

Article

## Forest Canopy LAI and Vertical FAVD Profile Inversion from Airborne Full-Waveform LiDAR Data Based on a Radiative Transfer Model

Han Ma <sup>1,2,3</sup>, Jinling Song <sup>1,2,\*</sup> and Jindi Wang <sup>1,2</sup>

<sup>1</sup> State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China; E-Mail: mahan@mail.bnu.edu.cn (H.M.); wangjd@bnu.edu.cn (J.W.)

<sup>2</sup> School of Geography, Beijing Normal University, Beijing 100875, China

<sup>3</sup> College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China

\* Author to whom correspondence should be addressed; E-Mail: songjl@bnu.edu.cn; Tel.: +86-10-5880-5452; Fax: 86-10-5880-5274.

Academic Editors: Lars T. Waser and Prasad S. Thenkabail

Received: 7 October 2014 / Accepted: 29 January 2015 / Published: 9 February 2015

---

**Abstract:** Forest canopy leaf area index (LAI) is a critical variable for the modeling of climates and ecosystems over both regional and global scales. This paper proposes a physically based method to retrieve LAI and foliage area volume density (FAVD) profile directly from full-waveform Light Detection And Ranging (LiDAR) data using a radiative transfer (RT) model. First, a physical interaction model between LiDAR and a forest scene was built on the basis of radiative transfer theories. Next, FAVD profile of each laser shot of full-waveform LiDAR was inverted using the physical model. In addition, the missing LiDAR data, caused by high-density forest and LiDAR system limitations, were filled in based on the inverted FAVD and the ancillary CHM data. Finally, LAI of the study area was retrieved from the inverted FAVD at a 10-m resolution. CHM derived LAI based on the Beer-Lambert law was compared with the LAI derived from full-waveform data. Also, we compared the results with the field measured LAI. The values of correlation coefficient  $r$  and RMSE of the estimated LAI were 0.73 and 0.67, respectively. The results indicate that full-waveform LiDAR data is a reliable data source and represent a useful tool for retrieving forest LAI.

**Keywords:** full-waveform LiDAR; LAI; FAVD; radiative transfer model

---

## 1. Introduction

Forest canopy leaf area index (LAI) is a critical variable in modeling climates and ecosystems and is required by many terrestrial ecosystem models [1]. LAI of high accuracy will promote accurate modeling of energy, carbon, water, and climate [2,3]. Remotely estimating forest structure potentially provides an ecologically significant advance over laborious ground-based estimation methods and has become widely used to characterize forest structure [4]. LAI is often retrieved from passive optical remote sensing, and there are two primary types of LAI inversion methods. The first is based on developing the empirical or semi-empirical relationship between the vegetation indices and the LAI, and the other is based on radiation transfer models using a look-up table (LUT) or a neural network (NN) [5], or an optimization method such as Levenberg–Marquardt algorithm [6]. However, estimated LAI values tend to saturate in areas of dense vegetation and high biomass.

Light Detection And Ranging (LiDAR), which is rapidly emerging as an active remote sensing technique, can provide direct measurement of the vertical structure of canopy that traditional optical data cannot, and has been successfully used to estimate the canopy height, LAI, biomass, and other variables [7–11] and is shown to be capable of measuring dense canopies [12].

Normally, LiDAR systems are classified as discrete-return or full-waveform depending on their analysis of the reflected signal. Many discrete-return systems capture between two and five values for every return laser pulse [13]. Full-waveform LiDAR provides more information regarding the structure and physical properties of the illuminated surface than discrete-return LiDAR. In the past, the full waveform data were often decomposed to produce 3D point clouds. Hofton *et al.* [14] decomposed the LVIS waveform into a series of Gaussian components using a nonnegative least-squares method (LSM), and Persson *et al.* [15] processed the Gaussian decomposition based on the expectation maximization (EM) algorithm for the TopEye Mark II system. Wagner *et al.* [16] tested LMS-Q560 data and demonstrated that the decomposition was successful for approximately 98% of waveform profiles. These clouds are subsequently used as the discrete data [17], or by building regression models using the calculated waveform metrics with field measurements to estimate forest parameters [18–20]. However, neither approach uses all of the information provided in the waveform data. Adams *et al.* [21] explored whether the waveform metrics of Gaussian peak height, half-height width, and exponential decay rate can distinguish the return types of foliage, understory, and ground. However, the result was that these three metrics could not determine the return type. The metrics alone did not bear a direct relationship to foliage density, even though foliage density correlated well with the decay rate using a simple model based on Beer-Lambert law.

The interaction between LiDAR and vegetation should be explored further to make better use of the waveform information. Several studies have been performed on modeling waveforms in forest areas, which is particularly difficult due to the complexity of the internal structure of the trees [22].

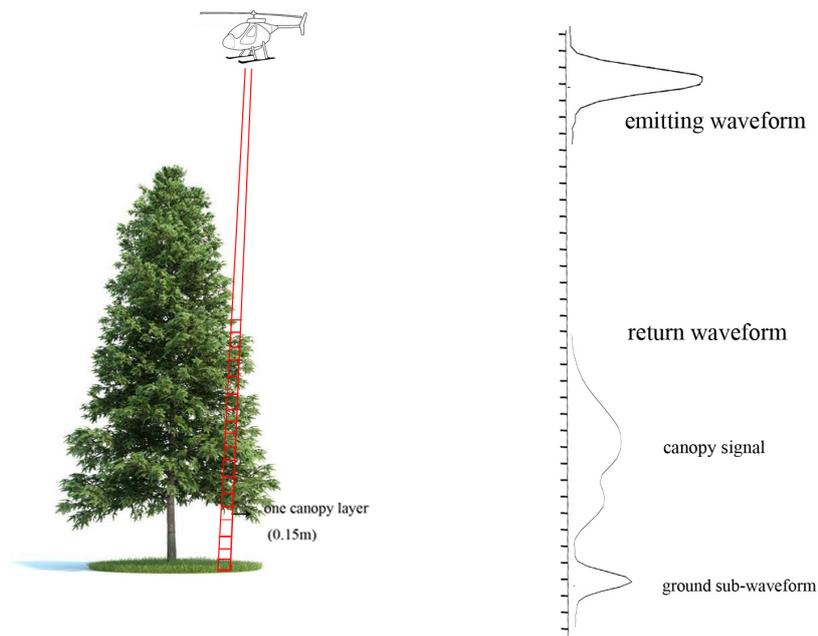
Blair and Hofton [23] first simulated the return waveform under the following assumption: the waveform represents the vertical distribution of the intercepted surfaces with individual backscattering

characteristics but with equal reflectivity, which is typical for dense forests, and their model is based on aggregating discrete return data rather than using an RT model. Sun and Ranson [24] developed a 3D model that successfully linked full waveform data to the structural and optical properties of forest by building a 3D forest stand scene divided into cells with specific characteristics. Although Koetz *et al.* [25] successfully retrieved LAI from a synthetic dataset using the 3D model, this study failed to retrieve the vertical foliage distribution of the forest. Ni-Meister *et al.* [26] developed a method to derive the gap probability and canopy cover from LiDAR waveforms by modifying a hybrid geometric optical-radiative transfer (GORT) model. The basic assumption of the model is that gap probability is the reverse of the vertical canopy profile because laser energy can only penetrate to the lower canopy layer or ground through gaps. However, due to the clumping effect for natural vegetation, the actual foliage profile cannot be determined; only the gap probability and apparent foliage profile can be directly retrieved from the LiDAR return [27]. Armston, *et al.* [28] and Chen, *et al.* [29] used a simple RT model for direct retrieval of canopy gap probability from waveform LiDAR. This model can improve the estimation of canopy gap probability from waveform LiDAR data and is self-calibrating for the ratio of the reflectance from crown and ground.

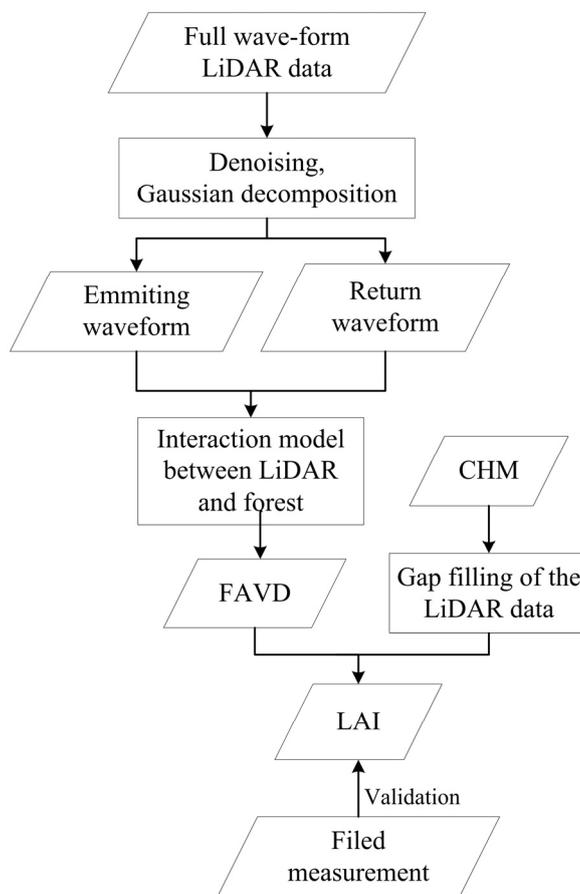
Morsdorf *et al.* [30] simulated LiDAR return waveforms by combining three different modeling components: the leaf optical properties, the tree structure, and the LiDAR measurement process (which builds on the open-source ray-tracing program POVRAY). Using the multi-spectral canopy LiDAR data, the NDVI profiles of the trees were obtained to monitor the chlorophyll content of the forests. Lewis [31] built a canopy reflectance model that can create a 3D structure model of individual plants using Monte Carlo ray tracing. North *et al.* [32] used the Monte Carlo radiative transfer model of GLAS waveforms for 3D canopies within the framework of the FLIGHT model; the integrated waveform energy and directly derived bidirectional reflectance factors (BRFs) from FLIGHT were in good agreement. Hancock *et al.* [33] used the Monte Carlo ray tracer from Lewis [31] to create a full waveform LiDAR simulator of dual wavelength to separate canopy from ground over steep topography. Though Monte Carlo ray tracers make far fewer assumptions than did Sun and Ranson [24] and Ni-Meister *et al.* [26] in their methods, they are computationally complex.

Figure 1 illustrates the basic process of the LiDAR–forest interaction. When the LiDAR pulse penetrates the canopy, the return waveform reflects the inner structure of the canopy. This study aims to map the spatial and vertical distribution of forest structural parameters using waveform LiDAR data on the basis of a physical inversion model. The canopy that the laser pulse shoot was divided into many layers and the model assumes that either foliage or air is in the slice. The LAI and foliage area volume density (FAVD) of each layer (shown in small cubes in Figure 1) is addressed in the inversion process, and the final result of the LAI is calculated based on the FAVD retrieval results. The main steps are illustrated in Figure 2. First, the full-waveform data are denoised and decomposed to rebuild the emitting and return waveforms using the Gaussian decomposition parameters. Then, FAVD of each LiDAR shot is inverted based on the physical radiative transfer model between LiDAR and the forest scene. Due to the sampling missing data between footprints of the waveform data, FAVD of the missing sampling shots is filled in using a canopy height model (CHM) of the study area, which was generated from the discrete LiDAR data concurrently obtained with the full-waveform data. Finally, LAI was retrieved using the FAVD at a 10-m resolution and compared with the CHM-derived LAI and the *in situ* LAI measurements. Note that in this paper, the leaves and the trunk are not considered

separately, so the FAVD and LAI are actually equivalent to plant area volume density and plant area index.



**Figure 1.** Basic process of the Light Detection And Ranging (LiDAR)–forest interaction.



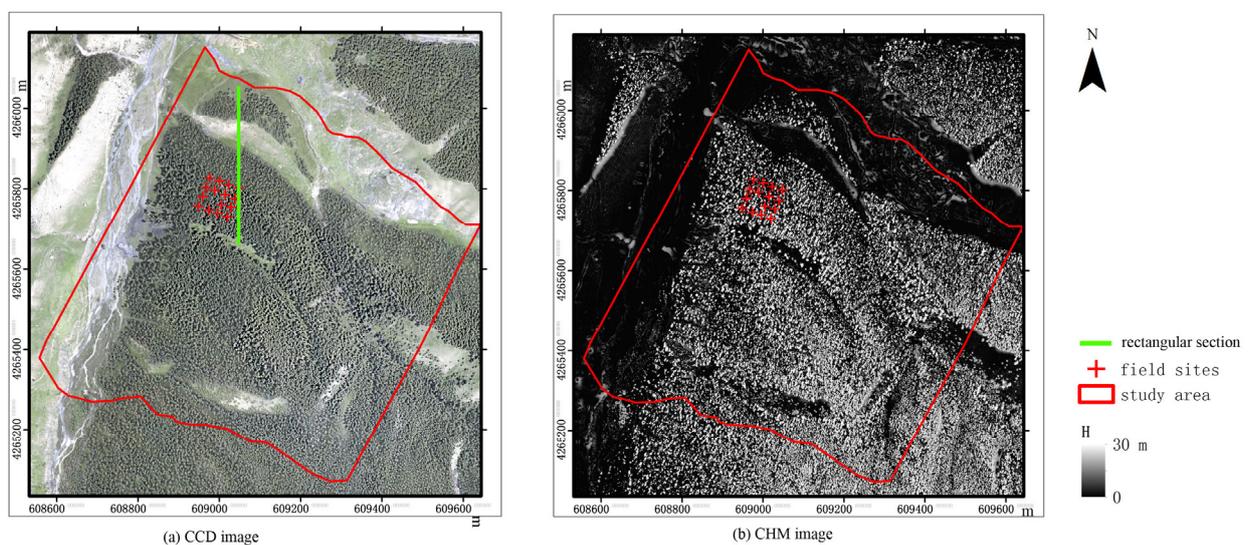
**Figure 2.** Flow chart of leaf area index (LAI) retrieval from full-waveform LiDAR data.

## 2. Data and Pre-Processing

### 2.1. Study Sites

The study area, Dayekou forest (38.6°N, 100°E), is located in the Qilian Mountain area of Gansu province in the arid region of northwestern China. The mean altitude of this study area is approximately 2400 m, and the slope of the study area is relatively gentle with a mean value of approximately 8.2°.

The forest is influenced by the temperate continental mountain climate, and the annual precipitation occurs mainly in summer. Prevalent vegetation types at this site are mountain pastures and forests, and the dominant forest type is Qinghai Spruce (*Piceacrassifolia*). Figure 3 shows an RGB photograph and CHM image of the study area with a spatial resolution of 0.5 m.



**Figure 3.** Distribution of the study area and field sites on (a) RGB photograph and (b) CHM image. Both images have a spatial resolution of 0.5 m.

### 2.2. Field Measurements

Field measurements were acquired during June 2008 in 15 sample plots with a size of 25 m × 25 m (shown as red crosses in Figure 3). LAI at each plot was measured using LAI-2000 Plant Canopy Analyzer (LAI-2000) [34] and Tracing Radiation and Architecture of Canopies (TRAC) [35], as described in detail in [36].

In our study, the true LAI was used as the test data, which is a combination of LAI measured using LAI-2000 and LAI measured using TRAC, and the ratio, namely foliage clumping index, which can convert effective LAI to true LAI, is derived from TRAC.

### 2.3. Airborne LiDAR Data

A full-waveform LiDAR system digitizes the complete waveform of the emitting laser pulse as well as the return laser pulse scattered back from the target. The discrete-return LiDAR system only records single or multiple returns often taken as the peak of a pulse. The airborne full-waveform LiDAR data

for the study area was acquired on June 28, 2008 using a RIEGL LMS-Q560 laser scanner; the LiDAR system and flight parameters are listed in Table 1 [37]. Though the scan angle range is  $\pm 30^\circ$ , only the scan route that was under nadir was chosen, *i.e.* the scan angle of chosen dataset was less than  $10^\circ$ .

In addition, a CHM (Figure 3b) that represents the forest canopy height distribution was built by filtering the discrete LiDAR data concurrently obtained by the scanner. The discrete raw data was processed using the Terrasolid software to classify the data into two types, *i.e.*, ground points and non-ground points, which can produce a digital elevation model (DEM) and a digital surface model (DSM), respectively. The CHM was created at a resolution of 0.5 m by subtracting the DEM from DSM.

**Table 1.** LMS-Q560 flight and system parameters.

Absolute Flight Height	Laser Wavelength	Laser Beam Divergence	Scan Angle Range
3550 m	1550 nm	$\leq 0.5$ mrad	$\pm 30^\circ$

#### 2.4. LAI Derived from CHM

A 10 m LAI map derived from CHM was used as a dataset to compare with LAI derived from the full-waveform data. Here LAI was calculated based on the Beer-Lambert law [38,39]:

$$LAI = -\frac{\cos \theta}{G(\theta) \cdot \Omega} \ln(1 - f_{cover}) \quad (1)$$

where  $G(\theta)$  is the foliage projection coefficient characterizing the foliage angular distribution, in a canopy with a spherical leaf angle distribution,  $G(\theta)$  is approximated by 0.5,  $\theta$  is the zenith angle of the incoming radiation, which is ignored since only data with scan angle  $< 10^\circ$  was used in this study,  $\Omega$  is the clumping index, since the study site is mainly covered by Qinghai spruce,  $\Omega$  was assigned to an constant 0.69 calculated from measurement by TRAC.  $f_{cover}$  is the canopy cover factor. In this study, CHM pixels greater than zero are considered to be canopy pixels, and  $f_{cover}$  is calculated as the ratio between the number of canopy pixels and the number of corresponding CHM pixels within the 10 m-resolution pixel.

### 3. Methods

#### 3.1. Decomposition of Waveforms

Gaussian decomposition provides estimates of the location and scattering properties of the targets within the travel path of the laser beam, and it assumes that the cross-sectional profile can be represented by a series of Gaussian functions [16]. The input parameters for the inversion model must be prepared by decomposition of the waveform data into a sum of Gaussian components. First, we assume the emitting laser pulse is Gaussian, and the return waveform is a series of Gaussian components:

$$f(x) = \sum_{i=1}^l a_i \exp(-(x - x_i)^2 / 2\sigma_i^2) \quad (2)$$

where  $l$  is the number of Gaussians, and  $\alpha_i$ ,  $x_i$ , and  $\sigma_i$  is the amplitude, position, and half-width of the Gaussian component, respectively. The essence of Gaussian decomposition is to solve  $3l$  unknowns of  $L$  observation equations ( $L > 3l$ ) using the nonlinear least-squares method (LMS).

Before decomposition of the waveform, the noise should be removed by calculating the threshold for each waveform. In our study, Persson’s [15] algorithm, which is based on computing the median absolute deviation (*MAD*), was adopted to calculate the noise threshold. The equation is as follows:

$$\sigma_{noise} = 1.4826 \cdot median(|wave - median(wave)|) \tag{3}$$

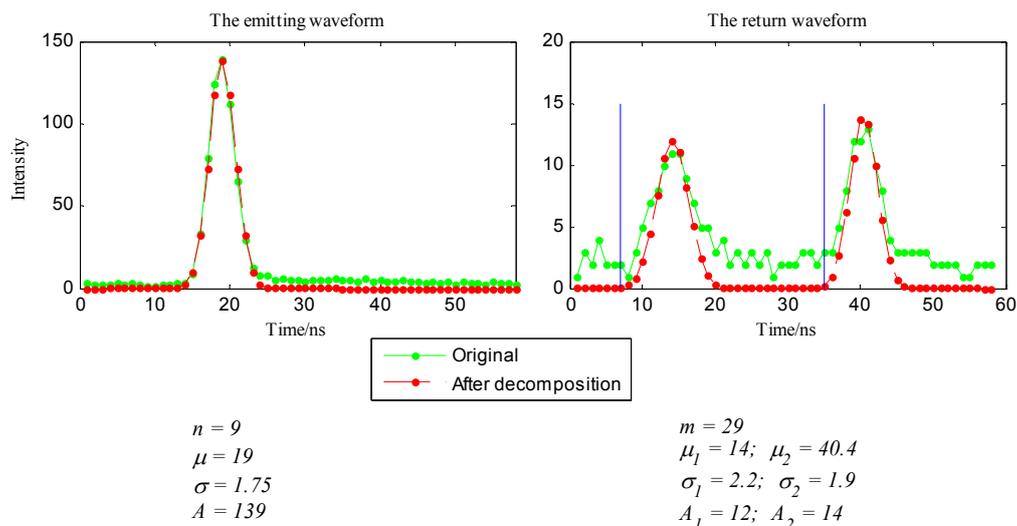
where *wave* is the original waveform, and *median* means calculation of the *MAD* from samples.

The samples of the original waveform below the noise threshold were set to zero. Then, to eliminate isolated values, samples were set to zero if both left and right neighbors were also zero. To better estimate the number, amplitude, position, and half-width of the sub-Gaussian components, the initial values of these parameters should be defined. The initial values of  $l$ ,  $x_i$ , and  $\alpha_i$  were calculated based on the number, position, and value, respectively, of the local maxima of the denoised waveform, and the initial value of  $\sigma_i$  was set to 2, because inspection of the emitting waves indicated that their half-widths were near 2.

Equation (4) is the object function of the LMS optimization algorithm, where  $f(x_i)$  is the observation value,  $y_i$  is the calculated value, and  $w_i$  is the observation error. Optimization is the process of constantly adjusting model parameters to minimize the objective function value. Sequential quadratic programming (SQP), which is an excellent optimization algorithm for solving nonlinear programming problems [40], is adopted to search for the cost function minimum.

$$object = \sum [(y_i - f(x_i))/w_i]^2 \tag{4}$$

Figure 4 shows a pair of emitting and return waveforms before and after Gaussian decomposition. The emitting wave can be modeled by a Gaussian wave using nine sampling values.



**Figure 4.** A pair of emitting and return waveforms before and after Gaussian decomposition.

According to Dubayah and Drake [41] and Mallet and Bretar [22], for small-footprint LiDAR systems, the laser beam has difficulty penetrating the vegetation and has a high possibility of missing

the ground under dense canopy. In this paper, we assume that the last return Gaussian wave from either ground or tree branch and is shown in the right figure of Figure 4 by the sub-Gaussian wave whose position is 40.4. The distance between the start of the entire signal and the start of the last wave is defined as the canopy range ( $H_r$ ), shown as the distance between the two blue lines in Figure 4. The number of canopy layers is defined as the ratio between canopy range and the vertical sampling resolution of the LiDAR system (0.15 m). Moreover, a series of parameters after decomposition should be determined as the input parameters of the FAVD inversion model: the number of canopy layers  $m$ , the number of emitting wave samples  $n$ , and the reconstructed emitting and return waveforms.

### 3.2. Model of LiDAR Waveform of Forest

The physical model of the interaction between LiDAR emitting laser pulse and forest was built based on RT theories and a 3D model [24]. In this model, the forest canopy is described as a turbid scattering medium that is parameterized by its FAVD,  $\mu$ , the Ross-Nilson G-factor, and the foliage reflectance,  $r_{leaf}$ . The background (soil) within the forest scene is parameterized by its reflectance,  $r_{soil}$ .

Because LiDAR is working under hot spot conditions, the bidirectional gap probability  $p$  of the canopy in the direction  $\Omega_i$ , at the canopy depth  $z$ , can be expressed as:

$$p(z, \Omega_i) = \exp\left(-\int_0^z u_L(z')G(z', \Omega_i)dz'\right) \quad (5)$$

where  $u_L(z')$  is the density of a one-sized leaf area at depth  $z'$ ,  $G(z', \Omega_i)$  is the G-factor, the mean projection of unit foliage area at a depth  $z'$  in the direction  $\Omega_i$ .

The emitting laser pulse was divided into  $n$  narrow pulses ( $I_{emit}(i), i = 1 \dots n$ ), with the duration time set to the time resolution of the system ( $\Delta t$  1 ns), and the canopy was divided into  $m$  layers from top to bottom, with the thickness  $\Delta z$  the same as the vertical range resolution of the system ( $\Delta z = \Delta t \times c/2 = 0.15$  m,  $c$  is the speed of light). Supposing the LiDAR return pulse begins at a time when the first sub-emitting pulse reaches the first canopy layer, then, when the  $i$ th sub-pulse reaches the  $j$ th sub-canopy layer, the time delay is:

$$t(i, j) = 2(i + j)\Delta z/c \quad (6)$$

In addition, the returned energy is:

$$I(i, j) = I_{emit}(i) \cdot R_j \cdot E_{j-1} \quad (7)$$

where  $R_j = \Gamma \cdot u_j \cdot \omega$  is the back-scatter factor of the  $j$ th canopy,  $\omega$  is a parameter determined by the LiDAR system,  $\Gamma$  is the scattering phase function of the canopy ( $\Gamma = r_{leaf}G/\pi$ ), and  $E_{j-1}$  is the extinction above this canopy layer ( $E_{j-1} = \exp(-\sum_{k=1}^{j-1}(u \cdot G)_k \cdot \Delta z)$ ) [24].

When the  $i$ th sub-pulse reaches the  $m+1$  soil layer, the returned energy is calculated from Equation (8).

$$I(i, m + 1) = I_{emit}(i) \cdot r_{soil} \cdot \omega \cdot E_m \quad (8)$$

To set the LiDAR system parameter,  $\omega$ , the soil return sub-Gaussian wave extracted from the full return waveform was used. So, in this model, the calibration of the LiDAR system is not necessary. The return signal is recorded in  $m + n$  digital bins according to the time delay.

### 3.3. Inversion of FAVD and LAI

The input parameters of this model include the reflectance of leaves, the reflectance of soil, and the G-factor. The reflectance of leaves and soil were set to 0.55, and 0.2, respectively, according to their field measurements [42]. G-factor was assumed to be 0.5, the same as in Section 2.4. The LMS method was used for the inversion process, and the cost function described in Equation (9) was used.

$$\text{object} = \sum_{i=1}^n (I_{\text{observe}}(i) - I_{\text{model}}(i))^2 \quad (9)$$

In Equation (9),  $I_{\text{observe}}$  is the LiDAR observed return signal,  $I_{\text{model}}$  is the result estimated by the interaction model, and  $n$  is the number of observation samples; moreover, the vertical FAVD within a laser shot,  $u_k (k = 1 \dots m)$ , was set to be a nonnegative to control the optimization process.

### 3.4. Filling in Missing LiDAR Data

After displaying the sampling positions of laser shots, it was found that many of the laser shots missed the denser canopy areas, as shown in Figure 5a. Note that all of the samples along the waveform are missing; this may occur either because the forest is so dense and high that the laser cannot penetrate or because the system fails to detect the return signal due to an excess of energy attenuation after penetration.

To obtain the complete forest information, the corresponding missing FAVD was filled in as described in this section. First, the number and position of the missing laser shots was calculated based on the temporal and geolocational information of the existing ones. The statistics of the temporal information for emitting laser shots indicate that the normal emitting time interval,  $\Delta t$ , is 0.00002 s; therefore, the missing laser shot number  $n$  and its corresponding coordinates  $(X_i, Y_i, Z_i), i = 1, 2 \dots n$  are:

$$n = \Delta T / \Delta t \quad (10)$$

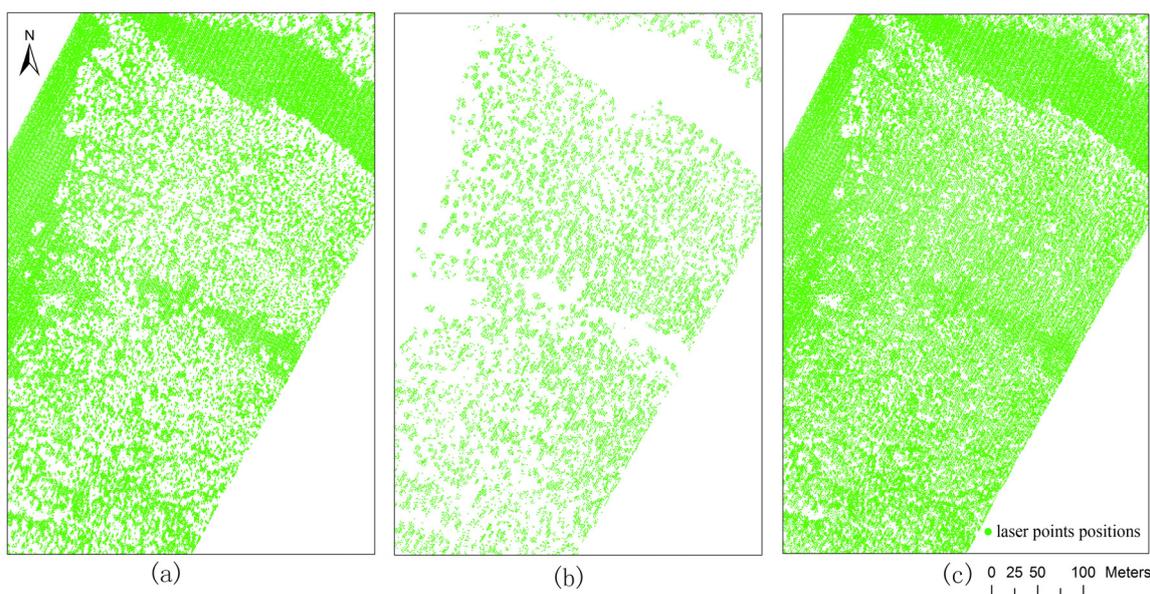
$$\begin{aligned} X_i &= X_1 + \frac{i}{n+1} \cdot (X_2 - X_1) \\ Y_i &= Y_1 + \frac{i}{n+1} \cdot (Y_2 - Y_1) \\ Z_i &= Z_1 + \frac{i}{n+1} \cdot (Z_2 - Z_1) \end{aligned} \quad (11)$$

where  $\Delta T$  is the time interval between two existing neighboring laser shots, and  $(X_1, Y_1, Z_1), (X_2, Y_2, Z_2)$  are their coordinates.

Then, the corresponding forest canopy height of the missing laser shot was searched on the CHM according to its coordinates. When the corresponding height was confined to the range [1.9, 50], the missing shot area was considered to be canopy, and in this case, FAVD of this shot could be filled in. The range [1.9, 50] is chosen based on the tree height measurements from the sample plots (1656 trees in total). To minimize the error, the nearest existing shot that had the same or similar canopy height value was found, and its inverted FAVD was assigned to the missing one. Following this step, the FAVD reconstruction for one missing shot was completed. Figure 5b shows the positions of the filled-in shots; the shapes of these filled-in shots are round circles corresponding to forest canopy.

Figure 5c shows the positions of the original and filled-in shots combined, and the number of empty holes is greatly reduced. Besides, among the filled FAVD profiles, about half of them have the height exceeding 8 m.

To validate this gap-filling method, 1500 out of 8000 waveforms were randomly removed from the site, and were filled up again using the above method. It was found that 939 waveforms (about 63%) can be filled. In reality, it is the areas with dense and high forest coverage that lacks the full-waveform data; this validation method cannot detect this bias. However, areas with tree height above 8 m were examined, and 82% waveforms can be filled. Besides, aggregated FAVD values of the 939 original and filled waveforms were compared, and results indicate that they have a correlation of 0.64.



**Figure 5.** Positions of the laser shots (a) original (b) filled using the ancillary CHM data (c) original and filled.

### 3.5. Inversion of LAI from FAVD

The diameter of the laser shot is considered a function of the laser incidence angle ( $\alpha$ ) and the relative flight height ( $H_f$ ) approximately expressed as:

$$D = H_f \cdot \alpha \quad (12)$$

In this equation,  $\alpha$  is determined by the laser beam divergence and the illuminating surface slope; because the laser shot is small, the surface slope effect is not considered in our study, and  $\alpha$  is assigned as  $0.5 \text{ mrad}$ , according to the laser beam divergence (see Table 1). According to the flight parameters around the sample plots,  $D$  is primarily in the range of (0.3~0.4 m).

To retrieve the 2D LAI information from the 3D FAVD of the inversion results, a cube was constructed, with the size determined by the coordinate range of the LiDAR samples. To guarantee that there were enough samples within the voxels in the cube, the size of the voxel was set to be  $10 \text{ m} \times 10 \text{ m} \times 15 \text{ m}$ . Supposing the LiDAR system takes a random sample after the reconstruction work, the FAVD of one voxel is the volume weighted average of the FAVD of the samples ( $u_1, u_2, \dots, u_k$ ) within the voxel, as described by Equation (16).

$$u_{voxel} = \frac{u_1v_1 + u_2v_2 + \dots + u_kv_k}{v_1 + v_2 + \dots + v_k} \tag{13}$$

$$v_k = \pi \left( \frac{\tan(\alpha) \cdot H_k}{2} \right)^2 \cdot \Delta z \tag{14}$$

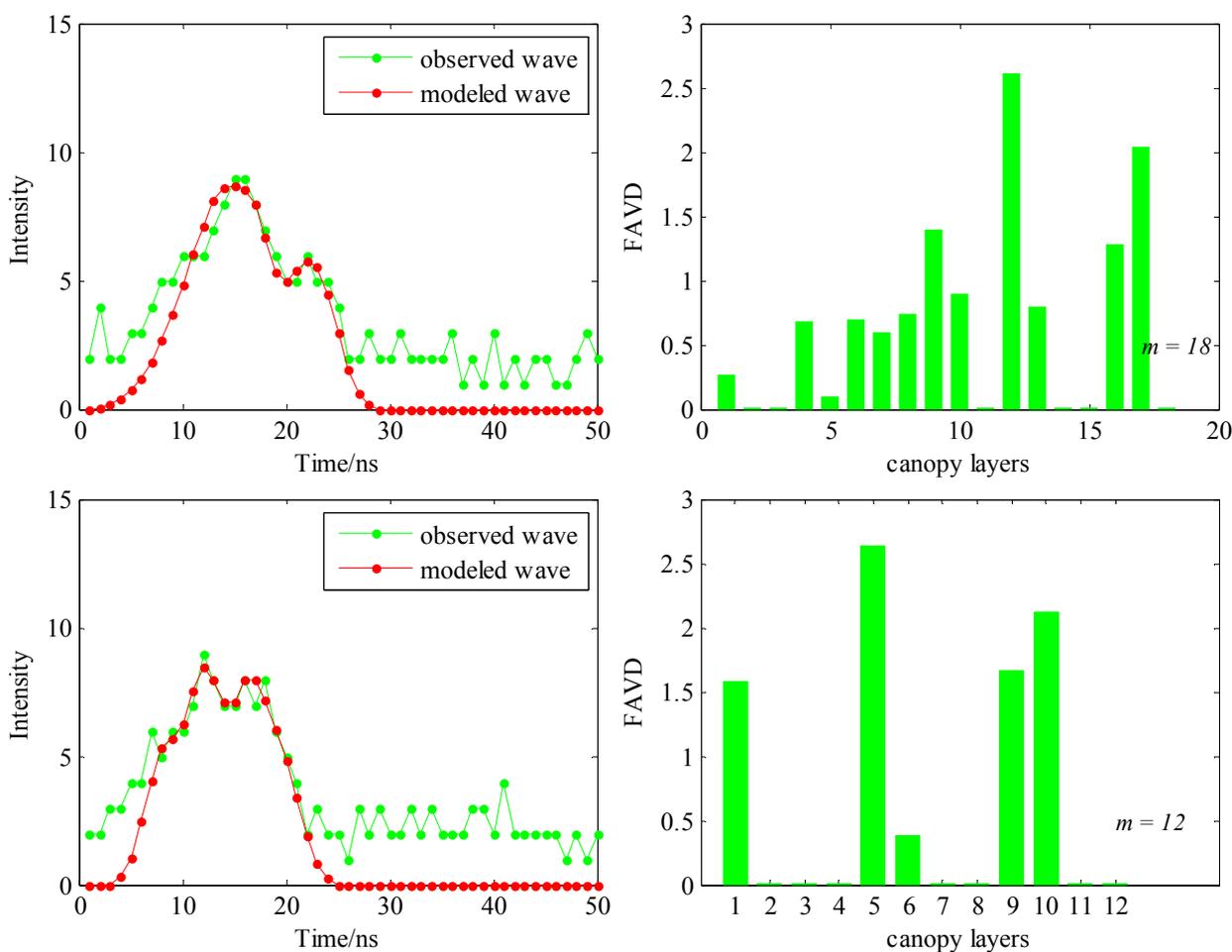
In Equation (14),  $H_k$  is the height from the sensor to the target, the value of which can be extracted from the LiDAR data.

After calculation of the voxel FAVD values, the 2D LAI information can be obtained by integration of FAVD values in the height.

### 4. Results

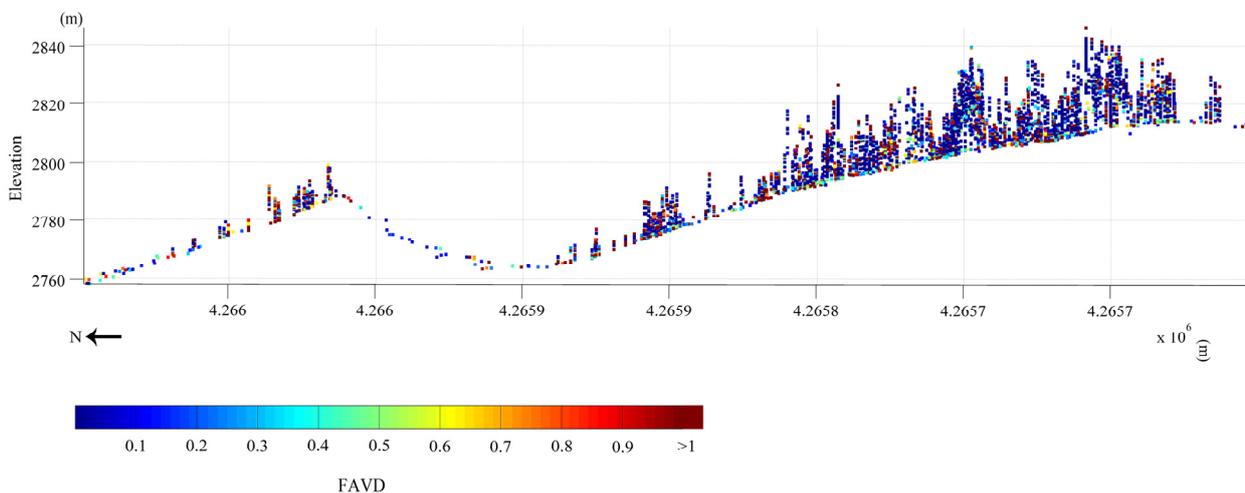
#### 4.1. Vertical FAVD Profile on Laser Shot Scale

A total of 40,022 return waveforms were inverted by the model in this paper. Figure 6 shows the FAVD inversion results of two waveforms: the left images are the original observed return waveforms and the modeled waveforms, and the right images are the results of FAVD of each canopy layer, with a total of 18 and 13 values, respectively. The red waves were modeled by the LiDAR-forest radiative transfer model. The modeled waveforms and the original waves align well.

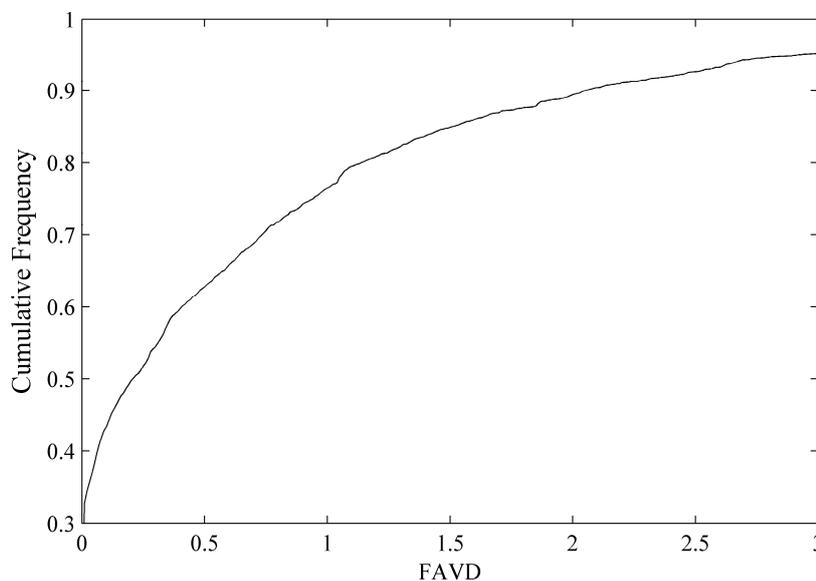


**Figure 6.** FAVD (Foliage Area Volume Density) inversion results of two different waveforms.

To better illustrate the inverted FAVD of our study area, Figure 7 gives a vertical profile of the FAVD distributed in the rectangular slice area in Figure 3a. It can be found from Figure 7 that trees are very sparse on the southern side of the mountain, which is consistent with the field survey and the photograph in Figure 3. Figure 8 is the cumulative frequency distribution of FAVD for the whole study area; 42.5% of FAVD values are below 0.1, 80.1% are below 1, and 94.7% are below 3.



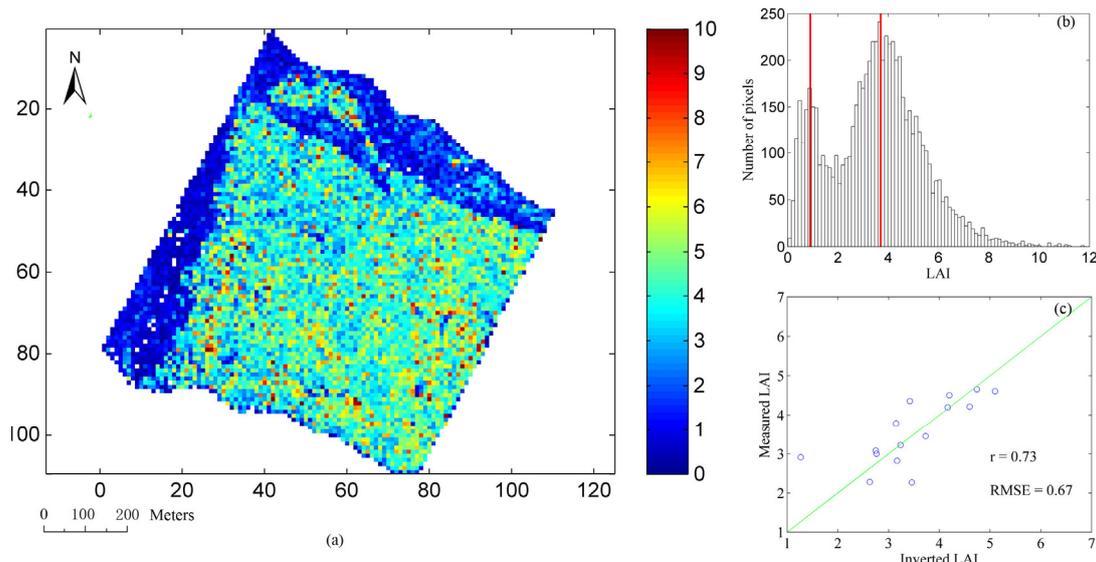
**Figure 7.** FAVD distribution in the green rectangle area in Figure 3. Trees on the southern side of the mountain are very sparse.



**Figure 8.** Cumulative frequency distribution of FAVD for the whole study area; 42.5% of FAVD values are below 0.1, 80.1% are below 1, and 94.7% are below 3.

4.2. LAI Inversion and Comparison with field measurements

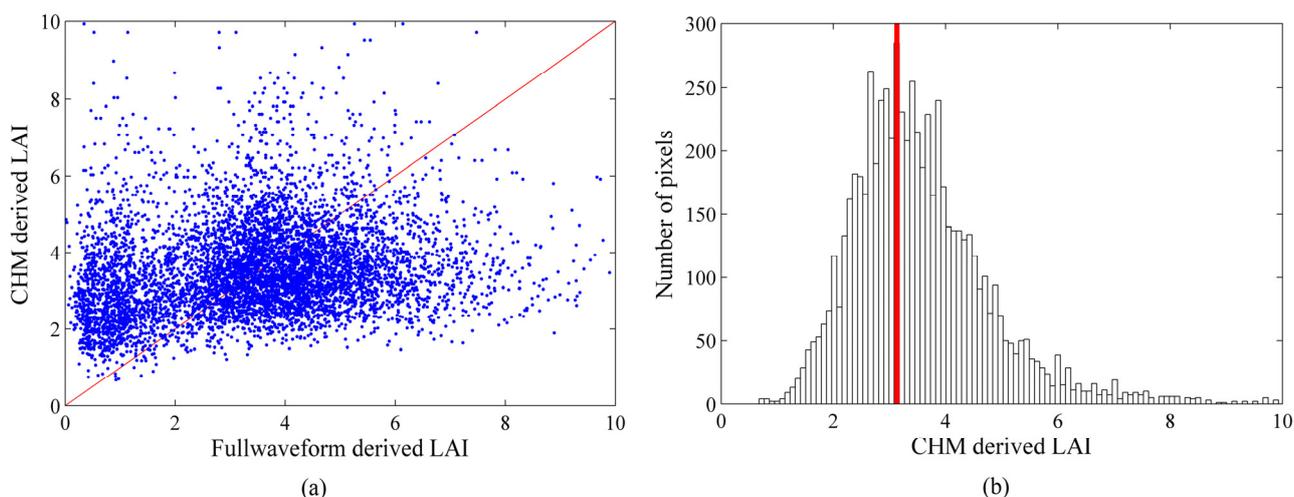
Figure 9a,b show the LAI inversion results of our study area. A comparison of (a) and (b) illustrates that LAI distribution is consistent with the canopy density. There are two LAI peaks (0.9 and 3.7) shown in Figure 9b corresponding to the two different landcover types in the study area, which are grass and Qinghai spruce, respectively.



**Figure 9.** (a) LAI inversion results of the study area (with the spatial resolution of 10 m). (b) Histogram of LAI. (c) Scatter plot of the inverted LAI and the field measured LAI ( $r = 0.73$ , RMSE = 0.67).

However, little correlation was observed between the CHM derived LAI and full-waveform derived LAI ( $r$  is 0.20, Figure 10a). The histogram of CHM derived LAI shows that there is only one LAI peak located at 3.13 (Figure 10b). This may result from using simply canopy cover to calculate LAI, ignoring the vertical information of the canopy, and thus cannot detect the differences between high trees and low grass. Moreover, full-waveform derived LAI is larger than CHM derived LAI at higher LAI values ( $LAI > 2.5$ ), and CHM derived LAI is more easily saturated.

In addition, the estimated LAI values were compared with the field measured LAI in 15 sample plots to address the accuracy of our method; as shown in Figure 9c, the value of the correlation coefficient,  $r$ , is 0.73, and the RMSE value is 0.67.



**Figure 10.** (a) Scatter plot of full-waveform derived LAI vs. CHM derived LAI (poor correlation, with  $r = 0.20$ ); (b) Histogram of CHM derived LAI, with peak located at 3.13.

## 5. Discussion

This study represents an attempt to retrieve vertical forest information from full-waveform LiDAR data. This technique can still be improved in several aspects. First, this method must be tested on different forest types and expanded to a larger study area; due to the lack of computing capacity, only a small area was chosen in this study. Second, due to the lack of field measurements of forest FAVD distributions in our study area, the FAVD inversion results should be validated in a future study. Furthermore, this LiDAR–forest interaction model is only suitable for use with small-footprint full-waveform LiDAR data; if large-footprint LiDAR data, such as GLAS [43] or LVIS [44] data are used, the slope effect must be considered and corrected [45]; besides, multi-scattering is ignored in this RT model since the laser beam divergence is less than 0.5 mrad, while for large-footprint LiDAR data, the multi-scattering effect should be included. Lastly, the gap filling method used in this paper will cause underestimation of LAI, since the true FAVD cannot be obtained due to the loss of full-waveform data in dense and high forest. In future work, a more detailed assessment of the potential biases of this method is needed. However, if high-density sampling LiDAR data is available, then this gap filling procedure and its potential biases could be avoided. With the development of LiDAR systems and the feasibility of LiDAR data, more attention will be devoted to studies concerning the physical interaction mechanism between LiDAR and the target, or the combination of LiDAR data with other remote sensing data sources, such as hyperspectral and high spatial resolution data.

## 6. Conclusions

Recently, various approaches have been undertaken to retrieve canopy structural information directly from LiDAR data sources, or by integration of optical remote sensing data [46]. However, current work has failed to make full use of the waveform information, nor to explore the physical mechanism of the interaction between LiDAR and vegetation. The main advantage of this study consisted in direct and rapid retrieval of forest canopy vertical FAVD profiles and LAI from airborne full-waveform LiDAR data, based on the physical LiDAR–canopy interaction model. The method discussed in this paper makes it possible to take advantage of the geolocational, temporal and intensity information of the full-waveform data, besides, with the aid of CHM data, the inconsistency and data gap could be eliminated to some extent. At last, we compared the full-waveform derived LAI with field measurements, the values of correlation coefficient  $r$  and RMSE were 0.73 and 0.67, respectively. This result indicated that LAI inverted from waveform LiDAR data has the ability to serve as one of the validation sources for other products and data.

## Acknowledgments

This research was supported in part by the Special Funds for Major State Basic Research Project (2013CB733403), the National Natural Science Foundation of China (No. 41171263), the National 863 Program (2013AA12A303, 2012AA12A301), and the Fundamental Research Funds for the Central Universities. The authors would also like to thank anonymous reviewers who gave valuable suggestions that have helped to strengthen the paper.

## Author Contributions

Jindi Wang and Jinling Song conceived and designed the study. Jinling Song provided the full-waveform LiDAR and field measurement data. Han Ma performed the experiments, designed the inversion model, wrote the paper, and integrated the contributions from all authors as a whole. Jinling Song and Jindi Wang reviewed and edited the manuscript. All authors read and approved the manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Neilson, R.P.; Drapek, R.J. Potentially complex biosphere responses to transient global warming. *Global Change Biol.* **1998**, *4*, 505–521.
2. Baldocchi, D.D.; Wilson, K.B.; Gu, L. How the environment, canopy structure and canopy physiological functioning influence carbon, water and energy fluxes of a temperate broad-leaved deciduous forest—An assessment with the biophysical model canoak. *Tree Physiol.* **2002**, *22*, 1065–1077.
3. Sellers, P.J.; Dickinson, R.E.; Randall, D.A.; Betts, A.K.; Hall, F.G.; Berry, J.A.; Collatz, G.J.; Denning, A.S.; Mooney, H.A.; Nobre, C.A.; *et al.* Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science* **1997**, *275*, 502–509.
4. Lefsky, M.A.; Cohen, W.B.; Parker, G.G.; Harding, D.J. LIDAR remote sensing for ecosystem studies. *Biosci.* **2002**, *52*, 19–30.
5. Liang, S. *Quantitative remote sensing of land surfaces*. John Wiley & Sons: Hoboken, NJ, USA, 2004; Volume 1.
6. Calders, K.; Lewis, P.; Disney, M.; Verbesselt, J.; Herold, M. Investigating assumptions of crown archetypes for modelling LiDAR returns. *Remote Sens. Environ.* **2013**, *134*, 39–49.
7. Dubayah, R.; Sheldon, S.; Clark, D.; Hofton, M.; Blair, J.; Hurtt, G.; Chazdon, R. Estimation of tropical forest height and biomass dynamics using LIDAR remote sensing at la Selva, Costa Rica. *J. Geophys. Res.* **2010**, *115*, G00E09.
8. Lefsky, M.A.; Harding, D.J.; Keller, M.; Cohen, W.B.; Carabjal, C.C.; Espirito-Santo, F.D.B.; Hunter, M.O.; de Oliveira, R., Jr, Estimates of forest canopy height and aboveground biomass using icesat. *Geophys. Res. Lett.* **2005**, *32*, L22S02.
9. Lefsky, M.A.; Keller, M.; Pang, Y.; de Camargo, P.B.; Hunter, M.O. Revised method for forest canopy height estimation from geoscience laser altimeter system waveforms. *J. Appl. Remote Sens.* **2007**, *1*, 013537.
10. Riaño, D.; Valladares, F.; Condés, S.; Chuvieco, E. Estimation of leaf area index and covered ground from airborne laser scanner (LiDAR) in two contrasting forests. *Agr. Forest Meteorol.* **2004**, *124*, 269–275.

11. Van Leeuwen, M.; Nieuwenhuis, M. Retrieval of forest structural parameters using LiDAR remote sensing. *Eur. J. Forest Res.* **2010**, *129*, 749–770.
12. Hofton, M.A.; Rocchio, L.E.; Blair, J.B.; Dubayah, R. Validation of vegetation canopy LiDAR sub-canopy topography measurements for a dense tropical forest. *J. Geodyn.* **2002**, *34*, 491–502.
13. Jaskierniak, D.; Lane, P.N.J.; Robinson, A.; Lucieer, A. Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions. *Remote Sens. Environ.* **2011**, *115*, 573–585.
14. Hofton, M.A.; Minster, J.B.; Blair, J.B. Decomposition of laser altimeter waveforms. *Geosci. Remote Sens., IEEE Transactions on* **2000**, *38*, 1989–1996.
15. Persson, Å.; Söderman, U.; Töpel, J.; Ahlberg, S. Visualization and analysis of full-waveform airborne laser scanner data. In Proceedings of ISPRS WG III/3, III/4, V/3 Workshop “Laser Scanning 2005”, Enschede, the Netherlands, 12–14 September 2005; pp. 103–108
16. Wagner, W.; Ullrich, A.; Ducic, V.; Melzer, T.; Studnicka, N. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *Int. J. Photogramm. Remote Sens.* **2006**, *60*, 100–112.
17. Lindberg, E.; Olofsson, K.; Holmgren, J.; Olsson, H. Estimation of 3D vegetation structure from waveform and discrete return airborne laser scanning data. *Remote Sens. Environ.* **2012**, *118*, 151–161.
18. Sun, G.; Ranson, K.; Kimes, D.; Blair, J.; Kovacs, K. Forest vertical structure from glas: An evaluation using lvis and srtm data. *Remote Sens. Environ.* **2008**, *112*, 107–117.
19. Farid, A.; Goodrich, D.; Bryant, R.; Sorooshian, S. Using airborne LiDAR to predict leaf area index in cottonwood trees and refine riparian water-use estimates. *J. Arid Environ.* **2008**, *72*, 1–15.
20. Zhao, K.; Popescu, S. LiDAR-based mapping of leaf area index and its use for validating globcarbon satellite lai product in a temperate forest of the southern USA. *Remote Sens. Environ.* **2009**, *113*, 1628–1645.
21. Adams, T.; Beets, P.; Parrish, C. Extracting more data from LiDAR in forested areas by analyzing waveform shape. *Remote Sens.* **2012**, *4*, 682–702.
22. Mallet, C.; Bretar, F. Full-waveform topographic LiDAR: State-of-the-art. *Int. J. Photogramm. Remote Sens.* **2009**, *64*, 1–16.
23. Blair, J.B.; Hofton, M.A. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophys. Res. Lett.* **1999**, *26*, 2509–2512.
24. Sun, G.; Ranson, K.J. Modeling LiDAR returns from forest canopies. *Geosci. Remote Sens., IEEE Trans.* **2000**, *38*, 2617–2626.
25. Koetz, B.; Morsdorf, F.; Sun, G.; Ranson, K.; Itten, K.; Allgower, B. Inversion of a LiDAR waveform model for forest biophysical parameter estimation. *Geosci. Remote Sens. Lett., IEEE* **2006**, *3*, 49–53.
26. Ni-Meister, W.; Jupp, D.L.; Dubayah, R. Modeling LiDAR waveforms in heterogeneous and discrete canopies. *Geosci. Remote Sens, IEEE Trans.* **2001**, *39*, 1943–1958.
27. Tang, H.; Dubayah, R.; Swatantran, A.; Hofton, M.; Sheldon, S.; Clark, D.B.; Blair, B. Retrieval of vertical LAI profiles over tropical rain forests using waveform LIDAR at la Selva, Costa Rica. *Remote Sens. Environ.* **2012**, *124*, 242–250.

28. Armston, J.; Disney, M.; Lewis, P.; Scarth, P.; Phinn, S.; Lucas, R.; Bunting, P.; Goodwin, N. Direct retrieval of canopy gap probability using airborne waveform LiDAR. *Remote Sens. Environ.* **2013**, *134*, 24–38.
29. Chen, X.T.; Disney, M.I.; Lewis, P.; Armston, J.; Han, J.T.; Li, J.C. Sensitivity of direct canopy gap fraction retrieval from airborne waveform LiDAR to topography and survey characteristics. *Remote Sens. Environ.* **2014**, *143*, 15–25.
30. Morsdorf, F.; Nichol, C.; Malthus, T.; Woodhouse, I.H. Assessing forest structural and physiological information content of multi-spectral LiDAR waveforms by radiative transfer modelling. *Remote Sens. Environ.* **2009**, *113*, 2152–2163.
31. Lewis, P. Three-dimensional plant modelling for remote sensing simulation studies using the botanical plant modelling system. *Agronomie* **1999**, *19*, 185–210.
32. North, P.; Rosette, J.; Suárez, J.; Los, S. A monte carlo radiative transfer model of satellite waveform LiDAR. *Int. J. Remote Sens.* **2010**, *31*, 1343–1358.
33. Hancock, S.; Lewis, P.; Foster, M.; Disney, M.; Muller, J.-P. Measuring forests with dual wavelength LiDAR: A simulation study over topography. *Agr. Forest Meteorol.* **2012**, *161*, 123–133.
34. Gower, S.T.; Kucharik, C.J.; Norman, J.M. Direct and indirect estimation of leaf area index,  $f_{apar}$ , and net primary production of terrestrial ecosystems. *Remote Sens. Environ.* **1999**, *70*, 29–51.
35. Chen, J.M.; Cihlar, J. Plant canopy gap-size analysis theory for improving optical measurements of leaf-area index. *Appl. Opt.* **1995**, *34*, 6211–6222.
36. Zhuo, F.; Jindi, W.; Jinling, S.; Hongmin, Z.; Huaguo, H.; Baisong, C. Comparison of three indirect field measuring methods for forest canopy leaf area index estimation, In Proceedings of Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009, 12–17 July 2009; pp IV-1011–IV-1014.
37. Hug, C.; Ullrich, A.; Grimm, A. Litemapper-5600-a waveform-digitizing LiDAR terrain and vegetation mapping system. In Proceedings of International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, Freiburg, Germany 3–6 October 2004; Volume XXXVI-8/W2, pp. 24–29.
38. Richardson, J.J.; Moskal, L.M.; Kim, S.-H. Modeling approaches to estimate effective leaf area index from aerial discrete-return LiDAR. *Agr. Forest Meteorol.* **2009**, *149*, 1152–1160.
39. Nilson, T. A theoretical analysis of the frequency of gaps in plant stands. *Agr. Meteorol.* **1971**, *8*, 25–38.
40. Boggs, P.T.; Tolle, J.W. Sequential quadratic programming. *Acta Numerica* **1995**, *4*, 1–51.
41. Dubayah, R.O.; Drake, J.B. LiDAR remote sensing for forestry. *J. For.* **2000**, *98*, 44–46.
42. Jinling, S.; Jindi, W.; Zhuo, F.; Bengyu, W.; Xin, A.T. *WATER: Dataset of spectral reflectance observations of the Picea crassifolia at the super site around the Dayekou Guantan forest station*. Institute of Remote Sensing Applications, Chinese Academy of Sciences: Beijing, China; Institute of Forest Resource Information Techniques, Chinese Academy of Forestry: Beijing, China, 2008.
43. Abshire, J.B.; Sun, X.; Riris, H.; Sirota, J.M.; McGarry, J.F.; Palm, S.; Yi, D.; Liiva, P. Geoscience laser altimeter system (glas) on the icesat mission: On-orbit measurement performance. *Geophys. Res. Lett.* **2005**, *32*, 1–4.

44. Blair, J.B.; Rabine, D.L.; Hofton, M.A. The laser vegetation imaging sensor: A medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *Int. J. Photogramm. Remote Sens.* **1999**, *54*, 115–122.
45. Breidenbach, J.; Koch, B.; Kändler, G.; Kleusberg, A. Quantifying the influence of slope, aspect, crown shape and stem density on the estimation of tree height at plot level using LiDAR and insar data. *Int. J. Remote Sens.* **2008**, *29*, 1511–1536.
46. Ma, H.; Song, J.; Wang, J.; Xiao, Z.; Fu, Z. Improvement of spatially continuous forest LAI retrieval by integration of discrete airborne LiDAR and remote sensing multi-angle optical data. *Agr. Forest Meteorol.* **2014**, *189–190*, 60–70.

© 2015 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/4.0/>).