NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

FORESTRY IDENTIFICATION WITH LIDAR WAVEFORM AND POINT CLOUDS

by

Andrew S. Davis

June 2018

Thesis Advisor: Richard C. Olsen
Co-Advisor: Jeremy P. Metcalf

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2018	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE FORESTRY IDENTIFICATION WITH LIDAR WAVEFORM AND POINT CLOUDS	**5. FUNDING NUMBERS**
6. AUTHOR(S) Andrew S. Davis	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

The aim of this study was to analyze discrete and waveform data to improve existing Terrain Classification (TERCAT) capabilities. Light Detection and Ranging (LiDAR) data were collected over the Point Lobos State Park, which contains various buildings, vegetation, and man-made surfaces. Data were used from two separate airborne LiDAR systems, Optech Titan and Airborne Hydrography AB (AHAB) Chiroptera II. Classic standard point cloud analysis techniques were used with the discrete data. Waveform data were analyzed following a gridding or rasterization process to enable visualization and processing. Analysis approaches used were ENVI classification tools such as Support Vector Machines (SVM), Spectral Angle Mapper (SAM), Maximum Likelihood, and K-means to classify returns. Through the use of this analog to hyperspectral data analysis to classify vegetation and terrain, the results are that, by using the Support Vector Machines with full waveform data, we can successfully improve low vegetation classifiers by 40%, and differentiate tree types (Pine/Cypress) at 40–60% accuracy.

14. SUBJECT TERMS LiDAR, full waveform LiDAR, remote sensing, terrain classification, laser altimetry, Support Vector Machines	**15. NUMBER OF PAGES** **81**
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

FORESTRY IDENTIFICATION WITH LIDAR WAVEFORM AND POINT CLOUDS

Andrew S. Davis
Lieutenant, United States Navy
BS, U.S. Naval Academy, 2013

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN SPACE SYSTEMS OPERATIONS

from the

NAVAL POSTGRADUATE SCHOOL
June 2018

Approved by: Richard C. Olsen
Advisor

Jeremy P. Metcalf
Co-Advisor

James H. Newman
Chair, Department of Space Systems Academic Group

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The aim of this study was to analyze discrete and waveform data to improve existing Terrain Classification (TERCAT) capabilities. Light Detection and Ranging (LiDAR) data were collected over the Point Lobos State Park, which contains various buildings, vegetation, and man-made surfaces. Data were used from two separate airborne LiDAR systems, Optech Titan and Airborne Hydrography AB (AHAB) Chiroptera II. Classic standard point cloud analysis techniques were used with the discrete data. Waveform data were analyzed following a gridding or rasterization process to enable visualization and processing. Analysis approaches used were ENVI classification tools such as Support Vector Machines (SVM), Spectral Angle Mapper (SAM), Maximum Likelihood, and K-means to classify returns. Through the use of this analog to hyperspectral data analysis to classify vegetation and terrain, the results are that, by using the Support Vector Machines with full waveform data, we can successfully improve low vegetation classifiers by 40%, and differentiate tree types (Pine/Cypress) at 40–60% accuracy.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

3D	Three-Dimensional
AHAB	Airborne Hydrography AB
AOL	Airborne Oceanographic LiDAR
CALIOP	Cloud-Aerosol LiDAR with Orthogonal Polarization
CCRS	Canada Center for Remote Sensing
CHP	Canopy Height Profile
DBH	Diameter at Breast Height
DRL	Discrete Return LiDAR
DSM	Digital Surface Model
DTM	Digital Terrain Model
EM	Electromagnetic
ENVI	Environment for Visualizing Images
FWHM	Full Width at Half Maximum
GLAS	Geoscience Laser Altimeter System
HBEF	Hubbard Brook Experimental Forest
ICESat	Ice, Cloud, and Land Elevation Satellite
IDL	Integrated Data Language
LAI	Leaf Area Index
LiDAR	Light Detection and Ranging
LSM	Least Squares Method
LVIS	Laser Vegetation Imaging System
MBLA	Multi-Beam Laser Altimeter
NASA	National Aeronautics and Space Administration
nDSM	normalized Digital Surface Model
NIR	Near-Infrared
QTM	Quick Terrain Modeler
ROI	Region of Interest
RTM	Radiative Transfer Method
SAM	Spectral Angle Mapper
SAR	Synthetic Aperture Radar

SLA	Shuttle Laser Altimeter
SLICER	Scanning LiDAR Imager of Canopies by Echo Recovery
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
SWIR	Shortwave Infrared
TERCAT	Terrain Classification
TIN	Triangulated Irregular Network
VIP	Very Important Points

ACKNOWLEDGMENTS

A special thank you to

Dr. Richard Olsen
Jeremy Metcalf

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. PURPOSE OF RESEARCH

The transition from aerial photography to the many new technologies of remotes sensing, since the 1960s have continued apace. Infrared imaging, Synthetic Aperture Radar (SAR), Scatterometers, Sounders, and Light Detection and Ranging (LiDAR) have emerged as useful technology for military and civilian uses, from air and space (Remote Sensors, 2018).

One attractive, underutilized, and available remote sensing tool is LiDAR. This research explores the uses of full waveform LiDAR data to help advance classification capabilities over various terrains, including areas containing dense canopies. Waveform data can potentially benefit the intelligence community because these systems can fill information gaps or provide additional insights to enhance the common intelligence picture. LiDAR systems can collect data while operating in various environments including space. Furthermore, ground forces rely heavily on topographic maps because terrain plays a significant factor in military planning, targeting, and mission execution. If using waveform LiDAR improves the accuracy of tactical mapping, especially in regions with dense canopy cover, it could enhance our understanding of the environment and lead to increased mission success.

LiDAR is an active remote sensing system, which sends a burst of light energy out to a target; using the return pulse information, we can calculate the location of that target. The preliminary uses of LiDAR have primarily focused on discrete return studies. Discrete return LiDAR studies accurately produce three dimensional (3D) topographic maps with intensity values for each recognized pulse return. Discrete LiDAR data can provide useful products, but, during the processing stage some of the return data is filtered or lost, reducing its accuracy. Full waveform studies have only gradually increased because only a limited amount of tools and resources can translate the complex data for analysis.

This thesis focuses on the post-processing stage, to include filtering and analyzing raw waveform and discrete data in order to improve current vegetation classification

methods. Waveform LiDAR data provides additional metrics, such as pulse width and backscattered area. These metrics can be calculated from the waveform data, since the pulse information is available.

B. OBJECTIVE

The objective of this thesis is to investigate the benefits of full waveform LiDAR. In order to investigate these benefits, this thesis processes and analyzes two data sets to compare classification capabilities. Additionally, this study focuses on distinguishing the various types of vegetation species using machine learning tools that filter the waveform data. In order to do this, we use waveform data sets, along with ground observation data, to validate results. The data we analyzed includes various vegetation returns from Point Lobos State Park, located in Carmel-by-the-Sea, CA.

II. BACKGROUND

The previous chapter discusses the objective and purpose of this thesis. This first section of this chapter will explain the basic differences between discrete and waveform LiDAR data. The second section discusses various full waveform studies which helped inspired and establish the analysis methods used in this thesis.

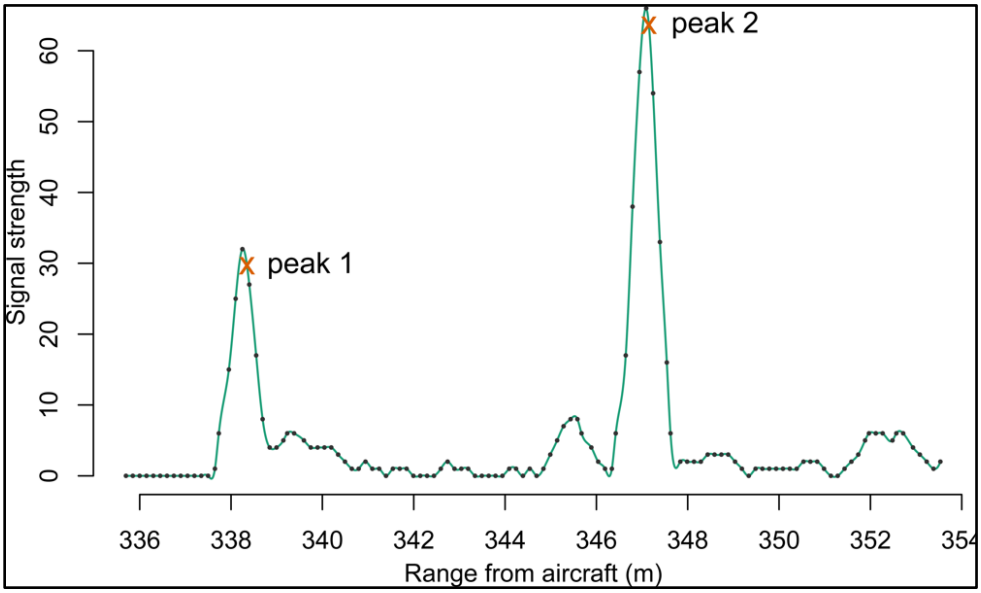
A. LIDAR BACKGROUND

LiDAR is a remote sensing tool which has provided an abundance of utility since its creation. Early uses of LiDAR systems were directed toward bathymetric, atmospheric and meteorological studies. Many terrestrial applications emerged with the evolving technology. One of the most common uses is to analyze and create 3D high-resolution maps. LiDAR systems, similar to radar in many ways, operate on shorter wavelengths than imaging radar systems. LiDAR wavelengths range from 500–2000 nm, which improves the detection accuracy of smaller objects such as low-lying vegetation and terrain variations (Hancock et al., 2012). LiDAR systems are capable of operating in environments ranging from underwater to space. One of the first well-known LiDAR space systems was a dedicated payload on the Apollo 15 mission, in which the system mapped the surface of the moon (Abshire, 2011). Research and development areas of studies within the LiDAR community are beginning to shift toward the exploitation of full-waveform LiDAR. This waveform exploitation shift can increase terrain characterization capabilities and future LiDAR uses.

B. FULL WAVEFORM FUNDAMENTALS

Discrete LiDAR returns are obtained by recording the peak points within a LiDAR pulse return. These peak points are then associated within a specific X, Y, and Z location along with an intensity level. Discrete LiDAR systems are capable of recording multiple returns for a single laser pulse, but, these discrete collection systems typically have a blind spot, due to the limitations of discrete data. These limitations of discrete LiDAR systems are discussed in the 2015 *Remote Sensing of the Environment Journal* by Sumnall, Hill, and Hinsley. The study discusses that these limitations are typically a 1.2 to 5 meter gap

between returns, during which other surfaces cannot be detected (Sumnall, Hill, & Hinsley, 2015). Full waveform LiDAR can address these limitations, since the signal is measured as a function of time. A full waveform system measures the full distribution of the laser's return energy, which is why additional metrics can be produced. These additional identification metrics applied to various studies have resulted in significant improvements, especially when conducting canopy understory collections (Anderson et al., 2016). Figure 1 shows a typical waveform return from a relatively modern system. The discrete returns will be picked off as "peak 1" and "peak 2" by the LiDAR processor, and identified as the first and second returns.

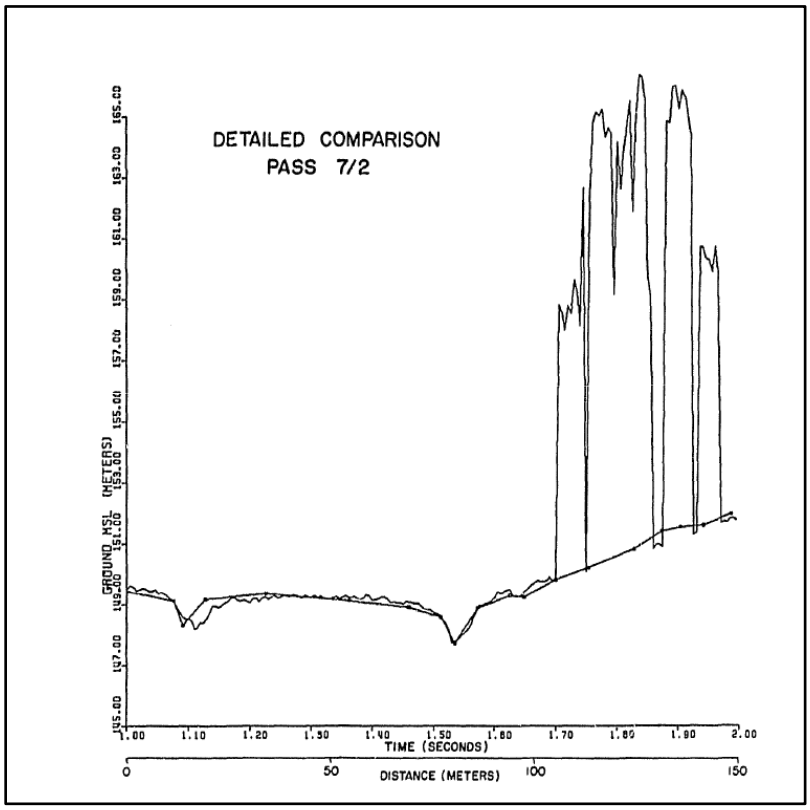


There are 120 waveform samples shown as circles. Peak detection distances for these returns were recorded at 338 and 347 meters, which are shown with "x's."

Figure 1. Data from an airborne laser scanner. Source: Vaughn et al. (2012).

Current operational systems primarily measure discrete point returns, but waveform systems have been in use since the initial days of LiDAR, particularly in the domain of bathymetric LiDAR. In 1980, a joint study involving the National Aeronautics and Space Administration (NASA) and the U.S. Army Corp of Engineers investigated NASA's Airborne Oceanographic LiDAR (AOL) to determine additional capabilities that LiDAR systems can offer. Figure 2 shows some of those first results, where the tree canopy is

identified by spiked returns. The results of the study discussed NASA's plans to further investigate the uses and experimentations of waveform LiDAR for forestry management information and other applications. The identified potentials included the ability to provide biomass density, tree height total, and stem height (Krabill et al., 1980; Krabill et al., 1984).



Tree elevations from 15–20 m above ground level are detected here in one of the first LiDAR forestry results.

Figure 2. Comparison of AOL data with ground truth.
Source: Krabill et al. (1980).

Figure 3 is an idealized representation of the laser pulse collection, corresponding to the data shown in Figure 1. Figure 3 illustrates the collection of both full waveform and discrete LiDAR data. The discrete data, or points, given here would correspond to the peaks in Figure 3. Waveform data are displayed in the right portion of the graphic and it is apparent that a more comprehensive vertical profile can be obtained when compared to the discrete return data in the graphic. The type of data shown in this figure can be used to

characterize tree density and identify canopy gaps (Jalobeanu & Goncalves, 2012). Additionally, the capabilities of full waveform data are improving because advanced processors and data storage capabilities allow systems to digitize returns at a higher rate (Vaughn et al., 2012). The increase in return collection capabilities will allow systems to replicate a digital wave more accurately, especially the data before and after the peak. Ultimately, waveform data can provide a complete vertical profile to associated returns, which will enhance the volume of data available for exploitation.

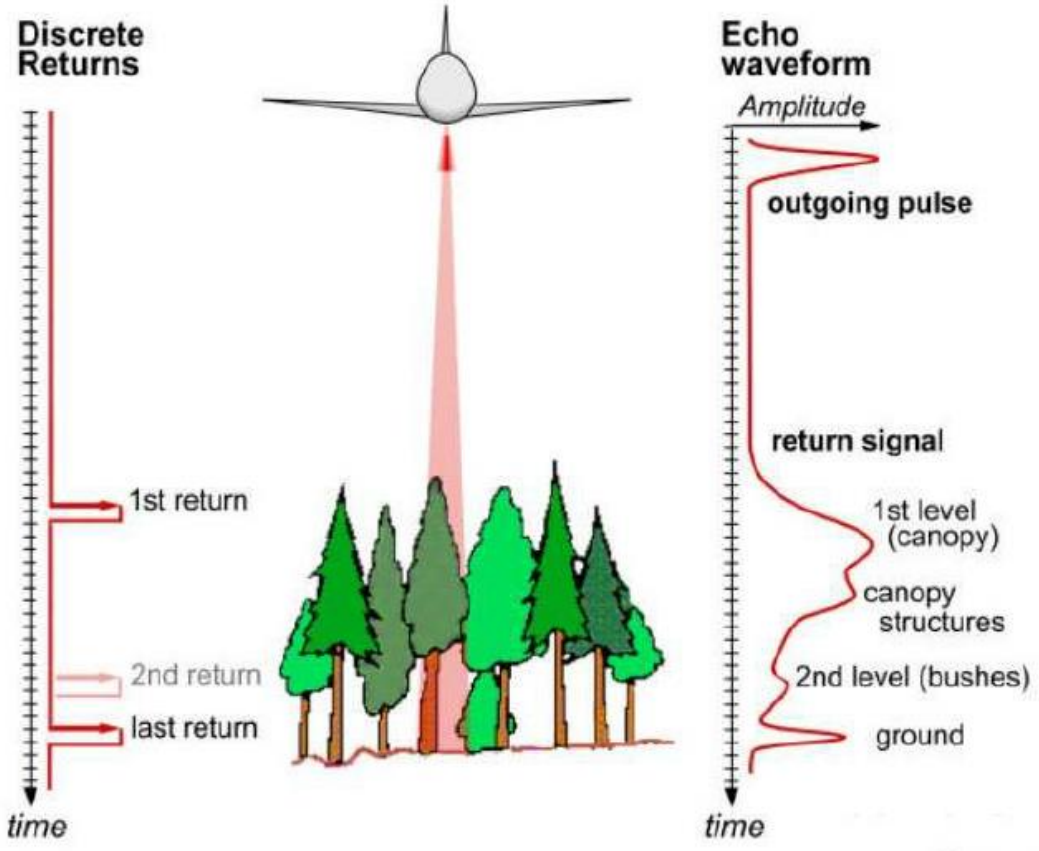


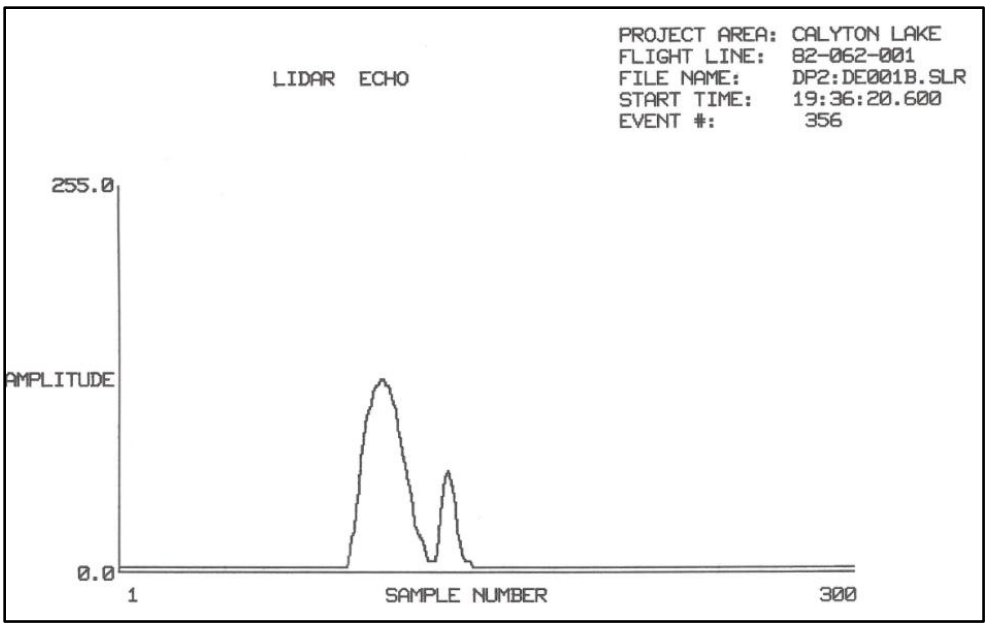
Figure 3. Graphic displaying discrete (left) and full waveform returns (right). Source: Ferraz et al. (2009).

Waveform data provide the standard discrete retrievable parameters, like range and amplitude, but also additional metrics, such as pulse width and pulse shape deviation. This makes additional classification of objects, such as low vegetation, more obtainable with

echo width information. Additionally, waveform data can be manipulated by performing a Gaussian decomposition to improve the classification of targets (Jalobeanu & Goncalves, 2012). Also, to further improve the accuracy of the data, we can use methods such as an iterative closest point algorithm, which will increase the reliability of each point (Ulrich, 2011).

1. Early LiDAR systems

To explore LiDAR system's capabilities, initial studies focused on comparing the accuracy and reliability of the results to other established systems and field measurements. One of the first terrestrial studies released that addressed full waveform data was conducted in 1985 by the Canada Center for Remote Sensing (CCRS). The center conducted a forest canopy study with ground observations to determine the applications and benefits of full waveform LiDAR data. The study followed the use of the Airborne Oceanographic LiDAR (AOL) system, which focused on bathymetric surveys. The CCRS found utility from the AOL data, but decided to focus, instead, on terrestrial studies. The study analyzed waveform data for individual pulse returns to gain more insight into the return pulse. After comparing the results to ground observations, the model produced tree height estimations within +/- 4.1 meters, with a 95% confidence level. The study also concluded that using the leading edge discriminators to determine height was optimum at 85% of the waveform max (Aldred & Bonnor, 1985). Following the CCRS study, a trend of studies began focusing on improving the accuracy of waveform data. Figure 4 displays a single pulse with two detected returns, the first being a strong return, while the second being weaker.

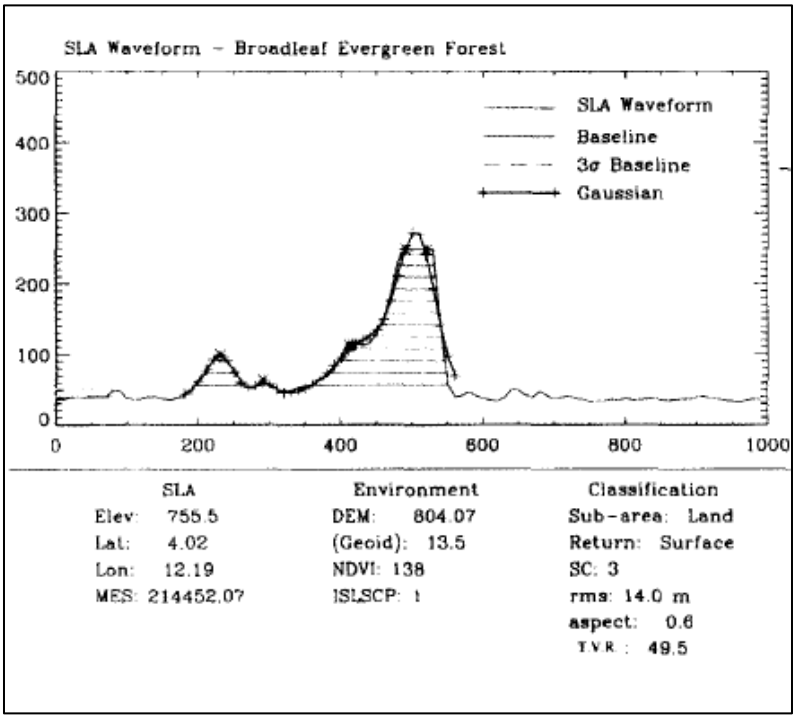


Shown are two returns, the strong first return (left) and weaker ground return (right).

Figure 4. A typical laser returns from a softwood stand.
Source: Aldred and Bonnor (1985).

Early space tests of full waveform LiDAR were conducted by NASA in January of 1997. The Shuttle Laser Altimeter (SLA) was launched on STS-72 and collected over 83 hours of LiDAR data. The sensor had a 100 meter diameter footprint that was equipped with surface and atmospheric LiDAR modes. The result of this experiment led to significant data collection over rugged terrain areas in Africa, South Asia, and South America (Bufton, Harding, & Garvin, 1999). Additionally, 57% of the data resulted in validated surface ranges, and, after further processing, the data, in some cases, accurately distinguished between ground and vegetation canopies (Garvin et al., 1998). This system set the stage for further STS missions equipped with the SLA, as well as supporting follow on systems such as the Ice, Cloud, and Land Elevation Satellite-1 (ICESat) and Multi-Beam Laser Altimeter (MBLA). ICESat-1 was eventually launched and will be discussed in more detail later in the chapter. Figure 5 displays an SLA echo that clearly distinguishes between the canopy and ground. The first peak is considered a canopy return just before 250 milliseconds, while the larger second peak is most likely a ground return. Additionally,

this graphic includes the Gaussian decomposition analysis, which smooths out the raw waveform data.



The figure is clearly defining the canopy with the first peak and the ground return with the last peak. The vertical axis is measuring amplitude and horizontal axis is in time using nanoseconds.

Figure 5. The SLA waveform data with two peaks.
Source: Garvin et al. (1998).

In 2000, NASA, United States Department of Agriculture, and the Smithsonian Environmental Research Center conducted a study to test methods for validation in a closed canopy environment. The study used an airborne system known as Scanning LiDAR Imager of Canopies by Echo Recovery (SLICER) to collect data to estimate backscattered energy returns from the canopy. The study also included an estimation method known as Canopy Height Profile (CHP). The data to create this model consisted of SLICER and ground points to calculate a CHP. The survey area was classified as a broad leaf deciduous forest in Eastern Maryland. The SLICER waveform data were able to produce an accurate vertical structure, which compared well with the ground observation data (Harding et al.,

2001). The SLICER footprint size was 10 meters. This size was chosen because the system could easily differentiate between ground and canopy returns at this setting (Harding et al., 2001). This thesis uses a small-footprint system, with a footprint size of 10's of centimeters. Focus on large-footprint systems dominated in the early systems, and this trend continues today.

2. Canopy Height Studies

The application of waveform LiDAR systems to forest canopies continued, and was extended by the use of the NASA Goddard Space Flight Center's Land, Vegetation, and Ice Sensor (LVIS) system, still in active use today. In 2002, Michael Lefsky and his colleagues completed a waveform study comparing the relationships between collected LiDAR waveform data to determine accurate canopy height measurements. The study targeted three unique terrestrial biomes, using SLICER and LVIS airborne LiDAR systems. The collected data were then compared to on-site field measurements for accuracy. The data were filtered using a stepwise multiple regression analysis approach, which included the canopy structure and field estimates for all three sites. The study resulted in a preliminary hypothesis that analysis of the three biomes produced a single equation which can be used to estimate above ground biomasses (Lefsky et al., 2002). Figure 6 displays the forest canopy structure for all three biomes the Lefsky study collected on. The vertical spikes represent the canopy ceiling which in each biome helped create an estimated canopy height profile. Additionally, the color scheme used corresponds to point density within each area, where red represents large densities. This graphic shows a large fraction of processed points were closer to the ground.

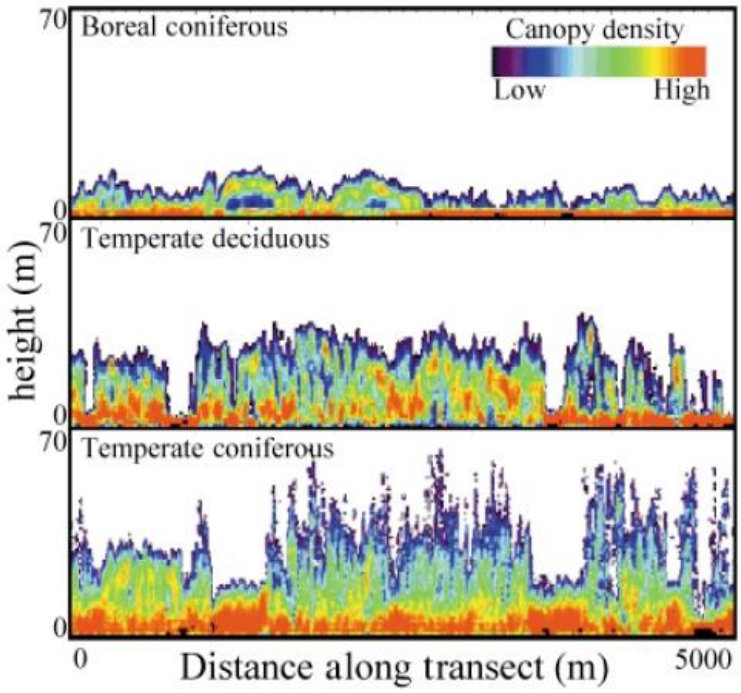


Figure 6. SLICER data measurements of the forest structure.
Source: Lefsky et al. (2002).

In 2003, ICESat, a free-flying satellite successor to the SLA sensor, became operational (Lefsky et al., 2005). The satellite had a primary mission of measuring ice sheets, aerosol levels, and topographical information. ICESat is equipped with the Geoscience Laser Altimeter System (GLAS), which is a full waveform sampling LiDAR system. The system has an elliptical footprint size which varies, but on average it is 53 x 97 m (Lefsky et al., 2005). Lefsky et al. (2002) applied the GLAS data to topographic studies. The study aimed to use the waveform data to improve the previous model created in 2002 to estimate the canopy height. This study utilized two space-based systems, ICESat and the Shuttle Radar Topography Mission (SRTM), to increase the size of the data set. The study also analyzed the extent of the waveform leading edge to minimize estimation errors (Lefsky et al., 2002).

The study's results show that the estimated forest canopy height achieved a 69% explanation of variance using a proposed three parameter equation (Lefsky et al., 2005). Additionally, using metrics obtained from the waveform data, including waveform extent,

terrain index, and scaling factors for the waveform, the study found a correlation between above ground biomass and maximum canopy height. The results increased the accuracy of the data overall, obtaining an explanation for 73% of the variation (Lefsky et al., 2005). Table 1 includes the three- and two-parameter equations used in the study to help improve the overall coefficient of determination value. Included in the table are the R^2 values. The R^2 values in the graphic correspond to the data and its relationship to a fitted regression line. If the R^2 value is a low percentage then there is a large amount of unexplained variance (Frost, 2017).

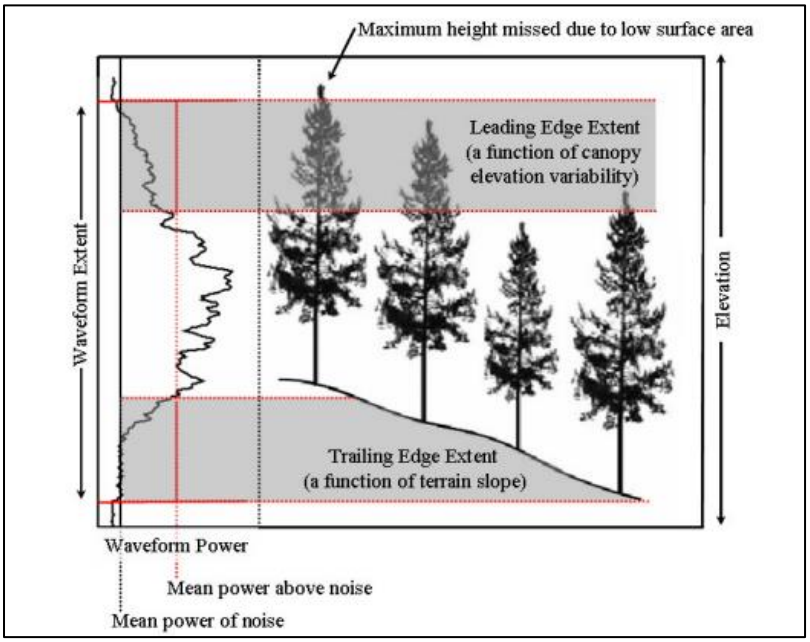
Table 1. Displayed are the three and two parameter equations results. Adapted from Lefsky et al. (2005).

Comparison between the two and three parameter equations							
Parameters	**R^2 (%)**	**B_0**	**B_1**	**B_2**	**Bias(m)**	**RMSE (m)**	**Count**
2	**59**	**.68778**	**.13517**	**-**	**.01**	**4.85**	**23**
3	**69**	**.62108**	**.36924**	**.41841**	**.01**	**4.21**	**23**

In the Lefsky et al. (2005) study, the three parameter equation includes the leading edge extent. Symbols included in the equation are B_0 (coefficient for the waveform), B_1 (coefficient for the terrain index), B_2 (coefficient for the leading edge), I (extent for the leading edge) and g (terrain index).

Two years later, a revised method for forest canopy height estimations was published, this time to address estimations on sloped terrain, which previously resulted in skewed estimations (Lefsky et al., 2007). On flat surfaces the returns can be differentiated because the first echo return can be determined using the difference between the average elevations of the ground returns. Previous studies, which incorporated sloped terrain, have shown that LiDAR returns have difficulties differentiating between the canopy and ground. This phenomenon occurs because some canopy returns are at the same elevation as higher-elevation ground points. This skews the vertical extent data and makes height measurements inaccurate. Figure 7 represents a scanning illustration from the study, which defines surveying an area on a slope. The two variables introduced include the leading edge extent and trailing edge extent. Also included in the study is an eccentricity recommendation, which can significantly impact the collected data if the value is large. A

larger eccentricity value will increase the footprint shape and orientation which will increase canopy height errors.



Graphic using waveform leading edge and trailing edge extent, which demonstrates how the slope of the terrain can affect ground and canopy returns.

Figure 7. Displayed are key terms used in waveform studies.
Source: Lefsky et al. (2007).

Since most vegetation areas are not flat surfaces and vary in ground height, Lefsky et al. (2007) conducted research to determine a sufficient algorithm to address the slope terrain issue. The results fared well after comparing them to the ground-observed results, showing RMSE and R^2 values which were consistent, with the exception of the New Hampshire site. Additionally, this study "found a limited 1:1 relationship between the trailing edge extent and topographic slope" (Lefsky et al., 2007). Later, in 2010, Lefsky conducted another study to map the forest, but on a global scale. This study involved the use of LiDAR and multispectral data. The study once again used statistical analysis, but involved Lorey's height to calculate the height estimations. Lorey's height is calculated using the mean height of the trees and basal area, the tree's cross-sectional area 4.5 meters from the ground, to estimate the mean tree height (Pourrahmati et al., 2017). This

estimation was implemented because it adds more weight to taller trees within the data. Using the Lorey height formula, the study calculated two height values, the mean and 90th percentile using the LiDAR data (Lefsky et al., 2010). Figure 8 displays the six forest types this study analyzed, where the mean height for the 90th percentile roughly feel between 20–30 meters, except for the boreal forest data.

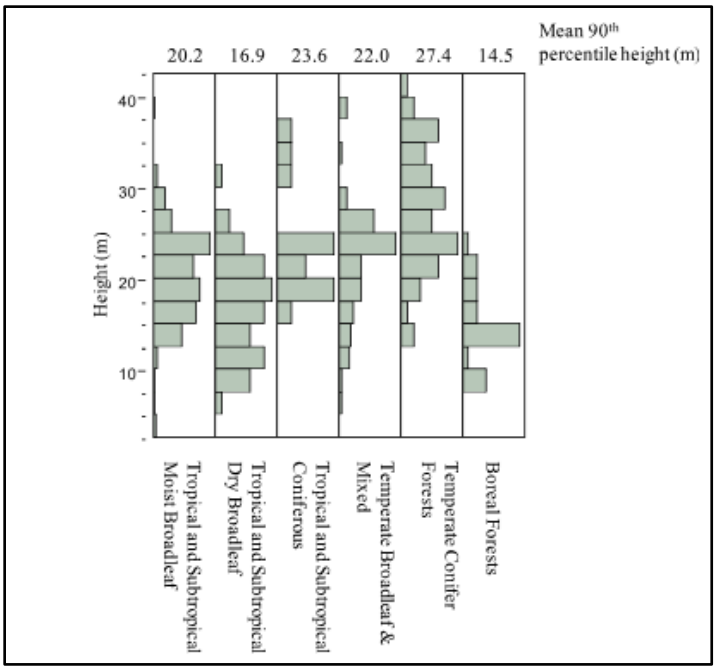


Figure 8. Distribution for the 90th percentile height estimations.
Source: Lefsky et al. (2010).

Height estimation studies have significantly improved over the years and now include directed studies analyzing various sections of canopies. In 2013, a study collected LVIS data over the Hubbard Brook Experimental Forest (HBEF) in New Hampshire. This study used both a small-footprint Discrete Return LiDAR (DRL) and full waveform data. The study used LVIS data measurement at 25m GSD to increase the accuracy of the estimated ground elevation (Whitehurst et al., 2013). The study concluded with results providing ample information to determine variation of the canopy's vertical structure. Figure 9 displays the mean foliage area profiles, which are separated by 3 meters. The graphics shows how the foliage concentration can be found in the midstory range between

6–15 meters, but an approximate peak height is apparent between 9 and 12 meters. The benefits of this study can improve layering information, which can be useful for habitat modeling and forestry management (Whitehurst et al., 2013).

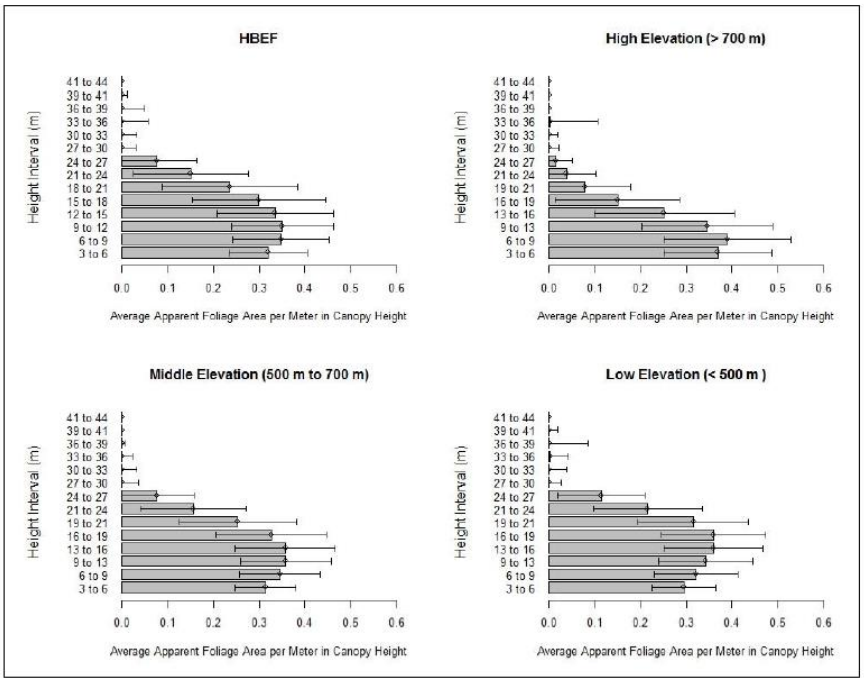
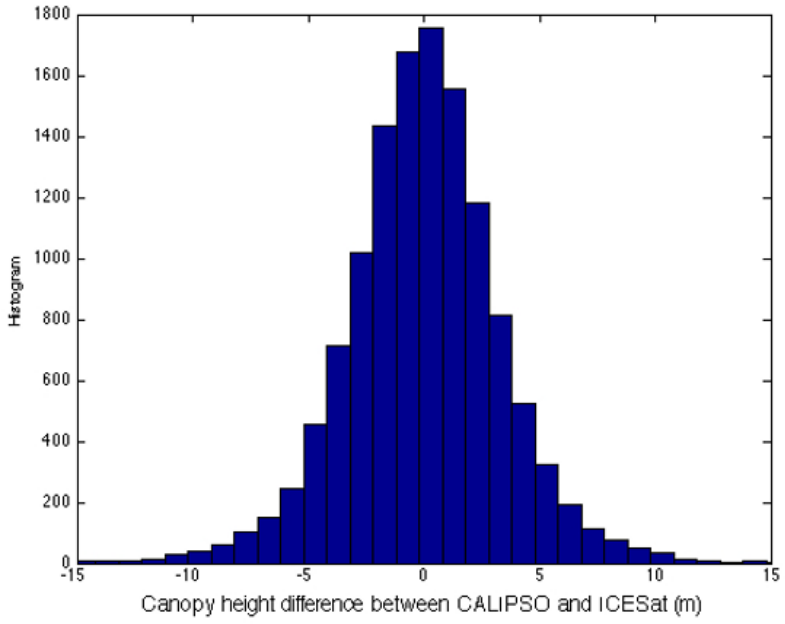


Figure 9. Height interval data from the HBEF dataset.
Source: Whitehurst et al. (2013).

Overall, the estimations for canopy height have improved significantly over the past 20 years, especially with new systems that utilize multiple wavelengths. For example, a study conducted in 2016 by NASA's Langley Research Center used the Cloud-Aerosol LiDAR with Orthogonal Polarization (CALIOP) system to determine forest canopy height estimations (Lu et al., 2016). The study analyzed the CALIOP data, which uses three receiver channels to determine height estimations, though with relatively poor altitude resolution. The objective was to analyze the penetration capabilities of two wavelengths (1064 and 532 nm) to improve the estimated forest canopy height. The data were then compared to the ICESat-estimated canopy height for validation. After comparisons were completed, the results proved the two sets of data were highly correlated, producing a

correlation coefficient of .89 (Lu et al., 2016). This proposed method provided the idea of using two wavelengths in the future to produce increasingly accurate vegetation height studies. Figure 10 displays the level of correlation for the combined data set, the peak value in the middle of the graph shows that the data is highly correlated, especially since most of the data is closer to the center.



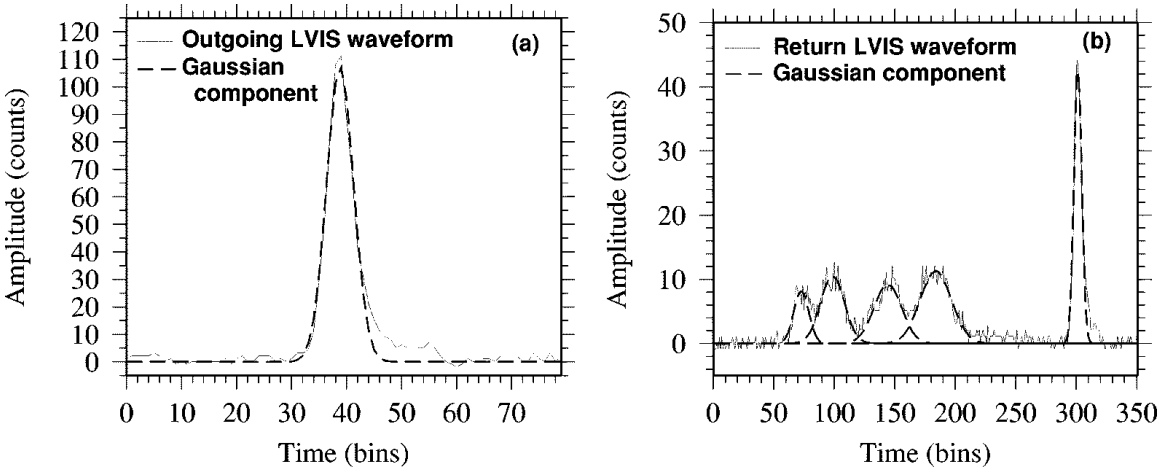
The majority of the data is concentrated near the middle which demonstrates the data is highly correlated.

Figure 10. Canopy height difference between CALIOP and ICESat data. Source: Lu et al. (2016).

3. Filtering Waveform Data

Although waveform data has an immense amount of useful data, some of the most prevalent problems involve filtering out noise and analyzing the large volume of data in a reasonable amount of time. A popular published research area involves using statistical models to reduce waveform data into reasonable data sets. One of the most common approaches in this area is to fit multiple Gaussian distributions to the waveform data (Hofton, Minster, & Blair, 2000). This approach is widely used by the LVIS researchers

with the large-footprint system. The method involves using a non-negative Least Squares Method (LSM) to first establish initial amplitude estimates; then, in order to filter out background noise, an importance factor for each Gaussian was determined using initial half width and amplitude estimates (Hofton et al., 2000). Once all the Gaussians were ranked by importance, the data were further reduced by using the Levenburg-Marquardt method, which determines step sizes while incorporating Newton's method to receive a best fit. After completing this process, the study targeted a specific accuracy to be achieved; if the results fell short, then supplementary Gaussians were re-optimized and included in the data (Hofton et al., 2000). The desired results were eventually achieved, which validated the method's practicality. Figure 11 shows the converted noisy return waveform data transformed into a sequence of Gaussian components which are represented with the dashed lines.



The return LVIS waveform data is the solid line, while the Gaussian is the dotted line. The Gaussian line is able to convert the complex returns into a simpler set of data, while maintaining accuracy.

Figure 11. Output Pulse (a) and return pulse (b).
Source: Hofton et al. (2000).

A subsequent significant study involved optimizing LiDAR and SAR data. Sun et al. (2011) explored methods to improve biomass mapping by combining the two remote sensing techniques. The study analyzed Laser Vegetation Imaging Sensor (LVIS) data to

predict the LiDAR biomass map. The study produced a referenced biomass map with an R^2 value of .71 and RMSE of 31.33 Mg/ha (Sun et al., 2011). The next part of the study used the biomass map to compare it with the co-located SAR data accuracy. The SAR data were capable of predicting LiDAR samples to an extent that resulted in an R^2 value of .63-71 and RMSE of 32–28.2 Mg/ha (Sun et al., 2011). These results were considered preliminary, but did provide enough information to prove that LiDAR and SAR data could be fused together to produce forest biomass mapping products.

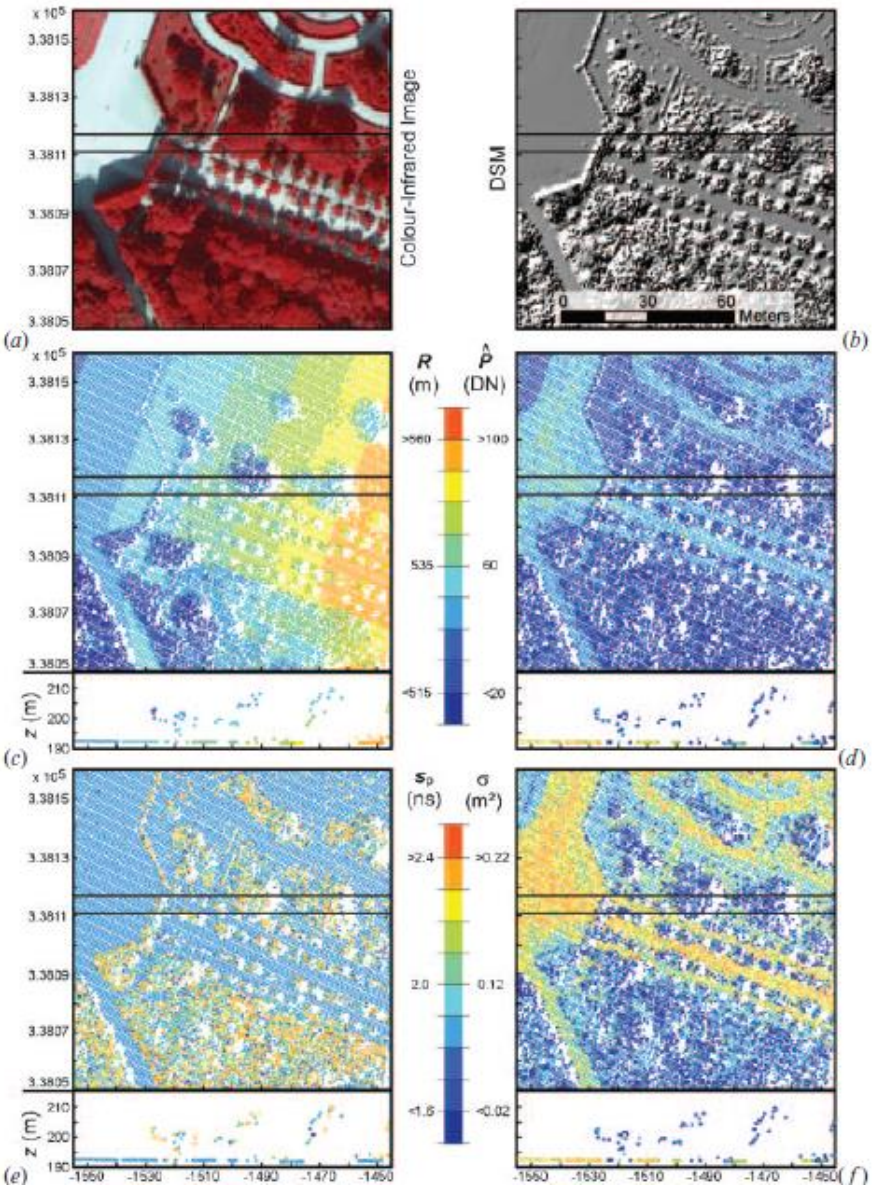
4. Classification LiDAR Studies

Classification studies relating to the identification of target types, specific tree species, and canopy layers have recently garnered more interest over the years, as full waveform data has become more available. The Institute of Photogrammetry and Remote Sensing within Vienna's University of Technology conducted a study to determine if scattering characteristics from full-waveform LiDAR data could be used to classify survey areas. Wagner et al. (2008) focused on full waveform data metrics, such as echo pulse width, backscattering cross-sectional area, amplitude and range. These measurements were then analyzed and filtered to help better understand the characteristics of full waveform data specific to vegetated areas (Wagner et al., 2008).

The LiDAR collection system used in this experiment was a Riegl LMS-Q560, which was flown at 500 m above ground level and calibrated with a scan rate of 66 kHz (Wagner et al., 2008). The surveying area consisted of the Schonbrunn Palace Park in Vienna, Austria, which was divided into two areas. The first area was the French garden, which is described as a baroque garden, and the second area was the English Garden, which closely resembles a naturally forested area. In terms of vegetation, the French Garden consisted mostly of Linden, Maple, and Sycamore trees, while the English garden consisted of five types of trees, including Oak, Ash, Maple, Linden, Cherry, and Hornbeam (Wagner et al., 2008).

The Wagner et al. 2008 study found that the backscattering pulse characteristic of terrain echoes are usually larger than canopy echoes. Also, in terms of using the scattering data, the study was able to classify between non-vegetation and vegetation with an

accuracy of 89.9% in a densely forested area, with even better results in the baroque garden, at 93.7% accuracy. This accuracy was obtained by using the four full waveform parameters, a color infrared image, and a shaded DSM over a small portion of the French Garden. The echo width parameter was used to accurately distinguish vegetation from smooth surfaces, like manmade objects, while the pulse width data were used to help filter out even the lower-level vegetation. The study also discovered that the cross-section of the vegetation echoes are generally lower than terrain echoes. On average, the terrain produced cross-sectional value of .18m^2 and grass with a .098m^2 value. Another interesting finding was how much the total cross-sectional area varied when it scanned various types of trees, which is depicted in Figure 12, and could later be used for tree classification metrics (Wagner et al., 2008). Of note, in Figure 12 product "f" shows the cross sectional area metric used in the study where the classification improved significantly.



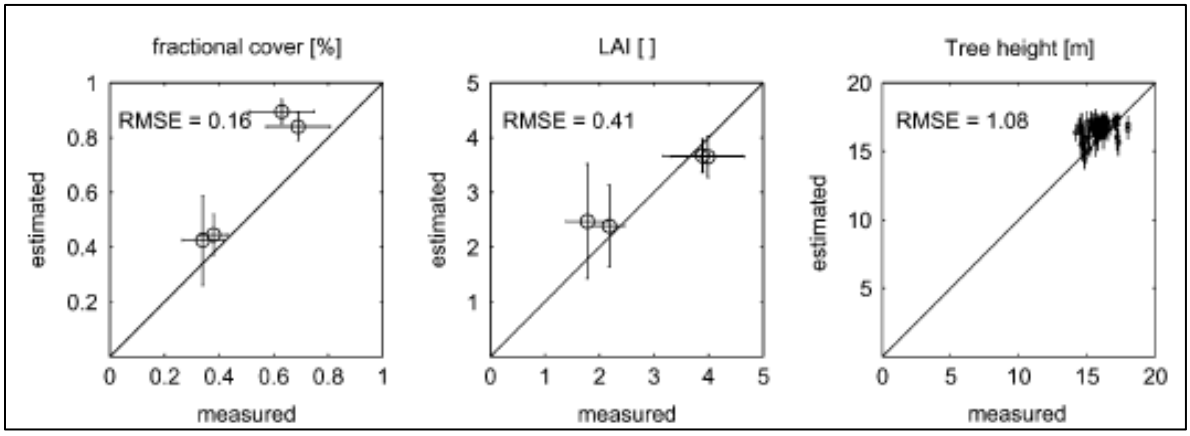
Six products which were used to compare the experimental data (a) color-IR image (b) DSM (c) uses range metrics to differentiate targets (d) uses returned echo amplitude in digital numbers (e) uses the returned echo width (f) uses the return echoes cross-section in m^2.

Figure 12. Results from the waveform data using various metrics.
Source: Wagner et al. (2008).

The study further filtered all the last echoes with widths larger than 1.9 ns and total cross sectional area less than .08m^2 to then automatically classify those targets as vegetation echoes. This non-complex approach allowed the researchers to identify bushes,

hedges and low-lying arrangements. This additional filtering produced results which accurately identified 93.7% of targets while using a k value of .86 in the French Garden (Wagner et al., 2008).

Modeling work has been done to investigate waveform characteristics over tree canopies. Koetz et al. (2006) applied a Radiative Transfer Method (RTM) to the problem of extracting biophysical parameters from large-footprint systems like LVIS. The theory was that a 3D RTM can represent the canopy structure by including LiDAR pulse interactions. Using a synthetic dataset, the study's RTM demonstrated that estimating the horizontal and vertical forest structure is possible, including estimations of the fractional cover, maximum tree height, and vertical extension of the crown layer. Unfortunately, the Leaf Area Index (LAI) and fractional cover were not as accurate in the study area because model assumptions and data processing proved to complicate estimations (Koetz et al., 2006). Figure 13 displays three metrics, fractional cover, LAI, and tree height, and compares the estimated values to the measured values. Of note, the tree height estimations and measurements were the most promising, with an RMSE value of 1.06.



Circles symbolize the median solution and bars are the associated uncertainties values of model.

Figure 13. Swiss National Park study results. Source: Koetz et al. (2006).

There has been a modest amount of work in the applications of small-footprint waveform systems to foliage analysis. Fieber et al. (2013) used airborne data from Riegl

LMS-Q560 scanner to determine a classification capability between orange trees, grass and the ground. Fieber, et al. (2013) used the backscattering cross-sectional area, its coefficient, and pulse width to determine various relationships for classification. In the case of a single peak return, the study compared backscattering cross-sectional areas against pulse width to differentiate between grass, ground and orange trees. If the data showed multiple returns, then the first and last return cross-sectional values were used to separate targets. The hardest classification type to distinguish was grass, because its reflectance value falls between that of the orange tree and the ground. After discovering this problem, the study classified targets as two types, one being the grass or ground and the other orange trees, and yielded an accuracy of 95%, which Table 2 displays (Fieber et al., 2013). The table displays how the ground and grass class were grouped, but proved to identify those returns with 94% accuracy.

Table 2. Results from the study that distinguished the ground and orange trees with 94.8% accuracy. Adapted from Fieber et al. (2013).

Gamma/width classification of all waveforms (single and last)				
Class	**Orange Tree**	**Grass/ Ground**	**Total**	**Producer's accuracy**
Orange trees	12851	586	13437	95.60%
Grass/ ground	1301	21749	23050	94.40%
Total	14152	22335	36487	Average 95%
User's accuracy	90.80%	97.40%	94.10 %	Overall Accuracy 94.8%

Larger vegetation species, such as orange trees and oaks, have continuously been studied because they are easier to determine height metrics. However, Zlinsky et al. (2014) conducted a study to analyze lower vegetation targets again using a small-footprint LiDAR system, the Riegl LMS-Q680. The study focused on various grasslands, including lowland hay meadows, which have much lower vegetation species and are widespread throughout Europe. Using random forest machine learning techniques, objects were classified into various groups. The study had two sets of classification guides, the first separating the

objects in 10 vegetation types and the second into five meadow land vegetation types (Zlinsky et al., 2014). The classification guide with five classes was able to accurately group 75% of the targets, while the classification guide with 10 classes reached an overall accuracy of 68%. Table 3 displays the study's results, and of significance the classification for scrubs was an astonishingly 96.7%. The highlighted values are the number of correct classifications for this type, while the un-highlighted values are incorrect categorizations of the vegetation (Zlinsky et al., 2014)

Table 3. Results of the experiment along with the various categories tested. Source: Zlinsky et al. (2014).

Table 3. Confusion matrix of 10 vegetation classes and accuracies

	Not vegetation	Shrub	Fringe	Abandoned	Meadow like	Lowland hay meadow	Wet high	Molinia	Dry meadow	Lawn	Total	User's Acc. (%)
Not vegetation	4106	47		25		793			203		5174	79.4
Shrub	2	2108	55	16							2181	96.7
Fringe	28	256	688	593	594	140	270	1			25702	26.8
Abandoned		48	105	2865	370	559	199	40	34		4220	67.9
Meadow like	142	1	59	379	2695	1201	107	72	267	2	4925	54.7
Lowland hay meadow	77		6	78	619	2271	19	579	586		4235	53.6
Wet high		2	11	150	228	7	2906	378	55		3737	77.8
Molinia			16	103	42	284	37	1823			2305	79.1
Dry meadow	1	78		437	612	833	90	65	4200	5	6321	66.4
Lawn	308				44	83				2750	3185	86.3
Total	4664	2540	940	4621	5997	5403	3628	2958	5345	2757		
Producer's Acc. (%)	88.0	83.0	73.2	62.0	44.9	42.0	80.1	61.6	78.6	99.7		
Cohen's Kappa	0.64										Overall Acc. (%)	68.0

THIS PAGE INTENTIONALLY LEFT BLANK

III. DATA SET AND PREPARATIONS

The previous chapters provided a background on LiDAR basics and previous studies focusing on waveform data. This chapter discusses the instruments, software, and data this thesis uses to analyze waveform data. Section A focuses on the two airborne LiDAR collection platforms while Section B discusses the software, data, and initial point cloud creation method.

A. INSTRUMENTS

This section will discuss the instruments and ground observation data this thesis needed in order to analyze discrete and full waveform data. Additionally, this section will explain how these two LiDAR systems are dissimilar.

1. Optech Titan Multispectral LiDAR

The Optech Titan multispectral LiDAR is one of two aerial collection platform that this thesis uses. The Optech Titan is unique because it has three different laser channels. The first channel is operating at 1550 nm, which covers the intermediate shortwave infrared (SWIR) range ("Optech titan," 2015). The system's second channel operates at 1064 nm, which covers the near infrared (NIR) spectrum. Finally, the third channel operates at 532 nm, which is in the visible spectrum (green). Figure 14 displays each channel, identified by the numbers at the top of the graphic. This figure shows where the channels are located on the electromagnetic (EM) spectrum and an expected reflectance range for vegetation, soil, and water within each region. We used these predefined reflectance characteristics to organize and recognized outlier returns.

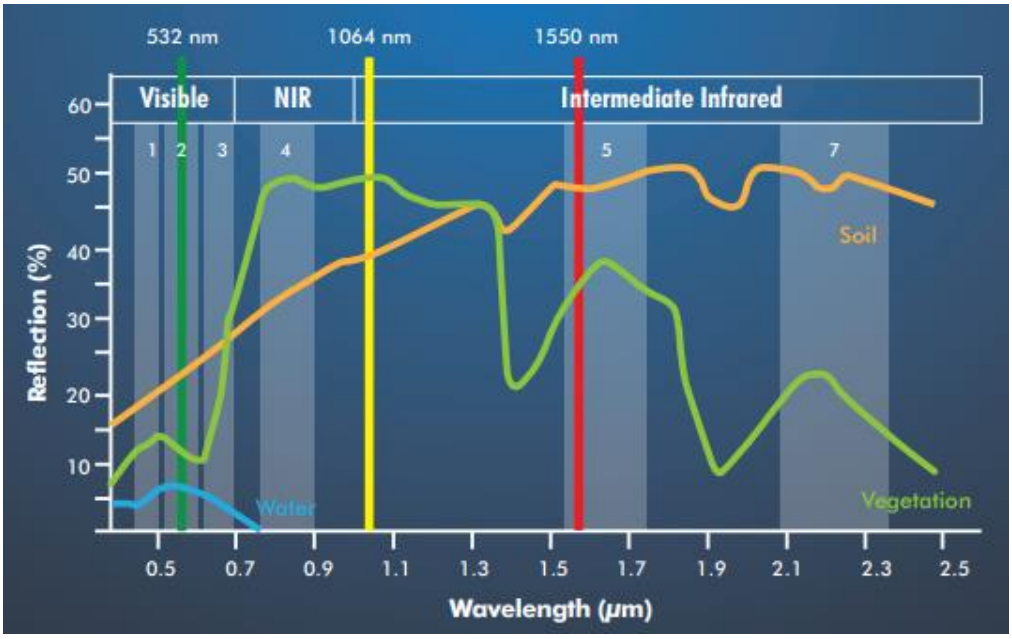


Figure 14. Included are three channels associated with the Optech Titan LiDAR system. Source: "Optech Titan" (2015).

2. AHAB—Airborne Hydrography AB LiDAR

The second LiDAR collection system this thesis uses is the Airborne Hydrography AB (AHAB) LiDAR system. The AHAB Chiroptera II scanner uses two channels, one bathymetric, 532nm, and one topographic channel, 1064 nm (Quadros, 2013). This topographic scanner is easily programmable and includes a 500 KHz receiver which is useful for land topology studies. The significance of this LiDAR system is the unique scanning pattern uses a dual head oblique scanner to replicate a palmer scanning pattern, which is displayed in Figure 15. Additionally, the onboard camera can be utilized for nearshore charting ("Leica chiroptera," 2018).

This scanning pattern helps improve coverage of tall objects by providing two separate angles for data comparisons.

Figure 15. Representation of the scan pattern used on the AHAB LiDAR. Source: "AHAB Bathymetric" (2014).

These multiple wavelength LiDAR systems are used for all types of LiDAR scanning missions such as bathymetry readings, vegetation studies, and 3D land classifications. The aim of this thesis is to use full waveform data to improve vegetation classification.

3. Ground Survey Data

To further validate data collected by the aerial LiDAR systems, we performed a ground survey of the vegetation and terrain. A ground survey is used to help identify and locate specific tree species within the collection area. We also needed a ground survey because without it there would not be any ground observation data to reference the LiDAR data with. The metrics included in the ground survey uses tree species type, height, diameter at breast height (DBH), and a Global Positioning System location. Ground

surveys are a common practice among LiDAR studies, but, since the ground survey covered a smaller area than the LiDAR collection area, this thesis focuses primarily on the ground surveyed areas.

B. DATA COLLECTION, VIEWING, AND PREPARATIONS

In 2016, LiDAR flights collected full waveform LiDAR data over Point Lobos State Park, which is located south of the Monterey Peninsula, CA. Point Lobos is a California State Park, which is designated as a natural and marine reserve containing over 550 acres of wilderness terrain. The park has over 300 plants, specifically highlighted in the ground observations were the Monterey Cypress, Monterey Pine, and Coast Live Oak trees. These species are densely populated throughout the Monterey peninsula and are a respectable group of species for the study's focus. Additionally, the total area collected by the systems included the surrounding areas outside of the park, to bring the total collected area to approximately 4.5 km^2 ("WSI applied remote sensing," 2013). The data, after initially viewing and processing, exhibited multiple high-resolution areas within the footprint of both sensors, where in some cases over 20 pulses per square meter were available. Figure 16 displays the coverage map which includes a highlighted collection area.

Additionally, processing was another key tool we used during the preparation phase. A variety of software tools were useful to work with the data, to include LAStools (Isenburg, 2018), Environment for Visualizing Images (ENVI) (Harris Geospatial Solutions, 2009), Quick Terrain Modeler ("QTM," 2018), and ENVI LiDAR (ENVI

LiDAR, 2018). By combining these software tools, we were able to visualize and analyze both sets of data.

Figure 16. Point Lobos, shaded region represents the collection area. Source: "WSI Applied Remote" (2013).

1. Initial Viewing and Filtering

After receiving the raw LiDAR data, we first needed to assess the data to ensure that no collection gaps or significant errors were present. We used this process of initially viewing the data because this is a quality control procedure to visualize the raw data in an unclassified point cloud. One of the many programs this thesis uses is Quick Terrain Modeler (QTM), which allows users to immediately view the data as long as it is in an acceptable format. This software has many tools associated with it to include the ability to view the return data's intensity, height, time, number of returns, and return number. Normally, LiDAR data is delivered in flight lines or tiles, depending on the specifics detailed in the LiDAR collection request. In this case, Optech Titan data were separated into 18 tiles varying in size while the AHAB data were separated into 21 flight lines, both representing similar coverage areas. Additionally, the AHAB collected data consisted of 42 individual flight lines of data because the channels were not merged together. Figure 17

shows the discrete data (point cloud) in a perspective view using a gray scale encoding of return intensities.

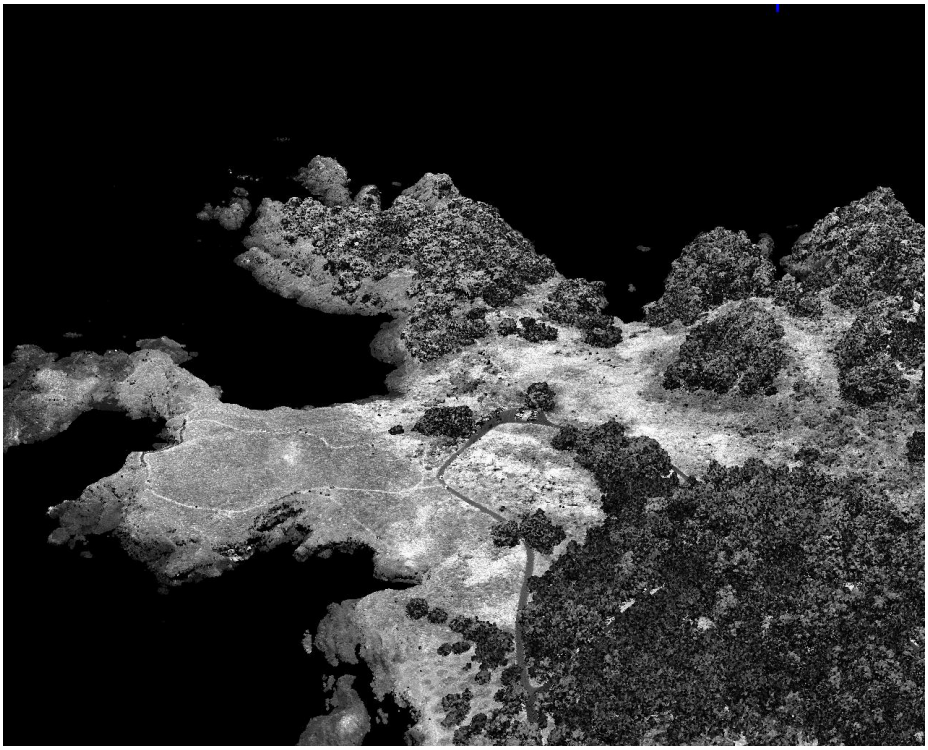


Figure 17. Unclassified AHAB data displaying returns based on intensity.

Figure 18 displays an unclassified Optech Titan discrete point cloud for tiles 593_4042 and 594_4042 (local designation), where the returns are separated by their return number. The white colored returns are mostly first returns or ground points, while the red, blue, and teal colored points represent multiple returns over areas such as vegetation. Figure 19 displays some flight lines where AHAB data were collected. This graphic uses a different filter that separates returns by time to determine if the overlap data was adequate.

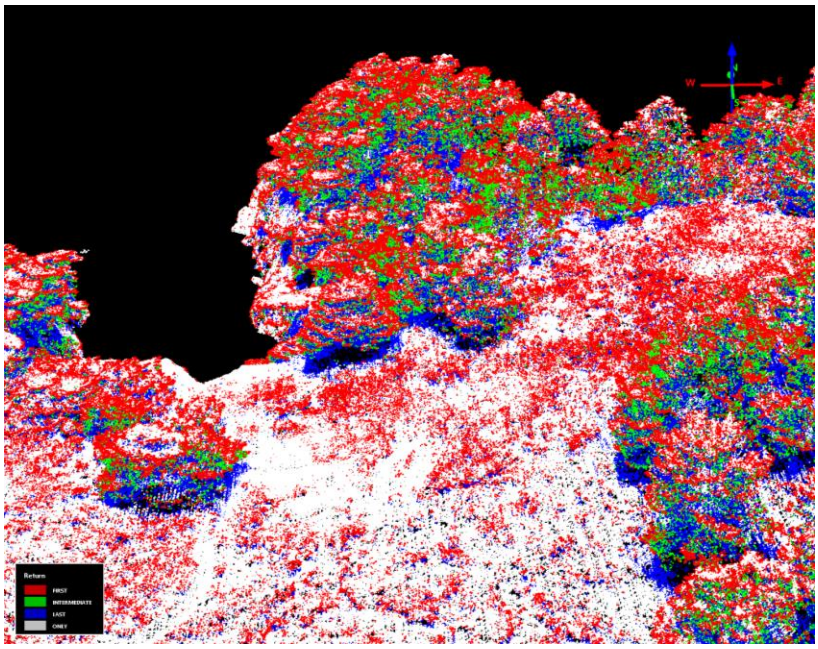
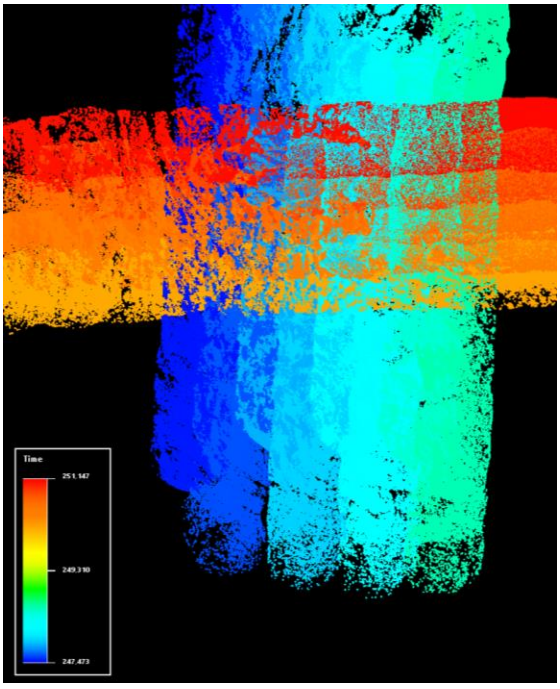


Figure 18. Unclassified raw data collected by
Optech Titan with a filter coloring various returns.



The overlapping coverage this data provided resulted in various return angles and increase point density.

Figure 19. AHAB Flight line data showing the
North-South and East-West flight lines.

After viewing the raw data, we first removed excessive data, such as water returns, since this thesis focuses on vegetation. Water returns normally consist of solo returns, but since there are thousands of points over the terrain with a single return, the typical methods we use could not filter out first returns without losing some significant data. Instead, we located the elevation for the water level data, to confirm water returns were no longer influencing the data. This filtering method allowed us to retain the most important data, the terrain and trees, while some of the underground terrain features at and below the water level were lost. Additionally, we removed the points well above the tree line during this stage since they are identifiable noise points.

2. Data Sizing

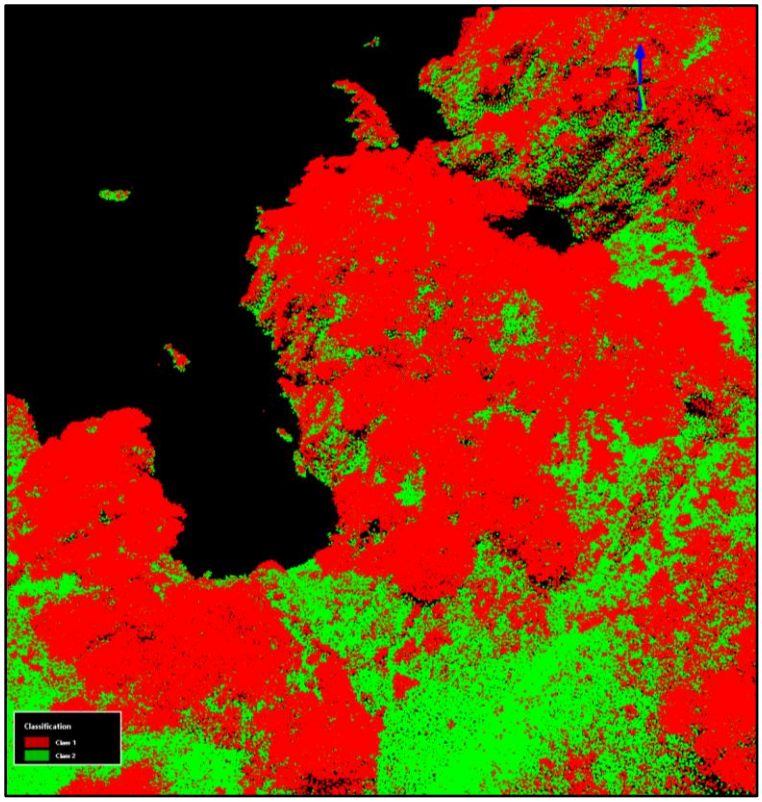
Another step after initially filtering out the unwanted data was to separate the tiles and flight lines into manageable segments. This includes using buffers to help increase the ground classification algorithms. We then ran the data through LAStile, where returns were separated into 500m x 500m tiles. This technique significantly reduces the size of various files, especially since some have over 40 million points in a single tile. Additionally, we used buffers to increase ground estimation accuracy, since not every tile will have a sufficient amount of ground points to reference.

3. Filtering with Algorithms

We then resized the unclassified data into usable tiles, where the focus is to filter out isolated noise points. Using LAStools, a software development by Dr. Martin Isenberg, we were able to run LASnoise to efficiently filter additional noise data. The algorithm Dr. Isenberg uses to make this program function looks for isolated points within a specified step size to create a block cell, which will compare itself to its surrounding data points (Isenburg, 2017). This step in the study allowed us to identify points that do not have enough surrounding points to validate a specific return. Next, the software classifies these uncorrelated points as code 7, which is defined as low or high noise. We then used various viewers to verify these points and make a determination to retain or delete those points.

At this point we identified the noise points, so the next step is to classify some of the returns. Before we can identify vegetation points, we needed to identify the ground

returns. In order to do this, we used another script called LASground, which identifies the majority of the ground points using another triangulation algorithm. The tool is designed to work well in mountainous and rugged terrains with large amounts of vegetation. The software uses a set bin size, which defaults to a 5m by 5m grid, to calculate the ground height by comparing the other ground points within that area to estimate the ground height within that grid (Isenburg, 2018). In order to increase the accuracy, LAStools allows users to set the grid to smaller sizes, but this tradeoff will increase the time to compute the data. Figure 20 displays a point cloud product after running LASground using the raw AHAB data. Class two objects are green, representing ground classified returns, while class one returns represent unassigned targets.



Red areas represent unclassified returns, while green areas represent ground returns.

Figure 20. AHAB point cloud data classified using LASground.

LASground allowed us to analyze the ground data and receiving a moderately accurate representation of the terrain. The next script, we ran was LASheight which identifies various objects by height. We determined the proper height classifications based on the ground observed data. The vegetation areas we are targeting are all above 2 meters, so, we filtered out any vegetation such as shrubs or low lying vegetation by classifying them as class 3. The algorithm that this program uses in LASheight is similar: it only analyzes the points above the ground, which we previously identified in order to produce a Triangulated Irregular Network (TIN). This method analyzes single points, where every point has a stored elevation value, and, the script uses these values to determine the Very Important Points (VIP) (Marcoe, 2007). This thesis, at first, aimed to focus on high elevation points and classifies them, if they fell into the defined range. The range that we selected was based on the ground observations with some flexibility. Using this, high vegetation was classified from 2 to 50 meters. Figure 21 displays a common LiDAR example to classify certain vegetation; additionally, the figure includes a legend, which is similar to the one this thesis uses.

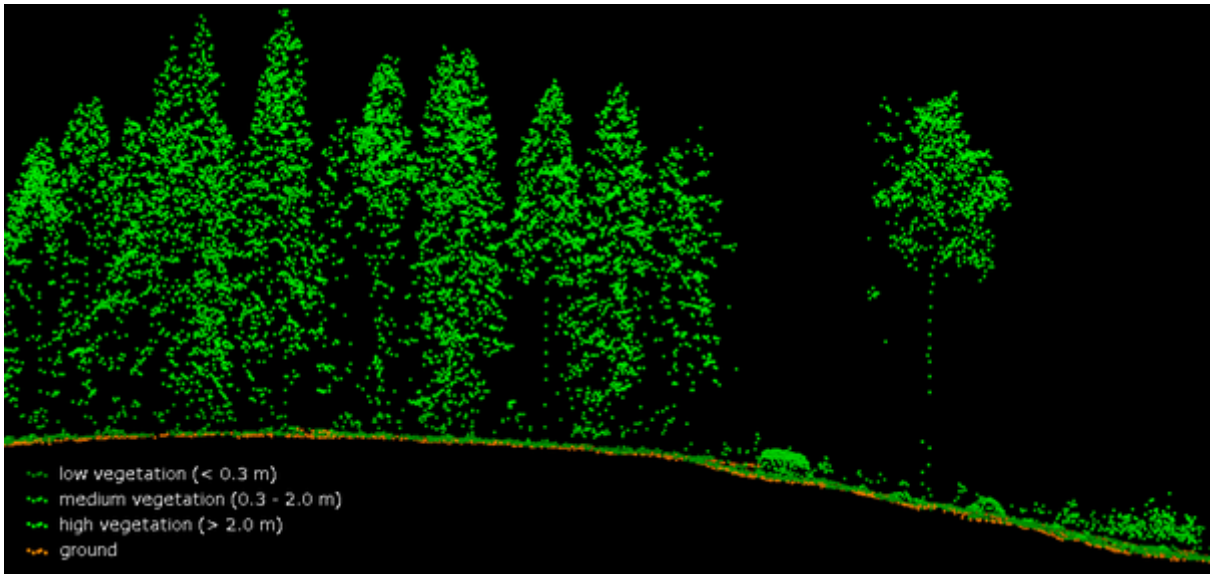
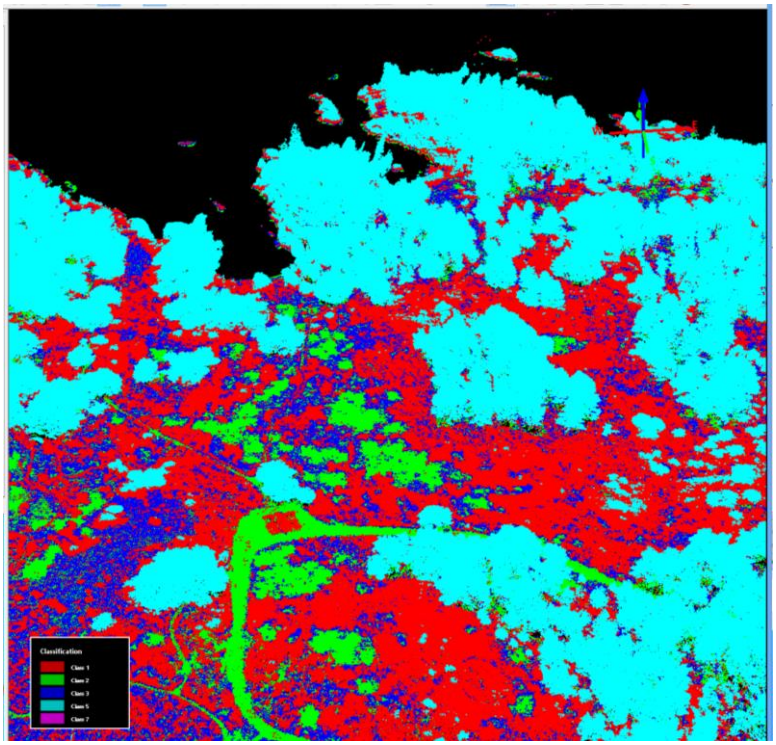


Figure 21. Example point cloud classification used to identify vegetation types by height. Source: “Solutions: Forestry” (2018).

After, we determined the height classifications using LAStools, we were then able to view and verify that the tool ran correctly in QTM. QTM allowed us to manually filter out any data without classifications if we deemed it necessary. Table 4 displays the various classification codes this study used throughout the filtering process for discrete data. These codes are standards among the LiDAR community and have been widely accepted. Figure 22 is a classified point cloud for the AHAB data after running various LAStools scripts, which were previously discussed, the figure displays classifications for various returns, including high and low vegetation targets.

Table 4. List of classification codes used in LASground and QTM.

Classification Code	Classification Type	Assigned Color
1	Unclassified	Red
2	Ground	Green
3	Low Vegetation	Blue
5	High Vegetation	Teal
7	Noise	Purple



The high vegetation is marked with light blue, while the road ground is green.

Figure 22. Point cloud classification after running LASground and LASheight.

C. WAVEFORM VIEWING AND ANALYSIS

Once the point cloud was fully functional and classified using standard codes, then waveform analysis could commence. This thesis uses PulseWaves and ENVI to view the pulses in order to determine if those returns can be classified accurately. By using the discrete point cloud data, we were able to classify the data into various categories. Next, we needed to determine the best way to compare the two sets of data, thanks to Dr. Olsen, we were able to export the waveform file into ENVI for analysis. Additionally, another waveform viewing software, known as PulseWaves, helped us verify that the waveform reader was processing the data accurately. These readers allowed us to visualize the filtered data at a granular level. Figure 23 is an example of a PulseWaves product. The bottom portion of Figure 23 is similar to what most spectrum analyzers display, while the top portion is representing the pulse path in 3D.

In summary, discrete data currently allows users to receive only location information and intensity values. This thesis will use the discrete information to create a classified point cloud. Once the point cloud was classified then complete that data, along with the waveform data, could be compared to explore terrain and vegetation classification capabilities.

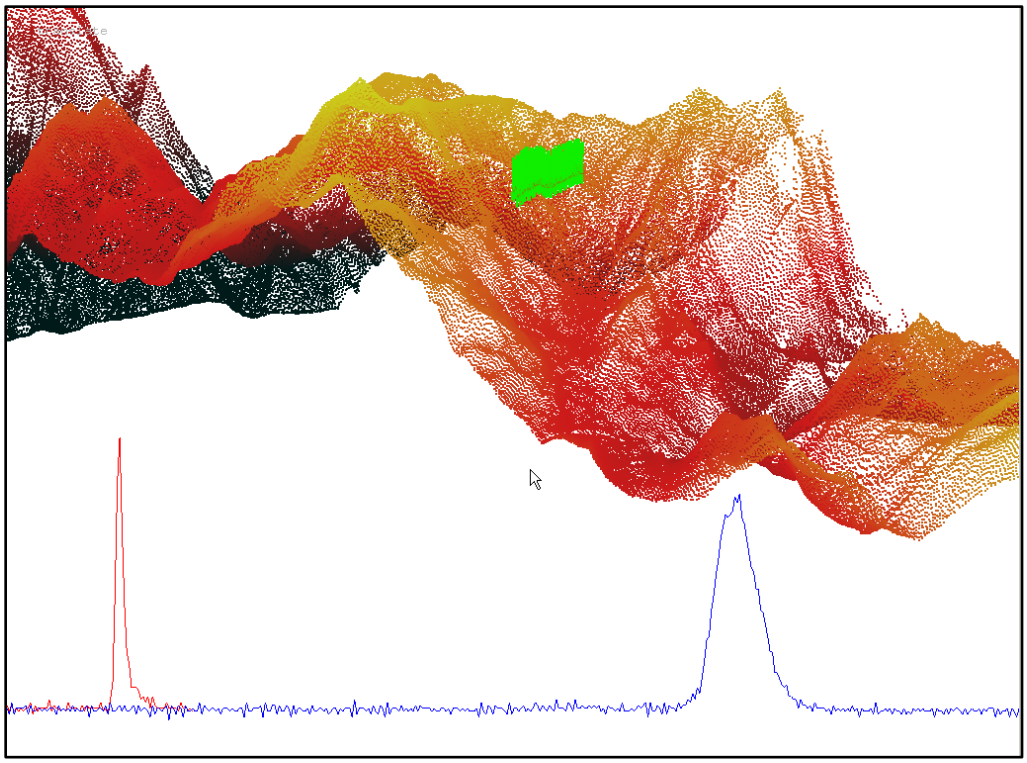


Figure 23. PulseWaves software displaying waveform returns.
Source: Isenberg (2012).

THIS PAGE INTENTIONALLY LEFT BLANK

IV. ANALYSIS AND RESULTS

Chapter III presented the data, instruments, and tools this thesis uses to prepare and examine LiDAR data. This chapter addresses the analytical approach and results. Additionally, this section details the methods this thesis used to classify the scenes with machine learning techniques while approaching the analysis portion with hyperspectral data methods. Figure 24 displays the workflow diagram this thesis uses to analyze LiDAR data.

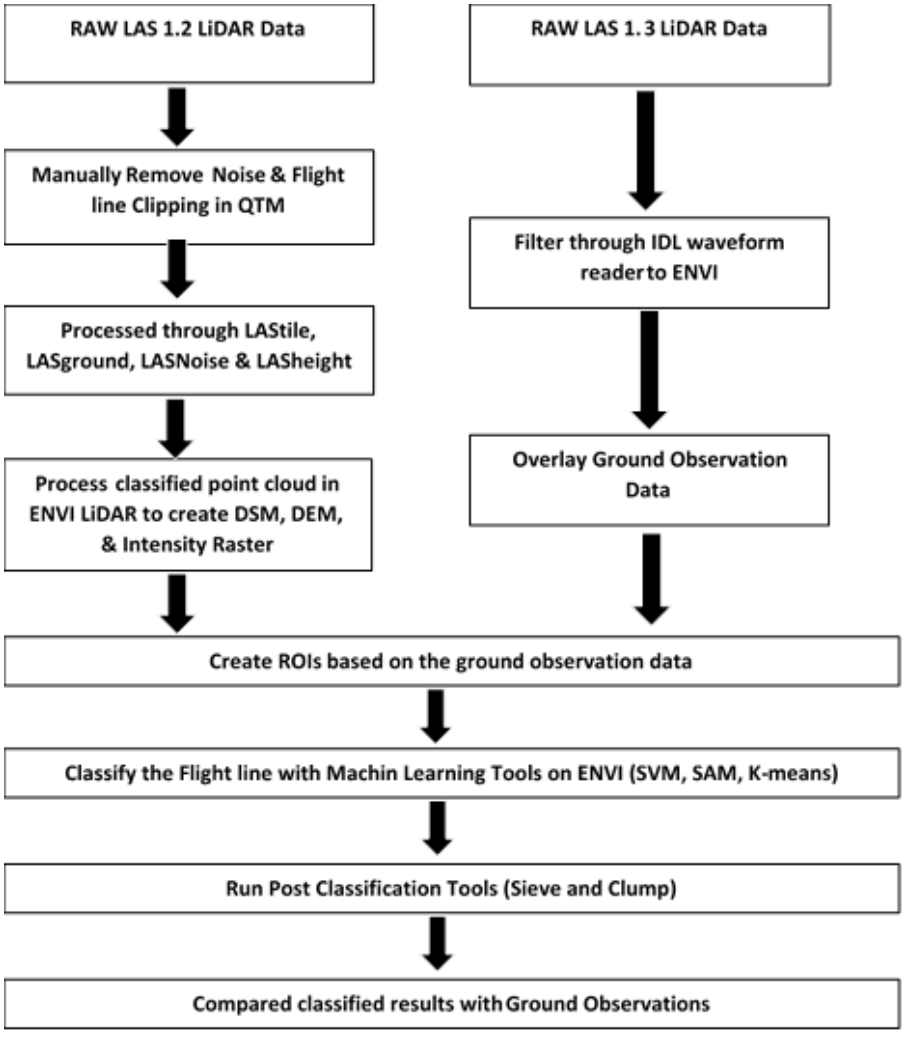


Figure 24. Analysis workflow diagram for discrete and full waveform LiDAR classification.

A. FORMAT AND READER

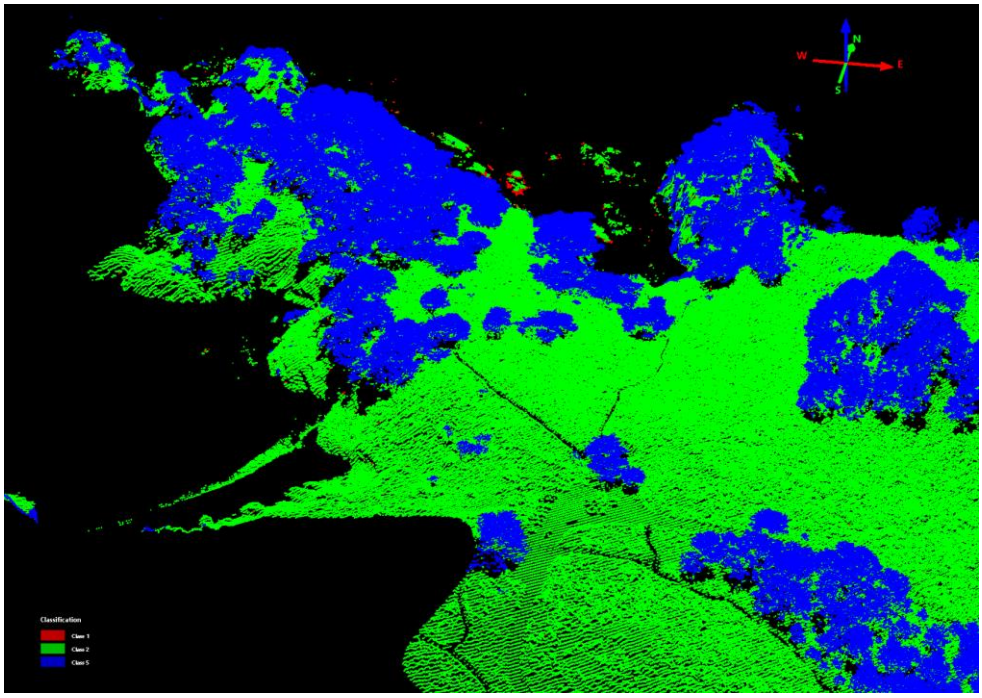
This thesis first created a classified point cloud with discrete data. The format in which the data were stored is known as LAS 1.2. These LAS 1.2 files only have the capability of providing discrete data, unlike LAS 1.3 format, which has the same fields as an LAS 1.2 file but also includes waveform data. Before we could evaluate the discrete and waveform data, we would need to find a way to analyze each data set consistently. As previously stated, waveform data formats are difficult to work with because the majority of LiDAR studies focus on discrete data. A small portion of researchers do conduct full waveform studies, but they normally create their own proprietary waveform readers. In order to read and analyze the waveform data, this thesis uses a waveform reader created by Dr. Richard Olsen. Through the use of an Integrated Data Language (IDL) script that Dr. Olsen created, we were able to convert the file into a readable data format. After running the files through this script, we were able to covert the data file onto ENVI for analysis.

B. ANALYSIS

This section first discusses the discrete LiDAR classification preparations and methods, we used to classify the flight lines. The following section discusses waveform data classification methods.

1. Classifying

After filtering the discrete data with various LAStools scripts, we then displayed the results using QTM to view the product and ensure the data were acceptable. QTM can display various metrics, but, in this case, we wanted to ensure that we properly identified our ground and high vegetation points. Figure 25 displays flight line 29 where the point cloud visibly classifies the ground and high vegetation returns.



Blue colored returns represent high vegetation, green returns as ground, and red as unassigned.

Figure 25. Point Cloud data for flight line 29.

2. Formatting

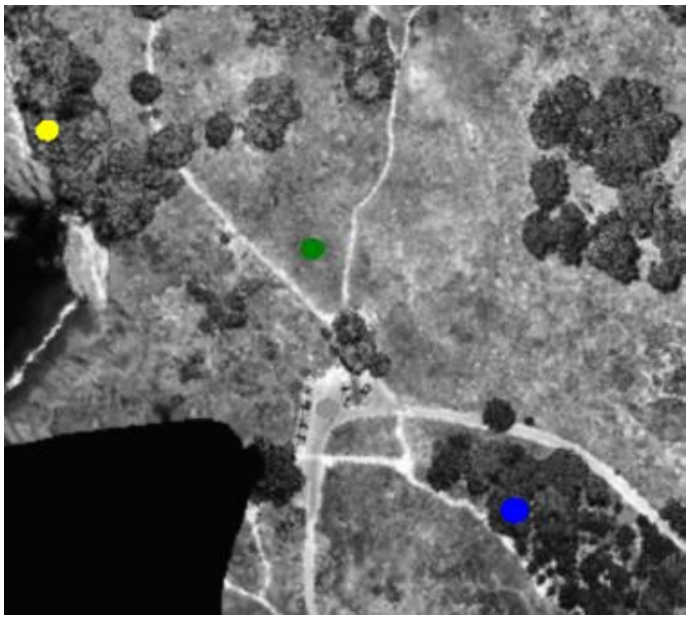
Next, we converted the LAS 1.2 files into various gridded products ENVI can read. The first product we exported from the discrete data was a Digital Surface Model (DSM), which displays the identified ground and extruding features, in this case, tree returns. We then converted the product from a vector image to a raster image for ENVI analysis. Next, we created a Digital Terrain Model (DTM), which uses all the ground returns to create a product that displays the bare earth surface in a gridded format ("Introduction to Light," 2018). Next, we used these two products, the DSM and DTM, to create a normalized Digital Surface Model (nDSM). An nDSM was valuable for analysis because it allowed us to retain the height information for all the objects above the ground in ENVI.

After exporting the nDSM into ENVI, we converted the LAS 1.2 file into a data file, which represent the various elevations for each return. Additionally, we created an intensity band using values exported from the point cloud to signify return intensities. We then imported this file into ENVI and overlaid it with the nDSM file. Overlaying these

models was an important step because the size of the two models needed to be correct, as well as the location accuracy within the pixels. After we merged these two files to build a combined file, the discrete data had all the significant information from the point cloud onto ENVI for analysis. Additionally, ENVI also allowed us to overlay the ground observation survey, which has the tree species and location data associated with those additional points.

3. Regions of Interest

The first part of the analysis uses basic regions of interest (ROI) to locate the targeted tree species and various terrain elements such as low vegetation and roads. In order to correctly label these regions of interest in ENVI, we used the ground observations and imagery to confirm their locations. The regions of interest consisted of a small sample set first to run basic statistics over these areas for identifiable information. Figure 26 displays the four regions of interest overlaid on the data file.



Each ROI associates with a target type: Pine (Blue), Cypress (Yellow), Road/Trail (Gray), and Low vegetation (Green).

Figure 26. Intensity data file displaying the four specified ROIs.

Initially, the statistics displayed some variations between the ROIs over the bands used. The cypress and pine trees had similar values, but there were some differences with their mean return intensity values. The intensity, which could later be used as a classification metric, was much larger for the cypress ROI than that of the pine tree ROI. Additionally, the "Road/Trail" ROI did not show a noticeable variation between the "low vegetation" ROI.

4. Discrete Classification

After initially verifying the different ROIs and running basic statistics, we then began classifying the flight line with ENVI. A variety of classification algorithms were applied to the gridded waveform data. The supervised algorithms included Support Vector Machines (SVM), Spectral Angle Mapper (SAM), and Maximum Likelihood. The unsupervised algorithm we used was K-means.

a. Unsupervised

The k-means tool in ENVI was applied to the discrete data. This tool does not depend on the defined ROIs. This tool creates classes based solely on the data file (Harris Geospatial Solutions, 2009). We limited the classes down to six because, if we had not imposed these limits, the classifier could have created several unnecessary classes that would skew the results. The resulting product showed a minimal amount of classification accuracy.

b. Supervised

The classification tools this thesis uses is a supervised machine learning tool. Using a supervised approach, the algorithm takes into account the specific inputs, in this case the ROIs, to make a classification decision for the final product. The first tool we used for discrete data is a Support Vector Machine (SVM). This tool runs an algorithm that classifies data based on the input data, or training regions, used. This algorithm is a linear classifier that identifies the best location to place a hyperplane which separates the ROIs and classifies the scene (Ben-Hur & Weston, 2010). The data points near the hyperplane are known as support vectors, and the separation between these data points are known as

the margin. Figure 27 displays how a hyperplane can be used to classify two sets of data. The red circles on the bottom left corner of Figure 27 represent a different class of data from the data on the top right. The hyperplane is the middle bold line separating the two data sets, and the margin size is represented by the two outside lines, which are defined by the support vectors, which are circled.

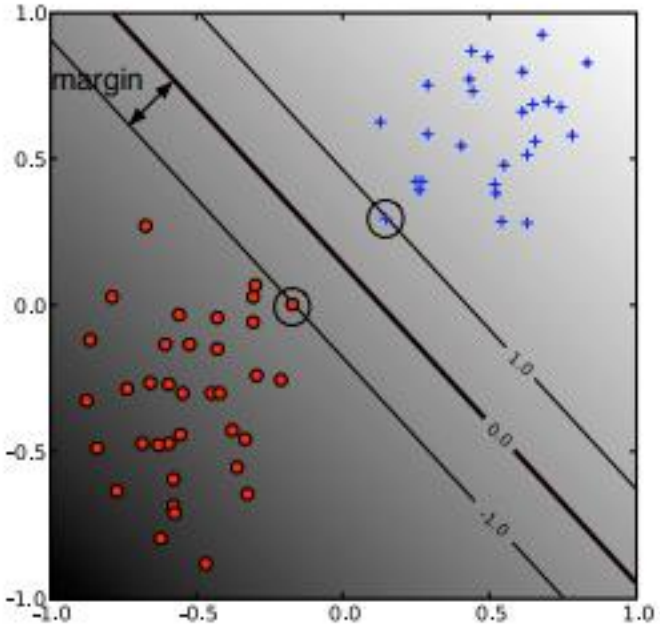


Figure 27. SVM which is classifying two sets of data, red and blue symbols. Source: Ben-Hur and Weston (2010).

Figure 28 displays the classified product after running the data through the SVM tool in ENVI. The results from this tool were unsuccessful for the goals of this study; the cypress trees were not distinguishable from the pine trees. Also in this model, the ground points were not assigned a classification because the threshold level was too high. We then lowered the threshold level to retrieve improved results, but, by doing this, we degraded the reliability of the classifier.

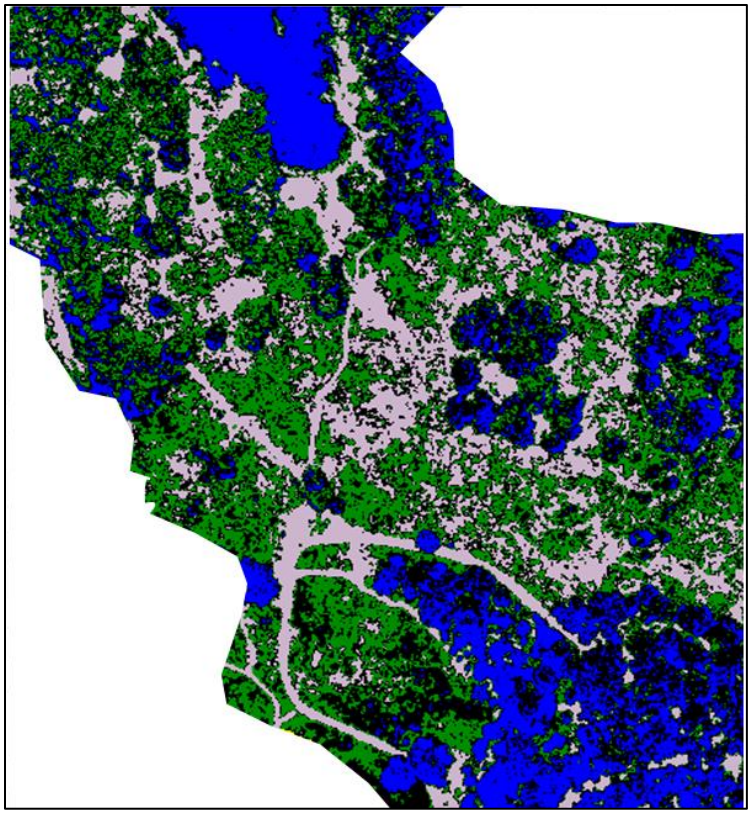
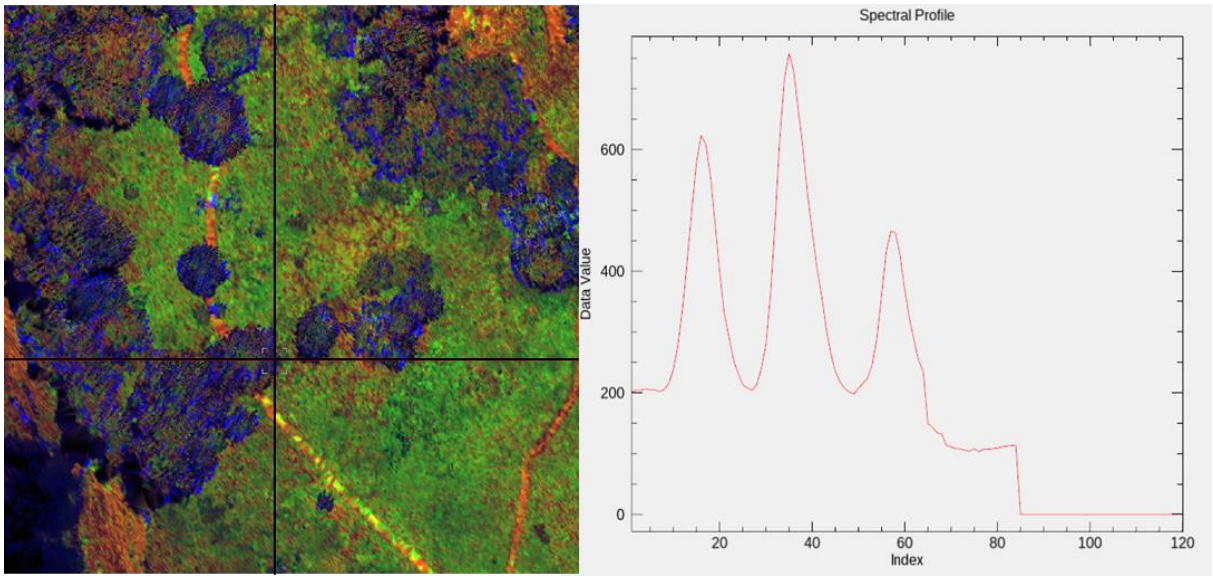


Figure 28. SVM tool used to classify discrete LiDAR data.

We also conducted a maximum likelihood classification with the discrete LiDAR data. The maximum likelihood classifier allows us to use the established ROIs as the training group with which we want to classify the data. The algorithm assumes that the data in each band shows a normal distribution pattern, which then allows it to classify each pixel based on the probability that it belongs to a certain class (Harris Geospatial Solutions, 2009). After running maximum likelihood tool multiple times, we concluded that the SVM was more accurate for the given data set. The results after using the support vector machine allowed us to accurately distinguish the difference between various objects when compared to the other classifying tools. These results showed that, although the SVM was the best general classifier, it was not successfully able to distinguish between the tree species within the flight line

5. Full Waveform Classification

We also were classifying the area using full waveform data. The associated waveform data uses 120 individual bands to represent the full distribution of each returns. Through the IDL script Dr. Olsen created, we were able to view the waveform data samples, which was narrowed down to the first 120 to make the files manageable. Figure 29 displays multiple returns over a tree's canopy. The amplitude information is available as well as the other data such as Full Width at Half Maximum (FWHM).



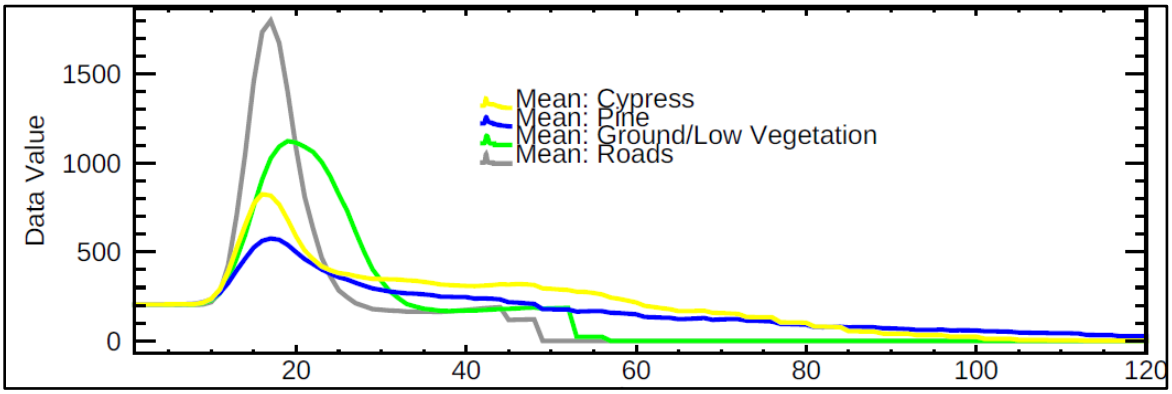
The pulses vary in intensity which is based off on their reflectivity properties. The graphic is displaying three returns over this tree.

Figure 29. A single LiDAR pulse showing multiple returns.

We first converted the waveform files onto ENVI, then we overlaid the regions of interest and processed some of the basic statistic. Figure 30 displays the basics ROI statistics where you can visibly see the mean data values associated with separate classes vary for each bands.

The spectra in Figure 30 show an abrupt transition noticeable around the 50–60 sample point for single return signals, e.g., roads and low vegetation. This is also present in Figure 36, which is displaying low vegetation returns. Additionally, in Figure 29 it is

present at sample 85. This is simply due to the truncation of the signal captured by the digitizer.



The x-axis represents the band number and y-axis are the associated return numbers for each band.

Figure 30. ROI statistic for waveform data.

a. Unsupervised Learning

Next, we classified the waveform data using the k-means unsupervised classification tool. This data proved to create a similar result to the supervised classification methods. The product was as anticipated: the pine tree and cypress trees were partially separated, as well as the ground and low vegetation. The minimum distance technique that the k-means tool uses helped assign some data points, which were left unassigned while using other classification tools (Harris Geospatial Solutions, 2009). Figure 31 displays one of the k-means classification products for flight line 29. This figure displays various classes, but it does start to distinguish vegetation in which the Pine trees on the bottom right of the graphic are colored blue and red.

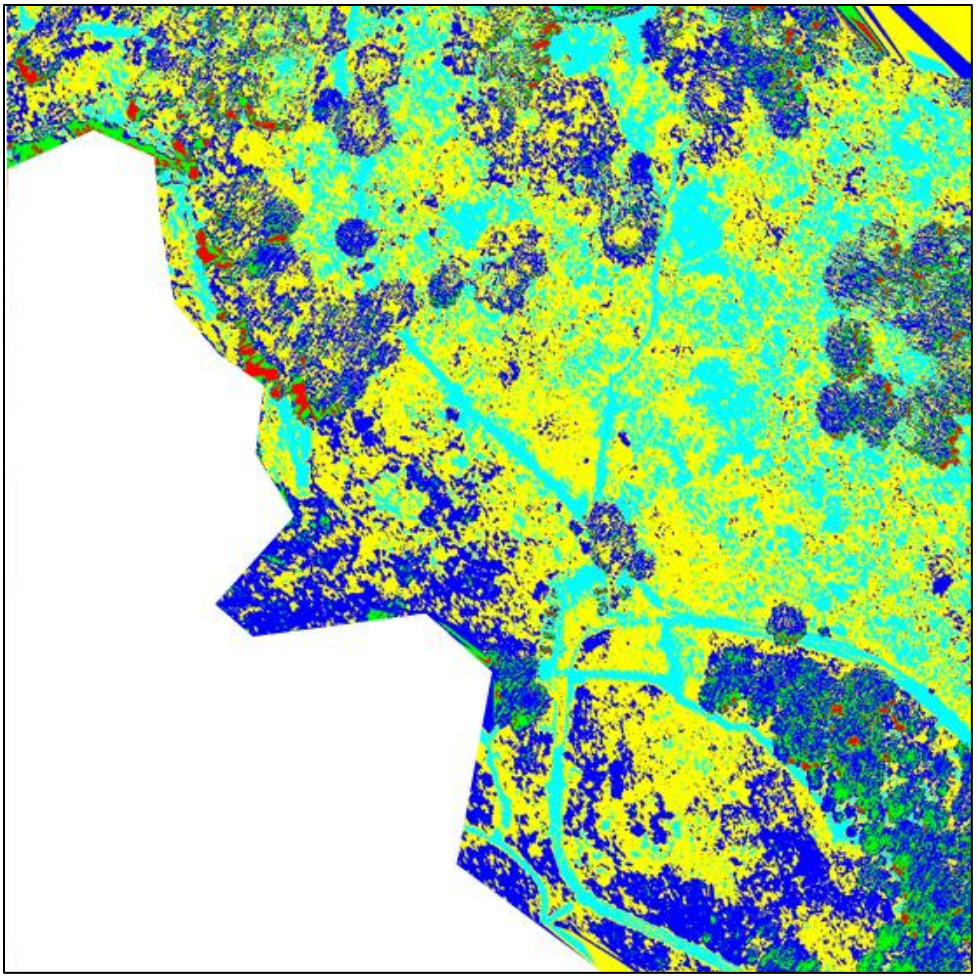


Figure 31. K-means classified product.

b. Supervised Learning

The first analysis tool we used with the waveform data was the SAM tool. This tool is used here because the rather intuitive nature of the algorithm, and its widespread use in optical imagery. This tool processed all 120 bands to classify the flight line with the specified training groups, or regions of interest. This tool allowed us to classify the flight line by using the spectral similarities between the objects. The data is grouped using an n-D angle where the angle between the training group mean vector and the pixel is compared (Harris Geospatial Solutions, 2009). If there is a large angle between two ROIs that will allow the tool to successfully separate and classify the data with more accuracy. Figure 32

displays the results from the SAM tool classifier with the waveform data. Green returns represent low vegetation, while blue and yellow returns indicate high vegetation returns.

Using the SAM, we analyzed various locations in the flight line. The first SAM classification was set to 0.5 radians, because this was the same value we used with the discrete data earlier. The initial results showed some distinguishing features over pine and cypress trees. After running multiple SAM classification with varying threshold angles, we were able to produce an enhanced product which labeled the majority of the general classifications features we identified.

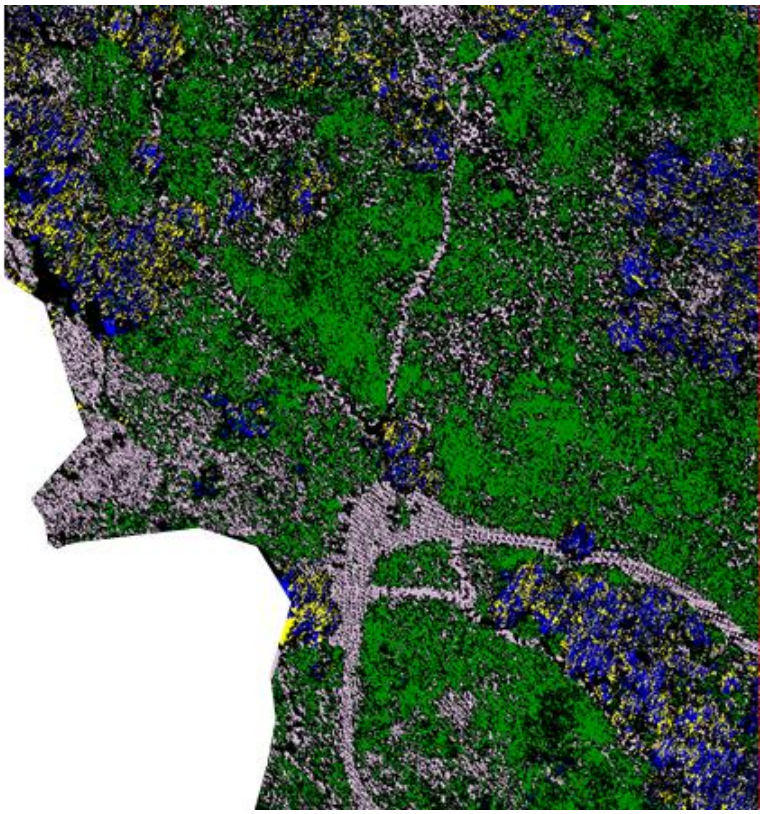


Figure 32. SAM Classification tool product using a 0.35 maximum radian threshold.

Following the use of the SAM classification tool, we then used the SVM tool. The SVM tool initially ran with a threshold level similar to the one used with the discrete file, 60 %, but, after improved results, we then began to increase the threshold level to determine

how accurately the scene could be classified. Figure 31 displays the SVM classification product after increasing the classification probability threshold to 0.7. Any data point that is less than the threshold level of 70% will be labeled as unclassified. Compare Figure 33 to Figure 28, which used discrete data, because by using this tool, we were able to successfully separate tree clusters and improve classification capabilities.

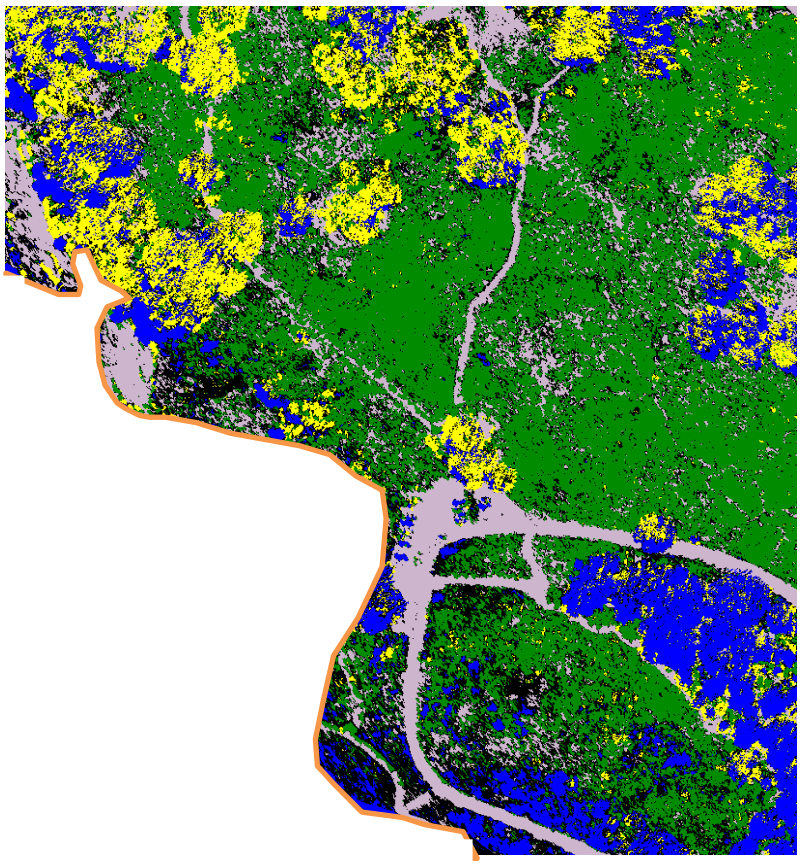


Figure 33. SVM classifying pine trees (blue),
low vegetation (green), cypress (yellow), and road/trail (gray).

After successfully distinguishing some of the known tree species groupings, we then created additional ROIs to quantify the data. There two ROIs were separated by geographical locations with known tree species based on the ground observations. In Figure 34, the mixed cypress and pine area is an equal ratio of pine and cypress tree according to estimations, while the pine area is roughly 70% pine trees. After gathering the class

statistics for the SVM classifier, we were able to obtain quantitative results. Table 5 displays all of the classes within the classifier and proved to classify 46% of the pine area as pine trees. This is significant because within the classifier there is a minimal amount of returns associated with cypress trees, which means the two trees can be discerned to a certain degree of accuracy.

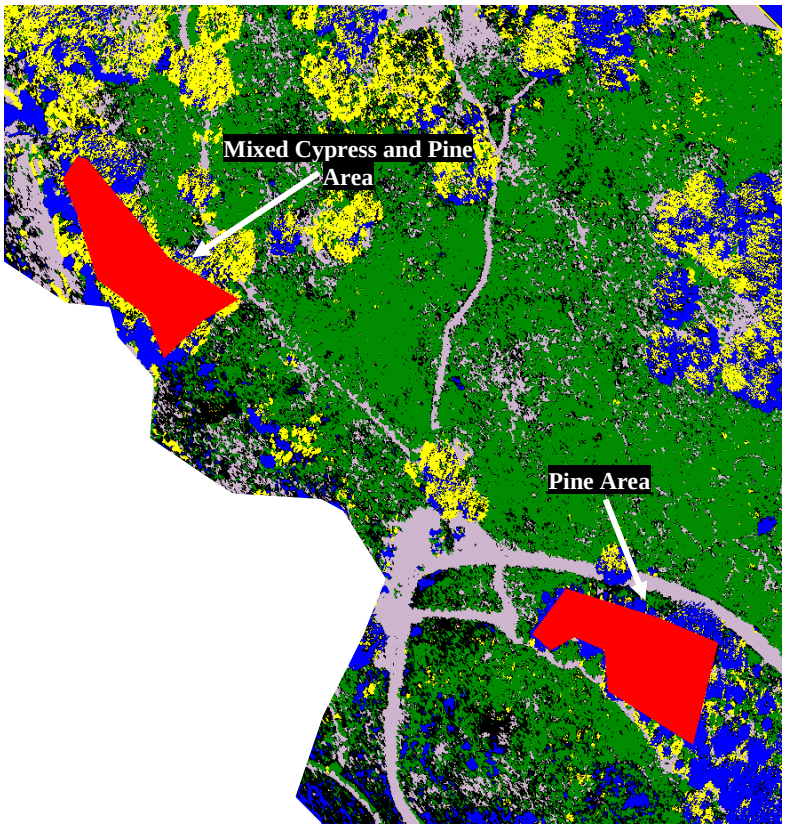


Figure 34. SVM results using new two regions of interest to compile class statistics.

Table 5. Class statistics for waveform SVM

Classes	Mixed Cypress & Pine Area	Pine Area
Unclassified	18.68	16.04
Cypress trees	32.65	6.41
Low Vegetation	18.25	24.4
Road/Trail	6.24	6.65
Pine Trees	24.15	46.56

Additionally, after comparing the two SVM products it was clear that the waveform data could also improve the classification capabilities for low vegetation. Figure 35 displays the targeted region with known low vegetation. Table 6 represents the statistic we compiled over a known low vegetation area which was conducted over both data sets. The waveform data improved the low vegetation classification by 40%.

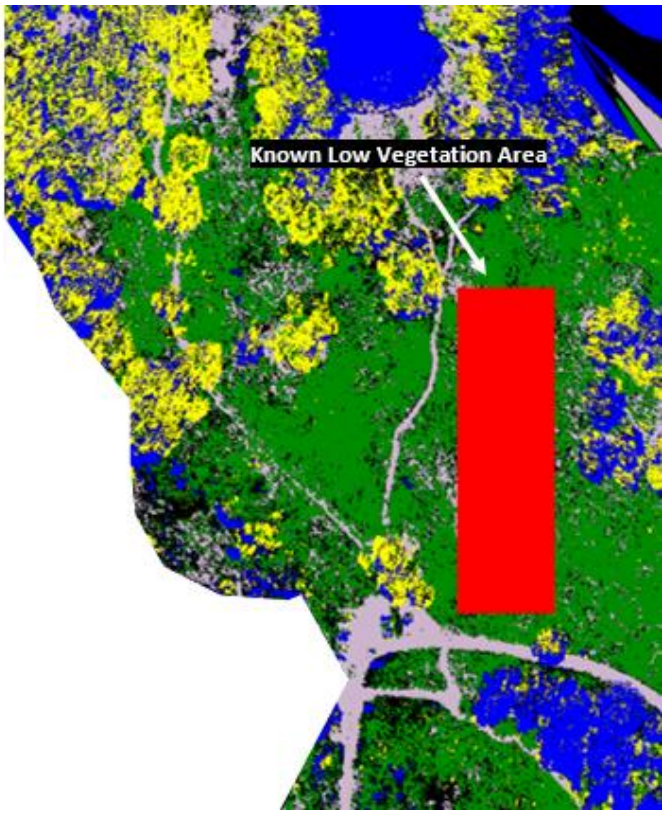


Figure 35. SVM results with new region of interest to compile low vegetation statistics

Table 6. SVM classification statistics over region consisting predominately of low-level vegetation.

Classification	Discrete (%)	Waveform (%)
Unclassified	17.1	15.3
Cypress	0.3	0.3
Low Vegetation	33	73.9
Road/Trail	49.6	10.3
Pine Tree	0	0

C. RESULTS

After utilizing various machine learning tools on ENVI to classify the flight lines, we came to a conclusion that the tree species were distinguishable with the given waveform data set. The support vector machines in both cases, for discrete and waveform data, proved to classify returns with the most accuracy. Additionally, after further comparison between the two products, the results proved that waveform data was able to increase the classification of the low-level vegetation significantly. Low vegetation returns using waveform data were much more detectable because the leading edge and trailing edge of the waveform display slight variations. The returns associated with ground points resembled spikes, while the waveform data had larger widths and in some cases two peaks. Discrete data processing would normally miss the waveform variations, and misclassify those returns as ground. Figure 36 displays two typical low vegetation returns within the data set. Both waveforms show a slight variation on the leading or trailing edges, which enabled the SVM to accurately classify this return. We ran a statistic using both SVM products for each data set over a specific region with known low level vegetation. The waveform data allowed this tool to increase the accuracy of the classification by 40 % for low vegetation.

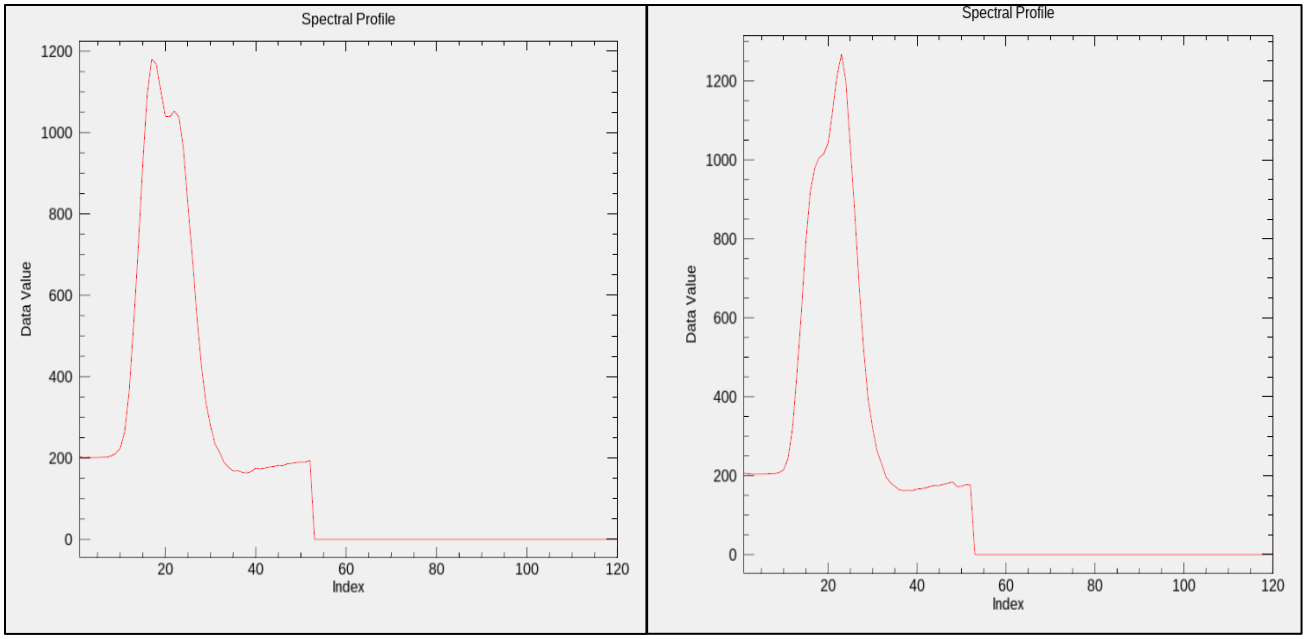


Figure 36. Waveform ground return recognized as low vegetation using the SVM tool.

V. CONCLUSION

The focus of this study was to explore the uses of full waveform data for classification studies were possible. We compared full waveform and discrete data using the same tools and flight lines to receive unbiased results. The first section of this chapter discusses the results while the second section discusses future work.

A. DISCUSSION

Vegetation classification can be improved to a higher degree of accuracy using Full Waveform LiDAR data and a small sample of ground observations. Our initial goal was to determine if tree types could be distinguishable with waveform data, which proved to be obtainable by using Support Vector Machines. Additionally, we also established that the SVM tool on ENVI can classify the difference between low vegetation and "trail/road" returns with a higher level of accuracy than discrete LiDAR data. This finding is in concurrence with the results in Wagner et al. (2008) and Zlinsky et al. (2014), in which the low vegetation classifiers improved using waveform data. Other testing involved tools such as the Spectral Angle Mapper, maximum likelihood and K-Means to classify the areas, but we concluded that these tools could not successfully distinguish the difference between the identified species with the data set given. While performing classification techniques, we recognized how essential the ground observation data was to validate tree species classifications. Future studies focusing on this objective should identify isolated trees while conducting ground observations to minimize mixed classifications of various tree species that can corrupt the classifiers.

Discrete LiDAR data can, by itself, identify various objects and categorize them into generic categories, but, through the use of full waveform LiDAR, we can improve this accuracy. Low vegetation elements are often misclassified as ground points while using discrete data because these classification methods normally group low vegetation returns, such as scrubs, with the ground or trail since their intensity levels and heights are similar. By using full waveform data to classify a scene, we can improve Terrain Classification (TERCAT) capabilities in future operations. When troops conduct mission planning

scenarios, a common element to consider is understanding the operational environment, which includes the terrain. Additionally, through the use of full waveform data we can improve tree density estimations, which can be used by troops to identify the best route for transporting ground forces on the move in over heavily canopied areas.

Through the use of full waveform LiDAR data, collected from space or airborne platforms, and machine learning classification tools, we can identify those specific terrain elements with a higher degree of accuracy to improve operations.

B. FUTURE WORK

Future researchers could identify the low vegetation plant species located within Point Lobos to determine if their spectral characteristics are unique from each other. In order to do this, we would need to conduct another accurate ground survey of the park to include location, species type, and roughly the size of the species. A key starting point would be to identify a test sample species that is isolated from other vegetation to increase the reliability of the sample.

Additionally, this thesis only included the Optech Titan collection platform when analyzing the full waveform data. Once the AHAB waveform data set can be converted into the readable ENVI file, researchers can conduct the same study over heavily vegetated areas within Point Lobos. The unique scanning pattern of the AHAB data can be particularly useful because the point density for some areas within the Optech Titan flight lines were low. Overlaying the AHAB data to balance the insufficient Optech flight lines can further validate the results or increase the classification capabilities even further.

LIST OF REFERENCES

Abshire, J. B. (2011, April). NASA's space LiDAR measurements of the earth and planets: A brief overview. Paper presented at Institute of Electrical and Electronics Engineers (IEEE) Photonics Society Meeting, College Park, MD.

AHAB bathymetric LiDAR system. (2014, May 27). Retrieved from http://www.bshc.pro/media/documents/lidar2014/04_140527_ AHAB_Bathymetric_LiDAR.pdf

Aldred, A.H., Bonner, G.M., & Petawawa National Forestry Institute (1985). *Applications of airborne lasers to forest surveys.* Petawawa National Forestry Institute, Chalk River, Ont. Canadian Forestry Service Office. Retrieved from http://cfs.nrcan.gc.ca/pubwarehouse/pdfs/4525.pdf

Anderson, K., Hancock, S., Disney, M., & Gaston, K. J. (2016). Is waveform worth it? A comparison of LiDAR approaches for vegetation and landscape characterization. *Remote Sensing in Ecology and Conservation*, *2*, 5–15. https://doi.org/10.1002/ rse2.8

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. *Methods in Molecular Biology 609*. Retrieved from https://link.springer.com/protocol/ 10.1007/978-1-60327-241-4_13

Bufton, J. L., Harding, D.J., & Garvin, J.B. (1999). *Shuttle Laser Altimeter: Mission results and Pathfinder accomplishments* (Report No. 19990087525). Retrieved from https://www.researchgate.net/publication/23812561_Shuttle_ Laser_Altimeter?ev=auth_pub

ENVI LiDAR. (2018). Retrieved from http://www.harrisgeospatial.com/docs/ IntroductionLidar.html

Ferraz, A., Bretar, F., Jacquemoud, S., & Goncalves, G. (2009). *The role of LiDAR systems in fuel mapping* (Report No. INESCC-13). Retrieved from https://www.researchgate.net/publication/323525770_The_Role_of_ Lidar_Systems_in_Fuel_Mapping

Fieber, K. D., Davenport, I.J., Ferryman, J.M., Gurney, R.J., Walker, J.P., & Hacker, J.M. (2013). Analysis of full-waveform LiDAR data for classification of an orange orchard scene. *ISPRS Journal of Photogrammetry and Remote Sensing 82*. Retrieved from https://www.sciencedirect.com/science/article/pii/ S0924271613001238

Frost, J. (2017, April 15). How to interpret R-squared in regression analysis. Retrieved from http://statisticsbyjim.com/regression/interpret-r-squared-regression/

Garvin, J.B., Bufton, J.L., Blair, J.B., Harding, D.J., Lutheke, S., Frawley, J., & Rowlands, D. (1998). Observations of the earth's topography from the Shuttle Laser Altimeter (SLA): laser-pulse echo-recovery measurements of terrestrial surfaces. *Physics and Chemistry of the Earth, 23*(9-10), 1053–1068. https://doi.org/10.1016/S0079-1946(98)00145-1

Hancock, S., Lewis, P., Foster, M., Disney, M., & Muller, J. (2012). Measuring forest with dual wavelength LiDAR: a simulation study of topography. *Agricultural and Forest Meteorology 161*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168192312001232

Harding, D.J., Lefsky, M.A., Parker, G., & Blair, J.B. (2001). Laser altimeter canopy height profiles: Methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment, 76*(3), 283–297. https://doi.org/10.1016/S0034-4257(00)00210-8

Harris Geospatial Solutions. (2009). *ENVI ex user's guide* (4th ed.). Broomfield, CO: Author. Retrieved from http://www.harrisgeospatial.com/SoftwareTechnology/ENVI.aspx

Hofton, M. A., Minster, J.B., & Blair, J.B. (2000). Decomposition of Laser Altimeter Waveforms. *IEEE Transactions on Geoscience and Remote Sensing 38*(4). Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=851780

Isenburg, M. (2012, July 9). PulseWaves: lvis-data2.png. Retrieved from https://github.com/PulseWaves/PulseWaves/blob/master/screen_shots/lvis_data2.png

Isenburg, M. (2017, July 4). Removing excessive low noise from dense-matching point clouds. Retrieved from https://rapidlasso.com/ 2017/07/04/removing-excessive-low-noise-from-dense-matching-point-clouds/

Isenburg, M. (2018). LAStools. Retrieved from https://rapidlasso.com/lastools/

Introduction to Light Detection and Ranging. (2018) Retrieved from https://www.neonscience.org/intro-lidar-r-series

Jalobeanu, A., & Goncalves, G. R. (2012, March 19–23). *The full-waveform LiDAR Riegl LMS-Q6801: from reverse engineering to sensor modeling*. Paper presented at the Annual Meeting of the American Society for Photogrammetry and Remote Sensing, Sacramento, CA. Retrieved from http://lsiit-miv.u-strasbg.fr/paseo/publis/j-asprs12.pdf

Koetz, B., Morsdorf, F., Sun, G., Ranson, K.J., Itten, K., & Allgower, B. (2006). Inversion of a Lidar Waveform Model for Forest Biophysical Parameter Estimation. *IEEE Geoscience and Remote Sensing Letters, 3*(1), 49–53. https://doi.org/10.1109/LGRS.2005.856706

Krabill, W. B., J. G Collins, L. E. Link, R.N. Swift, and Butler. M.L. (1984). Airborne Laser Topographic Mapping Results. *Photogrammetry Engineering and Remote Sensing 50*(6), 685–694. Retrieved from https://www.asprs.org/wp-content/uploads/pers/1984journal/jun/1984_jun_685-694.pdf

Krabill, W. B., Collins, J. G., Swift, R.N., & Butler, M.L. (1980). *Airborne Laser Topographic Mapping Results from the Initial Joint NASA/U.S. Army Corps of Engineers Experiment* (Report No. 19800019281) Retrieved from NASA https://ntrs.nasa.gov/search.jsp?R=19800019281

Lefsky, M. A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters, 37*(15), L15401, https://doi.org/10.1029/2010GL043622

Lefsky, M. A., Harding, D.J, Keller, M., Cohen, W.B., Carabajal, C.C., Espirito-Santo F.D., Hunter, M.O., & Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters, 32(*22), L22S02, https://doi.org/10.1029/2005GL023971

Lefsky, M. A., Keller, M., Pang, Y., Camargo, P.B., & Hunter, M.O. (2007). Revised method for forest canopy height estimations from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing 1*. Retrieved from https://www.fs.fed.us/global/iitf/pubs/ja_iitf_2007_lefsky001.pdf

Lefsky, M. A., Warren, D. C., Harding, D.J., Parkers, G.G., Acker, S.A., & Gower, T. (2002). LiDAR remote sensing of above-ground biomass in three biomes. *Global Ecology & Biogeography, 11*(5), 393–399. https://doi.org/10.1046/j.1466-822x.2002.00303.x

Leica chiroptera II bathymetric & topographic LiDAR sensor. (n.d.). Retrieved from https://leica-geosystems.com/en-us/products/airborne-systems/lidar-sensors/leica-chiroptera-ii

Lu, X., Yongxiang, H., Lucker, P.L., & Trepte, C. (2016). Forest Canopy Height Estimations from CALIPSO LiDAR Measurements. *EJP Web of Conferences, 119*, 22005. https://doi.org/10.1051/epjconf/201611922005

Marcoe, K. (2007). *TIN & surface interpolation* [Lecture]. Retrieved from http://web.pdx.edu/~jduh/courses/geog493f12/Week06.pdf

Optech titan multispectral LiDAR. (2015, May 15). Retrieved from https://www.teledyneoptech.com/wp-content/uploads/Titan-Specsheet-150515-WEB.pdf

Pourrahmati, R.M., Baghdadi N., Darvishsefat, A.A., Namiranian, M., Gond, V., Bailly, J., & Zargham, N. (2017) Mapping lorey's height over Hyrcanian forest of Iran using synergy of ICESat/GLAS and optical images. *European Journal of Remote Sensing, 51*(1), 100–115, https://doi.org/10.1080/22797254.2017.1405717

QTM - Quick Terrain Modeler: LiDAR exploitation software. (n.d.). Retrieved from http://appliedimagery.com/

Quadros, N.D. (2013). Unlocking the characteristics of bathymetric LiDAR sensors. *LiDAR magazine 3*(6), Retrieved from http://www.lidarmag.com/PDF/LiDARMagazine_Quadros-BathymetricLiDARSensors_Vol3No6.pdf

Remote Sensors. (2018, May 29). Retrieved from https://earthdata.nasa.gov/user-resources/remote-sensors

Solutions: forestry inventory. (2018, May 28). Retrieved from http://www.terrasolid.com/solutions/forest_inventory.php

Sumnall, M.J., Hill, R.A., & Hinsley, S.A. (2015). Comparison of small-footprint discrete return and full waveform airborne LiDAR data for estimating multiple forest variables. *Remote Sensing of Environment, 173,* 214-223. https://doi.org/10.1016/j.rse.2015.07.027

Sun, G., K. Ranson, K.J., Zhang, G., Montesano, P., & Kimes, D. (2011). Forest biomass mapping from LiDAR and radar synergies. *Remote Sensing of Environment, 115*(11), 2906–2916. https://doi.org/10.1016/j.rse.2011.03.021

Ulrich, A., & Pfennigbauer, M. (2011). Categorisation of full waveform data provided by laser scanning devices. *SPIE 8186*(09). Retrieved from http://adsabs.harvard.edu/abs/2011SPIE.8186E..09U

Vaughn, N.R., Moskal, M.L., & Turnblom, E.C. (2012). Tree Species Detection Accuracies Using Discrete Point Lidar and Airborne Waveform Lidar. *Remote Sensing, 4*(2), 377–403. https://doi.org/10.3390/rs4020377

Wagner, W., Hollaus, M., Briese, C., & Ducic, V. (2008). 3D vegetation mapping using small footprint full-waveform airborne laser scanners. *International Journal of Remote Sensing, 29*(5), 1433–1452. https://doi.org/10.1080/01431160701736398

Whitehurst, A.S., Swatantran, A., Blair, J.B., Hofton, M.A., & Dubayah, R. (2013). Characterization of Canopy Layering in Forested Ecosystems Using Full Waveform Lidar. *Remote Sensing, 5*(4). 2014–2036. https://doi.org/10.3390/rs5042014

WSI applied remote sensing and analysis. (2013, Nov. 22). Point Lobos State Reserve LiDAR & Imagery. Unpublished report.

Zlinsky, A., Schroiff, A., Kania, A., Deak, B., Mucke, W., Vari, A., Szekely, B., & Pfeifer, N. (2014). Categorizing grassland vegetation with full-waveform airborne laser scanning: A feasibility study for detecting natura 2000 habitat types. *Remote Sensing*, *6*(9), 8056–8087. https://doi.org/10.3390/rs6098056

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California