Article

Retrieval of Tree Height Percentiles over Rugged Mountain Areas via Target Response Waveform of Satellite Lidar

Hao Song [1], Hui Zhou [1,2,*], Heng Wang [1], Yue Ma [1], Qianyin Zhang [1] and Song Li [1,2]

1   School of Electronic Information, Wuhan University, Wuhan 430072, China; songhao@whu.edu.cn (H.S.);
    wh_els@whu.edu.cn (H.W.); mayue_eis@whu.edu.cn (Y.M.); zhangqianyin@whu.edu.cn (Q.Z.);
    ls@whu.edu.cn (S.L.)
2   Wuhan Institute of Quantum Technology, Wuhan 430206, China
*   Correspondence: zhouhui@whu.edu.cn

Abstract: The retrieval of tree height percentiles from satellite lidar waveforms observed over mountainous areas is greatly challenging due to the broadening and overlapping of the ground return and vegetation return. To accurately represent the shape distributions of the vegetation and ground returns, the target response waveform (TRW) is resolved using a Richardson–Lucy deconvolution algorithm with adaptive iteration. Meanwhile, the ground return is identified as the TRW component within a 4.6 m ground signal extent above the end point of the TRW. Based on the cumulative TRW distribution, the height metrics of the energy percentiles of 25%, 50%, 75%, and 95% are determined using their vertical distances relative to the ground elevation in this study. To validate the proposed algorithm, we select the received waveforms of the Global Ecosystem Dynamics Investigation (GEDI) lidar over the Pahvant Mountains of central Utah, USA. The results reveal that the resolved TRWs closely resemble the actual target response waveforms from the coincident airborne lidar data, with the mean values of the coefficient of correlation, total bias, and root-mean-square error (RMSE) taking values of 0.92, 0.0813, and 0.0016, respectively. In addition, the accuracies of the derived height percentiles from the proposed algorithm are greatly improved compared with the conventional Gaussian decomposition method and the slope-adaptive waveform metrics method. The mean bias and RMSE values decrease by the mean values of 1.68 m and 2.32 m and 1.96 m and 2.72 m, respectively. This demonstrates that the proposed algorithm can eliminate the broadening and overlapping of the ground return and vegetation return and presents good potential in the extraction of forest structure parameters over rugged mountainous areas.

Keywords: satellite lidar; received waveform; target response waveform; Richardson–Lucy deconvolution; height percentiles

1. Introduction

Forests are a significant component of ecosystems and play important roles in determining carbon storage, climate, and ecological functionalities. Monitoring the spatial and temporal characteristics of a forest's structure contributes to interpreting the global carbon cycle and environmental changes [1–4]. Remote sensing technologies have been proved to be highly effective for acquiring the forest structure at regional and even global scales. Passive optical sensor and active radar systems can mainly obtain canopy height and stem volume, but vertical forest structural information is still missing due to poor penetration [5–8]. Light detection and ranging (lidar) is an active remote sensing technique that can overcome such problems and accurately represent vertical forest structure and underlying topography.

The full-waveform satellite lidar can implement a global forest inventory, which is dependent on the amplitude of the received laser pulse signal reflected from the illuminated target surface. The received waveform provides an excellently resolved measurement of

the vertical distribution of the vegetation and ground within a laser footprint [9–12]. As one of the significant indicator parameters for characterizing forest structure, height percentiles are usually derived from the cumulative distributions of the received waveform [13,14]. Many studies have been carried out to predict forest characteristics using height percentiles. For example, height percentiles were used to develop a prediction system to calculate stem volume [15–19]. Also, based on the strong correlation between the canopy height and the height percentiles [17–22], Chen chose the 98th percentile of height to represent the highest canopy height and Pang et al. proved that the 75th percentile of height has the best correlation with actual canopy height [23,24]. In addition, Wang et al. employed canopy height percentiles as the independent variables to estimate the aboveground biomass with a determination coefficient of 0.90 [25]. The accurate extractions of the height percentiles are extremely significant for implementing such applications.

The key point in terms of obtaining canopy percentile heights is how to determine the ground elevation and signal start elevation from the received waveforms. Until now, there have been two common algorithms used to derive the canopy height percentiles. The first algorithm attempts to decompose the received waveform into multiple Gaussian components by using the Gaussian decomposition method [23,26,27]. The signal start elevation is identified as the first location where the waveform amplitude exceeds a threshold level. The ground elevation is assumed to be the peak location of the last Gaussian component. Thus, the height percentiles can be estimated from the cumulative distributions of the received waveform by their vertical distances relative to ground elevation. The Global Ecosystem Dynamics Investigation (GEDI) provides the relative height metrics in the level-2 product via a similar method. The difference is that the ground elevation is defined as the zero-crossing point of the first derivative in the denoised waveform [28–31]. The Gaussian decomposition method can be applicable for the received waveforms over the vegetated area with the flat ground. The second algorithm assumes that the ground return can be only modeled by one Gaussian function and fits the Gaussian ground return based on the ground surface slope calculated from Shuttle Radar Topography Mission (SRTM) data [25]. Meanwhile, the canopy height percentiles are redefined as the difference between height percentiles of the received waveform and height percentiles of the fitted ground component. This algorithm may have better adaptability compared to the Gaussian decomposition algorithm over the rugged mountain area, but it relies on the assumption of a single Gaussian ground return and the accurate surface slope derived from a priori knowledge about the underlying topographic relief.

In fact, the received waveform can be regarded as the convolution of the lidar system response and target response within a laser footprint [32,33]. However, the lidar system response represents the convolution of the transmitted laser pulse and lidar receiver impulse response, which can introduce smoothing and broadening effects for the received waveform. As for the steeply mountainous areas, the ground return will comprise multiple components and each component should be broadened and smoothed by the lidar system response. As a result, the ground return cannot be expressed by a single Gaussian model and may be aliased with the vegetation return, which makes it difficult to extract ground component or leads to a false identification of the ground component [34]. Meanwhile, the target response waveform (TRW) over the vegetated region can accurately represent the distributions of the vegetation and ground within the laser footprint of a satellite lidar and characterize the transmission properties of a laser pulse. Hence, the conventional algorithms used to derive the canopy height percentiles over rugged mountain area have non-negligible shortcomings: (1) due to the difference between the received waveform and the TRW, the cumulative distributions of the received waveforms are so dissimilar from the TRWs that the derived height percentiles deviate from the actual values; and (2) due to the inaccurate ground elevation introduced by the false extraction of a ground peak or low-precision characterization of the Gaussian ground return from SRTM data, the derived height percentiles may have some offsets.

To overcome such shortcomings, this research aims to propose a novel algorithm to derive height percentiles based on the resolved TRW taken from the received waveform without a prior topographic relief. The detailed objectives and contributions include (1) extracting the TRW from the received waveform using a Richardson–Lucy deconvolution algorithm for eliminating the broadening and smoothing effect of the lidar system response; (2) identifying the ground return and extracting accurate ground elevation information from the TRW for retrieving the high-precision height percentiles based on the shape distribution of the TRW; and (3) comparing the derived height percentiles with those from the conventional algorithms for evaluating the performance of the proposed algorithm. Our proposed algorithm is expected to possess good applicability in the extraction of height percentiles for complicated ground waveform returns.

2. Materials and Methods

2.1. Study Site

In this study, we employ the GEDI lidar data over a steeply mountainous vegetated area to validate the proposed method. The study site is located in the Pahvant Mountains of central Utah, USA, with the geolocation center at 38°56′55″N, 112°10′53″W. The location of the study site and four selected tracks of the GEDI lidar are presented in Figure 1. The ground tracks of the GEDI lidar are marked by red and green lines. Each ground track has an along-track length of around 18 km. The ground track information is listed in Table 1.
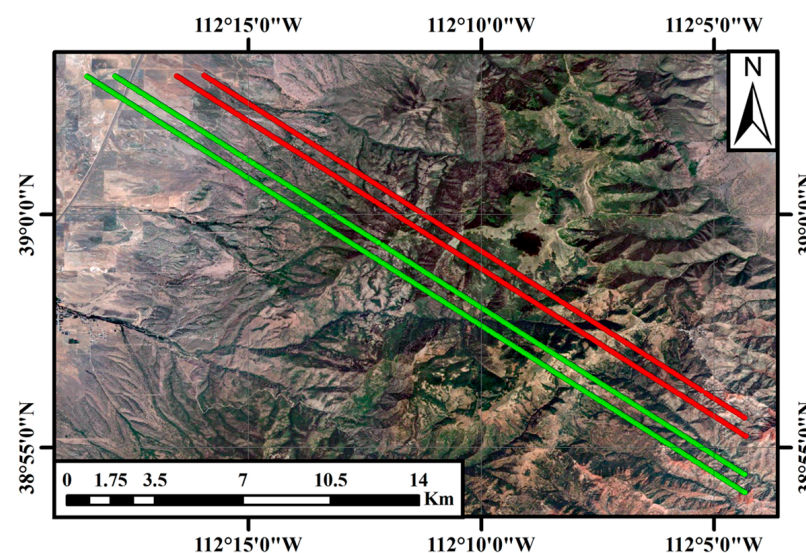
Figure 1. Location of the study site in the Pahvant Mountains of central Utah, USA, and the ground track data trajectories of the GEDI lidar. The green lines represent the ground tracks produced by GEDI full-power laser beams and the red lines represent the ground tracks produced by GEDI splitting laser beams.

Table 1. Topographic information within laser footprints for the selected ground tracks.

Beam Type	Ground Track	Laser Shots	Surface Slope (°)	Elevation (m)
Splitting beam	BEAM 0010	286	1.59~45.09	1785.9~2828.1
Splitting beam	BEAM 0011	291	2.28~41.60	1844.0~2787.4
Full-power beam	BEAM 0101	279	1.37~42.85	1918.7~2660.5
Full-power beam	BEAM 0110	296	1.36~63.15	1842.7~2876.2

The primary land covers of the study site are the rugged mountains and the wildwoods including the pine, cypress, other cold-resistant tree species, and some shrubs. Plant diversity is abundant, and the tree height distribute between 3 m and 27 m, indicating

the universality of the study. The local airborne lidar data are available in the study site. These indicate that the selected topography has a high relief with elevations ranging from 1785 m to 2829 m and surface slopes varying from 1° to 63°. The topography in the study site is rugged and complex, including flat areas, small-slope areas and large-slope areas, which may bring great limitations to the traditional height percentile extraction algorithms. The topographic information within a laser footprint and laser shots from four ground tracks are shown in Table 1. The complicated topography and biological diversity make the selected site suitable for evaluating the performance of the proposed algorithms of height percentiles.

2.2. Datasets

2.2.1. GEDI Lidar Dataset

The GEDI lidar onboard the international space station is designed to investigate the vertical forest structure for a better understanding of the role of forests in the global carbon cycle. The GEDI lidar comprises three lasers with a 16 ns pulsewidth and 242 Hz repetition frequency at a wavelength of 1064 nm [35,36]. The beams from two full-power laser beams, numbered from BEAM 0101 to BEAM 1011, produce four parallel ground tracks, and the splitting beams released from one coverage laser, numbered from BEAM 0000 to BEAM 0011, also produce four parallel ground tracks. The selected four ground tracks in this study root from two full-power laser beams and two splitting beams, which are marked with green and red lines, as illustrated in Figure 1, respectively. Adjacent laser footprints for the selected tracks have intervals of 60 m in the along-track direction and 600 m in the across-track direction.

The GEDI L1B standard data product provides the precise geographical coordinates for all laser footprints, the geolocated received waveforms, and corrected transmitted waveforms with a temporal resolution of 1 ns [37]. In this study, the selected GEDI L1B version 2 lidar datasets for about 1200 laser footprints were acquired on 20 May 2020 from Earth Science Data Systems (https://search.earthdata.nasa.gov/search?q=GEDI), accessed on 10 September 2023.

2.2.2. Airborne Lidar Dataset

The airborne lidar datasets used in this study were mainly collected from U.S. Geological Survey (https://apps.nationalmap.gov/downloader), accessed on 10 September 2023, and gathered with the Optech Galaxy PRIME lidar sensor by the Aerial Department of the Aero-Graphics on 4 August 2020. This airborne lidar dataset has the closest temporal window to the selected GEDI data in terms of maintaining the phenological consistency. The airborne lidar was mounted onboard the Bell helicopter and provided a scan line with a full scanning angle of 46°. The instrument can operate at a laser pulse repetition frequency of 300 kHz and a scanning frequency of 55.6 Hz. With a flight altitude of 1.6 km, the absolute horizontal and elevation errors of footprints on the ground are less than 0.16 m and 0.25 m, respectively. The minimum interval distance of the discrete points is less than 0.7 m and the average point density is around 3.3 pulses per square meter. The high-density airborne lidar data can accurately represent three-dimensional structures of the targets over the study site and provide a reference for simulating the actual target response.

2.3. Methods

In this study, we propose extracting the TRW from the received waveform and utilizing the cumulative distribution of the TRW to derive the height percentiles. Meanwhile, the reference height percentiles calculated from the airborne lidar data are introduced to evaluate the accuracy of the height percentiles derived from the TRW. The detailed flowchart of the proposed method is illustrated in Figure 2. The methodology comprises four processing steps which are described as the following Sections 2.3.1–2.3.4.
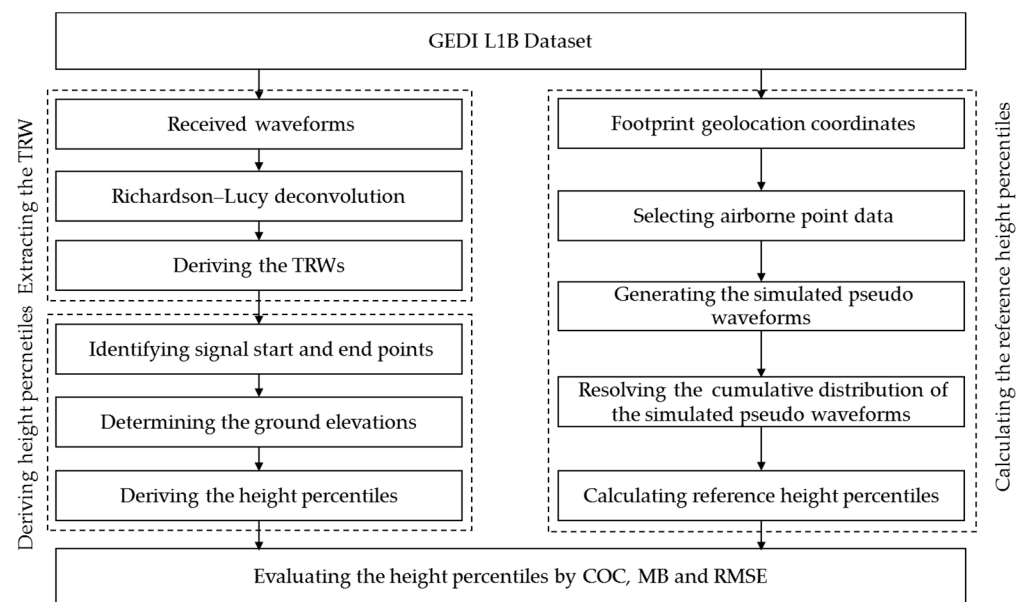
Figure 2. Detailed flowchart of the proposed algorithm. The procedure includes four processing steps: extracting the TRW from the received waveform; deriving height percentiles based on the resolved TRW; calculating the reference height percentiles from the simulated pseudo-waveform; and evaluating height percentiles.

2.3.1. Extracting the TRW from the Received Waveform

When a transmitted laser pulse projects into the forest, vertical canopy structures, including foliage, stems, branches trunks and the underlying ground, would definitely reflect the laser energy back to the satellite lidar receiver. The strength of the received waveform at a given depth of the canopy depends on the amount of the laser illumination penetrating into that depth and the reflectivity of the intercepted surface. Due to the pulse-broadening and smoothing effects of the lidar system response, the received waveform over the rugged mountain area would be a mixture of the overlapping vegetation and ground components. In addition, each component may deviate from the conventional Gaussian distribution and present an asymmetrical shape. It is very difficult to exactly extract all waveform components from the received waveform via the waveform decomposition method.

Owing to the convolution relationship between the TRW and lidar system response, we can utilize the deconvolution algorithm to eliminate the influence of the lidar system response on the received waveform. The Richardson–Lucy deconvolution algorithm based on Bayes' theorem has proved to be an effective time-domain iterative algorithm [38]. The iterative formation can be given by

$$m^{i+1}(t) = m^i(t)\left[\frac{R(t)}{m^i(t) * h(t)} * s(-t)\right], \tag{1}$$

where $m^i(t)$ is the resolved TRW after i-th iterations; $R(t)$ is the denoising received waveform obtained using a Gaussian filter [39]; and $s(t)$ represents the lidar system response that can be expressed by the convolution of the transmitted laser pulse and the receiver impulse response. In general, the lidar system response can be replaced by the received waveform returned from a flat ground surface in a clear sky [40].

The resolved TRW is heavily dependent on iterative times. Excessive iterative time can promote the accuracy of the deconvoluted TRW, but increases the running time of the

algorithm. As a result, an adaptive criterion is proposed to determine the iterative time in this study as

$$N = \left\{ i \left| \sqrt{\sum_{j=1}^{M} \frac{\left[W^i(t_j) - R(t_j)\right]^2}{M \cdot A^2}} < \delta_{th} \right. \right\}, \tag{2}$$

where $W^i(t)$ represents the received waveform after *i*-th iterations, as $W^i(t) = m^i(t) * h(t)$; M is the waveform sampling length; A is the maximal amplitude of the received waveform $R(t)$; and δ_{th} is the predefined threshold for determining the optimal iterative time. In this study, the threshold δ_{th} is set to be 1%.

2.3.2. Deriving the Height Percentile Based on the TRW

An illustration of the received waveform and the corresponding TRW is presented in Figure 3. The blue curve is the received waveform over the vegetated area and the red curve is the resolved TRW. It is apparent that both the shape and signal duration of the received waveform are different from those of the TRW, which may result in the difference in height percentiles derived from these two waveforms.
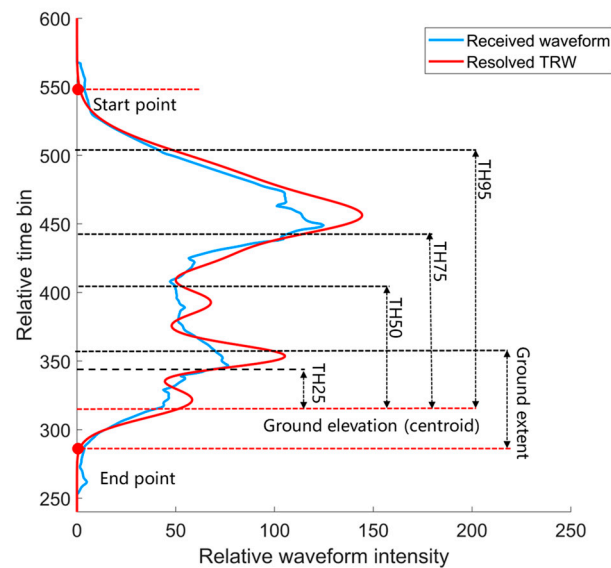
Figure 3. Illustration of the received waveform (blue curve) and TRW (red curve) extracted from the received waveform, and the definition of height percentiles from the TRW.

The calculation of the height percentile based on the TRW requires the elevations of the TRW signal start and end points and the ground elevation. The elevations of the TRW signal start and end points are identified by directly seeking the highest and lowest locations where the TRW amplitudes are greater than one percent of the maximums of the TRW, respectively. In this paper, we define the elevations of the signal start and end points as *sigsp* and *sigep*. Based on a minimum height threshold of 4.6 m, drawing from the US Forest Service's definition of a tree [41], we predefine that the ground's signal extent is the 4.6 m elevation extent above the end point (*sigep*). Hence, the ground elevation (*grd*) can be obtained by calculating the centroid of the TRW components within the ground signal extent, which can be expressed by

$$grd = \frac{\sum_{sigep}^{sigep+4.6} m(h) \cdot h}{\sum_{sigep}^{sigep+4.6} m(h)}, \tag{3}$$

where h represents the elevation of each bin of the resolved TRW. The relationship between the elevation and the sampling time in Equation (1) satisfies as $h = ct/2$, where c is the light

speed in a vacuum. The cumulative distribution of the normalized TRW energy between the start and end points is employed to generate the cumulative TRW, as presented by

$$\text{CM}_{\text{TRW}(h)} = \frac{\sum_{h}^{sigsp} m(h)}{\sum_{sigep}^{sigsp} m(h)} \tag{4}$$

The height metrics of the energy percentiles of 25%, 50%, 75%, and 95% for the cumulative TRW distribution are defined as TH25, TH50, TH75, and TH95 by their vertical distances relative to the ground elevation. Here, the energy percentiles represent the laser pulse energy levels distributed in different strata of the vegetation within a laser footprint.

2.3.3. Calculating the Reference Height Percentiles

The simulated pseudo-waveform is indicative of the vertical distribution of reflecting surfaces in the forest canopy and ground. Therefore, it is expected to closely resemble the actual target response. The simulated pseudo-waveform can be generated from the airborne lidar points and spatial energy distribution of the outgoing laser pulse on the ground. In our simulation, only the airborne lidar points within the satellite lidar field of view (FOV) are effectively acceptable. If the spatial energy distribution within a satellite lidar laser footprint is assumed to be a Gaussian function, the simulated pseudo-waveform can be described as [42,43]

$$\text{SW}(t) = \sum_{i \in U} I_i \exp\left[-\frac{(x_i - x_0)^2 + (y_i - y_0)^2}{2r^2}\right], \tag{5}$$
$$U = \left\{ i \middle| (x_i, y_i, z_i) \in \text{FOV and } \left|\frac{2z_i}{c} - t\right| \le \frac{\Delta t}{2} \right\}.$$

where I_i is the intensity of each airborne lidar point, (x_i, y_i, z_i) are the coordinates of each airborne lidar point, (x_0, y_0) denote the satellite lidar footprint center, r is the satellite lidar footprint radius (defined as half of $e^{-1/2}$ of the maximum), Δt is the temporal resolution of the received waveform, and c is light speed in vacuum.

Just like deriving the height percentiles from the TRW, the reference height percentiles can be calculated by the cumulative distribution of the normalized pseudo-waveform energy. The key difference is that the reference ground elevation is the weighted mean elevation of airborne lidar points within a laser footprint [42]. We define the height metrics of the energy percentiles of 25%, 50%, 75%, and 95% for the cumulative pseudo-waveform distribution as RH25, RH50, RH75, and RH95, respectively.

2.3.4. Evaluating the Height Percentiles by Different Methods

The performance of the height percentiles derived based on the TRW in this paper is compared with the height percentiles extracted by the conventional Gaussian decomposition (GD) method and the slope-adaptive waveform metrics (SWM) [23,25]. We introduce evaluation metrics, including the coefficient of correlation (COC), the mean bias (MB), and the root-mean-square error (RMSE), to assess the differences between the above height percentiles and the reference height percentiles. The COC, MB, and RMSE are given by

$$\text{COC} = \frac{\sum_{i=1}^{n} (\text{TH}_i - \text{THM})(\text{RH}_i - \text{RHM})}{\sqrt{\sum_{i=1}^{n}(\text{TH}_i - \text{THM})^2} \sqrt{\sum_{i=1}^{n}(\text{RH}_i - \text{RH}_i - \text{RHM})^2}}, \tag{6}$$

$$\text{MB} = \frac{1}{n}|\text{TH}_i - \text{RH}_i|, \tag{7}$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\text{TH}_i - \text{RH}_i)^2}{n - 1}}. \tag{8}$$

n denotes the number of the selected footprints, TH_i and RH_i denote the derived height percentile and the reference height percentile for i-th laser footprint, and THM and

RHM are the means of producing all derived height percentiles and reference height percentiles, respectively.

3. Results

3.1. Calculation and Analysis on the TRW

The GEDI laser footprint geolocations in the GEDI L1B data product may deviate from the actual positions due to laser pointing errors. Hence, in this study, the horizontal geolocations of all GEDI laser footprints are corrected via the waveform matching method [36]. We randomly choose nine laser footprints at the BEAM 0110 ground track and simulate the pseudo-waveforms by using the airborne lidar data within the corrected footprints. Furthermore, nine groups of the TRWs are extracted from the corresponding received waveforms based on the proposed method in Section 2.3.1. We provide visual comparisons of the received waveforms, the pseudo-waveforms, and the resolved TRWs for nine laser footprints in Figure 4.
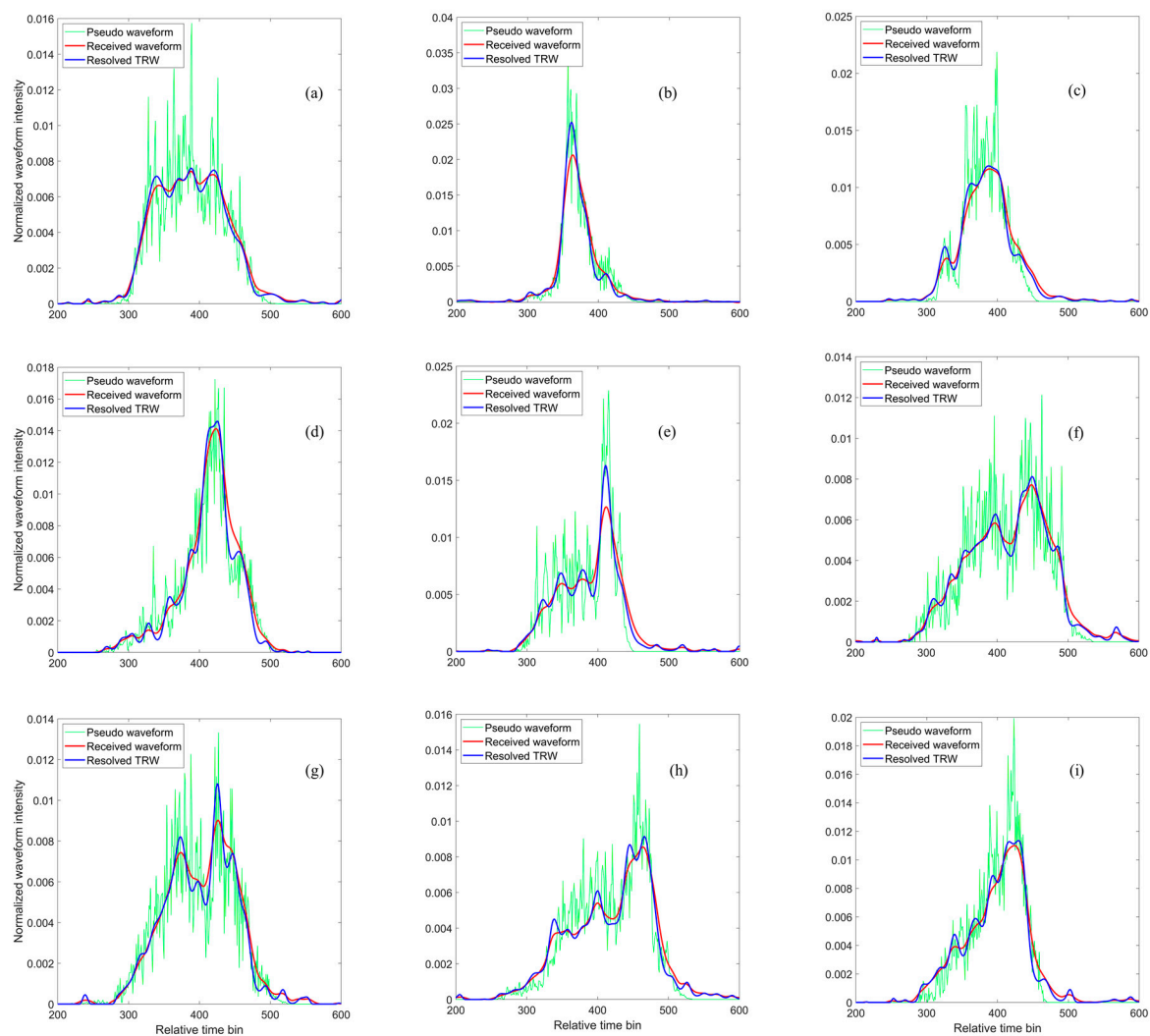
Figure 4. Comparisons of the received waveforms, the pseudo-waveforms, and the resolved TRWs for nine randomly selected laser footprints, (**a**–**i**) are the waveforms corresponding to nine randomly selected laser footprints. All waveforms are normalized by their own total energies. The start points of the relative time of all waveforms correspond to the first points of GEDI waveform, and relative time interval for each bin is 1 ns.

As shown in Figure 4, the received waveform shapes present multimodal distributions with multiple waveform components due to the complexity of vertical forest structures. Each waveform component is no longer described by the traditional Gaussian pattern. Meanwhile, the resolved TRWs are different from the received waveforms but closer to the pseudo-waveforms. This suggests that the resolved TRW can exactly characterize the laser transmission property into the forest. Furthermore, the ground components of the resolved TRW can be easily separated from the overlapping components of the received waveforms in Figure 4h,i, which is favorable for identifying the ground returns. The statistical metrics, including the COCs and total biases among the resolved TRW, the received waveform, and the pseudo-waveform, are calculated and listed in Table 2.

Table 2. Statistical metrics among the resolved TRW, the received waveform, and the pseudo-waveform for nine randomly selected laser footprints at the BEAM 0110 ground track.

Footprint Number	COC		Total Bias	
	TRW vs. Pseudo Waveform	Received Waveform vs. Pseudo-Waveform	TRW vs. Pseudo Waveform	Received Waveform vs. Pseudo-Waveform
60	0.94	0.89	0.0890	0.2623
70	0.97	0.94	0.0626	0.2710
92	0.94	0.90	0.0668	0.3060
119	0.95	0.93	0.0863	0.2512
123	0.94	0.82	0.0781	0.3858
156	0.93	0.87	0.1005	0.2879
166	0.94	0.90	0.0989	0.2704
180	0.94	0.91	0.1042	0.2676
244	0.96	0.93	0.0644	0.2417

The average COC between the resolved TRWs and the pseudo-waveforms in Table 2 is approximately 0.95, which indicates that the resolved TRWs have very strong correlations with the pseudo-waveforms. In addition, the maximum of total biases is less than 0.1042. Meanwhile, it is noticed that the received waveforms also correlate very well with the pseudo-waveforms, with an average COC of round 0.90. However, the total biases between the received waveforms and the pseudo-waveforms are obviously promoted, being almost 3.5 times the total biases between the TRWs and the pseudo-waveforms. Such a slight discrepancy between the TRWs and the pseudo-waveforms demonstrates that the resolved TRWs more closely resemble the pseudo-waveforms than the received waveforms. To thoroughly expound the relationship between the resolved TRWs and the pseudo-waveforms in the study site, we provide their histograms of the COC, total bias and RMSE for all selected footprints, as presented in Figure 5.
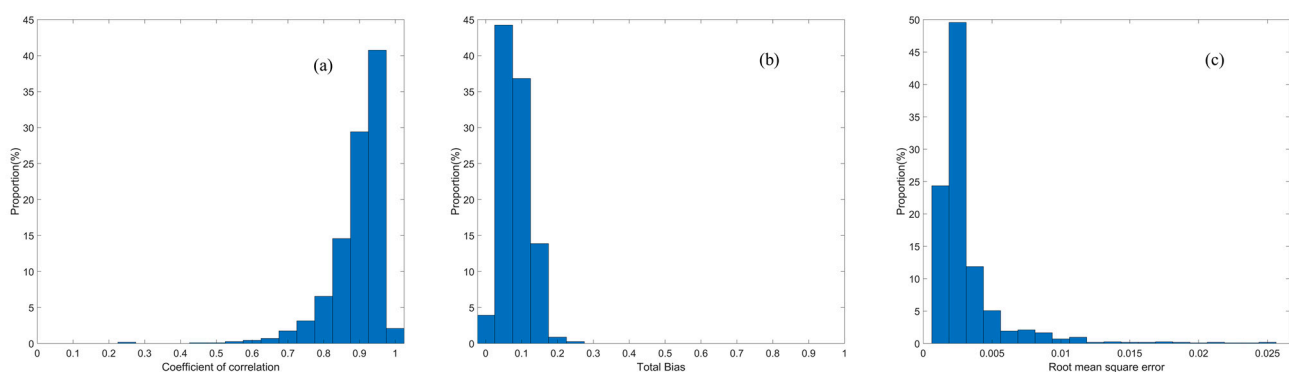
Figure 5. Histograms of the statistical metrics about the relationship between the resolved TRWs and the pseudo-waveforms for the selected footprints in the study site. (**a**) COC; (**b**) Total bias; (**c**) RMSE.

The correlation coefficients in Figure 5a are variable for each measurement due to the diverse forest structure within each laser footprint. We partition correlation coefficients into five segments with a range from 0 to 1 and an interval of 0.2 based on the definition of the correlation strength. When correlation strengths vary from very weak to very strong, we discover that the COC proportions rapidly increase and take values of 0.00%, 0.18%, 0.26%, 2.53%, and 97.03%. This indicates that 99.56% of the resolved TRWs correlate very well with the pseudo-waveforms. Meanwhile, the total biases in Figure 5b are distributed within a narrow range, and 94.86% of total biases are less than 0.15. The RMSE in Figure 5c presents analogous distribution, and 97.99% of RMSEs are less than 0.005. The detailed statistical results of the COC, total bias, and RMSE for all footprints in the study site are enumerated in Table 3.

Table 3. Statistical results of the COC, total bias, and RMSE between the resolved TRWs and the pseudo-waveforms for all footprints in the study site.

Metrics	COC	Total Bias	RMSE
Maximum	0.99	0.2351	0.0310
Mean	0.92	0.0813	0.0016
Minimum	0.29	0.0096	0.0006

The mean values of the COC, total bias, and RMSE in Table 3 are equal to 0.92, 0.0813, and 0.0016, respectively. Greater COC, smaller bias and RMSE reveal that the resolved TRWs have strong similarities with the pseudo-waveforms. Combining this with the histograms in Figure 5, we can conclude that the resolved TRWs are more applicable for the extraction of the height percentiles than the received waveforms.

3.2. Extraction and Analysis on Height Percentiles

Based on the cumulative distributions of the resolved TRWs and the received waveforms, we extract the height percentiles using our proposed algorithms, the Gaussian decomposition method (GD method) and the SWM method. These derived height percentiles are directly compared with the reference height percentiles from airborne lidar data to evaluate their differences. To investigate the contributions of the splitting beam and full-power beam to the height percentiles, we separate the height percentile comparisons into two groups. The scatter plots of the reference height percentiles of RH25, RH50, RH75 and RH95 versus the derived height percentiles of TH25, TH50, TH75, and TH95 from different algorithms are illustrated in Figures 6 and 7 for two groups of laser beams, respectively.

The results in Figures 6 and 7 indicate that the derived height percentiles from the three abovementioned algorithms present some differences with the reference height percentiles. When the energy percentiles vary from 25% to 95%, the gradients of the best fitting lines rapidly decline and nearly approach 1. This suggests that the derived height percentiles may trend towards the reference values with the increase in the energy percentiles. For quantitative comparisons, the evaluation metrics are listed in Table 4, including the COC, MB, and RMSE for the derived height percentiles calculated by different algorithms.

The COC values in Table 4 present a strong uptrend when the energy percentiles increase from 25% to 50%. The derived height percentiles from the GD method and our proposed algorithm have stronger correlations with the reference height percentiles than the SWM method. In addition, the accuracies of the derived height percentiles from our proposed algorithm have been significantly improved compared with the GD and SWM methods. As for the splitting beams, the MB values decrease by the mean values of 2.52 m and 2.22 m, and the RMSE values decrease by the mean values of 3.12 m and 2.51 m, respectively. While there are major declines in the MB and RMSE values due to the GD method for the full-power beams, the mean reductions in the MB and RMSE values are 1.19 m and 1.38 m relative to the splitting beams, respectively. It may be due to the higher energies and signal-to-noise ratios of the received waveforms for full-power beams. The

total MB and RMSE values calculated by our proposed algorithm for the selected tracks in the study site have average decreases of 1.68 m and 2.32 m and 1.96 m and 2.72 m.
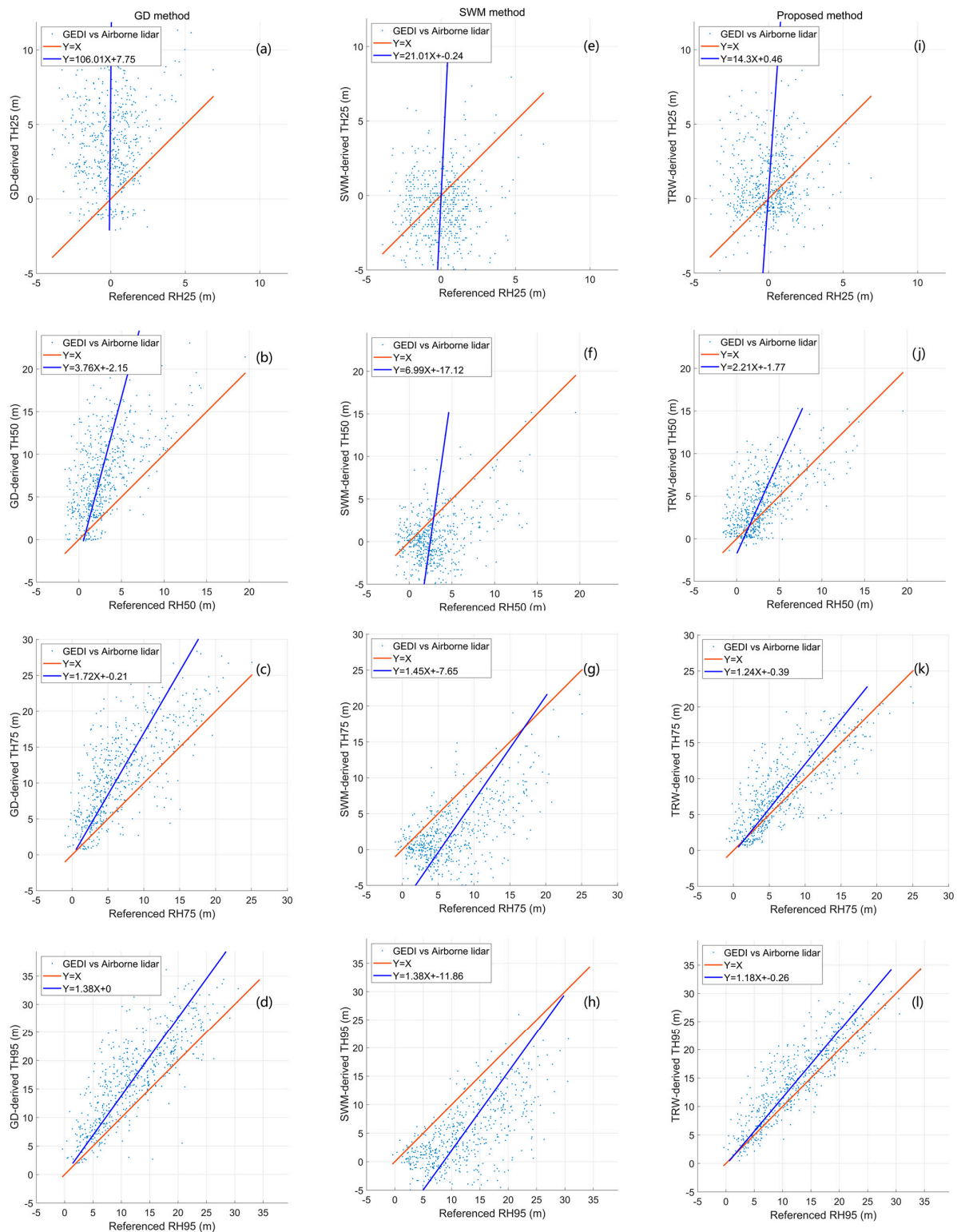
Figure 6. Scatter plots of reference height percentiles of RH25, RH50, RH75, and RH95 versus derived height percentiles of TH25, TH50, TH75, and TH95 from different algorithms for the GEDI splitting beams. (**a–d**) are the Gaussian decomposition method; (**e–h**) are slope-adaptive waveform methods; and (**i–l**) are the proposed method. Red lines and blue lines represent the 'Y = X' and the best fitting lines, respectively.
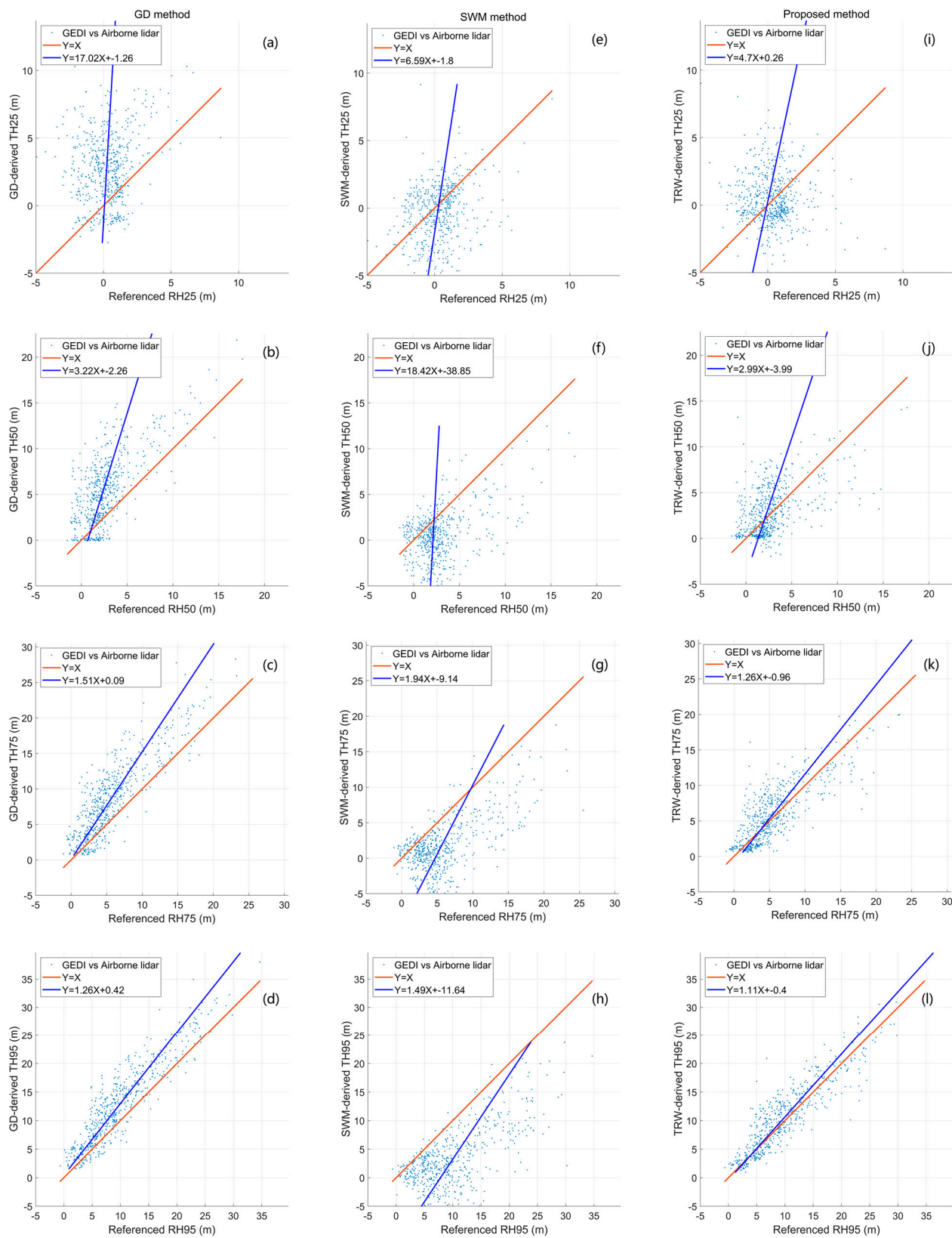
Figure 7. Scatter plots of reference height percentiles of RH25, RH50, RH75, and RH95 versus derived height percentiles of TH25, TH50, TH75, and TH95 from different algorithms for the GEDI full-power beams. (**a**–**d**) are results from Gaussian decomposition methods; (**e**–**h**) are from slope-adaptive waveform methods; and (**i**–**l**) are from the proposed method. Red lines and blue lines represent the 'Y = X' and the best fitting lines, respectively.

Table 4. Statistical metrics between the reference height percentiles and the derived height percentiles from different algorithms.

Beam	Method	COC				MB (m)				RMSE (m)			
		25%	50%	75%	95%	25%	50%	75%	95%	25%	50%	75%	95%
Splitting beam	GD	0.17	0.72	0.74	0.82	4.57	4.91	5.04	5.23	5.82	6.27	6.50	6.80
	SWM	0.18	0.60	0.58	0.61	2.25	3.42	5.16	7.71	2.89	4.38	6.36	9.29
	TRW	0.18	0.74	0.81	0.88	2.03	2.20	2.49	2.95	2.68	2.94	3.35	3.93
Full-power beam	GD	0.42	0.82	0.87	0.92	3.05	3.18	3.28	3.39	3.91	4.07	4.15	4.24
	SWM	0.49	0.78	0.57	0.57	1.84	3.07	4.67	7.29	2.53	4.12	6.04	9.09
	TRW	0.43	0.79	0.85	0.91	1.95	2.02	2.04	2.14	2.60	2.73	2.69	2.85

There are two principal reasons leading to the inferior derived height percentiles for the GD and SWM methods. The first reason is that the received waveform should not be modeled by one or multiple Gaussian functions, especially for ground components over the rugged mountain area. The second reason is that the ground elevation should not be determined by the last ground component peak from the GD method and the end point of the fitted ground component from the SWM method.

4. Discussion

Although the derived accuracies of the height percentiles from the resolved TRW were improved, there were still some differences between the derived height percentiles and the reference height percentiles from the pseudo-waveform, as presented in Figures 6 and 7. According to the abovementioned results in Section 3.1, the resolved TRWs closely resembled pseudo-waveforms. Hence, we speculate that the deviations in the derived height percentiles probably originated from the inaccurate ground elevation. To investigate the hypothesis, we specially selected one of laser footprints where the derived height percentiles had the average differences of 3.70 m for our proposed algorithm and 8.87 m for the GD method in Figure 6. The received waveform, the pseudo-waveform, the resolved TRW, and airborne point cloud within the selected footprint are displayed in Figure 8.
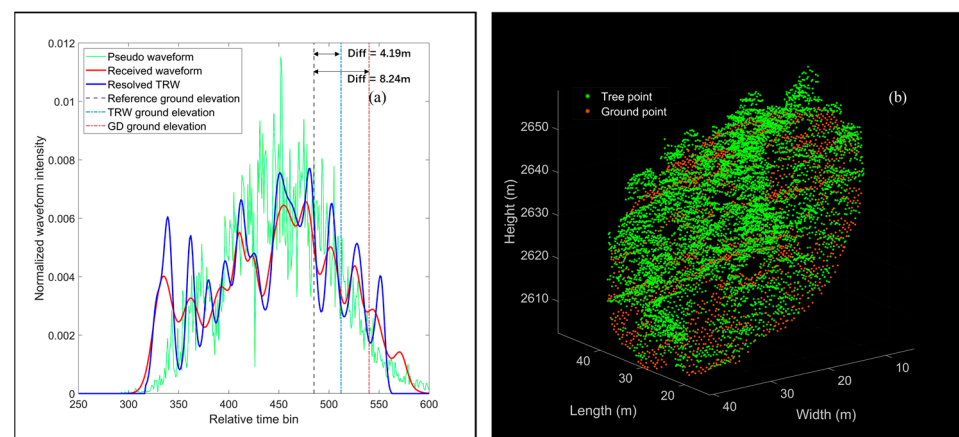
Figure 8. Illustration of satellite lidar data and airborne lidar data within a selected laser footprint. (**a**) GD waveform, pseudo-waveform and resolved TRW; (**b**) airborne point cloud. All waveforms are normalized by the total energy.

As shown in Figure 8, the received waveform comprises several vegetation and ground components due to the complicated forest structure within the selected laser footprint. Especially, there is an apparent inclination of around 53.08° and rugged relief for the ground in Figure 8b, leading to multiple components and a pulse-broadening effect in

the ground return. Even though the resolved TRW and the GD waveform seem to be basically in accordance with the pseudo-waveform, it is still impossible to exactly identify all ground components from the resolved TRW and GD waveform. Thus, the resolved ground elevations from the proposed algorithm and GD method deviate from the reference ground elevation. The locations of the resolved ground elevations are marked by blue and purple dot dashed lines and the reference ground elevation is marked by a black dashed line in Figure 8a. The differences of ground elevations for our proposed algorithm and GD method approximate to 4.19 m and 8.24 m, which are close to the average deviation (3.70 m and 8.87 m) of the derived height percentiles.

To explore the influence of ground elevation errors on the derived height percentiles produced by our proposed algorithm and GD method, we attempt to replace the resolved ground elevation by the reference ground elevation and recalculate the derived height percentiles. In this study, we define such recalculated height percentiles as the renewed height percentiles. The differences between the derived height percentiles, renewed height percentiles and reference height percentiles for the selected footprint are listed in Table 5.

Table 5. The differences between the derived height percentiles, renewed height percentiles and reference height percentiles by the GD method and proposed algorithm for the selected footprint.

| Energy Percentile		25%	50%	75%	95%
Difference for derived height percentile (m)	GD method	6.32	7.53	9.62	12.02
Difference for renewed height percentile (m)		−0.87	0.67	1.42	2.82
Difference for derived height percentile (m)	Proposed method	1.64	2.84	5.54	4.79
Difference for renewed height percentile (m)		0.27	−0.06	0.63	0.42

As seen from Table 5, the differences in the derived height percentiles versus reference height percentiles range from 1.64 m to 4.79 m (by the proposed method) and 6.32 m to 12.02 m (by the GD method), respectively. However, the differences between the corresponding renewed height percentiles versus reference height percentiles rapidly decrease and only range from −0.06 m to 0.63 m and −0.87 m to 2.82 m, respectively. The accuracies of the renewed height percentiles are greatly improved. To further validate the effect of the reference ground elevations on the height percentiles by the proposed algorithm, we plot the renewed height percentiles for all laser footprints as presented in Figure 9. The corresponding evaluation metrics of the derived height percentiles are listed in Table 6.

Table 6. Statistical metrics between the reference height percentiles and the renewed height percentiles from the GD method and proposed algorithm for all footprints.

Method	COC				MB (m)				RMSE (m)			
	25%	50%	75%	95%	25%	50%	75%	95%	25%	50%	75%	95%
GD	0.69	0.92	0.97	0.97	1.47	1.31	1.40	1.56	1.73	1.53	1.85	2.16
Proposed	0.72	0.92	0.97	0.97	1.12	1.06	1.15	1.30	1.32	1.25	1.58	1.74

Compared with the derived height percentiles by the proposed algorithm in Figures 6 and 7, the renewed height percentiles in Figure 9 present stronger correlation strength with the reference height percentiles. The MB and RMSE of the GD method decrease by mean values of 63.24% and 63.20%, and the MB and RMSE of the proposed method decrease by mean values of 47.90% and 50.82%, respectively. This suggests that the renewed height percentiles obtained by the GD method have preferable promotion than those acquired by our proposed algorithm. However, the accuracies of the renewed height percentiles by the proposed algorithm are still higher than those taken from the GD method. The average MB and RMSE decreased by 0.28 m and 0.35 m. The comparisons demonstrate that accurate ground elevation plays a significant role in determining the derived height percentiles,

especially using the GD method [23,26,27]. Nevertheless, the derived height percentiles produced by the GD method are more susceptible to the ground elevation. Therefore, it is worthwhile to carry out research on the extraction of the ground elevation using the resolved TRW or received waveform over rugged mountainous areas.
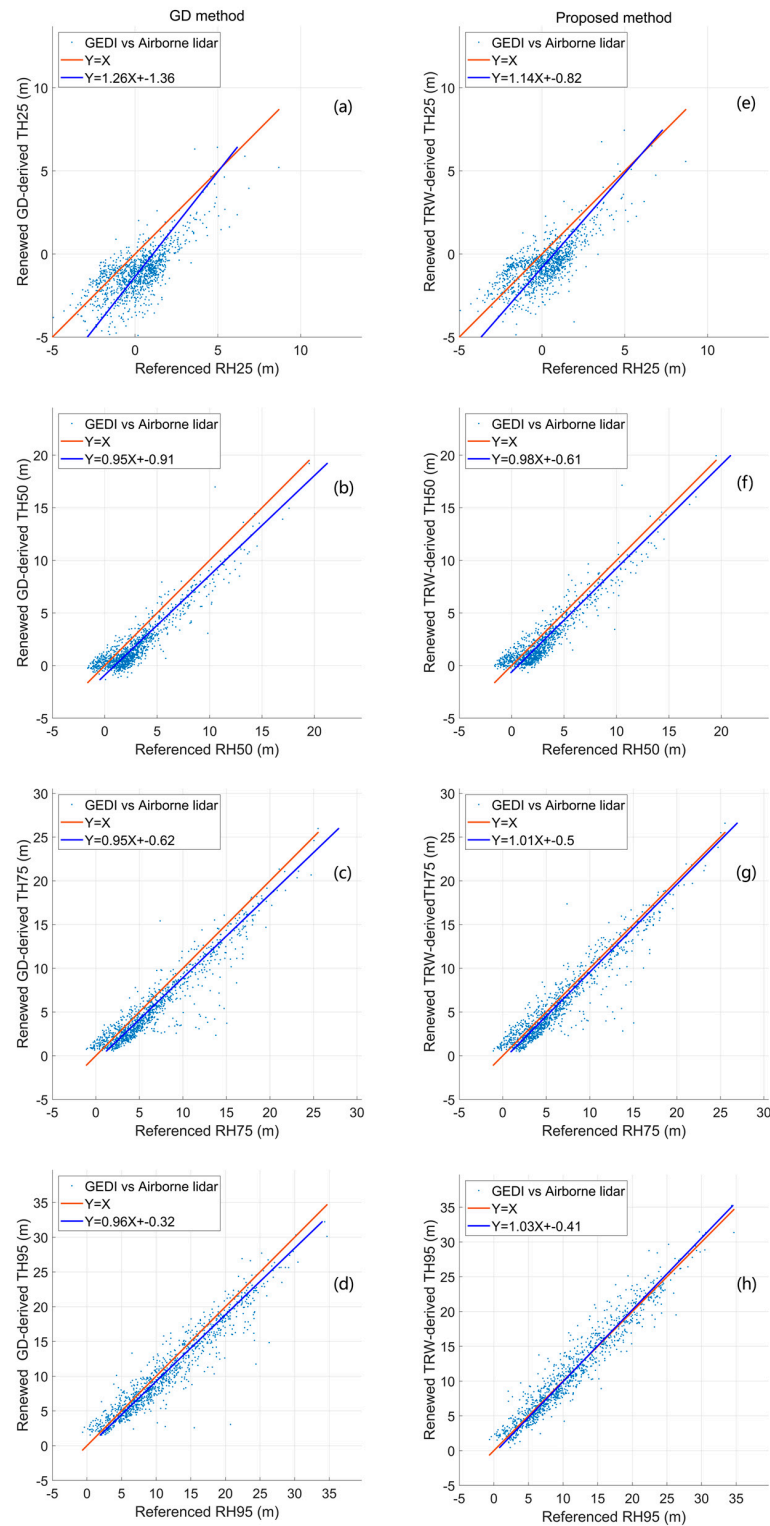
Figure 9. Scatter plot of renewed height percentiles from the GD method and proposed algorithm vs. reference height percentiles. (**a**,**e**) Renewed TH25 vs. RH25; (**b**,**f**) Renewed TH50 vs. RH50; (**c**,**g**) Renewed TH75 vs. RH75; (**d**,**h**) Renewed TH95 vs. RH95.

5. Conclusions

We propose a novel algorithm to derive the height percentiles using the TRW extracted from the received waveform with the Richardson–Lucy deconvolution algorithm. The TRWs resolved from the GEDI lidar waveforms in the study site have strong correlations and slight deviations with the actual TRWs (i.e., pseudo-waveforms) simulated by the coincident airborne lidar data. The resolved TRW can characterize laser transmission properties into the forest since the broadening and smoothing effects of a lidar system response on the received waveform are eliminated. More accurate target response and better identification of the ground return can provide higher-precision height percentiles. Compared with the Gaussian decomposition method and the slope-adaptive waveform metrics method, the accuracies of the derived height percentiles from the proposed algorithm are greatly improved. The MB and RMSE values of the derived height percentiles decrease by the mean values of 1.68 m and 2.32 m and 1.96 m and 2.72 m, respectively. Our proposed algorithm presents a good applicability in the extraction of the height percentiles for the complicated ground returns.

Based on the accuracy of height percentiles in this study, we discover that the inadequate identification of all ground components from the TRW over the rugged mountain area is the significant factor degrading the accuracies of the derived height percentiles. Although our proposed algorithm can improve the accuracy of ground elevation, there are some deviations relative to the ground truth. In future research, the question of how to extract a higher-precision ground elevation from the resolved TRW should be further explored.

Author Contributions: Conceptualization, H.Z. and H.S.; methodology, H.S., H.Z. and Q.Z.; software, H.S. and H.W.; validation, H.S., H.Z. and H.W.; formal analysis, H.Z. and H.S.; investigation, H.S.; resources, H.W. and H.S.; data curation, H.W.; writing—original draft preparation, H.Z. and H.S.; writing—review and editing, H.Z. and Y.M.; visualization, H.S.; supervision, H.Z. and S.L.; project administration, H.Z.; funding acquisition, H.Z. and Y.M. All authors have read and agreed to the published version of the manuscript.

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

References

1. Koetz, B.; Morsdorf, F.; Sun, G.; Ranson, K.; Itten, K.; Allgower, B. Inversion of a Lidar Waveform Model for Forest Biophysical Parameter Estimation. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 49–53.
2. Jonckheere, I.; Fleck, S.; Nackaerts, K.; Muys, B.; Coppin, P.; Weiss, M.; Baret, F. Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. *Agric. For. Meteorol.* **2004**, *121*, 19–35. [CrossRef]
3. Brodie, J.F.; Mohd-Azlan, J.; Chen, C.; Wearn, O.R.; Deith, M.C.M.; Ball, J.G.C.; Slade, E.M.; Burslem, D.F.R.P.; Teoh, S.W.; Williams, P.J.; et al. Landscape-scale benefits of protected areas for tropical biodiversity. *Nature* **2023**, *620*, 807–812. [CrossRef]
4. Jucker, T.; Gosper, C.R.; Wiehl, G.; Yeoh, P.B.; Raisbeck-Brown, N.; Fischer, F.J.; Graham, J.; Langley, H.; Newchurch, W.; O'Donnell, A.J.; et al. Using multi-platform LiDAR to guide the conservation of the world's largest temperate woodland. *Remote Sens. Environ.* **2023**, *296*, 113745.
5. Wulder, M.A.; Hall, R.J.; Coops, N.C.; Franklin, S.E. High Spatial Resolution Remotely Sensed Data for Ecosystem Characterization. *BioScience* **2004**, *54*, 511–521. [CrossRef]
6. Song, C.; Dickinson, M.B. Extracting forest canopy structure from spatial information of high resolution optical imagery: Tree crown size versus leaf area index. *Int. J. Remote Sens.* **2008**, *29*, 5605–5622. [CrossRef]
7. Lin, X.; Xu, M.; Cao, C.; Dang, Y.; Bashir, B.; Xie, B.; Huang, Z. Estimates of forest canopy height using a combination of ICESat-2/ATLAS data and stereo-photogrammetry. *Remote Sens.* **2020**, *12*, 3649. [CrossRef]

8. Adrah, E.; Jaafar, W.S.; Bajaj, S.; Omar, H.; Leite, R.V.; A Silva, C.; Cardil, A.; Mohan, M. Analyzing canopy height variations in secondary tropical forests of Malaysia using NASA GEDI. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *880*, 012031.

9. Duncanson, L.; Niemann, K.; Wulder, M. Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sens. Environ.* **2010**, *114*, 138–154. [CrossRef]

10. Hirschmugl, M.; Lippl, F.; Sobe, C. Assessing the Vertical Structure of Forests Using Airborne and Spaceborne LiDAR Data in the Austrian Alps. *Remote Sens.* **2023**, *15*, 664. [CrossRef]

11. Gupta, R.; Sharma, L.K. Mixed tropical forests canopy height mapping from spaceborne LiDAR GEDI and multisensor imagery using machine learning models. *Remote Sens. Appl. Soc. Environ.* **2022**, *27*, 100817. [CrossRef]

12. Wang, Y.; Fang, H.; Zhang, Y.; Li, S.; Pang, Y.; Ma, T.; Li, Y. Retrieval and validation of vertical LAI profile derived from airborne and spaceborne LiDAR data at a deciduous needleleaf forest site. *GISci. Remote Sens.* **2023**, *60*, 2214987. [CrossRef]

13. Harding, D.J.; Dabney, P.W.; Valett, S. Polarimetric, two-color, photon-counting laser altimeter measurements of forest canopy structure. In *International Symposium on Lidar and Radar Mapping 2011: Technologies and Applications*; SPIE: London, UK, 2011; Volume 8286.

14. Lee, S.; Ni-Meister, W.; Yang, W.; Chen, Q. Physically based vertical vegetation structure retrieval from ICESat data: Validation using LVIS in White Mountain National Forest, New Hampshire, USA. *Remote Sens. Environ.* **2011**, *115*, 2776–2785.

15. Maltamo, M.; Eerikäinen, K.; Packalén, P.; Hyyppä, J. Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry* **2006**, *79*, 217–229. [CrossRef]

16. Naesset, E. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sens. Environ.* **1997**, *61*, 246–253.

17. Mandl, L.; Stritih, A.; Seidl, R.; Ginzler, C.; Senf, C. Spaceborne LiDAR for characterizing forest structure across scales in the European Alps. *Remote Sens. Ecol. Conserv.* **2023**, *9*, 599–614. [CrossRef]

18. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.* **2021**, *253*, 112165. [CrossRef]

19. Næsset, E. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scand. J. For. Res.* **2004**, *19*, 164–179. [CrossRef]

20. Wang, S.; Liu, C.; Li, W.; Jia, S.; Yue, H. Hybrid model for estimating forest canopy heights using fused multimodal spaceborne LiDAR data and optical imagery. *Int. J. Appl. Earth Obs. Geoinform.* **2023**, *122*, 103431. [CrossRef]

21. Zhu, X.; Nie, S.; Wang, C.; Xi, X.; Wang, J.; Li, D.; Zhou, H. A noise removal algorithm based on OPTICS for photon-counting LiDAR data. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1471–1475.

22. Zhu, X.; Nie, S.; Wang, C.; Xi, X.; Lao, J.; Li, D. Consistency analysis of forest height retrievals between GEDI and ICESat-2. *Remote Sens. Environ.* **2022**, *281*, 113244. [CrossRef]

23. Chen, Q. Assessment of terrain elevation derived from satellite laser altimetry over mountainous forest areas using airborne lidar data. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 111–122.

24. Pang, Y.; Zhao, F.; Li, Z. Research on Average Tree Height Extraction from Airborne Lidar. *J. Remote Sens.* **2008**, *12*, 152–158.

25. Wang, Y.; Ni, W.; Sun, G.; Chi, H.; Zhang, Z.; Guo, Z. Slope-adaptive waveform metrics of large footprint lidar for estimation of forest aboveground biomass. *Remote Sens. Environ.* **2019**, *224*, 386–400.

26. 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.

27. 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. [CrossRef]

28. Hofton, M.; Blair, J.B.; Story, S.; Yi, D. *Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products*; University of Maryland: College Park, MD, USA, 2019; Volume 44.

29. Crockett, E.T.; Atkins, J.W.; Guo, Q.; Sun, G.; Potter, K.M.; Ollinger, S.; Silva, C.A.; Tang, H.; Woodall, C.W.; Holgerson, J.; et al. Structural and species diversity explain aboveground carbon storage in forests across the United States: Evidence from GEDI and forest inventory data. *Remote Sens. Environ.* **2023**, *295*, 113703. [CrossRef]

30. May, P.; McConville, K.S.; Moisen, G.G.; Bruening, J.; Dubayah, R. A spatially varying model for small area estimates of biomass density across the contiguous United States. *Remote Sens. Environ.* **2023**, *286*, 113420. [CrossRef]

31. Ni-Meister, W.; Rojas, A.; Lee, S. Direct use of large-footprint lidar waveforms to estimate aboveground biomass. *Remote Sens. Environ.* **2022**, *280*, 113147. [CrossRef]

32. Zhang, Z.; Xie, H.; Tong, X.; Zhang, H.; Tang, H.; Li, B.; Wu, D.; Hao, X.; Liu, S.; Xu, X.; et al. A Combined Deconvolution and Gaussian Decomposition Approach for Overlapped Peak Position Extraction from Large-Footprint Satellite Laser Altimeter Waveforms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2286–2303. [CrossRef]

33. Zhou, T.; Popescu, S.C.; Krause, K.; Sheridan, R.D.; Putman, E. Gold—A novel deconvolution algorithm with optimization for waveform LiDAR processing. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 131–150. [CrossRef]

34. Dubayah, R.O.; Sheldon, S.L.; Clark, D.B.; Hofton, M.A.; Blair, J.B.; Hurtt, G.C.; Chazdon, R.L. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *J. Geophys. Res. Biogeosci.* **2010**, *115*, G00E09.

35. Lang, N.; Kalischek, N.; Armston, J.; Schindler, K.; Dubayah, R.; Wegner, J.D. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. *Remote Sens. Environ.* **2022**, *268*, 112760. [CrossRef]

36. Lee, S.K.; Fatoyinbo, T.; Marselis, S.M.; Qi, W.; Hancock, S.; Armston, J.; Dubayah, R. Spaceborne data fusion for large-scale forest parameter estimation: GEDI LiDAR & Tandem-X INSAR missions. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: Piscataway, NJ, USA, 2019.

37. Xu, Y.; Ding, S.; Chen, P.; Tang, H.; Ren, H.; Huang, H. Horizontal Geolocation Error Evaluation and Correction on Full-Waveform LiDAR Footprints via Waveform Matching. *Remote Sens.* **2023**, *15*, 776.

38. Tai, Y.-W.; Tan, P.; Brown, M.S. Richardson-Lucy Deblurring for Scenes under a Projective Motion Path. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 1603–1618.

39. Duong, V.H. Processing and Application of ICESat Large Footprint Full Waveform Laser Range Data. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 8 June 2010.

40. Mallet, C.; Bretar, F. Full-waveform topographic lidar: State-of-the-art. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 1–16. [CrossRef]

41. Neuenschwander, A.; Pitts, K. The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sens. Environ.* **2019**, *221*, 247–259. [CrossRef]

42. Popescu, S.C.; Zhao, K.; Neuenschwander, A.; Lin, C. Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sens. Environ.* **2011**, *115*, 2786–2797.

43. Silva, C.A.; Saatchi, S.; Garcia, M.; Labriere, N.; Klauberg, C.; Ferraz, A.; Meyer, V.; Jeffery, K.J.; Abernethy, K.; White, L.; et al. Comparison of Small- and Large-Footprint Lidar Characterization of Tropical Forest Aboveground Structure and Biomass: A Case Study from Central Gabon. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3512–3526. [CrossRef]