

6-2017

Application of Iterative Noise-adding Procedures for Evaluation of Moment Distance Index for LiDAR Waveforms

Eric Ariel L. Salas

New Mexico State University, [easalas@nmsu.edu](mailto: easalas@nmsu.edu)

Sadichya Amatya

Geoffrey Henebry

South Dakota State University, [geoffrey.henebry@sdstate.edu](mailto: geoffrey.henebry@sdstate.edu)

Follow this and additional works at: https://openprairie.sdstate.edu/gsce_pubs

 Part of the [Geographic Information Sciences Commons](#), and the [Remote Sensing Commons](#)

Recommended Citation

Salas, Eric Ariel L.; Amatya, Sadichya; and Henebry, Geoffrey, "Application of Iterative Noise-adding Procedures for Evaluation of Moment Distance Index for LiDAR Waveforms" (2017). *GSCE Faculty Publications*. 66.
https://openprairie.sdstate.edu/gsce_pubs/66

This Article is brought to you for free and open access by the Geospatial Sciences Center of Excellence (GSCE) at Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in GSCE Faculty Publications by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact [michael.biondo@sdstate.edu](mailto: michael.biondo@sdstate.edu).



Research article

Application of Iterative Noise-adding Procedures for Evaluation of Moment Distance Index for LiDAR Waveforms

Eric Ariel L. Salas^{1,*}, Sadichya Amatya² and Geoffrey M. Henebry³

¹ Department of Fish, Wildlife and Conservation Ecology, New Mexico State University, Las Cruces, NM 88003-8003, USA

² Pennoni Associates, Virginia Beach, VA 23452, USA

³ Geospatial Sciences Center of Excellence (GSCE), South Dakota State University, Brookings, SD 57007-3510, USA

* **Correspondence:** easalas@nmsu.edu; Tel: +1-575-646-2691.

Abstract: The new Moment Distance (MD) framework uses the backscattering profile captured in waveform LiDAR data to characterize the complicated waveform shape and highlight specific regions within the waveform extent. To assess the strength of the new metric for LiDAR application, we use the full-waveform LVIS data acquired over La Selva, Costa Rica in 1998 and 2005. We illustrate how the Moment Distance Index (MDI) responds to waveform shape changes due to variations in signal noise levels. Our results show that the MDI is robust in the face of three different types of noise—additive, uniform additive, and impulse. In effect, the correspondence of the MDI with canopy quasi-height was maintained, as quantified by the coefficient of determination, when comparing original to noise-affected waveforms. We also compare MDIs from noise-affected waveforms to MDIs from smoothed waveforms and found that windows of 1% to 3% of the total wave counts can effectively smooth irregularities on the waveform without risking of the omission of small but important peaks, especially those located in the waveform extremities. Finally, we find a stronger positive relationship of MDI with canopy quasi-height than with the conventional area under curve (AUC) metric, e.g., $r^2 = 0.62$ vs. $r^2 = 0.35$ for the 1998 data and $r^2 = 0.38$ vs. $r^2 = 0.002$ for the 2005 data.

Keywords: Moment Distance Index; MD framework; waveform LiDAR; MDI; noise models

1. Introduction

The active remote sensing using LiDAR has seen rapid developments in the past two decades. With a promise of improved accuracy of biophysical measurements and the spatial analysis done in the third dimension, LiDAR could play an important role in atmospheric and environmental field of studies. In fact, NASA's future launch of the Global Ecosystem Dynamics Investigation (GEDI) LiDAR space mission in 2018 [1] could provide the waveform laser scanning technology more boost. The full-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, therefore showing position of individual targets, and finer details of the signature of intercepted surfaces or the proportion of the canopy complexity. Information associated with the illuminated object can be decoded from the generated backscattered waveform, as key features of the waveform such as the shape, area, and power are directly related to the geometry of the illuminated object [2–4]. The richness of the LiDAR waveform holds the promise to address the challenge of characterizing in detail the geometric and reflection characteristics of vegetation structure, e.g. the vertical canopy volume distribution [5].

Waveform LiDAR has been used comprehensively in various ecosystem-related studies. It has shown to accurately retrieve canopy height [6–10], tree form [11], terrain relief [12], canopy architecture of urban vegetation [13], and classify species [14–16]. It can reduce the cost of mapping large forest regions [17], rapidly record vertical canopy profiles [18], and provide a more defined vertical arrangement of forest structure from canopy top to ground surface [6,18,19]. Data from large-footprint waveform LiDAR systems have been utilized to estimate LAI and canopy cover [20–22], foliage density [23] and to improve the estimation of biomass—changes in LiDAR vertical canopy profiles and the mean canopy height metric were correlated with estimated aboveground biomass [5,18,20,24–28]. Recent studies used waveform LiDAR for mapping change of forest biomass [29], canopy layering [30], and foliage profiles [31].

Usually, the raw incoming/received waveform displays system noise [32–34]. The noise can easily overlap returns especially in complex forested areas where waveform peaks from ground and surface objects can be broadened and mixed caused by different layers of the vegetation, making the recognition of ground surface difficult [10,35]. In most cases, noise reduction or elimination may be conducted [36] to extract the waveform intensity and avoid the noise to be detected as signal. This is done, for instance, through smoothing with a Gaussian filter at a specified window size [9]. Smoothing, however, may pose major risk in eliminating important snags and understory shrubs features [37] that are usually found at the broadened end of the waveform near the ground return.

Detection of the ground surface return is needed to extract the canopy height using the direct method, which involves the identification of wave signal start (WSS) and wave ground peak (WGP) on the waveform. There is no widely accepted method for estimating the locations of the WSS and WGP. One approach currently utilized sets thresholds [10,36] above the mean background noise in the waveform. Thresholds vary from 3σ [9], 4σ [35], to 4.5σ [8], where σ is the standard deviation of the background noise. In either smoothed or unsmoothed waveforms, thresholding can pose a risk of removing the broadening effects at the extents of the waveform that may carry vital information about understory vegetation or structure. The waveform is also susceptible to extreme values in the background that could cause premature peaking, which eventually could lead to nonsensical estimates of canopy height [34]. Therefore, it is necessary to come up with a better threshold limit or a sufficient smoothing process that could improve data extraction and pull out the true signal without jeopardizing the information that may be available at every change of the morphology of the LiDAR waveform. Appropriate parametric functions may be applied to the waveform to reconstruct the shape and retrieve information about the object and characterize the properties. Conventional LiDAR methods include splines [38], the Gaussian mixture models [32,39], and the non-linear least-square approach [40]. In many mapping applications, Gaussian approximation has been shown to be satisfactory for signal modeling, although Gaussian fitting is less satisfactory for high amplitude pulses [41].

The preprocessing of full-waveform LiDAR data usually leads to using only part of the return signal. In most cases the full intensity of the LiDAR return is rarely used. Hopkinson & Chasmer [42] emphasized the importance of using the intensity of the LiDAR returns, especially in canopy fractional cover models, since the intensity values provide some quantification of the surface areas interacting with the laser beam. In characterizing the intercepted scenes, it is essential to know the full geometry (shape) and radiometry (power) of the signal as both could explain the geometry and radiometry of the detected object. Muss et al. [43] took advantage of the geometry and radiometry of the waveform to introduce shape-based metrics—centroid (C) and radius of gyration (RG)—for forest structure analysis. While these metrics demonstrated better relationships with estimated aboveground biomass (EAGB) than traditional height-based metrics such as height of median energy (HOME) (*e.g.*, [24,25]) and relative heights (RH) or height percentiles (*e.g.*, [9,44]), the centroid by itself cannot track changes of the waveform shape. Expressed as the root mean square of the sum of the distances from the centroid to all points on the waveform, RG is dependent on the centroid and cannot stand alone.

Here we use full-waveform LVIS datasets from La Selva, Costa Rica, to assess the strength of the new Moment Distance framework, which was first introduced for characterizing fine resolution spectrometer data [45,46], and later on seen to be more useful for detecting plastic greenhouses [47], mapping sparse vegetation [48] and estimating canopy heights using LiDAR waveform [49]. To study shapes we use the full unsmoothed LVIS Geolocated Waveform (.lgw) dataset. We assess how different types of noise and smoothing procedures impact the new metric in characterizing the shape of the full waveform extent and waveform subsets. As noise can be

expected to alter the shape of the waveform, we examine the sensitivity of the new metric to varying levels of uncertainty and in various subsets of the waveform, using metric obtained from the original waveform as the reference. The performance of our method is investigated by how the new metric performs against the canopy quasi-height. We smooth the waveform using various window sizes to demonstrate the relationship of the MDI against the canopy quasi-height and compared the results to the area under curve (AUC) method [50,51].

2. Materials and Method

2.1. Waveform LiDAR datasets

The full-waveform LiDAR datasets were acquired in 1998 and 2005 over the same period in La Selva, Costa Rica using the Laser Vegetation Imaging Sensor (LVIS), which is an airborne NASA laser scanning altimeter. We specifically used the LVIS geolocated waveform (.lgw) files from 1998 and 2005, with both having the same number of 431 wave counts. The LVIS laser device produces Gaussian optical pulses at a wavelength of 1064 nm [17]. We paired LiDAR waveform samples from 1998 and 2005 datasets based on the latitude and longitude coordinates. The pairings served as inputs for generating noise-affected waveforms in our analysis.

LVIS waveform data are easily converted into distance since the signal returns are measured as a function of time. Accounting both the times the laser pulse was emitted and returned could give a measure of the distance from the sensor to the intercepted surface. LVIS has a scan angle of about 12 degrees, and could cover 2 km swaths of surface from an altitude of 10 km, with 10 to 25 m footprint size. We estimated the canopy height (we refer it as quasi-height) from the waveform as the difference from the power of the first increase of return above the mean noise level to the center of the last pulse, which is designated as the ground return.

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. The approach takes advantage of the multiple 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. In addition, it means avoiding 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. It 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:

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

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

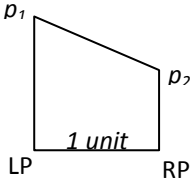
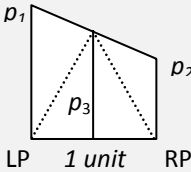
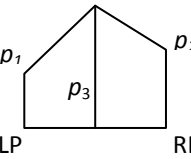
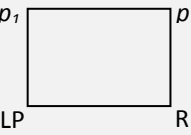
$$MDI = MD_{LP} - MD_{RP} \quad (3)$$

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 .

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 wave counts 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 multiple wave counts 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 ill perform and fail to define the waveform shape when there are only few points between pivots. Table 1 presents some limiting cases when computation of the MDs and MDI are not appropriate.

Figure 1 demonstrates the MDI when applied to lines. Illustrated simply, a curve opening down (Figure 1A) will differ from a curve opening up (Figure 1B) when defined by moment distances with varying pivot ranges. In Figure 1C, for instance, a two-peak, opening-down curve defined by fixing the early time wave count and increasing the range one wave count at a time (going from point 1 to point 2) a slope becomes evident when a second peak is contained within the pivot range. The dip around count 187 defines the largest difference of the change of shape detected by a particular pivot pair. A similar pattern is observed when fixing the late time wave count and increasing the range one count at a time (going from point 2 to 1). As shape dissimilarities are detected by comparing MD behavior from point 1 to MD behavior from point 2, the largest difference in shape is recorded when the earlier peak is covered by the range.

Table 1. Simple cases where the MD approach would not be suitable. Note that MDI exploits multiple bins and limiting the number of bins may fail to define the waveform shape.

Cases	Equations of MDI	Remarks
	$MD_{LP} = \sqrt{(1^2 + y_2^2)}$ $MD_{RP} = \sqrt{(1^2 + y_1^2)}$ <p style="text-align: center;">if:</p> $p_1 = p_2, MDI = 0$ $p_2 < p_1, MDI = \text{negative value}$ $p_2 > p_1, MDI = \text{positive value}$	<p>When the pivots are the only values, then a curve is not described. This defeats the purpose of using MDI to define the shape of a line that can be described by the slope direction. The sign of the magnitude of the MDI changes when $p_1 < p_2$.</p>
	$MDI = 0$	<p>Missing to include the backscatter powers p_1 and p_2, and only used p_3. No shape has been defined in this case. Missing significant returns of the waveform must be avoided.</p>
	$p_1 < p_2 < p_3$ $MDI = \sqrt{(4 + p_2^2)} - \sqrt{(4 + p_1^2)}$	<p>Too few points. Assuming p_3 as noise, this illustrates the capability of the MDI to eliminate unwanted returns.</p>
	$MDI = 0$	<p>MDI is not the appropriate approach in this type of curve involving only two returns with the same magnitude. This defeats the purpose of the new metric.</p>

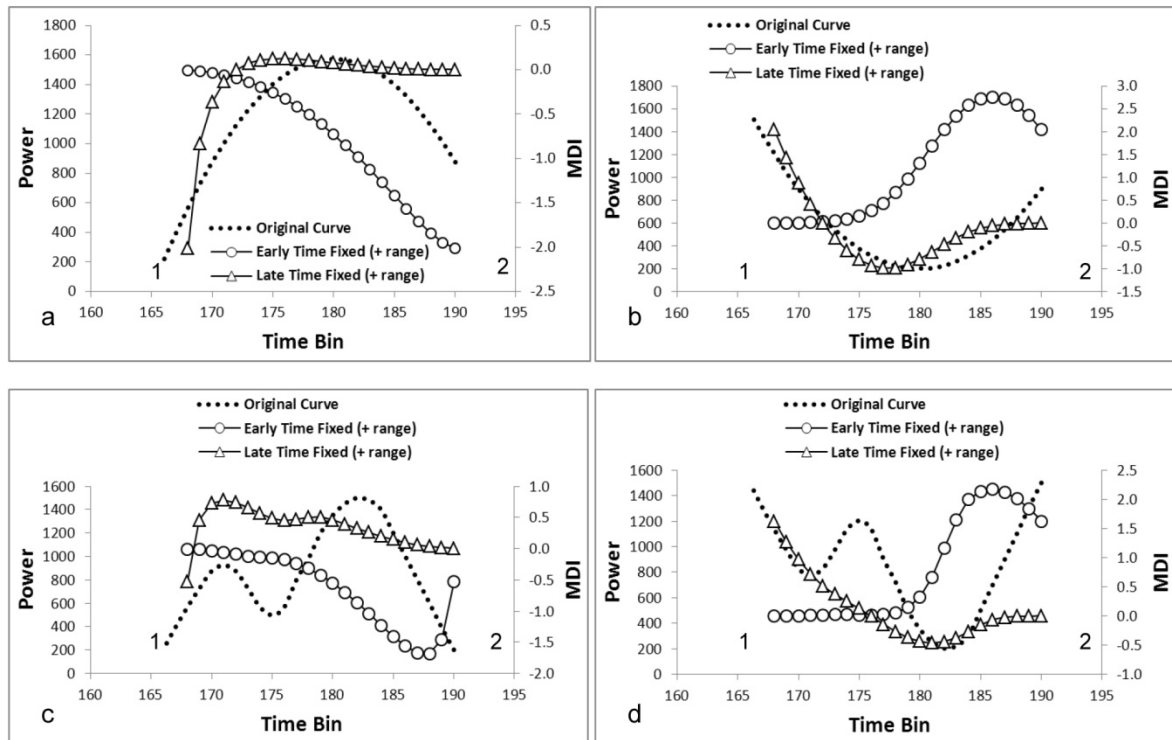


Figure 1. Sample illustrations of MDI applied to simple curves, with: (a) curve opening down, (b) curve opening up, (c) two curves opening down, and (d) two curves opening up. The figures demonstrate the changes of the MDI values with varying pivot ranges, moving from point 1 to 2, and vice versa. Maximum values are observed at maximum shape differences, usually occurring at the inclusion of peaks.

In a curve with two dips, such as in Figure 1D, the maximum differences in shape around bins 180 to 185 occur when the dips are included in the range. Minimal differences in shapes are expected for short pivot ranges. It is important to stress that the MDI detects the differences of curvature regardless of where the pivot is being fixed, such as the cases in Figure 1, provided that the important peaks are taken into consideration in the pivot range. Detecting the differences is crucial in the following analysis using the LiDAR waveforms as various peaks and dips may exist from the first signal to the last signal detected. Using the proper range that encompasses the significant peaks of the waveform could lead to the shape difference maxima: the maximum difference of the summation of distances from point 1 and the summation of distances from point 2. The value of the difference tells how the shape of the waveform as viewed from reference point 1 varies from the one viewed from reference point 2.

Figure 1 illustrates that the selection of LP and RP may not necessarily be exactly at the start of signal and end of signal, respectively. In equations 1 and 2, there is no fixed location of pivots; instead, an option is given to pick the range of wave counts. The locations of LP and RP do not have a strict limitation as to where they should be placed exactly along the wave count axis, thus, making MDI exploitable at different locations. The location of the MDI maxima often occurs right before or

right after signal peaking. Elimination of few points after the first detected signal and/or points before the last detected signal may not hurt the MDI nor affect the detection of shapes from any of the set fixed pivots.

2.3. Generation of noise-affected waveforms

We added three types of random noise to the LiDAR signal: additive (AD), uniform additive (UA), and impulse (IM). The additive Gaussian noise model [52] is the simplest noise model that consists in adding a realization of a zero mean random vector to a clean return signal. The impulse noise [53] is a different type of noise that consists of sparse impulsions, generated by a random distribution with slowly decaying probability. The additive uniform noise [54] is the kind of noise where noise is uniform in a given waveform interval. We define the resulting noisy waveform $I_n(t)$ by the sum of the original waveform $I_o(t)$ and the added noise $n(t)$ where t indicates time.

$$I_n(t) = I_o(t) + n(t) \quad (4)$$

We individually ran the additive Gaussian, uniform additive, and impulse error algorithms to the set of paired LiDAR waveforms from the 1998 and 2005 La Selva datasets. This process included generating 1000 noise realizations per type of noise model at four levels of uncertainty (5%, 10%, 15%, and 20%), and in two waveform subsets analyzed—leading and trailing subsets. More of these two subsets are discussed below. We observed waveform behaviors such as how it responds to noise and how certain types of waveform morphology changes to different types of noise. We then looked at the MDI as a function of the original quasi-height (calculated height from original waveform) to assess how the introduction of the levels of uncertainty to the waveforms using three different types of noise models changed the initial observed trend of the original MDI against the canopy quasi-height.

2.4. Smoothing waveforms

In addition to adding noise, we also looked at the behavior of the MDI when waveforms were smoothed to different degrees. We eliminated the small peaks using forward-moving average, backward-moving average, and centered-moving average window approaches as described by Savitzky & Golay [55] and Madden [56] by calculating the average of power of adjacent bins [12,57]. The forward-moving average smoothens the waveform by moving a specified window forward in time, while the backward-moving average smoothens the waveform as it moves backward in time. Since MD is shape-dependent, care was taken not to over-smooth the waveform (*i.e.*, using too large a window or percent wave counts). As much as possible, we preserved the significant peaks, such as the canopy and the ground [58]. Moving windows of 1% to 3% of the total wave count were used to smooth irregularities without degrading waveform morphologies.

2.5. Calculation of MDI from full waveform extent and waveform subsets

The full extent of the recorded waveform, plus two defined subsets—leading and trailing subsets [8] were used to illustrate the impacts of noise integration to the waveforms and to the value of the MDI (Figure 2). The full extent covers the range starting at the first detected pulse up to the last detected pulse. The leading subset covers the range of time bins from the start of the signal to the location of the maximum peak at early time (above the mean noise level). The trailing subset defines the range from the location of the maximum peak at early time to the maximum peak at a later time (ground response). It is worth noting that a subset can be any specific and narrow range and it is always defined by the locations of the left and right pivots (represented by gray dots in Figure 2) that necessary for the computation of MDI.

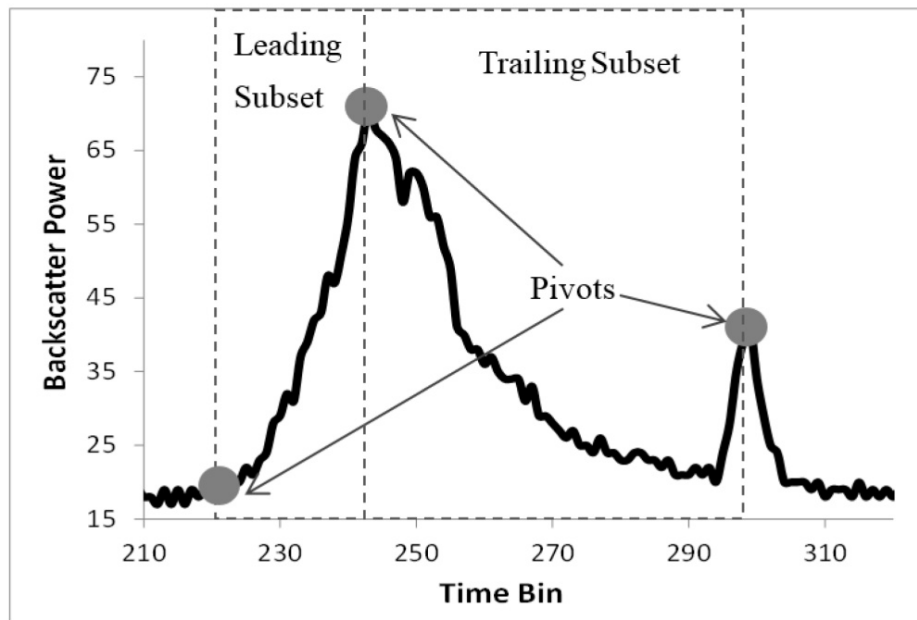


Figure 2. Sample of LVIS waveform and subsets used for MDI calculations. A pair of pivots defines the range of a subset. Leading subset is defined by the first pivot (signal start) to the second pivot (maximum peak early time). Trailing subset is defined by first pivot (maximum peak early time) to the second pivot (maximum peak later time). The abscissa is in terms of time (t) and the ordinate is the backscattered power (p).

Between the two subsets, the trailing subset includes two important waveform morphologies (maximum peaks at early and later times) in its range. These two waveform morphologies are important as both could dictate the magnitude of MDI [48]. In the later section of this chapter, we made use of canopy quasi-heights to describe how each subset, defined by range at key profile locations of the waveform, relates to our new metric.

3. Results

3.1. Generated noise-affected waveforms

Figure 3 give insights to a noise generated at 10% for a pair of waveform samples. Note that each pair in the dataset may illustrate different behaviors of the waves. Sample 1998 (LVIS waveform file ID 666, shot number 574836) shows two distinct peaks representing the canopy (maximum peak earlier time) and ground (maximum peak later time) return pulses. The corresponding pair in 2005 (LVIS waveform file ID 8983, shot number 2607680) also manifests the two peaks, with the addition of an earlier shorter peak. Among the three error algorithms, the impulse type differed in distribution of noise because of the appearances of spikes along the waveform.

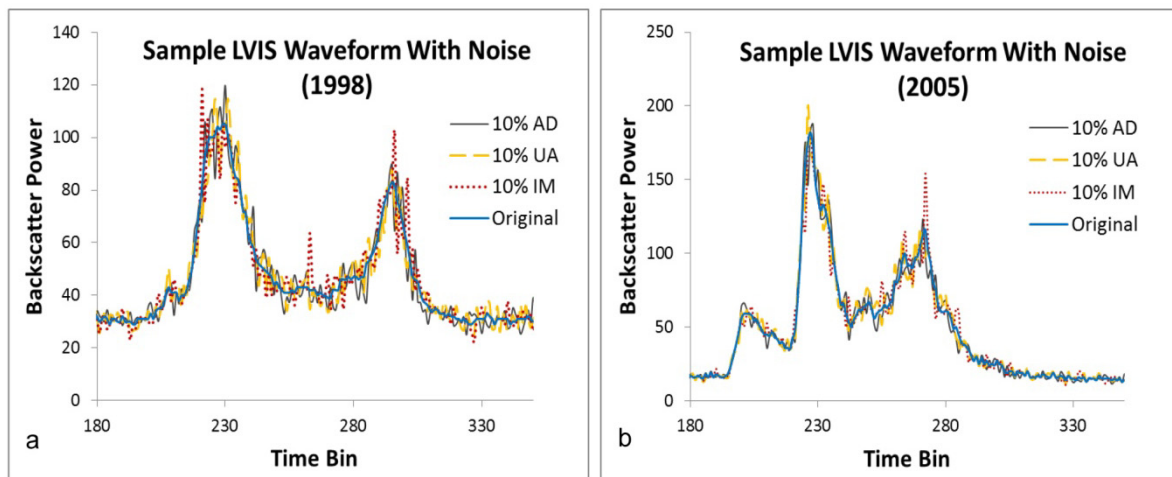


Figure 3. A pair of LVIS waveform samples from (a) 1998 (LVIS waveform file ID 666, shot number 574836) and (b) 2005 (LVIS waveform file ID 8983, shot number 2607680) datasets showing the original and generated waveforms with 10% noise. The noise is randomly simulated and applied throughout the wave extent. AD = Additive Noise; UA = Uniform Noise; IM = Impulse Noise.

3.2. Error effects on waveform shape and the MDI

Table 2 lists the statistics of the MDI from the averaged 1000 noise-affected waveforms for a sample pair. Using the full extent of the waveform (Table 2a), large MDI values for 1998 and 2005 were observed with the impulse error algorithm. Negative MDI values were seen using the full extent and leading subset, while positive values are generated by the trailing subset.

Table 2. MDI statistics of the original and 1000 generated waveforms for a pair of samples from 1998 (waveform file ID 666) and 2005 (waveform file ID 8983) datasets. Table 2a is for the full extent, 2b for the leading subset, and 2c for the trailing subset. Notice that the impulse showed the most variability estimated with CV, especially at high uncertainty levels.

2A. Full Extent: Original MDI (1998) = -742.33						Full Extent: Original MDI (2005) = -259.68					
	1998	5%	10%	15%	20%		2005	5%	10%	15%	20%
AD	Mean	-741.9	-742.2	-743.1	-744.6	AD	Mean	-259.6	-260.2	-261.4	-263.2
	CV	2.0	4.1	6.1	8.0		CV	2.4	4.7	7.0	9.2
UA	Mean	-743.4	-745.1	-747.2	-750.1	UA	Mean	-260.3	-261.5	-263.3	-265.8
	CV	2.1	4.3	6.4	8.4		CV	2.3	4.6	6.8	9.0
IM	Mean	-744.6	-748.3	-753.5	-761.3	IM	Mean	-261.2	-263.3	-269.9	-280.4
	CV	2.0	5.4	12.8	22.8		CV	2.3	14.2	44.7	83.1
2B. Leading Subset: Original MDI (1998) = -29.12						Leading Subset: Original MDI (2005) = -37.36					
	1998	5%	10%	15%	20%		2005	5%	10%	15%	20%
AD	Mean	-29.1	-29.2	-29.3	-29.5	AD	Mean	-37.3	-37.3	-37.4	-37.4
	CV	2.7	5.4	8.3	11.6		CV	3.4	6.9	10.6	14.6
UA	Mean	-29.1	-29.2	-29.4	-29.6	UA	Mean	-37.4	-37.5	-37.6	-37.7
	CV	2.9	5.8	8.8	12.0		CV	3.6	7.2	10.9	14.7
IM	Mean	-29.7	-30.3	-30.9	-31.4	IM	Mean	-38.1	-39.0	-39.7	-40.4
	CV	2.9	6.7	10.6	13.9		CV	3.5	7.8	12.1	15.7
2C. Trailing Subset: Original MDI (1998) = 92.17						Trailing Subset: Original MDI (2005) = 233.23					
	1998	5%	10%	15%	20%		2005	5%	10%	15%	20%
AD	Mean	92.3	92.2	92.0	91.5	AD	Mean	233.2	232.9	232.2	231.0
	CV	5.6	11.3	17.3	23.8		CV	2.0	4.1	6.3	8.7
UA	Mean	92.1	91.9	91.5	90.9	UA	Mean	233.1	232.6	231.7	230.5
	CV	5.8	11.7	17.8	24.1		CV	2.1	4.2	6.4	8.7
IM	Mean	90.7	88.8	86.9	84.3	IM	Mean	232.0	230.3	228.1	225.1
	CV	5.8	12.3	19.4	26.4		CV	2.1	4.4	7.0	9.4

Statistical results from the leading subset are presented in Table 2b. Among the three error algorithms, impulse showed the most variability, especially at high uncertainty levels. With 15% and 20% uncertainty levels, the coefficient of variations (CV) for 1998 are 10.6% and 13.9%, respectively; for 2005 are 12.1% and 15.7%, respectively.

For the trailing subset (Table 2c), MDIs exhibited positive values. Statistical calculations showed minimum MDI values were generated using impulse at different levels of errors, ranging from 8.17 to 74.0 for the 1998 sample. A large MDI range was manifested in the 2005 sample (139.36 to 217.23). Smaller ranges were observed from the additive (49.49 and 47.52, for 1998 and 2005 respectively)

and the uniform additive (60.50 and 54.84, for 1998 and 2005 respectively) for the trailing subset.

The mean MDI values of the additive and the uniform additive approaches are comparable at various error levels in both subsets (Table 2). In fact, increasing the level of noise from 5% to 20% showed no big differences of the MDI means between AD and UA in both years. Differences of MDI means were observed in AD vs IM, and UA vs IM, after significance testing for both the leading ($p < 0.05$ in all tests) and trailing subsets ($p < 0.05$ for 1998; $p < 0.10$ for 2005). The full extent of the waveform, however, showed no significant differences of the means in any of the one-to-one comparisons (AD vs UA, AD vs IM, UA vs IM). The MDI increased negatively in the leading and full-extent with every increase of uncertainty. The trailing subset, on the other hand, had decreasing positive MDI at each increase of uncertainty.

3.3. Standard Error (SE) and RMSE of MDI

Large MDI values as the noise levels increased is evident on the bar graphs of MDI and standard error of 1000 noise realizations (Figure 4, full extent of the wave). At lower noise levels, the SE is lower relative to higher noise levels. As the noise level increased the average MDI further deviates from the reference MDI, while the SE increased alongside. The impulse approach showed large increases of SE at differing noise levels. These increases can be observed for both the tested 1998 and 2005 samples.

Computed RMSE (Table 3) showed the performance of each noise approach using two pairs of waveform samples both from 1998 and 2005 datasets. The additive and the uniform additive approaches resulted to smaller values of RMSE compared to the impulse. However, when the level of noise is low (e.g., 5%), the three approaches were comparable.

Table 3. Computed RMSE of MDI for two pairs of samples using the full extent of the waveform in four percentage levels of noise: (A) 1998 waveform file ID 666; 2005 waveform file ID 8983 and (B) 1998 waveform file ID 9999; 2005 waveform file ID 1408. The highest RMSE values are found using the IM approach, especially at high levels of noise.

<i>A. 1998</i>	<i>5%</i>	<i>10%</i>	<i>15%</i>	<i>20%</i>	<i>A. 2005</i>	<i>5%</i>	<i>10%</i>	<i>15%</i>	<i>20%</i>
AD	15.12	30.16	45.06	59.06	AD	6.11	12.20	18.30	24.44
UA	15.95	31.90	47.81	63.65	UA	6.03	12.12	18.31	24.64
IM	14.90	40.63	97.04	174.80	IM	6.12	37.72	120.99	233.84
<i>B. 1998</i>	<i>5%</i>	<i>10%</i>	<i>15%</i>	<i>20%</i>	<i>B. 2005</i>	<i>5%</i>	<i>10%</i>	<i>15%</i>	<i>20%</i>
AD	11.90	23.75	35.51	47.13	AD	5.29	10.56	15.80	20.99
UA	12.96	25.91	38.82	51.66	UA	5.12	10.23	15.32	20.38
IM	12.15	36.87	94.80	173.91	IM	5.19	37.30	120.89	233.04

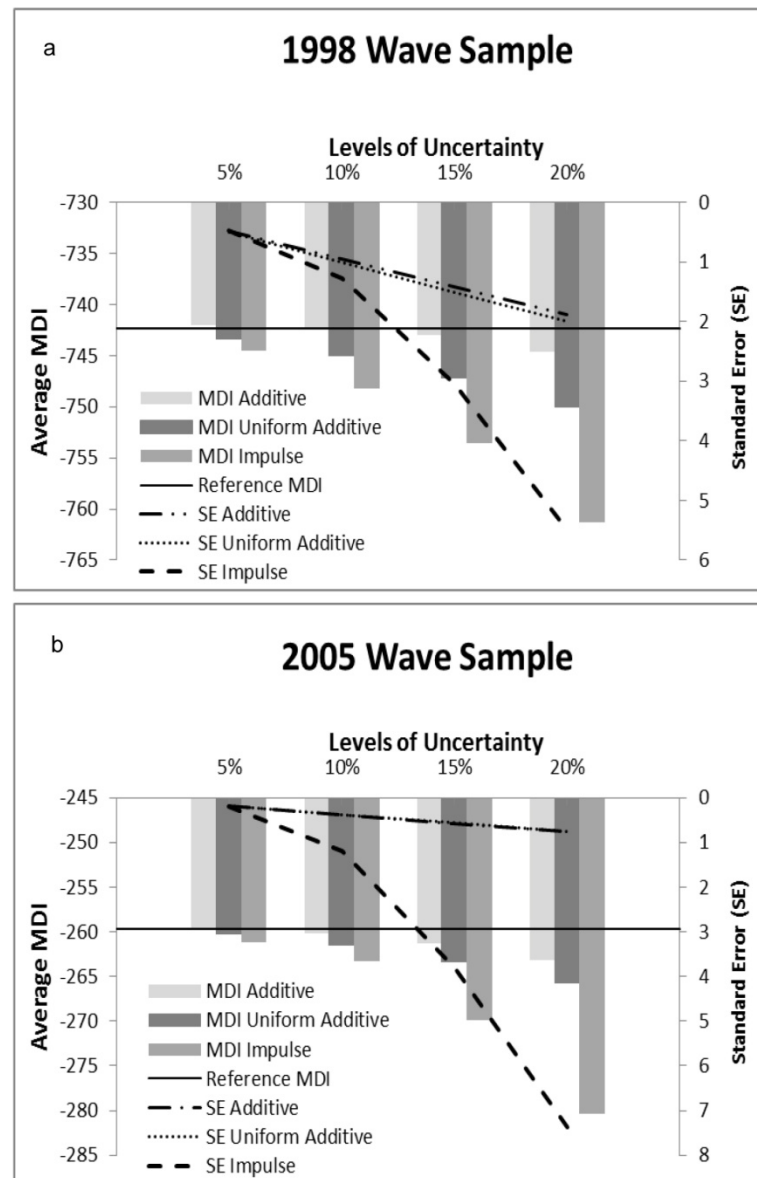


Figure 4. Average MDI and Error bar plots for 1000 noise realizations analyzed for each noise approach, using the full extent of the waveform: (a) for 1998 (waveform file ID 666) and (b) for 2005 (waveform file ID 8983). Take note of the increasing values of the MDI as the levels of noise are increased. The impulse noise shows abrupt increase of the standard deviation of the mean. The reference MDI is shown as horizontal line.

Figures 5 and 6 show the average MDI and the equivalent standard errors for the leading and trailing portions of the waveform, respectively. Similar to the results in Figure 4, the impulse model resulted to an average MDI with the highest offset from the reference value with respect to the other two noise models.

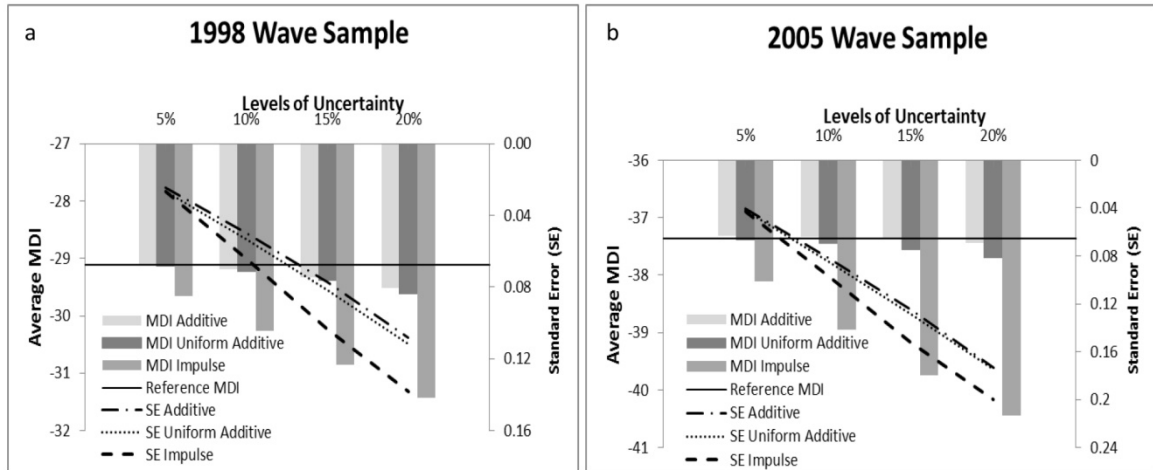


Figure 5. Using the leading subset of the waveform, the plots show the average MDI and the equivalent standard errors for 1000 noise realizations analyzed for each noise types at various levels of uncertainty. The left panel is for the 1998 sample (waveform file ID 666), while the right panel is for its matched pair in 2005 (waveform file ID 8983). All values of MDI are negative, with the impulse approach having the largest SE. The reference MDI is shown as horizontal line.

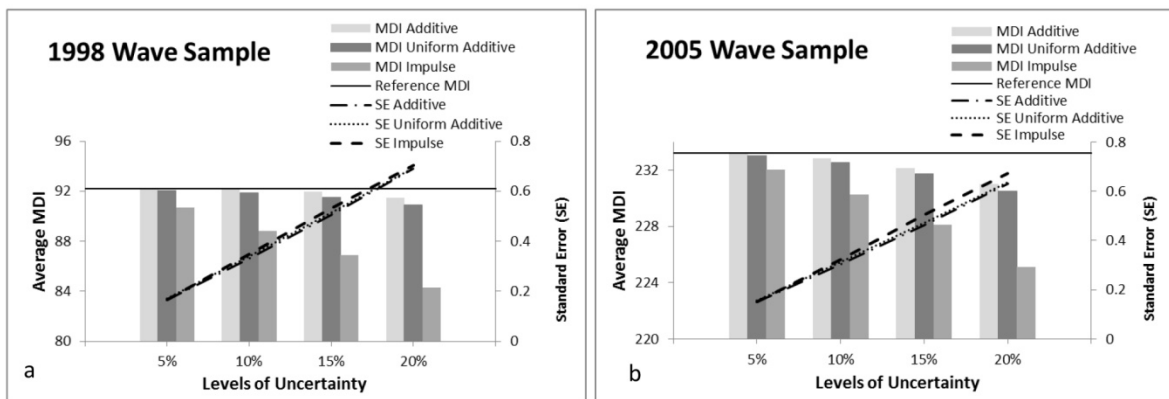


Figure 6. Using the trailing subset of the waveform, the plots show the average MDI and the equivalent standard errors for 1000 noise realizations analyzed for each noise types at various levels of uncertainty. The left panel is for the 1998 sample (waveform file ID 666), while the right panel is for its matched pair in 2005 (waveform file ID 8983). In this subset, all values of MDI are positive, with the impulse approach having the largest SE. The reference MDI is shown as horizontal line.

3.4. Moment distance index vs canopy quasi-heights

A positive relationship is manifested in Figure 7 between the MDI from full extent and the

derived canopy quasi-heights from 16 pairs of waveforms. As shown in Table 4, the relationship is true for both the years 1998 and 2005, with the results from 1998 exhibiting stronger linear relationship (averaged $r^2=0.62$ vs. averaged $r^2=0.38$). The trailing subset shows a similar trend (Figure 8) as the one using the full extent of the waveform, albeit with less explanatory power (averaged $r^2=0.30$ vs. averaged $r^2=0.22$). However, even at a high noise level (20%), the trend does not changed.

Table 4. Coefficients of determination (r^2) of the MDI vs canopy quasi-heights for the 1998 and 2005 data. Values are kept to four significant figures to show differences.

<i>1998 Full Extent</i>							<i>1998 Trailing Subset</i>						
Noise	Reference	5%	10%	15%	20%	Ave	Reference	5%	10%	15%	20%	Ave	
AD	0.6168	0.6169	0.6176	0.6218	0.6206	0.6192	0.2956	0.2952	0.2952	0.2956	0.2938	0.2950	
UA	0.6168	0.6168	0.6173	0.6272	0.6200	0.6203	0.2956	0.2953	0.2953	0.2949	0.2944	0.2950	
IM	0.6168	0.6175	0.6186	0.6170	0.6222	0.6188	0.2956	0.2926	0.287	0.2821	0.2778	0.2849	
<i>2005 Full Extent</i>							<i>2005 Trailing Subset</i>						
Noise	Reference	5%	10%	15%	20%	Ave	Reference	5%	10%	15%	20%	Ave	
AD	0.3756	0.3755	0.3756	0.3757	0.3760	0.3757	0.2236	0.2238	0.2239	0.2239	0.2238	0.2239	
UA	0.3756	0.3757	0.3758	0.3761	0.3765	0.3760	0.2236	0.2238	0.2239	0.2238	0.2237	0.2238	
IM	0.3756	0.3757	0.3759	0.3762	0.3765	0.3761	0.2236	0.2065	0.2043	0.2033	0.2001	0.2036	

Table 5. Computed RMSE of MDI with 16 pairs of samples from the 1998 and 2005 noise-affected waveforms using the full extent and the trailing subset. Notice that the highest RMSE values are found using the IM approach.

<i>Full Extent (1998)</i>					<i>Full Extent (2005)</i>				
Noise	5%	10%	15%	20%	Noise	5%	10%	15%	20%
AD	0.352	0.395	1.793	2.613	AD	0.0979	0.396	1.179	2.383
UA	0.957	2.457	5.180	7.232	UA	0.3969	1.197	2.407	4.026
IM	2.277	6.023	20.87	18.78	IM	1.2523	2.855	8.709	18.13
<i>Trailing Subset (1998)</i>					<i>Trailing Subset (2005)</i>				
Noise	5%	10%	15%	20%	Noise	5%	10%	15%	20%
AD	0.121	0.658	1.749	3.411	AD	0.310	1.441	3.457	6.417
UA	0.140	0.705	1.798	3.433	UA	0.344	1.510	3.537	6.464
IM	2.021	4.750	7.796	11.66	IM	1.819	4.560	7.983	12.47

From the 16 pairs of samples, we measured the magnitude of the error between the reference MDI and corresponding observed MDI from noise-affected waveforms using RMSE. Table 5 shows

the results for each approach using the full waveform and the trailing subset. Small RMSE values were related to 5% noise on the waveform while the high RMSE values were from the 20% noise level, regardless of the type of noise approach. The highest RMSE values were found using the IM approach.

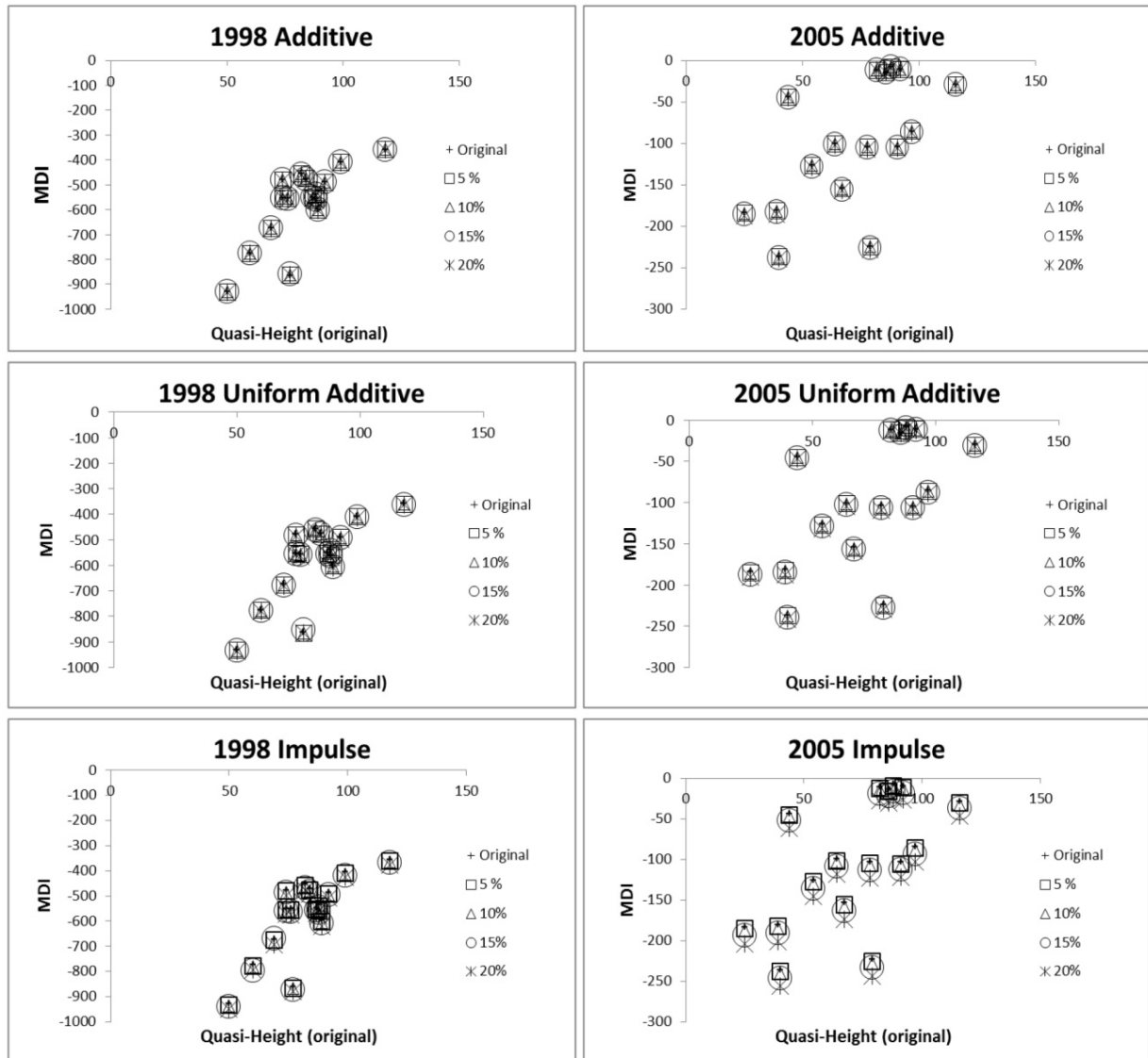


Figure 7. MDI (from full-extent waveform) as a function of canopy quasi-height for 16 pairs of the 1998 and 2005 La Selva LVIS datasets. First column: 1998 dataset with added noise levels and; Second column: 2005 dataset with added noise levels. A minimal effect of the noise is observed on the relationship of the MDI with the canopy quasi-height. See Table 4 for the coefficients of determination.

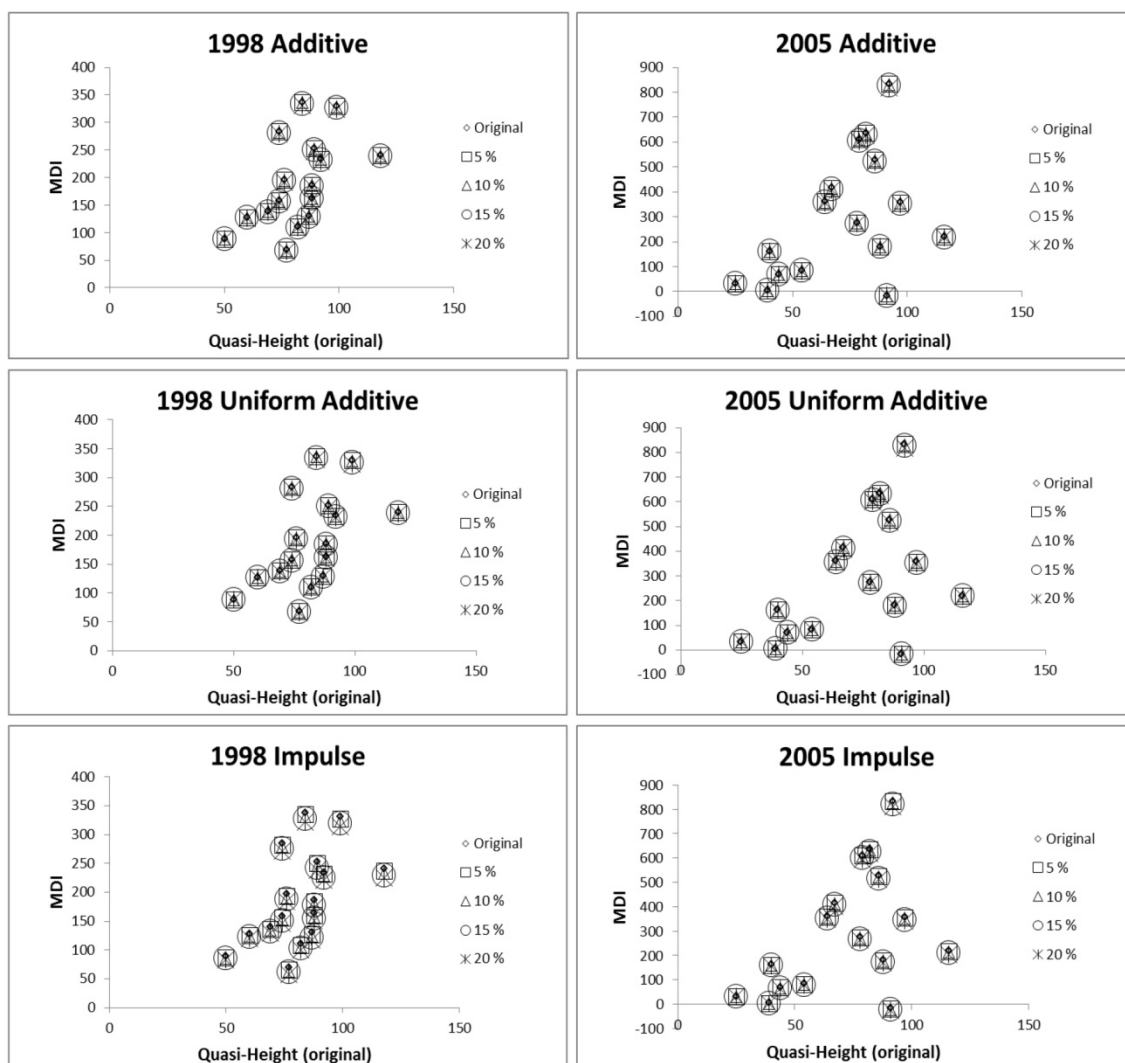


Figure 8. MDI (from waveform trailing subset) as a function of canopy quasi-height for 16 pairs of waveforms from the 1998 and 2005 La Selva LVIS datasets. First column: 1998 dataset with added noise levels and; Second column: 2005 dataset with added noise levels. See Table 4 for the coefficients of determination.

3.5. Waveform similarities

We compared waveform components between two signals using correlation analysis on the 1998 and 2005 pair of subsets at different noise levels. We tabulated in Table 6 the correlations between two waveforms (original and noise-affected) with varying signal powers, putting emphasis on the Spearman rank-order correlation coefficient, r_s values. Comparing between subsets in Table 6, the leading subsets for both years showed higher r_s than the trailing subset. For instance, the 2005 leading subset had a minimum $r_s = 0.88$ (Table 6B1), while its equivalent in the trailing subset had $r_s = 0.72$ (Table 6B2). In the 1998 leading subset, the IM showed an $r_s = 0.91$ (Table 6A1), while its equivalent trailing subset was only $r_s = 0.67$ (Table 6A2).

Table 6. Spearman correlation coefficients (r_s) for the leading (A1/B1) and trailing (A2/B2) subsets. The r_s measures the strength of the associations between the original waveform segment subset and the generated subsets with noise (significant at $p = 0.05$). Note that the statistics of the pair of samples are from 1998 (waveform file ID 666) and 2005 (waveform file ID 8983).

<i>A1. Leading Subset (1998)</i>					<i>A2. Trailing Subset (1998)</i>				
Noise	5%	10%	15%	20%	Noise	5%	10%	15%	20%
AD	0.97	0.93	0.91	0.85	AD	0.95	0.87	0.81	0.75
UA	0.98	0.96	0.94	0.91	UA	0.96	0.90	0.83	0.76
IM	0.97	0.94	0.92	0.91	IM	0.94	0.85	0.76	0.67
<i>B1. Leading Subset (2005)</i>					<i>B2. Trailing Subset (2005)</i>				
Noise	5%	10%	15%	20%	Noise	5%	10%	15%	20%
AD	0.98	0.95	0.92	0.89	AD	0.97	0.93	0.88	0.80
UA	0.97	0.95	0.92	0.88	UA	0.97	0.90	0.84	0.73
IM	0.98	0.97	0.95	0.92	IM	0.96	0.89	0.82	0.72

3.6. Comparison between noise levels

Comparing among the noise levels, waveforms with less noise showed a higher degree of similarity to the original curve than those with high noise levels. From 5% to 20% noise level, the waveforms with 5% additive, uniform additive, and impulse noise resulted in much closer agreements (e.g. trailing subset, Table 6A2—the additive 5% has $r_s = 0.95$; the 10% has $r_s = 0.87$; the 15% has $r_s = 0.81$; and the 20% has $r_s = 0.75$) to the original.

3.7. Comparison among additive, uniform additive, and impulse

Between additive and uniform additive, the effects of the two approaches to the waveform are comparable, returning equivalent r_s values as shown in Table 6. As shown in the previous section (Tables 5, 6, and 7), the distributions of the MDI for 1000 realizations of the waveforms for the additive and uniform additive were almost always equal. In the case of the impulse model, the appearance of intermittent spikes did not tremendously change the shape of the curve in general. In fact, based on the results of the leading subsets for the 1998 and 2005 datasets (Tables 6A1 and 6B1), the IM showed high r_s values (as high as $r_s = 0.98$), comparable to the results of the other two models, AD and UA.

3.8. Smoothing waveforms

Table 7 showed the comparison of the performances of the three smoothing window types measured in terms of RMSE. The forward-window approach gave a large RMSE of MDI, 83.93 and 53.44 for 1998 and 2005, respectively, based on a moving window of 3% of the total wave counts. Centered window smoothing appears to be the best way to smooth a waveform.

Table 7. Computed RMSE from 16 pairs of samples using the full extent of the smoothed waveforms. Curves were smoothed up to a moving window of 3% of total wave counts (size 15).

<i>Full Extent (1998)</i>				<i>Full Extent (2005)</i>			
Type	Size 5	Size 10	Size 15	Type	Size 5	Size 10	Size 15
Forward	23.58	53.60	83.93	Forward	16.18	35.18	53.44
Backward	22.02	48.58	74.67	Backward	14.77	32.92	50.84
Centered	12.36	9.41	18.55	Centered	7.58	4.47	8.79

Figure 9 showed the relationships of the MDI computed using smoothed waveforms against the quasi-heights. Trends from Figure 9 and Figure 7 are comparable with each other, having observed a stronger 1998 linear relationship between MDI and the canopy quasi-height than the 2005 dataset—averaged $r^2=0.62$ (1998) vs averaged $r^2=0.38$ (2005) for noise-affected waveform, while averaged $r^2=0.30$ (1998) vs averaged $r^2=0.22$ (2005) for smoothed waveform.

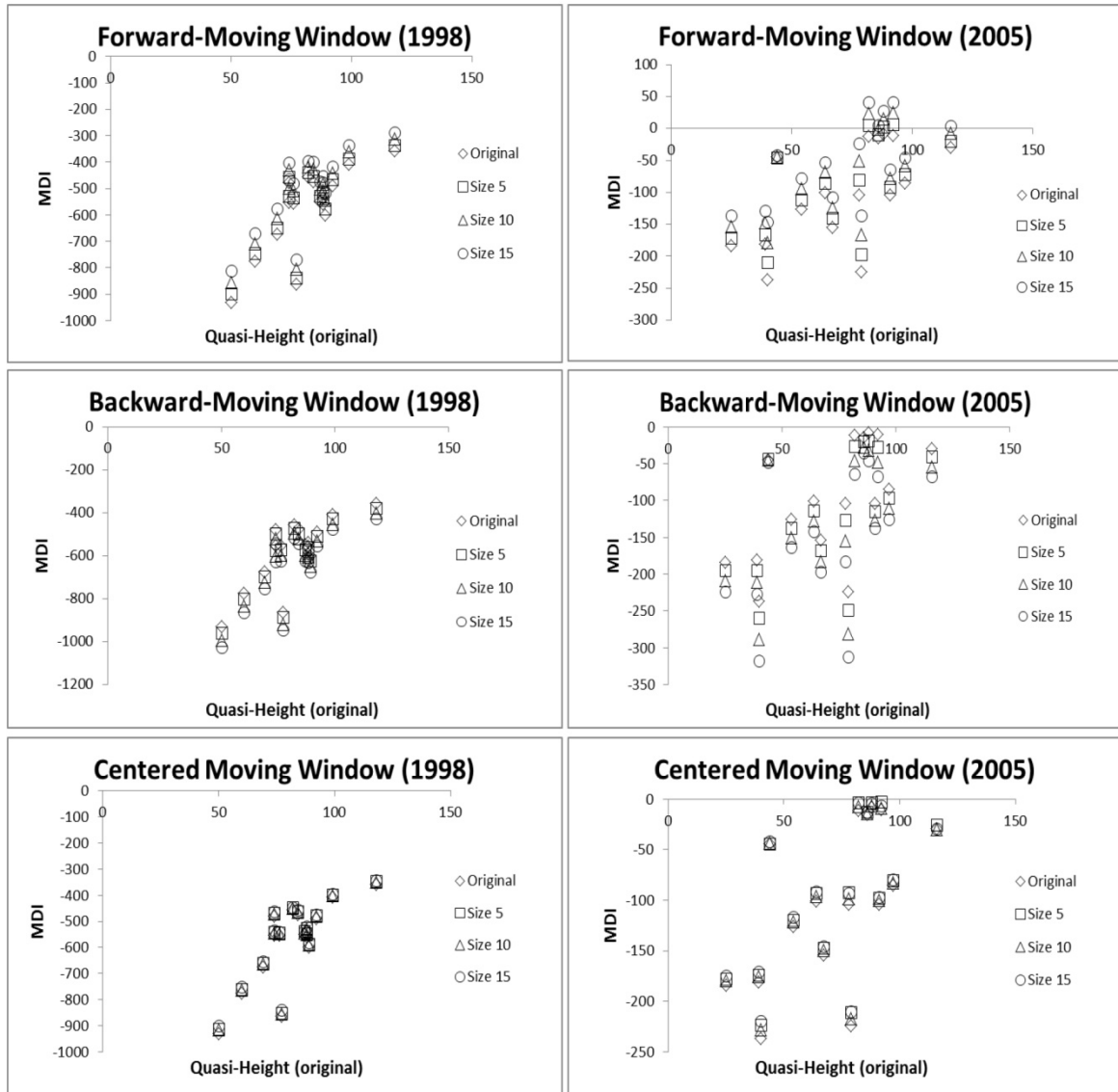


Figure 9. MDI (smooth full-extent waveform) as a function of quasi-height for the 1998 and 2005 La Selva LVIS datasets. First column: 1998 dataset at various window sizes and; Second column: 2005 dataset at various window sizes. The relationships shown between MDI and quasi-height in these figures can be compared with those found in Figure 7 from noise-affected waveforms.

3.9. Area under the curve (AUC)

Figures 10 and 11 illustrate the MDI as having a far better and stronger relationship with the quasi-height than the AUC (*e.g.* $r^2 = 0.62$ vs $r^2 = 0.35$ for the 1998 AD), especially if the full-extent of the waveform is analyzed. The same pattern holds for the UA and the IM models. Table 8 lists the coefficients of determination (r^2) for quasi-height vs the AUC and MDI. MDI shows better fit than AUC in most tests except in the 2005 trailing subset.

Table 8. Coefficients of determination (r^2) for quasi-height vs the AUC and MDI from 16 pairs of samples for 1998 and 2005 datasets using the full extent and trailing subset of the waveform.

<i>Full Extent (1998)</i>			<i>Full Extent (2005)</i>		
Type	AUC	MDI	Type	AUC	MDI
AD	0.35	0.62	AD	0.002	0.38
UA	0.35	0.62	UA	0.002	0.38
IM	0.34	0.62	IM	0.001	0.38
<i>Trailing Subset (1998)</i>			<i>Trailing Subset (2005)</i>		
Type	AUC	MDI	MDI	AUC	MDI
AD	0.27	0.30	AD	0.33	0.22
UA	0.27	0.30	UA	0.33	0.22
IM	0.26	0.29	IM	0.32	0.20

We computed the RMSE (Table 9) of the AUC for each noise approach using the full waveform extent. Small RMSE values were related to 5% noise on the waveform while the high RMSE values were from the 20% noise level, regardless of the type of noise approach. The highest RMSE values were found using the IM approach.

Table 9. Computed RMSE of AUC from 16 pairs of samples for both 1998 and 2005 datasets using the full extent of the noise-affected waveform.

<i>Full Extent (1998)</i>				<i>Full Extent (2005)</i>			
Type	Size 5	Size 10	Size 15	Type	Size 5	Size 10	Size 15
AD	5.38	10.75	16.13	AD	26.55	53.10	79.65
UA	11.58	23.17	34.77	UA	32.02	64.03	96.05
IM	14.32	28.64	42.97	IM	33.95	65.90	98.85

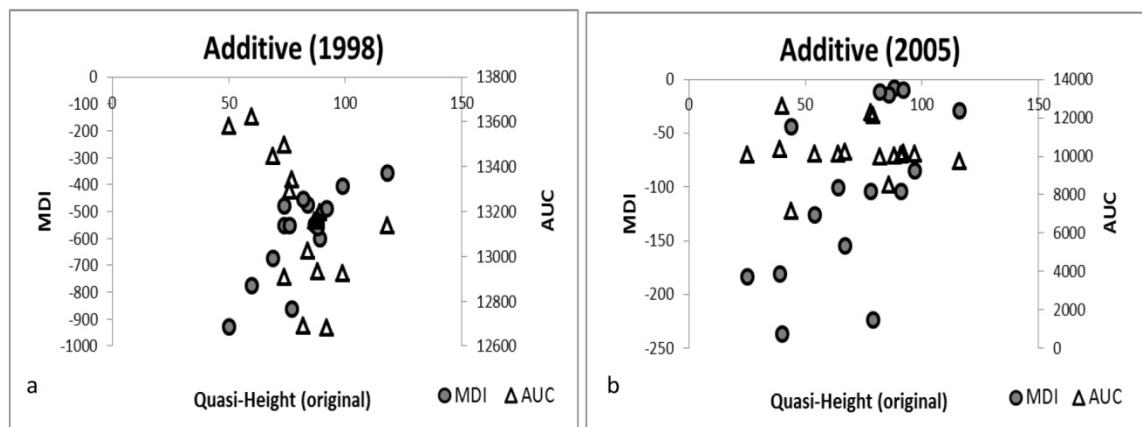


Figure 10. AUC and MDI (noise-affected full-extent waveform, using additive) as a function of quasi-height for the (a) 1998 and (b) 2005 La Selva LVIS datasets. Linear trend is observed between MDI and canopy quasi-height: $r^2 = 0.62$ (MDI) vs $r^2 = 0.35$ (AUC) for the 1998 AD; $r^2 = 0.38$ (MDI) vs $r^2 = 0.002$ (AUC) for the 2005 AD.

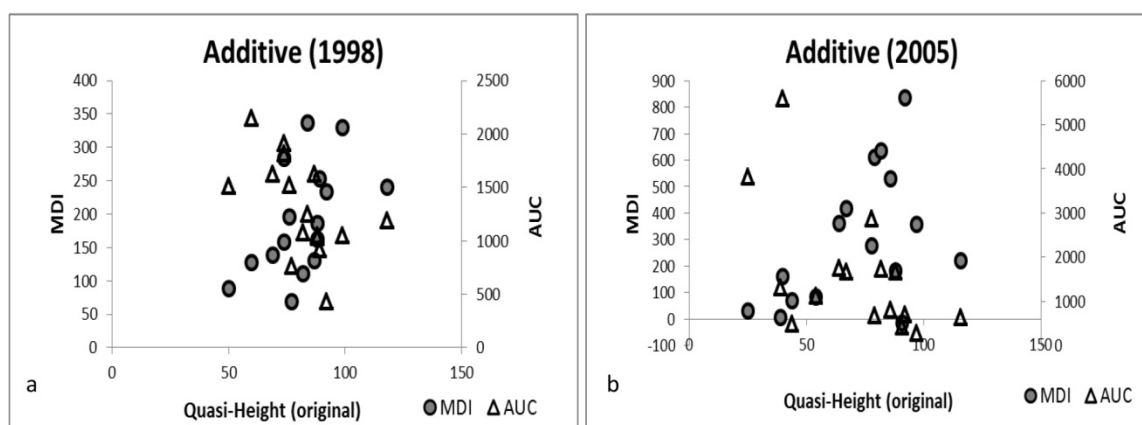


Figure 11. AUC and MDI (noise-affected smaller subset waveform, using additive) as a function of quasi-height for the (a) 1998 and (b) 2005 La Selva LVIS datasets. Linear trend is still observed between MDI and canopy quasi-height using a short range: $r^2 = 0.30$ (MDI) vs $r^2 = 0.27$ (AUC) for the 1998 AD; $r^2 = 0.22$ (MDI) vs $r^2 = 0.33$ (AUC) for the 2005 AD.

3.10. Temporal movement of MDI

For the final analysis, we looked at the movement of the MDI from the pairing of the 1998 to 2005 samples (Figure 12). Two groupings of paired samples were observed. The pairs from the first group showed decreasing negative MDI as quasi-heights increased, while pairs from the second group showed decreasing negative MDI as quasi-height decreased (Figure 12a).

The same groups of samples were observed in Figure 12b for the trailing subset waveform. The

first group of pair samples showed increasing MDI as quasi-heights increased, while pairs from the second group showed decreasing MDI as quasi-height decreased.

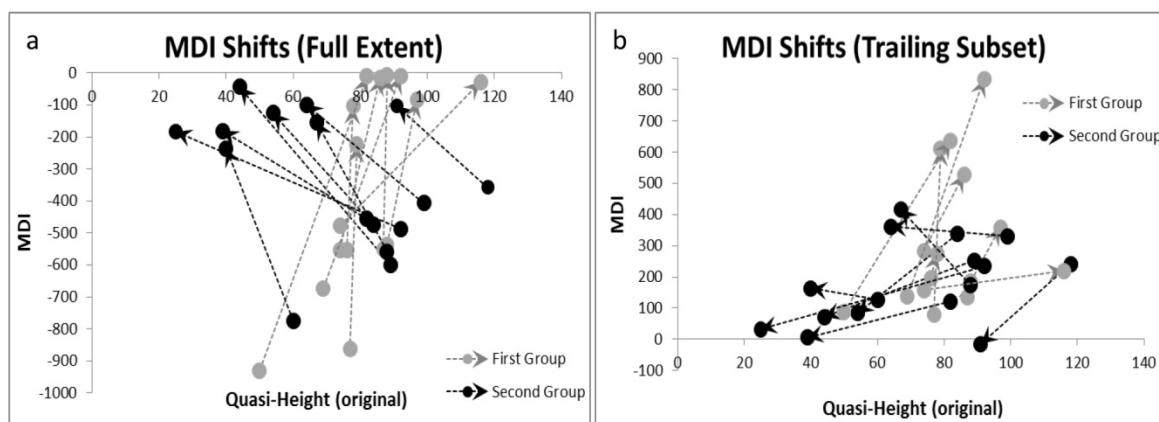


Figure 12. 1998 and 2005 values of MDI vs quasi-heights plotted to show the shifting of the MDI over the time period. The arrow shows the direction of the sample pairing from earlier year to a later year. Two groupings of paired samples were observed for both the (a) full-extent and (b) trailing subset of the waveform.

4. Discussion

The new MD approach is hardly affected by noise, as shown in our test results when waveforms corrupted with AD, UD, and IM noise have been fitted with our method. MDI appears to be sensitive to impulse noise, but robust to the other noise forms. Impulse noise tends to exaggerate the waveform by introducing spikes when a longer wave count gap exists between peaks (see the 1998 and 2005 shapes in Figure 3 for impulse). Results from these tests are important since we want the new metric to be robust to random noise in the waveform.

Even with the MDI changing in absolute values at several noise levels (up to 20%), the relationships with canopy quasi-height are minimally affected. Tests on the noise-affected waveforms resulted with AD, UD, and IM having the same averaged r^2 of 0.62 and 0.38, for the 1998 and 2005 data, respectively. Figures 7 and 8 show that good fitting results can be achieved even with waveforms corrupted with noise, regardless of whether the analysis involves the full extent or the subset of the waveform. We attribute this robustness of the MDI algorithm to its ability to define the shape of the wave from two points of perspective, rather than one. Accordingly, the MDI can minimize the effects of noise present between the left (MD_{LP}) and right (MD_{RP}) pivots. There were no significant differences of MDI means observed in all trials when we tested using ANOVA at 0.05 significance level ($p > 0.05$ for both 1998 and 2005 datasets). Although, MDI values differ in each trial, the values are so close that the symbols in Figures 7 and 8 look like they overlap. However, these minimal differences did not affect the relationship of MDI to the canopy quasi-height (cf. Table 4). Will the MDI work in the presence of major error spikes on the waveform?

As illustrated by the impulse model results, the MDI can attenuate the effects of spikes, but it is more susceptible to spikes than other noise forms.

The leading subset has very little to no relationship to the quasi-height. This could affirm the need for a 10% elimination of the upper canopy in height analysis [59]. In contrast, the trailing subset shows a positive yet low relationship against the quasi-height (Figure 8) with highest averaged r^2 of 0.30 for 1998 and 0.22 for 2005. From 5% to even a high noise of 20%, the pattern of MDI vs quasi-height remained to be similar for the AD, UD, and IM approaches. In Table 5, we computed the RMSE of MDI for each noise model using the full waveform and the trailing subset and found out that the small RMSE values are related to 5% error on the waveform.

In Table 6, the leading subsets for both years come with higher correlation coefficients, r_s , compared to the trailing subset. It implies that the waveforms generated with noise levels in the leading subsets exhibit shape similarity to the original waveform. This observation further suggests that the noise within the leading region may not have strong effects to the relationship of MDI with the canopy quasi-height. The above average values of r_s observed in the trailing subsets are caused by a much longer pivot range as well as the steep rise of the curves in between pivots.

While the spikes introduced by IM approach have been detected by the MDI approach as shown by the MDI distribution of the 1000 iterations of the impulse noise in the previous section (Figures 4, 5, and 6), the MDI is resistant to a single spike or two. The MDI be significantly affected only when a major spike is of considerable duration and thus be mistaken for a valid signal return or peak.

It should be noted that, in Table 6, a high correlation coefficient, say $r_s = 0.90$, does not necessarily signify that there will be the same MDI values for the original and the noise-affected waveforms. For example, a noise-affected waveform with a major spike that is near a pivot can result to a high correlation coefficient compared to the original waveform; however, the presence of the spike will result to two different values of MDI. Moreover, a low correlation coefficient does not signify a curve shape that is less effective in estimating canopy characteristics. It is shown that with the MDI, the noisy curves are still able to show a pattern of MDI-to-height correlations (Figures 7 and 8). It is important to note that waveform shapes are maintained despite the introduction of noise.

We also have shown the new approach to have an advantage over the area under curve especially for the full waveform. Our results have shown that the MDI has a stronger positive relationship with the canopy quasi-height than does the AUC ($r^2 = 0.62$ vs. $r^2 = 0.35$ for the 1998 AD). This improvement is also evident for the 1998 trailing subset. As shown in Table 8, however, AUC exhibited a better fit with the quasi-height for the 2005 trailing subset. It even came out with a comparable r^2 to the 1998 subset, AD approach ($r^2 = 0.27$ for AUC vs. $r^2 = 0.30$ for MDI). This better showing of the AUC against the MDI maybe caused by the shorter pivot range of the subset. The shorter the pivot, the fewer the points present to define the curve for MDI. Nonetheless, with a longer pivot such as the full extent, the AUC fails badly ($r^2 = 0.002$ for the 2005 AD), especially with the presence of noise that can amplify the areas under the curve.

When smoothing the waveforms, small smoothing moving windows of 1% to 3% of the total wave count is recommended. Using moving windows of 5 to 10 counts, even up to a maximum of

15 counts, can effectively smooth waveform irregularities without risking of leveling small important peaks, especially those found near the ground return. Also, centered window smoothing technique may be the best way to smooth a waveform with the least RMSE for 1998 and 2005 datasets, as shown in Table 7. Forward and backward smoothing may fail to work when the waveform consists of multiple peaks or small peaks in between the major peaks that may vanish due to the smoothing. This effect explains the large deviations of the MDI (forward and backward smoothed waveforms) from the reference MDI found in Figure 9 for 2005.

The groupings of the MDI that are seen in Figure 12 are due to the fact that the waveform is complex and can have different shapes. Three canonical waveform shapes include (a) a maximum early peak, when the first canopy peak is maximum, or (b) maximum late peak, when the ground peak is maximum, or (c) roughly equal peaks for both canopy and ground. Our results reveal that the MDI can capture aspects of temporal dynamics of canopy quasi-heights and group them based on the curve shapes. How this information can be used to classify the spatial and vertical heterogeneity of forest structure is the target of a future paper.

5. Conclusion

The Moment Distance framework we have developed is novel and its application to LiDAR data provides a new, computationally easy approach to characterizing waveform shape. The new approach decomposes LiDAR waveform returns into left moment (with pivot from early time wave count) and right moment (with pivot from late time wave count) components, and then computes the MDI metric. The decomposition allows us to look at how a selected pivot defines the strength of each point on the waveform from a single point of perspective. The summation of each *strength* defines the structural behavior of the asymmetrical curve from, again, a single standpoint. Having two pivots solidifies the concept of the MD as an approach for shape-characterization by defining the structural behavior of the waveform not only from a single standpoint, but two.

Moreover, the MDI is minimally affected by noise. It has an advantage over AUC with its stronger relationship against the canopy quasi-height. MD calculation is straightforward and thus MDI analysis is easy to replicate.

One important contribution of our metric to waveform analysis is that the customary use of the Gaussian modeling to fit multiple peaks and improve peak detection may be avoided. While current waveform optimization fitting schemes rely strongly on initial parameters, which they usually fail in identifying weak returns, the new approach uses the raw waveform itself to capture shape, without requiring parameter estimation.

In conclusion, MDI is a robust metric. The results shown in this paper allows us to put forth an argument on how the new metric will behave against the existing relative height (RH) metric. Also, the results warrant future tests of MDI using different types of LiDAR waveform shapes—maximum peak observed at an earlier time, maximum peak observed at a later time, and observed peaks are equal (roughly) in return magnitude—against the important key profile landmarks of the waveform.

We plan in a future paper to show how waveform shape and movement of the peaks, dips, and other landmarks respond to changes in canopy quasi-height. It is crucial to explore the behavior of the MDI in relation to the temporal changes of waveform shape and landmarks. In this way, we can illustrate how the new metric can capture temporal change in canopy height and, thus, provide a means to monitor forest growth and development for habitat assessment and carbon monitoring applications.

Acknowledgement

The authors would like to thank the anonymous reviewers for helping to improve the clarity 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. Garner R (2014) New NASA Probe Will Study Earth's Forests in 3-D. *NASA*, 08-Sep-2014. Available from: <http://www.nasa.gov/content/goddard/new-nasa-probe-will-study-earth-s-forests-in-3-d>.
2. Ni-Meister W, Jupp DLB, Dubayah R (2001) Modeling lidar waveforms in heterogeneous and discrete canopies. *IEEE Trans Geosci Remote Sens* 39: 1943-1958.
3. Koetz B (2007) Fusion of imaging spectrometer and LIDAR data over combined radiative transfer models for forest canopy characterization. *Remote Sens Environ* 106: 449-459.
4. Wu J (2009) Lidar waveform-based woody and foliar biomass estimation in savanna environments. *Proc Silvilaser*, 1-10.
5. 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.
6. Dubayah RO, Drake JB (2000) Lidar remote sensing for forestry. *J For* 98: 44-46.
7. Hurtt GC (2004) Beyond potential vegetation: combining Lidar data and a height-structured model for carbon studies. *Ecol Appl* 14: 873-883.
8. Lefsky MA, Keller M, Panga Y, et al. (2007) Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *J Appl Remote Sens* 1: 013537.
9. 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.
10. Chen Q (2010) Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry. *Remote Sens Environ* 114: 1610-1627.

11. Lefsky M, McHale MR (2008) Volume estimates of trees with complex architecture from terrestrial laser scanning. *J Appl Remote Sens* 2: 1-19.
12. Duncanson LI, Niemann KO, Wulder MA (2010) Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sens Environ* 114: 138-154.
13. Yan WY, Shaker A, El-Ashmawy N (2015) Urban land cover classification using airborne LiDAR data: A review. *Remote Sens Environ* 158: 295-310.
14. Reitberger J, Krzystek P, Stilla U (2008) Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *Int J Remote Sens* 29: 1407-1431.
15. Vaughn NR, Moskal LM, Turnblom EC (2012) Tree species detection accuracies using discrete point lidar and airborne waveform lidar. *Remote Sens* 4: 377-403.
16. Alonzo M, Bookhagen B, Roberts DA (2014) Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sens Environ* 148: 70-83.
17. 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.
18. Drake JB, Dubayah RO, Knox RG, et al. (2002) Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sens Environ* 81: 378-392.
19. Danson FM (2014) Developing a dual-wavelength full-waveform terrestrial laser scanner to characterize forest canopy structure. *Agric For Meteorol* 198–199: 7-14.
20. Zhao K, Popescu S (2009) Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA. *Remote Sens Environ* 113: 1628-1645.
21. Zhao F (2011) Measuring effective leaf area index, foliage profile, and stand height in New England forest stands using a full-waveform ground-based lidar. *Remote Sens Environ* 115: 2954-2964.
22. Tang H (2012) Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sens Environ* 124: 242-250.
23. Adams T, Beets P, Parrish C (2011) Another dimension from LiDAR – Obtaining foliage density from full waveform data. *Int Conf LiDAR Appl Assess For Ecosyst*, Oct. 2011.
24. Drake JB, Dubayah RO, Clark DB, et al. (2002) Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sens Environ* 79: 305-319.
25. Drake JB, Knox RG, Dubayah RO, et al. (2003) Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Glob Ecol Biogeogr* 12: 147-159.
26. Ni-Meister W, Lee S, Strahler AH, et al. (2010) Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *J Geophys Res Biogeosciences* 115: p. G00E11.
27. Powell SL, Cohen WB, Healey SP, et al. (2010) Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sens Environ* 114: 1053-1068.

28. Zolkos SG, Goetz SJ, Dubayah RO (2013) A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sens Environ* 128: 289-298.
29. Huang W, Sun G, Dubayah RO, et al (2013) Mapping biomass change after forest disturbance: Applying LiDAR footprint-derived models at key map scales. *Remote Sens Environ* 134: 319-332.
30. Whitehurst AS, Swatantran A, Blair JB, et al. (2013) Characterization of canopy layering in forested ecosystems using full waveform Lidar. *Remote Sens* 5: 2014-2036.
31. Zhao F, Yang X, Strahler AH, et al. (2013) A comparison of foliage profiles in the Sierra National Forest obtained with a full-waveform under-canopy EVI lidar system with the foliage profiles obtained with an airborne full-waveform LVIS lidar system. *Remote Sens Environ* 136: 330-341.
32. 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.
33. Mallet C, Bretar F (2009) Full-waveform topographic lidar: State-of-the-art. *ISPRS J Photogramm Remote Sens* 64: 1-16.
34. 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.
35. Lefsky MA, Harding DJ, Keller M, et al. (2005) Estimates of forest canopy height and aboveground biomass using ICESat. *Geophys Res Lett* 32: L22S02.
36. Hancock S, Armston J, Li Z, et al. (2015) Waveform lidar over vegetation: An evaluation of inversion methods for estimating return energy. *Remote Sens Environ* 164: 208-224.
37. Martinuzzi S, Vierling LA, Gould WA, et al. (2009) Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sens Environ* 113: 2533-2546.
38. Unser M (1999) Splines: a perfect fit for signal and image processing. *IEEE Signal Process Mag* 16: 22-38.
39. El-Baz A, Gimel'farb G (2007) EM-based approximation of empirical distributions with linear combinations of discrete Gaussians. *IEEE Int Conf Image Process* 4: IV-373-IV-376.
40. Hofton MA, Minster JB, Blair JB (2000) Decomposition of laser altimeter waveforms. *IEEE Trans Geosci Remote Sens* 38: 1989-1996.
41. Anderson K, Hancock S, Disney M, et al. (2016) Is waveform worth it? A comparison of LiDAR approaches for vegetation and landscape characterization. *Remote Sens Ecol Conserv* 2: 5-15.
42. Hopkinson C, Chasmer L (2009) Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sens Environ* 113: 275-288.
43. 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.
44. García A, 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.
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. 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.
47. Aguilar MA, Nemmaoui A, Novelli A, Aguilar FJ, García Lorca A (2016) Object-based greenhouse mapping using very high resolution satellite data and Landsat 8 time series. *Remote Sens* 8: 513.
48. Salas EAL, Boykin KG, Valdez R (2016) Multispectral and texture feature application in image-object analysis of summer vegetation in eastern Tajikistan Pamirs. *Remote Sens* 8: 78.
49. Salas EAL, Henebry GM (2016) Canopy height estimation by characterizing waveform LiDAR geometry based on shape-distance metric. *Geosci* 2: 366-390.
50. Cawse-Nicholson K, van Aardt J, Romanczyk P, et al. (2014) Below-ground responses in waveform lidar due to multiple scattering. Available from: ftp://cis.rit.edu/people/vanaardt/Temp/PostDoc/Cawse_Echo_letter_2014.pdf
51. Fieber FD, 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.
52. Radke RJ, Andra S, Al-Kofahi O, et al. (2005) Image change detection algorithms: a systematic survey. *IEEE Trans Image Process* 14: 294-307.
53. Henderson D, Hamernik RP (1986) Impulse noise: critical review. *J Acoust Soc Am* 80: 569-584.
54. Sripad A., Snyder D (1977) A necessary and sufficient condition for quantization errors to be uniform and white. *IEEE Trans Acoust Speech Signal Process* 25: 442-448.
55. Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36: 1627-1639.
56. Madden HH (1978) Comments on the Savitzky-Golay convolution method for least-squares-fit smoothing and differentiation of digital data. *Anal Chem* 50: 1383-1386.
57. Kotchenova SY, Song X, Shabanov NV, et al. (2004) Lidar remote sensing for modeling gross primary production of deciduous forests. *Remote Sens Environ* 92: 158-172.
58. Li Q (2008) Decomposition of airborne laser scanning waveform data based on EM algorithm. *Int Arch Photogramm Remote Sens Spat Inf Sci* 37: 211-218.
59. Popescu SC, Wynne RH (2004) Seeing the trees in the forest. *Photogramm Eng Remote Sens* 70: 589-604.



AIMS Press

© 2017 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>)