AdaptLidarTools: A Full-Waveform Lidar Processing Suite

Ravi Shankar
Department of Computer Science
Boise State University
Boise, ID, USA
ravishankar@u.boisestate.edu

Nayani Ilangakoon
Department of Geosciences
Boise State University
Boise, ID, USA
nayaniilangakoon@u.boisestate.edu

Aaron Orenstein
Treasure Valley Math and Science Center
Capital High School
Boise, ID, USA
aaronorenstein@u.boisestate.edu

Floriana Ciaglia
Department of Computer Science
Boise State University
Boise, ID, USA
flociaglia@u.boisestate.edu

Nancy F. Glenn
Department of Geosciences
Boise State University
Boise, ID, USA
nancyglenn@boisestate.edu

Catherine Olschanowsky
Department of Computer Science
Boise State University
Boise, ID, USA
catherineolschan@boisestate.edu

***Abstract*—AdaptLidarTools is a software package that processes full-waveform lidar data. Full-waveform lidar is an active remote sensing technique in which a laser beam is emitted towards a target and the backscattered energy is recorded as a near continuous waveform. A collection of waveforms from airborne lidar can capture landscape characteristics in three dimensions. Specific to vegetation, the extracted echoes and echo properties from the waveforms can provide scientists structural (height, volume, layers of canopy, among others) and functional (leaf area index, diversity) characteristics. The discrete waveforms can be transformed into georeferenced 2D rasters (images), allowing scientists to correlate field-based observations for validation of the waveform observations and fusing the data with other geospatial information. AdaptLidarTools provides an extensible, open-source framework that processes the waveforms and produces multiple data outputs that can be used in vegetation and terrain analysis.**

AdaptLidarTools is designed to explore new methods to fit full-waveform lidar signals and to maximize the information in the waveforms for vegetation applications. The toolkit explores first differencing, complementary to Gaussian fitting, for faster processing of full-waveform lidar signals and for handling increasingly large volumes of full-waveform lidar datasets. AdaptLidarTools takes approximately 30 min to derive a raster of a given echo property from a raw waveform file of 1 GB size. The toolkit generates first order echo properties such as position, amplitude, pulse width, and other properties such as rise time, fall time and backscattered cross section. It also generates other properties that current proprietary and open-source tools do not. The derived echo properties are delivered as georeferenced raster files of a given spatial resolution that can be viewed and processed by most remote sensing data processing software.

***Keywords*-lidar; full-waveform; PulseWaves; waveform characterization;**

I. Introduction

Scientists observe the earth's surface using several remote sensing methods including lidar (Light Detection and Ranging). Lidar systems emit a laser pulse and record the reflection. Full-waveform lidar data provide significant information about the structural characteristics of the landscape [1]. Existing packages process the data and provide products that describe a fixed set of landscape characteristics used to make observations. However, the intermediate data processing products are inaccessible. This limits the use of the data and hinders progress expanding the set of characteristics captured. This paper describes AdaptLidarTools[1], an open-source package that processes full-waveform lidar data. It is extensible and provides a platform to explore design trade-offs and new uses of intermediate data products.

Remote sensing is the process of detecting and monitoring physical and chemical characteristics of a target by means of reflected or emitted electromagnetic radiation at a distance from the target. Conventional methods to measure characteristics of vegetation in a landscape are time and resource intensive and sometimes are difficult to physically measure. Remote sensing provides complimentary information to these field-based methods. Remote sensing sensors can be mounted on drones, planes, or satellites and the information they capture is then used to estimate the conditions and the characteristics of the landscape. Lidar is an active remote sensing method that captures 3D landscape characteristics by measuring distance to targets with a pulsed laser and sensor that records the reflections as a function of time of flight [1].

Lidar remote sensing has a wide spectrum of applications in geography, geology, geomatics, engineering, geomorphology, archaeology, seismology, forestry, and atmospheric science and climate studies due to its ability to create high resolution, and high accuracy 3D representations of the landscape [1]. For example, lidar derived heights are used to generate Digital Terrain Models (DTMs), Digital Surface Models (DSMs), Canopy Height Models (CHMs) as well as specific information about the landscape and how it changes

[1]https://github.com/BoiseState-AdaptLab/adapt-lidar-tools

978-1-7281-2451-3/19/$31.00 ©2019 IEEE
DOI 10.1109/eScience.2019.00048

(e.g. canopy architecture, biomass [2], habitat, and post fire disturbance) [3]–[6]. Lidar intensity (strength of reflectance of the laser), with or without height, is used for vegetation classification [7]–[9]. Due to the penetrability through forest canopy, lidar is also used to detect archaeological sites within dense forests where access is difficult [10]. Further, lidar is used to detect landslides [11], erosion, earth deformations, atmospheric carbon dioxide, and aerodynamic roughness [5], [12]–[14].

Several discrete return lidar processing tools exist [15], [16], but there is a lack of tools for full-waveform processing. Many existing tools to process full-waveform lidar data are proprietary, non-extensible, and/or time intensive. Commercial software packages georegister full-waveform lidar signals and decompose them into discrete return records and there is no way to extract additional echo properties [17], [18]. Some open-source tools allow the user to ingest and process full-waveform data, however, the end products are limited to commonly used pulse attributes (amplitude, pulse width) and secondary attributes such as DTMs, DSMs, CHMs, and structural measures such as gap fraction, canopy openness, leaf area index and height percentiles [19], [20]. Further, many commercial and open-source tools that process full-waveform lidar struggle to efficiently process the inherently large volumes of full-waveform lidar.

The contributions of this work include the following:

- AdaptLidarTools: an open-source, extensible, full-waveform processing package
- A description of an extensible set of intermediate data values as raster files
- A verification and comparison framework for intermediate data products

This software package is an undergraduate led project; it provides valuable research experience to computer science undergraduate students and exposure to computationally intensive domain sciences. The lead scientist is a graduate student in the Geosciences department who is developing novel methods for full-waveform analysis of vegetation studies. She communicates requirements based on code she developed [21] to a team of computer science undergraduates who manage the implementation. The experience allows undergraduates to learn agile development approaches, work to understand the nomenclature of a different field, and build production quality software as a team. In addition to contributing full-waveform software capabilities, it is our hope that this project will serve as a model for other teams.

II. Background

This section describes and compares different kinds of lidar systems, and introduces Gaussian fitting as the current waveform fitting technique.

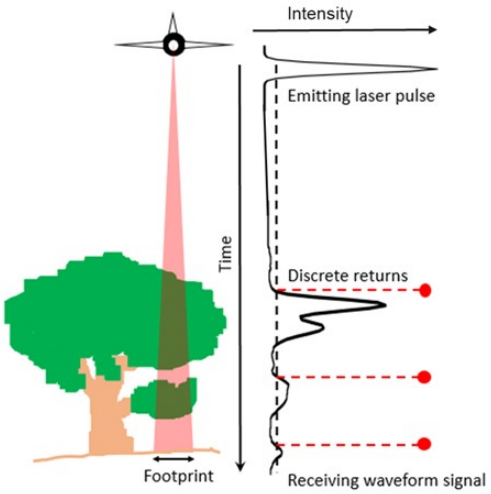


Figure 1: Principle of lidar systems. Full-waveform lidar signal (in black) and discrete return lidar (in red dots).

A. Lidar

Lidar systems use a laser pulse that typically emits either green or near infrared radiation [22]. There are two common types of lidar systems based on the method the reflections are recorded: discrete return and full-waveform. Full-waveform lidar records the whole return from a laser pulse as one continuous waveform [23] and discrete return lidar or point based lidar records the maximum reflections of echoes above a given reflection threshold as points from the waveform [6](see Figure 1). A discrete return lidar system typically records 1 to 5 returns from each laser pulse depending on the landscape (bare ground, dense forest, savanna, etc.) and recording system.

Lidar systems can also be differentiated as small footprint (0.2-1.0m) or large footprint (10–70m) by the target surface area covered by the laser beam [24]. Most discrete return lidar systems are small footprint and full-waveform systems can be either small or large footprint. Small footprint lidar systems mostly operate on ground, drone, or airborne platforms while most large footprint lidar systems are airborne (e.g. LVIS [25]) or spaceborne (e.g. GEDI [26]).

Recorded waveforms from lidar are mined for information about the target. Each waveform captures a series of echos. An echo appears as a peak in the waveform and the properties of the peak (or echo) are used by scientists to derive the size, orientation, and spatial arrangement, as well as radiometric characteristics of the target or scene [27]–[29]. Figure 1 shows a waveform with four echos (the first peak in the wave represents the outgoing pulse). The following properties can be extracted from each echo and are used to explain the below characteristics of the illuminated target [21].

Pulse width: The pulse width of an echo represents the surface roughness of a vegetation canopy or the slope of the ground [30].
Amplitude: The amplitude of an echo is the optical response of the target to the emitted lidar wavelength [30].
Backscattered cross section: The backscattered cross section derived from each echo shows the target area which interacted with the laser beam [29].
Rise time: The rise time represents the target structure in the vertical direction [31].
Number of echoes: The number of echoes in a waveform is an indication of how many targets were interacted with along the laser path and their position in 3D space [32]. For example, in a forest environment, the number of echoes represents canopy layers and their height distribution.

B. Gaussian Fitting

AdaptLidarTools was designed to allow a variety of waveform fitting approaches. The current version uses Gaussian fitting, because the amplitude distribution of the emitted lidar pulse is a Gaussian in most laser scanning systems (Optech, Leica, Riegl) [29], and produces a near Gaussian backscattered waveform. If multiple echoes were detected due to multiple distinct targets, the recorded backscattered waveform represents a sum of Gaussians. The following equation is used to express the sum of Gaussians $f(t)$. Where,

- P represents the number of peaks in each waveform,
- a is the amplitude of the the peak,
- t is the time,
- b is the time location at the peak amplitude, and
- c is the full width at half maximum (FWHM).

$$f(t) = \sum_{i=0}^{P} a_i e^{-[\frac{(t-b_i)^2}{2c_i^2}]} \quad (1)$$

Figure 2 shows a typical backscattered waveform decomposed into individual Gaussians. Gaussian fitting attempts to choose the parameters to the sum of Gaussians equation to create a waveform that matches the observed data. Each echo is then described using the corresponding single Gaussian.

III. Software Requirements

This software ingests binary lidar waveform data, extracts properties of each waveform, and uses the extracted properties to produce geolocated raster files. The geolocation is based on GPS data onboard the lidar sensing system. There are seven waveform properties currently defined, five of which are implemented (see Table II). The properties are summarized using one of several techniques. Users choose from over 90 raster products by configuring the desired property and summary technique. This section provides an overview of the available products.

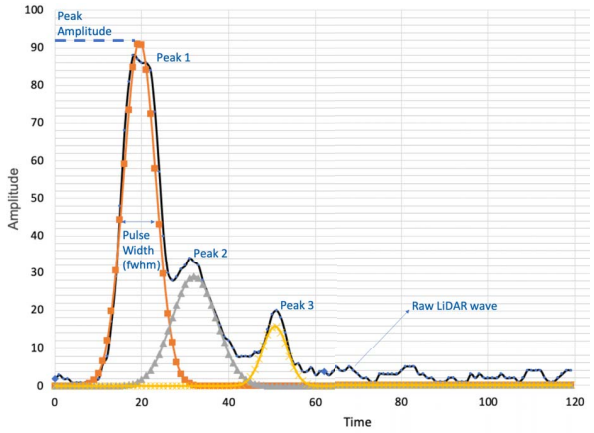


Figure 2: Gaussian decomposition of a waveform into its individual Gaussians

Geolocation is done per peak; each peak in a single waveform is assigned a location. The location of a peak is determined by the GPS coordinates of the inflection point in space as x, y, and z. The inflection point is the first recorded value at the beginning of the peak that exceeds the noise level (e.g. 6.25 in this study). The x and y describe the horizontal position of the peak in space while the z represents the altitude (elevation) with respect to mean sea level. The x and y coordinates of the inflection points are used to bin all other waveform lidar feature values into raster pixels. Peaks from a single waveform are typically binned into different locations depending on the relative distance of each peak in space. However, during rasterization many peaks from the same or different waveforms may be binned to a single pixel location in the raster file according to the selected spatial resolution. All the values binned into a single pixel are then combined to produce a single value for the pixel. Consider the case of an outgoing wave hitting a tree as in Figure 1. Several distinct pulses may reflect from different heights of the tree.

Raster files hold data values that have been binned by x,y location. For example, the z coordinate represents the elevation of the object hit by the outgoing waveform and can be saved to a raster file. The z values are binned into pixel locations and the maximum elevation observed at each pixel location can be the value output for a raster file. Other features that describe each peak such as the maximum amplitude of the wave (Peak Amplitude), the width of the wave at half of the maximum amplitude (Full Width at Half Maximum or FWHM), the time between the inflection point and the maximum amplitude of the peak (Rise Time), and the backscatter coefficient may also be saved in the raster file (see Table II).

For each echo in each waveform, the features/variables described in Table II are calculated.

Triggering Position (x,y,z): The x, y, and z of the triggering

Required parameters	Description	Symbol
Backscatter coefficient	The horizontal scattered cross-section of the target with respect to the deployed system wavelength and range that is normalized to the incident angle.	γ_i
Calibration constant	A constant calculated using waveforms of standard reflectance surfaces like black tarps/roads.	C_{cal} (9.660535682891269e-08 for this study)
Range	The distance from anchor to each peak triggering location in each waveform	R_i
Peak amplitude	Maximum point of each fitted peaks	$\hat{P}$
Pulse width	The full width at half maximum	W_p
Standard deviation of pulse width	The standard deviation of pulse width using peak with and the system pulse width	$s_{(p,i)}$
Amplitude of system pulse	Peak amplitude of the emitted waveform	$\hat{S}$
System pulse width	Pulse width of the emitted waveform	W_s
Atmospheric transmission factor	This tells the condition of the atmosphere during the data collection	η_{atm}

Table I: Parameters used in the calculation of the backscatter coefficient

Variable	Description
Peak Amplitude	The amplitude of given peak after Gussian fitting
Pulse width	The waveform width after applying Gaussian fitting
Rise time	The time from the first triggering location to the peak of the waveform
Backscatter coefficient	The horizontal scattered cross-section of the target with respect to the deployed system wavelength and range that is normalized to the incident angle.
Elevation	The first inflection point of each waveform above the noise level
Heights at percent energy	Heights at percent energies in each waveform. Absolute heights at cumulative energy percentiles (10th, 25th, 50th, 75th, 90th).
Energy at percent heights	Waveform energy at heights from first echo triggering location. Cumulative energy at height percentiles (10th, 25th, 50th, 75th, 90th)

Table II: Properties to extract from the full-waveform and used to characterize vegetation and the terrain.

locations (initial point above noise amplitude value (6.25) at the left limb of each peak). The values are in terms of GPS locations. The z-value provides elevation at that point.

Peak ID: The position of the peak (first, second, third, ... last) in the waveform. The position in the waveform indicates how many targets interacted with that particular waveform.

Amplitude: The peak amplitude value, that indicates the intensity of the returning wave. It tells us the reflectance of the target with respect to the laser beams' wavelength.

Pulse width: The peak width indicates the surface roughness of the target the laser beam interacted with (scatterer). See Figure 2.

Backscatter Coefficient: The horizontal scattered cross-section of the target with respect to the deployed system wavelength and range that is normalized to the incident angle. This is useful to differentiate targets such as ground vs. vegetation or different vegetation types in a landscape.

The backscatter coefficient (γ_i) requires further explanation. The equation is shown here.

$$\gamma_i = C_{cal} \frac{R_i^2 \hat{P} s_{(p,i)}}{\hat{S} \eta_{atm}} \quad (2)$$

C_{cal} is the calibration constant (9.660535682891269e-08 for this study). R_i is the range, which is the distance from the emitted pulse to the start of the peak. $\hat{P}$ is the peak amplitude. $s_{(p,i)}$ is the standard deviation of the pulse width. $\hat{S}$ is the amplitude of system pulse (emitted waveform). η_{atm} is the atmospheric transmission factor that describes the condition of the atmosphere during data collection. Each of the values used in this calculation are described in Table I.

Each of the three peaks in Figure 2 may be binned to different pixel locations based on their relative GPS positions. However, it is important that peak 1 is the first peak and that peak 3 is the last peak of that waveform. All other peaks in between the first and the last peaks are considered intermediate peaks. Moreover, it is possible for many peaks from different waveforms to fall into the same raster pixel. Summary values can be filtered to include only the first, last, or all peaks. The rasters can then be calculated as statistical moments of the feature values of those filtered peaks. This effectively means that 18 different raster files can be produced for each of the above waveform features using the following statistical moments.

- Average
- Minimum
- Maximum
- Standard Deviation
- Skewness

- Kurtosis

Percentiles (10th, 25th, 50th, 75th, and 90th) of waveform energy and heights are also calculated. These percentile energy and heights can be used to describe the heterogeneity of the 3D structure of the target (e.g. vegetation).

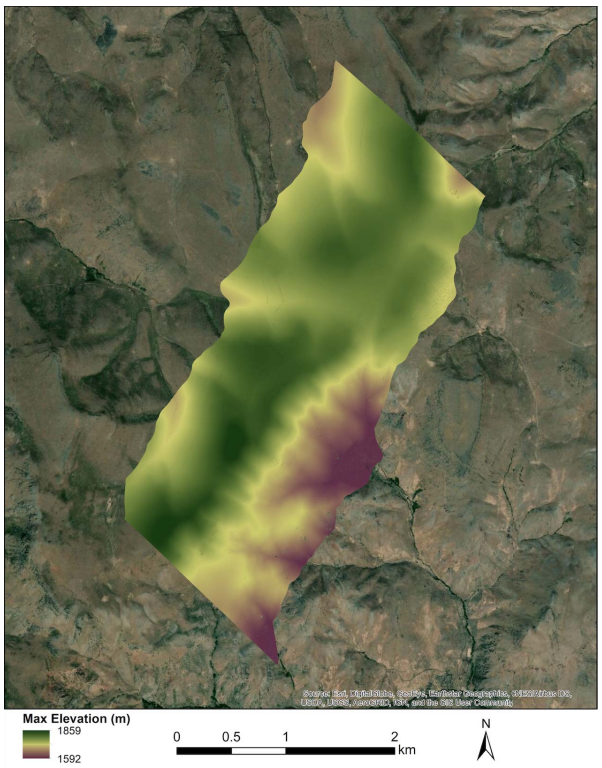


Figure 3: A georeferenced raster superimposed on a map using the ArcGIS software package.

The tool stores the raster files in geotiff format, which is georeferenced and can be viewed with common GIS software (Figure 3). The set of products is described by which characteristic of the waveforms is projected on the 2D surface. The 2D projection helps to display the continuous changes of waveform properties in relation to the surface characteristics (e.g. topography, vegetation) of the targeted landscape. Further, the 2D projected waveform lidar properties can be easily utilized in combination with other remote sensing data products (e.g. hyperspectral, multispectral or radar images) and climate data sets (e.g. precipitation, temperature) in raster format.

IV. Software Design

The software is designed to be modular, extensible, and have a small memory footprint. The overall data flow is from binary data representation to fitted waves to binned peaks, and finally to a raster file (Figure 4). The modularity ensures that different fitting mechanisms can be easily swapped out. Currently, first differencing and Gaussian fitting are implemented in the toolkit while fitting with splines is planned. The extensibility refers to the ability to change out binary readers and raster file writers. The dataflow is designed around avoiding a large memory footprint. Each wave is read and processed using the same small area of memory. The following section describes the binary format supported and the fitting methods implemented.

A. The PulseWaves format

The PulseWaves data exchange format is relatively new and is aimed at storing an entire digitized waveform in a fully georeferenced manner [33]. The PulseWaves format consists of a pair of binary files:

- Pulse file (*.pls): The Pulse file is a binary file consisting of a Header, any number of optional Variable Length Records, the Pulse Records, and one or more Appended Variable Length Records (Table III).
  - Pulse files contain the discrete data points that define each waveform and provide a geo-referenced origin and direction for aligning data with a physical location.
- Waves file (*.wvs): Each Waves file needs a corresponding Pulse file. The waves file contains the samples of the waveforms. Each pulse of the Pulse file contains a reference into the Waves file (Table IV).
  - Waves files store samples of the waveforms and contain the outgoing and returning waveforms.

B. Design Principles

The toolkit is written primarily in C++. The first stage of the process reads the binary files using the open-source PulseWaves library. Every outgoing pulse can have zero, one, or multiple returning waveforms (segments). The outgoing pulses that do not have returning waveforms can be safely ignored. The ones that do have returning waveforms are sent to a fitting function where Gaussian fitting is used to fit the waves and find the peaks. Gaussian fitting is achieved using the GNU scientific library (GSL). These peak

Header
Variable Length Records
Pulse Records
Appended Variable Length Records

Table III: Pulse(*.pls) file structure

Header
Waves of Pulse 0
Waves of Pulse 1
Waves of Pulse 2
Waves of Pulse 3
...
Waves of Pulse k

Table IV: Waves(*.wvs) file structure

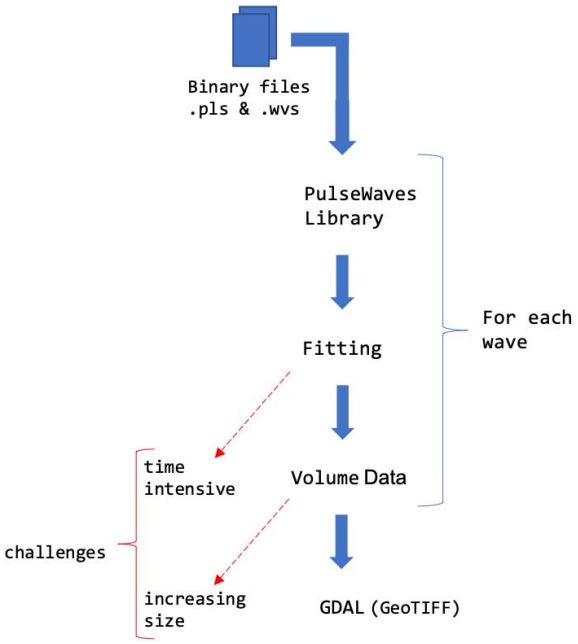


Figure 4: AdaptLidarTools data flow

values, along with additional information like the FWHM and activation time, are stored in a Peak data structure. Once all the peaks are found, we create a Volume data structure that represents these peaks in 2D space.

Some binary files can have more than one return for an outgoing pulse. In such cases, we observed that storing the entire data structure in memory led to significantly larger run-times and often caused the program to run out of memory and crash. To achieve the fastest run-times, it was necessary to minimize the memory footprint the program used. The solution streams data i.e. incrementally reads, processes, discards, and repeats the process. Only current data is in memory at each point.

C. Waveform Characterization

Waveform characterization refers to identifying the features of the waveform peaks as described in Section III. There are a variety of approaches (see Section VI); we have implemented Gaussian Fitting and first differencing.

1) Gaussian Fitting: Gaussian Fitting can be used to extrapolate the returning wave to form a better model of the pulse. Gaussian Fitting creates a Gaussian Function based on key aspects of the raw data, mainly the peak location, amplitude, and the FWHM. These functions are used to find a more accurate peak amplitude and location, for example, when data was collected on either side of the peak, but not at the peak itself, as shown in the first curve in figure 2. This method creates a model of the analyzed terrain that estimates the raw data itself.

The implementation uses the GNU Scientific Library (GSL) [34]. The software first estimates the number of peaks to be fitted and supplies GSL with approximate values of a (peak amplitude), b (time location of the peak amplitude), and c (FWHM) for each peak. We use the non-linear least squares fitting functions available from GSL.

2) First Differencing: Before Gaussian Fitting can be performed, estimates of peak locations and amplitudes as well as the FWHM must be found using first differencing. This method uses differences between data points and past trends in the pulse curves to determine where peaks are located. The FWHM is found by moving either backwards or forwards from the location of a peak until the height drops below half of the peak's amplitude. The information is either used to generate a raster or is used to complete Gaussian Fitting. When used to generate a raster without Gaussian Fitting, first differencing is significantly less time and resource intensive. Only the bare minimum information is extracted from each pulse and data are processed through simple truth statements rather than intensive mathematics. This bare minimum approach, however, compromises the accuracy, the degree to which can be analyzed by comparison with accepted results. First differencing seeks to provide a faster method for applications that do not require precise positioning of the area of interest.

D. Data Validation

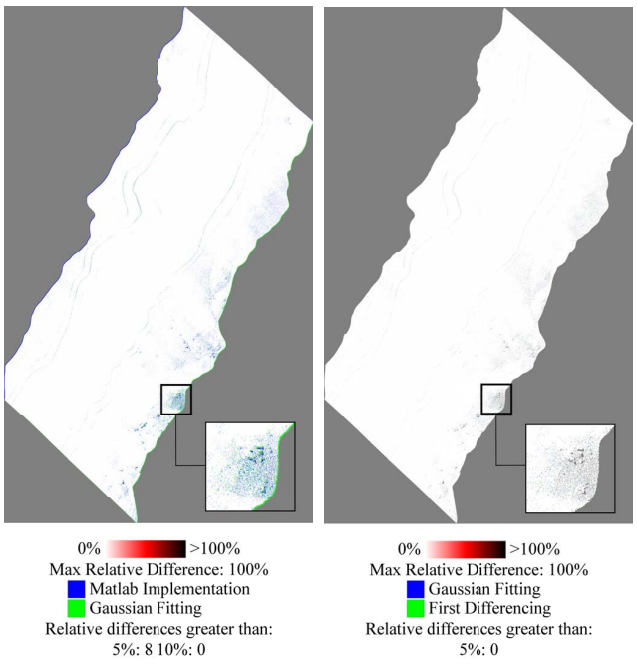
To validate data in the rasters generated by AdaptLidarTools, two Python scripts were developed. One of them generates a heatmap from the data and analyzes individual pixels of the raster file by creating a text dump. The other compares two geotiff raster files by creating a text dump of the raw data of each raster side by side, text file containing general information and statistics about the comparison, and heatmap of the differences between the two files.

V. Performance and Accuracy

This section compares AdaptLidarTools with an earlier version of the software [21] that was written in Matlab [35].

In the Matlab implementation, initial peak amplitudes and their locations were detected using the maxima of Savitzky-Golay smoothed second derivatives of the noise-filtered (6 DN threshold) backscattered waveform. A non-negative least squares fitting algorithm was used to fit the backscattered signals. From the best fit waveform, the number of peaks, their maximum amplitudes, inflection positions and pulse widths were found. This method was time intensive and averaged a day to process a PulseWaves dataset with a pulse file of size 1 GB.

AdaptLidarTools reduces run time considerably. The differences in analysis methods require confirmation of their respective resulting accuracy. Relative difference, in which two data points are compared by dividing their difference by some function of them, provided a key method of verifying the integrity of these new methods. Additionally counts of how many pairs of data points had relative differences



(a) Matlab (1 day) vs. AdaptLidarTools (33 min)

(b) AdaptLidarTools' Gaussian Fitting (33 min) vs. First Differencing (5 min)

Figure 5: Green and blue represent points where either one or the other method produced data. White represents extremely small or no difference. Gray represents no data for either.

greater than 5, 10, 25, 50, 75, and 100% were collected. Heatmaps were created to visualize the differences. Figure 5a shows the comparisons between the Matlab implementation and AdaptLidarTools' Gaussian fitting. Figure 5b shows the comparison between AdaptLidarTools' Gaussian fitting and first differencing. Both comparisons were performed on the same dataset. The comparisons found maximum relative differences of 5.97% and 6.11% respectively. In context, this is an average difference of 0.29m for the comparison in Figure 5a and 1.48m for the comparison in Figure 5b. Each method produced a range of average elevations of 1758m-1761m, which far eclipses the differences between methods. A difference in the location of the data produced by each method was also observed. This is represented by the blue and green in the heatmaps. However the offset is small, with similar features appearing within 1-3 pixels of each other, which translates to 1-3 meters. This is a result of different fitting methods producing marginally different locations for the various waveforms. This offset does not have a significant impact on the overall interpretation of the lidar data. For applications of this technology that do not require high precision, first differencing provides a fast, usable method. Gaussian fitting remains a viable option for those applications that require extremely accurate models of remotely sensed regions, sacrificing speed for accuracy.

VI. Related Work

A number of different approaches were developed to extract features from full waveform signals for vegetation studies. The processing method depends on pulse symmetry, computational efficiency, and the ability to extract maximal information from the waveforms for the required task. The simplest approaches do not use any fitting functions, but rather select the wave maximum [36], or sum amplitudes above a given threshold [37] from the original waves. In these methods, there are no attempts made to reconstruct the missing signal sections during digitization. Fitting simple functions such as quadratic [38] or spline [39] can also be observed to derive features from the lidar waveforms. The accuracy of these linear fitting functions are evaluated using least square fitting [39]. All of the above methods are computationally efficient, however, they produce only a single feature or value as the representation of the waveform.

More computationally expensive waveform fitting techniques include Gaussian fitting, lognormal fitting, and generalized Gaussian [29], [40], [41]. These methods are iterative, reconstruct missing data and mostly use the Levenberg-Marquardt algorithm to find the best solution [41]. Hence, they produce not only more waveform features, but also more accurate ones. While Gaussian fitting works best on symmetric pulses, the lognormal fitting can cope with both symmetric and asymmetric pulses. As an extension of the Gaussian approach, the generalized Gaussian methods provide more flexibility during waveform fitting for feature extraction [40].

VII. Conclusion

AdapatLidarTools achieves a 97.7% reduction in run-time as compared to the original Matlab implementation. An additional 84.8% (99.7% from the original run-time) improvement was observed when using only first differencing to fit the waves. However, a comparison between the geotiffs generated by Gaussian fitting and first differencing indicate that first differencing results in less accurate data. More research into comparison between these two methods over different landscapes is required before a definitive conclusion can be reached. This information and analysis also provided insight into the application of lidar technologies, specifically with how efficiency and accuracy may be balanced. Some vegetation studies may not require precise geolocation accuracy. However, a more localized analysis, such as measuring the height of trees or foliage characteristics likely requires greater accuracy.

This software package provides an open-source and extensible solution to enable remote sensing research based on full-waveform lidar data. The package follows community standards for input and output files and exposes interfaces to

experiment with a variety of approaches to characterizing the waveform data. Minimizing the memory footprint enables faster processing of much larger datasets, providing a much needed mechanism to process waveform lidar data as it becomes more readily available.

ACKNOWLEDGMENT

The authors would like to acknowledge high-performance computing support of the R2 compute cluster (DOI: 10.18122/B2S41H) provided by Boise State University's Research Computing Department.

Part of this work was supported by NASA Terrestrial Ecology [NNX14AD81G] and NASA Earth and Space Science Fellowship 2017 (NESSF 2017) [17-EARTH17F-0209].

REFERENCES

[1] F. Pirotti, "Analysis of full-waveform LiDAR data for forestry applications: a review of investigations and methods," *iForest - Biogeosciences and Forestry*, no. 3, pp. 100–106, 2011. [Online]. Available: http://iforest.sisef.org/contents/?id=ifor0562-004

[2] A. Li, N. F. Glenn, P. J. Olsoy, J. J. Mitchell, and R. Shrestha, "Aboveground biomass estimates of sagebrush using terrestrial and airborne lidar data in a dryland ecosystem," *Agricultural and Forest Meteorology*, vol. 213, pp. 138 – 147, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168192315001847

[3] T. Allouis, S. Durrieu, C. Vga, and P. Couteron, "Stem volume and above-ground biomass estimation of individual pine trees from lidar data: Contribution of full-waveform signals," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 924–934, April 2013.

[4] N. C. Coops, T. Hilker, M. A. Wulder, B. St-Onge, G. Newnham, A. Siggins, and J. A. T. Trofymow, "Estimating canopy structure of douglas-fir forest stands from discrete-return lidar," *Trees*, vol. 21, no. 3, p. 295, Jan 2007. [Online]. Available: https://doi.org/10.1007/s00468-006-0119-6

[5] S. B. DeLong, A. M. Youberg, W. M. DeLong, and B. P. Murphy, "Post-wildfire landscape change and erosional processes from repeat terrestrial lidar in a steep headwater catchment, chiricahua mountains, arizona, usa," *Geomorphology*, vol. 300, pp. 13 – 30, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169555X17304051

[6] K. Lim, P. Treitz, M. Wulder, B. St-Onge, and M. Flood, "Lidar remote sensing of forest structure," *Progress in Physical Geography: Earth and Environment*, vol. 27, no. 1, pp. 88–106, 2003. [Online]. Available: https://doi.org/10.1191/0309133303pp360ra

[7] L. Cao, N. C. Coops, J. L. Innes, J. Dai, H. Ruan, and G. She, "Tree species classification in subtropical forests using small-footprint full-waveform lidar data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 49, pp. 39 – 51, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0303243416300071

[8] A. Hovi, L. Korhonen, J. Vauhkonen, and I. Korpela, "Lidar waveform features for tree species classification and their sensitivity to tree- and acquisition related parameters," *Remote Sensing of Environment*, vol. 173, pp. 224 – 237, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425715301164

[9] J. Reitberger, P. Krzystek, and U. Stilla, "Analysis of full waveform lidar data for the classification of deciduous and coniferous trees," *International Journal of Remote Sensing*, vol. 29, no. 5, pp. 1407–1431, 2008. [Online]. Available: https://doi.org/10.1080/01431160701736448

[10] T. Inomata, D. Triadan, F. Pinzn, M. Burham, J. L. Ranchos, K. Aoyama, and T. Haraguchi, "Archaeological application of airborne lidar to examine social changes in the ceibal region of the maya lowlands," *PLOS ONE*, vol. 13, no. 2, pp. 1–37, 02 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0191619

[11] N. F. Glenn, D. R. Streutker, D. J. Chadwick, G. D. Thackray, and S. J. Dorsch, "Analysis of lidar-derived topographic information for characterizing and differentiating landslide morphology and activity," *Geomorphology*, vol. 73, no. 1, pp. 131 – 148, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169555X05002047

[12] M. Jaboyedoff, T. Oppikofer, A. Abellán, M.-H. Derron, A. Loye, R. Metzger, and a. Pedrazzini, "Use of lidar in landslide investigations: a review," *Natural Hazards*, vol. 61, no. 1, pp. 5–28, Mar 2012. [Online]. Available: https://doi.org/10.1007/s11069-010-9634-2

[13] A. Li, W. Zhao, J. J. Mitchell, N. F. Glenn, M. J. Germino, J. B. Sankey, and R. G. Allen, "Aerodynamic roughness length estimation with lidar and imaging spectroscopy in a shrub-dominated dryland," *Photogrammetric Engineering & Remote Sensing*, vol. 83, no. 6, pp. 415–427, 2017. [Online]. Available: https://www.ingentaconnect.com/content/asprs/pers/2017/00000083/00000006/art00012

[14] J. Mao, A. Ramanathan, J. B. Abshire, S. R. Kawa, H. Riris, G. R. Allan, M. Rodriguez, W. E. Hasselbrack, X. Sun, K. Numata, J. Chen, Y. Choi, and M. Y. M. Yang, "Measurement of atmospheric co_2 column concentrations to cloud tops with a pulsed multi-wavelength airborne lidar," *Atmospheric Measurement Techniques*, vol. 11, no. 1, pp. 127–140, 2018. [Online]. Available: https://www.atmos-meas-tech.net/11/127/2018/

[15] "Cloudcompare: 3d point cloud and mesh processing software: open source project." www.cloudcompare.org, 2019.

[16] TerraScan, "Software for lidar data processing and 3d vector data creation," http://www.terrasolid.com/products/terrascanpage.php, March 2019.

[17] Riegl, "Data processing software riprocess for riegl scan data," http://www.riegl.com/uploads/tx_pxpriegldownloads/11_Datasheet_RiProcess_2016-09-16_13.pdf, September 2019.

[18] M. Isenburg, "Pulsetools," https://github.com/PulseWaves/PulseWaves/tree/master/PulseTools, April 2019.

[19] OPALS, "Module fullwave," https://opals.geo.tuwien.ac.at/html/stable/ModuleFullwave.html, April 2019.

[20] Sorted Pulse Data Library, "The spdlib documentation wiki," http://www.spdlib.org/doku.php?id=spdlib:users:users, July 2019.

[21] N. T. Ilangakoon, N. F. Glenn, H. Dashti, T. H. Painter, T. D. Mikesell, L. P. Spaete, J. J. Mitchell, and K. Shannon, "Constraining plant functional types in a semi-arid ecosystem with waveform lidar," *Remote Sensing of Environment*, vol. 209, pp. 497 – 509, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425718300828

[22] R. N. Faux, J. M. Buffington, M. G. Whitley, S. H. Lanigan, and B. B. Roper, *Use of airborne near-infrared LiDAR for determining channel cross-section characteristics and monitoring aquatic habitat in Pacific Northwest rivers: A preliminary analysis*. Pacific Northwest Aquatic Monitoring Practitioners, 2009. [Online]. Available: https://www.fs.usda.gov/treesearch/pubs/33564

[23] C. Mallet and F. Bretar, "Full-waveform topographic lidar: State-of-the-art," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 1, pp. 1 – 16, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271608000993

[24] C. A. Silva, S. Saatchi, M. Garcia, N. Labrire, C. Klauberg, A. Ferraz, V. Meyer, K. J. Jeffery, K. Abernethy, L. White, K. Zhao, S. L. Lewis, and A. T. Hudak, "Comparison of small- and large-footprint lidar characterization of tropical forest aboveground structure and biomass: A case study from central gabon," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 10, pp. 3512–3526, Oct 2018.

[25] J. B. Blair, D. L. Rabine, and M. A. Hofton, "The laser vegetation imaging sensor (lvis): A medium-altitude, digitization-only, airborne laser altimeter for mapping vegetation and topography," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 54, pp. 115–122, 1999.

[26] "Gedi: Global ecosystem dynamics investigation," https://gedi.umd.edu/, 2019.

[27] A. Hovi and I. Korpela, "Real and simulated waveform-recording lidar data in juvenile boreal forest vegetation," *Remote Sensing of Environment*, vol. 140, pp. 665 – 678, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425713003714

[28] I. Korpela, A. Hovi, and L. Korhonen, "Backscattering of individual lidar pulses from forest canopies explained by photogrammetrically derived vegetation structure," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 83, pp. 81 – 93, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271613001391

[29] W. Wagner, A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka, "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 2, pp. 100 – 112, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271605001024

[30] K. D. Fieber, I. J. Davenport, J. M. Ferryman, R. J. Gurney, J. P. Walker, and J. M. Hacker, "Analysis of full-waveform lidar data for classification of an orange orchard scene," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 82, pp. 63 – 82, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271613001238

[31] G. Sun and K. J. Ranson, "Modeling lidar returns from forest canopies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 6, pp. 2617–2626, Nov 2000.

[32] J. Heinzel and B. Koch, "Exploring full-waveform lidar parameters for tree species classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 1, pp. 152 – 160, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0303243410001145

[33] M. Isenburg, "Pulsewaves," https://rapidlasso.com/pulsewaves/, January 2019.

[34] B. Gough, *GNU Scientific Library Reference Manual - Third Edition*, 3rd ed. Network Theory Ltd., 2009.

[35] MATLAB, *version 9.1 (R2016b)*. Natick, Massachusetts: The MathWorks Inc., 2016.

[36] R. Gaulton, F. Danson, F. Ramirez, and O. Gunawan, "The potential of dual-wavelength laser scanning for estimating vegetation moisture content," *Remote Sensing of Environment*, vol. 132, pp. 32 – 39, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425713000059

[37] G. Landi, "Properties of the center of gravity as an algorithm for position measurements," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 485, no. 3, pp. 698 – 719, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016890020102071X

[38] F. Blais and M. Rioux, "Real-time numerical peak detector," *Signal Processing*, vol. 11, no. 2, pp. 145 – 155, 1986. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0165168486900332

[39] A. Roncat, G. Bergauer, and N. Pfeifer, "B-spline deconvolution for differential target cross-section determination in full-waveform laser scanning data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 4, pp. 418 – 428, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271611000244

[40] A. Chauve, C. Mallet, F. Bretar, S. Durrieu, M. Pierrot-Deseilligny, and W. Puech, "Processing Full-Waveform Lidar Data: Modelling Raw Signals," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 2007*, Espoo, Finland, Jul. 2008, pp. 102–107. [Online]. Available: https://hal-lirmm.ccsd.cnrs.fr/lirmm-00293129

[41] S. Hancock, J. Armston, Z. Li, R. Gaulton, P. Lewis, M. Disney, F. M. Dansona, C. S. Alan Strahler, K. anderson, and K. J. Gaston, "Waveform lidar over vegetation: An evaluation of inversion methods for estimating return energy," *Remote Sensing of Environment*, vol. 164, pp. 208 – 224, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003442571500142X