



*Research article*

## Canopy Height Estimation by Characterizing Waveform LiDAR Geometry Based on Shape-Distance Metric

Eric Ariel L. Salas<sup>1,\*</sup> and Geoffrey M. Henebry<sup>2</sup>

<sup>1</sup> Department of Fish, Wildlife and Conservation Ecology, New Mexico State University, Las Cruces, NM 88003, USA

<sup>2</sup> Geospatial Sciences Center of Excellence (GSCE), South Dakota State University, Brookings, SD 57007-3510, USA

\* **Correspondence:** Email: [easalas@nmsu.edu](mailto:easalas@nmsu.edu); Tel: +1-575-646-2691

**Abstract:** There have been few approaches developed for the estimation of height using waveform LiDAR data. Unlike any existing methods, we illustrate how the new Moment Distance (MD) framework can characterize the canopy height based on the geometry and return power of the LiDAR waveform without having to go through curve modeling processes. Our approach offers the possibilities of using the raw waveform data to capture vital information from the variety of complex waveform shapes in LiDAR. We assess the relationship of the MD metrics to the key waveform landmarks—such as locations of peaks, power of returns, canopy heights, and height metrics—using synthetic data and real Laser Vegetation Imaging Sensor (LVIS) data. In order to verify the utility of the new approach, we use field measurements obtained through the DESDynI (Deformation, Ecosystem Structure and Dynamics of Ice) campaign. Our results reveal that the MDI can capture temporal dynamics of canopy and segregate generations of stands based on curve shapes. The method satisfactorily estimates the canopy height using the synthetic ( $r^2 = 0.40$ ) and the LVIS dataset ( $r^2 = 0.74$ ). The MDI is also comparable with existing RH75 (relative height at 75%) and RH50 (relative height at 50%) height metrics. Furthermore, the MDI shows better correlations with ground-based measurements than relative height metrics. The MDI performs well at any type of

waveform shape. This opens the possibility of looking more closely at single-peaked waveforms that usually carries complex extremes.

**Keywords:** Moment Distance Index; MD framework; waveform LiDAR; MDI; canopy height; LVIS; shape metric

---

## 1. Introduction

Canopy height has been regarded as an excellent predictor of the total mass of vegetation existing in a stand [1,2], for example, biomass and volume [3–6]. The waveform LiDAR has been utilized to accurately estimate canopy heights [7–9], utilizing the intricacies of the intercepted surfaces or the proportion of the canopy complexity that the waveform dataset offers. Lim et al. [10] took advantage of the peaks of the two most prominent modes in the amplitude waveform to estimate the height by translating the difference of the elapsed time between them. The modes signify the signal start after noise and the centroid of the last peak [11,12]. The detection of the prominent peaks, however, is sometimes problematic and insufficient.

The underlying waveform assumption is that the signal transmitted from the sensor behaves like that of a Gaussian structure [13] and that the received signal is a mixture of Gaussian distributions [12,14]. Hence, Gaussian functions have been used to decompose the waveform [13,15] for estimating vegetation height. The model is most sufficient for large-footprint LiDAR data [13,16], but not with small and medium-sized footprints. In addition, approximating the waveforms using a sum of Gaussians may not be an accurate representation depending on the application and the target [17]. Depending on the LiDAR system, the transmitted signal is not often Gaussian but can be somewhat contorted – asymmetric [18,19], flattened or peaked [20], and the asymmetry of the peaks hence may not be correctly adjusted [21]. Hancock et al. [22] suggested that this might be due to shadows from heterogeneous canopy cover. In the presence of noise in the waveform, there is a problem in finding the location and determining the number of Gaussians by inflection points as it relies in derivative; even smoothing the data does not help [23]. Also, automated algorithms may have difficulty interpreting peaks, especially with weak returns [24], oftentimes requiring the use of human visual interpretation.

Several other modeling functions have been carried out with the goal to isolate as many waveform peaks from the distorted returned signal as possible. One is the non-linear least-squares approach using Levenberg-Marquardt optimization algorithm [25,26]. The model has been utilized in satellite analysis such as Geoscience Laser Altimeter System (GLAS) waveforms [27], and airborne laser scanning altimetry [28].

The Maximum Likelihood approach using the Expectation Maximization algorithm [29] is a general technique to fit the signal to an assortment of Gaussian functions to uncover and

parameterize the peaks. Nevertheless, the gradient computation necessary in such models limits both the introduction of physical knowledge on the waveforms and the type of the chosen function [21]. The stochastic approach using Reversible Jump Monte Carlo Markov Chain [30] is another method that could fit terrestrial LiDAR waveforms with specific modeling functions [21,31]. Even so, existing waveform optimization fitting schemes rely strongly on initial parameters or priori assumption (e.g. [32]). This implies that various types of parameters must be approximated extensively to find the best model and prevent faulty outcomes. In addition, there is a need to investigate new robust possibilities of waveform signal processing to capture vital information from variety of complex waveform shapes in LiDAR.

While current waveform optimization fitting schemes rely strongly on initial parameters, in this study, we explore the possibilities of utilizing the raw waveform (to retain richness of the data). The waveform LiDAR system has the ability to record many returns per emitted pulse, as a function of time, within the vertical structure of the illuminated object. Key features of the waveform such as the shape and power are directly related to the geometry of the illuminated object [33]. Thus, in this study, we place importance on its asymmetrical shape and return power of the waveform to examine a new framework as a canopy height indicator without the customary use of the Gaussian modeling to fit multiple peaks or normalization procedures [34] or computing the radius of gyration [35]. First, using the raw waveform signal, vital information from the variety of complex waveform shapes in LiDAR can be captured without the various types of parameters that must be approximated using the current methods. Second, without any assumption that the ground return is symmetric (such as a Gaussian), we try to eliminate the potential source of tree height error. Our new method agrees with Hopkinson & Chasmer [36] that described the importance of the geometry and radiometry of detected objects.

This paper aims to extend the LiDAR analysis through characterization of the waveform based on its shape and return peak locations. We relate our index to the movements of the waveform shapes and key profile landmarks. We illustrate the feasibility of our approach as a canopy height indicator by means of synthetic and real LiDAR datasets. More importantly, we use extant field measurements from a field campaign to verify the performance of the new framework.

## **2. Materials and Method**

### *2.1. Waveform LiDAR and Field Datasets*

In the summer of 2003, Laser Vegetation Imaging Sensor (LVIS) data was acquired over an intensively studied forest near Howland, Maine. The dataset, which is available for download online through NASA LVIS website (<http://lvis.gsfc.nasa.gov>), displays varying LiDAR waveform shapes that are evident when the pulse return power is plotted as a function of temporal or range bins. LVIS has a scan angle of about 12 degrees, nominally with 25 m wide footprints, and could cover 2 km

swaths of surface from an altitude of 10 km. Further information regarding the instrument is available from Blaire et al. [37]. In addition, we also utilized the 2009 LVIS acquisition from Maine to confirm trends we observed in the 2003 data. The 2009 acquisition served as an excellent testbed to relate the LVIS results with the available data from the 2009 NASA DESDynI (Deformation, Ecosystem Structure and Dynamics of Ice) field campaign. The ground-based forest inventory measurements provided data needed to verify the efficacy and performance of the new shape-distance metric as a canopy-height indicator.

The field data from Howland, Maine (diameter at breast height, dbh) includes biomass from large stems ( $\text{dbh} \geq 10$  cm) and small stems ( $\text{dbh} < 10$  cm), mean heights from field plots, and stem density. Howland Research Forest was one of the experimental field stations for the 2009 forest biophysical measurements, as part of the North American Carbon Program (NACP). One of the goals of the field campaign was to compare the forest biophysical attributes with LVIS. Biomass was calculated using allometric equations. The biophysical measurement results and the biomass estimates were made available online and free for download via data center ORNL DAAC, in NASA's Earth Observing System Data and Information System (EOSDIS). We refer readers to Cook et al. [38] for the details of the field campaign (e.g., plot locations, descriptions of the dataset, and the link to download the data).

There were two stumbling blocks when we related the LVIS dataset and the field data. First, the two datasets did not exactly overlap geographically. Second, the location of the field sampling plots limited the number of points that can be compared to the LVIS data. To overcome these, we took the average of field data points surrounding an LVIS data point. Then we picked the number of observations we believe was satisfactory in providing evidence on the relationship of the MD against ground-based measurements.

The synthetic LiDAR dataset was generated through simulated landscapes at 25m footprint [39] built up from a 30m x 30m forest stand and grown using spatially explicit forest gap succession model, ZELIG [40]. For more information and details on how the synthetic data was generated, we refer you to Sun and Ranson (2000) [39]. The simulated dataset (over 500 years) frequently exhibits three distinct local peaks: a first canopy peak, a second canopy peak, and a ground peak. More on the segregation of waveforms based on peaks is discussed in a specific section below.

Following Lefsky et al. [41] and Harding & Carabajal [42], we derived the canopy height (reference height) from the waveform dataset as the difference between the power of the first increase of return (above the mean noise level) and the center of the last pulse (designated as the ground peak). Additionally, we calculated vegetation height quartiles [11] or height percentiles [43]—RH100, RH75, RH50, and RH25—representing the heights corresponding to 100%, 75%, 50% and 25% aboveground level energy return, respectively. The location of the ground peak is imperative in calculating quartiles, thus, it is crucial to detect the peak effectively. If the ground is found incorrectly, the RH metrics will have corresponding errors [24]. We then used the reference canopy height to compare against the values of MD and the RH. To guarantee robust results in our

comparisons for this study, we located the ground peak through labor-intensive manual checking of the waveform samples.

## 2.2. Moment Distance Framework

The Moment Distance is a new analytical framework that uses a computationally simple metric to capture the shape of the curve. First used to analyze the shape of the reflectance curve [44,45], this is the first time the framework is applied to waveform LiDAR. The approach takes advantage of the many returns of the waveform LiDAR to monitor changes in shape and its asymmetry—exploiting the range from first detected signal to last detected signal above the noise threshold. The formulation of the concept revolves around using the raw waveform to retain richness of the data. That means the framework avoids Gaussian fittings in our goal to detect changes of the waveform (e.g. widening of peaks, existence of complex extremes) with the change of canopy parameters, such as canopy height. The process involves fixing two points as references and has two aspects: the set of equations that generate the MD metrics and the choice of positions within the waveform to highlight. Assume that the waveform is displayed in Cartesian coordinates with the abscissa displaying time lapse  $t$  and ordinate displaying backscattered power  $p$ . Let the subscript LP denote the *left pivot* or earlier temporal reference point and subscript RP denote the *right pivot* or later temporal reference point. Let  $t_{LP}$  and  $t_{RP}$  be the time value observed at the left and right pivots, respectively. The MD framework is described in the following set of equations:

$$MD_{LP} = \sum_{i=t_{LP}}^{t_{RP}} (p_i^2 + (i - t_{LP})^2)^{0.5} \quad (1)$$

$$MD_{RP} = \sum_{i=t_{RP}}^{t_{LP}} (p_i^2 + (t_{RP} - i)^2)^{0.5} \quad (2)$$

$$MDI = MD_{LP} - MD_{RP} \quad (3)$$

The moment distance from the left pivot ( $MD_{LP}$ ) is the sum of the hypotenuses constructed from the left pivot to the power at successively later times (index  $i$  from  $t_{LP}$  to  $t_{RP}$ ): one base of each triangle is difference from the left pivot ( $i - t_{LP}$ ) along the abscissa and the other base is simply the backscattered power at  $i$ . Similarly, the moment distance from the right pivot ( $MD_{RP}$ ) is the sum of the hypotenuses constructed from the right pivot to the power at successively earlier times (index  $i$  from  $t_{RP}$  to  $t_{LP}$ ): one base of each triangle is the difference from the right pivot ( $t_{RP} - i$ ) along the abscissa and the other base is simply the backscattered power at  $i$ . Salas and Henebry [45] provided illustrations focusing on the computations of MD (note that the reflectance is equivalent to backscattered power and wavelength is equivalent to the time bins).

The MD Index (MDI) is an unbounded metric. It increases or decreases as a nontrivial function of the number of wave counts considered and the shape of the waveform that spans those contiguous

wave counts. The number of bins is a function of the temporal resolution of the LiDAR (digitization rate) and the length of the waveform (*i.e.* full extent or subsets) being analyzed. Depending on digitization rate, the matrix resulting from the calculations of the MDs within a range of waveform could be a massive set of numbers. As the MDI is designed to exploit the many bins and the asymmetry of the waveform, the new metric may lose its capability to detect shape changes or movements of wave morphologies when used improperly. Being resolution-dependent, MDI may perform poorly and fail to define the waveform shape when there are only few points between pivots

Using the terminology of Lefsky et al. [46] known as waveform extent, we computed the MDI with (a) LP at the left foot of the first peak and RP at bin zero for the synthetic dataset and (b) LP at the left foot of the first peak and RP at the right foot of the ground peak for the LVIS dataset. Further, we computed the MDI from four other pivot pairings using the location of the ground peak as RP, and paired with four LPs from locations of RH25 ( $MDI_{RH25}$ ), RH50 ( $MDI_{RH50}$ ), RH75 ( $MDI_{RH75}$ ), and RH100 ( $MDI_{RH100}$ ). The goal for this task was to contrast the resulting statistical relationships obtained from RH *vs* reference canopy height; and MDI (with pivot ranges equal to RH) *vs* reference canopy height.

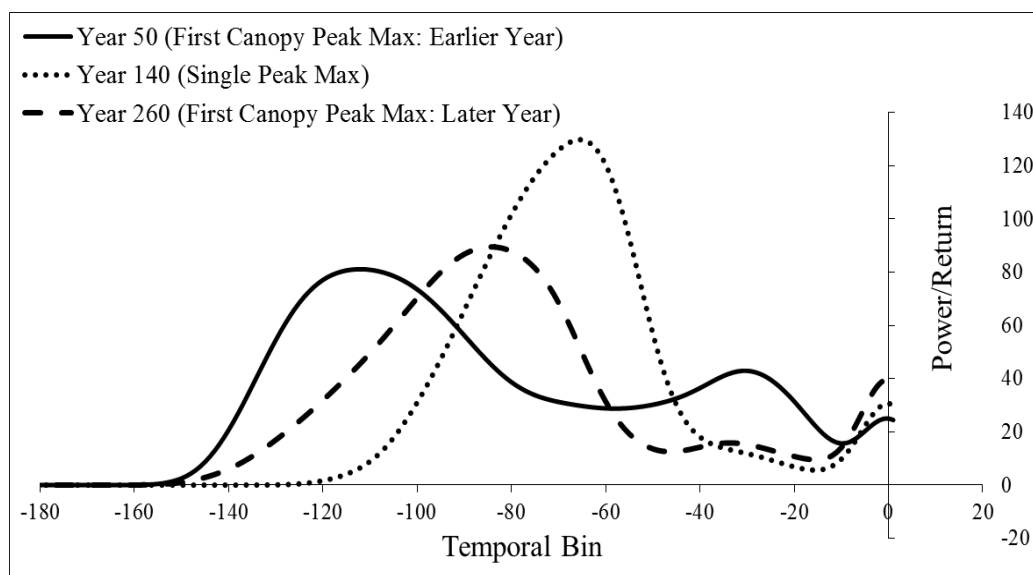
### 2.3. Segregation of Waveforms

For the synthetic LiDAR data, we identified four waveform shapes in terms of their temporal detection. For instance, the upper forest canopy is detected first by the sensor and we call it first peak, the midstory is the second peak, and the ground is the ground peak. The number of peaks detected or recorded depends on the characteristics of the forest. Here is how we classified the synthetic data: (1) *first canopy peak*, when the first canopy peak is maximum in a three-peak waveform (2) *second canopy peak*, when the second canopy peak is maximum in a three-peak waveform (3) *single peak*, when a single canopy peak is maximum in a two-peak waveform and (4) *ground peak*, when the ground/soil is maximum in a two-peak or three-peak waveform. A sample illustration of some of the behaviors of the curves is shown in Figure 1.

We analyzed the individual shapes in relation to MDI trends (*e.g.*, temporal MDI dynamics) and clustered them according to the generation of canopy stands. Every time a value of MDI is computed from a particular year, we plotted time *vs* waveform power to visually inspect the movement of the waveform peaks and how MDI is able to detect the movements (like shifting of positions of maximum returns at different time delays). In addition, we explored each generation stand through further differentiation of the peaks based on the waveform shape.

For the LVIS dataset we identified three shapes for the generally two-peak waveforms: (1) *early peak max*, when the canopy peak is maximum in return, (2) *equal (roughly) peak max*, when the canopy and the ground peaks are almost equal in return, and (3) *late peak max*, when the ground peak is maximum in return. The *early peak* is associated with a forest response, while the *late peak* is associated with a ground response. The variation of the backscattered power between the *early peak*

and the *late peak* depends on many things, canopy height is one. Segregation of waveforms was conducted, first to inspect the movement of the key profile landmarks (temporally as for the synthetic data), second to demonstrate how the new approach is able to detect the movements, such as shifting of positions of maximum peaks and bin locations, and third to explore the link between shape and the forest ground-based measurements.

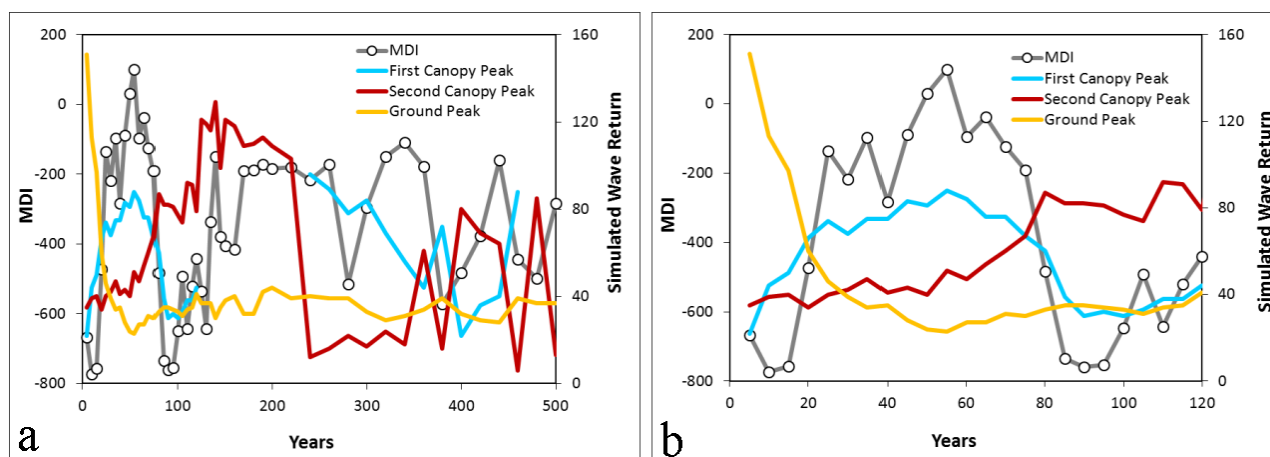


**Figure 1. Samples of synthetic LiDAR waveforms backscattered from a simulated forest stand and showing location of canopy and ground peaks.** Take note of the shifting of the maximum peak and the disappearance and reappearance of the minimum peak as the stand gets older.

### 3. Results

#### 3.1. Synthetic LiDAR Data

Figure 2 shows how the MDI and the magnitude of the respective peaks changed over the 500-year simulation. The *first canopy peak* disappeared for about a century starting around the 120th year. In Figure 2, three types of curves representing three canopy generations were observed. The first generation of stands covers the years 0 to around 120, the second is from years 120 to about 240, and the third generation covers the time range 240 to 500. Our results showed which *peak* has control over a certain generation of forest stands. The first generation was influenced most by the changes of the *first canopy peak* (from 5 to 120 years,  $r^2 = 0.81$ ) and *second canopy peak* (from 55 to 120 years,  $r^2 = 0.75$ ), after we put the magnitude of the peaks against the MDI.



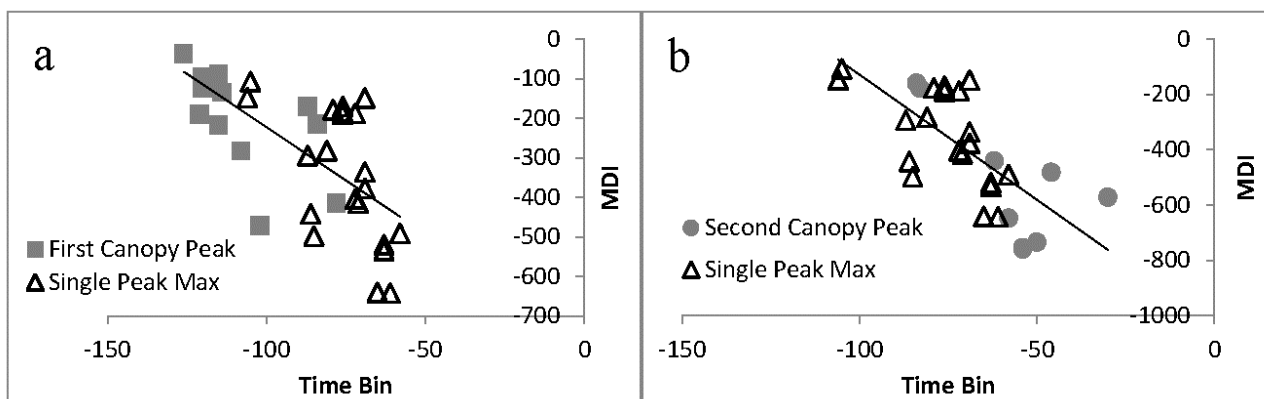
**Figure 2. Temporal dynamics of MDI tracking synthetic LiDAR waveforms from simulated forest stand: (a) with 500 years of simulations and (b) zooming into the first 120 years.** Notice in (2a) the disappearance of the *first canopy peak* around year 120 and its reappearance around year 240. Negative MDI signifies more MD accumulations on the RP.

Interestingly, the *ground peak* had the most influence on the second generation, in the absence of the *first canopy peak*. Ground returns gave an  $r^2 = 0.59$  when regressed against the MDI from second generation stands. The *ground peak* has the most influence on the later generation stands ( $r^2 = 0.47$ ). Neither the *first canopy peak* nor the *second canopy peak* was seen as responsible for the variations of the MDI in the later generation of stands.

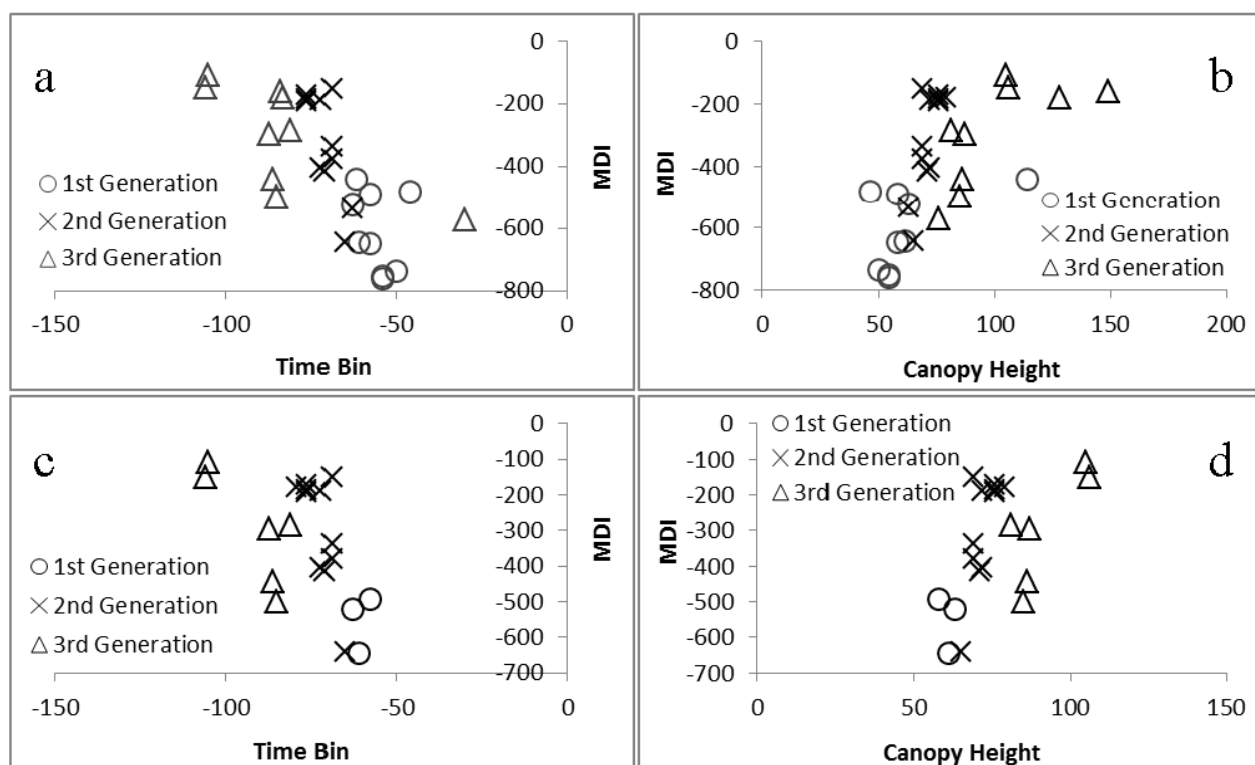
Looking closely into each peak of each time period by putting waveforms with maximum values at the *first canopy peak*, regardless of the generation, and grouping them with values of the *single peak*, we noticed the MDI clustering at low negative values when *first canopy peak* was at maximum, peaking at an earlier time. MDI showed to be much higher negatively when there is only a *single peak* on the waveform, peaking at much later time (Figure 3a). In addition, MDI of the *second canopy peak* showed to have higher negative values than the MDI of the *single peak* (Figure 3b). The *second canopy peak* occurred at an even much later time—later than *first canopy peak* and/or *single peak*.

Clustering of the MDI based on the generations was observed when we plotted all three-peak waveforms with maximum values found at the *second canopy peak* (Figure 4a). The same trends were seen when all waveforms showing *single peak* were combined and plotted (Figure 4c). Canopy height metrics decreased in increasing negative MDI. Large MDI values were linked to the first generation stands (Figures 4b and 4d).





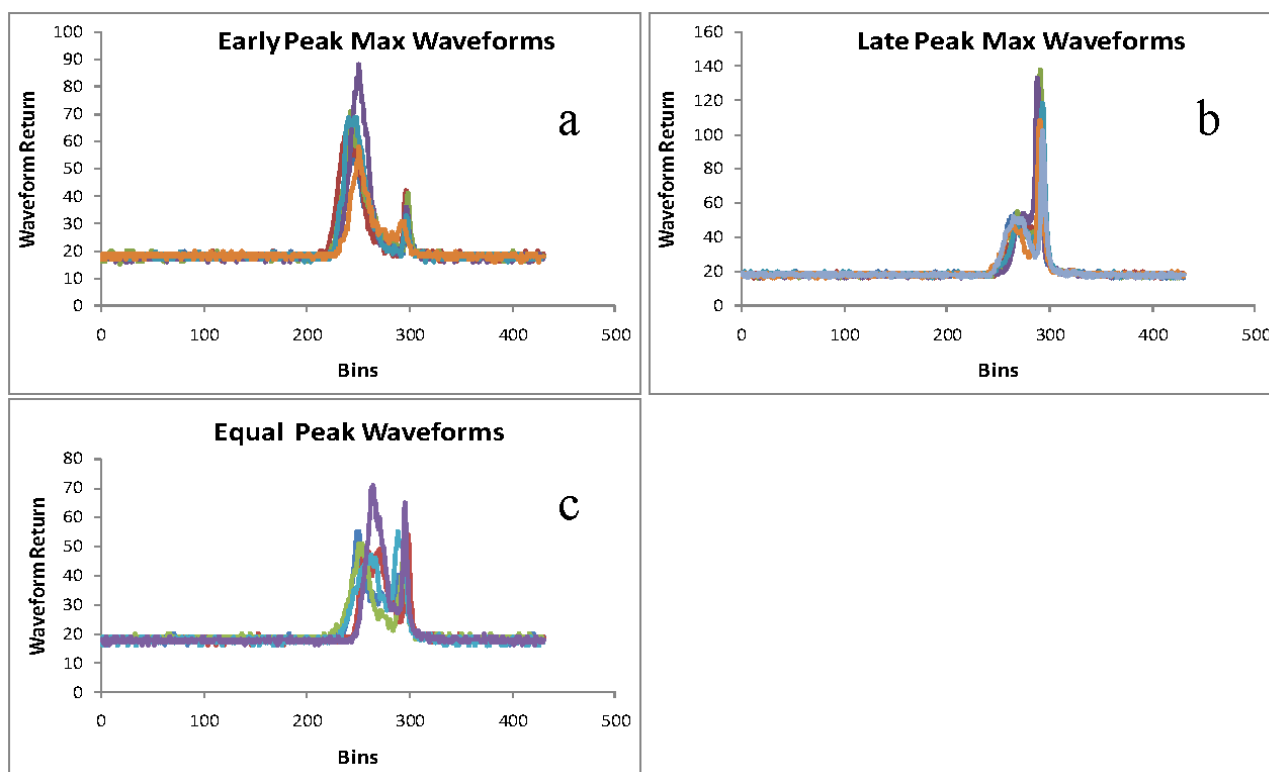
**Figure 3.** MDI against the time bin showing (a) *first canopy peak vs. single peak* and (b) *second canopy peak vs. single peak*. Groupings of the synthetic waveforms were based on the type of shape without consideration of the generation where the waveform belongs.



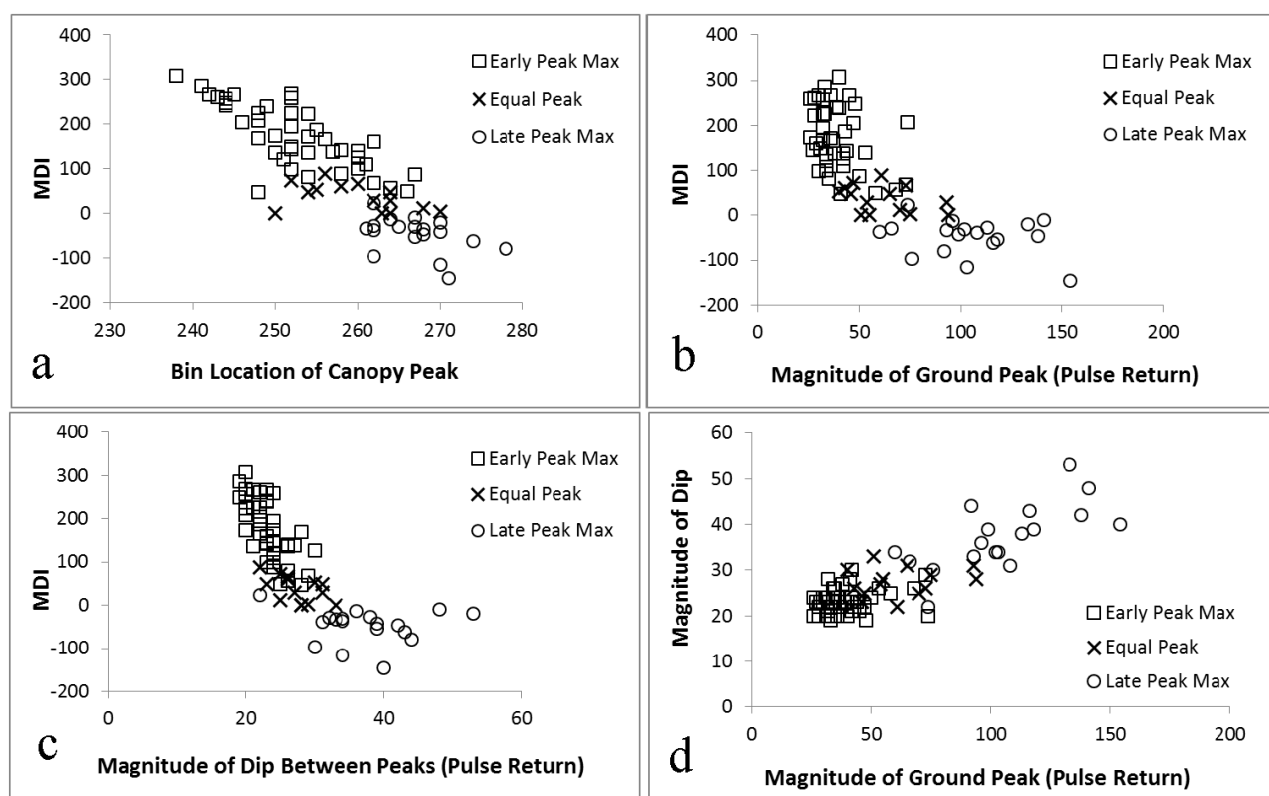
**Figure 4.** Looking at the tendency of MDI to cluster according to the generation period of stands for the synthetic waveforms. Plots (a) and (b) used the maximum values at the second canopy peak to plot MDI against time bins and reference canopy height, respectively. Plots (c) and (d) used the values at the single peak to plot MDI against time bins and reference canopy height, respectively.

### 3.2. LVIS LiDAR Dataset

Results of the waveform segregation are shown in Figure 5 with (a) *early peak max* (b) *late peak max* and (c) *equal (roughly) peak max* observed on the waveforms. For the description of each waveform shape, we refer the readers to section 2.3. In Figure 6, MDI showed evident trends against the waveform morphologies. Shifting of the maximum peaks and dip was detected by the shifting of the MDI. The major observation was on MDI being more sensitive to the location (time bin) changes of the canopy peak ( $r^2 = 0.73$ , Figure 6a) rather than the magnitude of the pulse return of the canopy peak ( $r^2 = 0.37$ ). In contrast, the magnitude of the pulse return of the ground peak showed higher relationship ( $r^2 = 0.65$ ) with MDI (Figure 6b) than the ground bin location ( $r^2 = 0.20$ ). Interestingly, the MDI detected the changes of the magnitude of the returns at the dip (in between the canopy peak and ground peak) ( $r^2 = 0.67$  in Figure 6c), but not the shifting of its bin ( $r^2 < 0.10$ ). Also, in Figure 6d, the magnitude of the dip and the magnitude of the ground peak were directly related ( $r^2 = 0.70$ ).



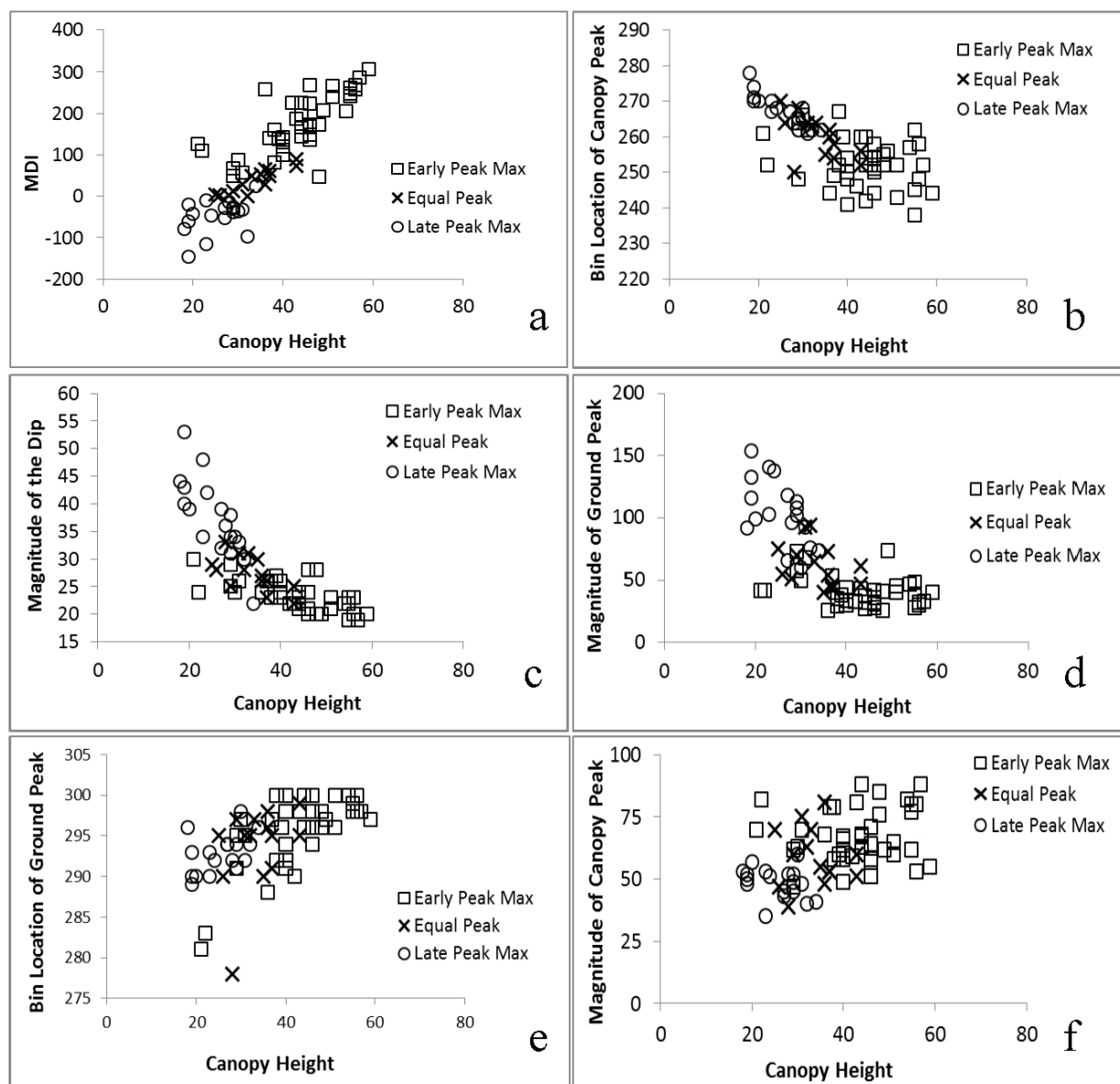
**Figure 5. Three waveform shapes of the LVIS dataset (a) early peak max observed (b) late peak max observed and (c) equal (roughly) peak max observed. Line colors represent different waveform samples.**



**Figure 6. Relationships between the MDI, the three waveform shapes, and the waveform landmarks (a) bin location of canopy peak, (b) and (d) magnitude of ground peak, (c) magnitude of dip between peaks, using the 2003 LVIS dataset. These relationships showed strong  $r^2$ : (a)  $r^2 = 0.73$ , (b)  $r^2 = 0.65$ , (c)  $r^2 = 0.67$ , (d)  $r^2 = 0.70$ .**

Results also demonstrated the relationship that is evident between the MDI and the reference canopy height ( $r^2 = 0.74$ , Figure 7a). The general trend confirmed previous results from the synthetic data associating MDI to the different types of the waveform shapes. In Figure 7b, as the location of the canopy peak shifts to a later time bin, the canopy height decreased.

The canopy height showed inverse proportionality with the three waveform properties: (1) the bin location of canopy peak in Figure 7b, (2) magnitude of the dip in Figure 7c, and (3) the magnitude of ground peak in Figure 7d. These were the same properties that illustrated strong relationships with MDI in Figure 6. Figure 7e and 7f illustrated the weak relationships of the canopy height against other profile landmarks such as the bin location of ground peak and the magnitude of the canopy peak, respectively.



**Figure 7. Relationships between the canopy height and the (a) MDI ( $r^2 = 0.74$ ), (b) bin location of canopy peak ( $r^2 = 0.85$ ), (c) magnitude of dip ( $r^2 = 0.67$ ), (d) magnitude of ground peak ( $r^2 = 0.55$ ), (e) bin location of ground peak ( $r^2 = 0.38$ ), and (f) magnitude of canopy peak ( $r^2 = 0.10$ ) using the Howland 2003 LVIS dataset. The bin location of the canopy peak is highly associated with the canopy height.**

Similarly, we analyzed the 2009 LVIS dataset and tabulated the  $r^2$  results in Table 1. The results showed the same linear relationships found using the 2003 LVIS dataset when we regressed MDI against the canopy height and the waveform properties.

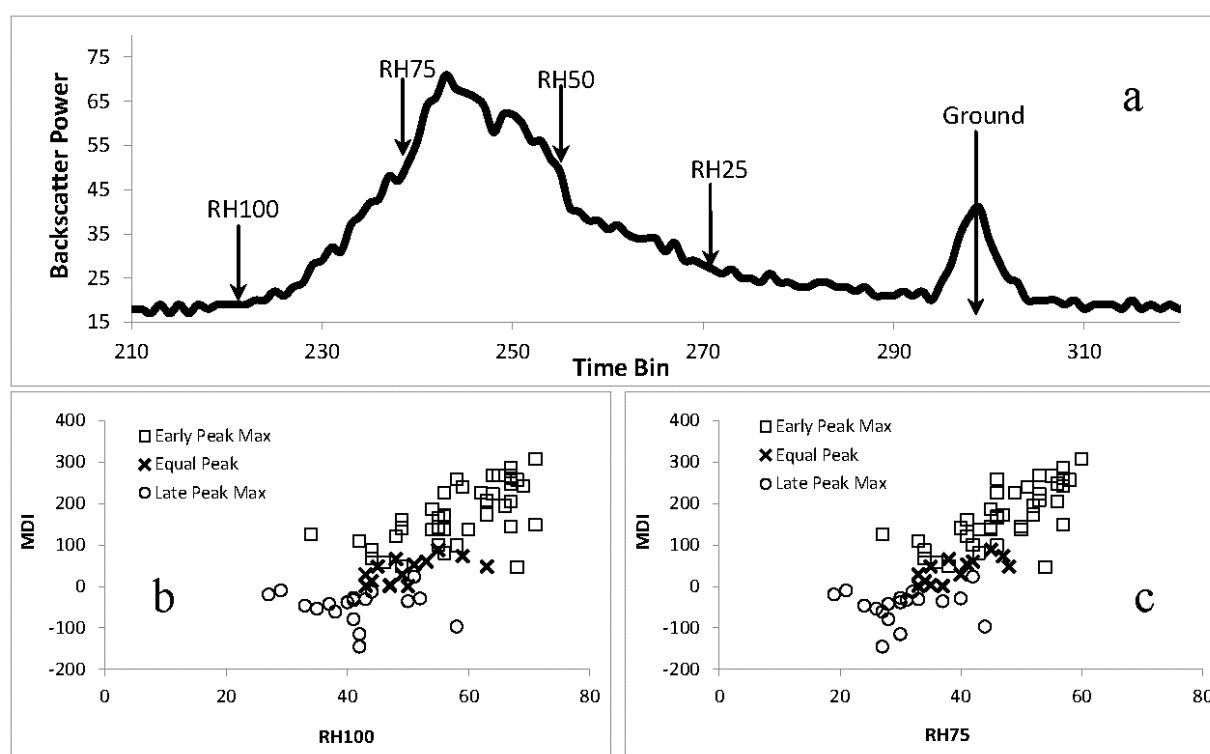
Further, we correlated the MDI against the computed height quartiles (Figure 8a). We found that the MDI is associated with all quartiles—RH100 (Figure 8b) and RH75 (Figure 8c) are shown. Regression analysis using these parameters produced average to high coefficients of determination:

$r^2 = 0.60$  and  $r^2 = 0.68$ , respectively for RH100 and RH75;  $r^2 = 0.79$  and  $r^2 = 0.86$ , respectively for RH50 and RH25. High degrees of correlations were also observed between RH pairings with correlation coefficients all above  $r^2 = 0.90$ . These high correlations led to all RH metrics to be strongly correlated with the canopy height, regardless of the location of the RH.

**Table 1. Coefficient of determination ( $r^2$ ) of MDI against the canopy height and the waveform properties for the 2009 and 2003 Howland LVIS datasets.**

Waveform Property	MDI	
	$r^2$ (2009)	$r^2$ (2003)
Canopy height	0.70	0.74
Bin location of canopy peak	0.67	0.73
Magnitude of ground peak	0.55	0.65
Magnitude of dip	0.56	0.67

The bin location of canopy peak, magnitude of the dip, and the magnitude of ground peak all show a strong relationship with the MDI.



**Figure 8. (a) Sample waveform with locations of relative height (RH) percentiles. RH represents the height (relative to ground peak) at which a certain percentage of the waveform energy occurs. The two bottom figures depict the relationships of (b) RH100 and (c) RH75 with MDI, showing the same trend as in Figure 7a. Note that RH50 and RH25 also had high degrees of correlations against the MDI (not shown) caused by strong intercorrelation among RH.**

The MDI computed with RH locations serving as pivots, in particular  $MDI_{RH25}$  and  $MDI_{RH50}$ , had much less correlation with the canopy height (Table 2). Unlike the RH, high correlations were only observed among high MDI height metrics (e.g.  $MDI_{RH75}$ ,  $r^2 = 0.83$ ; and  $MDI_{RH100}$ ,  $r^2 = 0.78$ ). In all waveform shapes, the  $MDI_{RH25}$  and  $MDI_{RH50}$  all came out less sensitive, if not insensitive, with the canopy quasi-height.

Note that among the three waveform shapes, the second peak max turned out to have the weakest relationship for MDI vs RH. Isolating each waveform type, the first peak max and the equal peak each returned an average to strong  $r^2$  (0.60 and 0.80, respectively). The same observations are evident in Table 2 when MDI is more sensitive to the canopy height in waveform shapes where the canopy peak stands out.

**Table 2. Coefficient of determination ( $r^2$ ) of MDI computed at locations of height quartiles and its relationship against the canopy height.**

	First Peak Max	Second Peak Max	Equal Peak
MDI ~ h	0.64	0.20	0.81
MDI100 ~ h	0.78	0.40	0.83
MDI75 ~ h	0.83	0.53	0.72
MDI50 ~ h	0.50	0.08	0.37
MDI25 ~ h	0.08	0.03	0.11

The first three rows, MDI calculated using a longer range from the ground peak, may be used to estimate the canopy height, with high  $r^2$  observed.

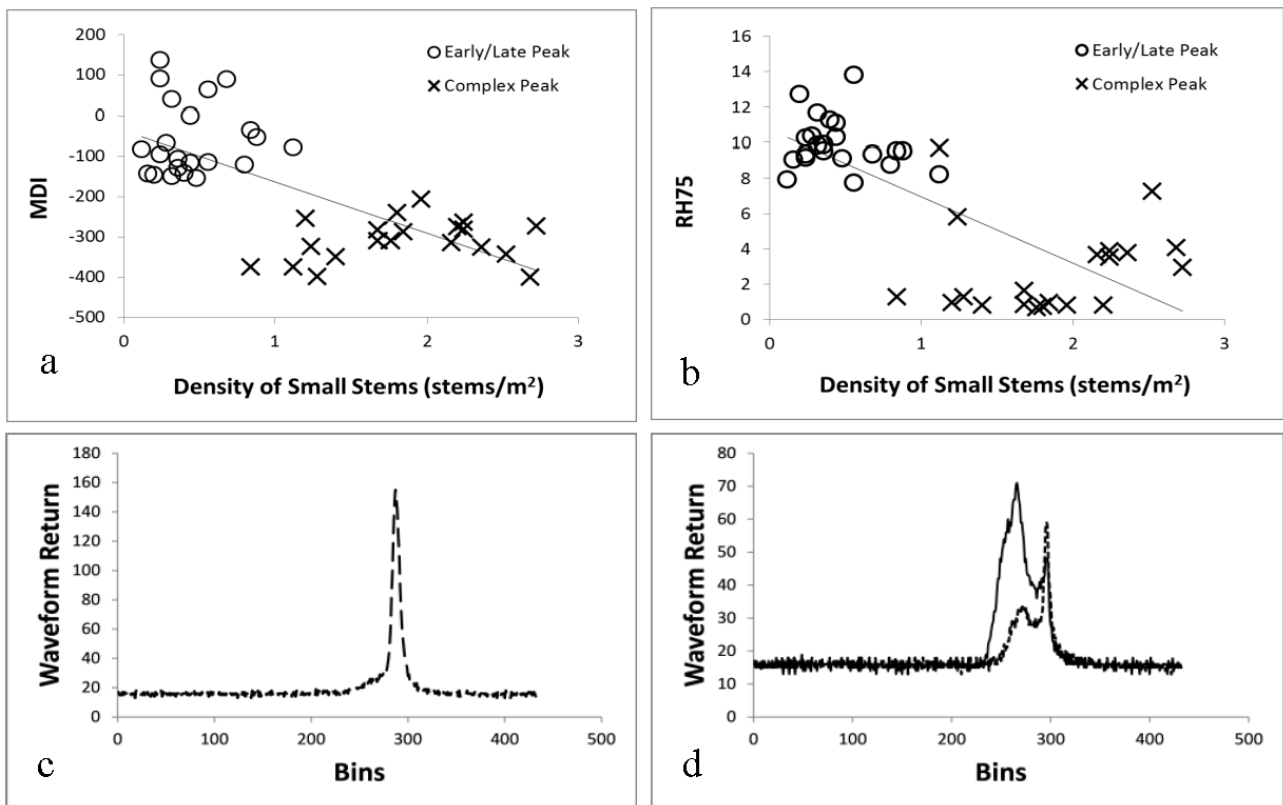
### 3.3. MDI vs Field Measurements

Table 3 summarized the ground-based forest inventory measurements from the 2009 DESDynI field campaign based on the waveform shape grouping provided in Figures 9c and 9d. The samples belonging to the *complex peak* group have low large-stem biomass and high small-stem biomass and density compared to those from the *early/late peak* group. Canopies in the early/late peak group have taller heights than those in the *complex peak*.

MDI showed a crisp clustering of the waveform shapes against the density of small stems than the relative height at 75% (RH75). The upper group, with mostly positive MDIs, was from the type of waveform shape having early and late peak (9c). The lower group, having mostly negative MDIs, was from the type of waveform shape having a complex peak, with existence of other peaks either from one side or both (9d). Figure 9 suggested that the type of waveform in Figure 9d may have been caused by a high density of small stems and less biomass from the large stems (clustering is also illustrated in Figure 10).

**Table 3. Statistical values of the field measurements grouped according to the LVIS waveform shapes.**

Biometry	With Early/Late Peak			With Complex Peak		
	Min	max	mean	min	max	mean
Large-stem biomass (Mg/subplot) (Jerkins et al. 2004 [47])	1.84	9.28	4.79	0.57	7.69	2.35
Large-stem biomass (Mg/subplot) (Young et al. 1980 [48])	1.63	8.88	4.21	0.58	9.37	2.16
Small-stem biomass (Mg/m <sup>2</sup> )	0.00001	0.0032	0.00097	0.0006	0.0088	0.00425
Small-stem density (stems/m <sup>2</sup> )	0.12	1.12	0.46	0.84	2.72	1.84
Mean of Maximum Height (m)	9.17	18.45	14.04	7.83	16.08	11.09



**Figure 9. Density of small stems (dbh < 10 cm) versus the (a) MDI and the (b) relative height at 75% (RH75). Observe how the MDI clustered the waveforms based on the shape type having (c) *early/late peak* and (d) *complex peak* (appearance of other peaks in either from one side or both).**

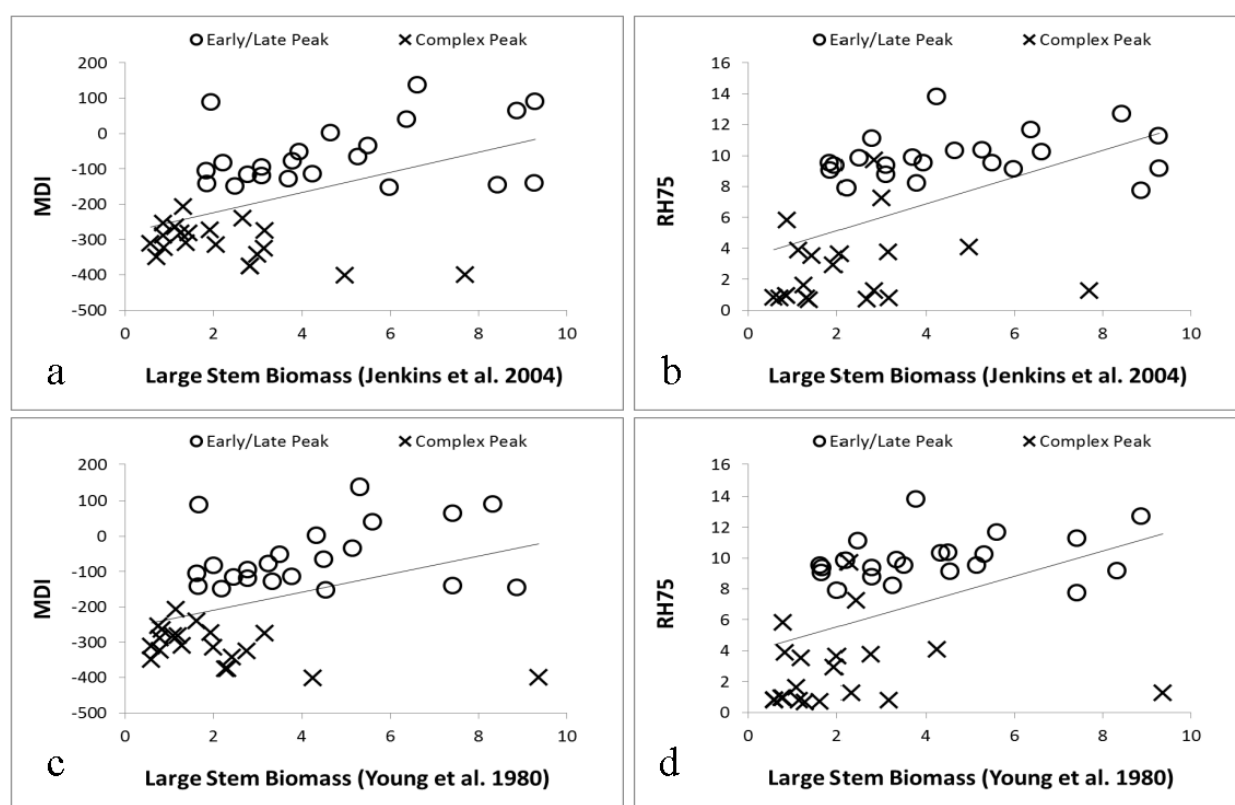
In Figure 10, the biomass from the large stems computed from Jerkins et al. (2004) [47] and Young et al. (1980) [48] both came out with comparable, however weak  $r^2$ , when regressed against the MDI ( $r^2 = 0.23$  and  $r^2 = 0.17$ , respectively) and the RH75 ( $r^2 = 0.27$  and  $r^2 = 0.21$ , respectively). Low negative MDIs were clustered at low large-stem biomass. Between MDI and RH, the MDI was

able to cluster better the two types of shapes.

Between MDI and RH, the former resulted to an  $r^2 = 0.4$  compared to the latter with  $r^2 = 0.3$  when regressed against the small-stem biomass. The trends observed in Figures 11a and 11b can be associated with Figures 9a and 9b, in terms of the clustering of the waveform shapes. The density of small stems is directly related to biomass of small stems ( $r^2 = 0.53$ ).

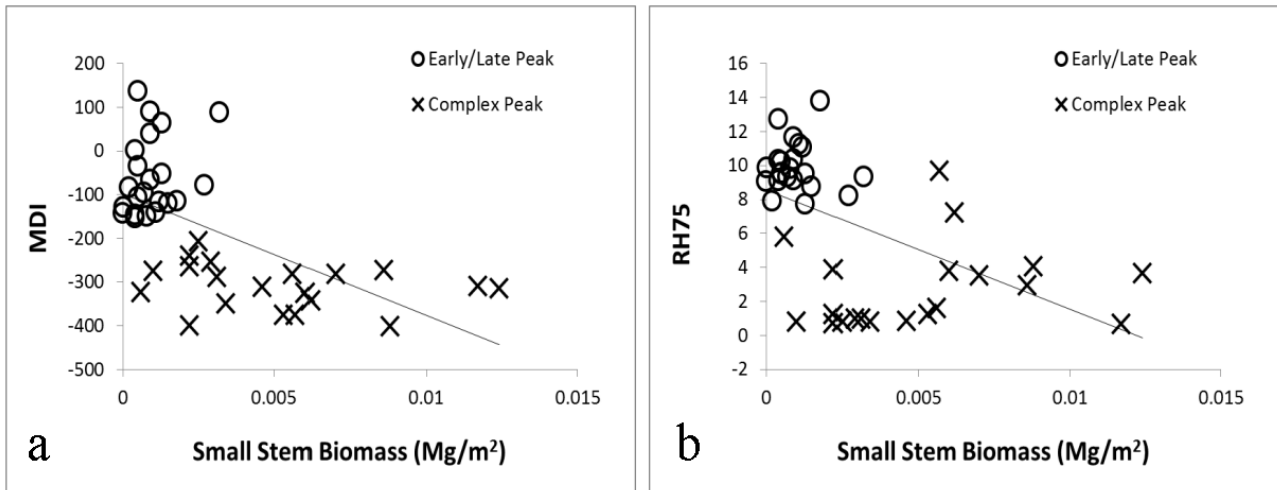
In Figure 12, the mean maximum height of the canopy has a higher correlation with the MDI ( $r = 0.43$ ), than with RH75 ( $r = 0.35$ ). The same linear trends were observed when using the synthetic data (Figure 4). The first generation stands in the synthetic data showed the same behavior (low negative MDI) as the group complex peak in the field data.

Isolating the group complex peak showed that MDI can still detect a trend of the low values of the large-stem biomass data (Figure 13a,  $r^2 = 0.40$  and Figure 13c,  $r^2 = 0.30$ ). RH75 failed to detect variations of the large-stem biomass.

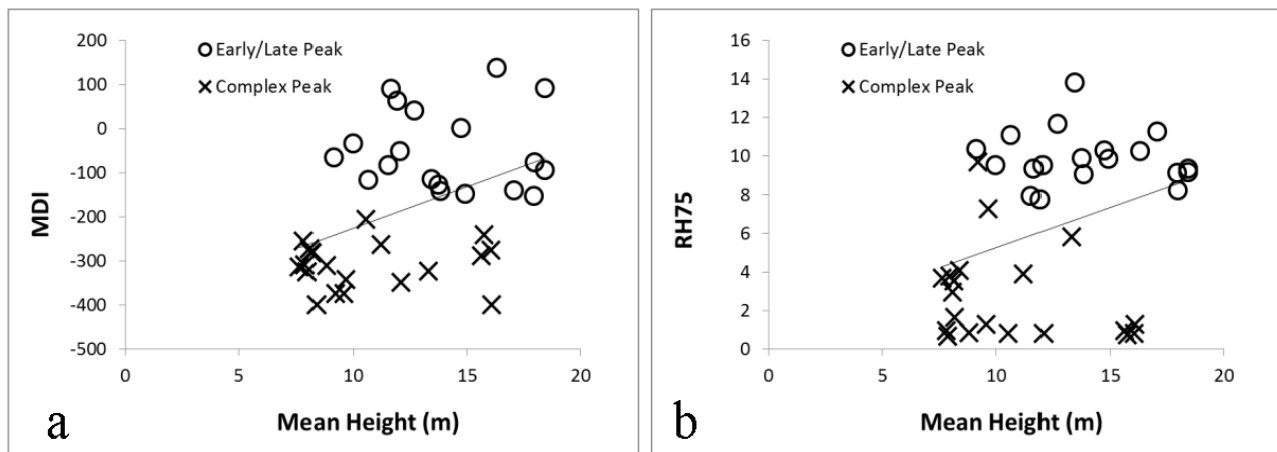


**Figure 10.** Biomass from large stems ( $\text{dbh} \geq 10 \text{ cm}$ ), calculated from Jenkins et al. (2004) and Young et al. (1980), versus the MDI (a & c, with  $r^2 = 0.23$  and  $r^2 = 0.17$ , respectively) and RH75 (b & d, with  $r^2 = 0.27$  and  $r^2 = 0.21$ , respectively). The low value MDIs (mostly negative) was from the type of waveform shape having complex peak as shown in Figure 9d.

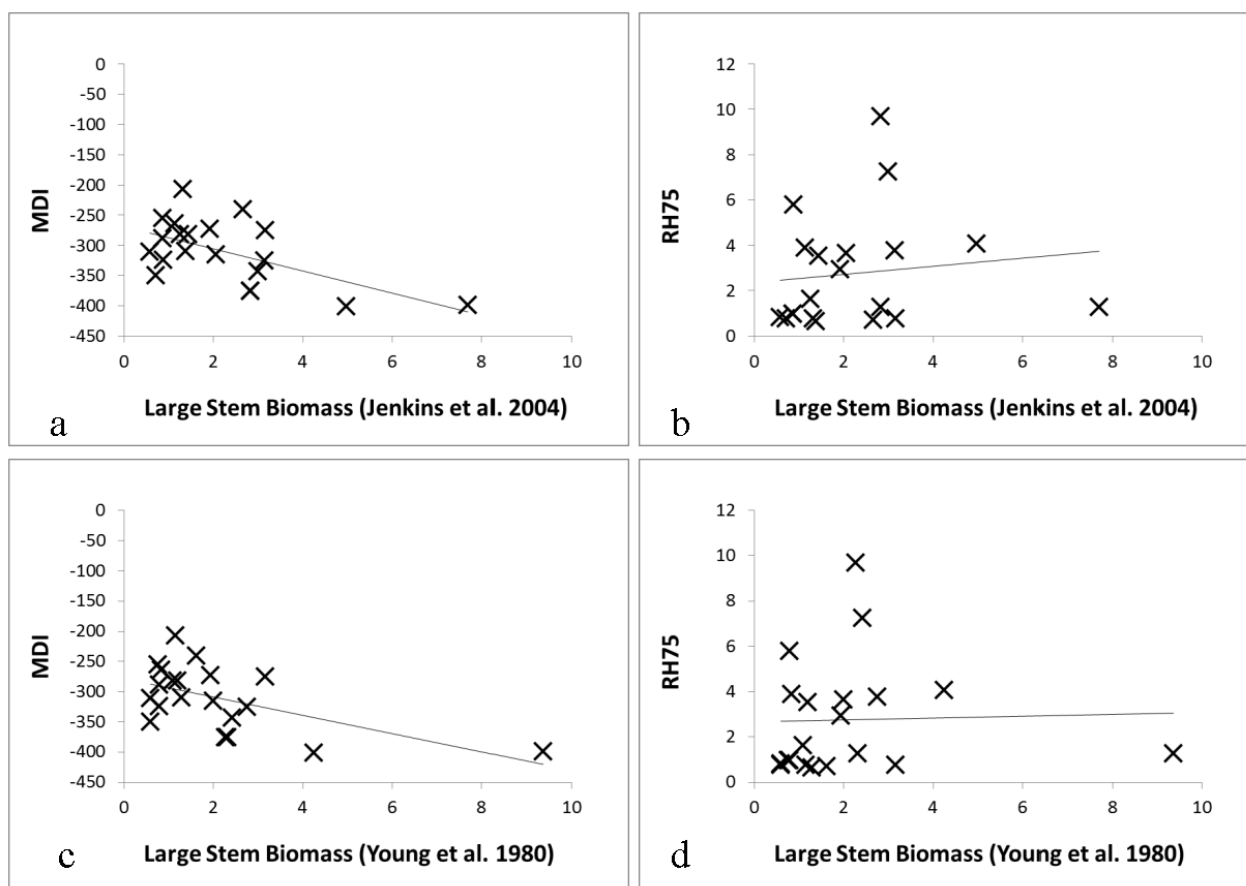




**Figure 11.** Biomass from small stems ( $\text{dbh} < 10 \text{ cm}$ ), calculated using the mixed hardwoods equation from Jenkins et al. 2004, was plotted against the (a) MDI, with  $r^2 = 0.40$  and (b) RH75, with  $r^2 = 0.30$ . Evident in the figure is the clustering of the waveform shapes with early/late peak showing the characteristics of small stem biomass.



**Figure 12.** Mean Height from field data against (a) MDI and (b) RH75. The mean height was computed from the three tallest trees in a subplot. The MDI had a higher linear relationship with the field data ( $r^2 = 0.20$ ) (albeit weak) compared to RH75 ( $r^2 = 0.10$ ). Note that for some samples, the height data were not available online.



**Figure 13.** Although MDI and RH75 showed direct relationship with large-stem biomass (Figure 10) in all waveform shapes, isolating the group with high density of small stems and less biomass from the large stems (group with complex peak) showed a different story. Only the MDI exhibited a trend, albeit weak ( $r^2 = 0.40$  for 13a, and  $r^2 = 0.30$  for 13c).

#### 4. Discussion

The Moment Distance Index computed from the LiDAR waveforms were found to provide satisfactory estimates of canopy height based on results we obtained using the synthetic data ( $r^2 = 0.40$ ) and the LVIS data ( $r^2 = 0.74$ ). The synthetic data may have shown a lower correlation for MDI against the canopy height, nevertheless, results exhibited that the MDI enabled the clustering of waveform samples according to the generation of stands. MDI values of the first and second generation of stands (year 0 to ~ year 120) highly correlated to the canopy peaks, represented as *first canopy peak* and *second canopy peak* in Figure 2. The relationships,  $r^2 = 0.81$  and  $r^2 = 0.75$ , suggested that changes of the canopy peaks, whether its magnitude of the return or the bin location, can be observed by the behavior of the MDI.

The simulated dataset presented a good way to look at the temporal dynamics of the MDI. The vanishing (~ year 120) and reappearing (~ year 240) of peaks at different time periods showed a

phenomenon where the younger canopy is “catching up” to the older canopy that caused a single canopy peak existing on the waveform. Also, the ground peak did not have any control on the first generation of stands, instead, the ground peak returns were observed to have effect on the MDI in much later years, ~ year 240 to ~ year 500.

The LVIS dataset provided a way for us to confirm the groupings of LiDAR waveform shapes found using the synthetic data. A pattern observed was the relationship between the MDI and the peak positions: stronger earlier peaks yield a positive MDI and stronger later peaks yield a negative MDI. However, in complex waveform shapes, MDI demonstrated that it could still separate the waveforms from the other non-complex ones (*e.g.* equal peak group). In the absence of peaks or in circumstances when peaks hardly exist, a big challenge is presented for methods employing the extraction of waveform peaks. The MDI, nevertheless, is not faced with the same difficulty since MDI is calculated based on the number of temporal bins included in the length of waveform being analyzed, regardless of the shape of the waveform or the number of peaks involved. This proved to work for height estimations minus the hassle of modeling Gaussian peaks.

The new approach performed well in our investigations against the canopy height. In fact, the results compared favorably with those of Cook et al. [49], in which the relative height metrics decreased in response to a more reflective ground surface. The greater energy of the ground peak reflected in our *late peak max* showed decreased canopy heights, *vis-à-vis*, the much higher values in the *early peak max* (Figure 7a). MDI was able to group the heights in terms of the three curve shapes. What were responsible for the height decreases were the movements of the key profile landmarks of the waveform, for instance, the bin locations of the *late peak max*.

The MDI is comparable with the relative height metrics. One reason for this is that the bins of one or two height quartiles (*e.g.* RH100, RH75) are close to the reference canopy heights, which is the bin difference of the canopy and the ground peak modes. The advantage of the MDI over the height metrics RH100, RH75, RH50, and RH25 is twofold. First, unlike in calculating height quartiles, there is no requirement in the MD approach to find the location of the ground peak. According to Rosette et al. [12], not detecting efficiently the ground signal can result into underestimation of vegetation height. In the case of the MDI, it can be calculated using any length of the waveform (*i.e.* full waveform, start of detected to end of detected signals, or any other useful subset of bins) and in any type of shapes. Second, even when ground peaks may be inexistent or waveforms having complex extremities, MDI can still become a valuable metric to estimate the canopy height. Using the start of detected to end of detected signals as waveform pivot range, MDI can be computed with no worry of detecting peak locations.

As explained in the previous paragraph, the value of the canopy height we used in the analysis is located close to RH75, hence the high correlation observed against the MDI. The high correlations of the other RH metrics were quite surprising, especially the RH25 versus the MDI. Seeing the RH25 located too far from the canopy peak and close enough from the ground peak, yet its correlation to canopy height is the highest among the RH metrics, makes the RH approach problematic. We see this

as a drawback of the RH method as a height estimator, as also shown in Andersen et al. [50] and Muss et al. [51].

The computed MDI with RH locations as pivots serves as a better alternative of the customary RH metric. For one, only the  $MDI_{RH100}$  and  $MDI_{RH75}$  are highly sensitive to the changes of the canopy height, while  $MDI_{RH50}$  and  $MDI_{RH25}$  are less sensitive to changes. Moment distance relative height metrics ( $MDI_{RH}$ ) indicated that they can be independent of one another, denoting that those quartiles (or any other relative heights) close to the ground peak (having shorter ranges) could isolate themselves from those close to the canopy peak (having longer ranges). Further, quartiles (or any other relative heights) close to the canopy peak tend to correlate with each other and, and eventually, with the canopy height.

We assert that while  $MDI_{RH100}$  and  $MDI_{RH75}$  may be good choices for canopy height estimations, they both still employ a process of locating the ground peak for appropriate RP and LP pivot assignment. To refrain from going through the process, especially when peaks are hardly identifiable, we recommend utilizing the MDI with RP at the first detected signal and LP at the last detected signal.

The field measurements provided an opportunity for us to test the behavior of the new framework. Examination of Figure 13 showed that MDI may work for any type of waveform, especially for a single-peak type where canopy height (canopy peak location minus ground peak location) cannot be estimated due to the absence of one peak. There are still questions to the method as to how Gaussian modeling is able to resolve waveforms with single peak or a peak with complex spread of extremes and how heights are computed in the absence of one of the important peaks in the waveform. The RH, which comes along any LVIS data available online, can be a good indicator of canopy height as well. But then again, in a shape dominated by high density of small stems and less standing biomass from the large stems, the relative height metric may not work. MDI can put an end to the doubt by not relying on peaks to compute heights, but rather on the shape of the curve.

We would have wanted to see a relationship of MDI vs height for small stems; however, the data only provided mean heights of the three tallest trees in a subplot. We believe that if more trees would have been used in the averaging of the canopy height, a much higher correlation and clearer relationship trend may have been obtained.

## 5. Conclusion

Without going through the process of curve modeling, which is sometimes problematic and insufficient, this study presented the potential of the moment distance index to estimate the canopy height using LiDAR waveform as tested on both synthetic and real datasets and verified by the field data measurements. This novel approach offers a way to analyze the different complex waveform shapes, which may be essential for studying canopy understory. MDI helps illustrate how the key profile landmarks of the waveform may change over time. Further, MDI aids future research in

explaining LiDAR waveforms more precisely based on the types of shapes.

The method of considering all the time bins within the length of the waveform being analyzed may be seen as an impediment to a swifter analysis. Nonetheless, this should not be considered a limitation, but rather an essential feature of the approach that could provide more information about the formation of the peaks and dips of the wave. Although parameterization is possible and efficient as shown in previous results, omitting too many bins may defeat the main goal of the approach. Additionally, the use of all time bins actually saves computational time as there is no need to find prominent peaks to compute heights. MDI only needs the beginning and ending temporal bins to analyze the LiDAR waveform and estimate the canopy height, regardless of the shape of the waveform or the number of peaks involved.

A future research direction is to study the spatial distribution of the MDI across a forested area, for instance. It will be interesting to know how the new framework relates to change across space such as spatial heterogeneity and spatial trends. Further, the potential of the MD to help improve characterization of forests by resolving both community composition and age structure can also be explored. Lastly, we hope to apply the new analytical framework to discrete return systems. Muss et al. [51] has shown that discrete return LiDAR dataset can be transformed into a pseudo-waveform that has many of the same characteristics of traditional waveform LiDAR data. The MDI can become an effective canopy height indicator for various types of canopy structures using pseudo-waveforms.

## Acknowledgements

The authors would like to thank Dr. Guoqing Sun for sharing with us the synthetic LiDAR data. The authors would also like to thank the anonymous reviewers for helping to improve clarifying of the presentation. This research was supported in part by NASA grant NNX09AL23G: Synergistic analyses of data from active and passive sensors to assess relationships between spatial heterogeneity of tropical forest structure and biodiversity dynamics.

## Conflict of Interest

The authors declare they have no conflicts of interest in this article.

## References

1. Lefsky MA, Harding D, Parker G, et al. (1999) Lidar remote sensing of forest canopy and stand attributes. *Remote Sens Environ* 67: 83-98.
2. Means JE, Acker SA, Harding DJ, et al. (1999) Use of large-footprint scanning airborne LiDAR to estimate forest stand characteristics in the western cascades of Oregon. *Remote Sens Environ* 67: 298-308.

3. Lefsky MA, Cohen WB, Harding DJ, et al. (2002) Lidar remote sensing of above-ground biomass in three biomes. *Glob Ecol Biogeogr* 11: 393-399.
4. Banskota A, Wynne RH, Johnson P, et al. (2011) Synergistic use of very high-frequency radar and discrete-return lidar for estimating biomass in temperate hardwood and mixed forests. *Ann For Sci* 68: 347-356.
5. Yao W, Krzystek P, Heurich M (2012) Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sens Environ* 123: 368-380.
6. McGlinchy J, Aardt JAN van, Erasmus B, et al. (2014) Extracting structural vegetation components from small-footprint waveform Lidar for biomass estimation in Savanna ecosystems. *IEEE J Sel Top Appl Earth Obs Remote Sens* 7: 480-490.
7. Dubayah R, Blair JB, Clark DB, et al. (1997) The vegetation canopy Lidar mission. In: Proc Conf Land Satellite Information in the Next Decade II. p. 100-112.
8. Hurtt GC, Dubayah R, Drake J, et al. (2004) Beyond potential vegetation: Combining Lidar data and a height-structured model for carbon studies. *Ecol Appl* 14: 873-883.
9. Fieber KD, Davenport IJ, Tanase MA, et al. (2015) Validation of canopy height profile methodology for small-footprint full-waveform airborne LiDAR data in a discontinuous canopy environment. *ISPRS J Photogramm Remote Sens* 104: 144-157.
10. Lim K, Treitz P, Wulder M, et al. (2003) LiDAR remote sensing of forest structure. *Prog Phys Geogr* 27: 88-106.
11. Sun G, Ranson KJ, Kimes DS, et al. (2008) Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sens Environ* 112: 107-117.
12. Rosette JAB, North PRJ, Suárez JC, et al. (2010) Uncertainty within satellite LiDAR estimations of vegetation and topography. *Int J Remote Sens* 31: 1325-1342.
13. Wagner W, Ullrich A, Ducic V, et al. (2006) Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS J Photogramm Remote Sens* 60: 100-112.
14. Mallet C, Bretar F (2009) Full-waveform topographic lidar: State-of-the-art. *ISPRS J Photogramm Remote Sens* 64: 1-16.
15. El-Baz A, Gimel'farb G (2007) EM-based approximation of empirical distributions with linear combinations of discrete Gaussians. In: 2007 IEEE International Conference on Image Processing. p. IV 373-376.
16. Zwally HJ, Schutz B, Abdalati W, et al. (2002) ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *J Geodyn* 34: 405-445.
17. Chauve A, Vega C, Durrieu S, et al. (2009) Processing full-waveform lidar data in an alpine coniferous forest: Assessing terrain and tree height quality. *Int J Remote Sens* 30: 27.

18. Hofton MA, Minster JB, Blair JB (2000) Decomposition of laser altimeter waveforms. *IEEE Trans Geosci Remote Sens* 38: 1989-1996.
19. Jutzi B, Stilla U (2006) Range determination with waveform recording laser systems using a Wiener Filter. *ISPRS J Photogramm Remote Sens* 61: 95-107.
20. Chauve A, Mallet C, Bretar F, et al. (2007) Processing full-waveform lidar data: modelling raw signals. In: International archives of photogrammetry, remote sensing and spatial information sciences [cited 2016 Jul 5]. p. 102-107. Available from: <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00293129/>.
21. Mallet C, Lafarge F, Bretar F, et al. (2009) A stochastic approach for modelling airborne lidar waveforms. Object Extr 3D City Models Road Databases Traffic Monit-Concepts Algorithms Eval CMRT 2009, 38(Part 3): W8.
22. Hancock S, Lewis P, Foster M, et al. (2012) Measuring forests with dual wavelength lidar: A simulation study over topography. *Agric For Meteorol* 161: 123-133.
23. Wallace A, Nichol C, Woodhouse I (2012) Recovery of forest canopy parameters by inversion of multispectral LiDAR data. *Remote Sens* 4: 509-531.
24. Dubayah RO, Sheldon SL, Clark DB, et al. (2010) Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *J Geophys Res Biogeosciences* 115(G2): G00E09.
25. Levenberg KA (1944) Method for the solution of certain problems in least squares. *Q Appl Math* 2: 164-168.
26. Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 11: 431-441.
27. Brenner AC, Zwally HJ, Bentley CR, et al. (2012) The algorithm theoretical basis document for the derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights [cited 2016 Jul 5]; Available from: <http://ntrs.nasa.gov/search.jsp?R=20120016646>.
28. Duong H, Lindenbergh R, Pfeifer N, et al. (2009) ICESat full-waveform altimetry compared to airborne laser scanning altimetry over The Netherlands. *IEEE Trans Geosci Remote Sens* 47: 3365-3378.
29. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B Methodol* 7: 1-38.
30. Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 711-732.
31. Hernandez-Marin S, Wallace AM, Gibson GJ (2007) Bayesian analysis of Lidar signals with multiple returns. *IEEE Trans Pattern Anal Mach Intell* 29: 2170-2180.
32. Jaskierniak D, Lane PNJ, Robinson A, et al. (2011) Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions. *Remote Sens Environ* 115: 573-585.

33. Wu J, Van Aardt JAN, Asner GP, et al. (2009) Lidar waveform-based woody and foliar biomass estimation in savanna environments. *Proc Silvilaser* 1-10.
34. Parrish CE, Rogers JN, Calder BR (2014) Assessment of waveform features for Lidar uncertainty modeling in a coastal salt marsh environment. *IEEE Geosci Remote Sens Lett* 11: 569-573.
35. Muss JD, Aguilar-Amuchastegui N, Mladenoff DJ, et al. (2013) Analysis of waveform Lidar data using shape-based metrics. *IEEE Geosci Remote Sens Lett* 10: 106-110.
36. Hopkinson C, Chasmer L (2009) Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sens Environ* 113: 275-288.
37. Blair JB, Rabine DL, Hofton MA (1999) The Laser Vegetation Imaging Sensor (LVIS): A medium-altitude, digitization-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS J Photogramm Remote Sens* 54: 115-122.
38. Cook B, Dubayah R, Griffith P, et al. (2011) NACP New England and Sierra National Forests biophysical measurements: 2008–2010 [cited 2016 Aug 16]; Available from: [http://daac.ornl.gov/cgi-bin/dsvviewer.pl?ds\\_id=1046](http://daac.ornl.gov/cgi-bin/dsvviewer.pl?ds_id=1046).
39. Sun G, Ranson KJ (2000) Modeling lidar returns from forest canopies. *IEEE Trans Geosci Remote Sens* 38: 2617-2626.
40. Urban DL (1990) A versatile model to simulate forest pattern. In: A User's Guide to ZELIG Version 1.0. Charlottesville, VA: The University of Virginia [cited 2016 Jul 5]. Available from: <https://daac.ornl.gov/BOREAS/bhs/Models/Zelig.html>.
41. Lefsky MA, Harding D, Cohen WB, et al. (1999) Surface Lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sens Environ* 67: 83-98.
42. Harding DJ, Carabajal CC (2005) ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophys Res Lett* 32: L21S10.
43. García M, Riaño D, Chuvieco E, et al. (2010) Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sens Environ* 114: 816-830.
44. Salas EAL, Henebry GM (2012) Separability of maize and soybean in the spectral regions of chlorophyll and carotenoids using the Moment Distance Index. *Isr J Plant Sci* 60: 65-76.
45. Salas EAL, Henebry GM (2013) A new approach for the analysis of hyperspectral data: Theory and sensitivity analysis of the Moment Distance Method. *Remote Sens* 6: 20-41.
46. Lefsky MA, Harding DJ, Keller M, et al. (2005) Estimates of forest canopy height and aboveground biomass using ICESat: ICESAT estimates of canopy height. *Geophys Res Lett* 32: 441.
47. Jenkins JC, Chojnacky DC, Heath LS, et al. (2004) Comprehensive database of diameter-based biomass regressions for North American tree species. U.S. For. Servo Gen. Tech. Rep. NE-319.
48. Young HE, Ribe JH, Wainwright K (1980) Weight tables for tree and shrub species in Maine. Life Sciences & Agriculture Experiment Station Miscellaneous Report 230.



49. Cook BD, Rosette J, North PR, et al. (2010) Effect of ground surface reflectance on LiDAR waveforms, height metrics and biomass estimation. AGU Fall Meet Abstract [cited 2016 Jul 5]. Available from: <http://adsabs.harvard.edu/abs/2010AGUFM.B33A0386C>.
50. Andersen H-E, McGaughey RJ, Reutebuch SE (2005) Estimating forest canopy fuel parameters using LIDAR data. *Remote Sens Environ* 94: 441-449.
51. Muss JD, Mladenoff N, Townsend PA (2011) A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sens Environ* 115: 824-835.



AIMS Press

© 2016 Eric Ariel L. Salas, et al., licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)