MODELING PLOT-LEVEL BIOMASS AND VOLUME USING AIRBORNE AND TERRESTRIAL LIDAR MEASUREMENTS

A Thesis

by

RYAN D. SHERIDAN

Submitted to the Office of Graduate Students of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2011

Major Subject: Forestry

Modeling Plot-level Biomass and Volume Using Airborne and Terrestrial Lidar Measurements

MODELING PLOT-LEVEL BIOMASS AND VOLUME USING AIRBORNE AND TERRESTRIAL LIDAR MEASUREMENTS

A Thesis

by

RYAN D. SHERIDAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Sorin Popescu
Committee Members,	Cristine Morgan
	Ross Nelson
Head of Department,	Steven Whisenant

May 2011

Major Subject: Forestry

ABSTRACT

Modeling Plot-Level Biomass and Volume Using

Airborne and Terrestrial Lidar Measurements. (May 2011)

Ryan D. Sheridan, B.S., University of Idaho

Chair of Advisory Committee: Dr. Sorin Popescu

The United States Forest Service (USFS) Forest Inventory and Analysis (FIA) program provides a diverse selection of data used to assess the status of the nation's forested areas using sample locations dispersed throughout the country. Airborne, and more recently, terrestrial lidar (light detection and ranging) systems are capable of producing accurate measurements of individual tree dimensions and also possess the ability to characterize three-dimensional vertical forest structure. This study investigates the potential of airborne and terrestrial scanning lidar systems for modeling forest volume and aboveground biomass on FIA subplots in the Malheur National Forest, eastern Oregon. A methodology for the creation of five airborne lidar metric sets (four point cloud-based and one individual tree based) and four terrestrial lidar metric sets (three height-based and one distance-based) is presented.

Metrics were compared to estimates of subplot aboveground biomass and gross volume derived from FIA data using national and regional allometric equations respectively. Simple linear regression models from the airborne lidar data accounted for

15% of the variability in subplot biomass and 14% of the variability in subplot volume, while multiple linear regression models increased these amounts to 29% and 25%, respectively. When subplot estimates of biophysical parameters were scaled to the plot-level and compared with plot-level lidar metrics, simple linear regression models were able to account for 60 % of the variability in biomass and 71% of the variation in volume. Terrestrial lidar metrics produced moderate results with simple linear regression models accounting for 41 % of the variability in biomass and 46% of the variability in volume, with multiple linear regression models accounting for 71% and 84%, respectively. Results show that: (1) larger plot sizes help to mitigate errors and produce better models; and (2) a combination of height-based and distance-based terrestrial lidar metrics has the potential to estimate biomass and volume on FIA subplots.

DEDICATION

To my wife, Andrea, for her love, patience, sacrifice, support, and encouragement throughout my education.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Sorin Popescu, for his inspiration, advice, support, and feedback throughout the course of my research. I am also grateful to my committee members, Dr. Ross Nelson and Dr. Cristine Morgan, for their valuable feedback and other advice. Also, I would like to show my gratitude to Dr. Ross Nelson for sparking my interest in lidar remote sensing.

I would like to thank Dr. Demetrios Gatziolis and Nian-Wei Ku for their help while collecting the data for this study in Oregon. I am indebted to Dr. Demetrios Gatziolis providing access to and an explanation of the data collected by the United States Forest Service Forest Inventory and Analysis Program.

I owe my deepest gratitude to the department faculty and staff for their continued support of my education. It is an honor for me to thank my friends and colleagues in the Spatial Sciences Laboratory for helping me learn and grow as an individual while studying at Texas A&M University.

Finally, thanks to my mother, father, mother in-law, father in-law, and wife for their incredible encouragement, patience, support, and love.

NOMENCLATURE

AGL	Above ground level
ALS	Airborne laser scanning
CHM	Canopy height model
DBH	Diameter at breast height
DEM	Digital elevation model
DSM	Digital surface model
FIA	Forest inventory and analysis
FS	Forest service
GPS	Global positioning system
Lidar	Light detection and ranging
QTM	Quick Terrain Modeler
TLS	Terrestrial laser scanning
USDA	United States Department of Agriculture
UTM	Universal Transverse Mercator
VIF	Variance inflation factor
WAAS	Wide area augmentation system

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

Page

Page

1 INTRODUCTION AND LITERATURE REVIEW

Light detection and ranging (lidar) is a laser-based, active remote sensing system, which collects ranging data utilizing the known speed of light and information about the flight time of a laser pulse (Lim et al., 2003). In this context, flight time refers to the time it takes for a given laser pulse to travel from a system, reflect off of an object, and return back to the system. A wide variety of lidar systems currently exist, and data has been successfully collected utilizing systems mounted to space-borne, aerial, and terrestrial (tripods or vehicle-based) platforms.

Over the past several decades the use of lidar remote sensing data in forestry has seen steady growth. The increased use of lidar systems to acquire data over forested areas can be attributed to their ability to cover extents of local or regional scales and accurately quantify the three-dimensional vertical structure of the forest. Previous studies have demonstrated the usefulness of lidar for: (1) Forest measurements (Nilsson, 1996; Næsset, 1997a; Næsset, 1997b; Næsset and Bjerknes, 2001; Næsset and Okland, 2002; Popescu et al., 2002; Holmgren et al., 2003); (2) habitat analysis (Hyde et al., 2006); (3) estimation of forest biophysical parameters (Cannell, 1984; Nelson et al., 1988; Lefsky et al., 1999; Holmgren, 2004; Lim and Treitz, 2004; Patenaude et al., 2004; Popescu, 2007); (4) change detection (yu et al., 2004); and (5) estimation of wild land fire parameters (Hall et al., 2005; Mutlu et al. 2008a; Mutlut et al., 2008b).

This thesis follows the style of *Remote Sensing of Environment*.

It should be noted that the ability to acquire three-dimensional data is not unique to lidar remote sensing systems. This type of data can also be obtained by radar systems (another active remote sensing system) or through the use of photogrammetric techniques in conjunction with stereoscopic image pairs collected by aerial or satellite systems. A variety of studies have provided comparison of lidar and radar forest measurements. For example, Sexton et al. (2009) used linear regression to examine lidar canopy height measurements and radar canopy height measurements and concluded that lidar provided more precise results ($R^2 = 0.83$). Hyde et al. (2007) used lidar, synthetic aperture radar (SAR), and interferometric synthetic aperture radar (InSAR) to individually and synergistically predict aboveground biomass for a southwestern ponderosa pine forest, and found, through individual comparison, that lidar predicted aboveground biomass best, accounting for almost 84% of the variability.

Airborne lidar systems, also known as airborne laser scanners (ALS) can be broadly grouped into two categories: discrete return and full waveform digitizers. These categories can be further specified by the type of system (profiling or scanning), laser footprint size, and the number of recorded returns for each laser pulse. Previous lidar studies have demonstrated that both large-footprint waveform and small-footprint discrete return lidar data, can be used to derive measurements such as tree height and crown dimensions at the stand level (Næsset and Bjerknes, 2001; Hall et al., 2005), plot level (Holmgren et al., 2003; Lim and Treitz, 2004; Popescu et al., 2004), or individual tree level (Coops et al., 2004; Yu et al., 2004; Holmgren and Persson, 2004; Roberts et al., 2005; Chen et al., 2006; Falkowski et al. 2006; Popescu, 2007). These direct lidar

measurements can then be used in conjunction with known allometric relationships or statistical analysis procedures to estimate parameters such as diameter at breast height (DBH), aboveground biomass, or volume.

Lidar research for forestry applications has largely focused on the development of methodologies to employ lidar data as a surrogate for various ground measurements. Airborne lidar data can be collected over larger areas with a reduced amount of effort when compared to traditional field measurements. However, the high level of complexity present within many forests (e.g. large number of species and variable canopy densities) can complicate the retrieval of such measurements. In Norway, researches have developed and implemented methods to produce measurements of interest for stand-based forest inventories, and were able to account for 84 to 89% of the variance when predicting stand volume (Næsset, 2007). Other stand-based variables of interest investigated by Scandinavian researchers include: mean tree height, dominant height, mean diameter, stem number, stand basal area, and stand volume (for a list of Scandinavian studies and a basic summary of their results, see Næsset, 2007).

Methods used to identify individual trees commonly begin using filters to detect local maxima, which are assumed to represent individual tree crown peaks, on a canopy height model (CHM). A CHM is a regular gridded surface representation of the forest canopy. This model is created by subtracting a digital elevation model (DEM), a gridded representation of the bare Earth's surface created from last return lidar data, from the Digital Surface Model (DSM), a gridded surface created using all of the lidar point cloud

data. After locating the individual trees on the CHM, tree crown measurements can be produced for each identified tree.

A number of methods for identifying local maximum in CHMs exist. Currently, local maxima have been identified in CHMs using both variable (Popescu et al., 2002; Popescu et al., 2004) and fixed (McCombs et al., 2003) filter window sizes. Popescu et al. (2002) identified tree crown peaks using a variable search window, where the window size was based on a relationship between tree height and crown width established from field measurements. McCombs et al. (2003) used a focal search function, based on a fixed and predetermined neighborhood size, to identify individual tree crown peaks in a Loblolly pine plantation. Both studies mentioned above assumed that the pixel at the center of the neighborhood was a crown peak if it was higher than all pixels within the search neighborhood. The height of each tree is the height value of the identified maxima pixel corresponding to the tree.

After identifying individual trees in the CHM, crown measurements can be made. A variety of methods can be used to obtain these measurements (e.g. Hyyppä et al., 2001; Popescu et al., 2003; Roberts et al., 2005; Falkowski et al., 2006; Qi et al., 2006). A Study by Hyyppä et al. (2001) used an image segmentation process where local maxima were defined as seed points, serving as inputs for an image-labeling algorithm along with a number of required user-defined variables set through trial and error procedures. Popescu et al. (2003) derived tree crown widths by calculating the mean of distances measured between two local minima identified on a forth degree polynomial fitted to two separate orthogonal canopy profiles of a tree. While Roberts et al. (2005)

estimated crown diameter by recording the distance from each identified tree crown peak to minima identified in 3.4 m linear search array in each cardinal direction from a tree crown peak, averaging the four radii measurements, and multiplying them by two. Qi et al. (2006) proposed a marker-controlled watershed segmentation method, where a CHM's (or other similar model) height values are inverted and flooded from the bottom up as "dams" are automatically constructed to keep water from neighboring areas separate, effectively separating neighboring tree canopies.

As illustrated above, the individual tree approach is capable of directly measuring tree height and canopy dimensions. However, it must be mentioned that tree height is often underestimated by airborne lidar systems. This underestimation is commonly attributed to the low probability of an individual laser pulse striking the apex of a tree. High canopy densities can result in the occlusion of laser pulses, and reduce the ability to create accurate DEMs (Hirata et al., 2009). Additionally, individual tree detection methods also commonly commit errors of commission and omission, and have been shown to have lower accuracy when used in forests with complex and dense canopy conditions (Falkowski et al., 2008).

Since ALS systems collect data looking down on the forest, forest measurements other than tree height or crown dimensions (e.g. diameter at breast height, biomass) must be indirectly estimated. Popescu 2007, used regression analysis to estimate the DBH of individual trees, using the lidar-derived height and crown diameter measurements provided by TreeVaW (a individual tree detection software package) as independent variables in a regression analysis. Individual tree detection algorithms implemented in

TreeVaW are described in Popescu and Wynne (2004). In traditional forestry, biomass estimations require destructive sampling, or the use of species-specific (Termikaelian and Korzukhin, 1997), regional, or national (Jenkins et al., 2003) allometric equations. Allometric equations can also be applied to lidar data, if the required information is available. Popescu (2007) outlines a method for obtaining individual tree biomass estimates using allometric equations and estimates of individual tree DBH from ALS data. A number of other studies have also predicted biomass using lidar data (Lefsky et al., 1999; Patenaude et al., 2004; Hyde et al., 2007; Zhao et al., 2009).

Measurements from terrestrial laser scanners (TLS) have been used for a relatively short time, starting in late 1990s. TLS systems are mounted on a terrestrial-based platform (e.g. tripod, motor vehicle) and in most cases require a portable computer and several large batteries or a portable generator to operate. Systems such as these are capable of acquiring and merging scans from multiple locations through the use of at least two stationary targets. This process requires more time scanning an area, but has the potential to provide a more complete picture of a study area.

A number of studies have provided methodologies for deriving forest measurements using data collected by TLS systems. Hopkinson et al. (2004) isolated individual trees for height and DBH measurement in merged lidar scans with two distinct forest types. Thies et al. (2004) merged lidar scenes, and produced detailed stem measurements of several large-diameter deciduous trees. A method for automatically identifying individual trees, measuring tree height, and tree DBH is presented in Maas et al. (2008). Henning and Radtke (2006) developed methods for identifying trees scanned

with a TLS system and measuring stem diameter. Their results showed the average error between the lidar-derived diameter estimates and caliper measurements, for sections below the base of live crown, to be less than 1 cm. Methods of deriving other forest measurements, such as aboveground biomass, have also been explored. Lefsky and McHale (2008) used high-density point cloud data for multiple urban trees with complex architecture in an attempt to develop allometric relationships for predicting species tree volume.

In the United States, the Forest Service (USFS) Forest Inventory and Analysis (FIA) program provides a diverse selection of data used to assess the status of the nation's forested areas. In the past, the FIA program used a periodic inventory system, where measurements on non-national forests were collected on a state-by-state basis in predetermined zones, and lead to inventory cycles of ranging from six to eighteen years (Gillespie, 1999). In 1998 legislation was passed (see the Agricultural Research, Extension, and Education Reform Act of 1998) that mandated the entire FIA program implement an annual inventory. This inventory method requires that the collection, analysis, and reporting of data at a state-level be completed every five years, meaning that under ideal conditions, 20% of plots in each state would be measured each year (Gillespie, 1999). The annualized FIA program allows for the collection of a variety of parameters of interest and consists of three phases: (1) remote sensing to identify forested and non-forested areas; (2) field samples located at intervals of about 1 plot every 6,000 acres, where forested sample areas are visited by field crews to collect ground measurements and non-forested areas are visited to quantify the frequency of

variables such as land use change; and (3) consists of visiting a subset of the plots in phase 2 (about 1 plot every 96,000 acres) to collect more detailed measurements (e.g. complete vegetation inventory, tree and crown condition, soil data) during the growing season (USFS, 2008).

The measurements collected by the FIA program can be scaled up to provide information about forest populations by aggregating plot statistics for specific populations. However, this is only possible if the population(s) of interest have been adequately sampled by the inventory. Many regional to national scale biomass and carbon budgets for the United States are based largely on the forest information provided by the FIA program, regional-level volume and biomass equations, and national-level allometric equations (Heath et al. 2008). Heath et al. (2008) also notes that in recent years the requests to the FIA for biomass, carbon, and volume information have continually increased.

1.1 Objectives

The overall objective of this study is to develop a methodology for modeling forest volume and aboveground biomass from ALS and TLS data by comparing subplot volume and aboveground biomass estimates derived from ground-based FIA measurements and allometric equations to metric sets created from spatially coincident lidar data. This will address the hypothesis that since lidar systems collect data that describe the three-dimensional vertical structure of the forest, the data can be used to estimate forest biophysical parameters of interest such as volume and aboveground biomass. Specific study objectives follow:

1. Develop a methodology to derive both area- and individual tree-based airborne lidar metrics related to forest biophysical parameters at the FIA subplot-level.
2. Develop a methodology to derive area-based terrestrial lidar metrics related to forest biophysical parameters at the FIA subplot-level.
3. Utilize simple linear and multiple linear regression analysis to help identify relationships between the lidar metric sets and FIA subplot estimates of forest volume and aboveground biomass calculated from FIA data.

1.2 Thesis Organization

This thesis contains a total of four major sections. An overall introduction and literature review for the thesis are presented here. The contents of Section 2 and Section 3 were created and organized to resemble individual manuscripts. Section 4 provides a summary of the conclusions presented in Sections 2 and 3.

2 COMPARISON OF AIRBORNE LIDAR METRICS AND GROUND-BASED ESTIMATES OF TOTAL ABOVEGROUND BIOMASS AND VOLUME

2.1 Introduction

Light detection and ranging (lidar) is a laser-based, active remote sensing system, which collects ranging data utilizing the known speed of light and information about the flight time of a laser pulse (Lim et al., 2003). In this context, flight time refers to the time it takes for a given laser pulse to travel from a system, reflect off of an object, and return back to the system. A wide variety of lidar systems currently exist, and data has been successfully collected utilizing systems mounted to space-borne, aerial, and terrestrial (tripods or vehicle-based) platforms.

Over the past several decades the use of lidar remote sensing data in forestry has seen steady growth. The increased use of lidar systems to acquire data over forested areas can be attributed to their ability to cover extents of local or regional scales and accurately quantify the three-dimensional vertical structure of the forest. Previous studies have demonstrated the usefulness of lidar for: (1) Forest measurements (Nilsson, 1996; Næsset, 1997a; Næsset, 1997b; Næsset and Bjerknes, 2001; Næsset and Okland, 2002; Popescu et al., 2002; Holmgren et al., 2003); (2) habitat analysis (Hyde et al., 2006); (3) estimation of forest biophysical parameters (Cannell, 1984; Nelson et al., 1988; Lefsky et al., 1999; Holmgren, 2004; Lim and Treitz, 2004; Patenaude et al., 2004; Popescu, 2007); (4) change detection (Yu et al., 2004); and (5) estimation of wild land fire parameters (Hall et al., 2005; Mutlu et al., 2008a; Mutlu et al., 2008b).

It should be noted that the ability to acquire three-dimensional data is not unique to lidar remote sensing systems. This type of data can also be obtained by radar systems (another active remote sensing system) or through the use of photogrammetric techniques in conjunction with stereoscopic image pairs collected by aerial or satellite systems. A variety of studies have provided comparisons of lidar and radar forest measurements to ground measurements. For example, Sexton et al. (2009) used linear regression to examine lidar canopy height measurements and radar canopy height measurements and concluded that lidar provided more precise results ($R^2 = 0.83$). Hyde et al. (2007) used lidar, synthetic aperture radar (SAR), and interferometric synthetic aperture radar (InSAR) to individually and synergistically predict aboveground biomass for a southwestern ponderosa pine forest, and found, through individual comparison, that lidar predicted aboveground biomass best, accounting for almost 84% of the variability.

Airborne lidar systems, also known as airborne laser scanners (ALS) can be broadly grouped into two categories: discrete return and full waveform digitizers. These categories can be further specified by the type of system (profiling or scanning), laser footprint size, and the number of recorded returns for each laser pulse. Previous lidar studies have demonstrated that both large-footprint waveform and small-footprint discrete return lidar data, can be used to derive measurements such as tree height and crown dimensions at the stand level (Næsset and Bjerknes, 2001; Hall et al., 2005), plot level (Holmgren et al., 2003; Lim and Treitz, 2004; Popescu et al., 2004), or individual tree level (Coops et al., 2004; Yu et al., 2004; Holmgren and Persson, 2004; Roberts et al., 2005; Chen et al., 2006; Falkowski et al. 2006; Popescu, 2007). These direct lidar

measurements can then be used in conjunction with known allometric relationships or statistical analysis procedures to estimate parameters such as diameter at breast height (DBH), aboveground biomass, or volume.

Lidar research for forestry applications has largely focused on the development of methodologies to employ lidar data as a surrogate for various ground measurements. Airborne lidar data can be collected over larger areas with a reduced amount of effort when compared to traditional field measurements. However, the high level of complexity present within many forests (e.g. large number of species and variable canopy densities) can complicate the retrieval of such measurements. In Norway, researches have developed and implemented methods to produce measurements of interest for stand-based forest inventories, and were able to account for 84 to 89% of the variance when predicting stand volume (Næsset, 2007). Other stand-based variables of interest investigated by Scandinavian researchers include: mean tree height, dominant height, mean diameter, stem number, stand basal area, and stand volume (for a list of Scandinavian studies and a basic summary of their results, see Næsset, 2007).

Methods used to identify individual trees commonly begin using filters to detect local maxima, which are assumed to represent individual tree crown peaks, on a canopy height model (CHM). A CHM is a regular gridded surface representation of the forest canopy. This model is created by subtracting a digital elevation model (DEM), a gridded representation of the bare Earth's surface created from last return lidar data, from the Digital Surface Model (DSM), a gridded surface created using all of the lidar point cloud

data. After locating the individual trees on the CHM, tree crown measurements can be produced for each identified tree.

A number of methods for identifying local maximum in CHMs exist. Currently, local maxima have been identified in CHMs using both variable (Popescu et al., 2002; Popescu et al., 2004) and fixed (McCombs et al., 2003) filter window sizes. Popescu et al. (2002) identified tree crown peaks using a variable search window, where the window size was based on a relationship between tree height and crown width established from field measurements. McCombs et al. (2003) used a focal search function, based on a fixed and predetermined neighborhood size, to identify individual tree crown peaks in a Loblolly pine plantation. Both studies mentioned above assumed that the pixel at the center of the neighborhood was a crown peak if it was higher than all pixels within the search neighborhood. The height of each tree is the height value of the identified maxima pixel corresponding to the tree.

After identifying individual trees in the CHM, crown measurements can be made. A variety of methods can be used to obtain these measurements (e.g. Hyyppä et al., 2001; Popescu et al., 2003; Roberts et al., 2005; Falkowski et al., 2006; Qi et al., 2006). A Study by Hyyppä et al. (2001) used an image segmentation process where local maxima were defined as seed points, serving as inputs for an image-labeling algorithm along with a number of required user-defined variables set through trial and error procedures. Popescu et al. (2003) derived tree crown widths by calculating the mean of distances measured between two local minima identified on a forth degree polynomial fitted to two separate orthogonal canopy profiles of a tree. While Roberts et al. (2005)

estimated crown diameter by recording the distance from each identified tree crown peak to minima identified in 3.4 m linear search array in each cardinal direction from a tree crown peak, averaging the four radii measurements, and multiplying them by two. Qi et al. (2006) proposed a marker-controlled watershed segmentation method, where a CHM's (or other similar model) height values are inverted and flooded from the bottom up as "dams" are automatically constructed to keep water from neighboring areas separate, effectively separating neighboring tree canopies.

As illustrated above, the individual tree approach is capable of directly measuring tree height and canopy dimensions. However, it must be mentioned that tree height is often underestimated by airborne lidar systems. This underestimation is commonly attributed to the low probability of an individual laser pulse striking the apex of a tree. High canopy densities can result in the occlusion of laser pulses, and reduce the ability to create accurate DEMs (Hirata et al., 2009). Additionally, individual tree detection methods also commonly commit errors of commission and omission, and have been shown to have lower accuracy when used in forests with complex and dense canopy conditions (Falkowski et al., 2008).

Since ALS systems collect data looking down on the forest, forest measurements other than tree height or crown dimensions (e.g. diameter at breast height, biomass) must be indirectly estimated. Popescu 2007, used regression analysis to estimation the DBH of individual trees, using the lidar-derived height and crown diameter measurements provided by TreeVaW (a individual tree detection software package) as independent variables in a regression analysis. Individual tree detection algorithms implemented in

TreeVaW are described in Popescu and Wynne (2004). In traditional forestry, biomass estimations require destructive sampling, or the use of species-specific (Termikaelian and Korzukhin, 1997), regional, or national (Jenkins et al., 2003) allometric equations. Allometric equations can also be applied to lidar data, if the required information is available. Popescu (2007) outlines a method for obtaining individual tree biomass estimates using allometric equations and estimates of individual tree DBH from ALS data. A number of other studies have also predicted biomass using lidar data (Lefsky et al., 1999; Patenaude et al., 2004; Hyde et al., 2007; Zhao et al., 2009).

The United States Forest Service (USFS) Forest Inventory and Analysis (FIA) program provides forest inventory measurements used to assess the status of the nation's forested areas. Forest resource managers and researchers commonly use these measurements to estimate forest biophysical parameters such as, volume, aboveground biomass, or carbon at local, regional, and national scales. This direct link between data provider and end user makes the FIA program responsible for many of the volume estimates, biomass budgets, and carbon budgets created for the United States.

The collection of forest inventory data at a national level is an enormous and complex undertaking. Models relating airborne lidar data to FIA parameters hold great potential to contribute to this task, by: (1) supplementing aging ground-based FIA measurements or biophysical parameter estimates with estimates produced from recently collected lidar data; (2) providing an increased amount of data for areas of interest that contain only a small number of FIA sample locations; or (3) aiding data collection in

remote areas where challenging environmental or terrain conditions make ground-based measurements exceedingly dangerous and time consuming.

The overall objective of this study is to develop a methodology for modeling forest volume and aboveground biomass at the FIA subplot-level, using a combination of previously developed area-based point cloud metrics and individual tree measurements. Subplot biomass and volume estimates, calculated using ground-based FIA measurement data and allometric equations, are compared to the previously mentioned metrics. This will address the hypothesis that because the data collected by ALS systems are capable of describing the three dimensional vertical structure of the forest, they can be used to estimate forest biophysical parameters of interest such as volume and aboveground biomass. Specific study objectives follow:

1. Develop a methodology to derive both area and individual tree-based airborne lidar metrics related to forest biophysical parameters at the FIA subplot-level.
2. Identify relationships between the lidar metric sets and FIA subplot estimates of forest volume and aboveground biomass calculated using a combination of regional and national-scale equations.

2.2 Materials and Methods

2.2.1 Study Area

The study area for this project is in the Malheur National Forest located in eastern Oregon, and covers approximately 105,936 hectares (Figure 1). Elevation ranges from 1,236 to 2,593 m, and slope varies from 0 to ~ 86 degrees. The general location of the study area, as defined by the NE and SW corners of a rectangle, is Universal

Transverse Mercator (UTM) Zone 11N 383297.6E, 4905767.9N and UTM Zone 11N 333344.5E, 4863102.6N. The site was selected because of access to FIA ground measurements, availability of recent airborne scanning LIDAR data, and the presence of a wide variety of forest conditions, such as slope and tree species. The forests located within the study area are composed of mostly Ponderosa pine (*Pinus ponderosa*), Douglas-fir (*Pseudotsuga menziesii*), western larch (*Larix occidentalis*), and grand fir (*Abies grandis*).

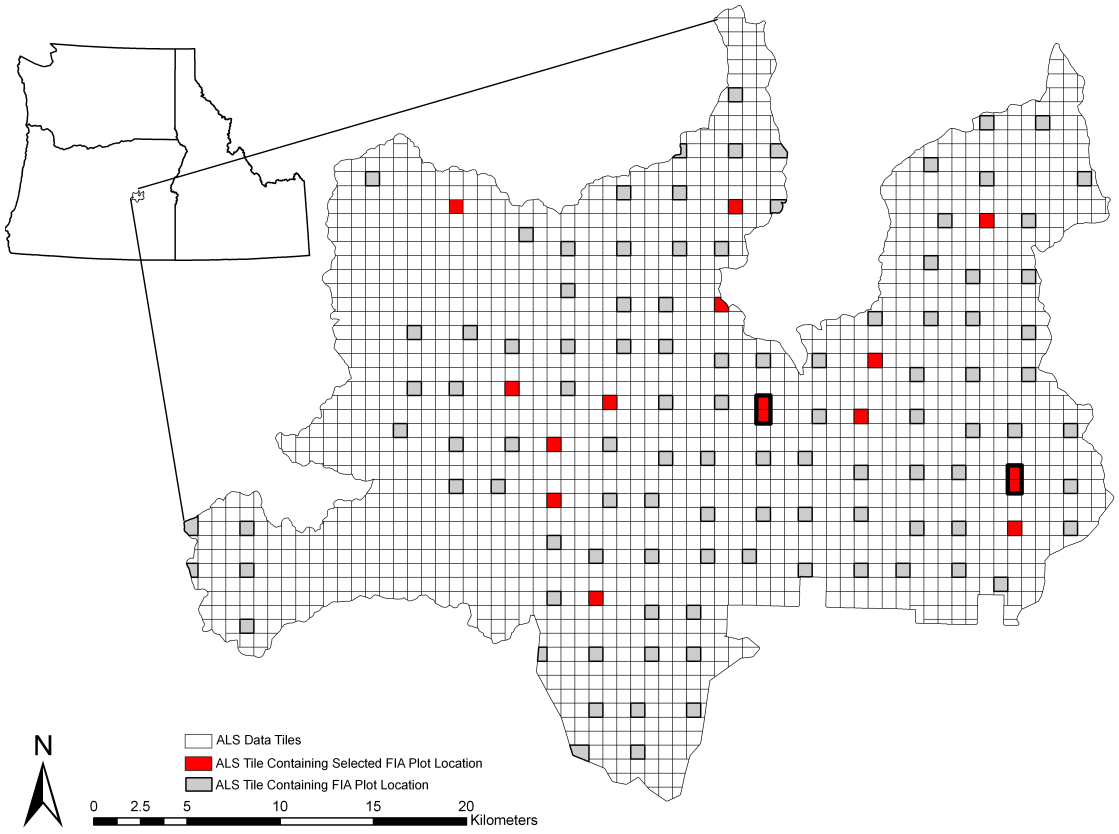

Figure 1: Malheur National Forest study area in eastern Oregon. Hollow squares represent individual airborne lidar tiles. Red squares represent tiles containing an FIA plot location selected for this study. Red squares with a thick outline identify selected tiles where the FIA plot location was located on the border of two tiles. Grey squares signify tiles containing an FIA plot location.

2.2.2 Data

This section of the study used several types of data to describe forest conditions at FIA plot locations including: (1) FIA ground crew *in situ* measurements; and (2) Discrete return, small-footprint, airborne lidar (ALS).

2.2.2.1 Forest Inventory and Analysis Data

The USFS provided FIA data for all FIA locations within the study area (91 locations, 364 subplots, and about 2,477 trees). This study will focus on FIA subplots since they are utilized in every FIA region. Each FIA location contains four circular ~0.016 hectare subplots (radius = 7.32 m). Subplot one is centered over the plot center for the entire FIA location. Subplots two, three, and four are located 36.58 m from the center of subplot at azimuths of 360°, 120°, and 240°, respectively (Figure 2). Individual trees are measured and recorded if they are located within the boundaries of subplot and have a DBH or diameter at root collar greater than 12.7 cm. Measurements collected for each of these trees include: DBH, height, tree condition (live/dead), crown class (open grown, dominant, codominant, intermediate, or overtopped), species, species group, azimuth to plot center, and distance to plot center.

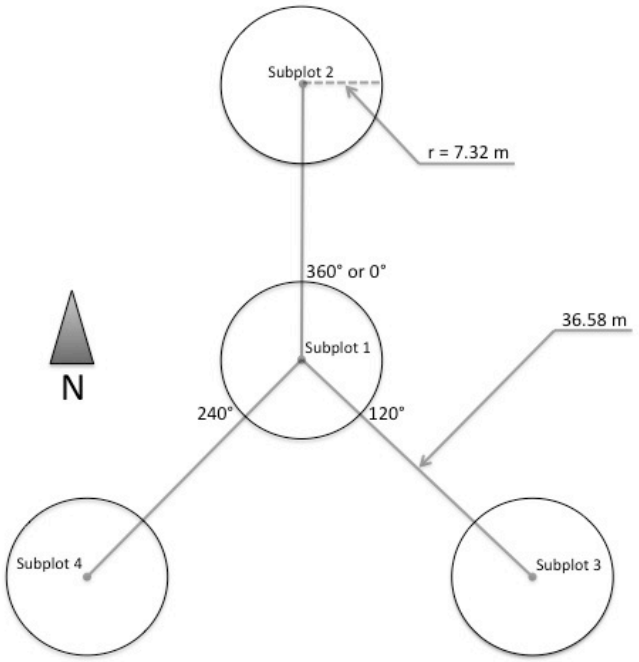

Figure 2: The location and dimensions of subplots at an FIA plot location.

The FIA program has provided estimates of volume for the majority of the trees in the study area. Estimates were only omitted for trees with a status code that listed the tree as dead, or for trees where the status code was completely absent. The regional equations used to calculate tree volume estimates can be found in Zhou and Hemstrom (2010). Estimates of individual tree total aboveground biomass (dry kg) were to also be included with the FIA data. However, the current high workload of the FIA data steward, and the time constraints of this study did not make this a feasable option. The national-scale total aboveground biomass equations, presented in Jenkins et al. (2003), were used to estimate the total aboveground biomass for individual trees in lieu of the estimates from regional FIA equations.

The use of the national-scale total above ground biomass equations, will result in different estimates of tree biomass than regional biomass equations. While, national-scale equations produce more generalized estimates, they are capable of providing standardized estimates for major species or species goups at large-spatial scales. The national-scale equations published by Jenkins et al. (2003), were created by reanalyzing reliable and raw tree measurement data collected throughout different regions and for different species by other scientists from previously published studies. Alternativly, regional equations are capable of more precise estimates, but are only representative for trees in a specific region. The equations are developed using data collected small study sites, and are thus representative of the species found near the area where data were collected. Hansen (2002) examined differences between allometric equations for estimating biomass and volume, and found differences between estimates of volume and biomass produced using the same set of tree measurements with different regional equatuions.

Subplot-level estimates for total aboveground biomass and volume were calculated by summing the total aboveground biomass and volume estimates for all of the trees in each subplot. FIA plot-level estimates for total aboveground biomass and volume were calculated by averaging the aboveground biomass and volume estimates for each of the four subplots.

2.2.2.2 Airborne Lidar Data

Two airborne lidar data acquisition missions were flown for the study area, the first covering the western half and a second covering the eastern half (Figure 3). Data were acquired for the western half of the study area between November 19, 2008 and December 11, 2008, while data for the eastern half of the study area were acquired between July 1, 2009 and July 5, 2009. Both data collection missions made use of multiple lidar sensors, a Leica ALS50 Phase II and a Leica ALS60, mounted in a single engine fixed-wing survey aircraft. The complete system is capable of collecting greater than or equal to 105,000 pulses per second at 900 m above ground level, and can record up to 4 returns for each pulse. For the study area, the lidar sensors used a maximum scan angle of ± 14°. A brief analysis of the lidar data for the study area shows an average pulse density of about 9 pulses m^{-2}. A total of 456 flight lines were required to cover the entire study area, with a side-lap of ≥ 50% for each of the N-S oriented flight lines.

Figure 3: The extent of the western (light grey) and eastern (dark grey) airborne lidar acquisitions.

2.2.3 Processing Approach

The overall processing steps used in this study are illustrated in Figure 4.

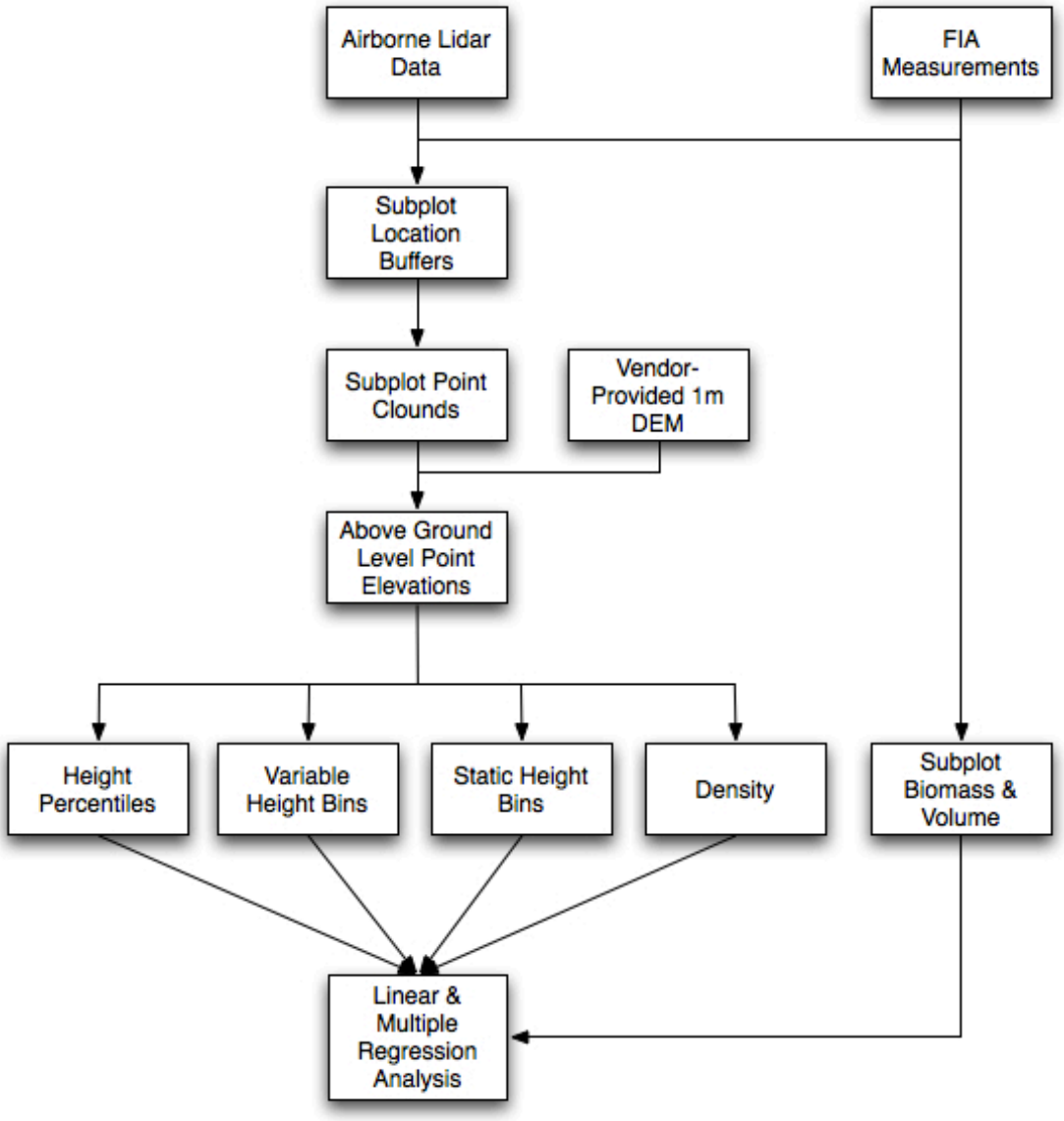

Figure 4: Flowchart of the airborne lidar data processing approach.

2.2.4 Point Cloud-Based Airborne Lidar Metrics

Three sets of point cloud-based ALS metrics were calculated for the 56 subplots at 16 FIA plot locations dispersed throughout the study area. Four of the FIA locations were chosen because TLS data had been collected for at least one subplot at a given location. The 12 other FIA locations were selected at random, without replacement, from the remaining 87 FIA plot locations in the study area. An initial review of the FIA data for the selected subplots identified five with no tree records. In this case, the subplots either contained only dead trees, or trees too small to meet the FIA measurement criteria mentioned in 2.2.2.1. The subplots from the selected FIA locations cover a range of slopes and other forest conditions present in the Malheur National Forest. Table 1 and Table 2 provide summaries of the data collection year for each subplot and the frequency of each tree species in each crown class. Descriptive statistics for FIA measured DBH and tree height, as well as FIA estimated total aboveground biomass and volume are presented in Table 3 through Table 5.

Table 1: Subplot data collection years.

	Sample year							
	2001	2002	2003	2004	2005	2006	2007	2008
Number of subplots	8	4	12	4	4	4	8	12

Table 2: Subplot tree species crown class frequencies.

		Crown Class				
Species	**n[a]**	**OG[b]**	**D[c]**	**CD[d]**	**I[e]**	**OT[f]**
Abies grandis	76	1	18	28	25	4
Juniperus occidentalis	5	0	2	1	2	0
Larix occidentalis	8	0	2	3	3	0
Pinus contorta	43	0	12	24	7	0
Pinus ponderosa	119	0	29	62	24	4
Pseudotsuga menziesii	50	0	21	25	3	1
Cerocarpus leditolins	11	0	0	2	9	0

[a] Number of trees
[b] Open grown crown class
[c] Dominant crown class
[d] Codominant crown class
[e] Intermediate crown class
[f] Overtopped crown class

Table 3: Subplot tree species DBH descriptive statistics.

		DBH (cm)			
Species	**n[a]**	**mean**	**min - max**	**SD[b]**	**CV(%)[c]**
Abies grandis	76	29.81	13.21 - 103.10	15.17	50.9
Juniperus occidentalis	5	22.86	13.72 - 45.72	13.44	49.9
Larix occidentalis	8	27.62	16.76 - 43.69	10.52	38.09
Pinus contorta	43	20.63	13.72 - 32.51	5.17	25.08
Pinus ponderosa	119	33.78	12.70 - 104.60	24.17	71.53
Pseudotsuga menziesii	50	34.46	14.73 - 82.55	17.53	50.87
Cerocarpus leditolins	11	26.92	13.72 - 120.60	12.29	115.98

[a] Number of trees
[b] Standard deviation
[c] Coefficient of variation percentage

Table 4: Subplot tree species height descriptive statistics.

		Height (m)			
Species	**n[a]**	**mean**	**min - max**	**SD[b]**	**CV(%)[c]**
Abies grandis	76	16.54	7.62 - 35.06	5.84	35.32
Juniperus occidentalis	5	11.65	6.10 - 15.55	4.21	36.15
Larix occidentalis	8	17.53	10.37 - 28.05	5.93	33.85
Pinus contorta	43	15.46	9.15 - 21.34	3.09	20
Pinus ponderosa	119	16.44	4.27 - 44.82	8.97	54.58
Pseudotsuga menziesii	50	18.87	3.66 - 38.11	6.84	36.28
Cerocarpus leditolins	11	6.125	4.27 - 8.54	1.15	18.82

[a] Number of trees
[b] Standard deviation
[c] Coefficient of variation percentage

Table 5: Subplot descriptive statistics for estimated biomass and volume.

	n[a]	**mean**	**min - max**	**SD[b]**	**CV(%)[c]**
Biomass (dry kg)	56	3558.2	0 - 25143.9	4505.1	126.6
Volume (m^3)	56	5	0 - 45.6	7.3	146.6

[a] Number of subplots
[b] Standard deviation
[c] Coefficient of variation percentage

2.2.4.1 Extraction of Subplot Point Clouds

Coordinates for the center of each subplot were included in the data provided by the FIA program. These coordinates were collected during field measurements using a WAAS (wide area augmentation system) enabled Trimble global positioning system (GPS) unit, and then post-processed using base station data. Proprietary Trimble software was utilized to compute the average accuracy of the reported coordinates. The average accuracy values are computed to represent a 95th percentile three-dimensional

distance threshold around the subplot center location. The resulting point data were provided as a shapefile, to allow for visualization and manipulation in ArcMap.

The geolocation accuracy of the post-processed subplot center coordinates can be quantified by examining the average accuracies of the GPS-collected subplot center coordinates provided in the shapefile (Figure 5 and Table 6). These data show that when working with the 56 subplots selected for this study, the geolocation error ranged from 0.50 to 3.78 m. It should be noted it is uncommon for FIA crews to utilize the survey-grade GPS units and advanced processing methods used to collect subplot coordinate information within this study area. Instead, crews frequently rely on recreational-grade GPS units, which, under the best data collection conditions, commonly provide positional uncertainties of at least three meters.

Table 6: Descriptive statistics for the average accuracies (m) of subplot center coordinates.

n[a]	mean	min - max	SD[b]	CV(%)[c]
56	1.36	0.50 - 3.78	0.66	48.48

[a] Number of subplots
[b] Standard deviation
[c] Coefficient of variation percentage

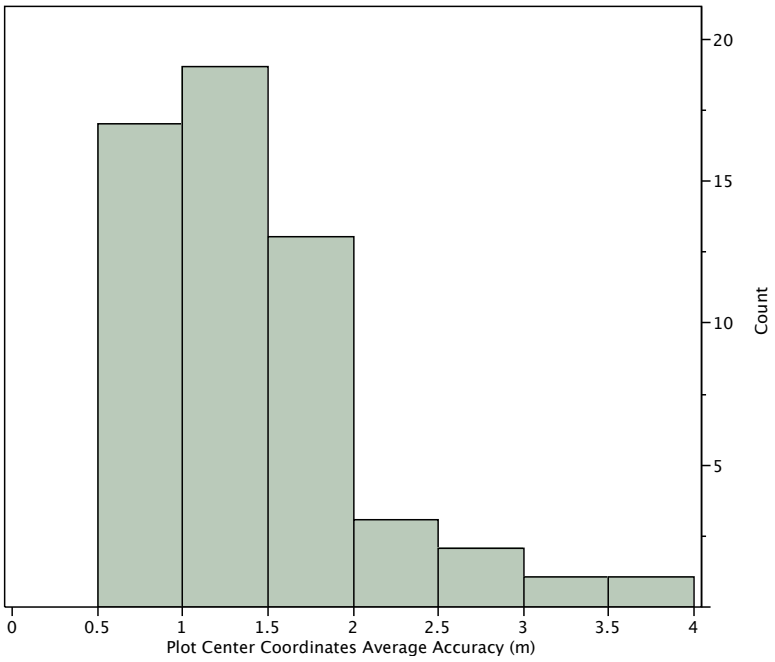

Figure 5: Histogram of average accuracies for subplot center coordinates.

The four subplots for each of the chosen FIA locations were selected and individually extracted from the provided shapefile. This process created a total of 56 shapefiles, each containing a single point, which represented the plot center coordinates for the selected subplot. A circular buffer (r = 7.32 m) was then created around each subplot center. This process provided polygon shapefiles covering the extent of each selected subplot.

To create individual subplot point clouds, the points within each subplot were extracted from the corresponding lidar data tile using Quick Terrain Modeler (QTM). In all but two cases, each selected FIA plot location, and subplot buffers were contained entirely within one lidar data tile. The extraction procedure for these exceptions was

identical to the procedure described above, except for the need to import the two neighboring lidar data tiles.

2.2.4.2 Calculation of Above Ground Level Elevations and Removal of Ground Points

All point cloud metrics in this study were calculated with above ground level (AGL) elevation values. QTM was used to compute AGL values for each subplot point cloud by subtracting the DEM values from the corresponding point elevations. The vendor hired to acquire the ALS data, Watershed Sciences, Inc., supplied a 1 m DEM of the study area, created from the raw point cloud data. In an operational context, the use of a vendor-provided DEM requires significantly less time than creating a DEM from the raw point cloud data. The methodology presented in this chapter made use of only the vendor-provided DEM.

Points under a predetermined height were excluded from the analysis. The exclusion of points in this manner helps remove possible effects of shrubs, large rocks, and other ground-related laser returns from lidar data. Mitigating these effects reduces the overall amount of data, but helps to insure that the majority of the remaining data represent the objects of interest, e.g. trees. A height cut off value of 0.5 m (AGL) was selected. All points with an elevation of less than this value were cropped from each subplot point cloud.

2.2.4.3 Height Percentiles

Height percentiles, calculated from ALS data have been previously used to relate ALS data to forest biophysical parameters (Holmgren, 2004; Lim and Treitz, 2004; and Patenaude et al. 2004). For this study, mean height, and the 25^{th}, 50^{th}, 75^{th}, 90^{th}, 95^{th}, and 100^{th} height percentiles were calculated for each subplot point cloud.

2.2.4.4 Variable Height Bins

Height bins were used to separate the vertical space within a subplot into five equaly-spaced intervals. The variable height bin approach calculates height bins by finding the distance between highest and lowest points in a plot, and dividing this distance by the number of desired bins (in this case five). The approach used to calculate bins size was similar to the method used by Næsset (2007) to determine bins intervals used when calculating canopy densities. This approach allowed a different height bin interval to be calculated for each subplot. After the height bin interval had been calculated for a given subplot point cloud, the height bin break points were set, and the number of points within each height bin was counted. The ratio of the number of points in a height bin to the total number of points in the subplot point was then calculated for each FIA subplot.

2.2.4.5 Static Height Bins

Static height bins were also used to separate the vertical space within a subplot into intervals. This height bin method was selected to highlight differences between subplots, because it uses the same height bins regardless of the minimum and maximum subplot height values. While conceptually similar to the variable height bin approach

(section 2.2.3.2), this method used constant height bin intervals for all of the subplots, and contained a total of six bins. The static height bin break points selected were: (hb1) 0.5 – 5 m; (hb2) 5 – 10 m; (hb3) 10 – 15 m; (hb4) 15 – 20 m; (hb5) 20 – 25 m; and (hb6) greater than or equal to 25 m. The total number of points within each height bin is counted, as well as the ratio of the number of points in a given height bin to the total number of points in the subplot.

2.2.4.6 Density

Density metrics, similar to those calculated by Næsset (2007), were also calculated for each subplot. The procedure used to calculate density metrics is closely related to the steps described for calculating height bins. Density metrics for this study were calculated using the same static height bin intervals mentioned in section 2.2.3.5 (where Næsset [2007] used a procedure similar to variable height bins, section 2.2.3.4). The major difference between this method and height bins is that all points greater than or equal to a given bin breakpoint are counted, e.g. d1 contains the count of all points greater than or equal to 0.5 m, d2 contains the count of all points greater than or equal to 5.0 m, etc. A ratio of the number of points in each count to the total number of points in the subplot was calculated.

2.2.4.7 Plot-level point cloud metrics

Plot-level point cloud metrics were calculated in conjunction with subplot-level point cloud metrics because of the locational error present in the subplot center coordinates (described in section 2.2.4.1). Previous studies have shown that geolocation

error can increase variation in ALS metrics as well as biophysical parameters estimated from these metrics (Gobakken and Næsset 2008a; Gobakken and Næsset 2008b). Frazer et al. (2011) investigated how plot size and geolocation error impacted estimates of forest stand biomass, and found that plot size could amplify or condense the severity of geolocation error. The authors also conclude that plots with larger areas are more robust to geolocation error because they provide increased amounts of overlap between ground and ALS data, capture more variability from ground measurements, and reduce the perimeter of the plot with respect to the area within the plot. The conclusions presented from these studies provided support for the inclusion of plot-level point cloud metrics since the increased plot size could mitigate the effect of geolocation errors.

The methods described in sections 2.2.4.1 through 2.2.4.2 were used to extract and preprocess FIA plot-level point clouds. The only significant changes to the methods were: (1) buffering was only performed on the subplot center coordinates for subplot number one at each FIA location; and (2) the buffer size was increased from 7.32 m to 56.42 m. These changes extracted all ALS returns within a one-hectare plot that included all FIA subplots and their immediate surrounding areas.

ALS point cloud metrics (height percentiles, variable height bins, static height bins, and densities) were calculated for each of the 14 FIA plot-level point clouds with the methods presented in sections 2.2.4.3 through 2.2.4.6.

2.2.5 Individual Tree Metrics

Methods for detecting and measuring individual trees in ALS data are also found throughout the literature, and are discussed in Section 1. Several of the main differences

between individual tree identification and point-could metrics are: (1) the use of a CHM instead of raw point cloud data; and (2) the resulting estimates of individual tree variables (such as height and crown width). The individual tree detection software TreeVaW was employed to identify trees within each selected subplot, and provide information on tree height, crown width.

2.2.5.1 Canopy Height Model Creation

TreeVaW cannot process raw point cloud data. Instead, this software required CHM. Lidar vendors do not typically provide CHMs to clients, so one must be created for the study area. In order to reduce the amount of time dedicated to generating a CHM, CHMs were only produced for the spatial extent of each lidar data tile containing one of the 16 selected FIA locations. CHMs were generated using subsets of the vendor-provided DEM and user-created DSMs for each lidar data tile. The relatively high point density of the ALS data (~ 9 pts m^{-2}) justified the creation of 0.5 m DSM. QTM was used to import the raw point data in a lidar tile as a 0.5 m gridded surface. This procedure assigns height values based on the maximum height of the points within each cell. Once imported, the DSM was saved, and the spatially coincident DEM subset was subtracted from the DSM to produce a CHM. Negative height values, caused by DSM cells with slightly lower values than the corresponding DEM cells, were set to zero, and the CHM was saved. This process was repeated for all lidar tiles containing a selected FIA location.

2.2.5.2 Calibration of TreeVaW

TreeVaW must be calibrated for different regions using *a priori* knowledge about the relationship between tree height and crown width. Differences between tree species morphology, and the heterogeneous forest species composition in the study area required the production of a general height to crown width relationship. This relationship was established by performing manual tree crown width measurements on trees identified by TreeVaW using its default settings. Crown widths and heights for a total of 100 trees (25 per CHM) were measured on the CHMs of the lidar tiles containing FIA locations 5708, 5992, 5993, and 6185 using image-processing software. Crown width measurements were recorded from North to South and East to West for each tree, and used to calculate an average crown width. A model representing the relationship between the average crown width and height of each measured tree was generated ($R^2 = 0.86$, Figure 6) using simple linear regression analysis. The resulting equation (eq. 1) was used to calibrate TreeVaW for this study area. Other TreeVaW settings that were adjusted included: minimum tree height (3.5 m) and maximum crown width size (15.0 m).

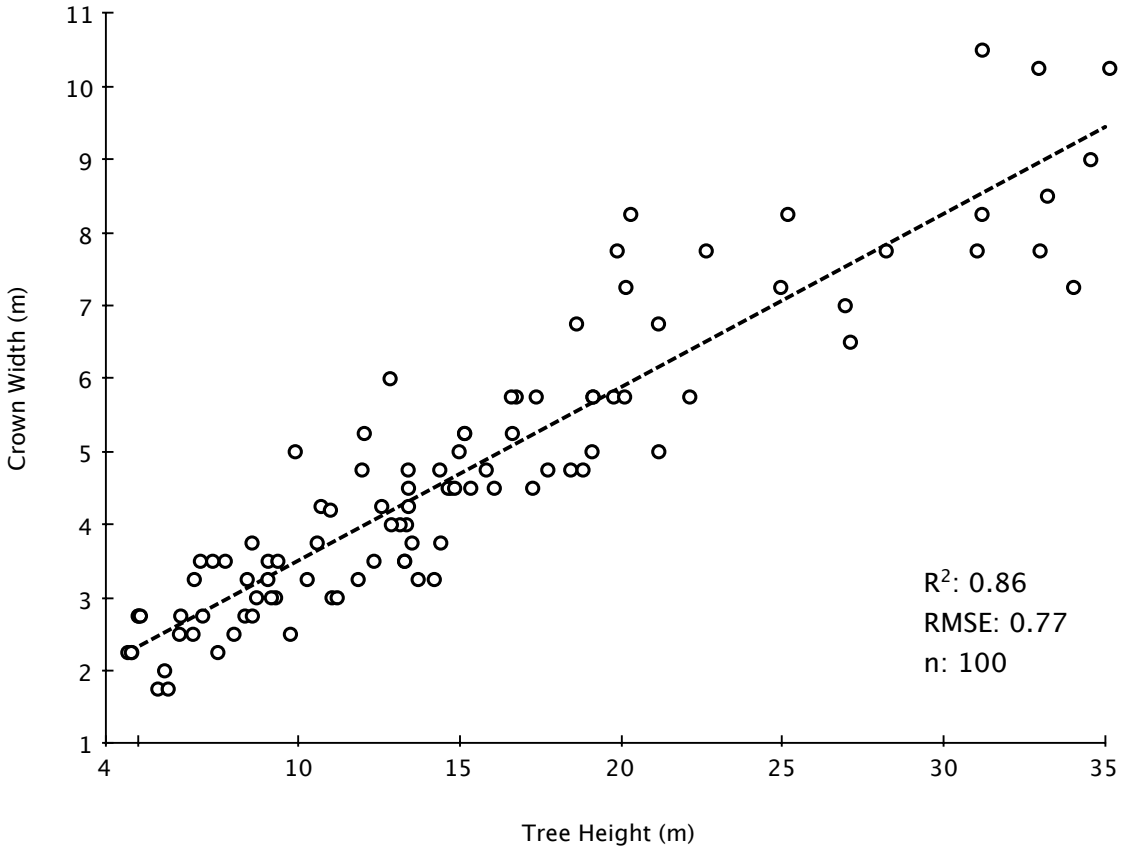

Figure 6: TreeVaW calibration scatterplot, TreeVaW identified tree height vs. average of manually measured tree crown widths, and simple linear regression line.

$$Y = 1.1396 + 0.2374 * x \quad \text{(eq. 1)}$$

2.2.5.3 Calibrated TreeVaW Measurements

After calibrating TreeVaW, the software was run on each of the CHMs. TreeVaW provides text-based output of tree location, tree height, and crown radii. ArcMap was used to create a point shapefile of tree locations. The subplot polygon shapefiles, used to extract the subplot point clouds for the point cloud-based lidar metrics (section 2.2.3.1), were used to extract the TreeVaW-identified trees for each subplot. Attribute tables for the resulting point shapefiles were exported from ArcMap as

CSV files. A list of all of the trees detected in each subplot as well as lists of trees for each individual subplot were created using Excel.

2.2.5.4 Subplot-Level Individual Tree Metrics

Average tree height, average crown width, and total number of stems were calculated for each subplot, using the CSV files exported from ArcMap. TreeVaW produces crown radii estimates from the average of two perpendicular crown diameter estimates for each detected tree. This process utilizes a fourth degree polynomial, centered on the tree location, and attempts to identify and measure the distance between critical points in the fitted function (Popescu et al., 2003). In some cases, the measurement algorithm employed by TreeVaW cannot successfully produce estimates of tree crown diameters. Popescu et al., (2003) attributes this problem to the natural complexity found in forest canopies, e.g. interactions between adjoining tree crowns. A similar problem was encountered when using TreeVaW for this study. A total of 408 trees were identified on the 56 selected subplots. TreeVaW was unable to produce crown radii estimates for 186 (approximately 46%) of the identified trees. The failure to estimate crown radii for such as large number of trees is believed to stem from the highly complex forest canopy structure present in the Malheur, such as: (1) tree crown overlap; (2) the variability of crown morphology between the species present; (3) differences between the crown morphology of species in the Malheur and the southern pine species to create TreeVaW; and (4) multi-story canopy conditions. In a similar study, Popescu et al., (2003) ignored trees with a reported crown radii of zero when calculating plot-level metrics such as average crown width. It should be noted that

percentage of total trees with zero values in Popescu et al., (2003) were much lower, 4.49% for a site with mixed deciduous and pine trees, and 8.78% for a site dominated by large deciduous trees. To minimize the loss of tree measurements due to crown radii values of zero, the general calibration equation (eq. 1) presented in section 2.2.4.2, was used to estimate crown widths for all trees with TreeVaW-reported crown widths of zero.

2.2.6 Regression Analysis

Simple linear regression models were used to examine the relationship between the lidar-derived metrics and the subplot-level estimates of total aboveground biomass and gross volume. The independent variables and methods used to calculate them were described in the previous sections. A summary of the independent and dependent variables used the regression analyses can be found in Table 7. Models were run for each independent variable. Information about the best model from each metric set, and the variable that produced the best model for each metric were provided.

Multiple regression analysis, using a stepwise selection method, was utilized to determine if models could be improved through the inclusion of additional variables within each metric set. A model created using stepwise selection and every independent variable was also created.

Table 7: Selected regression variables.

Metric sets of lidar-derived area-based independent variables	Lidar-derived individual tree independent variables	Predicted variables (FIA field measurements)
Percentiles	ave_ht	Aboveground Biomass (kg)
p25, p50, p75, p90, p95, p100, mean	ave_cw	Gross Volume (ft^3)
Variable Height Bins	num_trees	
vhb1, vhb2, vhb3, vhb4, vhb5		
Static Height Bins		
shb1, shb2, shb3, shb4, shb5, shb6		
Density		
d1, d2, d3, d4, d5, d6		

2.3 Results and Discussion

2.3.1 Subplot-level Point Cloud-Based Airborne Lidar Metrics

Individual simple linear regression models were created for aboveground biomass and gross volume using each of the point cloud-based metrics. Examination of the resulting models showed that all lidar-derived predictor variables were poorly related to the ground-based estimates of subplot aboveground biomass and volume. The best predictor variables from each point cloud-based metric set were the 100^{th} percentile (p100), variable height bin one (vhb1), static height bin four (shb4), and density four (d4, Table 8, Figure 7 to Figure 10 [top]). The best model used p100 as the predictor variable, and yielded R-square values of 0.16 for biomass and 0.14 for volume.

A common practice when outliers or points of inters are found, is to remove the points from the data set and rerun any analyses that were performed on the original dataset. A summary of the resulting regression analyses for the four best predictor variables can be found in Table 8. Metrics designated as “reduced” represent the data with the two points of interest removed.

Table 8: Simple linear regression analysis results for the four best point cloud predictor variables.

Point cloud-based metric	Dependent variable	R^2	RMSE	Parameters and p-values			
				β_0	P-value	β_1	P-value
Height Percentiles							
p100 (all)	Biomass	0.16	4172.11	-1606.41	0.3533	257.65	0.0024
p100 (all)	Volume	0.14	6.81	-2.95	0.2969	0.40	0.0040
p100 (reduced)	Biomass	0.30	2044.63	-951.37	0.2643	190.44	< 0.0001
p100 (reduced)	Volume	0.33	2.92	-2.00	0.1030	0.29	< 0.0001
Variable Height Bins							
vhb1 (all)	Biomass	0.10	4316.95	2025.42	0.0214	9869.66	0.0185
vhb1 (all)	Volume	0.09	7.03	2.64	0.0635	15.03	0.0270
vhb1 (reduced)	Biomass	0.00	2248.78	2870.94	< 0.0001	-266.21	0.9158
vhb1 (reduced)	Volume	0.00	3.57	3.94	< 0.0001	-1.03	0.7798
Static Height Bins							
sbh4 (all)	Biomass	0.02	4513.16	3018.25	0.0008	4581.63	0.3737
sbh4 (all)	Volume	0.01	7.31	4.18	0.0038	6.74	0.4187
sbh4 (reduced)	Biomass	0.12	2297.60	1999.10	< 0.0001	6963.61	0.0103
sbh4 (reduced)	Volume	0.11	3.36	2.61	0.0002	9.89	0.0127
Density							
d4 (all)	Biomass	0.02	4491.30	2964.10	0.0004	3109.28	0.2522
d4 (all)	Volume	0.03	4.97	3.95	0.0031	5.33	0.2248
d4 (reduced)	Biomass	0.16	2243.14	2005.68	< 0.0001	4257.99	0.0026
d4 (reduced)	Volume	0.20	3.19	2.45	< 0.0001	6.92	0.0007

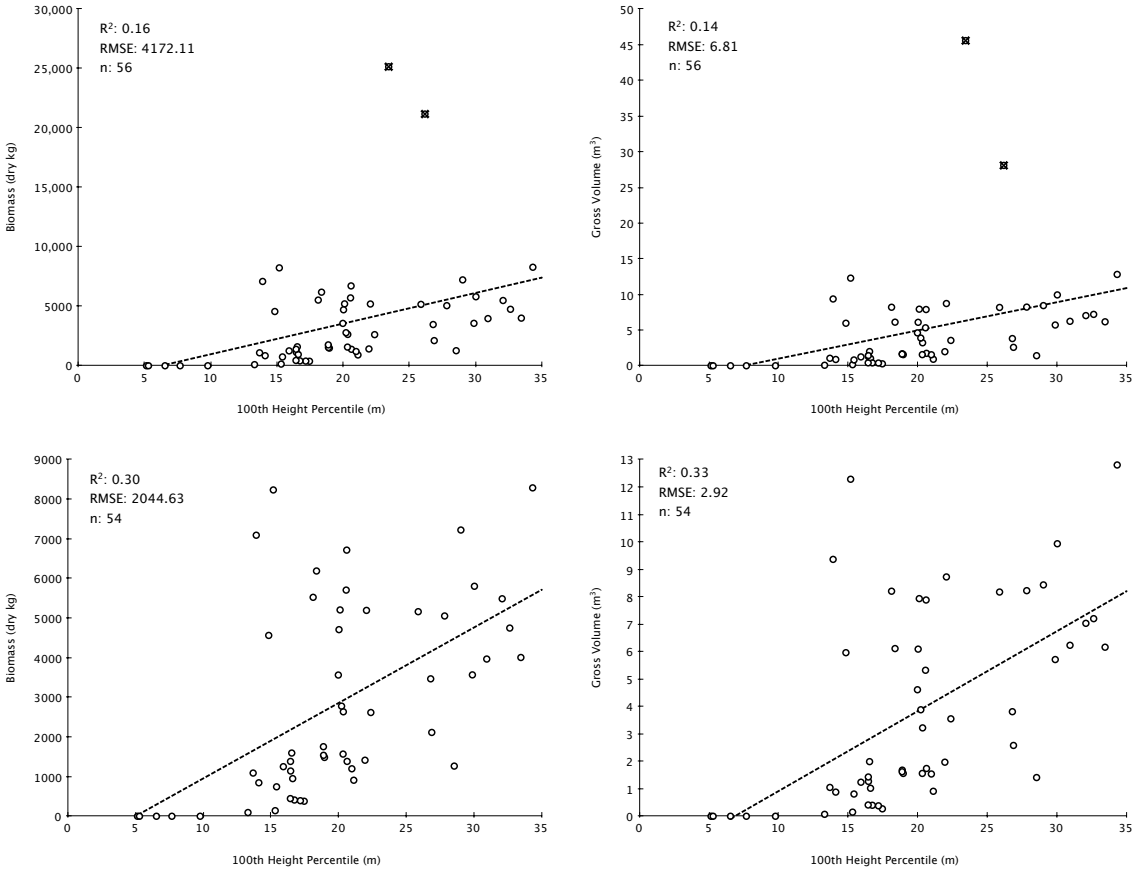

Figure 7: (Top) Scatter plots of aboveground biomass and gross volume vs. the lidar-derived 100th height percentile metric. Filled circles indicate points of interest. (Bottom) scatter plots of aboveground biomass and gross volume with points of interest removed.

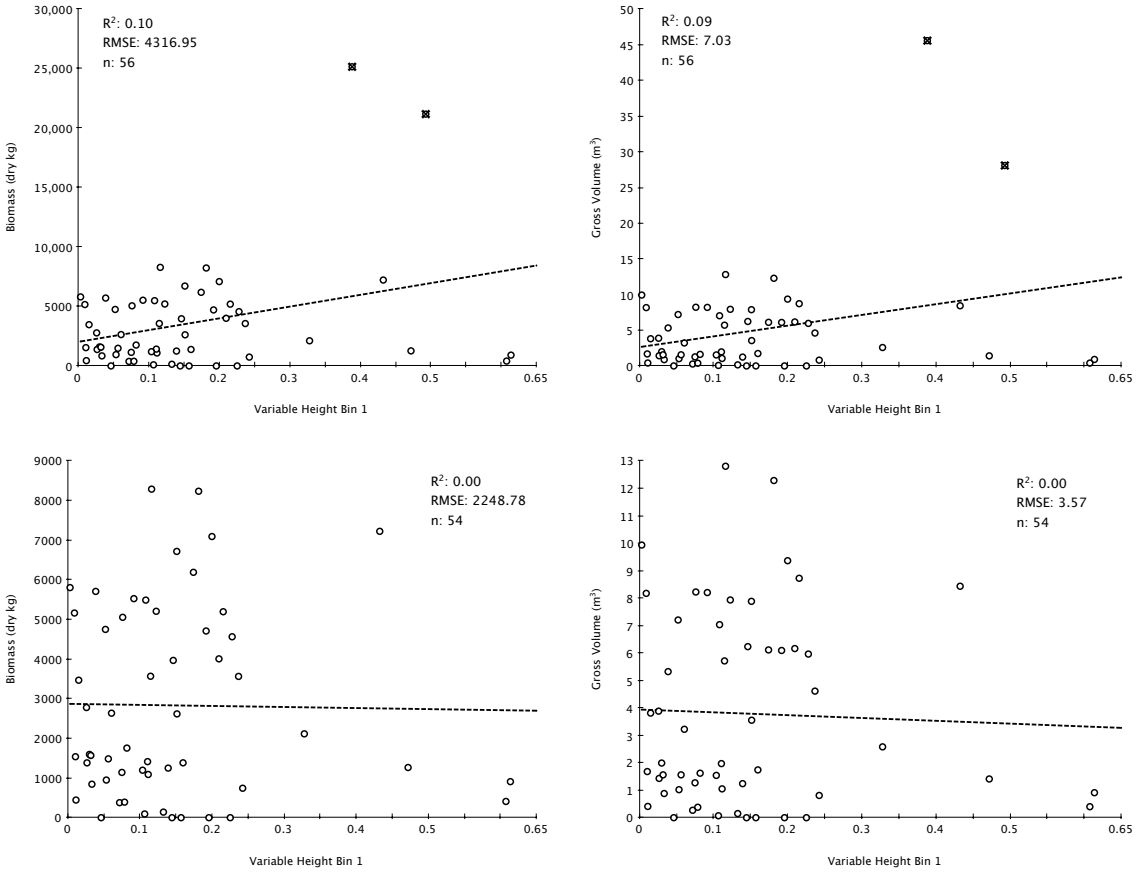

Figure 8: (Top) Scatter plots of aboveground biomass and gross volume vs. the lidar-derived variable height bin 1 metric. Filled circles indicate points of interest. (Bottom) scatter plots of aboveground biomass and gross volume with points of interest removed.

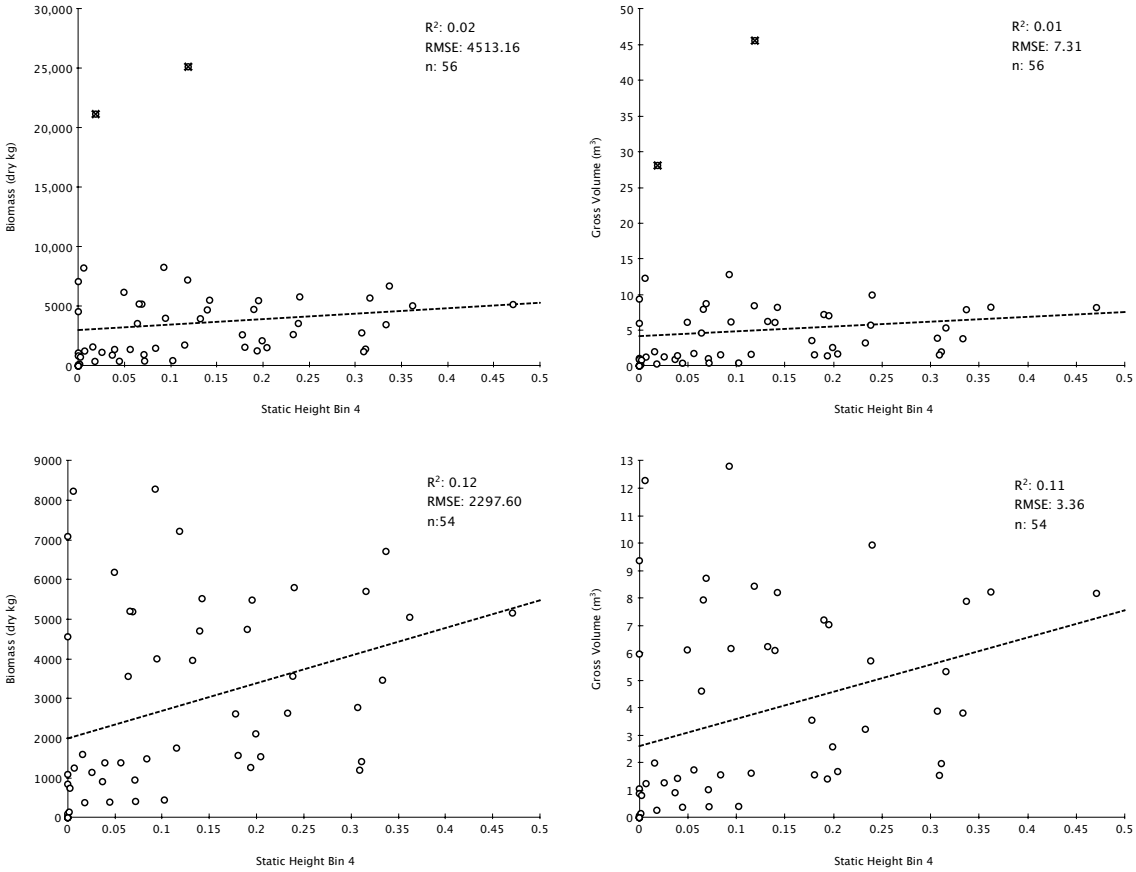

Figure 9: (Top) Scatter plots of aboveground biomass and gross volume vs. the lidar-derived static height bin 4 metric. Filled circles indicate points of interest. (Bottom) scatter plots of aboveground biomass and gross volume with points of interest removed.

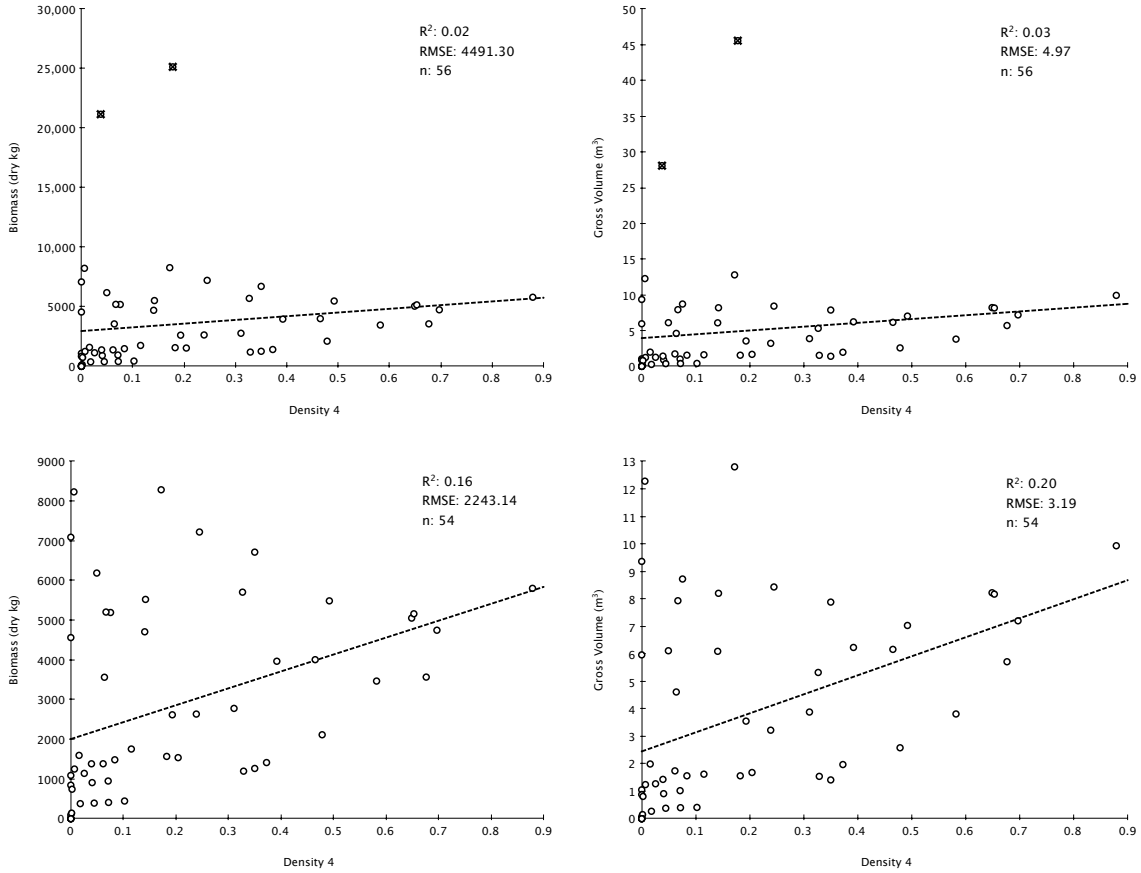

Figure 10: (Top) Scatter plots of aboveground biomass and gross volume vs. the lidar-derived density 4 metric. Filled circles indicate points of interest. (Bottom) scatter plots of aboveground biomass and gross volume with points of interest removed.

The low R-square values produced by these analyses are cause for some concern because previous literature has shown that metrics from the airborne lidar were capable of producing much better results. For instance, prediction of total biomass by Lim and Treitz (2004) produced R-square values of 0.82, 0.84, 0.89, 0.90, and 0.83 using the 100^{th}, 75^{th}, 50^{th}, 25^{th}, and 0^{th} height percentiles, respectively, for uneven-aged hardwood forest dominated by two species. While the results from the previously mentioned study, as well as others, do produce support for the 100^{th} percentile as a good indicator of biomass, the differences in achieved R-square values are very large. When comparing

the methods of Lim and Treitz (2004) to the methodology used for this study, there are several notable differences. One major difference was the use of a natural log transformation for both the predictor and response variables by Lim and Treitz (2004). This transformation is common when estimating forest biophysical parameters using lidar data (see Næsset and Bjerknes, 2001; Næsset and Okland, 2002; Næsset , 2007). However, when the natural log transformation was applied to our point cloud-based metrics and the ground-based estimates of subplot biomass and volume, only a slight improvement in the ability to estimate biomass or volume was seen. For example, simple linear regression models for biomass and volume, using the p100 predictor variable, provided R-square values of 0.28 and 0.31, respectively. It should be noted that five of the subplots had values of zero for biomass and volume, because no trees meeting FIA measurements criteria were located on the plot. Plots with zero values were removed from the dataset prior to the natural log transformation.

Examination of the scatter plots for all of the predictor variables identified two points of interest, both of which had values of aboveground biomass and gross volume that were much larger than the other subplots in the sample, identified in Figure 7 - Figure 10 as filled circles. These two data points were investigated further with a scatter plot of the FIA estimates of biomass and volume demonstrated a strong positive linear trend (Figure 11). Because this trend between the biophysical parameters is expected and clear, it is unlikely that errors were made when calculating individual tree or subplot-level biomass.

The resulting scatter plots and the best-fit regression lines are displayed in the bottom halves of Figures 6 through 9. The removal of the points of interest from the dataset provided moderate, if any, improvement in the R-square values of the models (Table 8).

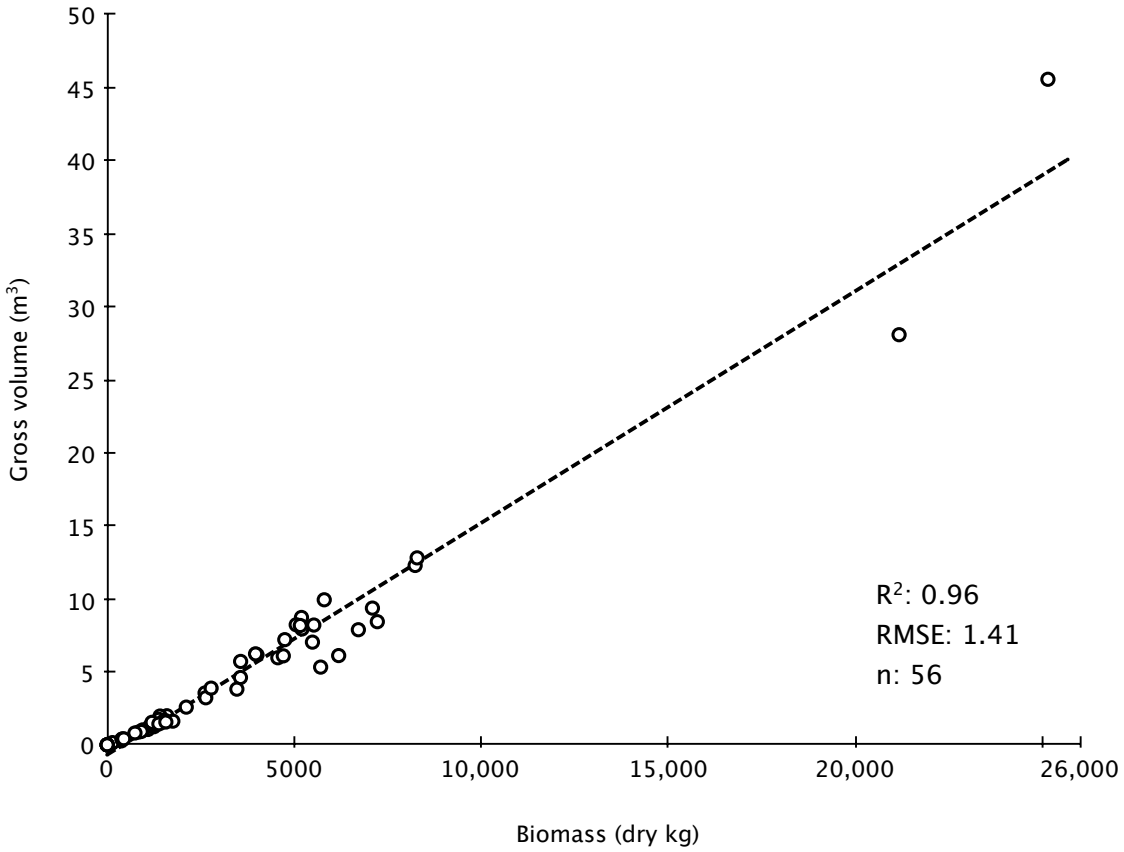

Figure 11: Ground measurement estimated biomass vs. ground measurement estimated gross volume.

A more detailed review of the characteristics of the individual FIA trees located on the two points of interest found that these subplots exhibited a high frequency of large trees. Five of the eleven trees at FIA location 5992, subplot one had DBH values greater than 70 cm, and similarly large heights (Table 9). At FIA location 6218, subplot one,

four of the six trees had DBH values of greater than or equal to 70 cm, and also had large height measurements (Table 10). The high frequency of trees with DBH values greater than or equal to 70 cm does not occur in any of the other selected subplots. A distribution of the measured values of DBH for all live trees meeting FIA measurement criteria in the selected subplots showed a positively skewed distribution (Figure 12). When the nine trees with DBH found on the two subplots mentioned above (DBH greater than or equal to 70 cm) were selected, it is easy to identify that they are in the top 8 to 10% of the trees contained in the entire subplot sample. Furthermore, a similar distribution, using all live trees meeting the FIA measurement criteria in all of the subplots within the study area demonstrates a similar positively skewed pattern, and shows trees with large DBH values are not as common in the study area (Figure 13).

The relatively small number of subplots selected for this study, makes it difficult to say if other subplots with similar conditions (e.g. a high frequency of large diameter trees in a single plot) are a common occurrence within this study area, or if the plots are anomalies. The existence and frequency of plot conditions such as these could be investigated further by increasing the number of subplot observations in the sample.

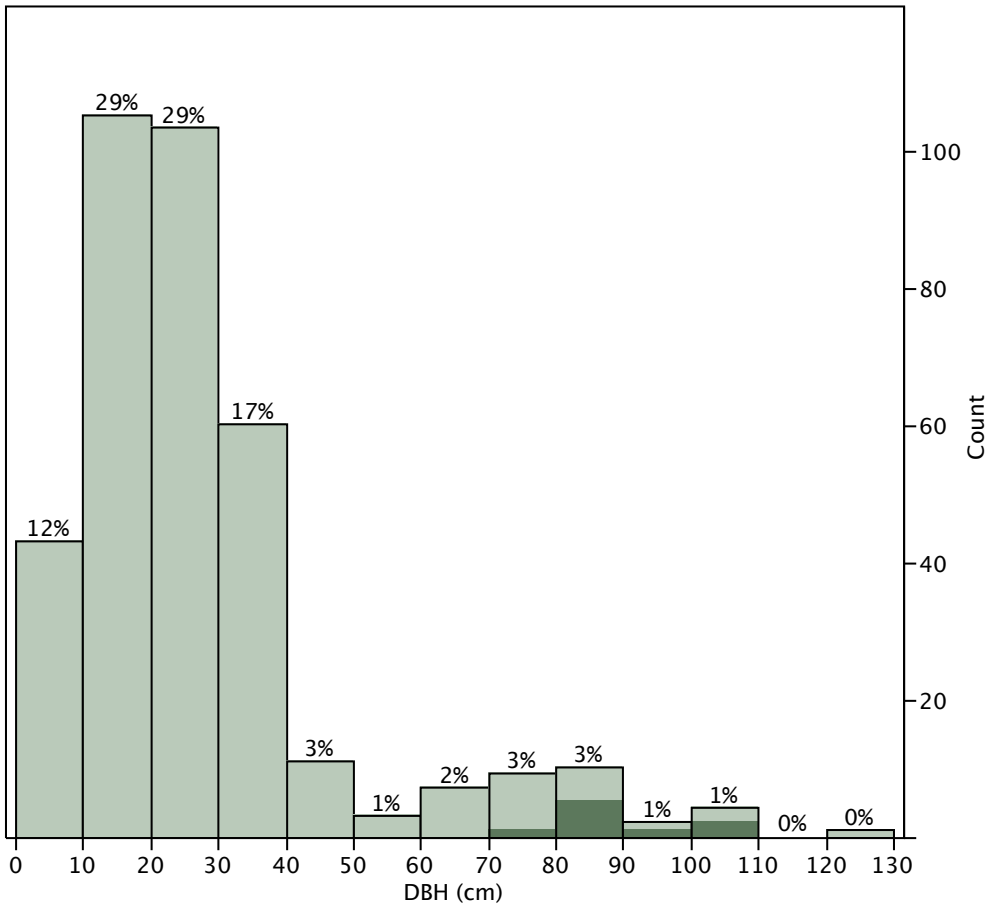

Figure 12: Distribution of FIA subplot tree DBH measurements for the selected subplots. Dark areas indicate the location of trees with recorded DBH values of greater than or equal to 70 cm at FIA locations 5992 subplot 1 and 6218 subplot 1.

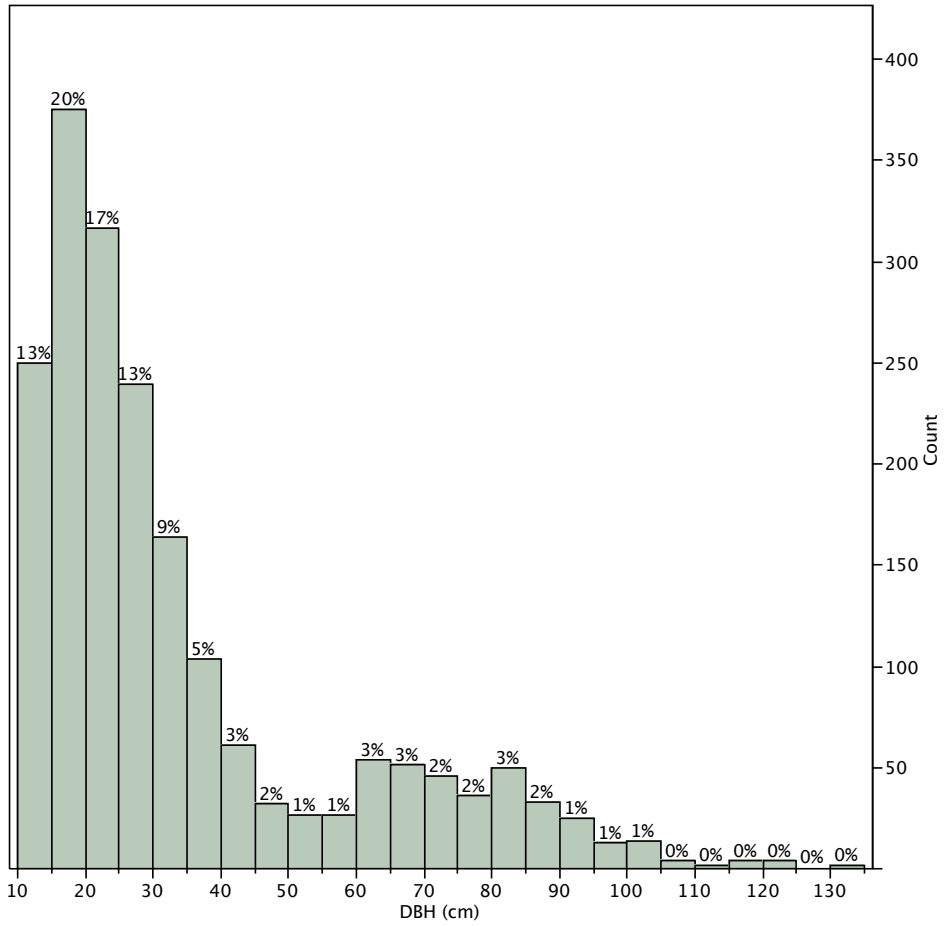

Figure 13: Distribution of the FIA subplot tree DBH measurements for the entire study area.

Table 9: Tree information FIA location 5992, subplot 1.

TREE	SPECIES	DIA (cm)	HT (m)	JTAGB (kg)	VOLGRS (ft^3)
100	*Pinus ponderosa*	12.95	3.95	40.49	0.74
101	*Pinus ponderosa*	13.21	4.03	42.46	0.61
102	*Pinus ponderosa*	27.43	8.36	251.66	8.86
103	*Abies grandis*	33.27	10.14	472.65	26.74
104	*Abies grandis*	20.83	6.35	147.80	6.12
105	*Pseudotsuga menziesii*	36.83	11.23	721.71	27.96
106	*Pinus ponderosa*	85.34	26.01	3990.40	257.66
107	*Pinus ponderosa*	87.12	26.55	4195.85	287.54
108	*Pinus ponderosa*	84.33	25.70	3875.71	253.63
109	*Pinus ponderosa*	104.65	31.90	6556.10	464.05
120	*Pinus ponderosa*	92.46	28.18	4849.07	275.09
n = 11				**25143.90**	**1609.00**

Table 10: Tree summary FIA location 6281, subplot 1.

TREE	SPECIES	DIA (cm)	HT (m)	JTAGB (kg)	VOLGRS (ft^3)
100	*Pinus ponderosa*	15.24	4.65	60.15	1.29
101	*Pseudotsuga menziesii*	19.56	5.96	153.71	4.53
104	*Pseudotsuga menziesii*	75.44	22.99	4161.42	168.70
105	*Pinus ponderosa*	84.84	25.86	3932.81	242.36
123	*Abies grandis*	103.12	31.43	7825.96	358.05
124	*Pseudotsuga menziesii*	81.53	24.85	5031.60	216.72
n = 6				**21165.65**	**991.65**

The stepwise selection method was used to identify prediction variables to be included in multiple linear regression models for each of the point cloud-based lidar metric groups. A probability threshold of 0.10 was used for a variable to enter the model as well as the to stay in the model. If more than one predictor variable was included in the final model, variance inflation factors (VIFs) were calculated to check for the presence of multicollinearity. Predictor variables with VIFs greater than five were considered an indicator of multicollinearity in the model.

The stepwise selection method described above, produced only two models, one for biomass and one for volume. The only predictor variables selected using this methodology were the 100^{th} height percentile, and the 75^{th} percentile. Both models were based on all 56 subplots. The resulting R-square values for the model of biomass and the model of volume were low, accounting for only 29% and 25% of the variance, respectively (Table 11). VIFs were calculated for each model to check for multicollinearity. An initial plot containing scatter plots for all of the height percentile predictor variables identified linear relationships between the variables. However, the resulting VIFs of 3.61 for both predictor variables in both models suggested that

multicollinearity between the selected predictor variables did not contribute to the increased R-square values of the multiple regression models.

Table 11: Point cloud-based multiple linear regression analysis results.

	Dependent variable	
	Biomass (p75, p100)	Volume (p75, p100)
R^2	0.29	0.25
Adj-R^2	0.27	0.22
RMSE	3854.46	6.43
β_0	-2108.72	-3.67
P-value	0.1912	0.1728
β_1 (p75)	-554.59	-0.80
P-value	0.0023	0.0080
VIF	3.61	3.61
β_2 (p100)	644.42	0.95
P-value	<.0001	0.0002
VIF	3.61	3.61

2.3.2 Plot-level Point Cloud-Based Airborne Lidar Metrics

Individual simple linear regression models were created for the plot-level aboveground biomass and gross volume estimates using each of the point cloud-based metrics. The best predictor variables from each point cloud-based metric set were the 95th percentile (p95), variable height bin two (vhb2), static height bin six (shb6), and density six (d6, Table 12 and Figure 14 to Figure 17). Examination of the best resulting model from each point cloud metric set, showed three models with R square values between 0.56 and 0.71. This is a large increase in R square values from the subplot-level point cloud metrics (section 2.3.1). The four best models were also produced using

different predictor variables than those used for the best models for the subplot-level point cloud metrics. All of the best models are significant at the $\alpha = 0.1$ level, except for the model using variable height bin two, which had R-square values of 0.10 for plot-level biomass and 0.09 for plot-level volume.

Table 12: Simple linear regression analysis results for the four best plot-level point cloud predictor variables.

				Parameters and p-values			
Point cloud-based metric	Dependent variable (plot-level)	R^2	RMSE	β_0	P-value	β_1	P-value
Height Percentiles							
p95	Biomass	0.56	1967.88	-10603.54	0.0132	655.77	0.0020
p95	Volume	0.67	2.71	-19.46	0.0022	1.13	0.0004
Variable Height Bins							
vhb2	Biomass	0.10	2815.28	-317.69	0.9268	11249.17	0.2631
vhb2	Volume	0.09	4.49	-0.76	0.8908	16.62	0.2977
Static Height Bins							
shb6	Biomass	0.60	1876.92	1455.60	0.0610	86476.42	0.0011
shb6	Volume	0.71	2.52	1.35	0.1793	149.07	0.0001
Density							
d6	Biomass	0.60	1876.99	1459.08	0.0603	86409.69	0.0011
d6	Volume	0.71	2.52	1.35	0.1772	148.96	0.0001

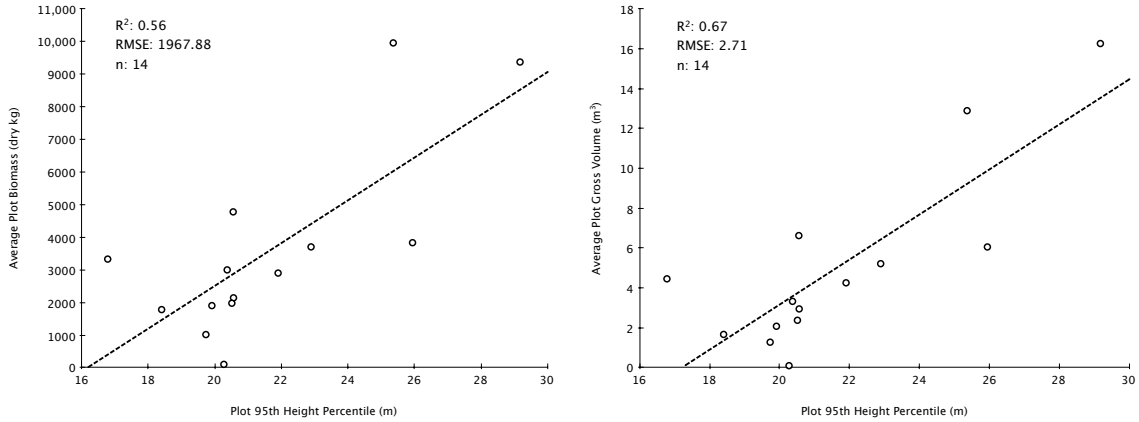

Figure 14: Scatter plots of plot-level aboveground biomass and gross volume vs. lidar-derived 95th height percentile metric.

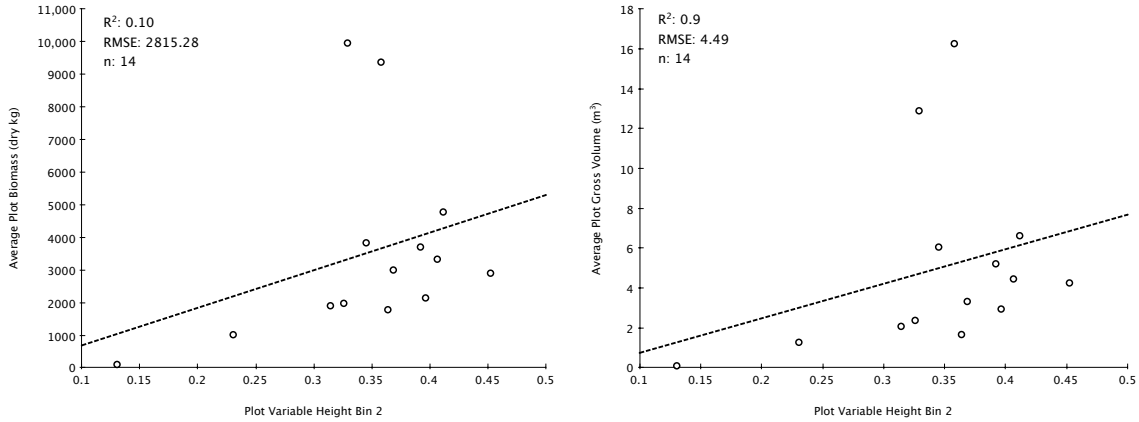

Figure 15: Scatter plots of plot-level aboveground biomass and gross volume vs. lidar-derived variable height bin 2 metric.

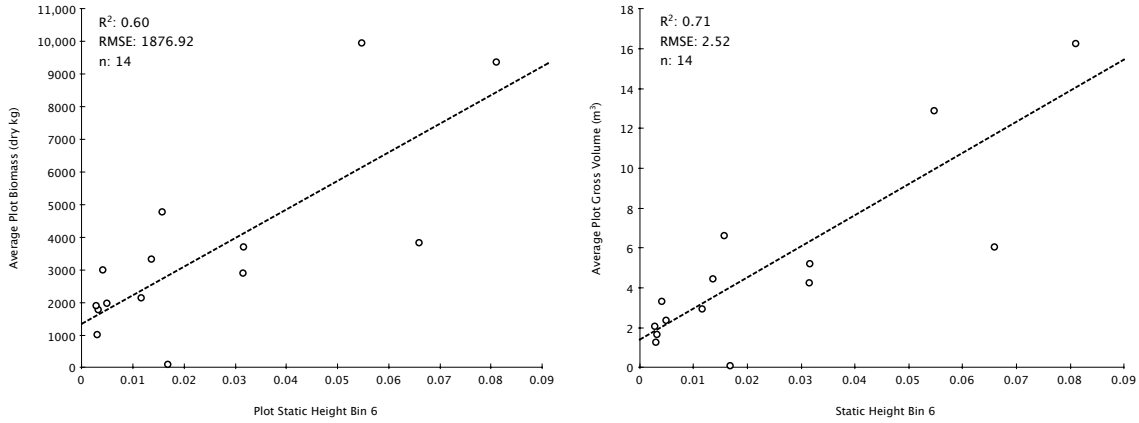

Figure 16: Scatter plots of plot-level aboveground biomass and gross volume vs. the lidar-derived static height bin 6 metric.

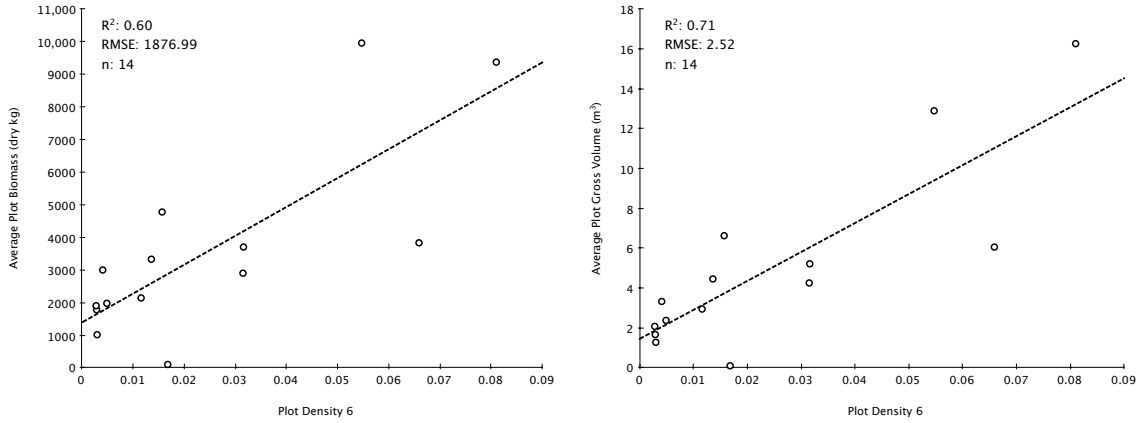

Figure 17: Scatter plots of plot-level aboveground biomass and gross volume vs. the lidar derived density 6 metric.

As previously mentioned, the large plot size used to calculate the plot-level point cloud metrics resulted in simple linear regression models with higher R-square values than the models produced using subplot-level point cloud metrics, by reducing the effects of geolocation error. The larger plot size also helped to mitigate edge effect,

where only a portion of ALS returns were utilized for trees with stems located at the edge of a subplot.

The stepwise selection method was used to identify prediction variables to be included in multiple linear regression models for each of the plot-level point cloud-based lidar metric groups. A probability threshold of 0.10 was used for a variable to enter the model as well as the to stay in the model. If more than one predictor variable was included in the final model, variance inflation factors (VIFs) were calculated to check for the presence of multicollinearity. Predictor variables with VIFs greater than five were considered an indicator of multicollinearity in the model. This selection method produced no models containing more than one predictor variable when plot-level point cloud-base lidar metrics were used.

2.3.3 Subplot-Level Individual Tree Metrics

Simple linear regression models using the predictor variables of average tree height, average crown width, and number of trees also produced poor results. When all 56 subplots were included in the model, the best predictor variable identified was number of trees. Models created using this predictor variable yielded extremely low R-square values ($R^2 = 0.06$ for biomass and $R^2 = 0.06$ for volume). With the two points of interest removed from the data set, the R-square values of all models increased, and the best predictor variable became average height ($R^2 = 0.14$ for biomass and $R^2 = 0.17$ for volume, Table 13).

Table 13: Subplot-level individual tree simple linear regression results for best predictor variables.

Individual tree metric	Dependent variable	R^2	RMSE	Parameters and p-values			
				β_0	P-value	β_1	P-value
num_trees	Biomass	0.06	4416.76	1712.20	0.1545	253.37	0.0782
num_trees	Volume	0.03	7.23	2.63	0.1806	0.32	0.1700
ave_ht (reduced)	Biomass	0.14	2275.15	768.24	0.3307	183.19	0.0059
ave_ht (reduced)	Volume	0.17	3.25	0.43	0.7010	0.30	0.0019

Stepwise multiple regression analysis was also performed, using the same procedure that was utilized in section 2.3.1. The stepwise selection method only selected one predictor variable (number of trees), so no multiple linear regression models were generated from the subplot-level individual tree metrics.

2.4 Conclusions

This study represents an initial attempt to model biophysical parameters of interest to the FIA program utilizing the standard FIA plot design, data from FIA ground crews, and ALS data. As previously mentioned, other studies have successfully modeled similar forest parameters with ALS data, but have done so under conditions different than those present in the Malheur National Forest. Issues such as the presence of a large number of tree species, complex terrain, dense forest conditions, and multi-story forest canopies, increased the difficult of modeling forest biophysical parameters in this study.

The area-based subplot-level point cloud and individual tree ALS metrics used in this study were shown to have a poor relationship to subplot estimates of biomass and volume estimated from FIA measurements. While the larger plot-level point cloud metrics exhibited a stronger relationship with the estimates of biomass and volume

produced using FIA measurements, resulting in models that could account for approximately 71% of the variation present in the data. Even though the R-square values for models using the plot-level predictor variables are lower those reported by previous studies, e.g. Lim and Treitz 2004, they represent a large improvement from the models developed using subplot-level point cloud and individual tree ALS metrics.

The poor results produced from the subplot-level and individual tree ALS metrics helped identify several sources of error that should be taken into account in future research, such as: (1) geolocation error for subplot center coordinates; (2) edge effect. Other possible sources of error, such as FIA measurement date and the large variety of tree species, could have negatively affected study results. For example, 44 of the subplots used in this study were measured by FIA crews at least two years before the lidar data were collected. Future research should utilize larger plots to reduce errors caused by geolocation and plot edge effect, as well as regional species-specific growth and yield models to mitigate error caused by out-of-date tree measurements.

The overlap of tree crowns within the complex multi-story forest canopy, coupled with the large number of tree species with crown morphologies differing from the southern pine species used to create TreeVaW most likely reduced the software's effectiveness. This problem was apparent when TreeVaW was unable to produce crown radii estimates for approximately 46% of the trees it identified. Future attempts to use TreeVaW to model forest biophysical parameters could possibly reduce these problems by manually matching TreeVaW identified trees, rather than averaging heights and

crown widths for trees identified within a subplot. Methods such as this could follow a methodology similar to Popescu (2007).

3 COMPARISON OF TERRESTRIAL LIDAR METRICS AND GROUND-BASED ESTIMATES OF TOTAL ABOVEGROUND BIOMASS AND VOLUME

3.1 Introduction

Light detection and ranging (lidar) is a laser-based, active remote sensing system, which collects ranging data utilizing the known speed of light and information about the flight time of a laser pulse (Lim et al., 2003). In this context, flight time refers to the time it takes for a given laser pulse to travel from a system, reflect off of an object, and return back to the system. A wide variety of lidar systems currently exist, and data has been successfully collected utilizing systems mounted to space-borne, aerial, and terrestrial (tripods or vehicle-based) platforms. Over the past several decades the use of lidar remote sensing data in forestry has seen steady growth. The increased use of lidar systems to acquire data over forested areas can be attributed to their ability to cover extents of local or regional scales and accurately quantify the three-dimensional vertical structure of the forest.

Measurements from terrestrial laser scanners (TLS) have been used for a relatively short time, starting in late 1990s. TLS systems are mounted on a terrestrial-based platform (e.g. tripod, motor vehicle) and in most cases require a portable computer and several large batteries or a portable generator to operate. Systems such as these are capable of acquiring and merging scans from multiple locations through the use of at

least two stationary targets. This process requires more time scanning an area, but has the potential to provide a more complete picture of a study area.

A number of studies have provided methodologies for deriving forest measurements using data collected by TLS systems. Hopkinson et al. (2004) isolated individual trees for height and DBH measurement in merged lidar scans with two distinct forest types. Thies et al. (2004) merged lidar scenes, and produced detailed stem measurements of several large-diameter deciduous trees. A method for automatically identifying individual trees, measuring tree height, and tree DBH is presented in Maas et al. (2008). Henning and Radtke (2006) developed methods for identifying trees scanned with a TLS system and measuring stem diameter. Their results showed the average error between the lidar-derived diameter estimates and caliper measurements, for sections below the base of live crown, to be less than 1 cm. Methods of deriving other forest measurements, such as aboveground biomass, have also been explored. Lefsky and McHale (2008) used high-density point cloud data for multiple urban trees with complex architecture in an attempt to develop allometric relationships for predicting species tree volume.

In the United States, the Forest Service (USFS) Forest Inventory and Analysis (FIA) program provides a diverse selection of data used to assess the status of the nation's forested areas. In the past, the FIA program used a periodic inventory system, where measurements on non-national forests were collected on a state-by-state basis in predetermined zones, and lead to inventory cycles of ranging from six to eighteen years (Gillespie, 1999). In 1998 legislation was passed (see the Agricultural Research,

Extension, and Education Reform Act of 1998) that mandated the entire FIA program implement an annual inventory. This inventory method requires that the collection, analysis, and reporting of data at a state-level be completed every five years, meaning that under ideal conditions, 20% of plots in each state would be measured each year (Gillespie, 1999). The annualized FIA program allows for the collection of a variety of parameters of interest and consists of three phases: (1) remote sensing to identify forested and non-forested areas; (2) field samples located at intervals of about 1 plot every 6,000 acres, where forested sample areas are visited by field crews to collect ground measurements and non-forested areas are visited to quantify the frequency of variables such as land use change; and (3) consists of visiting a subset of the plots in phase 2 (about 1 plot every 96,000 acres) to collect more detailed measurements (e.g. complete vegetation inventory, tree and crown condition, soil data) during the growing season (USFS, 2008).

The measurements collected by the FIA program can be scaled up to provide information about forest populations by aggregating plot statistics for specific populations. However, this is only possible if the population(s) of interest have been adequately sampled by the inventory. Many regional to national scale biomass and carbon budgets for the United States are based largely on the forest information provided by the FIA program, regional-level volume and biomass equations, and national-level allometric equations (Heath et al. 2008). Heath et al. (2008) also notes that in recent years the FIA program has seen a continual increase in requests, from forest resource managers and researchers, for biomass, carbon, and volume information. The direct link

between data provider and end user makes the FIA program responsible for many of the volume estimates, biomass budgets, and carbon budgets created for the United States.

This overall objective of this study is to provide a novel approach for estimating aboveground biomass and gross volume at the FIA subplot-level, using height and distance-based TLS point cloud metrics. Subplot biomass and volume estimates derived from ground-based FIA measurements and allometric equations will be compared to a number of subplot TLS point cloud metrics. This will address the hypothesis that because data collected by a TLS system are capable of describing the three dimensional vertical structure of the forest, they can be used to estimate forest biophysical parameters of interest such as volume and aboveground biomass. The specific objectives of this study include:

1. Development of a methodology to derive height and distance-based terrestrial lidar metrics related to forest biophysical parameters at the FIA subplot-level.

2. Utilize simple linear and multiple linear regression analysis to help identify relationships between the lidar metric sets and FIA subplot estimates of forest biomass and volume calculated from FIA data.

3.2 Materials and Methods

3.2.1 Study Area

The study area for this project is in the Malheur National Forest located in eastern Oregon, and covers roughly 105,936 hectares (Figure 18). Elevation ranges from 1,236 to 2,593 m, and slope varies from 0 to ~ 86 degrees. The general location of the

study area, as defined by the NE and SW corners of a rectangle, is Universal Transverse Mercator (UTM) Zone 11N 383297.6E, 4905767.9N and UTM Zone 11N 333344.5E, 4863102.6N. The site was selected because of access to FIA ground measurements and plot locations, and the presence of a wide variety of forest conditions, such as slope and tree species. The forests located within the study area are composed of mostly pine species, including: Ponderosa pine (*Pinus ponderosa*), Douglas-fir (*pseudotsuga menziesii*), western larch (*Larix occidentalis*), and grand fir (*Abies grandis*).

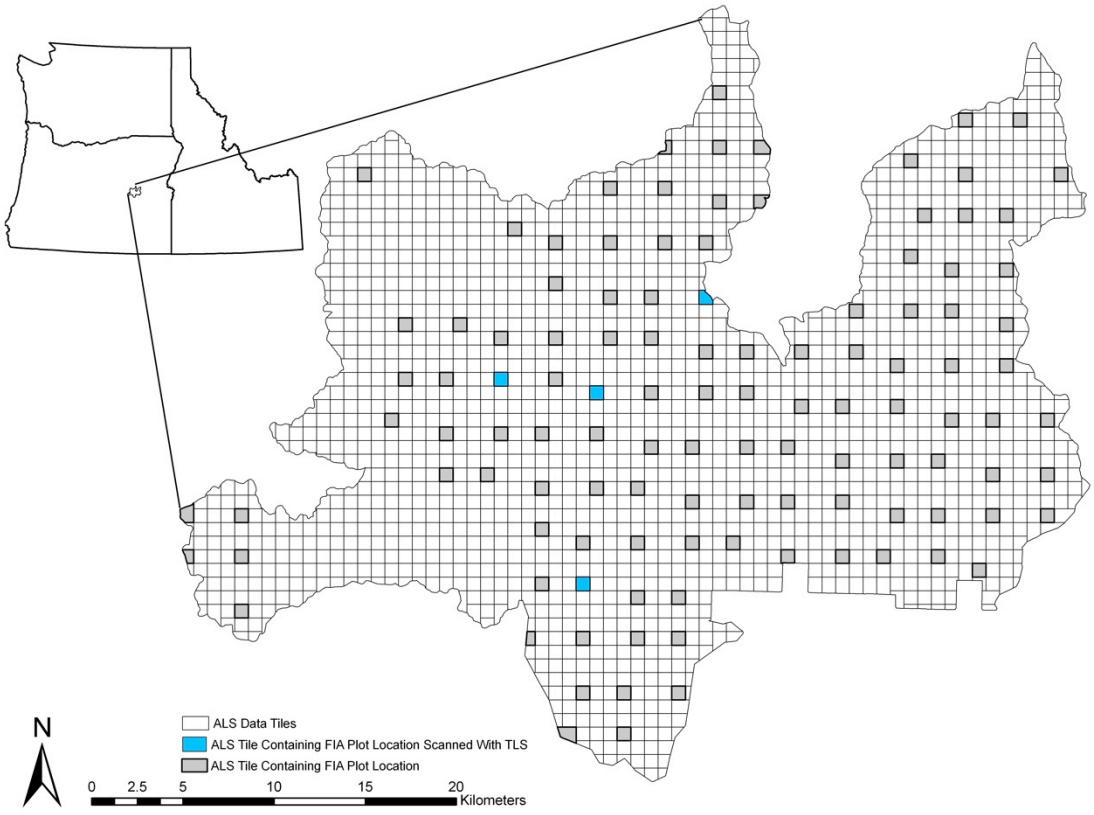

Figure 18: Malheur National Forest study area in eastern Oregon. Hollow squares represent individual airborne lidar tiles. Blue squares represent tiles containing an FIA plot location where at least one subplot was scanned with the terrestrial lidar scanner. Grey squares signify tiles containing an FIA plot location.

3.2.2 Data

This section of the study used several types of data to describe forest conditions at FIA plot locations in the study area including: (1) FIA ground crew *in situ* measurements; and (2) single return, small footprint, terrestrial laser scanners (TLS).

3.2.2.1 Forest Inventory and Analysis Data

The USFS provided FIA data for all FIA locations within the study area (91 locations, 364 subplots, and about 2,477 trees). This study will focus on FIA subplots since they are utilized in every FIA region. Each FIA location contains four circular ~0.04 acre subplots (radius = 7.32 m). Subplot one is centered over the plot center for the entire FIA location. Subplots two, three, and four are located 36.58 m from the center of subplot at azimuths of 360°, 120°, and 240°, respectively (Figure 2.2, section 2.2.2.1). Individual trees are measured and recorded if they are located within the boundaries of subplot and have a DBH or diameter at root collar greater than 12.7 cm. Measurements collected for each of these trees include: DBH, height, tree condition (live/dead), crown class (open grown, dominant, codominant, intermediate, or overtopped), species, species group, azimuth to plot center, and distance to plot center.

The FIA program provided estimates of volume for the majority of the trees in the study area. Estimates were only omitted for trees with a status code that listed the tree as dead, or for trees where the status code was completely absent. The regional equations used to calculate tree volume estimates can be found in Zhou and Hemstrom (2010). Estimates of individual tree total aboveground biomass were to also be included with the FIA data. However, the current high workload of the FIA data steward, and the

time constraints of this study did not make this a feasable option. The national-scale total aboveground biomass equations, presented in Jenkins et al. (2003), were used to estimate the total aboveground biomass for individual trees in lieu of the estimates from regional FIA equations. Plot-level estimates for total aboveground biomass and volume were calculated by summing the total aboveground biomass and volume estimates for all of the trees in each subplot.

3.2.2.2 Terrestrial Lidar Data

Terrestrial lidar data were collected for nine FIA subplots at four FIA plot locations (Table 14 through Table 19). Data were not collected on seven of the 16 subplots. Several issues were encountered that required the collection of data from only a subset of the available subplots, including: (1) finite amount of power available for operating the scanner; (2) major obstruction of the TLS field of view (FOV), by vegetation near or on subplot center (Figure 19 and Figure 20); or (3) inability to locate a FIA subplot center monument. Data were collected using a Leica ScanStation 2 (Figure 21). This system is capable of recording 50,000 pulses per second, has a maximum range of 300 m, a minimum point spacing of 1 mm apart (horizontally and vertically), and can collect data 360° horizontally and 270° vertically. For this study the scanner was placed directly over the center of an FIA subplot (marked by a permanent monument), and performed a full 360° by 270° scan, horizontally and vertically, (from here on referred to as a 360° scan) with a point density of 10 cm by 10 cm at a distance of 50 m.

Table 14: Data collection years for selected subplots.

	Sample year			
	2003	2005	2006	2008
Number of subplots	2	1	3	3

Table 15: Selected subplot crown class frequencies.

		Crown Class				
Species	**n[a]**	**OG[b]**	**D[c]**	**CD[d]**	**I[e]**	**OT[f]**
Abies grandis	33	1	10	8	11	3
Pinus contorta	11	0	4	5	2	0
Pinus ponderosa	38	0	7	25	4	2
Pseudotsuga menziesii	1	0	0	1	0	0
Cerocarpus leditolins	8	0	0	1	7	0

[a] Number of trees
[b] Open grown crown class
[c] Dominant crown class
[d] Codominant crown class
[e] Intermediate crown class
[f] Overtopped crown class

Table 16: Descriptive statistics for tree DBH.

		Diameter at breast height (cm)			
Species	**n[a]**	**mean**	**min - max**	**SD[b]**	**CV(%)[c]**
Abies grandis	33	30.08	13.21 - 70.36	15.46	51.41
Pinus contorta	11	21.17	16.51- 29.97	3.97	18.73
Pinus ponderosa	38	27.41	12.70 - 101.30	21.99	80.24
Pseudotsuga menziesii	1	14.73	NA	NA	NA
Cerocarpus leditolins	8	16.95	13.72 - 22.35	3.17	18.71

[a] Number of trees
[b] Standard deviation
[c] Coefficient of variation percentage

Table 17: Descriptive statistics for tree height.

		Height (m)			
Species	**n[a]**	**mean**	**min - max**	**SD[b]**	**CV(%)[c]**
Abies grandis	33	17.64	7.62 - 27.44	6.06	35.38
Pinus contorta	11	14.77	13.11 - 18.90	1.94	13.13
Pinus ponderosa	38	14.51	7.62 - 38.72	7.62	52.52
Pseudotsuga menziesii	1	14.02	NA	NA	NA
Cerocarpus leditolins	8	6.17	4.27 - 8.54	1.35	21.89

[a] Number of trees
[b] Standard deviation
[c] Coefficient of variation percentage

Table 18: Descriptive statistics for estimated subplot biomass and volume.

	n[a]	**mean**	**min - max**	**SD[b]**	**CV(%)[c]**
Biomass (dry kg)	9	3884.77	1141.28 - 7089.45	1888.1	48.6
Volume (m^3)	9	5.55	1.27 - 9.36	2.92	52.6

[a] Number of plot
[b] Standard deviation
[c] Coefficient of variation percentage

Table 19: Terrestrial lidar scanning summary.

FIA location 5992				
Subplot #:	1	2	3	4
TLS Scan:	No	Yes	No	Yes
FIA location 5993				
Subplot #:	1	2	3	4
TLS Scan:	Yes	No	No	No
FIA location 5708				
Subplot #:	1	2	3	4
TLS Scan:	Yes	No	Yes	Yes
FIA location 6185				
Subplot #:	1	2	3	4
TLS Scan:	No	Yes	Yes	Yes

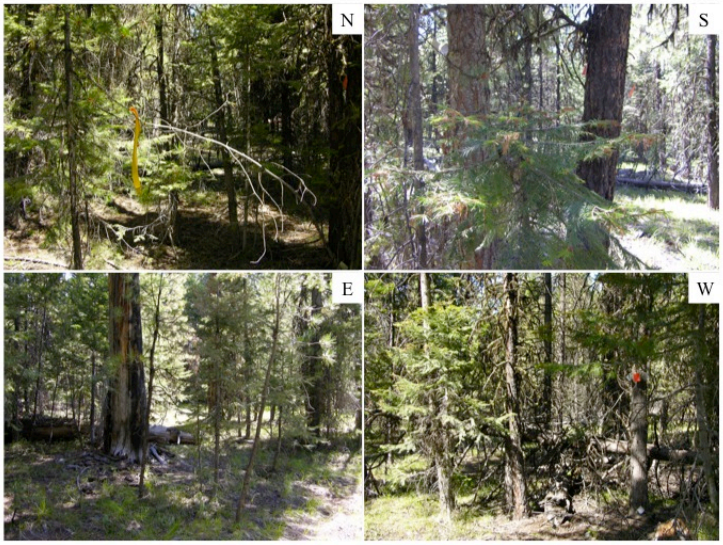

Figure 19: Example of a subplot not selected for scanning because of high levels of obstruction close to the center of the plot when facing North or South from subplot center.

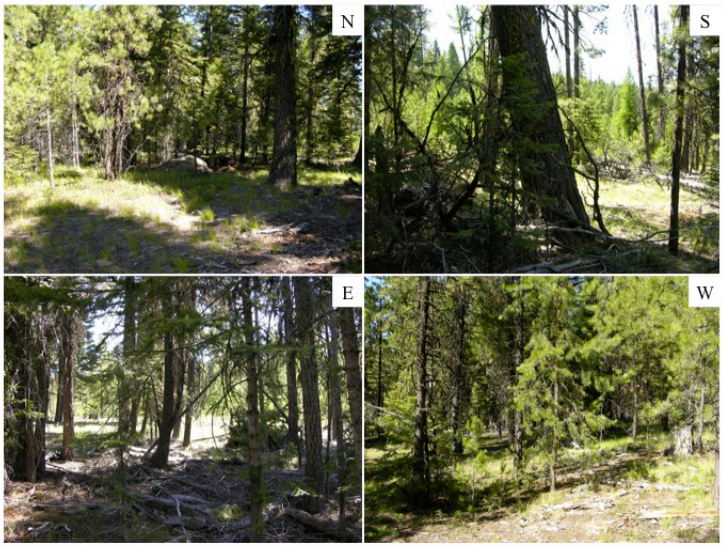

Figure 20: Example of selected subplot with area near subplot center clear of obstructions.

Figure 21: Leica ScanStation 2 terrestrial lidar scanner, located over the center of an FIA subplot, collecting data with a 360° scan.

3.2.3 Point Cloud-based Terrestrial Lidar Metrics

3.2.3.1 Visual Summary of Data Processing

A flowchart of the data processing steps used in this study is shown in Figure 22.

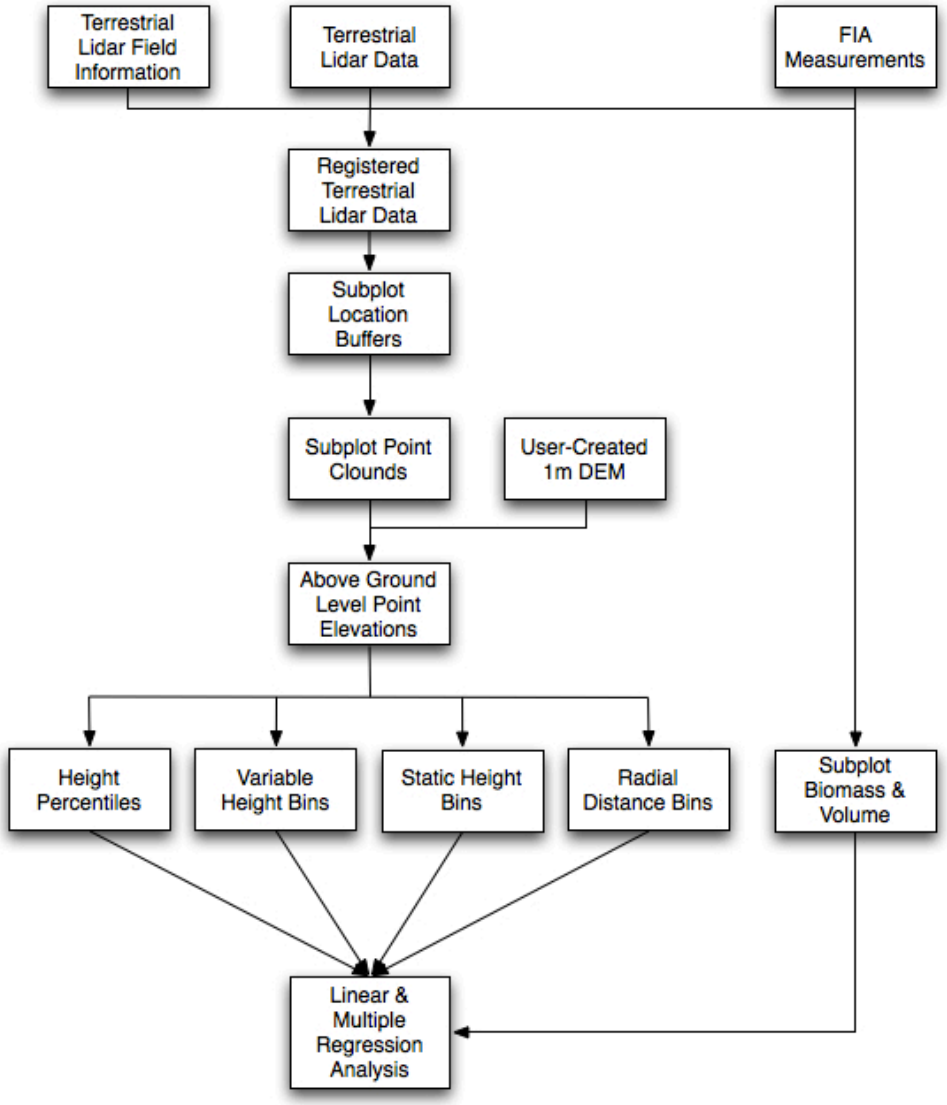

Figure 22: Flowchart of the terrestrial lidar data processing approach.

3.2.3.2 Registration of Terrestrial Lidar Data

An initial inspection of the TLS scans established that the points in each subplot were not correctly georeferenced. Instead, each point was assigned coordinates relative to the scanner location of (0.0, 0.0, 0.0). In the Leica software package Cyclone, a point cloud can be easily registered if the x, y, and z coordinates for one point are known, and an azimuth to a second point is also known. This information was available for each scan because the coordinates for the scanner location (FIA subplot center) and the azimuth from the scanner to a reference target were recorded at each scan location. However, there is no point in the point cloud, which represents the scanner. A "scanner point" had to be manually added at (0.0, 0.0, 0.0) before the registration could take place. A third piece of information, critical to the registration process, was the height of the scanner above the ground. Without this information, a systematic error in the height values of each point would be present after registration. At the time of data collection, scanner height measurements were recorded for all but one subplot. These values were added to their corresponding scanner location elevation (z value). The eight subplots with sufficient reference information were registered successfully.

To accurately register the subplot with the missing scanner height record, a scanner height estimate needed to be produced. An old unpaved forest service road within the scan provided a easily identifiable bare Earth surface that could be used in conjunction with a spatial coincident 1m ALS-derived DEM to estimate scanner height. In order to compare the two datasets, the TLS data was registered without any scanner height correction and overlayed on the ALS DEM. An estimate of the systematic error

was generated from measurements of the difference between the DEM and TLS points for 14 locations, at one-meter intervals, along a 15 m height profile transect created in Quick Terrain Modeler (QTM). An average of the 14 measurements was taken and provided a mean height correction of 2.25 m with a standard deviation of 0.04 m. The scanner height estimate was applied to the subplot data, and a correctly registered point cloud was created (Figure 23).

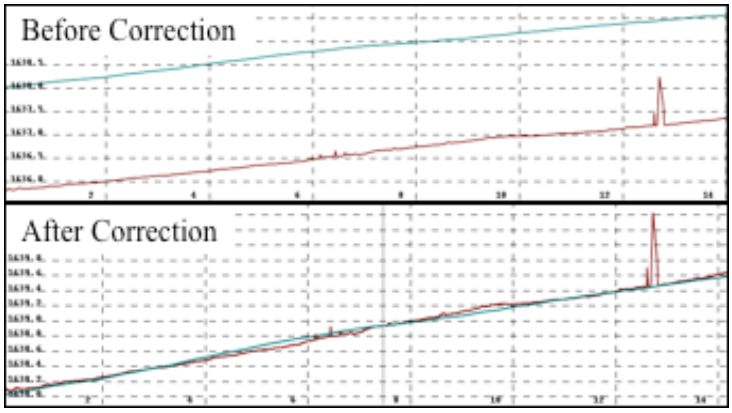

Figure 23: QTM generated transect height profile. The blue line represents the 1m ALS-derived DEM, and the red line represents the registered TLS data missing scanner height correction.

3.2.3.3 Extraction of Subplot Point Clouds

Coordinates for the center of each subplot were included in the data provided by the FIA program. These coordinates were collected during field measurements using a WAAS (wide area augmentation system) enabled Trimble GPS unit, and then post-processed using base station data. Proprietary Trimble software was utilized to compute the average accuracy of the reported coordinates. The average accuracy values are computed to represent a 95th percentile three-dimensional distance threshold around the

true subplot center location. The resulting point data were provided as a shapefile, to allow for visualization and manipulation in ArcMap.

The TLS-scanned subplots from each of FIA location were selected and individually extracted from the provided shapefile. This process created a total of nine shapefiles, each containing a single point, which represented the subplot center coordinates for the selected subplot. A circular buffer (r = 7.32 m) was created around each subplot center. This process provided polygon shapefiles covering the spatial extent of each subplot. To create individual subplot point clouds, the points within each subplot were extracted from the corresponding subplot scan.

3.2.3.4 Calculation of Above Ground Level Elevations and Removal of Ground

All point cloud metrics were calculated with above ground level (AGL) elevation values. QTM was used to compute AGL values for each subplot point cloud by subtracting the DEM values from the corresponding point elevations.

When considering TLS systems from an operational standpoint, locations where data are collected may not always have spatially coincident ALS data, and as such, users may not have access to a high resolution DEM. This hypothetical situation establishes the need to generate a DEM from the data collected by a TLS system. Several characteristics of TLS systems lend support to the idea of creating a DEM from TLS data, such as: (1) current TLS systems have the potential to collect data with much higher resolutions than ALS systems; and (2) TLS systems collect data from underneath a forest canopy, which allows the system to collect a larger number of ground returns for the area directly surrounding the scanner than an ALS system.

As previously mentioned, several characteristics of TLS systems provide a substantial increase in the amount of ground points collected by the system. The increased amount of available data can be both beneficial and detrimental. For instance, the high resolution of the scan provides the ability to create higher resolution DEMs, capable of capturing minute variations in the terrain, commonly generalized by the courser resolution DEMs generated from ALS data. Conversely, the high resolution of the terrestrial scans, often means that points close to the ground are collected for grasses, forbs, course woody debris, or other non-ground materials within the subplot. While information such as this could be useful for some studies, it has the potential to cause spurious results when generating a DEM.

For this study, the *AGL Analyst* tool in the Quick Terrain Modeler software package will be used to derive an estimated DEM for each subplot point cloud. Several limitations, stemming from the software and the mechanics of the TLS system itself were recognized during this process. The most restricting limitation of QTM's *AGL Analyst* tool was the 1m-resolution threshold for DEM creation. Since a finer resolution, better suited for the high resolution TLS data, is not allowed in QTM only a generalized DEM was produced. A positive aspect of this generalization was the previously mentioned effect of returns from non-ground materials was reduced. The limitation imposed by the TLS system stems from its limited vertical FOV (Figure 24), which produced a small circular section in each scan where no data are collected (Figure 25). Areas of no data can be detrimental to DEM creation, since the surface of the Earth must be interpolated from the available data surrounding the area. In some cases, overhang

from tree crowns caused large spikes in the DEM. These spikes were rectified using the *Smooth Area* tool, which used information from surrounding cells, selected by a user-defined polygon, to remove the anomalous spikes. The final TLS-derived DEMs for each subplot were subtracted from the corresponding subplot point cloud data to obtain AGL height values.

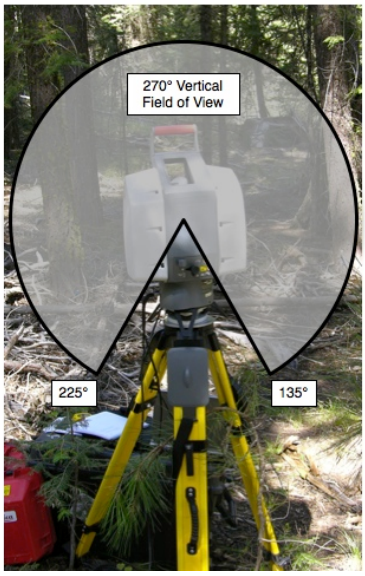

Figure 24: Conceptual illustration of the limited vertical field of view (270°) of the ScanStation 2.

Figure 25: Visualization of area of no data collection as a result of the limited vertical field of view.

A search of the literature reveals that when using point cloud metrics, points under a specified height are often excluded from an analysis. The exclusion of points in this manner helps remove possible effects of shrubs, large rocks, and other ground-related laser returns from lidar data. Mitigating these effects reduces the overall amount of data, but helps to insure that the majority of the remaining data represent the objects of interest, e.g. trees. Since the TLS system is capable of collecting a large amount of returns from the ground and vegetation other than trees, the removal of ground points is of even greater importance. A height cut off value of 0.5 m (AGL) was selected. All points with an elevation of less than this value were cropped from each AGL subplot point cloud.

3.2.3.5 Height Percentiles

Height percentiles were calculated for each subplot. A similar approach has been used with ALS data to relate it to forest biophysical parameters (Holmgren, 2004; Lim and Treitz, 2004; and Patenaude et al. 2004). For this study, mean height, and the 25^{th}, 50^{th}, 75^{th}, 90^{th}, 95^{th}, and 100^{th} height percentiles were calculated for each subplot point cloud.

3.2.3.6 Variable Height Bins

Height bins were used to separate the vertical space within a subplot into five equal-spaced intervals. The variable height bin approach calculates height bins by finding the distance between highest and lowest points in a plot, and dividing this distance by the number of desired bins (in this case five). The approach used to calculate

bins size was similar to the method used by Næsset (2007) to determine bins intervals used when calculating canopy densities from ALS data. An attractive element of this approach is it permitted a different height bin interval to be calculated for each subplot. After the height bin interval had been calculated, height bin break points were set, and the number of points within each height bin was counted. The ratio of the number of points in a height bin to the total number of points in the subplot point cloud was then calculated. This normalization procedure enables direct comparisons between the same height bin values from other subplots.

3.2.3.7 Static Height Bins

Static height bins were also used to separate the vertical space within a subplot into intervals. This height bin method was selected to emphasize differences between subplots, since it uses the same height bins regardless of the minimum and maximum subplot height values. While conceptually similar to the variable height bin approach (section 3.2.3.7), this method used constant height bin intervals for all of the subplots, which divided the vertical space into a total of six bins. The static height bin break points selected were: (shb1) 0.5 – 5 m; (shb2) 5 – 10 m; (shb3) 10 – 15 m; (shb4) 15 – 20 m; (shb5) 20 – 25 m; and (shb6) greater than or equal to 25 m. The total number of points within each height bin is counted. The counts were also normalized by dividing each count by the total number of points in the subplot.

3.2.3.8 Radial Distance Bins

Radial distance bins were created in an effort to calculate a TLS metric not related to height. This metric takes advantage of the unique data collection characteristics of a TLS system, e.g. the stationary system location, and the predefined scanning resolution. The method disregards the height information for each point in the subplot point cloud, transforming a three-dimensional point cloud to a two-dimensional data set. Linear distance of each point from the subplot center (scanner location) is calculated using the distance formula (Eq. 2). The two-dimensional subplot area (r = 7.32 m) is then separated into 8 radial bins. Bin break points were set at 1 m intervals, resulting in following radial bins: (rdb1) 0 – 1 m; (rdb2) 1 – 2 m; (rdb3) 2 – 3 m; (rdb4) 3 – 4 m; (rdb5) 4 – 5 m; (rdb6) 5 – 6 m; (rdb7) 6 – 7; and (rdb8) greater than or equal to 7 m (Figure 26). The total number of points within each of the radial distance bins was counted.

Unlike the previously mentioned height-based metrics, the counts for each radial distance bin were not normalized. The use of raw point counts is justified because of the previously described TLS system characteristics. The stationary system location during the scan, and the predefined scan resolution (set by the user), mean the total number of pulses emitted by the system at each location should be theoretically identical, and allow for direct comparisons between scans so long as scan settings remain constant. Furthermore, AGL point elevations and the removal of all points below 0.5 m ensure most ground related returns are removed from the point cloud, leaving only vegetation returns. With only vegetation returns remaining, point counts directly relate to the

amount of vegetation within the subplot, and thus can be related to biophysical parameters.

The point counts present in each of the radial distance bins are functionally related to the number a size of stems. The set resolution used for each TLS scan causes objects that are closer to the scanner to be scanned at a higher resolution than objects that are further away. Extending this principle, larger objects that are closer to the scanner will have higher point densities than smaller objects at the same distance. In some instances, lower portions of tree canopies can also be scanned with higher point densities, but since the canopy is composed of branches and foliage, numerous gaps are present. This porosity inherent in the canopy structure means that when a section of the canopy and a tree stem are equidistant from the scanner, a lower point density will be returned for the canopy than the tree stem.

Several similarities exist between the concept of TLS radial distance bins and metrics derived from hemispherical photography using programs such as HemiView. The analysis of hemispherical photography uses angles to project a hemispherical photo onto a plane. Angles are used because in the photographs the radial component of distance is related to the zenith angle (HEMIVIEW MANUAL). A similar task is performed when radial distance bins calculate distance without regard for point height, which essentially transforms the three dimensional lidar data to a two dimensional plane. A second similarity deals with how HemiView and radial distance bins separate the space within a hemispherical photo or plot. For example, when HemiView is used to calculate gap fraction for a skymap, the space is divided into radial bins as well as wedge

shaped sections defined by zenith and azimuth angle ranges, respectively. Radial distance bins perform a similar task, resulting in only radial bins based on horizontal distance.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad \text{(Eq. 2)}$$

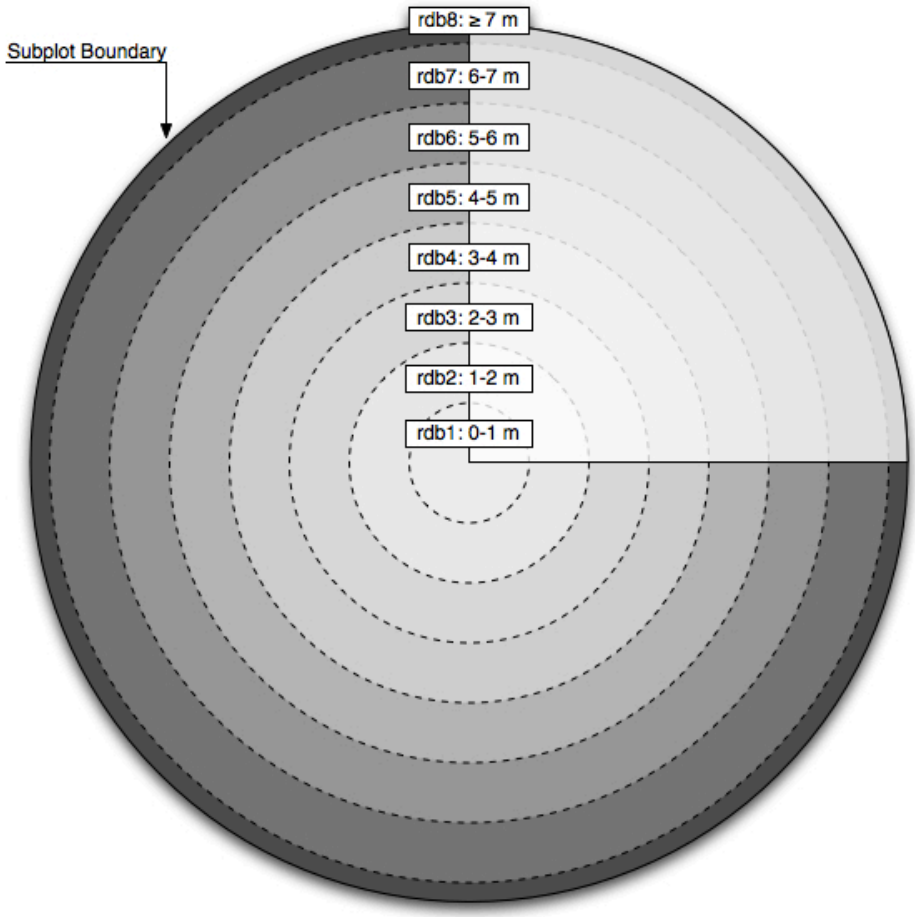

Figure 26: Conceptual example of a FIA subplot separated into radial distance bins.

3.2.3.9 Regression Analysis

Simple linear regression models were used to examine the relationship between the lidar-derived metrics and the subplot-level estimates of total aboveground biomass and gross volume. The predictor variables and methods used to calculate them were described in the previous sections. A summary of the predictor and response variables used the regression analyses can be found in Table 20. Models were run for each set of predictor variables. Information about the best model from each metric set, and the variable that produced the overall best model were provided.

Multiple regression analysis, with a mixed stepwise selection method was used to identify predictor variables to be included in multiple linear regression models for each of the point cloud-based lidar metric groups. The same analysis was used to identify predictor variables using the distance-based metric set and each of the height-based metrics sets individually (e.g. radial distance bins with height percentiles, radial distance bins with variable height bins, and radial distance bins with static height bins). A probability threshold of 0.1 was used as the probability of a variable to enter the model, as well as the probability of a variable to stay in the model. If more than one predictor variable was included on in the final model, variance inflation factors (VIFs) were calculated to check for the presence of multicollinearity. Predictor variables with VIFs greater than five were considered an indicator of multicollinearity in the model.

Table 20: Regression variables.

Sets of lidar-derived area-based predictor variables	Response variables (FIA field measurements)
Height Percentiles	Aboveground Biomass (kg)
p25, p50, p75, p90, p95, p100, mean	Gross Volume (ft^3)
Variable Height Bins	
vhb1, vhb2, vhb3, vhb4, vhb5	
Static Height Bins	
shb1, shb2, shb3, shb4, shb5, shb6	
Radial Distance Bins	
rdb1, rdb2, rdb3, rdb4, rdb5, rdb6, rdb7, rdb8	

3.3 Results and Discussion

3.3.1 Point Cloud-Based Terrestrial Lidar Metrics

Individual simple linear regression models were created for aboveground biomass and gross volume using each of the point cloud-based metrics. Examination of the resulting models showed that most lidar-derived predictor variables were poorly related to the ground-based estimates of subplot biomass and volume. The best predictor variables from each point cloud-based metric set were the 100^{th} percentile (p100), variable height bin one (vhb1), static height bin one (shb1), and radial distance bin four (rdb4, Table 21, Figure 27 through Figure 30). The best model used rdb4 as the predictor variable, and yielded R-square values of 0.40 for biomass and 0.46 for volume. It should be noted that the intercepts and slopes for the selected height percentile and variable height bin models, were not significant at the 0.1 level, providing evidence to fail to reject the null hypothesis (coefficients are equal to zero). The slopes of the selected radial height bin models were found to be significant at the 0.1 level, while p-values for the intercepts of both models did not provided enough evidence to reject the null

hypothesis the intercepts were equal to zero. The intercepts and slopes for the selected static height bin models were significant at the 0.1 level.

Table 21: TLS simple linear regression models for the best predictor variable from each metric set.

				Parameters and p-values			
Individual tree metric	Dependent variable	R^2	RMSE	β_0	P-value	β_1	P-value
Heght Percentiles							
p100	Biomass	0.11	1901.92	1714.76	0.4969	97.35	0.3784
p100	Volume	0.28	2.66	0.29	0.9344	0.24	0.1464
Variable Height Bins							
vhb1	Biomass	0.23	1768.24	1873.96	0.2520	3718.06	0.1886
vhb1	Volume	0.28	2.64	2.11	0.3775	6.36	0.1395
Static Height Bins							
sbh2	Biomass	0.41	1548.03	6793.95	0.0019	-11192.02	0.0624
sbh2	Volume	0.37	2.47	9.84	0.0033	-16.51	0.0799
Radial Distance Bins							
rdb4	Biomass	0.40	1557.48	1385.96	0.3070	0.07	0.0655
rdb4	Volume	0.46	2.29	1.42	0.4687	0.00	0.0436

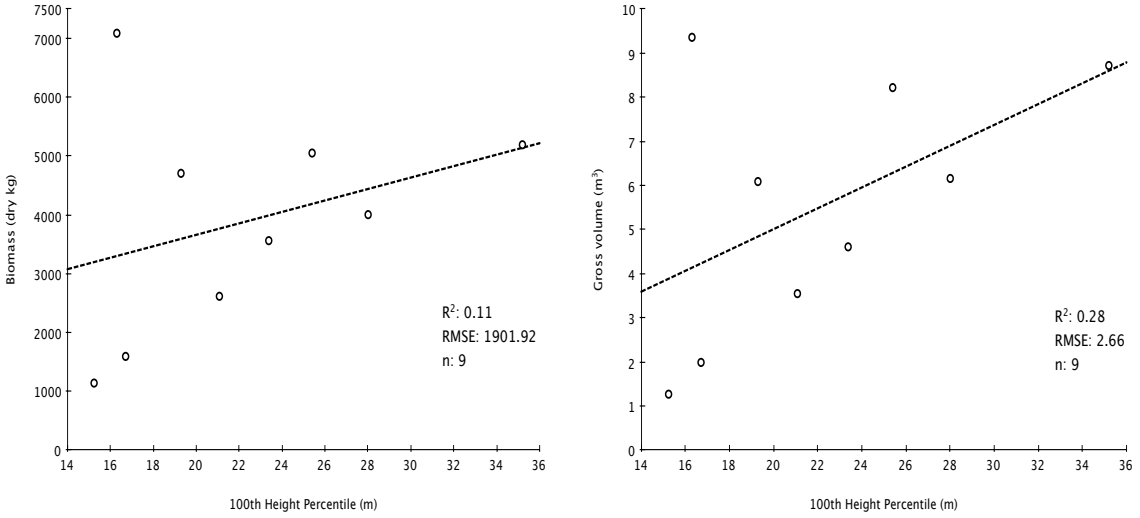

Figure 27: Scatter plots of aboveground biomass and gross volume vs. the lidar-derived 100th height percentile.

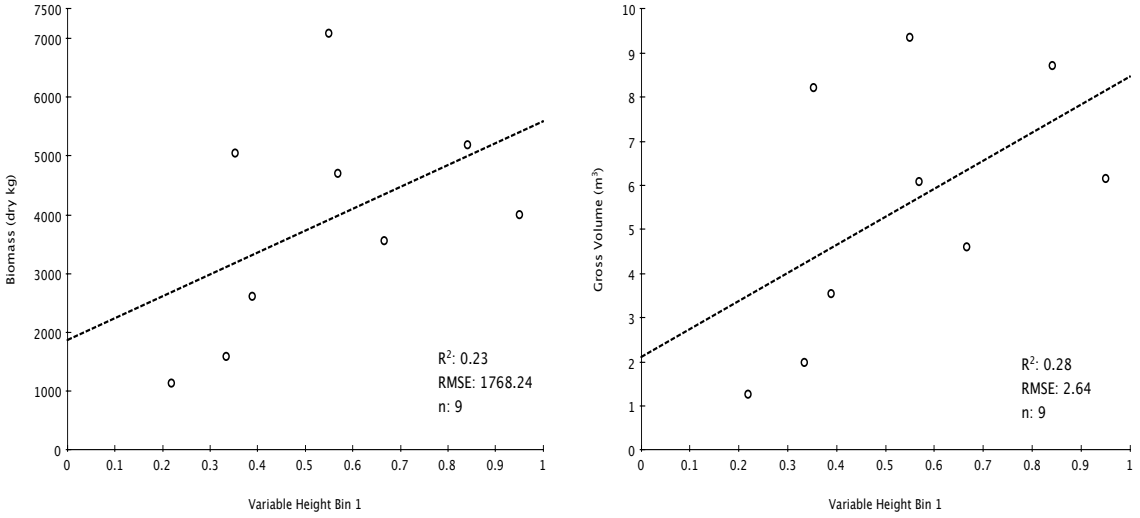

Figure 28: Scatter plots of aboveground biomass and gross volume vs. the lidar-derived variable height bin one.

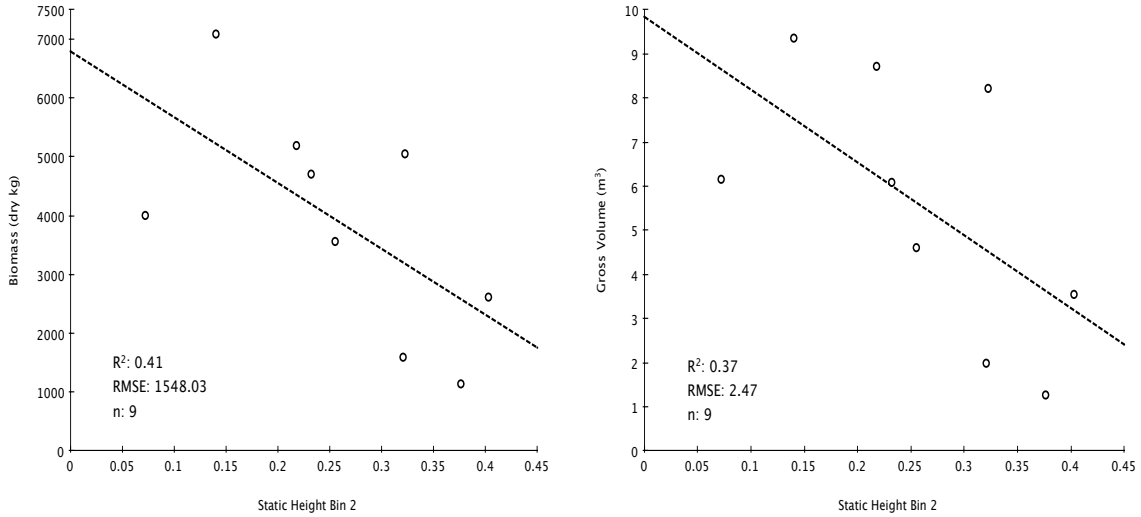

Figure 29: Scatter plots of aboveground biomass and gross volume vs. the lidar-derived static height bin two.

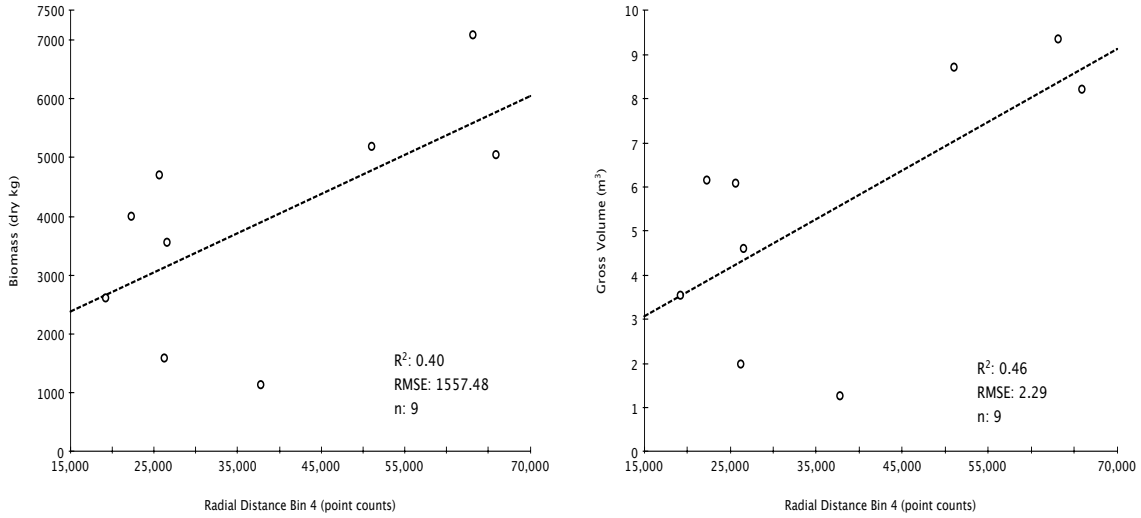

Figure 30: Scatter plots of aboveground biomass and gross volume vs. the lidar-derived radial distance bin four.

The mixed stepwise selection method used to identify predictor variables for multiple regression analysis from the same metric set, described in section 3.2.3.8, never selected more than one predictor variable. A probability threshold of 0.1 was used for a

variable to enter the model as well as and stay in the model. No variables were selected from the height percentile metrics or variable height bin metrics. Static height bin two and radial distance bin four were selected from the static height bin and radial distance bin metric sets respectively.

When the stepwise selection method was utilized to select predictor variables using the distance-based metric set and each of the height-based metrics sets individually, a total of three models were created. In two out of the three selections (radial distance bins with percentiles and radial distance bins with variable height bins) one distance-based bin and one height-based bin were selected as predictor variables for each model. However, when the stepwise selection method was used on the radial height bin metrics and the static height bin metrics, a total of four predictor variables were selected. Diagnostic plots and VIFs for each parameter in the models were examined to determine the most suitable model for predicting biomass and volume. The selected multiple regression models for predicting biomass and volume used radial distance bin four and variable height bin one, and provided R-square values of 0.71 and 0.84 respectively (Table 22 and Figure 31). VIFs for both predictor variables were only 1.01 indicating that multicollinearity was not a concern. It should be noted that the residual vs. fitted plot for the volume model (Figure 31[bottom left]) does show possible heteroscedasticity. However, the small sample size (n = 9) used in this study makes it difficult to verify assumptions used in multiple regression analysis.

The frequent selection of a radial distance bin by the stepwise selection method can be explained when the distances of trees to plot center and DBHs of trees are

analyzed. Over the nine subplots, the average distance of a tree to subplot center was approximately 4.64 m. the selected radial distance bin was radial distance bin four, which included points between three and four meters from plot center. FIA measurement protocol requires that this distance be measured from the pith (or center) of each tree to plot center. The average DBH all FIA trees in the nine subplots was 20.75 cm, which means that the average distance to the outside of the tree stems is shorter than the reported average distance and moving the distance closer to those included in radial distance bin four. Ground measurement errors for the distance between trees and the subplot center can also affect the measured distance.

Table 22: Selected multiple regression models from mixed stepwise selection procedure.

	Dependent variable	
	Biomass (rdb4, vhb1)	Volume (rdb4, vhb1)
R^2	0.71	0.84
Adj-R^2	0.62	0.79
RMSE	1164.09	1.35
β_0	-1198.30	-2.99
P-value	0.4188	0.1123
β_1 (p75)	0.07	0.00
P-value	0.0189	0.0039
VIF	1.01	1.01
β_2 (p100)	4322.03	7.36
P-value	0.0432	0.0096
VIF	1.01	1.01

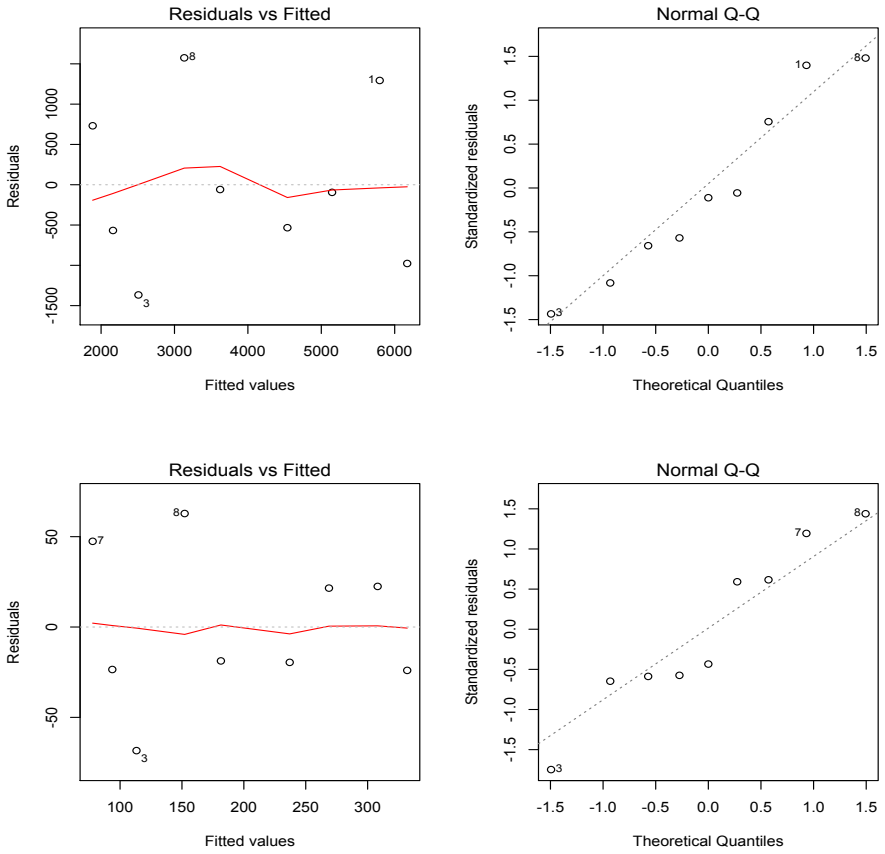

Figure 31: Diagnostic plots for selected multiple regression models for biomass (top) and volume (bottom).

There are several possible reasons for the poor predictive ability of the individual metrics used in this study, which include: (1) very small sample size; and (2) the chosen point cloud metrics. From a statistical standpoint the samples size used for this analysis (nine subplots) is small, which makes it difficult to account for the natural variability found in the data, as well as to produce estimates. A visual analysis of the scatter plots for the selected simple linear regression models indicates that individually, the lidar-derived metrics chosen for this study might not provide enough information about the subplots to enable the prediction of plot-level biomass or volume.

Many of the metrics used have produced good results when applied to ALS data (Lim and Trietz, 2004; Næsset and Bjerknes, 2001; Næsset and Okland, 2002; Næsset , 2007). It is clear that more research on the development of TLS metrics is needed. For instance, one of the major differences between ALS and TLS systems is system perspective. While ALS systems collect data looking down on the forest canopy, TLS systems collect data from underneath or within the forest canopy. In this case, the TLS system is also fixed at one location. The data collection differences between ALS and TLS systems, suggest that common ALS point cloud metrics do not provide enough information to successfully estimate plot-level biomass or volume. The main problem with using height in TLS metrics is that in many cases the real height of trees cannot be identified in the TLS data. This is because branches or other obstructions often occlude laser pulses from reaching the top of the tree. The results from this study suggest that combinations of horizontal distance-based and height-based metrics provided more information about the density of the vegetation within a plot, and lead to better prediction of biomass and volume.

It also may be the case that the high resolution subplots scans would be better utilized by some sort of individual tree measurement method, where individual tree measurements could be scaled-up to the plot level after their collection. Currently, options for automatically obtaining measurements such as number of stems, DBH, crown width, etc. are very limited. This means time consuming manual measurements, based on point cloud data, would be required to collect forest measurements from TLS data. However, the use of TLS technology is still in the early stages of development, and

future advances leading to the automated measurements of individual trees in point cloud data could one day become a possibility.

3.4 Conclusions

This study represents the first attempt to model biophysical parameters of interest to the FIA program utilizing elements of the standard FIA plot design, data from FIA ground crews, and TLS data. It is one of the first known attempts to apply commonly used ALS point cloud metrics to TLS data in order to model forest biomass and volume. Radial distance bin metrics, while similar to some hemispherical photography analysis metrics, were developed and helped account for the location and size of tree stems within TLS scans.

A TLS lidar system used to collect a single 360° scan was proven to be able to provide high-resolution data for small FIA subplots. This permanent record of the subplot conditions could be beneficial for a number of different uses. The utilization of commonly used ALS metrics, e.g. height percentiles, variable height bins, static height bins, to predict aboveground biomass or volume was not successful. It is hypothesized that metrics such as these support specific data collection characteristics of ALS systems better than TLS systems when trying to identify relationships between TLS data and plot-level biomass or volume. A combination of horizontal distance-based and height-based metrics better supports the data collection characteristics of TLS systems, and improved the ability to predict biomass and volume. Further investigations should focus on other lidar metrics designed to complement the data collection characteristics of the TLS systems and possible methods for combining spatially coincident ALS and TLS

data in an effort to provide a more complete view of the three-dimensional structure of the forest. The ability of a TLS system to predict biomass or volume should be subsequently addressed after this process using an increased sample size.

One possible source of error not addressed by this study were the FIA data collection dates. The FIA data for six of the nine subplots used to produce estimates of biomass and volume for this study were last collected at least two years before the lidar data were collected. Since TLS systems have the ability to collect information about tree stems, outdated DBH measurements could also negatively affect results. The use of regional species-specific growth and yield charts could help correct this deficiency for future studies.

Another future study could potentially investigate various alternatives to isolating the point cloud for subplots, by using a weighted distance method to include the influence of laser points hitting branches of trees with stems located within the subplot boundary, but with branches extending outside of it. It is also possible that branches from trees located outside the subplot boundary extend into the vertical space above the plot footprint and would therefore provide laser returns located in the subplot point cloud. A situation like this may compensate for missing branch returns of trees within the plot, but may also introduce unwanted variance from very large trees within the vicinity of the plot.

4 CONCLUSIONS

This study represents an initial attempt to model biophysical parameters of interest to the FIA program utilizing the standard FIA plot design, data from FIA ground crews, and ALS data. As previously mentioned, other studies have successfully modeled similar forest parameters with ALS data, but have done so under conditions different than those present in the Malheur National Forest. Issues such as the presence of a large number of tree species, complex terrain, dense forest conditions, and multi-story forest canopies, increased the difficult of modeling forest biophysical parameters in this study.

The area-based subplot-level point cloud and individual tree ALS metrics used in this study were shown to have a poor relationship to subplot estimates of biomass and volume estimated from FIA measurements. While the larger plot-level point cloud metrics exhibited a stronger relationship with the estimates of biomass and volume produced using FIA measurements, resulting in models that could account for approximately 71% of the variation present in the data. Even though the R-square values for models using the plot-level predictor variables are lower those reported by previous studies, e.g. Lim and Treitz 2004, they represent a large improvement from the models developed using subplot-level point cloud and individual tree ALS metrics.

The poor results produced from the subplot-level and individual tree ALS metrics helped identify several sources of error that should be taken into account in future research, such as: (1) geolocation error for subplot center coordinates; (2) edge effect. Other possible sources of error, such as FIA measurement date and the large variety of tree species, could have negatively affected study results. For example, 44 of the

subplots used in this study were measured by FIA crews at least two years before the lidar data were collected. Future research should utilize larger plots to reduce errors caused by geolocation and plot edge effect, as well as regional species-specific growth and yield models to mitigate error caused by out-of-date tree measurements.

The overlap of tree crowns within the complex multi-story forest canopy, coupled with the large number of tree species with crown morphologies differing from the southern pine species used to create TreeVaW most likely reduced the software's effectiveness. This problem was apparent when TreeVaW was unable to produce crown radii estimates for approximately 46% of the trees it identified. Future attempts to use TreeVaW to model forest biophysical parameters could possibly reduce these problems by manually matching TreeVaW identified trees, rather than averaging heights and crown widths for trees identified within a subplot. Methods such as this could follow a methodology similar to Popescu (2007).

This study represents the first attempt to model biophysical parameters of interest to the FIA program utilizing elements of the standard FIA plot design, data from FIA ground crews, and TLS data. It is one of the first known attempts to apply commonly used ALS point cloud metrics to TLS data in order to model forest biomass and volume. Radial distance bin metrics, while similar to some hemispherical photography analysis metrics, were developed and helped account for the location and size of tree stems within TLS scans.

A TLS lidar system used to collect a single 360° scan was proven to be able to provide high-resolution data for small FIA subplots. This permanent record of the

subplot conditions could be beneficial for a number of different uses. The utilization of commonly used ALS metrics, e.g. height percentiles, variable height bins, static height bins, to predict aboveground biomass or volume was not successful. It is hypothesized that metrics such as these support specific data collection characteristics of ALS systems better than TLS systems when trying to identify relationships between TLS data and plot-level biomass or volume. A combination of horizontal distance-based and height-based metrics better supports the data collection characteristics of TLS systems, and improved the ability to predict biomass and volume. Further investigations should focus on other lidar metrics designed to complement the data collection characteristics of the TLS systems and possible methods for combining spatially coincident ALS and TLS data in an effort to provide a more complete view of the three-dimensional structure of the forest. The ability of a TLS system to predict biomass or volume should be subsequently addressed after this process using an increased sample size.

One possible source of error not addressed by this study were the FIA data collection dates. The FIA data for six of the nine subplots used to produce estimates of biomass and volume for this study were last collected at least two years before the lidar data were collected. Since TLS systems have the ability to collect information about tree stems, outdated DBH measurements could also negatively affect results. The use of regional species-specific growth and yield charts could help correct this deficiency for future studies.

Another future study could potentially investigate various alternatives to isolating the point cloud for subplots, by using a weighted distance method to include the

influence of laser points hitting branches of trees with stems located within the subplot boundary, but with branches extending outside of it. It is also possible that branches from trees located outside the subplot boundary extend into the vertical space above the plot footprint and would therefore provide laser returns located in the subplot point cloud. A situation like this may compensate for missing branch returns of trees within the plot, but may also introduce unwanted variance from very large trees within the vicinity of the plot.

REFERENCES

Applied Imagery, 2010. Quick terrain modeler (Version 7.0.0) [Software]. Available from http://www.appliedimagery.com/.

Cannell, M., 1984. Woody biomass of forest stands, *Forest Ecology and Management*, *8*(3-4):299-312.

Coops, N., M. Wulder, D. Culvenor and B. St-Onge, 2004. Comparison of forest attributes extracted from fine spatial resolution multispectral and lidar data, *Canadian Journal of Remote Sensing*, *30*(6):855-866.

Falkowski, M., A. Smith, A. Hudak, P. Gessler, L. Vierling and N. Crookston, 2006. Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data, *Canadian Journal of Remote Sensing*, *32*(2):153-161.

Falkowski, M. J., A. M. S. Smith, P. E. Gessler, A. T. Hudak, L. A. Vierling and J. S. Evans, 2008. The influence of conifer forest canopy cover on the accuracy of two individual tree measurement algorithms using lidar data, *Canadian Journal of Remote Sensing*, *34*:S338-S350.

Frazer, G. W., S. Magnussen, M. A. Wulder and K.O. Neimann, 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass, *Remote Sensing of Environment*, *115*:636-649.

Grobakken, T. and E. Næsset, 2008a. Assessing effects of laser point density, ground samping intensity, and field sample plot size on biophtsical stand properties

derived from airborne laser scanner data, *Canadian Journal of Remote Sensing*, *38*:1095-1109.

Grobakken, T. and E. Næsset, 2008b. Assessing effects of sample plot possitiong errors on biophysical stand properties derived from airborne laser scanner data, *SilviLaser 2008, Sept. 17-19, 2008, Edinburgh, UK*, 14-21.

Gillespie, A., 1999. Rationale for a national annual forest inventory program, *Journal of Forestry*, *97*(12):16-20.

Hall, S., I. Burke, D. Box, M. Kaufmann and J. Stoker, 2005. Estimating stand structure using discrete-return lidar: An example from low density, fire prone ponderosa pine forests, *Forest Ecology and Management*, *208*(1-3):189-209.

Heath, L., M. Hansen, J. Smith, P. Miles and B. Smith, 2008. Investigation into calculating tree biomass and carbon in the FIADB using a biomass expansion factor approach *Forest Inventory and Analysis (FIA) Symposium*, October 21–23, 2008: Park City, UT.

Henning, J. and P. Radtke, 2006. Detailed stem measurements of standing trees from ground-based scanning lidar, *Forest Science*, *52*(1):67-80.

Hirata, Y., N. Furuya, M. Suzuki and H. Yamamoto, 2009. Airborne laser scanning in forest management: Individual tree identification and laser pulse penetrationin a stand with different levels of thinning, *Forest Ecology and Management*, *258*:752-760.

Holmgren, J., M. Nilsson and H. Olsson, 2003. Estimation of tree height and stem volume on plots-using airborne laser scanning, *Forest Science*, *49*(3):419-428.

Holmgren, J., 2004. Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning, *Scandinavian Journal of Forest Research*, *19*(6):543-553.

Holmgren, J. and Å. Persson, 2004. Identifying species of individual trees using airborne laser scanner, *Remote Sensing of Environment*, *90*(4):415-423.

Hopkinson, C., L. Chasmer, C. Young-Pow and P. Treitz, 2004. Assessing forest metrics with a ground-based scanning lidar, *Canadian Journal of Remote Sensing*, *34*(3):573-583.

Hyde, P., R. Dubayah, W. Walker, J. Blair, M. Hofton and C. Hunsaker, 2006. Mapping forest structure for wildlife habitat analysis using multi-sensor (lidar, sar/insar, etm+, quickbird) synergy, *Remote Sensing of Environment*, *102*(1-2):63-73.

Hyde, P., R. Nelson, D. Kimes and E. Levine, 2007. Exploring lidar-radar synergy - predicting aboveground biomass in a southwestern ponderosa pine forest using lidar, sar and insar, *Remote Sensing of Environment*, *106*(1):28-38.

Hyyppa, J., O. Kelle, M. Lehikoinen and M. Inkinen, 2001. A segmentation-based method to retrieve stem volume estimates from 3-d tree height models produced by laser scanners, *Geoscience and Remote Sensing, 39*(5):969-975.

Jenkins, J., D. Chojnacky, L. Heath and R. Birdsey, 2003. National-scale biomass estimators for united states tree species, *Forest Science*, *49*(1):12-35.

Lefsky, M., W. Cohen, S. Acker, G. Parker, T. Spies and D. Harding, 1999. Lidar remote sensing of the canopy structure and biophysical properties of douglas-fir western hemlock forests, *Remote Sensing of Environment*, *70*(3):339-361.

Lefsky, M. and M. McHale, 2008. Volume estimates of trees with complex architecture from terrestrial laser scanning, *Journal of Applied Remote Sensing*, *2*:1-19.

Lim, K., P. Treitz, M. Wulder, B. St-Onge and M. Flood, 2003. Lidar remote sensing of forest structure, *Progress in Physical Geography*, *27*(1):88-106.

Lim, K. and P. Treitz, 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators, *Scandinavian Journal of Forest Research*, *19*(6):558-570.

Maas, H. G., A., Bienert, S. Scheller and E. Keane, 2008. Automatic forest inventory parameter determination from terrestrial laser scanner data, *International Journal of Remote Sensing*, *29*(5):1579-1593.

McCombs, J., S. Roberts and D. Evans, 2003. Influence of fusing lidar and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation, *Forest Science*, *49*(3):457-466.

Mutlu, M., S. Popescu, C. Stripling and T. Spencer, 2008a. Mapping surface fuel models using lidar and multispectral data fusion for fire behavior, *Remote Sensing of Environment*, *112*(1):274-285.

Mutlu, M., S. Popescu and K. Zhao, 2008b. Sensitivity analysis of fire behavior modeling with lidar-derived surface fuel maps, *Forest Ecology and Management*, *256*:289-294.

Næsset, E., 1997a. Estimating timber volume of forest stands using airborne laser scanner data, *Remote Sensing of Environment*, *61*(2):246-253.

Næsset, E., 1997b. Determination of mean tree height of forest stands using airborne laser scanner data, *ISPRS Journal of Photogrammetry and Remote Sensing*, *52*(2):49-56.

Næsset, E. and K. Bjerknes, 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data, *Remote Sensing of Environment*, *78*(3):328-340.

Næsset, E. and T. Okland, 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve, *Remote Sensing of Environment*, *79*(1):105-115.

Næsset, E., 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in scandinavia, *Scandinavian Journal of Forest Research*, *22*(5):433-442.

Nelson, R., W. Krabill and J. Tonelli, 1988. Estimating forest biomass and volume using airbone laser data, *Remote Sensing of Environment,* 24:247-267.

Nilsson, M., 1996. Estimation of tree heights and stand volume using an airborne lidar system, *Remote Sensing of Environment*, *56*(1):1-7.

Patenaude, G., R. Hill, R. Milne, D. Gaveau, B. Briggs and T. Dawson, 2004. Quantifying forest above ground carbon content using lidar remote sensing, *Remote Sensing of Environment*, *93*(3):368-380.

Popescu, S., R. Wynne and R. Nelson, 2002. Estimating plot-level tree heights with lidar: Local filtering with a canopy-height based variable window size, *Computers and Electronics in Agriculture*, *37*(1-3):71-95.

Popescu, S., R. Wynne and R. Nelson, 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass, *Canadian Journal of Remote Sensing 29*(5):564-577.

Popescu, S., R. Wynne and J. Scrivani, 2004. Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in virginia, USA, *Forest Science*, *50*(4):551-565.

Popescu, S., 2007. Estimating biomass of individual pine trees using airborne lidar, *Biomass and Bioenergy*, *31*(9):646-655.

Qi, C., D. Baldocchi, P. Gong and M. Kelly, 2006. Isolating individual trees in a savanna woodland using small footprint lidar data, *Photogrammetric Engineering and Remote Sensing*, *72*(8):923-932.

Roberts, S., T. Dean, D. Evans, J. McCombs, R. Harrington and P. Glass, 2005. Estimating individual tree leaf area in loblolly pine plantations using lidar-derived measurements of height and crown dimensions, *Forest Ecology and Management*, *213*(1-3):54-70.

Sexton, J., T. Bax, P. Siqueira and J. Swenson, 2009. A comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeastern north america, *Forest Ecology and Management*, *257*:1136-1147.

TerMikaelian, M. and M. Korzukhin, 1997. Biomass equations for sixty-five north american tree species, *Forest Ecology and Management*, *97*(1):1-24.

Thies, M., N. Pfeifer, D. Winterhalder and B. Gorte, 2004. Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser

scanning of standing trees, *Scandinavian Journal of Forest Research*, *19*(6):571-581.

USFS 2008. Field instructions for the annual inventory of California, Oregon and Washington. United States Forest Service Pacific Northwest Research Station. 348 pp. Available from www.fs.fed.us/pnw/fia/publications/fieldmanuals.shtml

Yu, X., J. Hyyppa, H. Kaartinen and M. Maltamo, 2004. Automatic detection of harvested trees and determination of forest growth using airborne laser scanning, *Remote Sensing of Environment*, *90*(4):451-462.

Zhao, K., S. Popescu and R. Nelson, 2009. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers, *Remote Sensing of Environment*, *113*(1):182-196.

Zhou, X. and Hemstrom, M. A. 2010. Timber volume and aboveground live tree biomass estimations for landscape analyses in the pacific northwest. Gen. Tech. Rep. PNW-GTR-819. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 31 p.

VITA

Name: Ryan D. Sheridan

Address: Spatial Sciences Lab, 1500 Research Parkway, Suite B217,
College Station, TX 77843

Email Address: ryan.sheridan@tamu.edu

Education: B.A., Forest Resource Management, University of Idaho, 2009
M.S., Forestry, Texas A&M University, 2011