)] [[PubMed](#)]
92. Réjou-Méchain, M.; Muller-Landau, H.C.; Detto, M.; Thomas, S.C.; Le Toan, T.; Saatchi, S.S.; Barreto-Silva, J.S.; Bourg, N.A.; Bunyavejchewin, S.; Butt, N.A.; et al. Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. *Biogeoscience* **2014**, *11*, 6827–6840. [[CrossRef](#)]
93. Milena, S.; Markku, K. Allometric models for tree volume and total aboveground biomass in a tropical humid forest in Costa Rica. *Biotropica* **2005**, *37*, 2–8.
94. Miyakuni, K.; Heriansyah, I.; Heriyanto, N.M.; Kiyono, Y. Allometric biomass equations, biomass expansion factors and root-to-shoot ratios of planted *Acacia mangium* Willd. Forests in West Java, Indonesia. *J. Forest Plan.* **2004**, *10*, 69–76.
95. Stegen, J.C.; Swenson, N.G.; Valencia, R.; Enquist, B.J.; Thompson, J. Above-ground forest biomass is not consistently related to wood density in tropical forests. *Glob. Ecol. Biogeogr.* **2009**, *18*, 617–625. [[CrossRef](#)]



© 2017 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/>).