

Article

# **Correlating the Horizontal and Vertical Distribution of LiDAR Point Clouds with Components of Biomass in a** *Picea crassifolia* **Forest**

Wang Li<sup>1,2</sup>, Zheng Niu<sup>1</sup>, Shuai Gao<sup>1,\*</sup>, Ni Huang<sup>1</sup> and Hanyue Chen<sup>3</sup>

- <sup>1</sup> The State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, CAS Olympic S&T Park No. 20 Da Tun Road P.O. Box 9718, China; E-Mails: lwwhdz@sina.com (W.L.); niuzheng@radi.ac.cn (Z.N.); huangn@irsa.ac.cn (N.H.)
- <sup>2</sup> University of Chinese Academy of Sciences, Beijing 100039, China
- <sup>3</sup> Fujian Agriculture and Forestry University, Fuzhou 350002, China;
   E-Mail: chenhanyue.420@163.com
- \* Author to whom correspondence should be addressed; E-Mail: gaoshuai@radi.ac.cn; Tel./Fax: +8610-6488-9215.

Received: 18 December 2013; in revised form: 24 May 2014 / Accepted: 24 July 2014 / Published: 5 August 2014

Abstract: Light detection and ranging (LiDAR) has been widely used to estimate forest biomass. In this study, we aim to further explore this capability by correlating horizontal and vertical distribution of LiDAR data with components of biomass in a Picea crassifolia forest. Airborne small footprint full-waveform data were decomposed to acquire higher density point clouds. We calculated LiDAR metrics at the tree level and subplot level and correlated them to biomass components, including branch biomass (BB), leaf biomass (LB) and above-ground biomass (AGB), respectively. A new metric (Horiz<sub>cv</sub>) describing the horizontal distribution of point clouds was proposed. This metric was found to be highly correlated with canopy biomass (BB and LB) at the tree level and subplot level. Correlation between AGB and Horiz<sub>cv</sub> at the subplot level is much lower than that at tree level. AGB for subplot is highly correlated with the mean height metric (H<sub>mean</sub>), canopy cover index (CCI) and the product of them. On the other hand, the relationship between the vertical distribution of LiDAR point and biomass was explored by developing two types of vertical profiles, including LiDAR distribution profiles and a biomass profile. Good relationships were found between these two types of vertical profiles and assessed by Pearson's correlation coefficient (R) and the area of overlap index (AOI). These good correlations possess potential in predicting the vertical distribution of canopy biomass. Overall, it is concluded that not only the vertical, but also the horizontal distribution of LiDAR points should be taken into account in estimating components of biomass by LiDAR.

Keywords: horizontal and vertical distribution; LiDAR; biomass; Picea crassifolia

## 1. Introduction

Biomass is commonly used as an important parameter for the assessment of forest productivity and carbon sequestration rates [1]. It is generally estimated based on ground measurement [2], which is time-consuming and sometimes inefficient. Using remotely sensed data has become a preferable way to estimate the amount and spatial distribution of forest biomass [3]. Biomass shows significant correlation with the structure of forest, such as diameter at breast height (DBH) and tree height. Light detection and ranging (LiDAR), as an active remote sensing technology, exhibits immense potential in detecting the vertical and horizontal structure of forests. Therefore, LiDAR data has been widely used in the estimation of different types of forest biomass at various sites. The types of forest biomass cover tropical rainforest [4–6], deciduous and conifer forest in temperate area [3,7,8], and so on [9,10]. Height metrics, such as mean of height (H<sub>mean</sub>), mode of height (H<sub>mode</sub>) and percentile height, are the most commonly used LiDAR metrics to estimate biomass, because height metrics can be directly measured by LiDAR and can give insights into the vertical structure of individual trees and forest stands. Numerous allometric equations have been developed for different forest types in various areas when using LiDAR metrics to estimate biomass. However, most of the published literature has emphasized the effects of the vertical structure of forest stand on forest biomass. The horizontal structure of forest was described by the canopy cover index (CCI) or penetration index calculated using the simple ratio of canopy return or ground return [4,5,7–9]. Ni-Meister [11] indicated that the gap fraction combined with height metrics, which is closely related to the woody volume, can improve the aboveground biomass estimation at large scales by using large-footprint waveform LiDAR. The woody volume approximately equals the product of basal area and top 10% tree height [11]. Both the gap fraction and the woody volume take the vertical (tree height) and horizontal (basal area) distribution of canopy components into account.

For most studies, biomass was estimated at scales no smaller than plot sizes (approximately 0.01 hectare). Fewer studies estimated the biomass of individual trees [1,2,12,13]. However, calculating individual tree biomass is the first step for calculating biomass at larger scales in most studies. In these studies, height and crown diameter were identified from an interpolated raster surface (canopy height model, CHM) by some individual tree identification algorithms before biomass estimation. Not all of the trees could be identified, because some returns from the inner canopy under the raster surface were missing [14,15]. However, with the success of individual tree identification, it was possible to achieve a high resolution and to precisely map the biomass distribution, which can be used as a reference dataset instead of laborious ground measurements [13].

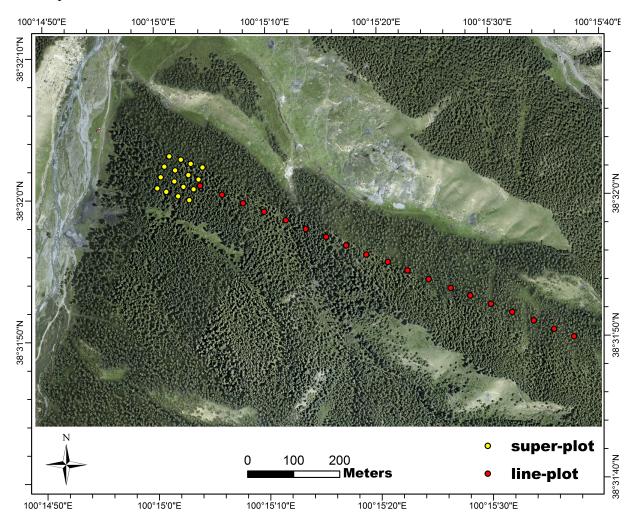
On the other hand, precisely estimated components of forest biomass improve the calculation of carbon storage. Branch biomass (BB) and leaf biomass (LB) are of vital importance to the estimation of canopy fuel load and monitoring of forest fire behavior. They are commonly used in the calculation of canopy fuel parameters, such as canopy base height, canopy bulk density and canopy fuel weight [16,17]. Good correlations were found between species-specific biomass components and LiDAR metrics [18]. Compared with aboveground biomass, canopy biomass, including BB and LB, mostly is distributed within the tree crown. Usually, the horizontal distribution range of the tree crown is comparable to or even wider than its crown depth. Therefore, LiDAR metrics that can reflect the horizontal distribution of the canopy possibly can give an indication on the canopy biomass. For instance, the canopy cover index (CCI) calculated using the simple ratio of canopy return has been found to be sensitive to the degree of canopy openness and shows good correlation with aboveground biomass at the plot level [4,5]. However, the relationship between CCI and canopy biomass has not been explored. To the best of our knowledge, the number of LiDAR metrics, which can give insight into the horizontal structure of the canopy, is quite limited among all of the traditional metrics. In addition, the horizontal distribution of the LiDAR point cloud has not been widely investigated. For instance, when tree heights reach a certain height, the canopy increases the speed of growth in the horizontal direction. This leads to the increment of canopy biomass (BB and LB). Different trees with similar vertical structures may have different horizontal structures. In this case, the variation of the horizontal distribution of LiDAR point clouds could be sensitive to the difference or variation of canopy biomass. We expected that when taking the horizontal characterization into account, the biomass estimating models could be improved by using LiDAR metrics. Therefore, investigation is needed to further explore the links between the distribution (both vertical and horizontal) of the LiDAR point cloud and biomass components.

In this paper, we explored the relationship between the distribution variation of LiDAR point cloud and biomass at the tree level and subplot level. Vertical and horizontal variations of the point distribution were investigated to assess their correlations with biomass of different canopy components. New LiDAR metrics were developed and used to estimate biomass components. The results were compared with those estimated using commonly selected metrics alone, which can be considered to deliver good supplementary information to previous studies in this study area [19,20].

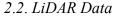
#### 2. Experimental Section

#### 2.1. Study Sites

The study area is a water conservation forest site located in Qilian Mountain Dayekou, Zhangye City, Gansu Province, China (Figure 1). It is a study area chosen by an airborne, satellite-borne and ground-based remote sensing experiment, Watershed Allied Telemetry Experimental Research (WATER) [21]. It covers three vegetative climate zones and significantly affects soil/water conservation and biodiversity protection. The area is the main body of the ecosystem in the Qilian Mountains, and the Hexi Corridor Oasis and is located at an elevation of 2500 m to 3800 m above sea level. The center coordinates of the area are 100°15'E and 38°32'N. The main species in the study area is *P. crassifolia*, occupying about 95% of the whole forestland over the test site [3].



**Figure 1.** Airborne ortho charged coupled device (CCD) image for the study area and the field plots.



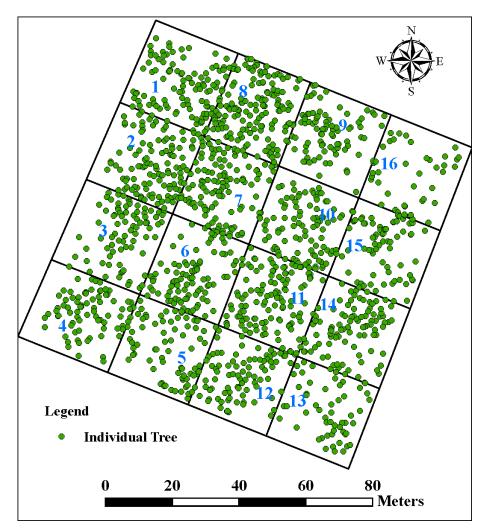
Small footprint full-waveform LiDAR data were acquired for the Dayekou study area on June 23, 2008. The LiteMapper-5600 airborne laser scanning system was used and configured to emit laser pulses of 3.5 ns at 50 kHz in the near-infrared band (wavelength = 1550 nm) with a beam divergence fixed with 0.5 mrad. The average footprint diameter was 38 cm. Geographical coordinates (easting, northing and elevation) were recorded by a dual-frequency differential Global Positioning System with an inertial measurement unit. The horizontal and vertical accuracy of the instrument was 0.1 m and 0.3 m, respectively. Repetitive flights were carried out over the sample study area with a nominal height above ground of 700–800 m [22], leading to a pulse density of approximately 3.43 pts/m<sup>2</sup>. For each pulse, the transmitted pulse and the returned pulse were recorded in the full-waveform data. Point clouds were generated from the full-waveform data by the commercial software provided by the vendor. All of the points were geo-referenced to a projection system of universal transverse Mercator (UTM) Zone 47N/WGS-84. In order to increase the point density in this study, we produced another set of point clouds from the original waveform data by waveform decomposing. This is described in Section 2.5. The point cloud with the relatively lower point density described above was not used in this study.

## 2.3. Field Measurement

1914

During the flight campaign, field measurements were conducted on two field plots distributing over the study area (Figure 1). The field plots were denoted as super-plot and line-plot. The super-plot was a 100 m  $\times$  100 m area with a slope of less than 20 degrees. It was divided into 16 subplots measuring 25 m  $\times$  25 m. The line-plot was a sample line consisting of 19 subplots measuring 20 m  $\times$  20 m distributed along the direction of flight. The distance between the subplots in the line-plot was 50 m. The total number of subplots was 35. The center location of each subplot was positioned using a differential global positioning system (DGPS) station. The DGPS mobile station was placed within the open space of the standing forest to ensure the positioning accuracy. The individual trees were located by the survey total stations with equipment model of TOPCON GTS-602 and TOPCON GPT-7002 (Figure 2). Tree height (H) and first branch height (FBH) were measured using a laser hypsometer; diameter at breast height (DBH) was measured with a tape measure, and crown size (R) in two orthotropic directions for each individual tree was measured using a ribbon tape. High correlations were found between the ground-measured features, as shown in Table 1 and Figure 3.

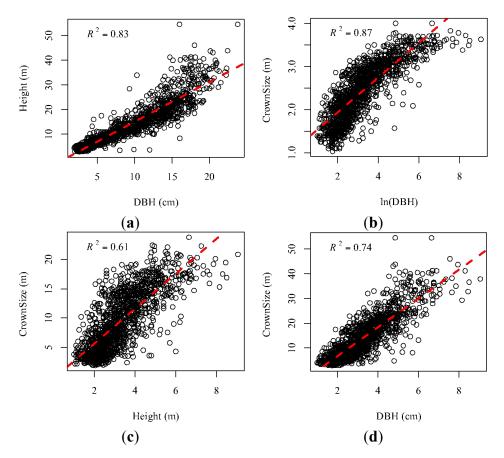
**Figure 2.** The spatial distribution of the field-measured individual trees in the super-plot with 16 subplots.



Dependent Variables (y)	Equation	$R^2$	Root Mean Squared Error (RMSE) (m)
Height	$y = 0.5171 \times \text{DBH} + 2.1268$	0.83	1.24
	$y = 7.3496 \times \ln(\text{DBH}) - 8.581$	0.87	1.18
Crown Size	$y = 0.2014 \times \text{Height} + 1.3742$	0.61	0.77
	$y = 0.1264 \times \text{DBH} + 1.4894$	0.74	0.71

 Table 1. Regression models of different ground-measured features.

**Figure 3.** Scatter plots for the field-measured attributes (**a**–**d**). Height is the tree height; DBH is the diameter at breast height;  $\ln(DBH)$  is the log transformation of DBH;  $R^2$  is the coefficient of determination.



## 2.4. Biomass Calculation

The biomass of different forest components was calculated based on their DBH and height according to the empirical relative growth equations [23] as follows:

stock biomass = 
$$0.0478 \times ((DBH)^2 \times H)^{0.8665}$$
 (1)

branch biomass = 
$$0.0061 \times ((DBH)^2 \times H)^{0.8905}$$
 (2)

leaf biomass = 
$$0.2650 \times ((DBH)^2 \times H)^{0.4701}$$
 (3)

fruit biomass = 
$$0.0342 \times ((DBH)^2 \times H)^{0.5779}$$
 (4)

The aboveground biomass (AGB) of each individual tree was calculated as the sum of biomass of all of the above components (stock, branch, leaf and fruit). Subplot biomass is calculated as the sum of ABG of each individual tree in each subplot.

## 2.5. Processing of LiDAR Data

In order to acquire a higher point density, we used a waveform decomposition approach [24] to decompose the original waveform data into another point cloud dataset. This decomposition approach has been successfully applied to another LiDAR dataset collected in the urban area of Zhangye City around the same point in time during which the WATER experiment was conducted. It approximates the LiDAR waveform as a Gaussian mixture model. A stepwise strategy was used to implement waveform decomposition by defining the maximum iteration number and error expectations before decomposition [24]. The final point density was increased to 6.22 pts/m<sup>2</sup> from 3.43 pts/m<sup>2</sup>. In the next step, ground points were filtered from the decomposed point clouds to generate a digital elevation model (DEM) at 0.5 m resolution. Heights of non-ground points were normalized by calculating the difference between the non-ground point heights and the corresponding interpolated DEM cell heights beneath the points. A digital surface model (DSM) at 0.5 m resolution was generated from the high density point clouds. The maximum height value of the points within each pixel was selected as the value of the interpolated DSM raster cell. Then, a canopy height model (CHM) was derived as the difference between the DEM and DSM.

## 2.6. Individual Tree Identification

The CHM was smoothed by a 3 m  $\times$  3 m Gaussian filter before being used to identify the individual trees. The search radius (3 m) was chosen based on the mean of all of the measured tree crown sizes. Then, the watershed algorithm [25] was applied to the smoothed CHM. The watershed algorithm firstly inverts the CHM to a raster converting peaks into depressions. Then, the canopy basins are delineated on the depression raster by watershed segmentation. Finally, the delineated basins raster is converted to segments. A height break (1.5 m) was set to extract the tree crown segments from the obtained segments. Then the crown segments were transferred into polygonally shaped files in Arcgis9.3. The highest pixel value obtained by zonal statistics within each canopy polygon was assigned as the tree height and its location as the tree location. This automatic delineation greatly helped to extract the individual trees, but manual intervention was still needed to remove some mistakes. Errors of omission and commission were unavoidable in individual tree delineation. It can clearly be seen that there are several field-measured trees clustering within one segment or that no field-measured trees are located in a particular segment. The center geographic location of each segment and its height were matched to the field measured trees accompanied by some visual interpretation. Before matching, visual interpretation and corrections were conducted on the segments containing more than one field-measured tree. Specifically, if the field-measured trees in the segment were too close to separate visually, the segment enclosing them was not selected in order to assure the quality of the crown segment. The difficulty in separating this kind of trees was caused either by tree clustering or the understory trees being sheltered by the tall trees. In the remaining cases, the field-measured trees were separated by splitting the segment manually according to the texture of CHM and the ortho CCD image. On the other hand, segments containing no field-measured trees were not selected. Finally, only the segments with heights and locations closest to the field-measured trees were selected as the final matched crown segments. The quality of tree delineation directly affects the accuracy of LiDAR metrics at tree level in this study. Differences in the locations and heights were limited to certain confidence intervals. Only the final matched crown segments were selected to clip the point clouds to calculate LiDAR metrics at tree level.

#### 2.7. LiDAR Metrics Calculation

The decomposed point clouds within each extracted crown shape files were clipped from the original points. Each point contained attributes, including the geographic coordinates (X/Y), the normalized height (H), returned pulse amplitude, width, standard deviation and returns. The mean height of each individual tree point cloud was calculated using the conventional LiDAR height metrics. The coefficient of height variation (Hcv) is a commonly used LiDAR metric and is used to describe the height variation in the Z-axis [10]. The coefficient of variation (CV) is defined as the ratio of the standard deviation to the mean. The height Z is normalized by subtracting the DEM from DSM. Based on this, we expected to develop a similar, but slightly differing coefficient of variation in the X/Y-axis to describe the horizontal distribution of LiDAR points. Before calculating the new metric, the X/Y coordinates of the point clouds within each tree crown segment were normalized. The original X/Y coordinates were transformed to x/y as follows:

$$x_{i} = (X_{i} - X_{\min}) / (X_{\max} - X_{\min})$$
(5)

$$y_i = (Y_i - Y_{\min}) / (Y_{\max} - Y_{\min})$$
(6)

Then, the coefficient of variation for x/y,  $x_{cv}/y_{cv}$  was calculated:

$$x_{cv} = \sigma_x / \overline{x} \quad \sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2}$$
(7)

$$y_{cv} = \sigma_y / \overline{y} \quad \sigma_y = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \overline{y})^2}$$
(8)

where  $\sigma_x$  and  $\sigma_y$  are the standard deviations for x and y, respectively. Finally, the metric Horiz<sub>cv</sub> was defined as:

$$Horiz_{cv} = x_{cv} \times y_{cv} \tag{9}$$

The normalized x/y coordinates can be seen as the projections of the point clouds in the horizontal plane. They can describe the relative spatial locations of the point clouds within the crown segment in the horizontal plane. The coefficient of x/y variation can reflect how the point cloud distributions vary in the x/y direction. They actually are independent of the absolute locations. Theoretically, this works for tree point clouds (points within the tree crown segment) with coordinates characterized by different coordinate systems. For example, there are two LiDAR point clouds collected from two trees with the exact same structure and biomass located at different geographic positions. The interactions between the LiDAR pulse and the trees are assumed to be the same, and the only difference is the shift in x/y

coordinates. There should be a transformation between the coordinates in the x/y plane. For example, we shift the x/y coordinates using the following formulas:

$$\begin{aligned} x_{shift_i} &= x_i \times a + b \\ y_{shift_i} &= y_i \times a + b \end{aligned} \tag{10}$$

where  $x_{shift_i}$  and  $y_{shift_i}$  are the shifted coordinates, *a* and *b* are the shifting coefficients. The normalized coordinate x/y are the same as those before coordinate shifting, because the normalization process has removed the effect of the absolute locations.

Horiz<sub>cv</sub> for each subplot was calculated by summing all of the Horiz<sub>cv</sub> of the individual trees within the subplot. All of the first-returned echoes within each subplot were selected to obtain the canopy cover index (CCI). CCI was calculated by dividing the number of canopy first-returned echoes (H > 1.5 m) by the total number of first-returned echoes. The mean of the point cloud heights (H<sub>mean</sub>) was also calculated, since biomass is closely related to tree height.

### 2.8. Vertical Biomass Profile Derivation

To assess how well the vertical distribution of point clouds correlated with the vertical distribution of biomass, three vertical profiles were calculated: canopy point density profile (CPD), canopy intensity profile (CIP) and canopy biomass density profile (CBP). The first two profiles are LiDAR-derived, and the third one is derived from ground-measured data. Trees in each subplot were sectioned into a series of small height bins (0.5 m). The number/intensity of LiDAR points in each height bin was summed to obtain CPD/CIP. The canopy biomass profiles (CBP) were derived during the derivation process of canopy bulk density (CBD). CBD is a measurement indicating the distribution of canopy biomass or fuel load. It is the dry weight of available canopy fuel per unit canopy volume [17], and it is a bulk property of the stand, not an individual tree. Canopy bulk density is derived following the steps [17]: (1) calculating the leaf and branch biomass for each tree in the subplot; (2) dividing that biomass equally into a 0.5 m bin from the base of the tree crown to the top of tree; (3) calculating the biomass contributed by each tree in the subplot for each 0.5 m bin (a vertical profile of biomass is created after this step, which we refer to as CBP); and (4) using a running mean to get the maximum value in the vertical profile (CBP), whereby the maximum value is selected as CBD. During the quantification of CBP, it is assumed that canopy biomass is distributed uniformly in the vertical direction within the tree crown [17]. Based on this, the vertical biomass distribution within the subplot or stand is approximately calculated by summing the canopy biomass contributed by each tree in each vertical bin. However, we have to admit that it is just an approximate way to describe the vertical distribution of biomass at the plot level or stand level rather than at the individual tree level. In fact, it is very hard or even unfeasible to directly measure the vertical distribution of tree biomass in the field.

Since height differences occur between LiDAR-derived heights and ground-measured heights, there should be height gaps between the CPD/CIP and CBP. Therefore, we limited the pairs for comparison to a height difference to 2 m in order to prevent potential errors. Different types of profiles possess different magnitudes and units. Before comparing and correlating different pairs of vertical profiles, all of the profiles were normalized by dividing the value for each vertical bin by the sum of those for all of the bins.

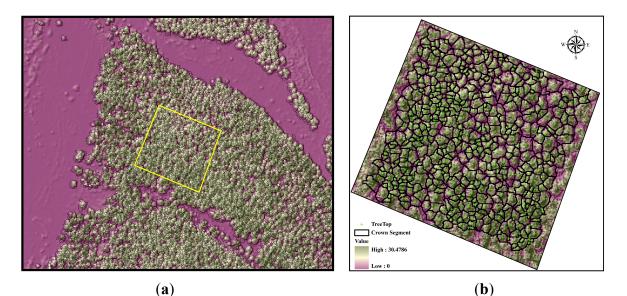
In order to correlate the LiDAR-derived profiles with the ground-measured profiles, we chose Pearson's correlation coefficient (R) and the area of overlap index (AOI) to assess the correlation. Pearson's correlation coefficient indicates how well two profiles match. The area of overlap index is defined as the fraction of area shared by two profiles, and it has been successfully used to compare the waveforms and profile data [5].

## 3. Results and Discussion

# 3.1. Individual Tree Identification

Ground-measured tree *X/Y* locations and their heights were compared to those delineated by the watershed segmentation method, as shown in Figure 4. Trees were matched if the difference between the coordinates and the difference between the tree heights were within certain confidence intervals, accompanied by visual interpretations. Errors of omission and commission for individual trees occurred as in the other studies [14,26,27]. These errors are highly related to the spatial distribution of the trees. Omission error is likely to occur when the heights of some trees are below the dominant canopy nearby [27]. Another important reason for errors of commission is that the only forest species, *P. crassifolia*, in this study is likely to cluster. In this study, we finally obtained 561 matched trees out of 1468 (38.22%) field-measured trees in the super-plot, as shown in Table 2. This ratio (38.22%) was higher than the 15.75% (117/743) obtained in [13] and lower than the 48% (84/175) obtained in [26]. The statistics of different extracted tree features are shown in Table 3.

**Figure 4.** Canopy height model overlaid with tree crown segments derived from watershed segmentation: (a) near the super-plot (within yellow boundary); (b) within the super-plot together with measured tree tops.



Subplot	Total Segments	Matched Segments	Mismatched Segments	Field Measured	Percentage (%)
1	53	38	15	98	38.78
2	63	46	17	114	40.35
3	29	22	7	77	28.57
4	54	39	15	80	48.75
5	50	38	12	70	54.29
6	31	24	7	91	26.37
7	68	55	13	120	45.83
8	92	75	17	156	48.08
9	27	20	7	87	22.99
10	28	20	8	100	20.00
11	55	43	12	111	38.74
12	49	39	10	94	41.49
13	38	27	11	62	43.55
14	45	33	12	94	35.11
15	37	28	9	77	36.36
16	18	14	4	37	37.84
Total	737	561	176	1,468	38.22

Table 2. Statistics of watershed segmentation results for the 16 subplots in a super-plot.

**Table 3.** Statistics of light detection and ranging (LiDAR) extracted tree features matched to those from the ground measured trees.

Statistics\Error	X (m)	Y (m)	H (m)	<b>D</b> (m)
Mean (abs)	0.28	0.30	1.02	0.46
SD (abs)	0.38	0.41	1.19	0.53
Number of Trees		561 (38	8.22%)	

Note: X = error of X-coordinate; Y = error of Y-coordinate; H = error of tree height; D = horizontal distance between LiDAR extracted tree and ground-measured tree; Mean = mean of errors; SD = standard deviation of errors; abs = absolute value.

#### 3.2. Correlating Horizontal Variation of Point Clouds with Biomass

Linear regression models were developed to indicate the correlation between the group of LiDAR metrics and the forest components of biomass. The regression models were evaluated using the coefficient of determination ( $R^2$ ), root mean squared error (RMSE) and relative root mean squared error (rRMSE).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p_i - \hat{p}_i)^2}$$
(11)

$$rRMSE = \frac{RMSE}{\stackrel{\wedge}{p_i}}$$
(12)

where  $p_i$  is the observed value and  $p_i$  is the predicted value. The value of RMSE is related to the magnitude of the observed variables. Therefore, we also selected rRMSE, a relative value, to compare

the performance of different regression models [18]. A lower rRMSE value often indicates a better regression performance. We first explored the relationships between the conventional height metrics and branch biomass (BB), leaf biomass (LB) and aboveground biomass (AGB) at the individual tree level, respectively. Results (Figure 5a-c) show that biomass for each component of individual trees initially increases with the increasing mean point heights (Hmean) and is saturated when Hmean continues increasing. This means that when the tree height is comparably low, tree height contributes greatly to the amount of biomass. Whereas when tree heights reach a certain height, the canopy increases the speed of growth in the horizontal direction, which also increases the canopy biomass (branch and leaf). Therefore, the horizontal distribution of branches and leaves directly affects the amount of canopy biomass. High correlations were found between the new metric (Horizcv) and the biomass components with high  $R^2$  and low RMSE and rRMSE, as shown in Figure 5 and Table 4. Horiz<sub>cv</sub> and AGB for the tree show a good relationship with  $R^2$  equal to 0.72, but a higher rRMSE (0.53) than that of BB (0.36) and LB (0.40). Specifically, a higher value of Horiz<sub>cv</sub> means more dispersion of the LiDAR points in the horizontal XY plane. Since branches and leaves rarely grow in only one direction, a greater dispersion probably reflects a wider horizontal distance between the branch and leaf growth and a larger weight of canopy components. Combined metrics were developed by combining the horizontal metrics (CCI and Horiz<sub>ev</sub>) with the vertical metric (H<sub>mean</sub>). Both CCI and Horiz<sub>ev</sub> can give an insight into the horizontal distribution of point clouds. Correlations were improved when using the combined metrics (H<sub>mean</sub>\*CCI and Horiz<sub>cv</sub>\*CCI) compared to when using the single metric alone, as shown in Table 4 and Figure 6a-f. The RMSEs and rRMSEs also decreased. Although improvements for the combined metric Horizev\*CCI is smaller than that of Hmean\*CCI, the regression models for Horizev\*CCI obtained higher  $R^2$  and lower RMSEs and rRMSEs. This is due to the high correlation between Horiz<sub>cv</sub> and the biomass components. However, no significant improvements were obtained when combining H<sub>mean</sub>, Horiz<sub>cv</sub> and CCI, as shown in Table 4.

Tree	Metrics	<b>Regression Equations</b>	$R^2$	RMSE	rRMSE
	H <sub>mean</sub>	$y = 0.0054 e^{0.6842x}$	0.45	9.78	0.97
	CCI	y = 45.674x - 5.7033	0.53	8.02	0.81
BB	H <sub>mean</sub> *CCI	y = 4.114x - 4.3411	0.59	7.47	0.76
DD	Horiz <sub>cv</sub>	y = 69.523x - 0.3626	0.73	6.11	0.36
	$H_{mean}$ *Horiz <sub>cv</sub>	y = 6.1408x + 0.3972	0.73	6.07	0.59
	H <sub>mean</sub> *Horiz <sub>cv</sub> *CCI	y = 0.9624x + 3.7116	0.68	6.64	0.67
	H <sub>mean</sub>	$y = 0.2487 e^{0.3612x}$	0.45	5.94	0.54
LB	CCI	y = 31.734x + 0.242	0.52	5.26	0.48
	H <sub>mean</sub> *CCI	y = 2.8721x + 1.1408	0.60	4.82	0.49
	Horiz <sub>cv</sub>	y = 43.369x + 4.5629	0.67	4.39	0.40
	$H_{mean}$ *Horiz <sub>cv</sub>	y = 3.8244x + 5.0467	0.67	4.36	0.38
	H <sub>mean</sub> *Horiz <sub>cv</sub> *CCI	y = 0.6429x + 6.9487	0.63	5.14	0.52

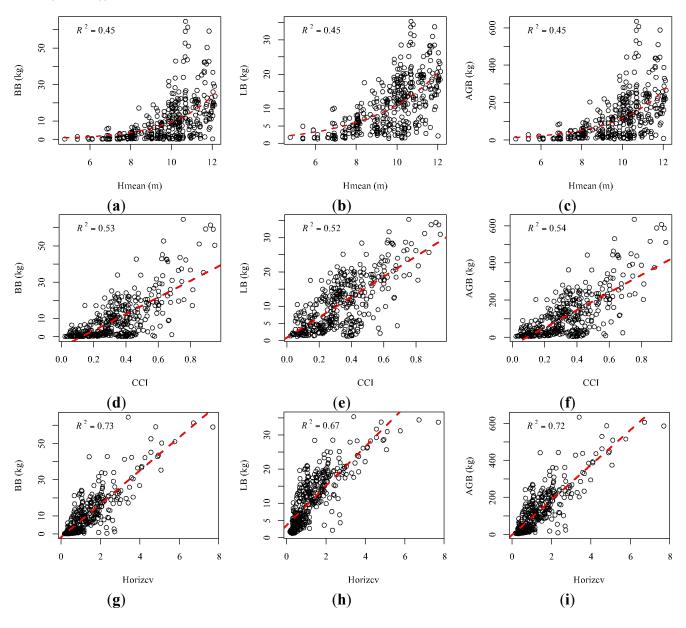
**Table 4.** Regression models developed from biomass components and LiDAR metrics at the tree level.

Tree	Metrics	<b>Regression Equations</b>	$R^2$	RMSE	rRMSE
	H <sub>mean</sub>	$y = 0.3638e^{0.5356x}$	0.45	98.43	0.82
	CCI	y = 472.47x - 42.536	0.54	81.81	0.69
AGB	H <sub>mean</sub> *CCI	y = 42.693x - 28.919	0.61	75.65	0.64
AGD	Horiz <sub>ev</sub>	y = 711.61x + 13.855	0.72	63.43	0.53
	$H_{mean}$ *Horiz <sub>cv</sub>	y = 62.835x + 21.663	0.72	62.86	0.50
	H <sub>mean</sub> *Horiz <sub>cv</sub> *CCI	y = 9.689x + 56.558	0.65	71.07	0.59

Table 4. Cont.

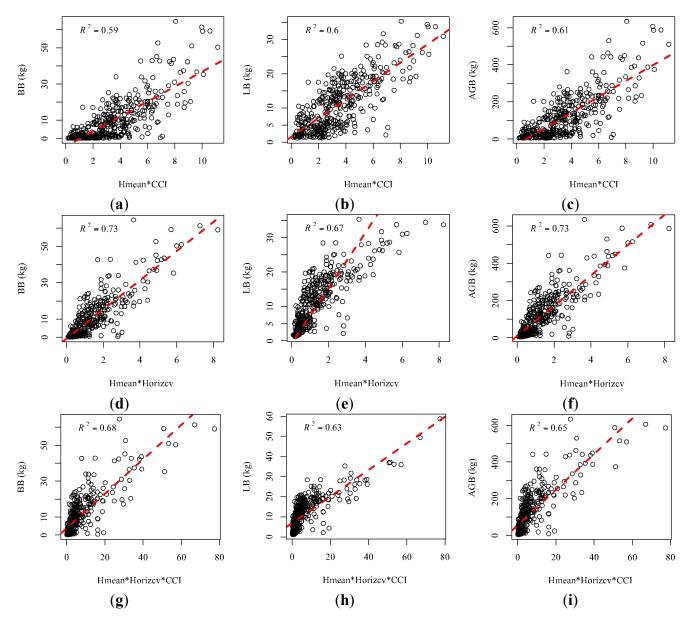
Note: x = independent variable (LiDAR metrics); y = dependent variable (biomass); BB = branch biomass; LB = leaf biomass; AGB = aboveground biomass; CCI = canopy cover index;  $R^2$  = coefficient of determination; RMSE = root mean squared error; rRMSE = relative root mean squared error.

Figure 5. Scatter plots for biomass components (branch biomass (BB) ( $\mathbf{a}$ ,  $\mathbf{d}$ ,  $\mathbf{g}$ ), leaf biomass (LB) ( $\mathbf{b}$ ,  $\mathbf{e}$ ,  $\mathbf{h}$ ) and aboveground biomass (AGB) ( $\mathbf{c}$ ,  $\mathbf{f}$ ,  $\mathbf{i}$ )) and LiDAR metrics (mean height ( $H_{mean}$ ) canopy cover index (CCI) and horizontal distribution of point clouds (Horiz<sub>cv</sub>)) at the individual tree level.



In this study, the accuracy of LiDAR metrics calculated at the tree level highly depends on the quality of individual tree delineation. We only selected the crown segments with a high delineation quality to clip the point cloud to calculate the metrics at tree level in order to reduce the influence from error of tree delineation. Therefore, the LiDAR metrics at tree level were based on the degree of quality of the (selected) crown delineation. Based on this, an irregular delineated crown segment indicated an irregular canopy shape. Vegetation hits from different parts of the crown segment were dispersed, leading to a relatively high coefficient of variation. The more regular the canopy shape, the lower the coefficient of variation for the point cloud within the crown segment.

**Figure 6.** Scatter plots for biomass components (BB ( $\mathbf{a}$ ,  $\mathbf{d}$ ,  $\mathbf{g}$ ), LB ( $\mathbf{b}$ ,  $\mathbf{e}$ ,  $\mathbf{h}$ ) and AGB ( $\mathbf{c}$ ,  $\mathbf{f}$ ,  $\mathbf{i}$ )) and the combined LiDAR metrics ( $H_{mean}*CCI$ ,  $H_{mean}*Horiz_{cv}$  and  $H_{mean}*Horiz_{cv}*CCI$ ) at the individual tree level.



By summing the biomass of different forest components for each individual tree, three types of biomass (BB, LB and AGB) were calculated at the subplot level. Their relationship with LiDAR metrics was explored as shown in Table 5. Results indicate that the mean height ( $H_{mean}$ ) shows good correlation with the three types of biomass, especially with AGB ( $R^2 = 0.65$ ). The canopy cover index (CCI) shows a good relationship with AGB ( $R^2 = 0.60$ ). When considering the product of  $H_{mean}$  and CCI as one metric,  $R^2$  for the regression model developed from this combined metric and AGB improved ( $R^2 = 0.71$ ), and RMSE and rRMSE decreased. This improvement was a consequence of taking both horizontal and vertical effects of the forest structure on AGB into account, which, in the end, agreed with Ni-Meister's study [11].

Subplot	Metrics	<b>Regression Equations</b>	$\mathbf{R}^2$	RMSE	rRMSE
	H <sub>mean</sub>	y = 101.71x - 426.78	0.49	178.86	0.31
	CCI	y = 1,188.4x - 411.79	0.35	202.00	0.35
BB	H <sub>mean</sub> *CCI	y = 74.479x - 44.884	0.48	179.27	0.32
DD	Horiz <sub>cv</sub>	y = 76.639x + 7.0569	0.75	126.17	0.22
	$H_{mean}$ *Horiz <sub>cv</sub>	y = 6.4054x + 95.374	0.84	99.76	0.18
	H <sub>mean</sub> *Horiz <sub>cv</sub> *CCI	y = 6.4395x + 160.2	0.83	103.42	0.18
	H <sub>mean</sub>	y = 70.079x - 62.366	0.42	227.66	0.37
	CCI	y = 1,034.2x - 229.79	0.25	223.4	0.36
LB	H <sub>mean</sub> *CCI	y = 56.211x + 160.45	0.26	221.39	0.36
LD	Horiz <sub>cv</sub>	y = 84.016x + 7.792	0.84	101.92	0.16
	$H_{mean}$ *Horiz <sub>cv</sub>	y = 6.4469x + 147.11	0.8	114.54	0.18
	Hmean*Horizcv*CCI	y = 6.4181x + 216.37	0.78	122.02	0.2
	H <sub>mean</sub>	y = 523.44x - 1,120.8	0.65	665.19	0.17
	CCI	y = 6,984.2x - 1,760.1	0.6	709.83	0.18
AGB	H <sub>mean</sub> *CCI	y = 404.22x + 672.21	0.71	597.16	0.15
AGD	Horiz <sub>cv</sub>	y = 238.08x + 2,257.8	0.36	894.31	0.22
	$H_{mean}$ *Horiz <sub>cv</sub>	y = 22.58x + 2,334	0.52	772.42	0.19
	H <sub>mean</sub> *Horiz <sub>cv</sub> *CCI	y = 24.377x + 2,456.1	0.59	712.3	0.18

**Table 5.** Regression models developed from biomass components and LiDAR metrics at the subplot level.

The accumulated Horiz<sub>cv</sub> for the subplot still showed significant correlation with the branch biomass and leaf biomass, while not obviously with AGB. The combined metric  $H_{mean}$ \*Horiz<sub>cv</sub> even improved the performance of the BB prediction model with  $R^2$  increasing from 0.75 to 0.84, RMSE decreasing from 126.17 to 99.76 and rRMSE decreasing from 0.22 to 0.18. This indicates that the horizontal distribution of branches and leaves still greatly affects the canopy biomass (BB, LB) at the subplot level. The poor correlation between Horiz<sub>cv</sub> and AGB at the subplot level might be due to the following: (1) individual trees were divided into several groups according to the area of extent of each subplot (Horiz<sub>cv</sub> for each plot is calculated as the sum of those area extents of the trees within the subplot); (2) the total biomass of the whole tree (AGB) not only benefited from the contribution of the canopy components (branches and leaves), but also from that of other tree components, such as stock, root and fruit; and (3) a vital important structure attribute has been out of discussion when we have assessed the effects from the horizontal and vertical structure variations on biomass: the diameter at

breast height (DBH). According to the relative growth equation, DBH greatly contributes to the estimation of AGB. However, up to now, airborne LiDAR still cannot directly measure the size of DBH. Finally, for the calculation of Horiz<sub>cv</sub>, either at the tree level or subplot level, the influence of the LiDAR data should be considered, for instance the influence of the LiDAR flight properties, *i.e.*, overlapping strips, different flying heights, scan pattern, scan angles, *etc*.

## 3.3. Correlating Vertical Distribution of Point Clouds with Biomass

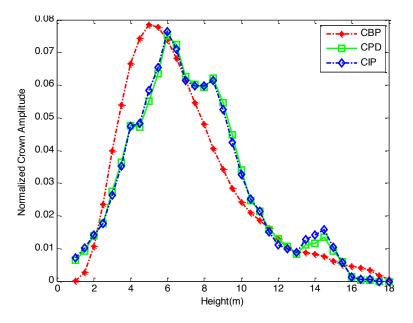
Table 6 shows the correlation between two groups of vertical profiles for each subplot in the study area. The Pearson's correlation coefficient (R) values for the correlation between CBP and CPD ranged between 0.39 and 0.93 (mean = 0.71, SD = 0.15), and 28 pairs of 35 had values  $\geq 0.6$ . For pairs of CBP and CID, R varied from 0.19 to 0.92 (mean = 0.62, SD = 0.19), and 21 pairs of 35 obtain values ≥0.6. On the other hand, AOI values for pairs of CBP and CPD varied from 0.50 to 0.84 (mean = 0.64, SD = 0.08), and 21 pairs of 35 had values  $\geq 0.6$ . CIP played a weaker role than CPD: only 17 of 35 CBP/CIP pairs had values above 0.6 (mean = 0.61, SD = 0.09). The results of the correlations indicate that CBP representing the vertical distribution of biomass is highly correlated with the vertical distribution of LiDAR points. It is likely that both CPD and CIP contain the information of tree height. For the two LiDAR-derived profiles, CPD performs better than CIP. CPDs and CIPs for all of the subplots show very similar curve shapes. The profiles for Subplot S7 are shown in Figure 7. The similarity in the curve shapes for all of the subplots is mainly attributed to the similar forest structure in this study area. According to the previous studies [28], the Picea crassifolia forest in this study area consists primarily of half-mature and mature stands, and the complexity of the canopy vertical distribution and the differentiation grade of canopy height were low. In order to increase the usefulness of the outcome, the approaches should be tested in different forest conditions in the future. The distances between the peaks of CPD (CIP) and CBP for different subplots differ. The height at which these two pairs of profiles start to overlap relates to the stem density and the canopy structure of the Picea crassifolia in the subplot. Specifically, the higher the stem density, the more LiDAR pulses are reflected by the canopy and the more precisely the CPD represents the distribution of the canopy components. In addition, when branches and leaves within the canopy clump at a certain height, the peak of accumulated biomass for all of the height bins and the peak of the LiDAR point density are very likely to overlap around this height. Calculating the vertical biomass profiles is one of the steps to calculate the canopy fuel parameter, canopy bulk density (CBD). It is assumed that the canopy biomass is distributed uniformly in the vertical direction within the tree crown [17]. Based on this, the vertical biomass distribution within the subplot or stand is approximately calculated by summing the canopy biomass contributed by each tree in each vertical bin. However, we have to admit that it is just an approximate way to describe the vertical distribution of biomass at the plot level or stand level, rather than at the individual tree level. According to [17], in fact, canopy fuels have never been measured directly. However, we believe that the potential of predicting CBD, especially when using the vertical detecting capacity of LiDAR, deserves to be tested.

	Pearson's Correlation Coefficient (R)		Area of Overlap Index (AOI)		
Plot	CBP/CPD CBP/CIP		CBP/CPD	CBP/CIP	
S1	0.67	0.65	0.59	0.59	
S2	0.87	0.86	0.69	0.67	
S3	0.86	0.84	0.66	0.64	
S4	0.82	0.81	0.58	0.58	
S5	0.79	0.58	0.66	0.61	
S6	0.8	0.78	0.57	0.54	
S7	0.91	0.92	0.77	0.77	
<b>S</b> 8	0.8	0.74	0.62	0.58	
S9	0.72	0.7	0.68	0.68	
S10	0.83	0.71	0.79	0.77	
S11	0.67	0.66	0.55	0.54	
S12	0.69	0.66	0.57	0.55	
S13	0.78	0.73	0.6	0.58	
S14	0.62	0.56	0.6	0.57	
S15	0.39	0.19	0.59	0.54	
L1	0.77	0.74	0.62	0.6	
L2	0.4	0.31	0.56	0.53	
L3	0.75	0.64	0.72	0.69	
L4	0.76	0.7	0.59	0.56	
L5	0.79	0.58	0.76	0.7	
L6	0.46	0.33	0.56	0.47	
L7	0.84	0.82	0.67	0.66	
L8	0.93	0.91	0.75	0.74	
L9	0.74	0.55	0.62	0.56	
L10	0.88	0.81	0.75	0.71	
L11	0.45	0.27	0.66	0.61	
L12	0.73	0.68	0.56	0.54	
L13	0.49	0.38	0.5	0.45	
L14	0.4	0.32	0.54	0.51	
L15	0.56	0.44	0.58	0.53	
L16	0.71	0.38	0.84	0.77	
L17	0.86	0.81	0.77	0.75	
L18	0.66	0.6	0.65	0.62	
L19	0.73	0.61	0.65	0.62	
L20	0.71	0.6	0.71	0.66	
Mean	0.71	0.62	0.64	0.61	
SD	0.15	0.19	0.08	0.09	

Table 6. Correlation between different pairs of vertical profiles.

Note: S = subplot in super-plot; L=subplot in line-plot; CPD = canopy point density profile; CBP = canopy biomass profile; CIP = canopy intensity profile; Mean = mean of  $R^2$  or AOI for all of the plots; SD = standard deviation; numbers in bold represent the maximum or minimum values.

**Figure 7.** Vertical profiles for Subplot S7. CBP = canopy biomass profile; CPD = canopy point density profile; CIP = canopy intensity profile; normalized crown amplitude means the normalized values for CBP, CBD and CIP.



## 4. Conclusions

This paper tried to explore the correlation between the distribution of LiDAR point clouds and the biomass of different forest components. A new metric (Horiz<sub>cv</sub>) describing the horizontal variation of LiDAR points was proposed and proven to be highly correlated with the canopy branch biomass and leaf biomass at both the individual tree and the subplot levels. Aboveground biomass (AGB) at the subplot level was more strongly correlated with the conventional LiDAR metrics (H<sub>mean</sub>, CCI). The correlation was improved by combining these two metrics (H<sub>mean</sub>\*CCI) into one in order to predict the biomass components. On the other hand, CPDs and CIPs can be good curves to represent the vertical distribution of LiDAR points. Good relationships between the LiDAR-derived profiles (CPD, CIP) and the ground-measured-derived biomass profile (CBP) were found. These good correlations can be used to predict the vertical distribution of the canopy biomass. If this prediction can be put into practice, a new and promising way to obtain canopy bulk density (CBD) will be available, which is of critical significance for forest fire management. Above all, this study demonstrates that when we estimate forest biomass by using LiDAR metrics, not only the vertical, but also the horizontal distribution of LiDAR points should be taken into account.

### Acknowledgments

This research was funded by China's Special Funds for Major State Basic Research Project (2013CB733405, 2010CB950603), and the project was supported by the National Natural Science Foundation of China (41201345). We thank all of the people working on the projects, especially for their efforts in the WATER experiment. We are deeply grateful to the four anonymous reviewers for the very constructive and insightful comments that greatly improved this paper.

## **Author Contributions**

Wang Li was the lead author of the article and made the analyses with Zheng Niu and Shuai Gao. Ni Huang contributed expertise in statistical analysis and spatial analysis. Hanyue Chen provided expertise in forest structure and management. The article was improved by the contributions of all of the co-authors at various stages of the analysis and writing process.

# **Conflicts of Interest**

The authors declare no conflict of interest.

# References

- 1. Allouis, T.; Durrieu, S.; Vega, C.; Couteron, P. Stem Volume and Above-Ground Biomass Estimation of Individual Pine Trees From LiDAR Data: Contribution of Full-Waveform Signals. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 924–934.
- 2. Bortolot, Z.J.; Wynne, R.H. Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. *ISPRS J. Photogramm. Remote Sens.* 2005, *59*, 342–360.
- He, Q.-S.; Cao, C.-X.; Chen, E.-X.; Sun, G.-Q.; Ling, F.-L.; Pang, Y.; Zhang, H.; Ni, W.-J.; Xu, M.; Li, Z.-Y.; *et al.* Forest stand biomass estimation using ALOS PALSAR data based on LiDAR-derived prior knowledge in the Qilian Mountain, western China. *Int. J. Remote Sens.* 2012, 33, 710–729.
- 4. Drake, J.B.; Dubayah, R.O.; Clark, D.B.; Knox, R.G.; Blair, J.B.; Hofton, M.A.; Chazdon, R.L.; Weishampel, J.F.; Prince, S.D. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sens. Environ.* **2002**, *79*, 305–319.
- 5. Drake, J.B.; Dubayah, R.O.; Knox, R.G.; Clark, D.B.; Blair, J.B. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sens. Environ.* **2002**, *81*, 378–392.
- 6. Clark, M.L.; Roberts, D.A.; Ewel, J.J.; Clark, D.B. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sens. Environ.* **2011**, *115*, 2931–2942.
- 7. Lefsky, M.A.; Cohen, W.B.; Harding, D.J.; Parker, G.G.; Acker, S.A.; Gower, S.T. Lidar remote sensing of above-ground biomass in three biomes. *Glob. Ecol. Biogeogr.* **2002**, *11*, 393–399.
- 8. Ahmed, R.; Siqueira, P.; Hensley, S. A study of forest biomass estimates from lidar in the northern temperate forests of New England. *Remote Sens. Environ.* **2013**, *130*, 121–135.
- 9. Huang, W.; Sun, G.; Dubayah, R.; Cook, B.; Montesano, P.; Ni, W.; Zhang, Z. Mapping biomass change after forest disturbance: Applying LiDAR footprint-derived models at key map scales. *Remote Sens. Environ.* **2013**, *134*, 319–332.
- Næsset, E.; Bollandsås, O.M.; Gobakken, T.; Gregoire, T.G.; Ståhl, G. Model-assisted estimation of change in forest biomass over an 11year period in a sample survey supported by airborne LiDAR: A case study with post-stratification to provide "activity data". *Remote Sens. Environ.* 2013, *128*, 299–314.

- Ni-Meister, W.; Lee, S.; Strahler, A.H.; Woodcock, C.E.; Schaaf, C.; Yao, T.; Ranson, K.J.; Sun, G.; Blair, J.B. Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *J. Geophys. Res.* 2010, 115.
- 12. Popescu, S.C. Estimating biomass of individual pine trees using airborne lidar. *Biomass. Bioenerg.* 2007, *31*, 646–655.
- 13. Zhao, K.; Popescu, S.; Nelson, R. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sens Environ*. 2009, *113*, 182–196.
- Popescu, S.C.; Wynne, R.H.; Nelson, R.F. Estimating plot-level tree heights with lidar: Local filtering with a canopy-height based variable window size. *Comput. Electron. Agrc.* 2002, *37*, 71–95.
- 15. Barraud, J. The use of watershed segmentation and GIS software for textural analysis of thin sections. *J. Volcanol. Geoth. Res.* **2006**, *154*, 17–33.
- 16. Andersen, H.-E.; McGaughey, R.J.; Reutebuch, S.E. Estimating forest canopy fuel parameters using LIDAR data. *Remote Sens. Environ.* 2005, *94*, 441–449.
- Scott, J.H.R.; Elizabeth, D. Assessing Crown Fire Potential by Linking Models of Surface and Crown Fire Behavior. Res. Pap. RMRS-RP-29; U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2001; p. 59.
- Pang, Y.; Li, Z.-Y. Inversion of biomass components of the temperate forest using airborne Lidar technology in Xiaoxing'an Mountains, Northeastern of China. *Chin. J. Plant Ecol.* 2013, *36*, 1095–1105.
- 19. He, Q.; Chen, E.; An, R.; Li, Y. Above-Ground Biomass and Biomass Components Estimation Using LiDAR Data in a Coniferous Forest. *Forests* **2013**, *4*, 984–1002.
- Tian, X.; Su, Z.; Chen, E.; Li, Z.; van der Tol, C.; Guo, J.; He, Q. Estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold and arid area. *Int. J. Appl. Earth Observ. Geoinf.* 2012, 14, 160–168.
- Li, X.; Li, X.; Li, Z.; Ma, M.; Wang, J.; Xiao, Q.; Liu, Q.; Che, T.; Chen, E.; Yan, G.; et al. Watershed Allied Telemetry Experimental Research. J. Geophys. Res. 2009, 114, doi:10.1029/ 2008JD011590.
- Cao, C.; Bao, Y.; Xu, M.; Chen, W.; Zhang, H.; He, Q.; Li, Z.; Guo, H.; Li, J.; Li, X.; *et al.* Retrieval of forest canopy attributes based on a geometric-optical model using airborne LiDAR and optical remote-sensing data. *Int. J. Remote Sens.* **2012**, *33*, 692–709.
- 23. Wang, J.Y.; Ju, K.J.; Fu, H.E.; Chang, X.X.; He, H.Y. Study on biomass of water conservation forest on North Slope of Qilian Mountains. *J. Fujian Coll. For.* **1998**, *18*, 319–323.
- 24. Qin, Y.; Li, B.; Niu, Z.; Huang, W.; Wang, C. Stepwise decomposition and relative radiometric normalization for small footprint LiDAR waveform. *China Ser. D* **2010**, *54*, 625–630.
- 25. Vincent, L.; Soille, P. Watersheds on digital spaces-An efficient algorithm based on immersion simulations. *IEEE Trans. Patt. Anal.* **1991**, *13*, 583–598.
- 26. Liu, Q.; Li, Z.; Chen, E.; Pang, Y.; Wu, H. Extracting height and crown of individual tree using airborne LIDAR data. *J. Beijing For. Univ.* **2008**, *30*, 83–89.
- 27. Edson, C.; Wing, M.G. Airborne Light Detection and Ranging (LiDAR) for Individual Tree Stem Location, Height, and Biomass Measurements. *Remote Sens.* **2011**, *3*, 2494–2528.

28. Li, W.; Niu, Z.; Gao, S.; Qin, Y.C. Analyzing and retrieving structure information of Picea crassifolia based on airborne light detection and ranging data. *J. Remote Sens.* **2013**, *6*, 1612–1626.

 $\bigcirc$  2014 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 license (http://creativecommons.org/licenses/by/3.0/).