Title
Using lidar in wildfire ecology of the California Sierra-Nevada forests

Author
Jakubowski, Marek K.

Publication Date
2012

Peer reviewed|Thesis/dissertation

Using lidar in wildfire ecology of the California Sierra-Nevada forests

By

Marek K. Jakubowski

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Environmental Science, Policy, and Management
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Maggi Kelly, Chair
Professor Peng Gong
Professor John Radke

Fall 2012

Using lidar in wildfire ecology of the California Sierra-Nevada forests

by

Marek K. Jakubowski

Abstract

Using lidar in wildfire ecology of the California Sierra-Nevada forests

by

Marek K. Jakubowski

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Maggi Kelly, Chair

California's fire suppression policy has dramatically changed Sierra Nevada forests over the last century. Forests are becoming more dense and homogenous, leading to fire regime changes that increase the potential of stand-replacing wildfires over large, continuous areas. To mitigate this problem on public lands, the US Forest Service has proposed to implement strategically placed forest fuel reduction treatments. These treatments have been proved effective in modeled and simulated environments, but their efficacy and impact in real forests is not known. The research described in this dissertation is part of a large multidisciplinary project, known as the Sierra Nevada Adaptive Management Project (SNAMP), that aims to evaluate strategically placed landscape area treatments (SPLATs) in two forests of the Sierra Nevada mountains. Specifically, in this thesis, I investigate the feasibility of using an airborne light detection and ranging (lidar) system to gain accurate information about forest structure to inform wildfire behavior models, forest management, and habitat mapping.

First, I investigate the use of lidar data in predicting metrics at the landscape level, specifically to derive surface fuel models and continuous canopy metrics at the plot scale. My results in Chapter 2 indicate that using lidar to predict specific fuel models for FARSITE wildfire behavior model is challenging. However, the prediction of more general fuel models and continuous canopy metrics is feasible and reliable, especially for metrics near the top of the canopy.

It is also possible to derive canopy parameters at the individual tree level. In Chapter 3, I compare the ability of two processing methods—object-based image analysis (OBIA) and 3D segmentation of the lidar point cloud—to detect and delineate individual trees. I find that while both methods delineate dominant trees and accurately predict their heights, the lidar-derived polygons more closely resemble the shape of realistic individual tree crowns.

Acquiring remotely sensed data at high resolution and over large areas can be expensive, especially in the case of lidar. In Chapter 4, I investigate the ability of lidar data to reliably predict forest canopy metrics at the plot level as the data resolution declines. I show that canopy metrics can be predicted at a reasonable accuracy with data resolutions as low as one pulse per squared meter. These findings will be useful to land managers making cost benefit decisions when acquiring new lidar data.

Collectively, the results of this dissertation suggest that remote sensing, and in particular lidar, can reliably and cost-effectively provide forest information across scales—from the individual tree level to the landscape level. These results will be useful for the fire and forest management community in general, as well as being key to the goals of the SNAMP program.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This dissertation is over 90 pages long, but writing this acknowledgements page has been the most rewarding. First and foremost I want to give a big thanks to my advisor Maggi Kelly, who has been an amazing guide through this journey. She pushed me when I needed it (even when I didn't realize it), but also gave me the freedom to explore my own research questions. Thank you for your guidance, motivation, and support when I needed them, and thank you for being patient through all these years. I also extend my gratitude to my dissertation committee, John Radke and Peng Gong, for their insight and advice.

Publishing papers is critical in graduate school, and to that extent I want to thank all of my mentors and co-authors who have helped me with analyses and in writing all of the manuscripts: Qinghua Guo, Wenkai Li, Brandon Collins, Scott Stephens, and Sam Blanchard.

The work described here would not have been possible without the tremendous help, determination, sweat (literally), and hard work of everyone who helped me collect the extensive ground truth data over two hot summer field seasons: Gary Roller, Alessandro Montaghi, Sarah Lewis, Kevin Krasnow, John Connors, Michael Rawlins, Zak Thomas, Stephen Keller, Lauren Grand, Kim Taylor, John Dvorak, Christine Byrn, and Danny Fry. Thank you for hiking through endless, steep, thick forests, carrying my heavy equipment, waking up before sunrise, and most importantly, holding the laser "puck."

I am very grateful for the financial support of the Sierra Nevada Adaptive Management Project (SNAMP). I also want to thank John Battles who, through his position as head PI of SNAMP, has shown me that it takes more than intelligence and a sound statistical background to be a good scientist: it takes a good character. SNAMP annual meetings wouldn't be the same without you.

Graduate school has been an absolutely transformative experience for me, in every sense of the word: as a scientist and, more importantly, as a person. But it would not have happened, and certainly would not have been as enjoyable, without all of the people who gave me great support. I want to give an enormous thank you to everyone who was not involved in my research, and yet, had to pay the price for it:

My parents, Grzegorz and Regina, who would have preferred to hear from me more often during my years in Berkeley and yet continually gave their complete support; my siblings Adam and Aleksandra, who always have been a great inspiration; my dear friend Bradley Balukjian, who has given me great advice and listened to my complaints; Jane White: living in your house has contributed to building my character in ways that I never anticipated—thank you for being a true friend; Felix Mormann and Alberto Amengual, for sharing a roof and being amazing friends (and thanks Alberto for introducing me to WEKA!); Tim De Chant, for igniting my passion for design through *Berkeley Science Review*; and finally, I would like to especially thank Hania. Thank you for the countless hours of looking at my dissertation drafts, making my life better, your attention, and smiles. All of these things and many more have been crucial and so very helpful. Thank you for creating and sharing the best moments in my Berkeley experience.

Thanks to you all.

Chapter One

Lidar remote sensing in wildfire forest ecology

Prior to this research, the light detection and ranging (lidar) remote sensing literature was unclear regarding (1) the optimal lidar pulse density at which stand metrics can be derived reliably, (2) whether lidar can be used consistently to derive spatial input for wildfire behavior models, and (3) how raster-based and 3D lidar point segmentation compare in delineating individual trees. To address these questions, I present a new method to model multiple-return discrete lidar data at lower pulse densities and then estimate their ability to derive stand metrics in a mountainous mixed-conifer Sierra Nevada forest. I show that lidar can predict general fuel types and continuous fuel stand metrics, but predicting specific standard surface fuel models is still challenging. Finally, I compare two segmentation approaches to delineate individual trees: both methods accurately predict tree heights, but the 3D segmentation approach delineates tree crowns that are more similar to those present in a real forest. The motivation for this research comes from a growing need to accurately map forests, especially in the context of wildfire ecology in California.

Influences of fire and fire suppression on Sierra Nevada forest ecosystems in California

Over a century of fire exclusion policy has led to dramatic changes in California's forests (Agee and Skinner, 2005; Taylor, 2001). Fire, a fundamental component of the Mediterranean-climate ecosystem, was prevalent in California prior to European settlement in the 1850s (Collins *et al.*, 2008; Skinner and Chang, 1996; Stephens and Collins, 2004; Stephens *et al.*, 2007; Sugihara *et al.*, 2006). This fire exclusion policy has directly changed fire regime parameters (i.e. decrease in return interval, earlier burn season, and larger fires), which in turn have fundamentally changed the structure and composition of forests (Taylor, 2001). Specifically, California forests are becoming more dense and homogeneous (Stephens and Collins, 2004), and ladder fuels (low and midaerial fuels that enable fire to climb from the forest floor into the crowns) which often lead to canopy wildfires are increasing (Kilgore and Sando, 1975; Menning and Stephens, 2007). All of these changes have, particularly in the American West, lead to an increase in stand-replacing (high severity) wildfires that affect large, continuous areas (Stephens and Collins, 2004). These more common high severity fires impact horizontal and vertical vegetation patterns, and thus directly impact forest processes such as wildlife habitat availability, carbon sequestration, hydrology, nutrient cycles, and other critical forest ecosystem processes (Miller *et al.*, 2009). Additionally, these fires impact humans through direct wildfire suppression costs, loss of forest products, personal property, and fatalities (Stephens *et al.*, 2012).

In 2004, in part because of the changed fire regime and its effects, and in part as a result of the tragic aftermath of the 2003 California fire season, the Regional Forester signed a new Record of Decision (ROD) for the Sierra Nevada Forest Plan Amendment. This ROD had made a number of policy changes regarding to management of wildlife habitats, old forests, and watersheds, but the changes most relevant to this work pertain to forest fuel treatments. The ROD recommended that strategic fuel reduction and tree thinning be deployed across the USFS forests. The concept behind these treatments is that if area treatments are placed strategically, the spread rate and intensity of the fire over the entire area burned will likely be reduced (Finney, 2003). These strategic area treatments are collectively known as strategically placed land area treatments (SPLATs). The key aim of SPLATs is to decrease high severity fires (Agee and Skinner, 2005), reduce danger to firefighting efforts (Agee *et al*., 2000; Moghaddas and Craggs, 2007), and facilitate ecological restoration of structure and function (Collins *et al*., 2010; McKelvey *et al*., 1996; North *et al*., 2007; Weatherspoon and Skinner, 1996). The aim is to change the forest structure from a dense (>60% canopy closure), small tree, Sierra mixed-conifer forest to an open cover (25-39% canopy closure), medium or large tree dominated forest stand. At the time when SPLATs were proposed they had largely been tested only in a modeling framework, and no applied evidence to confirm their practical efficacy on the ground, and their effect on an ecosystem, long- or short-term, were undetermined. The first step necessary to understand the impact of SPLATs on forest function is to accurately measure and map these forests. My work is part of a larger multidisciplinary project that aims to evaluate SPLATs in two forests of the Sierra Nevada mountains. This larger project is called the Sierra Nevada Adaptive Management Project (SNAMP).

Sierra Nevada Adaptive Management Project

In 2005, a key agreement was reached among US government agencies, public stakeholders, and the University of California science team (UCST) to work in concert on what is now known as SNAMP. The project's main objectives are to assess the impact of SPLATs as well as their effects on forest ecology, hydrology, and wildlife—specifically the California spotted owl (*Strix occidentalis occidentalis*) in the Last Chance study area in the Tahoe National Forest, and the Pacific fisher (*Martes pennanti pacifica*) in the Sugar Pine study area in the Sierra National Forest. These sites are comprised of mixed-conifer forests on steep topography. The Last Chance study area covers 9,950 ha and is distant from urban areas, and the Sugar Pine study area covers 3,930 ha and is partially in a wildland-urban interface. My research focuses primarily on the Last Chance study area. In each of the study areas, SPLATs were designed by the USFS to diminish the effects of wildfires by reducing the fire spread rate and intensity over the entire area (Finney, 2003; Stephens, 1998). The SPLATs were implemented in both study areas by the USFS, and the UCST is analyzing treatment effect across the forested landscape in a standard Before-After-Control-Impact (BACI) framework. In this framework, all remotely sensed data and measurements pertaining to wildlife, hydrology, fire and forest health were acquired prior to treatments, and the same measurements will be taken after the treatments are completed. The larger project is focused on the following forest parameters: water quality and quantity, forest health and fire behavior, and wildlife (California spotted owl and Pacific fisher); all with an evaluation of the importance of public participation in the forest management

process. The results will be used to inform forest management decisions following an adaptive management framework. My dissertation focuses on geospatial research in support of the forest processes, especially fire behavior.

The SNAMP research plan relies on empirical science and integrated spatial models. Spatial data—primarily remotely sensed imagery and lidar—and analysis are key integrators in this multi-disciplinary project: providing inputs to models, scaling biophysical processes from plot to landscape level, estimating uncertainty, and visualizing future scenarios. The research results rely on modeled outputs, but these models are parameterized and tested against a dense network of field plots. Circular, 500 m^2 plots were established (n = 411 in the northern, and n = 284 in the southern study area) across the study areas, spanned across an even grid, to collected data about tree characteristics and health, shrub and ladder fuels, and surface fuel loadings (Jakubowski *et al.*, 2013). These field data are linked with remote sensing through a suite of analytics: regression, machine learning, classification and object-based approaches, but all rely and build on a long history of remote sensing technology used to map and understand forests.

Remote sensing in forest science

Remote sensing has been an instrumental tool in many fields for gathering knowledge at local, landscape, and global scales. Its roots in forest science date back to 1960s, when the first Forestry Remote Sensing Laboratory was established at the University of California, Berkeley (Jensen, 2007). Remote sensing has gone through a technological metamorphosis, in terms of data gathering, as well as the data type, quality, and quantity. Early sensors gathered data at course resolutions, e.g. TIROS and GOES (Melesse *et al.*, 2007). As better airborne and spaceborne platforms became available, and as sensor materials technology improved, so did the ability to collect images across a wider range of spectral, temporal, and spatial resolutions and ranges.

The spectral content of remotely sensed imagery has been especially valuable as it allows for various types of analyses: from measuring atmospheric NO_2 in the ultraviolet (0.350 μm), through vegetation monitoring in the visible-near infrared (VNIR, 0.4-1.4 μm), to deriving relationships between land surface temperature and vegetation indices in the thermal infrared (1.5-12.5 μm (Chance, 2005; Jensen, 2007; Quattrochi and Luvall, 1999; Weng, 2009). Increasing the spectral resolution and range (hyperspectral imagery) is particularly useful in environmental analysis as it helps in distinguishing vegetation species and soil types (Gong *et al.*, 1997; Okin *et al.*, 2001; Roberts *et al.*, 1998; Schott, 2007). Perhaps the most noticeable range is in the spatial domain, in which satellite sensors can gather data anywhere from very low (OrbView-2: 1 km pixels), through medium (Landsat 7: 15-60 m), to high (WorldView-2: 48 cm) ground sample distance (GSD) (Kramer, 2001; Padwick *et al.*, 2010). Airborne- and unmanned aerial vehicle-mounted (UAV) sensors may push that limit even further depending on the operational altitude above the ground (e.g. using a four spectral band Leica Geosystems ADS40 sensor to produce VNIR imagery at 1 m GSD, or below 10 cm GSD using an ultra-light sensor mounted on a UAV), creating imagery potentially beyond a point that is currently useful (Nebiker *et al.*, 2008).

Although this rapid increase in spatial sampling has been exciting, is has become increasingly clear that algorithms developed for courser imagery often perform poorly at these finer scales due to high local heterogeneity (Cleve *et al.*, 2008; Guo *et al.*, 2007; Kelly *et al.*, 2011). As a result, finer resolution imagery has necessitated a change in the way imagery is processed: from pixel-based to object-based image analysis (OBIA). The OBIA approach first segments imagery into meaningful objects through the use of spectral and spatial context prior to classification. This allows for image processing and Geographic Information System (GIS) functionality (Blaschke, 2010). The OBIA approach is inherently multi-scalar, as image segments are captured and operate hierarchically. A number of algorithms have been design to segment imagery in the OBIA framework, for example eCognition and Berkeley ImageSeg (Clinton *et al.*, 2010; Trimble, 2012).

OBIA results have been encouraging across scales, applications, and data inputs. At the landscape scale, studies have successfully extracted biophysical forest inventory parameters from imagery (Chubey *et al.*, 2006) and used lidar to classify land covers (Antonarakis *et al.*, 2008; Brennan and Webster, 2006). At smaller scales, Guo *et al.* (2007) used high resolution imagery to delineate and identify dead trees, Holt *et al.* (2009) successfully delineated and classified individual cars within San Francisco, and (Cleve *et al.*, 2008) showed that classifying imagery in wildland-urban interface using OBIA can lead to significantly higher accuracies than with a pixel-based approach.

Optical imagery has been and continuous to be an invaluable tool in remote sensing of forests. However, it also has considerable limitations, especially in densely vegetated, complex forests such as my Sierra Nevada study area, where the passive energy detected by the sensors is incapable of penetrating the thick forest canopy. Fortunately, there are other remote sensing technologies that can better map the horizontal and vertical heterogeneity and structure of such forest stands. Radio detection and ranging (radar) and lidar are at the forefront of this endeavor. All of the work in my dissertation heavily relies on lidar data, as this was the primary focus of my research.

Although lidar has been invented in the late 1950s (Rocadenbosch, 2003), majority of the practical developments that contribute to its wide use in remote sensing have been accomplished in the last decade. Lidar works by emitting a laser pulse towards a target and measuring the time that it takes for it to return. The time is converted into distance using the simple time-distance-speed relationship, where the speed is the speed of light in the given medium (air). Because lidar platforms are also equipped with a high-accuracy global positioning system (GPS) and an inertial measurement unit (IMU), the travel distance can be easily transformed into a geolocated 3D path, from the sensor to a point on the ground. Along the 3D path is the signal amplitude: the amount of light that was returned as the laser pulse traveled through the canopy and back toward the sensor (waveform lidar). This amplitude is often quantized into a number of points, or returns, that can be converted into (*x,y,z*) locations (dicrete lidar). A lidar scanner repeats this process millions of times as it is flown over an area in order to collect a point (Baltsavias, 1999a, b). The point cloud can then be used in numerous ways, from transformation into a 2D raster surface where traditional image processing and GIS techniques can be applied, to fully utilizing the complexity of the 3D point configuration.

Because of lidar data's accuracy, detail, ability to penetrate and measure the canopy, and because the cost of its acquisition has been declining, the usage of lidar data in forest remote sensing has risen steeply in the last decade. In many ways, lidar has fundamentally changed the field of remote sensing (Dubayah and Drake, 2000). For example, with access to high quality, dense lidar data, extracting accurate heights of dominant trees, which is a very complex task for photogrammetric methods (Miller *et al.*, 2000; Sheng *et al.*, 2001), has become almost routine with lidar data, even in a dense forests (Erdody and Moskal, 2010; Miller *et al.*, 2000; Popescu and Wynne, 2004). Both waveform (Hyde *et al.*, 2005; Mallet and Bretar, 2009) and discrete (Hyyppä *et al.*, 2009)data have been used successfully to map a wide range of forest structure parameters (Hyyppä *et al.*, 2009; Lefsky *et al.*, 2002), including tree heights (Popescu and Wynne, 2004), basal area (Chen *et al.*, 2007; Jakubowski *et al.*, Accepted), canopy fuel weight and bulk density (Andersen *et al.*, 2005; Persson *et al.*, 2002), leaf area index (Riaño *et al.*, 2004), biomass (Zhao *et al.*, 2012a), and detection or delineation of individual trees in two- (Chen *et al.*, 2006; Koch *et al.*, 2006; Morsdorf *et al.*, 2004; Popescu *et al.*, 2004; Yu *et al.*, 2011) and three dimensions (Lee *et al.*, 2010; Li *et al.*, 2012).

Remote sensing has had a decades-long role in forest mapping, but it is only in the last decade, with application of lidar, that the vertical and horizontal characterization of forests has been possible. Despite this positive assessment, numerous questions remain about the applicability of lidar for forest science in general, and for the study of specific fire and fuel-related parameters of forests. This introductory chapter sets the stage for an evaluation of the use of remote sensing in forest ecology. In this dissertation, I develop new methods, compare existent methods, and test the limits of lidar and optical data in order to contribute to the fields of remote sensing and forest science, and to support the specific goal of understanding the impacts of forest fuel treatments across a range of forest processes.

Chapter overview

In the following chapters, I illustrate a suite of new approaches, uses and analyses of lidar data to provide detailed mapping of forests. My work has direct links to spatially explicit process models such as fire behavior models, I also highlight practical and applied considerations, and my work also takes advantage of integrated geoinformatic technologies.

One of the main questions of SNAMP is to address the possible modification of fire behavior across a fireshed with fuel treatments. The main way to test the impact of SPLATs on fire is through wildfire behavior modeling (Finney, 2001). Modeling the fire behavior requires a number of spatial inputs including topography (elevation, slope, and aspect), forest structure (canopy cover, canopy height, crown base height, and crown bulk density), and surface fuel characterizations (referred to as fuel models). While most of these inputs are well understood and estimated from ground data, the fuel models are subjective and sometimes altered iteratively by the analyst to make the wildfire behavior model produce realistic results. Commonly, fuel models are assigned using field data as recommended by a guide Scott and Burgan (2005) in addition to expert opinion. In Chapter 2, I investigate the use of lidar for fire behavior models. Specifically, I ask: *How well can capture the fuel build-up*

using lidar? How well can lidar be used to feed input into fire behavior models across landscapes? Can lidar predict individual variables that feed fire behavior models? I predict these fuel models using lidar and optical imagery, as well as the metrics that directly feed into the wildfire behavior model, and the metrics that are used by fire experts to assign the fuel models. I found that stand structure metric predictions generally decreased with increased canopy penetration. For example, from the top of canopy, I predicted canopy height ($r^2 = 0.87$), canopy cover ($r^2 = 0.83$), basal area ($r^2 = 0.82$), shrub cover ($r^2 = 0.62$), shrub height ($r^2 = 0.59$), combined fuel loads ($r^2 = 0.48$), and fuel bed depth ($r^2 = 0.35$). While the general fuel types were predicted accurately, specific surface fuel model predictions were poor (76 and <50 percent correct classification, respectively) using all algorithms. These fuel components are critical inputs for wildfire behavior modeling, which ultimately supports forest management decisions. This work is the first comprehensive evaluations of the relative utility of lidar and optical imagery for fire behavior modeling in a mixed-conifer Sierra Nevada forest.

In Chapter 3 I take on the task of delineating individual trees from lidar data, and ask: *Are there reliable methods for identifying individual trees from lidar data? How do these algorithms perform across forest* density? The successful identification and delineation of individual trees is critical in forest science, allowing for multi-scale analysis of tree demography, carbon storage, nutrient cycling and wildlife habitat. There are many possible methods to delineate trees with remotely sensed data. In this chapter, I compare two segmentation algorithms for individual tree delineation: a lidar-derived method that uses 3D lidar points, and an OBIA approach that make use of the lidar canopy height model. I compared these two methods in *x-y* space in terms of their agreement to ground referenced tree heights ($r^2 = 0.93$ and $r^2 = 0.96$, respectively), and tree detection across crown class and tree density. There were important similarities between the methods with respect to their ability to delineate trees in less dense forests, and with larger, overstory trees. There were also important differences. The two types of objects were different in terms of polygon area and shape complexity. The OBIA objects were more likely to over-segment while the lidar objects were more likely to under-segment the trees, although the latter produced polygons more similar in shape to real tree crowns. Further research is necessary to automate the OBIA segmentation process and to improve 3D segmentation of the point cloud in dense forests.

In Chapter 4 I build on my work and ask an ultimately practical question of the lidar data: *How much lidar data is necessary to map forest variables adequately in a mixed-conifer forest?* Discrete lidar is increasingly used to analyze forest structure, and technological improvements in lidar sensors have led to the acquisition of increasingly high pulse densities. In this chapter, I systematically investigated the relationship between pulse density and the ability to predict several commonly used forest measures and metrics at the plot scale. The accuracies of predicted metrics were largely invariant to changes in pulse density at moderate to high densities. In particular, correlations between metrics such as tree height, diameter at breast height, shrub height and total basal area were relatively unaffected until pulse densities dropped below 1 pulse/m^2. Metrics pertaining to coverage, such as canopy cover, tree density and shrub cover, were more sensitive to changes in pulse density, although in some cases high prediction accuracy was still possible at lower densities. These findings did not depend on the type of predictive algorithm used, although I found that support vector regression (SVR) and Gaussian processes (GP) consistently outperformed multiple regression across a range of pulse

densities. I also found that SVR yielded higher accuracies at low densities (<0.3 pl/m^2), while GP was better at high densities (>1 pl/m^2). These results suggest that low-density lidar data may be capable of estimating typical forest structure metrics reliably in some situations, and thus provide practical guidance to forest ecologists and land managers who are faced with tradeoff in price, quality and coverage, when planning new lidar data acquisition.

Finally, in Chapter 5, I highlight the key results of my dissertation as they relate to forest ecology and management. I also give my perspectives about future research, challenges encountered through my work, and their potential solutions. Together, the findings presented in this work will be useful for forest and fire management communities and they are key to the larger SNAMP program.

Chapter Two

Predicting surface fuel models and fuel metrics using lidar and CIR imagery in a dense, mountainous forest

Introduction

Fire is an important component of forest ecosystems in the Sierra Nevada, and was prevalent before widespread settlement in California (Collins *et al.*, 2008; Skinner and Chang, 1996; Stephens and Collins, 2004; Stephens *et al.*, 2007; Sugihara *et al.*, 2006). Many common Sierran plants exhibit fire-adapted traits such as thick bark and fire-stimulated flowering, sprouting, seed release and germination (Sugihara *et al.*, 2006). In the last century, however, fuel loads in many Sierra Nevada forests have increased, likely as a result of fire suppression policies, warmer and moister climatic conditions, and the effects of past harvesting (Collins *et al.*, 2011a; Stephens and Ruth, 2005), putting the forests at risk of catastrophic fire (Miller *et al.*, 2009; van Wagtendonk *et al.*, 1998). In response to this risk, the US Forest Service is employing an approach that involves placing discrete fuel reduction treatments strategically across landscapes such that fire intensity is reduced not only within treated areas, but throughout the entire landscape (Finney, 2001; Moghaddas *et al.*, 2010).

In the US, modeling wildfire behavior and associated planning of fuel reduction treatments across landscapes are typically performed using FARSITE (Finney, 1998) and FlamMap (Finney, 2006). Both models rely heavily on Rothermel's fire spread model developed at the US Forest Service Fire Sciences Laboratory in Missoula, Montana (Rothermel, 1972). FARSITE and FlamMap require a standard suite of spatial data layers in order to run. These data layers include topography (elevation, slope, and aspect), forest structure (canopy cover, canopy height, crown base height, and crown bulk density), and surface fuel characterizations (referred to as fuel models). These data layers, co-registered and resolved at the same spatial resolution, are a fundamental input of most fire behavior models.

While all of these inputs influence the fire behavior models, surface fuel models are particularly important in determining fire behavior, as the Rothermel fire spread model predicts behavior of surface fire only. These fuel models (FM) consist of a set of surface fuel characteristics that describe stands based on dominant fire carrying fuel types (grass, shrub, tree litter, etc.) that include: fuel loads by fuel particle size class, fuel bed depth, surface-area-to-volume ratios for fuel particles, heat content, and dead fuel extinction moisture. This work proceeds under the assumption that surface fuel loads are related to forest composition and structure. Although this relationship has not been comprehensively studied, recent work in the Sierra Nevada has demonstrated robust relationships

between fuel deposition rates and forest type/structure (van Wagtendonk and Moore, 2010). Surface FMs are divided into broad fuel types (e.g., shrub (SH), timber-understory (TU), and timber-litter (TL)), and then further into specific FMs (e.g., "low load compact conifer litter" (TU1 181)). Thirteen models were first described by Anderson (1982) in a technical report, and more recently this list was expanded to 40 models (Scott and Burgan, 2005). In this paper, the updated 40-model set was utilized since these models best describe the landscape in the study area.

In practice, a forest stand is assigned to one of these discrete surface FMs through a combination of field data, expert knowledge, and some type of either imagery interpretation or more automated remote sensing analysis (Collins *et al.*, 2010). The field data includes plot-based descriptions of vegetation (height, diameter at breast height (DBH), species, density of trees and shrubs, and percent canopy cover), fuel load (the amount of litter, fuel bed depth, and fuel load noted as 1-, 10-, and 100-, hour fuels), photographs, and a general description of the area (Bertolette and Spotskey, 1999; Keane *et al.*, 1998). Expert knowledge typically involves iteratively selecting FMs such that predicted fire behavior is consistent with first-hand experience of fire behavior in actual fires (Collins *et al.*, 2011b). Imagery is generally used to aid in FM selection across larger planning areas. While there is guidance aiding users in FM selection (Anderson 1982, Scott and Burgan 2005), the process of assigning FMs can be fairly subjective due to the common practice of incorporating expert opinion based on observed fire behavior. As a result, assignment errors can be introduced when FM descriptions are adjusted in order to produce fire behavior predictions more consistent with observed fire behavior (Varner and Keyes, 2009). This subjectivity can create inconsistency in FM assignments between users, resulting in substantial differences in predicted fire behavior (Cruz and Alexander, 2010; Varner and Keyes, 2009).

Wildfire behavior models and remote sensing

In typical circumstances where forest managers are faced with assessing fire behavior on a large scale, the cost, time, and technical challenges required to collect field data and to assign FMs make complete coverage of a forest prohibitive. This is particularly true in areas with steep topography or in areas with limited access. In part to address these challenges, some recent efforts have experimented with more automated approaches using multispectral imagery (Riaño *et al.*, 2002; van Wagtendonk and Root, 2003), hyperspectral imagery (Jia *et al.*, 2006), or a combination of remotely sensed imagery and topography, climate, and disturbance data (Moghaddas *et al.*, 2010). Yet, when FMs are estimated using remote sensing imagery, accuracy can be limited by the inability of sensors to capture the three-dimensional structure of the forest, especially when the tree density or stand height is high. Although hyperspectral image analysis offers more potential for species differentiation than multispectral data (Jia *et al.*, 2006), neither of these passive remote sensing methods can accurately map 3D forest structure. Lidar presents advantages in this context as it is capable of describing the vertical structure of a forest stand and has successfully been used to map detailed forest parameters.

Previous research

Lidar data is increasingly used to characterize forests across scales with direct measurements such as tree height, and derived measures such as biomass or leaf area index (LAI) (Wulder *et al.*, 2008a). At the stand scale, lidar has been used for canopy structure information (Lim *et al.*, 2003). For example, Hollaus *et al.* (2006) correlated mean tree stand height, Hyyppä *et al.* (2001) extracted stem volume, Hudak *et al.* (2008) concentrated on the basal area and tree density, while Næsset and Gobakken (2008) and Popescu *et al.* (2003) extracted biomass. At finer scales, a number of studies have focused on extracting forest parameters from lidar at the individual tree level. For example, Lin *et al.* (2011) delineated individual trees in a mountainous forest from small-footprint lidar. Chen *et al.* (2006) and Koch *et al.* (2006) tested a variety of watershed segmentation approaches to detect individual trees, while Persson *et al.* (2002) and Popescu *et al.* (2004) detected individual trees and correlated their measured height and/or crown diameter, respectively. Kaartinen *et al.* (2012) provide a good review of tree detection methods. Individual tree detection algorithms have typically been able to capture between 70 and 80% of individual trees (e.g. Persson *et al.*, 2002).

More recent lidar studies (often combined with optical imagery) focus on extracting fuel metrics across forest landscapes. Here, I refer to fuel metrics as continuous variables that contribute to the amount of overall fuel present at a site (both canopy and surface fuels): canopy height, canopy cover, crown base height (CBH), crown bulk density (CBD), total basal area (BA), shrub height, shrub cover, fuel bed depth, and fuel loads that compose of litter, 1-, 10-, and 100-hour fuel loads. The vast majority of this work focuses on extracting metrics that can be used in the previously mentioned fire behavior models, FARSITE and FlamMap. One of the first examples of a complete protocol for using lidar data in a fire modeling context is provided by Riaño *et al.* (2003). Andersen *et al.* (2005) used lidar to predict crown fuel weight, CBD, CBH, and canopy height of a western hemlock forest in Washington, USA at or above $r^2 = 0.77$. More recently, Erdody and Moskal (2010) used combination of lidar and color infrared (CIR) imagery to predict estimates of canopy fuel metrics in a mixed-conifer forest in Washington. This study correlated height ($r^2 = 0.94$), CBH ($r^2 = 0.78$), CBD ($r^2 = 0.83$), and available canopy fuel ($r^2 = 0.89$) using lidar data alone. Although they did not test for statistical significance, all correlation coefficients were higher when imagery was added to the analyses. Similarly, Peterson (2005) and Peterson *et al.* (2005) analyzed lidar data from Sierra Nevada forests to extract CBH ($r^2 = 0.59$) and CBD ($r^2 = 0.71$). Skowronski *et al.* (2007) computed metrics in the Pinelands of New Jersey using lidar data alone and reported much higher variability in biomass estimation than the previously mentioned studies. The prediction accuracies depended on the forest types; specifically, the accuracies in the lowlands (pine and hardwoods) ranged from $r^2 = 0.59$-0.74, while in the highlands (conifer and deciduous mix) they ranged from $r^2 = 0.11$-0.33. Mitchell *et al.* (2011) and Gatziolis (2011) detected shrub height ($r^2 = 0.86$) and forest cover type (75.6% accuracy) using lidar in a shrub and forest environments, respectively.

Despite this progress, there are few examples demonstrating the efficacy of using lidar to derive surface FMs. One recent case is provided by Mutlu *et al.* (2008), who predicted seven of the original Anderson (1982) FMs using a Gaussian Maximum Likelihood classifier based on fusion of discrete-return lidar (2.6 points/m^2) and QuickBird imagery. The study reported very high accuracy levels:

90% with fusion of lidar and imagery, and 70% with imagery alone. However, the tested environment (small study area on mostly flat ground at an average elevation of less than 200 meters above sea level) is not typical of the mountainous forests in the western USA. The latter forests are characterized by dense mixed-conifer vegetation at high elevations across large, topographically-complex areas. Since these more complex forests face increasing fire risk, evaluating the potential role lidar can play in mapping fuel models is valuable.

This research evaluates the use of small footprint, discrete return lidar data (alone and fused with optical remotely sensed imagery) to extract canopy fuel information from dense mixed-conifer forests across complex topographical terrain. The objectives are to determine: (1) whether relationship exists between surface FMs derived from measured field data and remotely sensed lidar, optical imagery, or combinations thereof; and (2) whether fuel metrics (both those that are directly used in the assignment of FMs, and those that directly feed fire behavior models) can be predicted reliably from lidar, optical imagery, or combinations of these data. The study domain in this research is larger than the previous studies of the same focus (e.g. Erdody and Moskal, 2010; Mutlu *et al.*, 2008) and uses higher density lidar pulse posting, resulting in a more complete representation of the forest. It is also part of a larger study, Sierra Nevada Adaptive Management Project (SNAMP), a multi-discipline collaborative effort among land managers, researchers, and interested stakeholders, designed to explore the effects of coordinated landscape fuel treatments (Collins *et al.*, 2011b).

Methods

I analyzed relationships among lidar, imagery, surface fuel models, and forest structural metrics in a classification and regression contexts. I used a range of classification methods to evaluate the relationship as well as the methods' relative performance. To predict FMs I used simple *k*-means classifier, two regression trees (random forest and classification and regression trees), and support vector machine (SVM) algorithms. To predict fuel metrics I used linear and additive regression models, and regression-based SVM. Based on preliminary results, I used all lidar returns in the analyses described below. The workflow is shown in Figure 2-1.

Study area

My study area is located in the northern part of the Sierra Nevada of California (centered at 39° 07' N, 120° 36' W) in Tahoe National Forest, about 35 km west of Lake Tahoe. It encompasses 99.5 km^2 of topographically complex and steep terrain with elevations ranging from 600 m to 2186 m above sea level (Figure 2-2). The average precipitation since the record began in 1990 is 1182 mm/year in this Mediterranean climate. Almost the entire study area is forested: only 7% of the area is non-conifer forest according to the Tahoe National Forest criteria (<10% of coniferous tree crown area). Fire scars recorded in tree rings from adjacent areas indicate low-severity fires with high-frequency occurrence from 5 to 15 year intervals (Stephens and Collins, 2004).

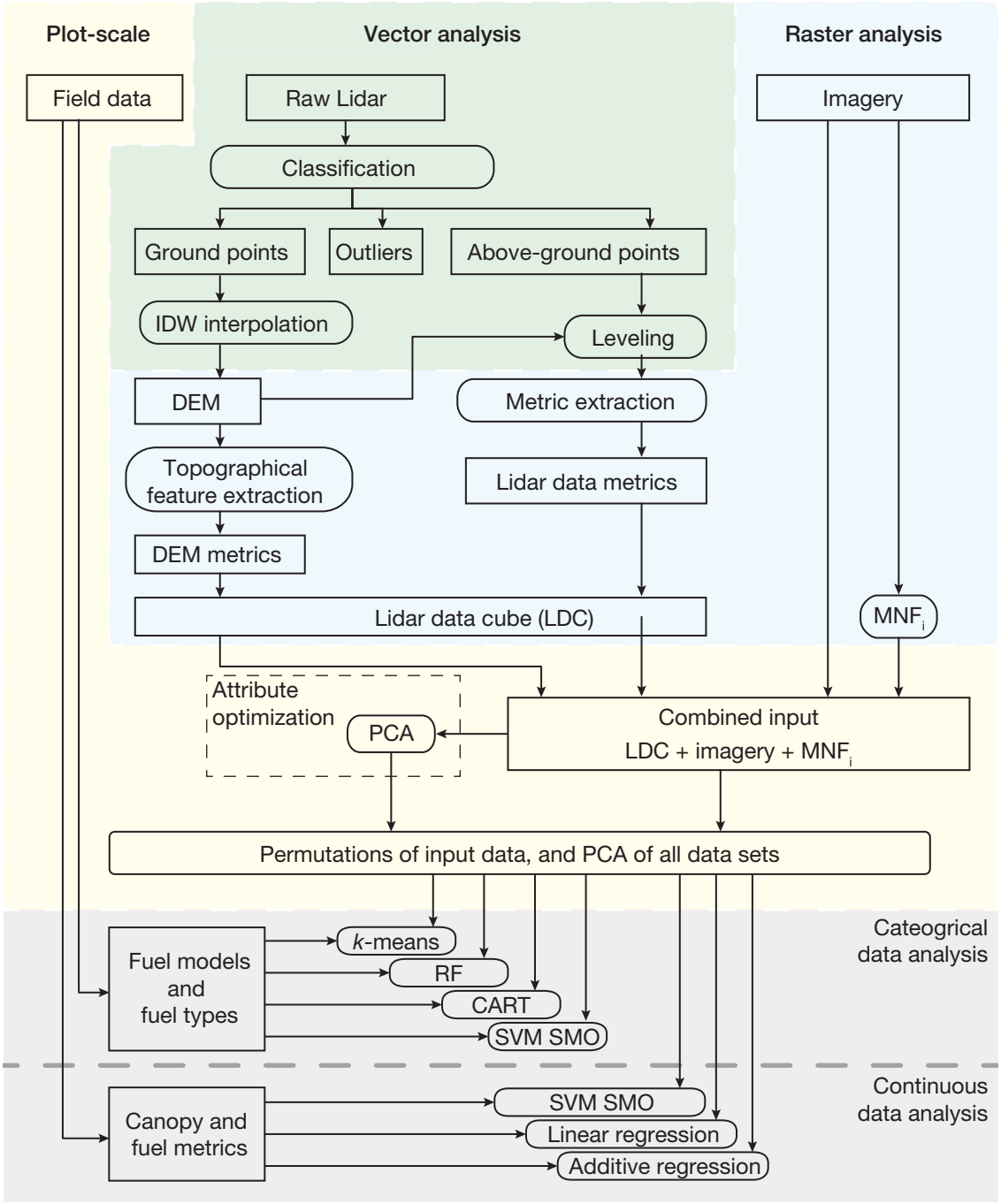

Figure 2-1. The data sources and analysis framework. The analysis included vector-, raster-, and plot-based analysis, as indicated by the background shade. All final analysis was conducted on a plot-scale level.

The dominant vegetation is dense mixed-conifer forest where dominate tree species include sugar pine (*Pinus lambertiana*), ponderosa pine (*P. ponderosa*), incense-cedar (*Calocedrus decurrens*), red and white firs (*Abies magnifica* and *A. concolor*), as well as California black oak (*Quercus kelloggii*). The study area is under management of the US Forest Service, with a few private inholdings scattered throughout.

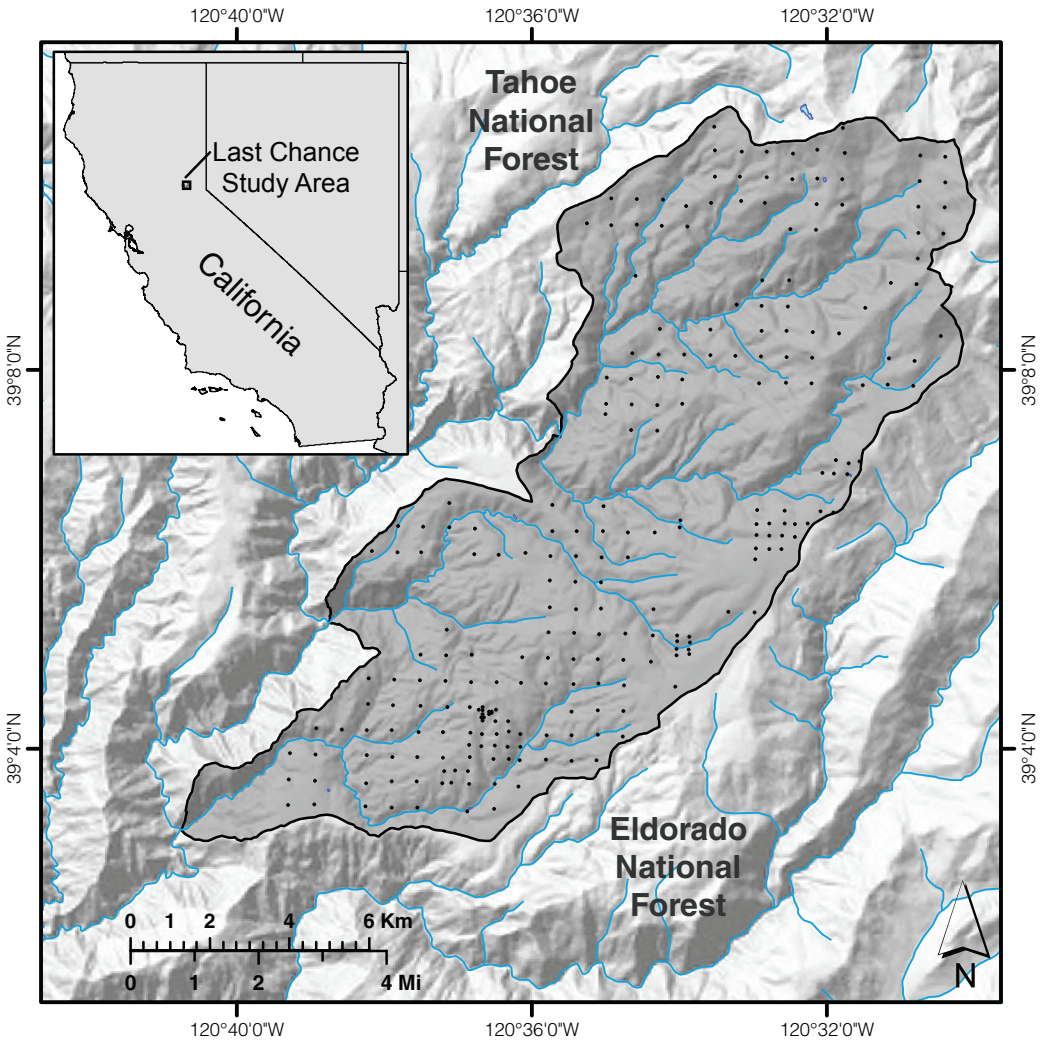

Figure 2-2. The study area in the Sierra Nevada, CA. The points indicate locations of the ground-reference plots.

Field data

Field data characterizing a range of forest parameters was collected at 248 circular plots covering the study area. The plot centers were distributed across UTM grid with 500 m spacing at even coordinate junctions. The plot centers were offset by 25 m in random direction when the desired coordinates fell on road surfaces, landings, rivers, or otherwise physically inaccessible locations. The field data was collected over the course of two summer field seasons in 2008 and 2009; in total, 2340 trees were measured. The plots were extensively mapped for validation of lidar data. Each plot covers 500 m^2 area with 12.62 m radius. All trees above 19.4 cm DBH were tagged with a unique numerical ID. Tree height, DBH, height to live crown base (HTLCB), species, and crown class were recorded along with the unique ID. HTLCB is defined in this study as the lowest extent of vertically continuous live crown on an individual tree. In instances where the crowns of smaller trees are touching the crowns of larger trees the HTLCB measurement includes the smaller tree. Trees with DBH between

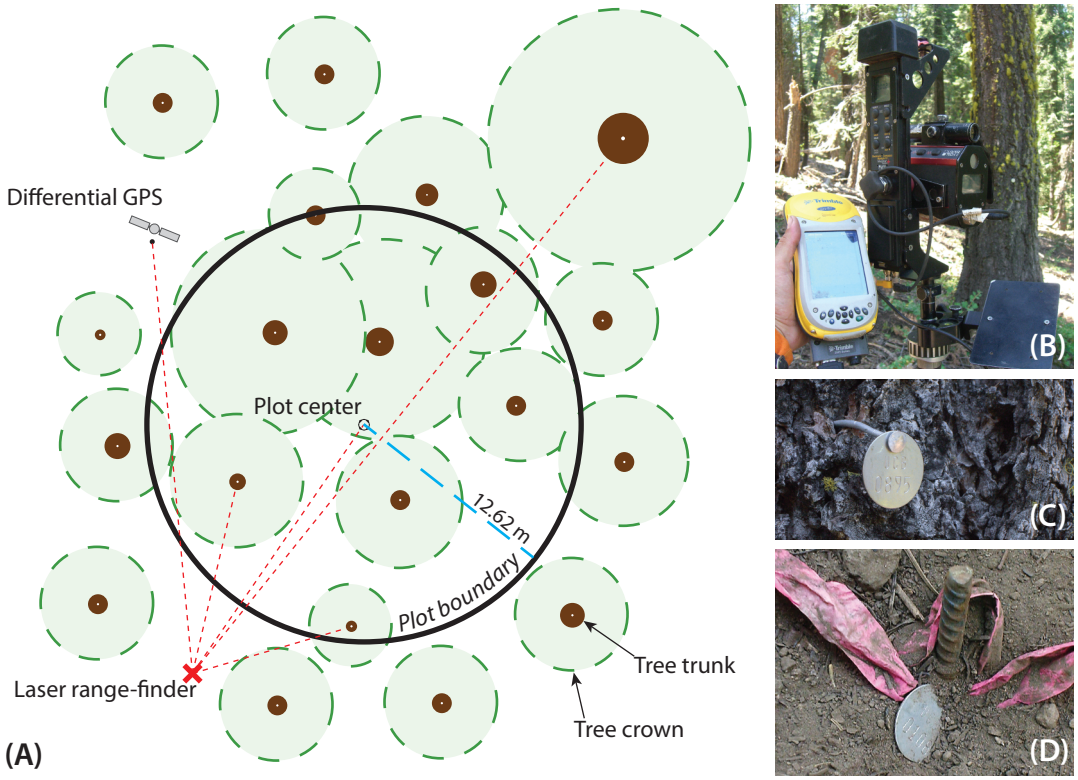

Figure 2-3. Conceptual diagram of field method: (A) example plot showing trees and lidar ground-reference methodology; (B) equipment used to collected positions of trees (C) example of individual tree markers; (D) example of rebar used to mark plot centers.

5 and 19.5 cm were also measured using the above protocol but no numerical IDs were assigned. I also sampled shrub cover and surface/ground fuels using three randomly chosen transects on each plot. The shrub information includes height, percent cover, and species. The fuel information includes tallies of fuel intersections with each transect by fuel particle size class (1-, 10-, 100-, and 1000-hour time lag), as well as, litter, duff, and total fuel heights. For more information regarding fuel inventory methodologies see (Brown and Roussopoulos, 1974; Collins *et al.*, 2011b). Five digital photographs were also taken at each plot in four directions and towards the sky.

In addition to the measurements described above, I generated shapefiles with precise locations and IDs for all tagged trees. I used a combination of Trimble's GeoXH global positioning system (GPS) with an Impulse Laser Rangefinder and an Impulse Electronic Compass to create the stem map shapefiles. At first, two approximately perpendicular angles were used to obtain positions of all trees; however, my analysis throughout the field season indicated that two angles did not sufficiently improve the positional accuracy of the trees to justify their collection at each plot. The positions of the trees were measured and recorded using a laser range-finder and electronic compass combination, and georeferenced to a differential GPS position. The locations of large marker trees were recorded outside of the plot boundary to improve the positional accuracy with respect to lidar data. The individual trees were marked with steel markers, whereas the plot centers were established by rebars hammered into the ground (Figure 2-3). Each plot shapefile includes positions of all tagged trees and their ID,

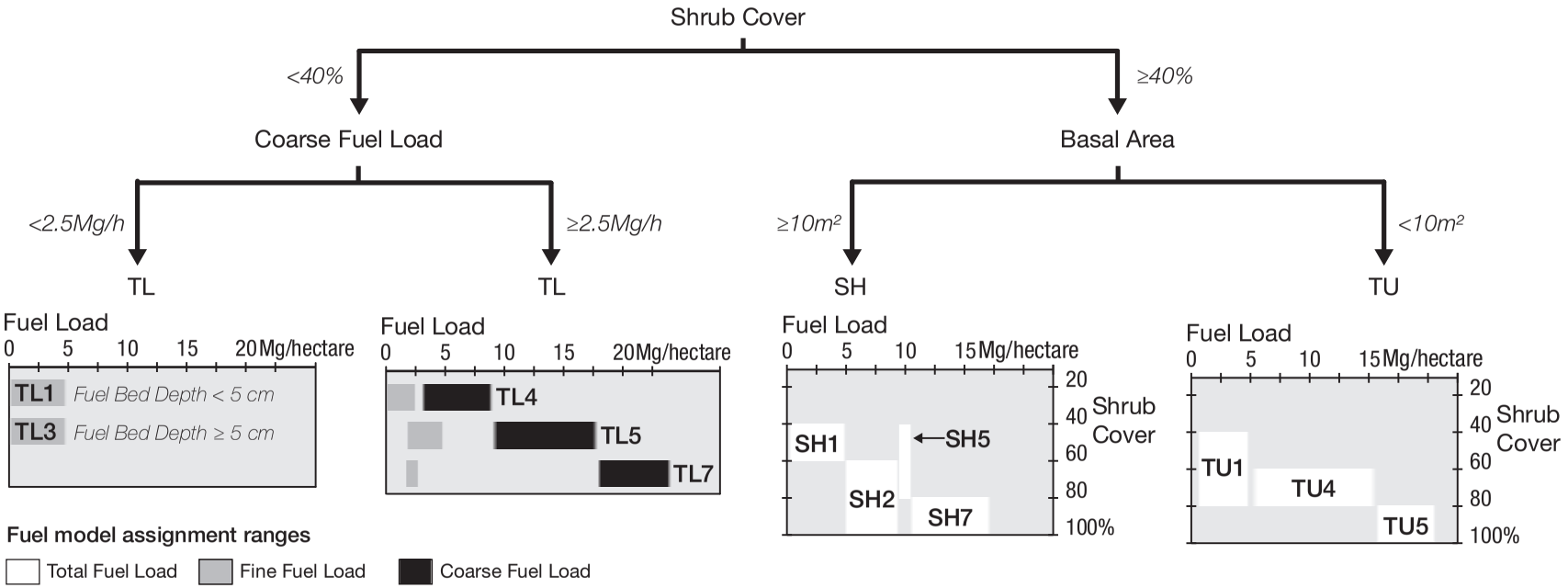

Figure 2-4. Surface fuel model assignment protocol for the study area, adapted from Scott and Burgan (2005). The most critical variables and their criteria are illustrated above. All plots are divided into fuel types (TL, SH, TU) based on shrub cover and then basal area, as illustrated by the decision tree. The TL models are further divided based on coarse fuel loads. The final assignment of plots into specific surface fuel models is based on metrics calculated from ground reference data, including fine fuel loads, fuel bed depth, shrub cover, and total fuel loads.

position of the GPS antenna, and the plot center. The tree positions were later adjusted to true tree centers based on the measured DBH (Figure 2-3). The tree structure data described above (height, species, etc.) were merged to each tree location based on the assigned ID number.

To georeference the stem map shapefiles, I positioned the GPS at most 30 m away from the plot center (typically within the plot or at the plot center) where there was relatively little canopy above to obtain a small positional dilution of precision (PDOP); I filtered the obtained GPS data to a maximum PDOP value of 5. For each GPS position, I collected at least 300 measurements, although majority of positions included about 1,000 and up to 7,700 measurements recorded at 1 second interval. I used Trimble GeoXH differential GPS with Trimble Zephyr Antenna on top of a 3 m GPS antenna pole to minimize multipath problems. Continuously Operating Reference Stations (CORS) and University NAVSTAR Consortium (UNAVCO) stations less than 20 km away from all field measurements were used for differential GPS post-processing. Finally, large, easily identifiable "marker" trees up to 50 m from the plot center were measured and spatially located to increase the accuracy and matching between lidar and ground reference data.

Fuel metrics

In this work, "fuel metrics" comprise the following continuous metrics: canopy height (maximum and mean), canopy cover, BA, CBH, CBD, shrub height, shrub cover, combined fuel load, 1000-hour fuel load, and fuel bed depth. Using the information collected from fuel transects I calculated surface fuel loads for each plot. These calculations were based on the species-specific coefficients reported in Wagtendonk *et al.* (1996; 1998), weighted by the proportion of BA of each species (Stephens, 2001). I summarized these fuel loads into three pools: (1) combined fuel load composed of litter and smaller woody fuels (1- to 100-hour fuel loads), (2) coarse woody fuels (1000-hour), and (3) total fuel. In addition to surface fuel loads, I calculated two canopy fuel metrics at the stand scale: CBH and CBD. CBH is defined as the mean lowest height above the ground at which there is sufficient available canopy fuel to propagate fire vertically through the canopy (Scott and Reinhardt, 2001). CBD is the oven-dry mass of available canopy fuel per unit volume (Scott and Reinhardt, 2001). I calculated these using established allometric equations (Reinhardt *et al.*, 2006a; Reinhardt *et al.*, 2006b) via Fuels Management Analyst (Carlton, 2005) and the collected plot tree measurements.

Fuel models

The field data described above were used to assign FMs. The FMs were assigned using a decision tree protocol adapted for my study area from Scott and Burgan (2005). The basic summary of the protocol is illustrated in Figure 2-4. In particular, fuel load, fuel depth, shrub coverage and height, tree composition, and general forest stand structural information were used to establish repeatable criteria for these metrics. The criteria were then used to systematically assign one of the 40 FMs to each plot. Two classes, "low load dry climate timber-grass-shrub (model TU1)" and "very high load, dry climate timber-shrub (model TU5)," had many more samples (n = 77 and n = 33, respectively),

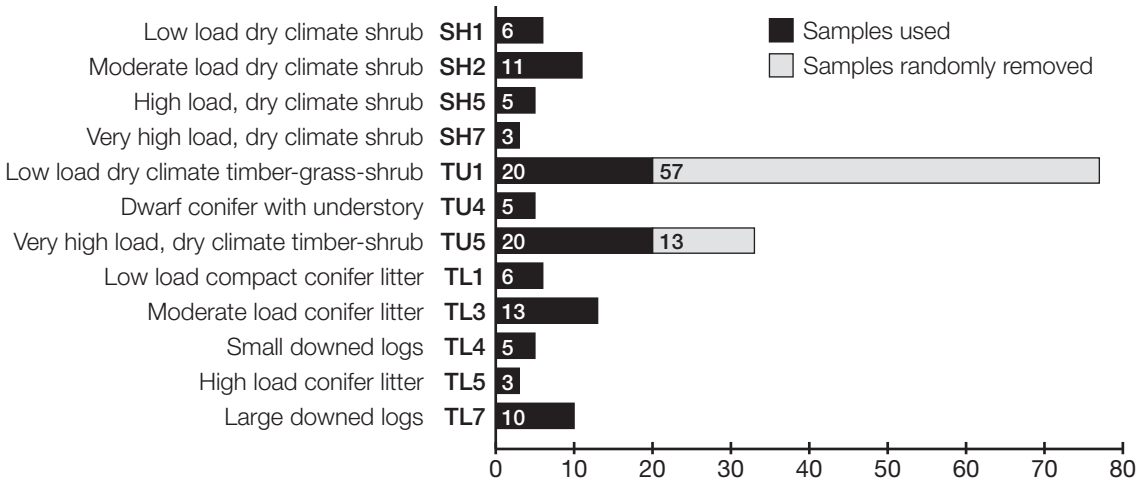

Figure 2-5. List of the standard fuel models within the study area. 70 of the plots were randomly removed from the analysis to equalize the models' selection probabilities.

than the rest of the FM classes. I randomly removed samples from these two FMs to equalize their selection probabilities with respect to the other classes and remove selection bias. In the end I used 12 FMs with N = 107 (Figure 2-5). Sample photographs taken at the time of ground reference data collection (Figure 2-6) depict a few FMs within the study area.

Lidar data

Lidar data was collected by National Center for Airborne Laser Mapping (NCALM) using the Optech GEMINI Airborne Laser Terrain Mapper (ALTM) sensor at an altitude of approximately 1000 m above ground level (AGL). The data was collected in five survey flights from September 18 to September 21, 2008 in leaf-on conditions. The relative horizontal and vertical accuracy were reported and had been independently confirmed to be approximately 5.5-10 cm and an average of 7.5 cm, respectively. Up to four returns per each laser pulse were recorded along with 12-bit dynamic range intensity. Due to steep topography, the pulse rate frequency (PRF) was limited to 70 kHz and the scan angle to ±20 degrees. To increase point density, the aircraft flew twice over the area with a large overlap, such that every ground point was acquired from at least three—and mostly four—angles to yield an average of 9 and minimum of 6 pulses/m^2. Since the lidar system records up to four returns per pulse, the total return density in heavy canopy was often greater than 20 points/m^2. I applied a buffer around the study area to ensure all parts were surveyed and consequently surveyed total area of 107 km^2. The data was delivered in Universal Transverse Mercator (UTM) coordinate system with respect to 1983 North American Datum (NAD83); orthometric heights were computed using NGS GEOID03 model in North American Vertical Datum of 1988.

Lidar pre-processing

The raw lidar data was processed by NCALM using TerraSolid's TerraScan software (Soininen, 2012) to remove obvious outlier points, including isolated point removal (points with no neighbors within 5 meters) and "air point" removal, where points clearly above the canopy are compared to their neighbors. The point cloud was then classified to ground, above-ground, and outlier points using an iterative triangulated surface model. My preliminary analyses indicated that the combination of ground and above-ground points leads to best results thus I used both classes in the analyses. A digital elevation model (DEM) was processed at 1 meter resolution using Inverse Distance Weighted (IDW) interpolation based on suggestions from past research (Guo *et al.*, 2010).

To ensure maximum accuracy, I extracted and only used lidar points within circular area above each plot (12.62 m radius). I developed a set of MATLAB functions to extract lidar metrics in a raster format at a user-defined, horizontal spatial resolution. The lidar metrics (Table 2-1) include descriptive metrics (e.g. *maximum height*, or *number of points from 0.5 to 1 m*) and statistically-based metrics (e.g. *0.05 percentile* and *standard deviation*). The metrics were calculated with respect to ground level. For example, *maximum height* describes the distance between the highest recorded lidar point within a moving window cell and the ground elevation as defined by the DEM. Similarly, *point density 0.5 to 1 m* is the number of lidar returns recorded between 0.5 m and 1.0 m normalized by the total number of returns within a raster cell above the DEM elevation. The plot rasters include a set of bands, each band describing a different lidar data metric. In practice, the user may choose any pixel size to generate a wall-to-wall map, although in this case the pixel size was irrelevant since all lidar data above each plot were summarized into a single point. While there may be a slight difference between results based on a circular footprint and a square pixel, I believe that it would be insignificant.

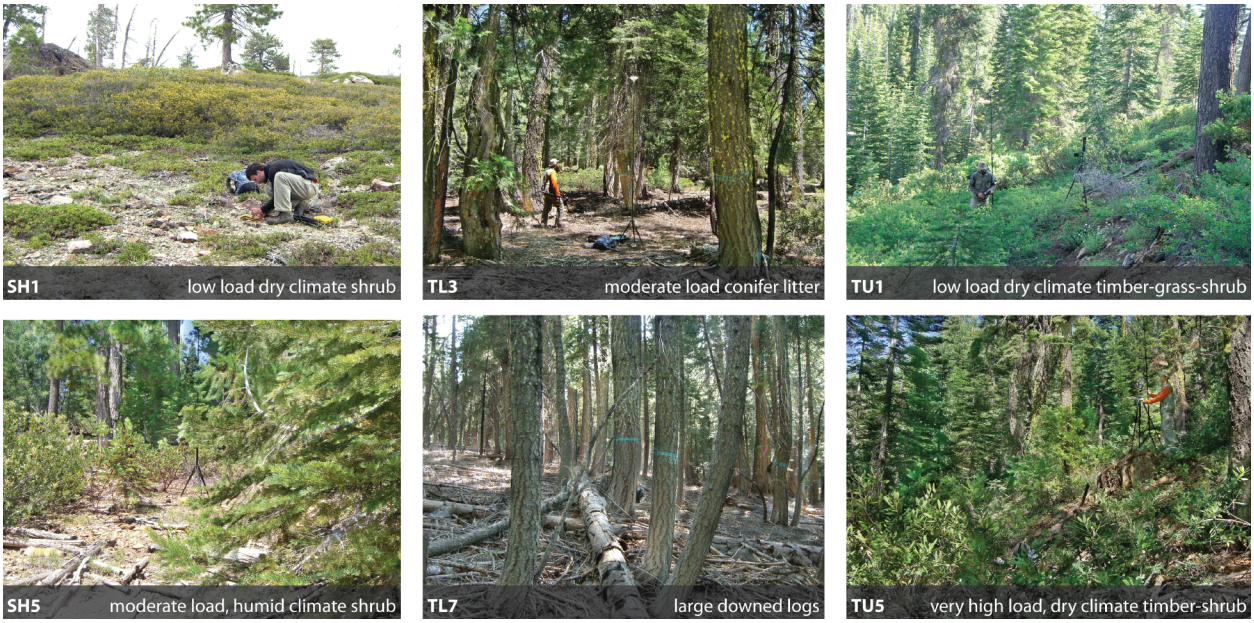

Figure 2-6. Examples of specific fuel models within the study area. The photographs were taken at the time of ground reference data collection.

I then extracted topographical information based on DEM derived from lidar data classified as ground points. All topographical measures (Table 2-1) were derived from the DEM using ITT's ENVI 4.7 Topographical Modeling feature (ITT Visual Information Solutions, 2009). This step was processed at 20 m resolution to be consistent with the remaining spatial and ground reference data. The plot raster data described above were combined with the topographical information into a raster dataset (lidar data cube, or the LDC) with a set of bands similar to a hyperspectral image cube, where each band describes a different lidar or topography metric. The LDC was saved into a 32-bit, floating-point Tagged Image File Format (TIFF) raster format to increase compatibility with external analysis software. A TIFF worldfile and an ENVI header file were generated to preserve metadata and description of each metric. The band data for all plots were then extracted to a table for analysis within a data mining software, as described in the Analysis section.

Imagery

augmented the LDC by adding four-band visible near infrared (VNIR) multispectral imagery from the National Agricultural Imagery Program (NAIP). The 1 m ground sample distance (GSD) imagery was collected and orthorectified in 2009. In preliminary analysis I simply added the multispectral values or their Principal Component Analysis (PCA) and/or Minimum Noise Fraction (MNF) transforms at the pixel level; however, this did not affect the prediction rates of the dependent variables. As a result, resampled the imagery to 20 m using simple averaging to keep all data resolutions consistent. Analysis of the eigenvalues indicated that using four MNF bands was appropriate.

Analysis

I conducted all analyses using 4 combinations of the LDC, raw multispectral imagery and MNF of the imagery (MNF_i) (Figure 2-1). Specifically, I used the following four combinations of the input data: (1) LDC, (2) LDC + multispectral imagery, (3) LDC + MNF_i, and (4) PCA of the combination of LDC and multispectral imagery. All further analyses were performed in Waikato Environment for Knowledge Analysis (WEKA), a set of data mining algorithms compiled into a software package at the University of Waikato, New Zealand (Frank *et al.*, 2010).

Feature optimization

Preliminary analysis using the complete dataset yielded low prediction rates due to overfitting. Thus, I optimized each of the four data combinations by using the optimization routine suggested by WEKA and designed for the subsequent analysis. For example, for data input to random forest I optimized input data with filtered subset evaluator using greedy stepwise search method. In each of these processes, the overall dataset was narrowed to include the most influential input variables in the subsequent analysis. In addition, I ran a PCA on the LDC and imagery dataset and analyzed the 6 principal components based on the eigenvalues which explained 98.04% of the dataset variability.

Table 2-1. Variables that contributed to analyses. The metrics that directly summarize lidar were derived from pulse returns within the field plot footprints; all imagery metrics were calculated on a 20 meter size grid. The lidar heights are calculated relative to the ground level, as indicated by the lidar-derived DEM.

Lidar data cube (LDC)	
Elevation	Point density 0 to .5 m
Slope	Point density .5 to 1 m
Aspect	Point density 1 to 1.5 m
Height: minimum	Point density 1.5 to 2 m
Height: mean	Point density 2 to 3 m
Height: maximum	Point density 3 to 4 m
Height: standard dev.	Point density 4 to 5 m
Percentile 0.01	Point density 5 to 10 m
Percentile 0.05	Point density 10 to 15 m
Percentile 0.10	Point density 15 to 20 m
Percentile 0.25	Point density 20 to 25 m
Percentile 0.50	Point density 25 to 30 m
Percentile 0.75	Point density 30 to 35 m
Percentile 0.90	Point density 35 to 40 m
Percentile 0.95	Point density 40 to 45 m
Percentile 0.99	Point density 45 to 50 m
Total number of returns	Point density 50 to 55 m
	Point density 55 to 60 m

Imagery variables	
NAIP: blue band	MNF component 1
NAIP: green band	MNF component 2
NAIP: red band	MNF component 3
NAIP: NIR band	MNF component 4

Fuel model analysis

I used a number of classification algorithms to evaluate the performance of simple and complex models (e.g. clustering and SVM algorithms, respectively) in predicting FMs using the 4 combinations of input data. The training data was randomly selected from the input data using a 10-fold validation method. I used all the methods described below to classify specific surface FMs as described in Figure 2-5 (e.g. TU1, TU4, TU5, SH1, etc.), as well as generalized surface fuel model types (SH, TU, and TL).

First, I used a simple *k*-means classifier to cluster the input spatial data into FMs. The *k*-means algorithm distributes data into *k* classes by calculating the Euclidean distance between each data point and an estimated class mean in feature space. The means are then reassigned iteratively to minimize the distances among data points and the class means (Duda *et al*., 2001). Next, I used two regression tree algorithms: random forest (RF) and SimpleCART. The RF algorithm uses an

ensemble classifier (multiple models) to obtain better predictive performance. The classifier generates a large number of decision trees at random, a subset of which is chosen to construct the final model (Breiman, 2001). SimpleCART is WEKA's implementation of a classification and regression tree (CART) algorithm (Feldesman, 2002). CART is a nonparametric algorithm that divides a large number of input variables and their interactions based on goodness of split criteria. The process generates a large number of splitting decisions and then applies a set of rules to reduce them (i.e. to prune the tree). CART makes no assumptions about the statistical distribution of its independent or dependent variables, which also means that it cannot generate a probability level or confidence interval (Feldesman, 2002). One of the differences between CART and RF is that RF bootstraps the data to provide a more honest and conservative assessment.

Finally, I used the sequential minimal optimization (SMO) learning algorithm, a type of SVM. In general, SVM algorithms are based on statistical learning theory; they classify datasets by fitting a hyperplane in the feature space to only the data points closest to the class boundaries. These data points are the support vectors. SVM uses structural risk minimization, which minimizes the probability of misclassification based on probability distributions. By utilizing only the points closest to the boundary hyperplane, the SVMs are well suited for classification problems of high-dimensional datasets with a small training sample size (Hsu *et al.*, 2003). Most specifically, the SMO algorithm globally replaces all missing values and transforms categorical attributes into binary classes (Keerthi *et al.*, 2001).

Fuel metric analysis

I used simple multiple linear regression, additive regression, and SVM models to predict the fuel metrics. The additive regression model is an iterative process in which the residuals of a multiple linear regression model are used to construct a new model. All iteration predictions are added to calculate the final prediction model. I reduced the learning rate parameter to prevent overfitting the data (Frank *et al.*, 2010; Friedman, 2002). In this instance, I used WEKA's regression implementation of the SMO in combination with the parameter learning algorithm, RegSMOImproved (Keerthi *et al.*, 2001). I used 10-fold cross-validation as suggested by previous research (Kohavi, 1995).

Results

Fuel model results

I analyzed lidar metrics and CIR imagery to predict the assigned surface FMs using four different classification algorithms. I report the results as the percentage of correctly predicted models based on a comparison to the field data because the dependent variable is categorical (Table 2-2). The best classification was obtained using a combination of lidar and MNF transform of the imagery with the CART algorithm; however, all specific surface FM results were poor (below 50%). The simple *k*-means classifier correctly predicted 24% of the FMs, based on the lidar data alone. The regression

tree classifications algorithms, RF and SimpleCART, predicted between 31% and 45% of the FMs correctly, depending on the data input and the classifier. The CART algorithm most consistently outperformed the RF. The machine learning algorithm, SMO, correctly predicted between 21% and 38%, depending on the data input and the classifier. I used various data input combinations,

Table 2-2. Percent correctly classified standard fuel models and generalized fuel types, given various data inputs and classification methods. The best fit for a given data input is in bold.

	Specific standard fuel models			
	Clustering	**Regression Tree**		**Machine Learning**
Data input	*k*-means	RF	CART	SMO
Lidar alone	24	35	35	**37**
Lidar + imagery	23	35	**40**	37
Lidar + MNF_i	23	42	**45**	38
PCA(lidar + imagery)	15	31	**44**	21
	Generalized fuels types (SH, TU, and TL classes)			
Lidar alone	46	65	70	**75**
Lidar + imagery	49	69	72	**73**
Lidar + MNF_i	46	73	69	**76**
PCA(lidar + imagery)	50	64	66	**75**

including the LDC, imagery, combination of LDC and imagery, and their transforms; I found that with most methods, either lidar data alone or lidar data with MNF of the imagery performed best. The PCA of the entire dataset (LDC + imagery) performed poorly in nearly all cases.

I also classified the broader fuel types: SH, TU, and TL. These results (Table 2-2) improved, suggesting that while lidar may not be able to accurately predict specific surface FMs, it is capable of assessing fuel types. The best overall predictor was the SMO machine learning algorithm (76% correctly classified). The difference among various data inputs was small, ranging from 73% to 76%, and the simplest data input (lidar data alone) attributed to 75% correct classification. Although the regression trees predicted these classes relatively well—RF's best at 75% and CART's best at 72% correctly classified—no algorithm has predicted the fuel types as consistently and accurately as the SMO. I found that the results were not improved after feature optimization of the input data. Consequently, all input variables (within the LDC and derived from the imagery) were used.

Fuel metric results

I analyzed the same set of input data to predict 12 fuel metrics using linear- and additive-regression models, and the SMO algorithm (Table 2-3). I split the metrics into two categories: those that are used to assign FMs (BA, shrub height, shrub cover, combination of 1-, 10-, and 100-hour fuel loads, and fuel bed depth) and those that directly feed fire behavior models such as FARSITE (canopy height, canopy cover, CBH, and CBD). The range in prediction rates is large across models and data inputs but there are trends (Table 2-4). The best prediction rates were associated with maximum canopy

Table 2-3. Correlation coefficients obtained from numerous iterations of remote sensing data, classification algorithms, and measured or calculated fuel metrics. The best fit between model and data input per metric is in bold. The top half of the table lists fuel metrics that are used by analysts to assign fuel models while those in the bottom half of the table are direct inputs to fire behavior models.

			Lidar	Lidar + imagery	Lidar + MNF_i	PCA(lidar + imagery)
Fuel model constituents	Total basal area	Linear Reg	**0.82**	**0.82**	**0.82**	0.67
		Additive Reg	0.73	0.72	0.73	0.71
		SMO	0.81	0.80	0.79	0.81
	Shrub height	Linear Reg	0.54	0.54	0.54	0.46
		Additive Reg	0.52	0.52	0.50	0.40
		SMO	0.58	0.57	0.56	**0.59**
	Shrub cover	Linear Reg	0.48	0.48	0.48	0.40
		Additive Reg	0.52	0.58	0.55	0.47
		SMO	0.59	**0.62**	0.60	0.57
	Combined fuel loads	Linear Reg	0.39	0.39	0.39	0.37
		Additive Reg	0.26	0.28	0.33	0.36
		SMO	0.40	0.43	0.43	**0.48**
	1000 hour fuel loads	Linear Reg	0.16	0.16	0.16	0.23
		Additive Reg	0.29	0.29	**0.31**	0.17
		SMO	0.27	0.27	0.27	0.18
	Fuel bed depth	Linear Reg	0.24	0.24	0.24	0.31
		Additive Reg	0.34	0.20	0.25	0.27
		SMO	0.25	0.29	0.29	**0.35**
Direct fire behavior model inputs	Canopy height (max)	Linear Reg	**0.87**	**0.87**	**0.87**	0.69
		Additive Reg	0.84	0.83	0.82	0.79
		SMO	0.86	0.85	0.85	0.82
	Canopy height (mean)	Linear Reg	**0.60**	**0.60**	**0.60**	0.44
		Additive Reg	0.55	0.54	0.53	0.54
		SMO	0.53	0.51	0.50	0.51
	Canopy cover	Linear Reg	0.62	0.62	0.62	0.74
		Additive Reg	0.72	0.72	0.73	0.76
		SMO	0.78	0.79	0.81	**0.83**
	CBH	Linear Reg	0.39	0.39	0.39	0.35
		Additive Reg	0.36	0.34	0.31	0.34
		SMO	0.32	0.33	0.34	**0.41**
	CBD	Linear Reg	0.14	0.14	0.11	0.13
		Additive Reg	0.20	0.18	**0.25**	0.19
		SMO	0.11	0.12	0.15	0.13

height (r^2: 0.69-0.87), BA (r^2: 0.67-0.82), and canopy cover (r^2: 0.62-0.83), followed by shrub cover (r^2: 0.40-0.62) and shrub height (r^2: 0.40-0.59). All other prediction rates (and all directly related to measures of fuel), were below $r^2 = 0.5$ (combined fuel loads, CBH, fuel bed depth, 1000-hour fuel loads, and CBD, in decreasing order of prediction rates). All other individual fuel load categories (e.g. litter, 1-, 10-, 100-hour loads) yielded very poor correlation coefficients and are not reported here.

Table 2-4, arranged by best Pearson's correlation coefficient and the associated method, provides a simplified summary of all the fuel metric results. The SMO algorithm outperformed the other models in 5 out of the 11 cases using the PCA-transformed data. The SMO performed well mostly with metrics that are more difficult to assess using lidar and imagery: metrics that describe the understory and fuel loads. Multiple linear regression model worked best when predicting less complex metrics (canopy height and BA), in which case data transformations made little or no difference. The additive regression performed best only in cases where the results were poor (1000-hour fuel loads and CBD). Although I performed feature optimization on all input data, I found that the results were not improved. Consequently, all input variables (within the LDC and derived from the imagery) were used.

Discussion

The results of forest structural metrics in general meet or exceed those from previous studies. For example, Lefsky *et al.* (1999) reported r^2 of 0.49 when explaining BA with lidar data compared to my regression model which resulted in an r^2 of 0.82. Lefsky *et al.* (2002) analyzed waveform lidar to

Table 2-4. Summary of continuous fuel and canopy metric results, sorted by best-obtained correlation coefficient. The columns indicate the method and contributing data input used to obtain the most accurate result.

	Best correlation	Method	Lidar	Lidar + imagery	Lidar + MNF_i	PCA(Lidar + imagery)
Canopy height (max)	0.87	L Reg	×	×	×	
Canopy cover	0.83	SMO				×
Total basal area	0.82	L Reg	×	×	×	
Shrub cover	0.62	SMO		×		
Canopy height (mean)	0.60	L Reg	×	×	×	
Shrub height	0.59	SMO				×
Combined fuel loads	0.48	SMO				×
CBH	0.41	SMO				×
Fuel bed depth	0.35	SMO				×
1000 hour fuel loads	0.32	A Reg				
CBD	0.25	A Reg			×	

predict canopy cover and reported mean r^2 of 0.84, 0.63, and 0.11 from boreal coniferous, temperate coniferous, and temperate deciduous forests, respectively; however, when the entire dataset was considered, their estimates decreased to r^2 of 0.37 (as compared to r^2 = 0.83 in my work). Riaño *et al.* (2007) predicted shrub height using lidar at r^2 = 0.48-0.65, which is comparable to my work (shrub height r^2 = 0.59; shrub cover r^2 = 0.62).

In contrast, a few studies predicted CBD with higher accuracy, including Andersen *et al.* (2005), Erdody and Moskal (2010), and Saatchi *et al.* (2007) with r^2 of 0.86, 0.83, and 0.84, respectively. Similarly, Erdody and Moskal (2010), and Andersen *et al.* (2005) predict CBH and ln(CBH) with r^2 of 0.78 and 0.77, respectively. There are a few possible reasons for why the results in my study differ from the these two studies. First, the forest composition and topography of my study area is more complex than the previously mentioned projects, which could affect the penetration of the laser. Second, the CBD metric is quite sensitive to modeling assumptions used to derive it, including the CBH and stand height (Andersen *et al.*, 2005), which may have contributed to the lower correlation coefficients in present work. Third, the Andersen *et al.* (2005) study did not measure crown base heights or tree heights for many trees in the field and therefore the reported results may contain significant errors. Finally, Saatchi *et al.* (2007) summarized his findings based on a radar, not lidar, system. Thus, it would be worth investigating whether radar or waveform lidar systems are more suitable in assessing FMs than a discrete small-footprint lidar system. Radar may be more reliable because of its ability to penetrate and describe some depth of soil, although at much coarser resolution. Waveform lidar may also be beneficial because of its ability to better describe the vertical structure of the forest, and in particular, of shrub and surface level vegetation.

I should mention that in addition to the analysis and results presented here, I have also used lidar + PCA of imagery as an input, and in the FMs analysis, I used another regression tree (C4.5 J48 classifier) and an SVM (LibSVM) algorithm. These results are not presented for brevity; however, in all cases, these results produced neither best nor worst results when predicting either the FMs or the forest metrics.

The fuel metric results varied in accuracy. One contributing factor to the varied success rate is the relative vertical position of the measured metric within the canopy. The metrics near the top of the canopy (e.g. canopy height, r^2 = 0.87) are predicted with better accuracy than metrics near mid-canopy (e.g. shrub height, r^2 = 0.59), or the still worse, near-ground level metrics (e.g. fuel bed depth, r^2 = 0.35). This is further depicted in Figure 2-7. Weakening laser pulse as it penetrates through the canopy towards the ground level and then back up, especially in dense forests, most likely plays the key role in this phenomenon. Further, in general, the fuel metric results indicate that (1) the less complex metrics (e.g. tree height) are best predicted by simple lidar data input and multiple linear regression, while (2) the more complex metrics (related to understory or fuel loads) are best predicted by the SMO support vector machine learning algorithm using PCA transform of lidar and imagery.

Analyses of specific surface FMs, as defined by Scott and Burgan (2005), show that these are difficult to reliably predict from lidar and multispectral imagery. The CART algorithm performed best, although these results were poor and should be considered with caution. There are three main

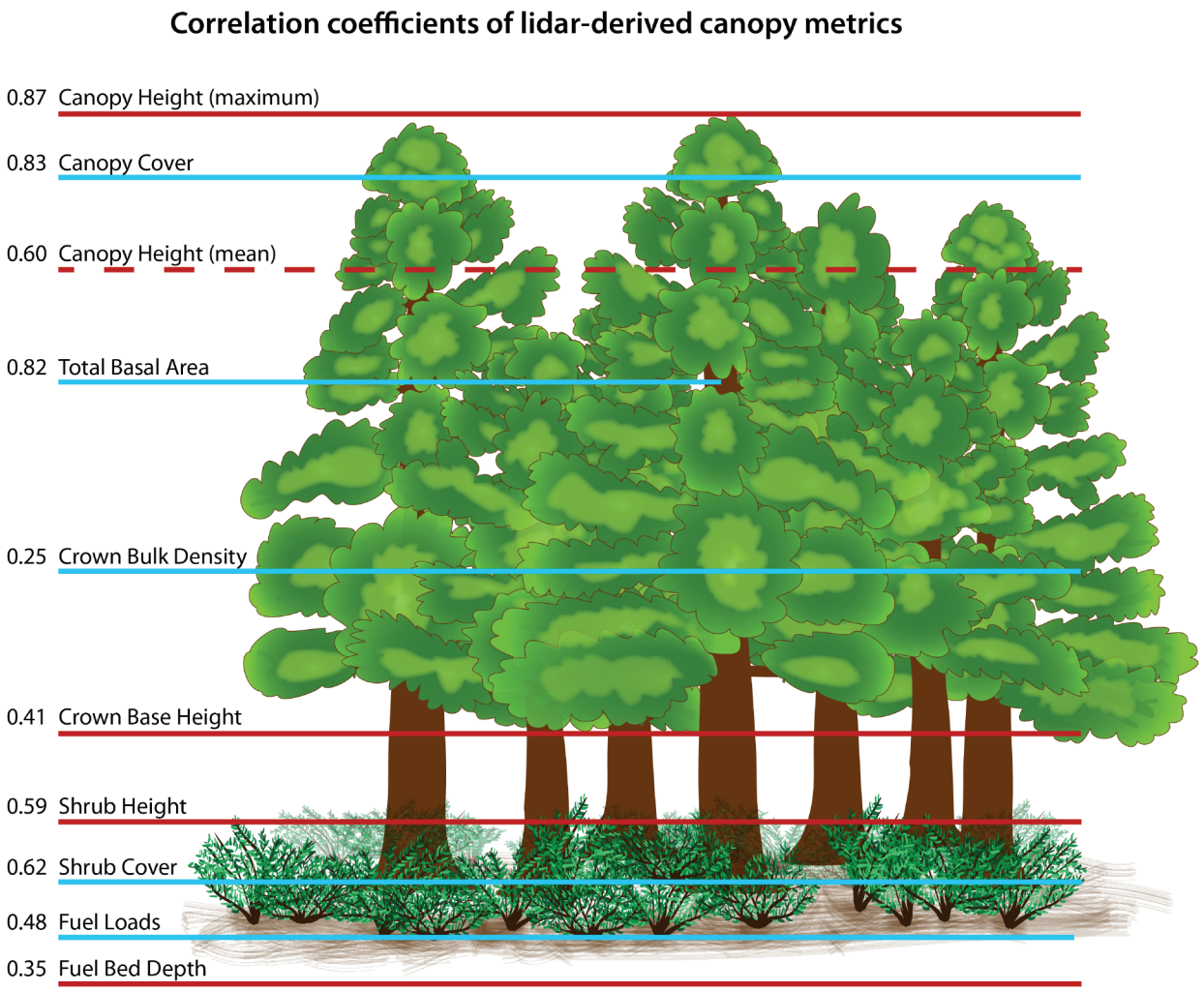

Figure 2-7. The accuracy of the continuous metrics depends, in part, on their vertical position within the canopy.

reasons for this. The first has to do with the study area: steep terrain in combination with dense, mixed-conifer vegetation, especially in comparison to previous studies of similar topics, make characterizing the forest floor problematic. Previous studies have shown that steep slopes and dense vegetation decrease the accuracy of lidar (Hollaus *et al.*, 2006; Su and Bork, 2006; Yu *et al.*, 2005). I hypothesize that the good statistical fits reported by Mutlu *et al.* (2008) between surface FMs and lidar data were in part due to a near-sea-level, flat forest study area consisting of tree plantations and old-growth pine stands. Dense vegetation may influence poor detection rates of FMs because the lidar laser pulse may not adequately penetrate the canopy structure in these situations—critical step in accurately assessing surface fuel structure. It may be worth investigating whether a lidar-based approach works for other terrains on less extreme slopes or sparser vegetation.

The second has to do with the physics of the lidar instrument. The FM assessment as described by Anderson (1982) and Scott and Burgan (2005) depends largely on the amount of dead fuel on

the ground, fuel bed depth, and the moisture of extinction of dead fuels. The detection of these characteristics directly from airborne lidar technology is very challenging. For the most part, these are not directly measured but approximated by the surrounding environment, slope, aspect, and vegetation type.

Finally, FM assignment depends heavily on expert knowledge and is often an iterative process. The expert knowledge may include familiarity with the study area, fire behavior and FMs in the area, and potential reassignment of FMs based on preliminary fire behavior model outputs. As a result, there is not always good correspondence between FMs and the original data from which they were derived. This is further shown by the fact that lidar can predict the general fuel types but not the specific FMs. Predicting FMs using lidar in a classification context may therefore not be desirable.

I do not want to discount the potential benefits of using lidar in FM assessment. As demonstrated in this study, using lidar and/or imagery to detect general fuel model types (SH, TU, and TL) works reasonably well: up to 76% correct classification. The SMO algorithm consistently performed best in predicting these fuel types in comparison to statistical clustering or regression models. Further, lidar is capable of mapping some canopy metrics well (canopy cover, tree height, BA, etc), and this information, in addition to general fuel types and field data, could be used by experts to improve future FM assignment. While lidar data may not provide the perfect solution in fuel mapping, it may be the best available option and can still provide a more reliable answer than the currently accepted and mostly unreported methods used for fuel model assignment.

Conclusions

In this study, I used a range of algorithms and combinations of discrete, small-footprint lidar and multispectral imagery to predict standard surface fuel models, their constituent inputs, and fuel metrics typically used in fire behavior models such as FARSITE. I performed an extensive analysis to determine what is the optimal combination of data inputs and methods in order to predict each of the above forest metrics and/or fuel models.

Results indicate that the specific surface fuel models are difficult to predict reliably using lidar and imagery data in a dense forest within complex terrain, regardless of the input data transformation or the methods used. However, more general fuel types were detected at a reasonable rate of 76% accuracy using the SMO version of SVM algorithm. The SVM predicted above 73% of fuel types using all data inputs. My analyses show that deriving canopy stand structure or continuous fuel metrics is repeatable and accurate. In general, I found that to derive the less complex metrics (e.g. tree height), lidar data alone with simple multiple linear regression works best, while for the metrics that are more difficult to measure (e.g. fuel loads), best results are obtained through more sophisticated analysis with SVM using PCA transform of lidar and imagery as the input. Further, as the lidar pulse weakens when it penetrates down through the canopy, so does the pulse density and the prediction accuracy level of the associated metrics. In particular, maximum canopy height, canopy cover, and BA were described with up to 0.87, 0.83, and 0.82 Pearson's correlation coefficient,

respectively. As I move deeper into the canopy, the ability to reliably predict fuel metrics declines. The best shrub prediction was 0.62, while the prediction of ground-based fuels declined below 0.50 correlation coefficient.

Chapter Three

Delineating individual trees from lidar data: a comparison of segmentation methods

Introduction

Remote sensing has been established as one of the primary tools for broad-scale analysis of forest systems. Metrics relating to forest structure such as tree height, canopy cover and biomass can be estimated using remote sensing (Hyyppä *et al.*, 2009; Jensen, 2007; Lefsky *et al.*, 2002). More recently, as sensor technology has improved and expanded to yield higher resolution optical data as well as light detection and ranging (lidar) data, it has become increasingly possible to detect individual trees. The successful detection and delineation of individual trees is critical in forest science, allowing for studies of individual tree demography and growth modeling (Clark *et al.*, 2004; Falkowski *et al.*, 2008; Vepakomma *et al.*, 2011), understanding of wildlife habitat and behavior (Garcia-Feced *et al.*, 2011a; Zhao *et al.*, 2012b), and more precise measures of biomass in forests (Popescu *et al.*, 2003; Zhao *et al.*, 2012a).

Many approaches have been developed to detect and delineate individual trees from remotely sensed data (Figure 3-1). Early on, studies focused on assessing individual trees based on optical imagery. Sheng *et al.* (2001) used stereoscopic aerial photographs to identify, model, and measure a single coniferous tree using photogrammetric techniques. Later studies used a variation of local maxima (Maltamo *et al.*, 2004) and adaptive binarization (Pitkänen, 2001) for tree *detection*, as well as region-growing (Wang *et al.*, 2004), edge detection (Brandtberg and Walter, 1998), and valley-following (Gougeon and Leckie, 2006) for tree *delineation*.

With the wide introduction of lidar into remote sensing, an increasing number of studies have undertaken individual tree detection. Through time, these studies have shown increased complexity of analyses, increased accuracy of results, and a focus on the use of lidar data alone. A combination of variable window size filtering (Chen *et al.*, 2006; Popescu and Wynne, 2004), spatial wavelet analysis (Falkowski *et al.*, 2008), *k*-means clustering (Morsdorf *et al.*, 2004), and morphological analyses (Kwak *et al.*, 2007) were used to identify trees and estimate their biophysical parameters. Brandtberg *et al.* (2003) Gaussian-smoothed a canopy height model (CHM) at numerous scales to create a scale-space structure and then detected trees which were represented as "blobs" in the scale-space structure. However, most commonly, a variation of region-growing algorithms was

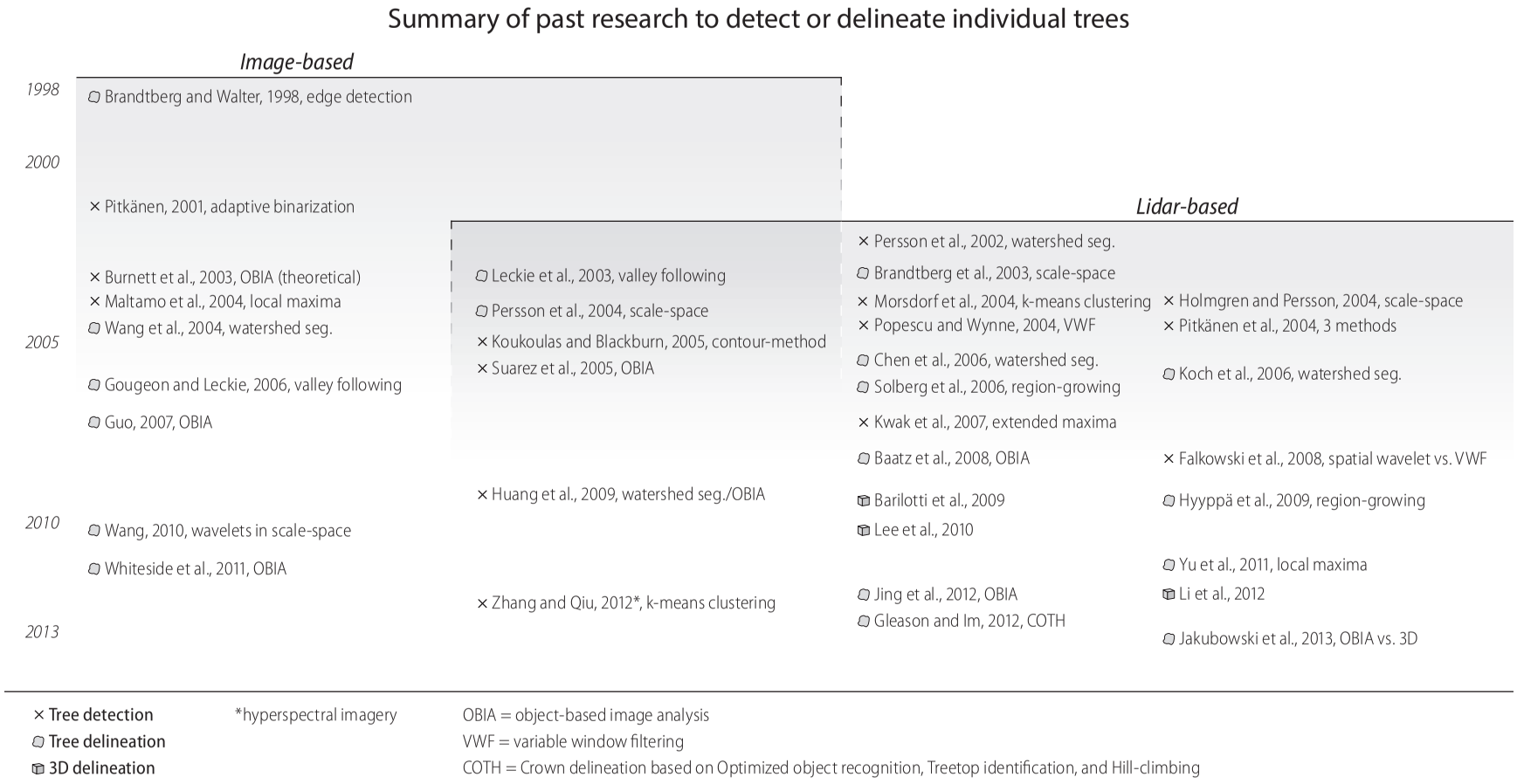

Figure 3-1. Past studies that detect or delineate individual trees from either imagery, lidar, or their combination. The published work is arranged by year, from top to bottom.

applied for tree delineation from CHM (Chen *et al.*, 2006; Hyyppä *et al.*, 2001; Koch *et al.*, 2006; Solberg *et al.*, 2006; Yu *et al.*, 2011). The latter approach was later used to predict tree volume on a semi-individual tree level (Breidenbach *et al.*, 2010). All of the region-growing studies relied on first identifying the tree tops with a local maximum filter.

More recent studies focus on full exploitation of the point cloud and do not transform the lidar data into a raster. Lee *et al.* (2010) implemented an adaptive clustering algorithm similar to watershed segmentation except that the method was applied to the 3D lidar points. Their algorithm relies on training data to perform segmentation based on supervised learning. Li *et al.* (2012) designed and tested an individual tree segmentation approach that uses the full lidar point cloud. The algorithm exploits the spacing between the tops of trees to identify and group points into a single tree based on simple rules of proximity and likely tree shape.

Simultaneous with developments in lidar technology has been the increase in spatial resolution of optical image sensors. Classification algorithms developed for courser imagery often perform poorly at fine scales due to high local heterogeneity (Cleve *et al.*, 2008; Guo *et al.*, 2007; Kelly *et al.*, 2011), and thus finer resolution imagery has necessitated a change in the way imagery is processed: from pixel-based to object-based image analysis (OBIA). The OBIA approach first segments imagery into meaningful objects through the use of spectral and spatial context prior to classification. This allows for image processing and Geographic Information System (GIS) functionality (Blaschke, 2010). The OBIA approach is inherently multi-scalar, as image segments are captured and operate hierarchically. At the landscape scale, studies have successfully extracted biophysical forest inventory parameters from imagery (Chubey *et al.*, 2006) and used lidar to classify land covers (Antonarakis *et al.*, 2008; Brennan and Webster, 2006). At stand-scale, Ke *et al.* (2010) used a combination of lidar and optical imagery to classify tree species. A number of geospatial image segmentation programs are available, including BerkeleyImageSeg (Clinton *et al.*, 2010) and Trimble eCognition (Trimble, 2012).

Less research has been done using OBIA to extract individual trees (Baatz *et al.*, 2008; Guo *et al.*, 2007), and to date, there have been no comparisons of OBIA-derived approaches and lidar point cloud segmentation. In this chapter, I compare two paradigms of segmentation for individual tree delineation. The first approach uses 3D lidar point clouds classified into individual trees as developed by (hereafter called "lidar-derived"), and the second uses an OBIA approach with surfaces derived from lidar data (hereafter called "OBIA"). The latter analysis uses a novel method for correcting artifacts in a CHM by filling in erroneous height values. I evaluate both types of tree polygons in terms of their area, shape index, compactness, and ability to predict ground reference tree height, all across a gradient of tree densities. Finally, I classify the tree objects into species and species type, and validate the results against ground reference data using a 10-fold validation.

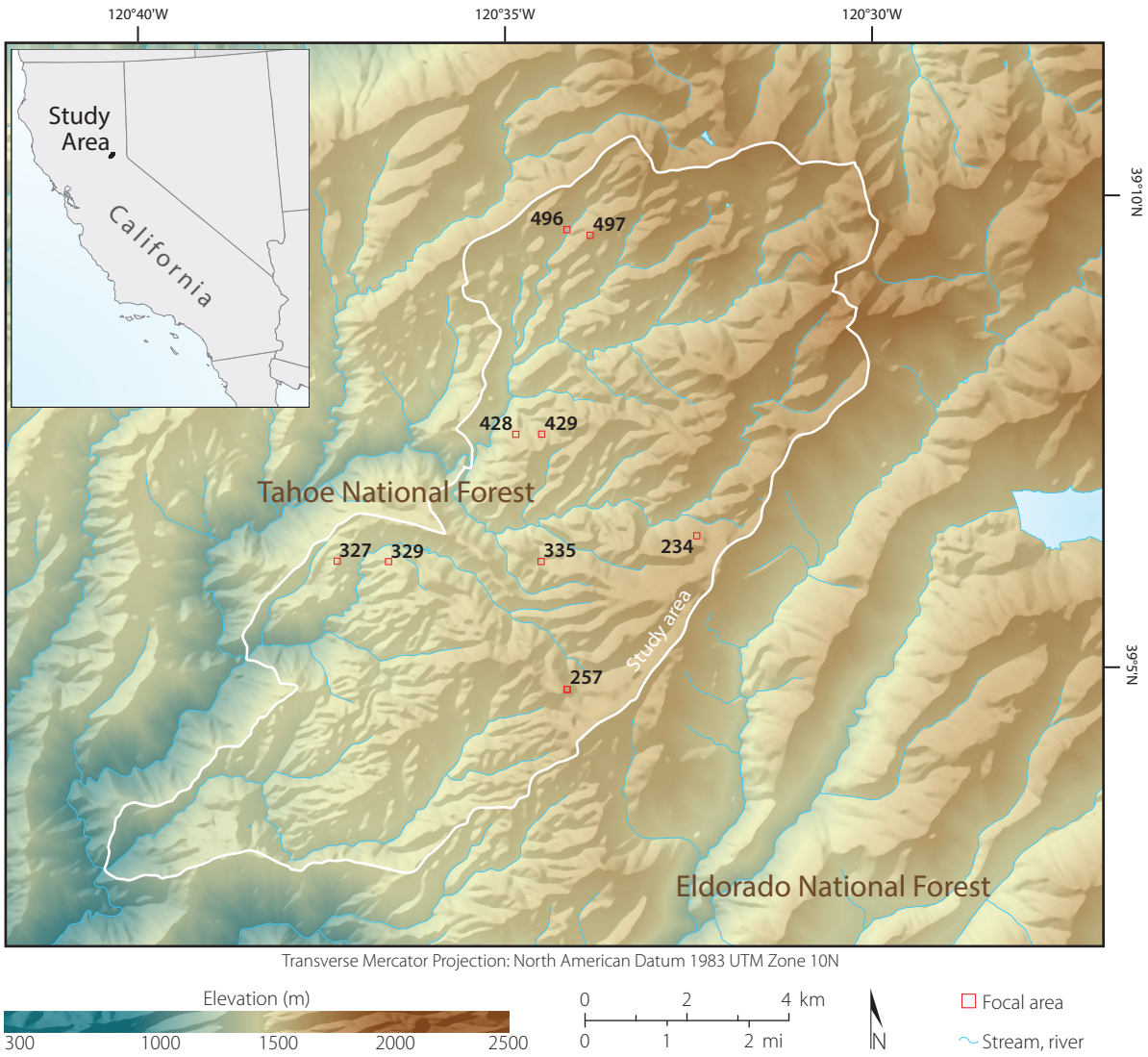

Figure 3-2. The study area (9,950 ha), mostly within Tahoe National Forest, is covered by mixed conifer forest, and covers topographically complex terrain. The nine 90m squares are indicated by red lines.

Methods

Study area description

The study area (9,950 ha) is comprised of topographically complex mixed-conifer forest within California's Sierra Nevada mountain range (center: 39° 07' N, 120° 36' W). The elevation varies from 600 to 2,186 m above sea level and much of the land is characterized by steep slopes (Figure 3-2). The forest is mostly managed by the US Department of Agriculture, Forest Service, with a few small private inholdings. The forest composition is predominantly conifer, with the dominant species being red and white fir (*Abies magnifica* and *Abies concolor*, respectively), sugar pine (*Pinus lambertiana*), ponderosa pine (*Pinus ponderosa*), and incense-cedar (*Calocedrus decurrens*). Less than 10 percent

of the forest is non-conifer, dominated by the California black oak (*Quercus kelloggii*). Overall, the area experiences typical Mediterranean climate conditions: an average precipitation of 1182 mm/year and frequent (5 to 15 year intervals) low-severity fires (Stephens and Collins, 2004).

Ground reference data

This experiment is part of a larger study—Sierra Nevada Adaptive Management Project—in which 411 circular plots (area = 500 m^2) were established throughout the study area to characterize the fire and forest health ecology. The plots are centered at even coordinates (e.g. 700000 m E, 4333000 m N; 700500 m E, 4333500 m N; etc.) according to the Universal Transverse Mercator (UTM) system. In the work described here, I chose nine plots (Figure 3-3) that are representative of three overstory tree densities: low (<200 trees/ha), medium (≥200 to <400 trees/ha), and high (≥400 trees/ha). Plots were classified according to tree density how. I define an overstory tree as one whose diameter at breast height (DBH) ≥19.5 cm. Within each plot, I geolocated trees according to a spatial protocol (described below), and collected measurements relating to trees (height, species, dominance, DBH, height to live crown), shrub (species, and height), and fuels (fuel loads and ladder fuels) More information can be found in (Jakubowski *et al.*, Accepted).

The trees were divided into three classes: overstory (DBH ≥19.5 cm), mid-size (5 ≤ DBH < 19.5) and small (DBH < 5cm). All overstory trees were tagged with a unique identification number and georeferenced according to the spatial protocol described below. Although all trees within the plots were measured and described, in this work I only consider the overstory trees (n = 141) since the position of the smaller trees were not recorded. In addition to the overstory trees, I selectively measured and georeferenced large and/or isolated "marker" trees (n = 57) outside of the plot, for the purpose of improving the match between lidar and ground reference data. Both, the overstory and the marker trees (n = 198) were used in the analyses.

The plots and the trees were georeferenced consistent with a standardized spatial protocol. I stabilized a Trimble GeoXH differential Global Positioning System (dGPS) in the most open canopy area within or near the plot (≤30 m from the plot center), where it recorded at least 300 measurements below positional dilution of precision (PDOP) of 5. In most circumstances, I recorded between 1000 and 7000 measurements for each plot position. I used a Trimble Zephyr antenna elevated 3 m from the ground to reduce multipath issues. The dGPS data were post-processed using University NAVSTAR Consortium (UNAVCO) and Continuously Operating Reference Stations (CORS), all located ≤20 km from the field measurements. To georeference the trees, I used an Impulse Laser Rangefinder and Impulse Electronic Compass. The electronic compass was carefully stabilized and calibrated on each plot before recording the trees' positions. I used the "filter" option in the range finder, so that it would only record distances to a designated target (positioned on the measured trees). Since the rangefinder requires a clear line of sight to its target, measurements were collected from two and sometimes three positions in plots with very high tree or shrub density. They were later combined in GIS. Since the error of the laser rangefinder is ≤2 cm, and the compass degree error ≤0.5 degrees, I only considered marker trees <50 m from the measuring unit. For the present work,

Table 3-1. WorldView-2 specifications of the spectral bands.

Band	Spectral region	GSD (nadir)	GSD (20° off nadir)
1 (Coastal)	400 - 450 nm	185 cm	207 cm
2 (Blue)	450 - 510 nm	185 cm	207 cm
3 (Green)	510 - 580 nm	185 cm	207 cm
4 (Yellow)	585 - 625 nm	185 cm	207 cm
5 (Red)	630 - 690 nm	185 cm	207 cm
6 (Red edge)	705 - 745 nm	185 cm	207 cm
7 (NIR1)	770 - 895 nm	185 cm	207 cm
8 (NIR2)	860 - 1040 nm	185 cm	207 cm
Panchromatic	450 - 800 nm	46 cm	52 cm

all tree positions were manually verified and in a few cases moved to increase the correspondence to the lidar data.

Multispectral imagery

I used DigitalGlobe's WorldView-2 data in the species prediction part of this work (described below). The imagery includes eight spectral and one panchromatic band (Table 3-1). A single image, acquired on June 14th, 2010 at 12:14 local time (sun elevation = 71.3°), was used for the analysis. The data was delivered in LV1B product level: 16-bit and radiometrically corrected. The sky was clear (cloud cover = 0.077%) and no snow was present in any of the analyzed plots.

The data was orthorectified and then pan-sharpened before the analysis. The orthorectification was performed in EXCELIS's ENVI 5 (EXELIS, 2012) software using a combination of a Rational Polynomial Coefficients (RPC) sensor model and a 0.5 m digital elevation model (DEM) generated from the lidar. The image and the DEM both used bilinear resampling. I used –24.557 m as the geoid offset based on National Geodetic Survey's GEOID99 model computed at the center of the image scene. The data was projected into the UTM (zone 10N) coordinate system with respect to 1983 North American Datum (NAD83) at 50 cm pixel ground sample distance (GSD) to match the remainder of the data. I pan-sharpened the spectral data cube using the Gram-Schmidt algorithm and bilinear resampling (Laben and Brower, 2000).

Lidar data

Lidar data was acquired between September 18th and 21st, 2008 by the National Center for Airborne Laser Mapping (NCALM) using the Optech GEMINI Airborne Laser Terrain Mapper (ALTM) sensor. Flying at an average height of 1000 m AGL, the sensor collected up to four returns per pulse at an average of nine and a minimum of 6 pulses/m^2. I confirmed the horizontal and vertical accuracy to be 5.5-10 cm, and an average of 7.5 cm, respectively. The instrument was operated at a 70 kHz pulse rate frequency (PRF), with a 40 Hz scan frequency, and a ±20 degree scan angle. All even-numbered flight lines were flown twice in order to obtain high point density. The final point

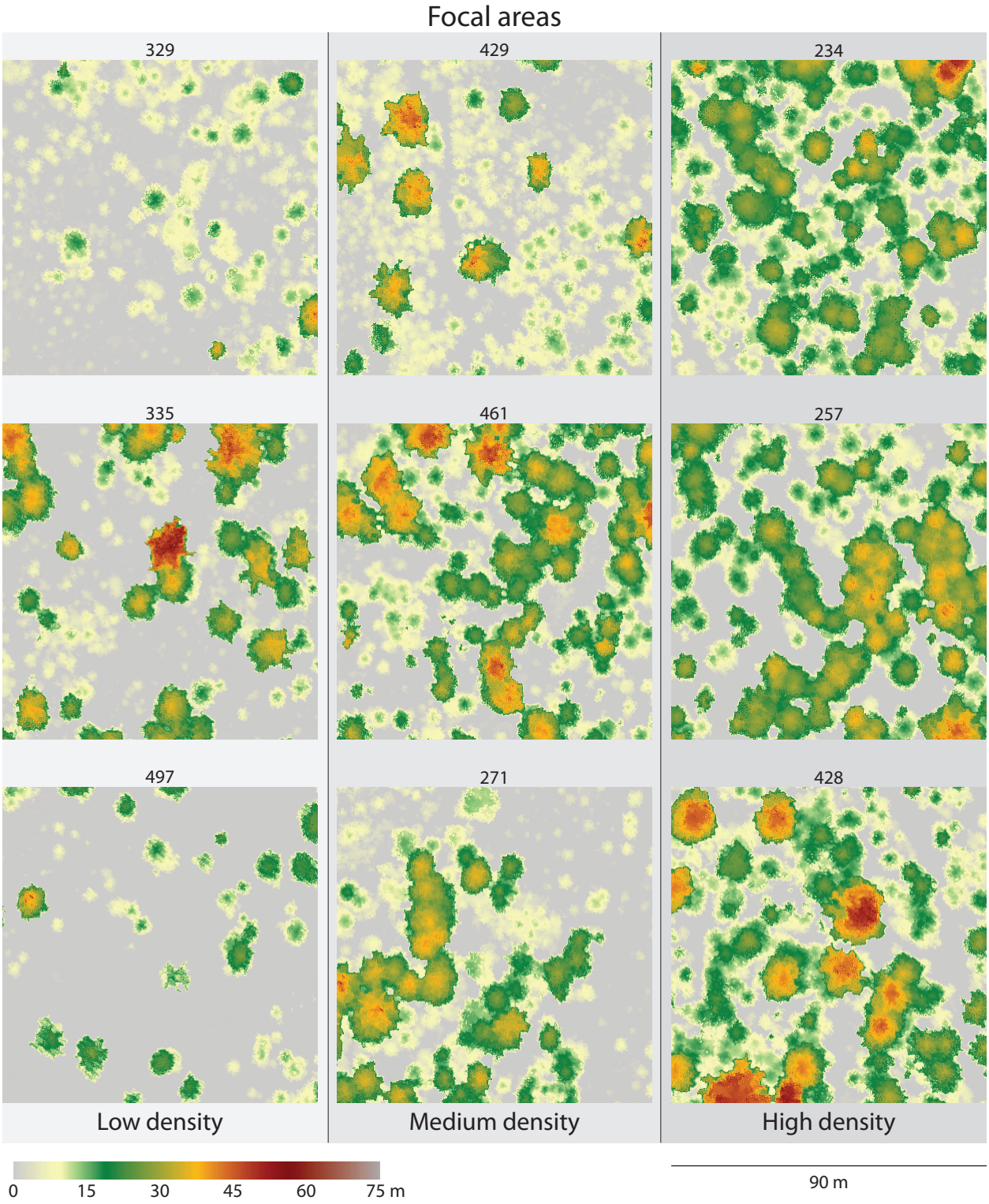

Figure 3-3. Nine 90m square focal areas were segmented using the lidar-derived and OBIA-derived methods. The areas varied across tree density, from low (left column), to high (right column).The color above corresponds to corrected CHM height values.

cloud often exceeded 20 points/m^2 because of dense vegetation. All data were delivered in the UTM coordinate system with respect to NAD83 and the 1988 North American Vertical Datum using the NGS GEOID03 model. NCALM preprocessed the data to remove erroneous below-surface and air points. A 5 m search radius was used to identify and remove isolated points. The points were classified as ground and above-ground using TerraScan's iterative triangle-building surface model classification (Chang *et al.*, 2008). From the entire point cloud, I extracted 100 m squares centered at plot centers.

Segmenting the point cloud into individual trees

Individual trees were s using a tree segmentation algorithm (Li *et al.*, 2012). The algorithm exploits the spacing between the tops of trees to identify and group points into a single individual tree. An adaptive spacing threshold was used to improve the segmentation accuracy: dt = 2 and 1.5 m when tree heights were ≥15 and <15 m, respectively. I did not use a minimum spacing rule or the shape index threshold as it did not improve the accuracy of the classification. Prior to the segmentation, the point cloud was normalized by subtracting the DEM ground values (Lee *et al.*, 2010). Following the segmentation, I converted the segmented point cloud into two dimensional tree polygons by calculating the concave hull around the lidar points identified as a single tree (Figure 3-4). Further, each tree's maximum-height point was extracted for later analysis.

Preparing the data for image segmentation

I generated a DSM from the 100 m square point clouds using LAStools. The DEM was generated only from points classified as ground. I determined that triangulated irregular network (TIN) interpolation was sufficient for generating these surfaces due to the very high spatial sampling (20 cm GSD). Preliminary analysis indicated that it was an appropriate resolution for accurate crown delineation—for example, 50 cm GSD surfaces were too course to create meaningful individual tree segments. I then generated a CHM by simple image subtraction:

$$CHM = DSM - DEM \tag{1}$$

Correcting for CHM artifacts

An accurate delineation of the tree crowns required generating a CHM at a high resolution (20cm GSD), such that there was not always a lidar point within each pixel. As a result, the CHM contained a number of artifacts that prohibit successful segmentation using the OBIA framework. In particular, the CHM contains many severe elevation drops, especially in the middle of the canopy. The elevation drop artifacts are directly related to the scanning pattern of the lidar sensor, i.e. the z-value

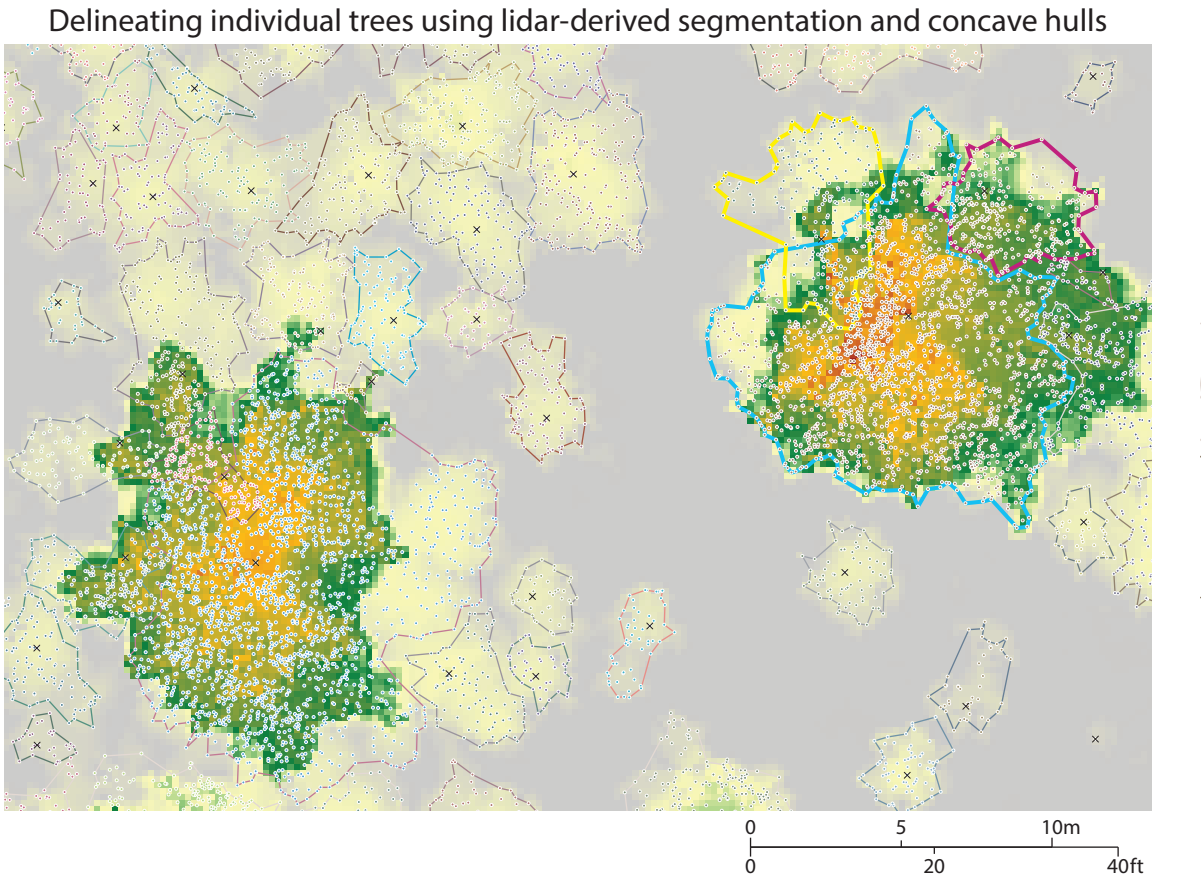

Figure 3-4. The normalized lidar point cloud was segmented into individual trees using 3D point classification routine (left). The points classified by individual trees were then delineated by a concave hull in the x-y space (right). The lidar-segmentation algorithm allows for polygons to overlap.

thin troughs are located in between the lines of lidar scans (Figure 3-5). The challenge is further complicated since the artifact pixels are *not* no-data values; rather their value is well within a typical CHM. This problem has been noted before, especially when attempting to delineate individual trees (Leckie *et al.*, 2003). Filtering the image using a blurring or a median filter is one solution, albeit with negative consequences: the effective sharpness and thus the ability identify the trees' edges diminishes substantially (typically at least a 5×5 or a 7×7 kernel must be used to minimize the artifacts). Another option is to interpolate the surfaces at a higher pixel size, although that approach also suffers from the above consequences. I identified a new method to mitigate this problem.

First, a CHM was smoothed such that its continuity across tree crowns is restored and, simultaneously, the crowns' distinct edges are preserved. This can be achieved with an adaptive enhanced Lee filter (Lee, 1980; Lopes *et al.*, 1990) or through morphological filtering (Haralick *et al.*, 1987). The enhanced Lee filter reduces speckle while preserving texture. It uses the local coefficient of variation to classify pixels as either homogeneous, heterogeneous, or point targets, and then replaces the targeted pixels based on the classification. The second method uses two cycles of the morphological closing operator (dilation followed by erosion). Closing filtering fills the small holes and fuses thin troughs. In this work I used the morphology filter approach (Equation 2) as it produced better results with my data.

$$CHM_{smooth} = (CHM \bullet k) \bullet k, \text{ where } k = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{2}$$

Although the smoothed CHM is an improvement over the original, strive to use the original CHM values as much as possible and thus the next few steps develop a technique that replaces the artifact pixels while preserving the surrounding original correct height values. Ultimately, I derived a continuous mask that was used to mix the two images: the correct pixels from *CHM* and the replaced artifact pixels from CHM_{smooth}.

A simple mask was generated by subtracting the artifact CHM from the smoothed CHM:

$$Mask_{diff} = CHM_{smooth} - CHM \tag{3}$$

The simple mask was normalized to make the process reproducible for landscapes where the CHM values may vary to a smaller or larger degree:

$$Mask_{norm} = \frac{Mask_{diff}}{\text{MAX}(Mask_{diff})} \tag{4}$$

A binary mask was created by reducing near-zero values to zero and all other values to one:

$$Mask_{Boolean} = \begin{cases} 0, Mask_{norm}(x,y) \leq a \\ 1, Mask_{norm}(x,y) > a \end{cases}, \text{ where } a \in [0, 0.1] \tag{5}$$

The Boolean mask was convolved with a simple matrix and the normalized mask was added. To increase the impact of the normalized mask, I multiplied it by a constant, *b*. The convolution part of this operation ensures that no hard edges are generated when the smoothed CHM pixels are added to the original CHM. Setting *b* to a very high number decreases the feathering effect when adding back the smoothed CHM pixels.

$$Mask_{smooth} = Mask_{Boolean} * \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + Mask_{norm} \cdot b, \text{ where } b \in [0, 10) \tag{6}$$

The smoothed mask was normalized to create a standard mask range from zero to one:

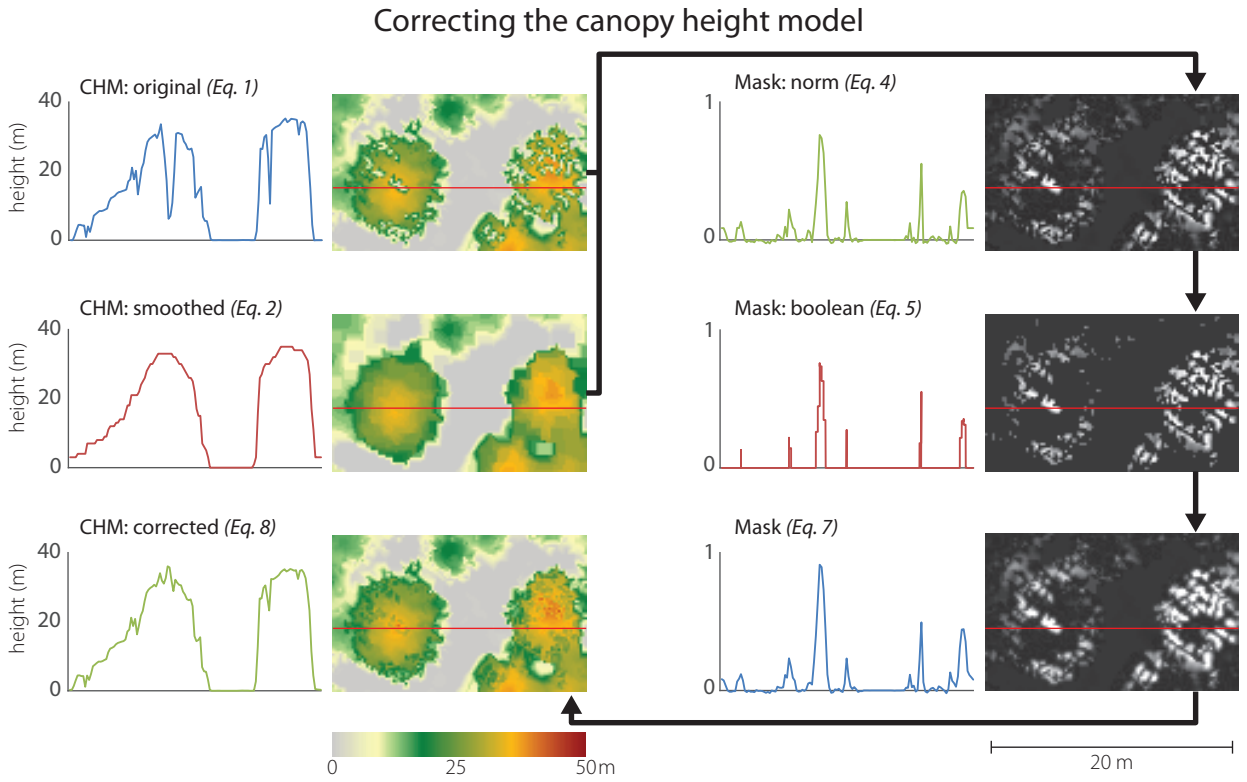

Figure 3-5. A number of steps were taken to correct the original CHM with artifacts. The horizontal profiles across canopy and ground (right) depict how the mathematical steps altered the raster surface.

$$Mask = \frac{Mask_{smooth}}{\mathrm{MAX}(Mask_{smooth})} \tag{7}$$

The final CHM was compiled by removing the artifact pixels (via the inverse of *Mask*) and filling in smoothed CHM values where *Mask* permits (Equation 8). Constant *c* was used to adjust the corrected values such that they match the original CHM. The process is further described and illustrated in Figure 3-5.

$$CHM_{corrected} = CHM \cdot (-1 \cdot Mask + 1) + CHM_{smooth} \cdot Mask \cdot c, \text{ where } c \in [1, 2] \tag{8}$$

CHM segmentation

I used the multi-resolution segmentation approach in eCognition 8.8 (Trimble, 2012)to divide the CHM into individual trees. The segmentation process relies on user-specified parameters regarding the scale, shape and spectral criterion of homogeneity, and a compactness ratio. I used the estimation of scale parameter (ESP) tool to determine the optimal parameters (Drăguţ *et al.*, 2010); however, the

shape and the compactness were determined manually. All three parameters were altered depending on the particular segmentation resolution or purpose (e.g. segmenting small isolated trees vs. large clumped tree stands).

The CHM was first segmented into large, coarse objects, which were then subdivided into individual tree crowns and ground. The subdivision into tree crowns was an iterative process of classification and segmentation. The classification in this case was a means to an end rather than an attempt to produce correctly classified objects. For example, objects that were likely short trees surrounded by ground pixels (low mean *z*-values) were identified and segmented further into very small objects to discriminate between actual ground objects and the trees. The small tree objects were then merged together to produce meaningful tree polygons. Next, medium-height tree objects were identified at the coarse-level and similarly, steps were taken to transform them into meaningful polygons. This processed continued until all possible objects of interest were segmented. Baatz *et al.* (2008) previously described this approach as "object-orient" OBIA, where meaningful objects are segmented, one object type at a time instead of trying to segment the entire image in one step. Various object metrics were used to isolate these trees prior to each sub-segmentation (Table 3-2: segmentation). A single complicated rule set (74 steps) was developed and applied to the entire study area.

During preliminary analysis I determined that optical imagery did not improve the delineation of the tree crowns. I also did not use the lidar intensity as it is inconsistent across sensor acquisitions and varies depending on the gain control settings of the lidar sensor (Im *et al.*, 2008; Leonard, 2005), the incidence angle, atmospheric dispersion and absorption, and bidirectional reflectance coefficients (Baltsavias, 1999a). The segmentation was based on 90 m squares extracted from the corrected CHMs at plot centers to avoid any edge effects due to the raster filtering. The segments were smoothed by eCognition during to remove the staircase-boundary effect.

Predicting the species

I used Waikato Environment for Knowledge Analysis (WEKA) software (Frank *et al.*, 2010) to classify the segments into tree species and species types (pines vs. firs). WEKA has proven to be an effective tool in classification, especially when many predictive attributes are available (Jakubowski *et al.*, 2013). The objects were classified into four species: *Abies concolor, Pinus lambertiana, Pinus ponderosa, Psuedotsuga menziesii*. Two species, *Calocedrus decurrens*, and *Quercus kelloggii*, were eliminated from the analysis because of low sample size (n = 4 each).

I extracted a number of object metrics to feed the classification routines, including object shape, lidar, and multispectral image metrics (Table 3-2: classification). Additionally, I ran a feature optimization algorithm based on Pearson's correlation to identify the most influential metrics and to avoid overfitting the classification models. All metrics were normalized prior to the classification. I identified two classifiers that best predicted the objects: LibLINEAR for individual species and the radial basis function (RBF) neural network for species type. LibLINEAR is a library for large-scale linear classification that supports logistic regression and linear support vector machines (SVM)

Table 3-2. Polygon attributes were calculated and were imported at three different stages of data processing: during the segmentation, classification of species, and feature optimization.

Object metric	Segmentation	Classification	Optimized features
Border shape		×	
Compactnes		×	
Density		×	
Length/width ratio	×	×	×
SqP	×	×	
Roundness		×	
Shape Index		×	
Self defined*	×	×	
Area	×	×	
Neighbord proximity (ground)	×		
Maximum lidar height	×	×	×
Mean lidar height	×	×	×
Minimum lidar height	×		
Mean difference to neighbor	×	×	×
8 metrics: mean(WV2 band i)		×	
8 metrics: standard deviation(WV2 band i)		×	
NDVI (maximum)		×	×
NDVI (mean)		×	×

*Self defined = (border index) × (shape index) × (compactness) × (roundness)

(Fan *et al.*, 2008). I used the SVM classifier for predicting the tree species. An SVM classifier uses data close to the support vectors (class boundaries) to fit a hyperplane to the training data within the feature space (Keerthi *et al.*, 2001). The RBF network uses *k*-means clustering to determine basis functions and then calculates logistic regression based on these functions. Multivariate Gaussian functions are fitted to all classes (Buhmann, 2000).

Comparison of segments

I compared the two segmentation approaches to each other and against the ground reference data. In validating the results to georeferenced trees, I evaluated the estimation of tree heights, species, and how tree detection varies across tree crown classes (e.g. dominant, intermediate, etc.).

In order to compare the two types of polygons, I first matched their tree height predictions to ensure that correct pairs of polygons were used for the comparison. Next, I compared the polygon shape characteristics by considering their area, shape index, and compactness. Shape index (Equation 9) is a widely applicable measure of shape complexity as compared to a square: a higher index indicates a more complex polygon (Forman and Godron, 1986; McGarigal *et al.*, 2012). Jiao and Liua (2012) found that, among other shape metrics, shape index is an effective descriptor of complexity. Similarly, compactness (Equation 10) is a measure of complexity as compared to a circle (Trani and Giles Jr, 1999).

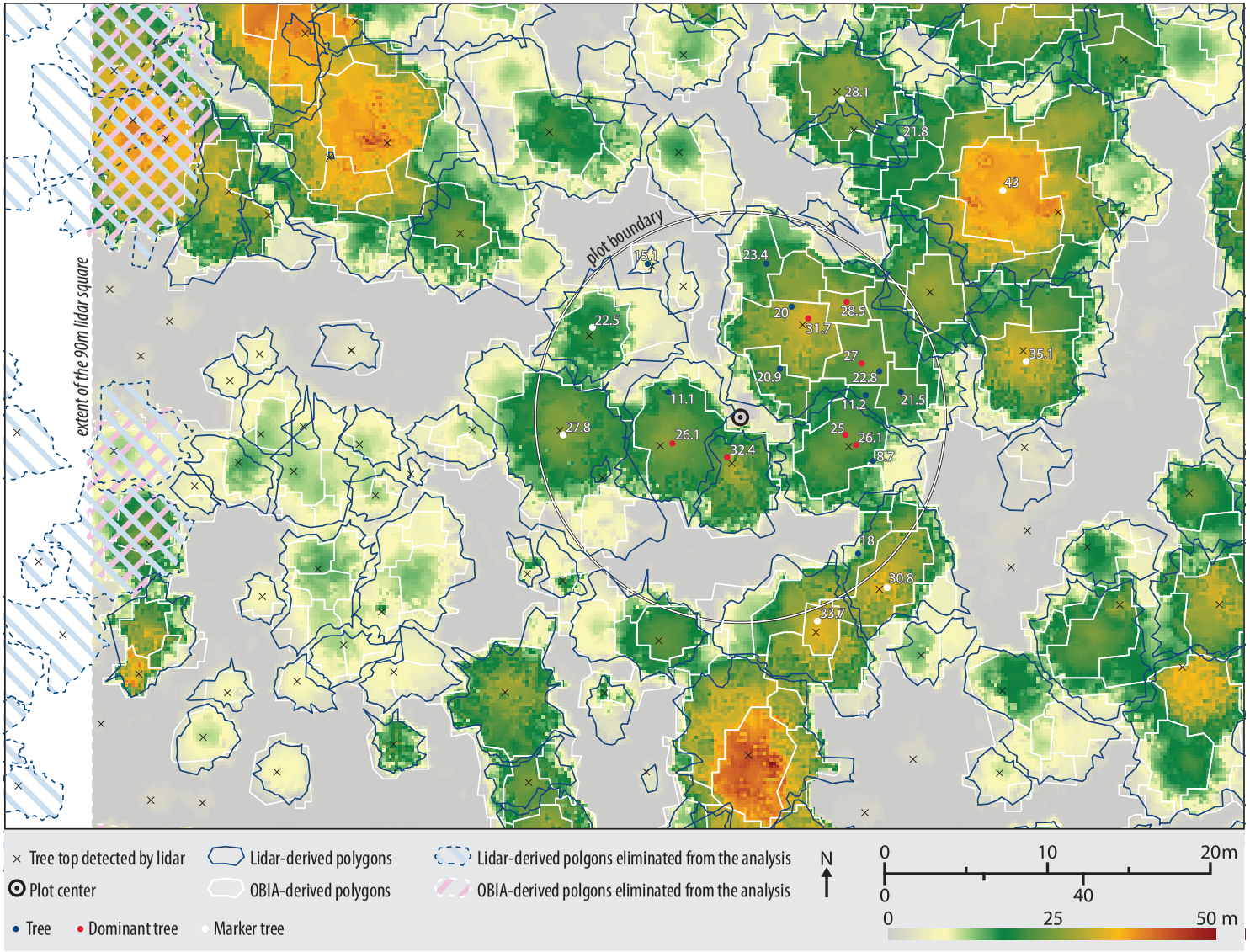

Figure 3-6. Each 90m focal area was centered at the plot center. The ground reference data included overstory trees (blue dots), overstory dominant and co-dominant trees (red dots), and marker trees (white). The blue (lidar-derived) and white (OBIA-derived) polygons show results of the data segmentation, while the crosses indicate the positions of tree tops as detected by the 3D segmentation algorithm. Tree heights (meters) are indicated next to ground-referenced trees.

$$SI = \frac{P}{4\sqrt{A}}, \text{ where } SI \in \left[\frac{\sqrt{\pi}}{2}, \infty\right) \quad (9)$$

$$C = \frac{4\pi A}{P^2}, \text{ where } C \in [0, 1] \quad (10)$$

where SI is shape index, C is compactness, P is perimeter, and A is area.

Since the lidar segmentation generated fewer polygons than OBIA, I used the former as a basis for comparison in order to avoid double-counting of the polygons. In particular, the tree top points as detected by the lidar segmentation algorithms were used to intersect and collect information from both types of polygons. All polygons that contact the edges were eliminated from the analysis resulting in n = 949.

Results

I used an OBIA approach to segment the CHM and a point cloud segmentation algorithm to segment the lidar data into individual trees. Segments at the edges of the considered areas were excluded from any of the analyses (Figure 3-6). Across all nine 90 x 90 m areas, a total of 2,875 objects were delineated using OBIA and 949 based on the lidar point cloud analysis. The OBIA method over-segmented many large trees, although in some cases, it correctly delineated smaller trees that the lidar approach had missed. This was especially true when the smaller trees were surrounded by large, taller species. Overall, the lidar-derived method produced larger polygons, especially for the tall and large trees. The agreements of tree height estimation between the ground reference data and the generated polygons were high for both methods. All dominant ground referenced trees were detected by both of the approaches; the tree detection rates dropped considerably for intermediate and suppressed trees, as well as for dead trees. Tree height accuracy declined in areas of high tree density.

Polygons vs. ground reference data

I intersected the ground reference tree points with lidar and OBIA polygons to validate the results. I cross referenced the tree heights measured on the ground against the maximum lidar z points within the lidar polygons, and the maximum pixel value of the corrected CHM within the OBIA polygons. In cases where there were more than one ground reference points within a polygon, I kept only the point that best matched the CHM height (n = 140). Three outlier points were excluded from the lidar polygons analysis. Pearson's squared correlation coefficients for the lidar polygons and for the OBIA polygons vs. ground measurements were $r^2 = 0.9641$ and $r^2 = 0.9309$, respectively (Figure 3-7).

I also analyzed the rate of tree detection by both methods in various crown classes (Figure 3-8). All

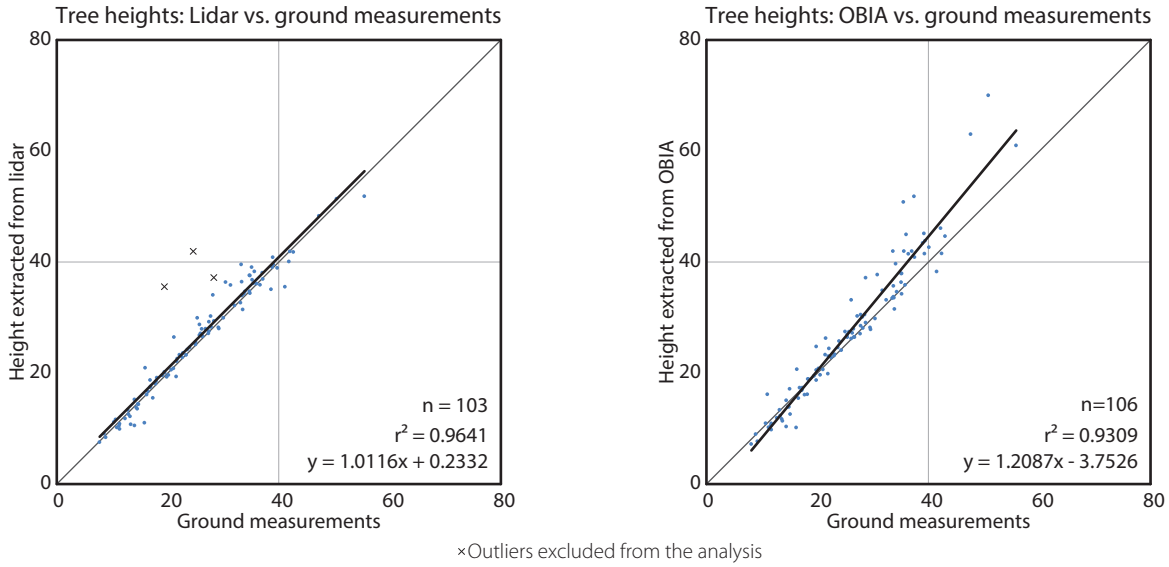

Figure 3-7. There was high agreement between the ground reference data vs. the lidar- (left) and the OBIA-derived (right) tree heights.

dominant trees were detected by both of the segmentation approaches. The rate of detection dropped when trees were more occluded by taller or bigger trees. OBIA detected a larger percentage of the non-dominant trees. Both methods detected less than 50 percent of dead trees.

I used the tree-top points, as detected by the lidar point cloud segmentation algorithm, to cross reference the lidar and OBIA polygons. I present the results as a function of (1) the method used to derive the polygons and (2) tree density.

The tree heights derived by both methods correlate very well ($r^2 = 0.9545$) indicating that the polygons detected the same trees. The OBIA approach estimated higher tree heights when predicting the tallest trees (Figure 3-9). The area, shape index, and compactness metrics differed significantly between the two methods. The area plot (Figure 3-9) clearly reveals that the lidar-derived polygons tend to be larger than the OBIA polygons. The range of lidar polygons is twice that of the OBIA polygons. In general, the compactness and shape index plots indicate that the lidar polygons are more complex. Both approaches produced polygons that were much more complex than simple geometric polygons (circle or a square).

Tree density effects

As expected, the tree density also affected tree height correlations (Figure 3-10). I summarize these effects with squared correlation coefficients as a function of tree density. Both methods performed best in areas with medium tree density when predicting the tree height. However, the OBIA method worked slightly better in low density areas, and considerably worse in high density areas.

The agreement between tree height predictions slightly declined in high tree density areas, but for the most part remained constant ($r^2 = 0.9596$ vs. $r^2 = 0.9401$ in low and high tree densities, respectively). The shape characteristics of the two polygon types were less correlated than the tree heights. The area correlation declined at high tree densities, where the lidar polygons were consistently larger than the OBIA polygons. OBIA polygons increased in complexity at higher tree densities and were therefore less similar to the lidar polygons.

Predicting tree species

I used all polygon features (Table 3-2) to predict individual species using the SVM algorithm from the LibLINEAR approach. The individual species predictions were poor: only 56 percent were correctly classified. However, the RBF neural network algorithm was able to correctly classify up to 75 percent of the species types. In the latter case, I used only six predictive metrics based on the suggestion from a feature optimization algorithm.

Discussion

The high Pearson's squared correlation coefficients between tree polygons and the ground reference data indicate that both lidar and OBIA methods detected individual trees well: $r^2 = 0.9641$ for the

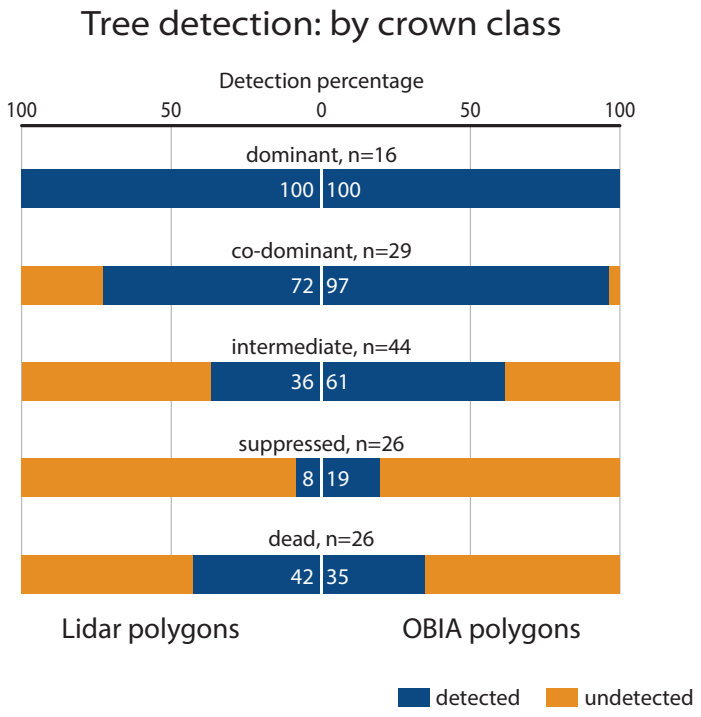

Figure 3-8. The perfect (100 percent) detection rates of the dominant trees dropped considerably as the less-dominant trees were considered. Dead trees were detected less than 50 percent of the time by both methods.

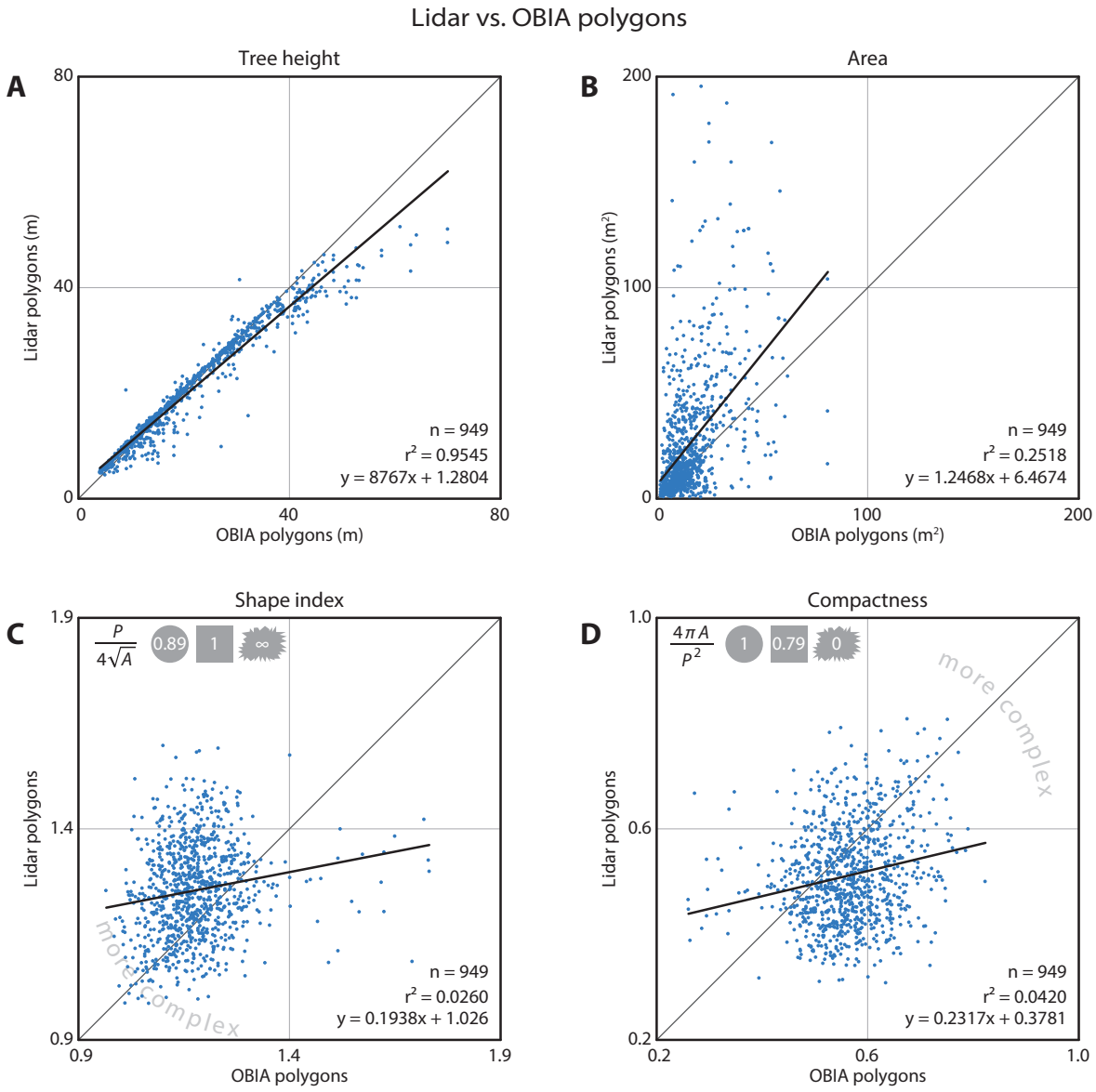

Figure 3-9. The tree-height agreement between polygons derived by the two methods was high (A), but their shape differed in terms of the area (B), and shape complexity (C and D).

lidar-derived and $r^2 = 0.9309$ for the OBIA-derived polygons. Further, the good fit with the OBIA approach also indicates that the CHM correction method worked well. While the lidar segmentation polygons were usually detected and delineated correctly, the locations of the tops of trees did not always match those of the ground reference data (Figure 3-6). This may be a result of trees that are leaning or not straight, as ground-reference positions were measured at the tree trunks, at breast-height level (1.3 m from the ground), as suggested by (Brokaw and Thompson, 2000).

All of the dominant, and most of the co-dominant trees were detected by both methods (Figure 3-8). Although the OBIA approach had a higher detection rate for non-dominant trees than the lidar

approach, this is likely due to the oversegmentation found with the OBIA approach. Consequently, these optimistic OBIA results should be explored further. This caution is supported by the tree height vs. ground data fits: the overall fit of the lidar-derived tree heights is tighter (Figure 3-7). The dominant tree detection rates of both methods were higher than those previously reported (Solberg *et al.*, 2006) possibly due to the higher point density in my data.

In comparing the ability of each method to delineate trees, one conclusion clearly stands out: while the detection rate and evaluated tree heights were comparable, the shapes of the derived objects were very different. Besides a few outlier points, the main discrepancy between lidar- and OBIA-derived tree heights was due to the tallest trees (>50 m), where the OBIA approach tended to overestimate height. This is likely due to the constants, *a, b* and *c*, used to correct the CHM surface. Previous studies (Falkowski *et al.*, 2008; Popescu and Wynne, 2004) have shown that tree heights derived from traditional CHM tend to be underestimated and thus, future research should evaluate whether the tall-tree heights derived from the corrected CHM are true.

It is clear from visual inspection that the lidar and OBIA polygons differ in shape (Figure 3-6). This was further verified by the comparison of their areas and shape complexities (Figure 3-9 and Figure 3-10). One explanation for the compactness of the OBIA polygons is that the segmentation algorithm does not allow polygon overlap. In contrast, the lidar-derived tree canopy polygons often overlapped because lidar points classified as one tree may be on top of, underneath, or within another tree in the *x-y* space (Figure 3-3). This is a more realistic representation of real forest canopy, especially in complex, dense forests such as those in my study area. Another reason for why the

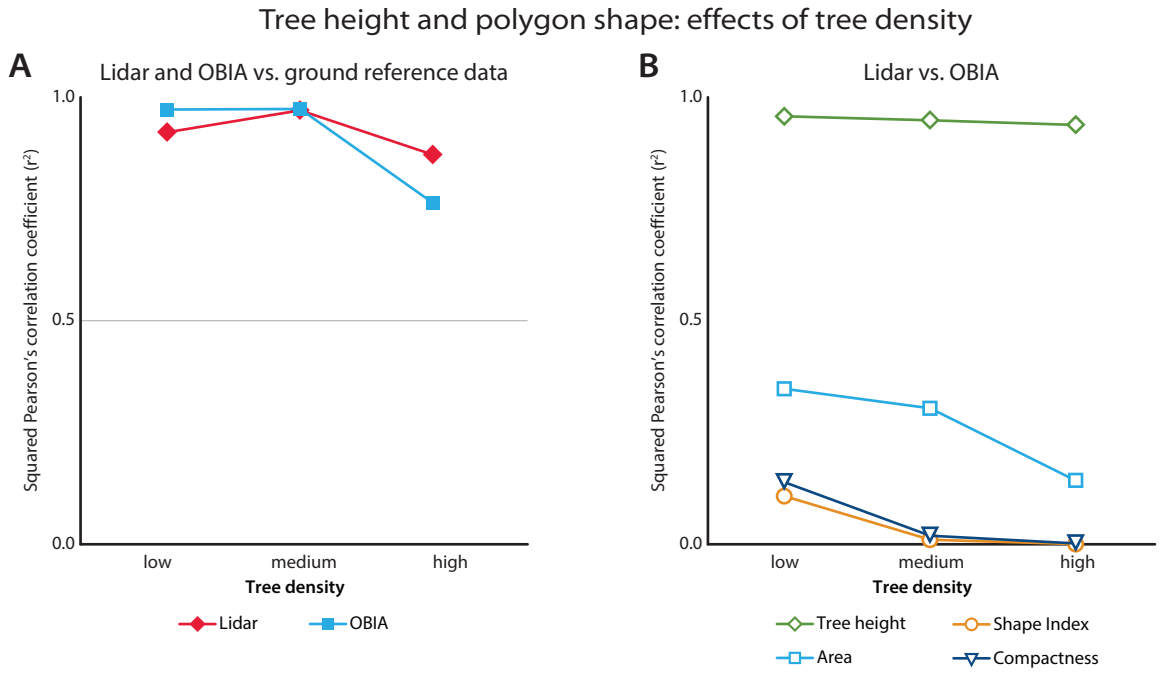

Figure 3-10. The agreements between tree heights measured on the ground vs. the lidar- and OBIA-derived methods were high (A). They dropped slightly in the high tree-density areas. The difference between the shape of the lidar- and OBIA-derived polygons grew in the high tree-density areas (B).

lidar polygons were more complex may be because of the way the concave hulls were constructed. Concave hulls that do not follow the lidar points as closely (by changing the α parameter when generating an alpha-shape) would lead to a simpler tree crown boundary and thus a simpler shape.

In general, the lidar polygons were larger than the OBIA polygons, as is evident in the area-comparison plot (Figure 3-9). Effectively, this means that the lidar segmentation method tended to under-segment and under-detect trees, while the OBIA method over-segmented the trees. In particular, the lidar segmentation algorithm often grouped trees into one segment when the trees were very close together (i.e. their crowns were merged or intertwined) or when the tops of the trees were not clearly detectable (Li *et al.*, 2012). Correct segmentation in such circumstances (especially when canopy cover >50 percent) is still problematic regardless of the used methods and it is an active area of research (Falkowski *et al.*, 2008). This is further evident when taking the tree density into consideration. In high tree density areas where trees grow in clusters, the disagreement between the area of the lidar- and OBIA-derived polygons increased and the overall accuracy of tree heights (as compared to ground truth) decreased (Figure 3-10). This is consistent with previous research where individual tree detection accuracy decreased in densely vegetated areas (Yu *et al.*, 2011).

The low accuracy of individual tree species classification is not surprising considering the study area. Even on the ground, the four classified species (Douglas fir, white fir, sugar pine, and ponderosa pine) can be easily misidentified by an untrained eye. On the other hand, the overall shape of firs vs. pines *is* more discernable, and more accurately classified: the firs are more conical and the pines are more oval, especially in mature species—and this is evident in the improved accuracy of tree type classification for firs vs. pines. Including hyperspectral data as a predictor would likely improve the results even further, and possibly help with the individual tree species classification. I should note that many other algorithms were tested during the preliminary analysis, including rule-based, decision tree and other machine learning classifiers; only the best results are presented here.

One of key challenges during the OBIA CHM segmentation process was the assignment of the correct scale, shape, and compactness parameters to obtain meaningful objects. Using previously established methods to determine the scale factor (Drăguţ *et al.*, 2010) helped in the choice of these parameters; however, as (Suarez *et al.*, 2005) discusses, designing an effective rule-set is still a largely manual process with potentially unlimited trial-and-error iterations. Efforts in future OBIA research will likely concentrate on designing processes to make the segmentation process less reliant on analyst input, and thus more deterministic and repeatable.

Conclusions

The successful detection and delineation of individual trees is critical in forest science, allowing for multi-scale analysis of the role of trees in forest functioning. There are many possible methods to delineate trees with remotely sensed data. In this chapter, I compare two segmentation algorithms for individual tree delineation: a lidar-derived method that uses 3-D lidar points, and an OBIA approach that make use of the lidar canopy height model. First, I segmented a lidar point cloud into trees

using a 3D segmentation algorithm. Next, I developed a new technique to produce a CHM without common lidar artifacts in order to segment this corrected CHM using an OBIA-derived method. The two methods were compared in *x-y* space in terms of their agreement to ground referenced tree heights, as well as tree detection across crown class and tree density. The overall tree height agreements of the OBIA- and the lidar-derived polygons were high ($r^2 = 0.9309$ and $r^2 = 0.9641$, respectively) and decreased in densely vegetated areas. The two types of objects were different in terms of polygon area and shape complexity. The OBIA objects were more likely to over-segment while the lidar objects were more likely to under-segment the trees, although the latter produced polygons more similar in shape to real tree crowns. Further research is necessary to automate the OBIA segmentation process and to improve 3D segmentation of the point cloud in dense forests.

Chapter Four

Tradeoffs between lidar pulse density and forest measurement accuracy

Introduction

The use of airborne lidar in forest studies has been on the rise in the last decade as increasing numbers of researchers use these data to predict directly measurable canopy characteristics (e.g. tree height, shrub height, etc.) as well as derived metrics (e.g. total basal area, biomass, carbon inventory, etc.) (Wulder *et al.*, 2008b). Since airborne lidar acquisition over large areas is costly, lidar missions necessarily involve tradeoffs. Researchers must decide between cost and coverage; between cost and lidar pulse density, between density and coverage; and at the same time maintain accuracy. Practical questions arise for scientists and managers using lidar to map forests: how large an area can I cover with a set budget? What is the minimum lidar pulse density that I can use and still get accurate results? In this paper, first consider the tradeoffs between cost, coverage and lidar pulse density, and then describe an experiment that directly examines the relationship between density and accuracy for various forest structure metrics at plot scale.

The cost of lidar

From a cost perspective, lidar is a viable option for remote sensing forest analysis. Lidar can already be less expensive or comparable to image data analysis when the entire cost of research (data acquisition, analysis, personnel, software, etc.) is considered, especially for large areas (Johansen *et al.*, 2010). Johansen *et al.* (2010) analyzed 26,000 km of a stream network using three datasets: lidar (3.98 pt/m^2), QuickBird imagery, and SPOT-5 imagery; the total research cost using these three approaches were approximately $3.8M, $6.4M, and $2.6M (Australian dollars), respectively. Tilley *et al.* (2004) also reported economies of scale; costs per ha declined from $37/ha to approximately $1/ha for large areal coverage (400 km^2).

Acquisition costs are related to coverage and lidar pulse density[1] (Baltsavias, 1999a; Lovell *et al.*, 2005), although the precise relationship between these variables is difficult to determine as most

1 Lidar pulse density is discussed in the literature in many ways, often with pulses, points, returns, and echoes used interchangeably. In this work, define a pulse ("pl") as the laser signal sent out from the lidar system towards the ground. A point ("pt"), also referred to as a return or an echo, is the signal, or multiple signals, reflected from target(s) back towards the lidar system (e.g. up to four points can be recorded from each pulse sent out by a typical lidar system).

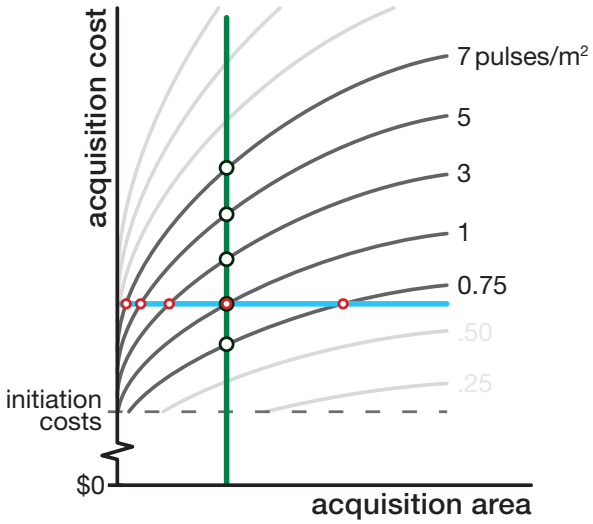

Figure 4-1. Conceptual diagram of tradeoffs between cost, data density, and area surveyed. The light blue, horizontal line represents decision space for managers with a fixed amount to spend: the area they can cover will increase with decreasing densities. The dark green line represents decision space for managers with a fixed target area: the price of collection will increase with pulse density. The cost of all pulse densities starts at a significant initial cost (mobilization of the aircraft, mission planning, etc.). The rate at which price increases for a larger study area decreases, i.e. cost per-square-area decreases as the study area increases. Using this logic, a constant budget (horizontal line) could afford a slightly larger area at 7 pl/m² and a much larger area at 1 pl/m² when compared to 9 pl/m².

studies do not report cost. Higher pulse densities can be obtained by independently increasing the pulse rate frequency (PRF), scanning rate, flight path overlap (or repeated flights), or by decreasing the flight altitude or the aircraft speed. Ultimately, the pulse density is related to the time that the aircraft/sensor spends flying, which translates to direct effect on acquisition costs. As a result of the tradeoffs between cost, coverage and density, land managers of vast areas (e.g. national norests) must often decide between ordering a low pulse density data that covers a large area or a higher pulse density data that concentrates on a subsection of their forest. Figure 4-1 conceptually illustrates the decision space for managers with a fixed amount to spend, or with a fixed target area. The cost per-square-area decreases as the study area increases. For example, managers with a constant budget (horizontal line) could afford a slightly larger area at 7 pl/m² and a much larger area at 1 pl/m² when compared to 9 pl/m².

Is high-density data necessary?

An important question to this work is whether high-density data are necessary to obtain accurate results in plot-scale lidar forestry. As technology improves, there has been a trend to acquire data at increasingly higher densities, reflecting the belief that this will improve accuracies. Researchers have speculated that higher pulse densities could improve the delineation of individual trees (Brandtberg *et al.*, 2003), the derivation of elevation and tree heights (Clark *et al.*, 2004), as well as predominant

height and LAI (Lovell *et al.*, 2003).

However, lidar data were not always gathered at high densities. A number of early lidar studies reported relatively high measures of correlation between lidar and field data using low-density data (<1pl/m^2). Hodgson and Bresnahan (2004) derived a digital elevation model (DEM) and validated the elevation error to be below 25.9 cm using 0.25 pt/m^2 data. Takahashi *et al.* (2010) derived stand volume and estimated a 10-39% root-mean squared error (RMSE) at $r^2 = 0.75$ using variable pulse density data between 0.125 and 0.25 pt/m^2. In one case, Thomas *et al.* (2006) has reported that using lower (0.035 pt/m^2) rather than higher (4 pt/m^2) lidar density resulted in higher correlation coefficients when predicting canopy metrics. Although the difference was small, they found that 8 out of the 11 tested metrics improved (root total biomass, stem bark, stem wood, foliage and all tree height metrics), root basal area remained the same, while only crown closure and live branches decreased in accuracy with the lower density data.

Further, in large projects, lidar data are often acquired at low pulse densities. The ongoing initiative to obtain nationwide lidar data for the continental United States is one example. In a few states the acquisition has already begun or has been completed. For example, Louisiana collected lidar data at 0.06-0.11 pt/m^2 (Cunningham *et al.*, 2004), Ohio at 0.04-1 pt/m^2, and Iowa at 0.11 pt/m^2 (Veneziano *et al.*, 2002). These efforts are targeted at the creation of digital elevation models (DEM), and thus the densities are much lower than typical forest research lidar data acquisitions (1-9 pl/m^2). The relevant question that address here is whether the density of this already-collected data could yield sufficiently high accuracies to reliably predict forest structure metrics *in addition* to DEMs.

Density and accuracy

The tradeoff between lidar pulse density and accuracy has been examined in several ways. Several researchers have investigated this experimentally by collecting data at multiple flying heights (i.e. above ground level or AGL); in general they found that AGL has little or no effect on the accuracy of predictions. Goodwin *et al.* (2006) analyzed the accuracy of the generated DEM, tree height, crown area, and tree volume, and found that with the exception of individual tree extraction, the predictions exhibited little to no effect based on the sensor altitudes—1,000, 2,000, and 3,000 m AGL. Takahashi *et al.* (2010) reported similar results at three altitudes—500, 1,000, and 1,500 m AGL—with very high and simulated low pulse densities—57, 25, and 9 pl/m^2 and 0.25-0.125 pl/m^2, respectively. In an earlier study, they found that increasing the AGL decreases penetration and intensity with little effect on accuracy (Takahashi *et al.*, 2008). Thomas *et al.* (2006) found higher AGL/lower density yielded slightly better results, while Næsset (2004) found no significant effect on stand height, volume, or basal area when comparing the same pulse density obtained at two AGLs.

Several studies considered the effect of lidar point (not pulse) density on the accuracy of predicted products. However, most these studies concentrate on the accuracy of a generated DEM. Although many use dissimilar measures for quantifying the accuracy, most of the studies agree that the error increases exponentially as the point spacing increases. This is true across topographic spectrum:

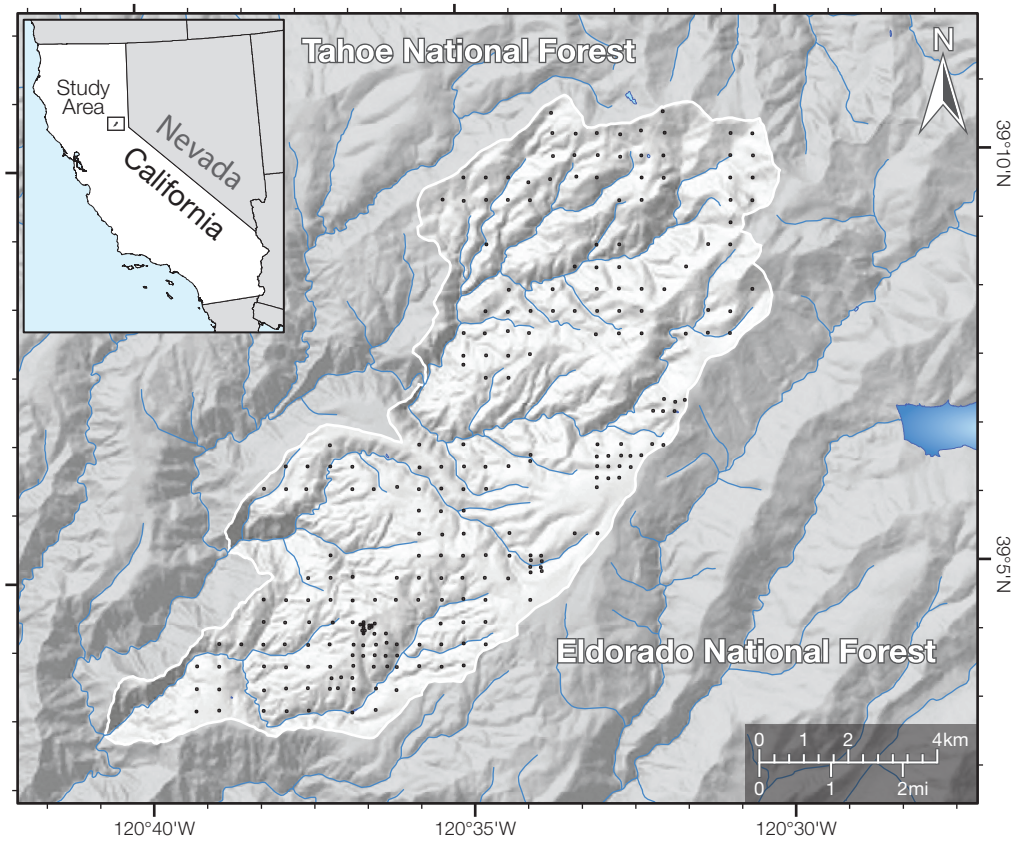

Figure 4-2. The study area (9,950 ha), mostly within Tahoe National Forest, is covered by mixed conifer forest, and covers topographically complex terrain.

in flat sites using low (Anderson *et al.*, 2006) and medium pulse density (Olsen *et al.*, 2009), in low-relief or undulating hills at various elevations (Goodwin *et al.*, 2006; Liu and Zhang, 2008; Sanii, 2008), and in steep, mountainous terrain using medium (Olsen *et al.*, 2009) to high pulse density data (Guo *et al.*, 2010; Pirotti and Tarolli, 2010).

In the few studies that examined canopy metrics the results are more variable. When considering tree height and/or volume, research has found little increase in accuracy as a function of point density, whether simulated (Tesfamichael *et al.*, 2010), or collected at different altitudes (Goodwin *et al.*, 2006; Takahashi *et al.*, 2010), unless the densities are simulated at very low levels (0.004 p/m^2) in which case the error grows exponentially (Magnusson *et al.*, 2007). Pulse density *does* affect individual tree detection, however, as reported by Takahashi *et al.* (2010) and Goodwin *et al.* (2006). Similarly, stream network extraction becomes problematic at high pulse spacing (Pirotti and Tarolli, 2010). However, while these studies are informative, they typically do not realistically simulate lower density data as though they were collected by a multiple-return lidar sensor at a higher AGL or lower PRF. Instead, the original data are typically thinned by even percentages (e.g. 75, 50, and 25% of the original point cloud) (Anderson *et al.*, 2006; Guo *et al.*, 2010; Liu and Zhang, 2008; Olsen *et al.*, 2009; Pirotti and Tarolli, 2010). Such results are not easily used by new researchers or land mangers that must determine the appropriate pulse density for their new lidar acquisitions.

In this study, simulate how predictions of various forest canopy and understory metrics vary with lidar pulse density. My objective was to determine the lowest pulse density that enables derivation of forest structure metrics at the plot scale at an acceptable accuracy level. report all findings in terms of the simulated pulse density—the typically-used specification when ordering lidar data. This practical and realistic approach yields results that are important not only to the scientific lidar community but also for land managers making decisions about lidar data acquisition parameters. My results will enable land managers to more reliably predict canopy metrics of interest with lidar data, while maximizing the coverage area given a set budget.

Materials and methods

This is a comprehensive study of the effect of lidar pulse density on extraction of various canopy and shrub metrics in the range from 0.01 to 9 pulses/m^2. extracted ten forest structure metrics that cover the vertical profile of a forest stand: maximum and mean tree height, total basal area, tree density, mean height to live crown base (HTLCB), canopy cover, maximum and mean diameter at breast height (DBH), and shrub cover and height. To ensure that the visible effect is a result of pulse density and not a particular algorithm, used three different approaches for canopy metric prediction: a simple multiple regression model, a support vector regression machine learning algorithm, and Gaussian processes.

Study area

The study area is located in northern California, USA, in the Tahoe National Forest (39° 07' N, 120° 36' W) (Figure 4-2). It spans over 9,950 ha on topographically steep and complex terrain between 600 m and 2,186 m above sea level. Since 1990, the average precipitation is 1,182 mm/year. The study area is 93% forested according to the Tahoe National Forest definition (Collins *et al.*, 2011b). Tree-ring analysis indicates high-frequency (5-15 years), low-intensity fires (Stephens and Collins, 2004), which is consistent and expected in this mixed-conifer, Mediterranean-climate forest. The dominant species include red and white fir (*Abies magnifica* and *Abies concolor*), ponderosa pine (*Pinus ponderosa*), sugar pine (*Pinus lambertiana*), incense-cedar (*Calocedrus decurrens*), and California black oak (*Quercus kelloggii*). The majority of the area is managed by the United States Department of Agriculture, Forest Service, with a limited number of private inholdings.

Field data

In this study analyzed data from 248, 0.05 ha circular plots (radius = 12.62 m). The plot centers were regularly spaced at 500 m intervals at even UTM coordinates (e.g. 710500 m E, 4329000 m N; 710500 m E, 4329500 m N; etc.). The few plots that were physically inaccessible, intersected rivers, road surfaces, or landings, were offset by 25 m in random directions. The ground truth data were collected during 2007 and 2008 summer field seasons according to a rigorous field protocol

previously described in Collins *et al.* (2011b) and MathWorks (2011). In summary, all trees' height, DBH, species, and HTLCB were measured. define HTLCB as the lowest extent of the live canopy; small trees below the canopy are included if there is no complete separation. All trees within the plot with DBH ≥19.5 cm were identified by a unique number. Smaller trees (5 ≤ DBH ≤ 19.5 cm) were measured on one third of each plot, designated by a random azimuth angle from north. A total of 7,284 trees were used in this analysis, including 4,581 with unique IDs and precise coordinates. The distributions of the trees' DBH and height are typical of a mixed-conifer Sierra Nevada forest. also characterized the understory by measuring shrub percent cover, height, and species. Five photographs were captured (toward the plot center from north, east, south, and west of the plot, and directly up from the plot center) to help characterize the forest structure of the plots.

All plots were georeferenced using a spatial accuracy protocol. First, find an optimal area to collected differential global positioning system (dGPS) signal—i.e. open canopy—ideally at or up to 30 m from the plot center. There, position a Trimble Zephyr antenna elevated 3 m from the ground and attached to Trimble GeoXH GPS receiver. record at least 300 (typically 1,000, and up to 7,000) measurements at positional dilution of precision (PDOP) <5. The dGPS measurements are post-processed using data from Continuously Operating Reference Stations (CORS) and University NAVSTAR Consortium (UNAVCO) stations. All CORS and UNAVCO stations used in post-processing were less than 20 km from all field measurements. use a combination of Impulse Laser Rangefinder, Impulse Electronic Compass, and a Trimble GeoXH receiver to georeference plot centers and tree locations. The range-finder is placed between the GPS antenna and the plot, such that the laser distance measurement error (≤2cm) and the compass degree error (≤0.5 degree) are minimized.

Lidar data

The lidar data were collected in five flights from September 19-21, 2008 (leaf-on conditions) at an average flying height of 950 m AGL. The acquisition area (10,700 ha) was larger than the study area to ensure complete, high-density data coverage. An Optech GEMINI Airborne Laser Terrain Mapper (ALTM) was used by the National Center for Airborne Laser Mapping (NCALM) to collect all data. Up to 4 returns were recorded from each pulse. The lidar data were collected at a minimum of 6 and an average of 9 pulses/m^2. Because of dense vegetation, the lidar system often recorded multiple returns with more than 20 points/m^2. All data were processed, recorded, and delivered in the Universal Transverse Mercator (UTM) coordinate system in the 1983 North American Datum (NAD83) and the North American Vertical Datum of 1988 (NGS GEOID03 model).

Lidar pre-processing

The raw lidar data were pre-processed before the outlined workflow (Figure 4-3). In particular, the data provider, NCALM, used TerraSolid's TerraScan software (Soininen, 2012) to remove isolated, "air," and below-surface points. For example, isolated points were defined as those with no neighbors within 5 m. The point cloud was separated into above-ground and ground points using an iterative

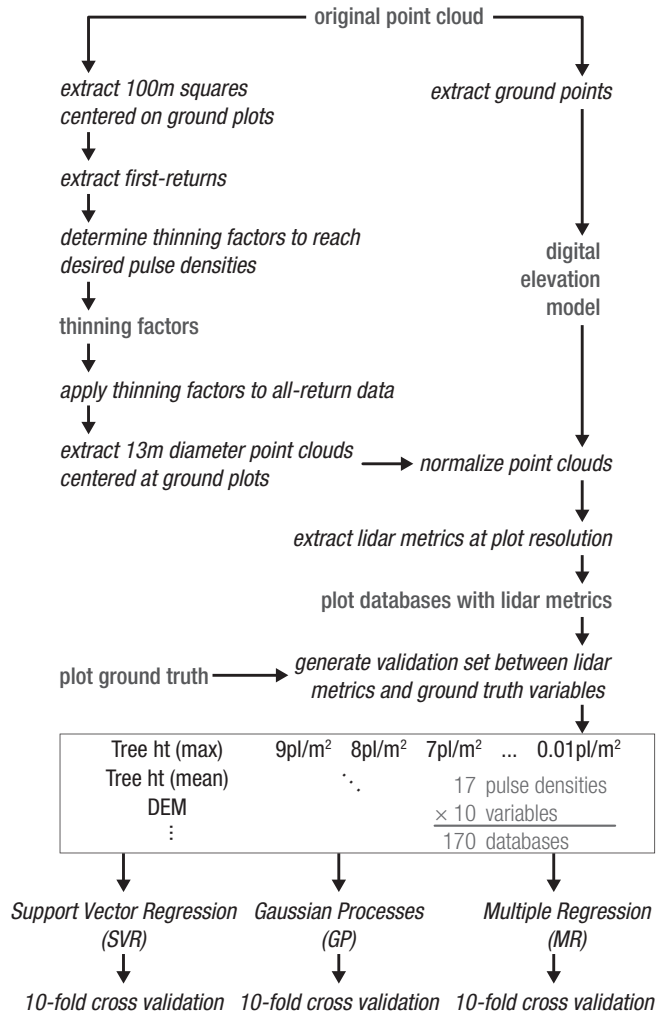

Figure 4-3. Workflow used to derive all forest metrics.

triangle-building surface model classification routine (Chang *et al.*, 2008). All digital elevation models (DEM) were generated using inverse distance weighted interpolation as suggested by Guo *et al.* (2010).

Simulating lidar pulse densities

The data were reduced to 17 pulse densities: 9, 8, 7, 6, 5, 4, 3, 2, 1, 0.75, 0.50, 0.25, 0.10, 0.075, 0.040, 0.025, and 0.01 pulses/m^2 (process illustrated in Figure 4-4). There are a number of ways to simulate lower density data; for instance, Garcia-Feced *et al.* (2011b) used a static, 1 m moving window to randomly choose points from the point cloud to reach the desired point density. Because wanted to simulate data obtained at specific pulse densities, my goal was to replicate the multiple-return effect as best as possible—in other words, sensor setting of x pulse/m^2 posting density will produce a point clouds with y points/m^2, where $y > x$. To simulate this process, the original lidar data were

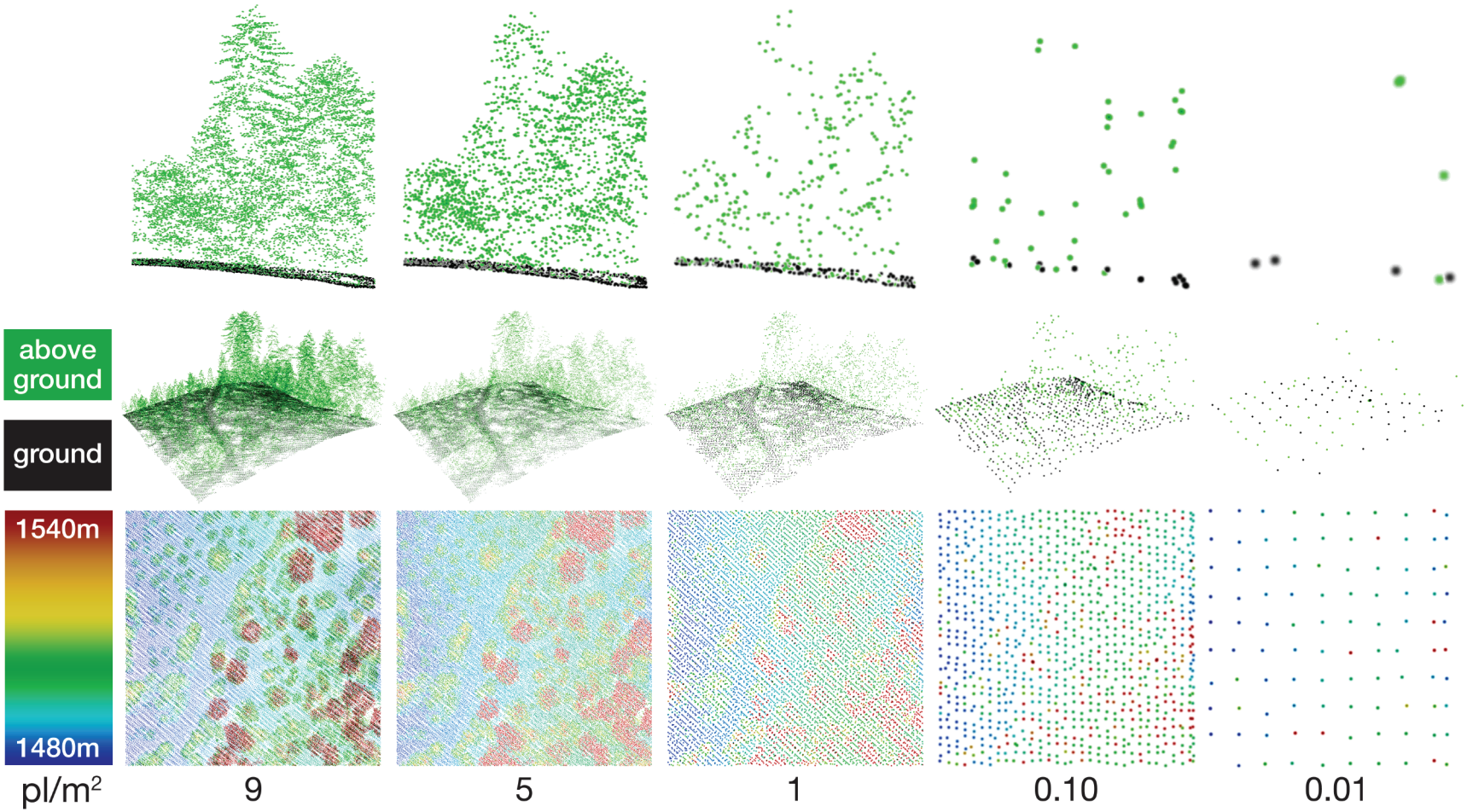

Figure 4-4. Simulation of pulse densities (as indicated by the number of pulses/m²). Only five out of the 17 simulated densities are shown here. The side and oblique, 3-D perspective (middle panel) shows the point cloud as classified into ground and above-ground points. In the bottom panel the same point clouds are visualized from above; the color corresponds to elevation of each point. The relative size of lidar points in lower density clouds is exaggerated for visual clarity.

divided into 100 m square tiles, each centered at a ground-reference plot. From these square tiles, only the first return points are saved and their average point density is recorded. The first-return points are assumed to correspond to the number of pulses sent out, as at least first-returns are recorded from each originating pulse. An optimization script was developed in combination with *lasthin* from *lastools* (Isenburg, 2011) to calculate a correct thinning factor for each 100 m square tile (n = 248), at each pulse density. The thinning is achieved by gridding the lidar data at different spatial resolutions and then choosing a random point within each cell. 4,216 files were processed this way to obtain 248 plot lidar point clouds at 17 different pulse densities. The reduced first-return point clouds were manually checked to ensure that proper pulse densities were obtained. Finally, the obtained thinning factors are used to reduce the original, all-return lidar data. This simulates lidar point clouds at multiple pulse densities with up to four returns per pulse.

Extraction of lidar metrics

From each 100 m square tile, extracted all points within 13 horizontal meters of the plot center to match the plot's footprint. used 13 m radius to account for any possible misalignment due to lidar or dGPS positional error. The 13 m horizontal radius point cloud was normalized by subtracting the DEM-derived elevation from each point's *z*-value yielding height above ground and removing any topographic effects. The point cloud is summarized by calculating descriptive statistics of the entire vertical profile (e.g. *mean height, percentiles, standard deviation*, etc.) and point densities throughout the vertical column (e.g. *number of points between 3 and 4 m above the ground*). The list of extracted metrics is provided in Table 4-1. Because the focus of this work was to compare accuracies obtained from various pulse densities, only the lidar points directly above the plots are considered in validation. The validation procedure described here is equivalent of deriving such products at a 26 m pixel size.

Forest structure prediction and validation

The lidar metrics described above were calculated for all plots, at all pulse densities, and then copied into a plot database. A series of files were generated—one for each forest metric and at all densities—to derive forest structure predictions and to cross-validate them. All lidar metrics were used in the prediction algorithms. used Waikato Environment for Knowledge Analysis (WEKA) software (Frank *et al.*, 2010), developed at University of Waikato, New Zealand to produce the results. WEKA is a software package designed to utilize various data mining algorithms. use three example algorithms—a simple multiple regression (MR), support vector regression (SVR) with sequential minimal optimization (SMO), and Gaussian processes (GP)—to compare their performance and to ensure that the effect see is a result of pulse density and not of using a particular validation method. The use of SVR and GP algorithms in forestry lidar has been successfully documented in the literature (MathWorks, 2011; Zhao *et al.*, 2011).

SVR algorithms use data points close to class boundaries (the support vectors) within feature space to fit a hyperplane to the training data (Keerthi *et al.*, 2001). SVR methods are particularly suitable

Table 4-1. Descriptive and statistical metrics extracted from the lidar data. The point density voxels were extracted based heights above the DEM.

Variables extracted from lidar	
Elevation	Point density: 0 to 0.5 m
Height: minimum	Point density: 0.5 to 1 m
Height: mean	Point density: 1 to 1.5 m
Height: maximum	Point density: 1.5 to 2 m
Height: standard deviation	Point density: 2 to 3 m
Total number of returns	Point density: 3 to 4 m
Quintile 0.01	Point density: 4 to 5 m
Quintile 0.05	Point density: 5 to 10 m
Quintile 0.10	Point density: 10 to 15 m
Quintile 0.25	Point density: 15 to 20 m
Quintile 0.50	Point density: 20 to 25 m
Quintile 0.75	Point density: 25 to 30 m
Quintile 0.90	Point density: 30 to 35 m
Quintile 0.95	Point density: 35 to 40 m
Quintile 0.99	Point density: 40 to 45 m
	Point density: 45 to 50 m
	Point density: 50 to 55 m
	Point density: 55 to 60 m

in high-dimensional datasets when the training sample size is small because only the data points closest to the hyperplane boundary are used (Hsu *et al.*, 2003). There are many variations of the SVR algorithms; in this paper, use the regression-based SMO implemented within WEKA, which uses a loss function to penalize errors that are greater than certain threshold on the training data (Frank *et al.*, 2010; Keerthi *et al.*, 2001; Smola and Schölkopf, 2004).

GP method is a robust, Bayesian method for non-parametric regression. A Gaussian process is a stochastic process for which any linear combination of samples has a joint Gaussian distribution. It is defined as a collection of random variables and can be described by its mean function—a vector—and a covariance function—a matrix (Rasmussen, 2004). GPs are appealing because they can model arbitrary functions (i.e. they are non-parametric), they yield simple linear algebra implementations, and the uncertainty of their predictions is quantifiable. used WEKA's standard implementation of GP in this work.

Each algorithm was validated using a 10-fold validation procedure as suggested by Kohavi (1995). In addition, calculated the RMSE for each predicted metric and at all pulse densities. My final report consists of the average correlation coefficients and the range of the RMSE values for each relationship between the lidar data and a forest structure and/or shrub metric, each as a function of pulse density. As an example, for tree height, report the average correlation coefficient and RMSE between field data across all plots and lidar data at 17 densities using 3 different prediction algorithms.

Results and discussion

extracted lidar metrics to analyze how pulse density affects prediction of various forest structure metrics. evaluated ten forest variables across the vertical profile, at 17 pulse densities, using three different predictive models (MR, SVR, and GP), to ensure that the visible trends result from pulse density and not from using a particular algorithm. The 10-fold correlation coefficients based on cross-validation of the predicted values and the ground data are plotted in Figure 4-5. All accuracies decreased as the pulse density approached 0.01 pl/m^2. The accuracies of some of the metrics remained relatively high until very low densities (in many cases up to about 1 pl/m^2). Similar trends were found by He and Li (2012) and Lovell *et al.* (2005), although the latter study simulated the entire point cloud and the forest environment to study these effects. He and Li (2012) analyzed lidar densities between 0.23 and 5.23 pt/m^2 to extract tree height and crown diameter in a Chinese, coniferous forest. The RMSE increased as the pulse density decreased in a similar pattern for all metrics. Because the RMSE trend was similar for all predictive models, report the range of the errors instead of the individual values (Figure 4-6). have identified a number of trends in the results, as described below.

In general, the relative overall accuracy of a forest metric was similar across all three predictive models. In all models, maximum tree height had the highest correlation coefficients followed by mean HTLCB and total basal area. The accuracy of the predicted forest structure metrics decreased roughly with its vertical position within the canopy; maximum tree height yielded highest accuracy, the accuracy declined for mean DBH, while shrub cover and height consistently produced the lowest accuracies.

The accuracy of the forest structure metrics increased as a function of pulse density at various rates. Further, each accuracy response line (Figure 4-5) reaches a point at which the accuracy either levels off or approaches maximum accuracy (the "turning point" in the trend). These turning points and best accuracy results are summarized in Figure 4-7. Collectively, these results indicate that beyond a certain density level, accuracy does not increase significantly.

The correlation coefficients of tree density and canopy cover increased significantly and consistently without a plateau from lowest to highest pulse densities (e.g. 0.271 to 0.753 at 0.01 and 9 pl/m^2 for tree density using GP, respectively). This indicates that maximum achievable accuracy of these two metrics may be beyond 9 pl/m^2. Others (maximum tree height and DBH metrics) significantly improved in accuracy within the lower pulse densities followed by consistent yet small increase. These are good examples of metrics that require a threshold density to achieve reasonable accuracy, but that do not benefit significantly from very high lidar density. This asymptotic behavior is very clear in maximum tree height and total basal area, where the turning points are reached at 1 and 2 pl/m^2, respectively. For example, the accuracy of total basal area using SVR drops significantly below 2 pl/m^2 (r = 0.789), but thereafter remains approximately equal to its maximum (0.796) and the 9 pl/m^2 accuracy (0.782). Similarly, the accuracy of maximum tree height using SVR rises steadily until 1 pl/m^2 (0.860), at which point it is similar to its maximum correlation coefficient (0.869). This level of accuracy is similar to previous studies that derived tree heights using lidar pulse density of 2.6 pl/m^2 (Erdody and Moskal, 2010) and >4 pl/m^2 (Zhao *et al.*, 2011), although my results were not as accurate, most likely due to the more complex study area.

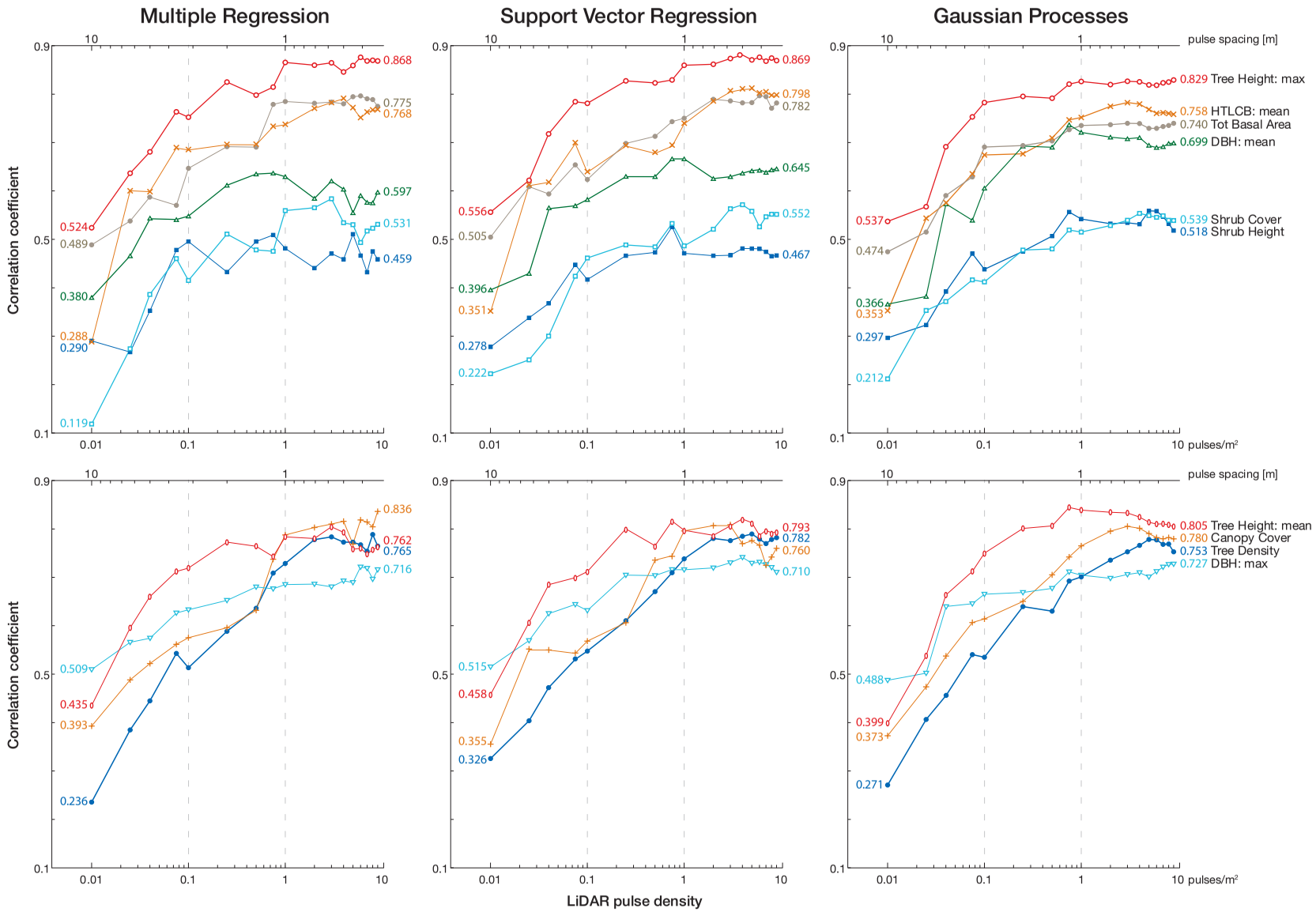

Figure 4-5. Ten-fold, cross-validation correlation coefficients of a number of predicted forest structure metrics using three algorithms: MR, SVR, and GP. The results are plotted as a function of pulse density on a logarithmic scale. The coefficient values are provided for lowest and highest pulse densities. Pulse density is provided in number of pulses per square meter (bottom axes) as well as the equivalent pulse spacing (top axes).

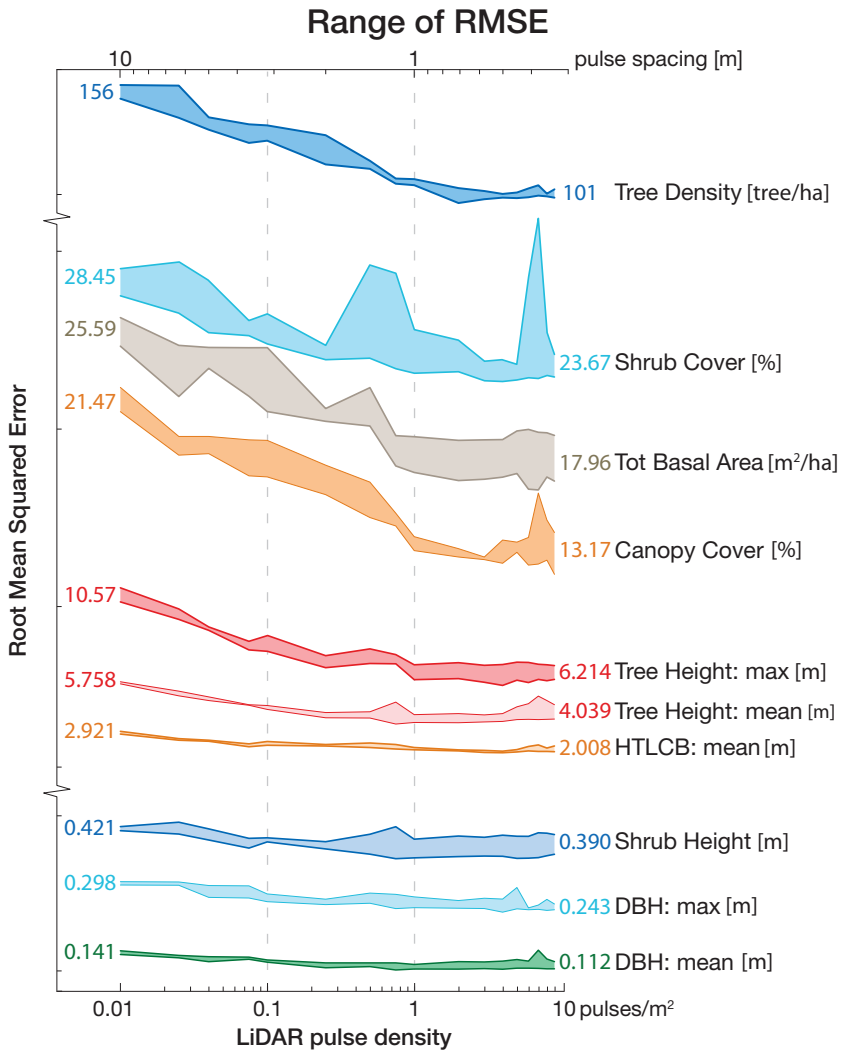

Figure 4-6. Ranges of the root mean squared errors (RMSE) for all three models (MR, SVR, and GP) as a function of lidar pulse density. Pulse spacing is indicated on top. The values directly to the left and right of the ranges are the average RMSE for the minimum and maximum pulse density. Units are indicated after the metric names on the right. The y-axis is linear while the x-axes are logarithmic.

The turning point depends on the forest metric and/or the predictive model. For tree density, canopy cover, and shrub cover, the turning point is reached at or above 1 pl/m^2 across all models. This indicates that "cover" metrics require relatively higher number of lidar returns to be mapped accurately. Conversely, mean tree height, mean DBH, and shrub height reach their turning points between 0.075 and 0.75 pl/m^2, indicating that less dense data can be used to estimate these metrics and still obtain reasonable results (should note here that although the turning point for shrub height occurs at a very low density, that metric is predicted at an overall low accuracy level). The predictive algorithms also influence the turning point. In general, SVR outputs reliable results at lower pulse densities than MR and GP. This implies that when working with low-density data, it may be best to use SVR in order to obtain reasonable results. MR almost always requires relatively high pulse density.

The calculated RMSE steadily increased for all metrics as the pulse density decreased. For example, the mean tree height RMSE as derived by the SVR was 3.92 and 5.67 m at 9 and 0.01 pl/m^2, respectively. The maximum tree height and maximum DBH RMSE were both higher than their corresponding mean metric RMSE. One metric, shrub cover, exhibited high outlier RMSE values at high pulse densities when estimating using the MR. The combination of relatively high RMSE with low and unstable correlation coefficient of this metric leads us to believe that MR is unsuitable for estimating shrub cover in high tree-density mixed-conifer forest.

Although the focus of this work is not algorithm comparison, there were a few noteworthy results with regard to method. In general, the three algorithms performed similarly for most of the variables. In a few cases, the SVR or the GP yielded better results: GP outperformed SVR and MR in estimating the mean DBH, shrub height, and mean tree height, while SVR outperformed the other algorithms in estimating the maximum tree height, mean HTLCB, and tree density (Table 4-2). In comparison to the other algorithms, MR performed particularly poorly in estimating the mean DBH, mean tree height, and maximum DBH. The mean DBH was most significantly affected by the algorithm used, especially at pulse densities above 0.1 pl/m^2: GP outperformed SVR and MR.

MR requires the most lidar points to obtain high correlation coefficients, followed by GP and SVR. The SVR and GP were most commonly the best performers. The SVR often performed better in lower densities (maximum and mean tree height, maximum DBH, and shrub height, total basal area, and mean HTLCB), while in some cases the GP performed decidedly better in the high pulse densities (mean tree height and DBH, and shrub height). This result indicates that the use of SVR may be preferable when working with low-density lidar data (below 0.3 pl/m^2).

Overall, my results indicate that very high lidar pulse density may not be necessary to predict typical forest structure metrics at the plot scale. From the best correlation coefficients obtained by the three algorithms, five out of ten best metrics were calculated from the lowest, and four were from the highest pulse density—the best-obtained correlation coefficient for each metric is indicated in Figure 4-7. This finding is particularly important for land managers that need to survey a large area with a specific forest metric and accuracy in mind. My conceptual model of the tradeoffs among

Table 4-2. Best-performing algorithms categorized into three groups of pulse densities: low, medium, and high. The noted algorithm performed best within the given pulse density category. ~ indicates mixed results.

Forest metric	<0.1 pl/m^2	0.1-1 pl/m^2	>1 pl/m^2
Tree height (max)	SVM	SVM	MR/SVM
Tree height (mean)	SVM	GP	GP
Total basal area	SVM	~	MR/SVM
Tree Density	~	SVM	SVM
HTLCB (mean)	SVM	MR/GP	SVM
Canopy cover	~	~	MR/GP
DBH (max)	SVM/GP	SVM	SVM/GP
DBH (mean)	~	GP	GP
Shrub cover	~	SVM	SVM
Shrub height	~	GP	GP

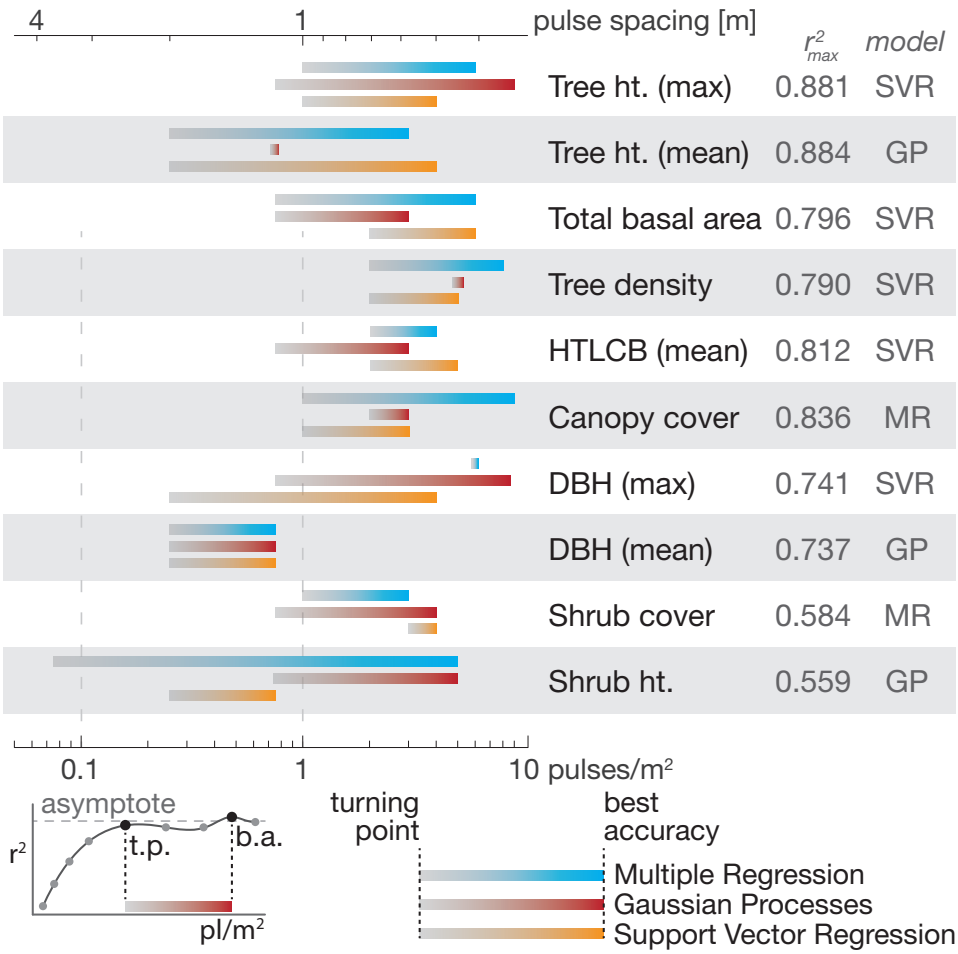

Figure 4-7. Synthesis of the observed results. The horizontal colored lines indicate pulse density ranges where the accuracies were deemed reasonable for the given metric. The left end of each accuracy range corresponds to its turning point— pulse density at which the metric accuracy plateaus or is comparable to its maximum. The right end of the range marks the pulse density at which the given model yields absolute maximum correlation coefficient. The three color lines correspond to individual algorithms, in order from top: MR (blue), GP (red), and SVR (orange). The two columns on the right of each metric reflect the maximum r^2 of all three algorithms and the model used to obtain it. For example, the turning point and the best accuracy for maximum tree height were located at 1 and 6 pl/m², respectively; SVR predicted the metric with the highest accuracy: $r^2 = 0.881$.

cost, coverage and density (Figure 4-1) can help guide this kind of process. For example, if tree height is the most important metric to estimate at a reasonable accuracy level, it may be sufficient to acquire lidar data at about 1 pl/m². On the other hand, if it is critical to derive the average HTLCB at a high accuracy (e.g. in a wildfire-prone forest where ladder fuels play an important role in forest dynamics), then it may be advisable to acquire a much higher pulse density—at least 2 to 4 pl/m². Previous research (Goodwin *et al.*, 2006; Næsset, 2004; Takahashi *et al.*, 2010; Thomas *et al.*, 2006; Zhao *et al.*, 2012b)

indicates that AGL—and therefore footprint size—affects the characteristics of the lidar point cloud and the absolute values of biophysical forest predictions. However, it has little effect on the prediction's accuracy. The studies listed here collected data at varying AGLs. It should be reiterated however that this work uses simulated lidar pulse densities, and not actual differences in data from multiple missions.

Unlike most of the previous lidar forestry studies, the analyses presented in this paper were performed in a topographically complex, steep terrain covered by dense, mixed conifer forest. For example, Tesfamichael *et al.* (2010) and Magnusson *et al.* (2007) assessed the effects of point density on a few forest metric but in a much simpler terrain and forest composition: South African rolling foothills and Sweden's low-relief coniferous forests, respectively. expect that the derived accuracies would improve in less challenging areas. Further, all analyses were completed at the plot-scale (equivalent to a 24 m pixel size), and would expect that the required pulse density for accurate identification and quantification of individual trees from lidar data would be larger than that required for plot-scale analysis. For plot scale analysis of forests, there might be cost savings and/or economies of scale in acquiring discrete return lidar data at lower densities.

Conclusions

The accuracy of predicted forest structure metrics decreases as the pulse density decreases, but it remains relatively high until low densities (e.g. 1 pl/m^2). This effect occurred with all tested predictive algorithms. The accuracy of tree height (mean and max), total basal area, DBH (mean and max), and shrub height remained *relatively* high until the pulse density decreased to 1 pl/m^2. Tree density, canopy cover, and shrub cover, and mean HTLCB typically required >1 pl/m^2. Support vector machine (SVR) and Gaussian processes (GP) algorithms consistently outperformed simple multiple regression model (MR). MR requires the highest pulse density to deliver good results; SVR and GP outperform other algorithms when the pulse density is low (<0.3 pl/m^2) and high (>1 pl/m^2), respectively. This work shows that low-density lidar data are capable of estimating the typical forest structure metrics at a plot level (~24 m pixel size) reliably. This finding indicates that land managers faced with a constrained budget may be able to cover larger areas at a lower cost when trying to estimate one of metrics tested in this work. However, higher density data are necessary for "cover" metrics and individual tree analyses.

Chapter Five

Conclusions and directions for future research

Fire is a fundamental component of most Sierra Nevada forest ecosystems and was prevalent in California prior to European settlement in the 1850s. Beginning around the turn of the last century, new policies to suppress fire were implemented broadly across the American West, and the century of fire suppression policy has led to dramatic changes in California's forests. Fire exclusion has directly changed fire regime parameters (i.e. decrease in return interval, earlier burn season, and larger fires), which in turn have fundamentally changed the vertical and horizontal structure of forests. California forests are becoming more dense and homogeneous, and ladder fuels, which often lead to canopy wildfires, are increasing.

Recent US Forest Service policies seek to reduce the impact of fire by reducing the amount of fuel in a forest. The concept behind the new treatments is that if area treatments are placed strategically, the spread rate and intensity of the fire over the entire area burned will likely be reduced. The impact of such strategically placed land area treatments (SPLATs) on water quality and quantity, forest health and fire behavior, and wildlife is unclear. This dissertation is part of a large multi-disciplinary research project, the Sierra Nevada Adaptive Management Project (SNAMP), which seeks to evaluate some of these impacts. Specifically, my dissertation focuses on geospatial and in particular remote sensing approaches to understanding forest processes, and can be used to inform forest management decisions.

The first step necessary in understanding the effects of SPLATs on forest function is to accurately measure and map forests. Spatial data—primarily remotely sensed imagery and lidar—are key integrators in this multi-disciplinary project: providing inputs to models, scaling biophysical processes from plot to landscape, estimating uncertainty, and visualizing future scenarios. The research results rely on modeled outputs, but these models are parameterized and tested against a dense network of field plots. These field data are linked with remote sensing through a suite of analytics: regression, machine learning, classification and object-based approaches, but all rely and build on a long history of remote sensing technology used to map and understand forests. In this dissertation, I develop new methods, compare existing methods, and test the limits of light detection and ranging (lidar) and optical data in order to contribute to the fields of remote sensing and forest science, and to support the specific goal of understanding the impacts of forest fuel treatments across a range of forest processes. Below, I highlight key findings from my research and discuss areas of future research and refinement.

Lidar is a powerful tool for mapping forest structure

Remote sensing has been used to map forests for several decades, but it is only in the last 10 years, with the application of lidar, that vertical and horizontal characterization of forests has been possible. In this dissertation, I explore the applicability of lidar to mapping forest structure from several different angles. Chapter 1 provides context, and defines key outstanding questions about the use of lidar for forest mapping. In Chapter 2, I show that lidar can effectively predict a wide range of forest stand metrics, especially those that pertain to parameters near the top of the canopy. Predicting fuels and fuel metrics is also possible to some degree, but challenging using discrete lidar. In Chapter 3, I compare the effectiveness of two very different remote sensing approaches to individual tree detection and delineation, and find that both OBIA and 3D lidar point cloud segmentation produce high tree-height agreement, and a low tree-shape agreement. Finally, in Chapter 4, I show that surprisingly sparse (and therefore less expensive) lidar data can provide reliable estimates of canopy and shrub structure at the plot level.

Stand metrics close to the top of the forest canopy can be effectively predicted using lidar

In Chapter 2, I provide an extensive analysis of data input types, data input transformations, and algorithms used to derive fuel models and fuel metrics. I found a clear association between the accuracy of stand metric prediction and the predicted metric's vertical position within the stand: the closer a metric is to the forest floor, the more difficult it is to predict it, and the more complex analysis is required for optimal classification.

Detailed forest fuel models are challenging to predict with lidar data, but it is possible to predict fuel types

I also found that precise surface fuel models are still challenging to predict using lidar, although the broader fuel types can be classified with reasonable accuracy (up to 76% correct classification). Although the approach may have worked better in a flatter, less complex, and sparser forest, ultimately it is difficult to predict surface fuel models that do not always correspond with the field data used to assign them. The prediction of continuous stand metrics used for wildfire behavior was variable. Maximum stand height, total basal area, and canopy cover yielded good results (all $r^2 > 0.82$). Shrub height and cover yielded reasonable results ($r^2 = 0.59$ and $r^2 = 0.62$) and therefore could act as a good predictor of ladder fuels: critical component of a forest stand in the wildfire context.

Lidar-derived segmentation over-predicted and OBIA under-predicted tree size

I segmented lidar data into individual tree polygons using two approaches: (1) 3D lidar point cloud segmentation and (2) object-based image analysis (OBIA) derived from a corrected canopy height model (CHM). The novel CHM correction method made OBIA segmentation possible,

even at relatively very high spatial resolution—i.e. higher than the point spacing of the original point cloud.

Validating with ground reference data, I found that both lidar-derived and OBIA-derived methods produced high tree height agreement ($r^2 = 0.9641$ and $r^2 = 0.9309$, respectively). The success rate for detecting dominant trees was perfect (100%) for both methods, and dropped considerably for suppressed trees. The OBIA-derived method had more success with the non-dominant trees, although this is likely due to the OBIA over-segmentation. The two methods produced polygons of different shape character. OBIA-derived polygons typically over-segmented the trees (producing smaller, less-complex polygons), while the lidar-derived polygons under-segmented the trees (producing larger, more complex polygons). I found that the lidar polygons were more true to real tree shapes, despite their occasional under-detection of suppressed or intertwined tree crowns.

Sparse lidar data is capable of predicting forest metrics at the stand level

As expected, the accuracy of predicted forest structure metrics decreases at reduced lidar pulse densities. The more surprising finding is that most forest metrics were still predicted at a reasonable level at relatively low pulse densities of approximately 1 pulse/m^2. This finding is important for future forest management decisions, and country-wide or global lidar data acquisitions. The conclusions provide clear guidelines as to where, in the pulse density space, lidar beings to fail in generating accurate estimates of forest structure.

I also found that machine learning algorithms and Gaussian processes consistently outperformed multiple regression models when predicting forest stand metrics. This was found to be true across all stand metrics, but it is particularly critical at the lower pulse densities. In other words, while sparse lidar data is capable of predicting forest metrics at a reasonable level, careful analysis should be applied when using such data. This point is especially relevant to land managers who make landscape-scale management decisions based on analysis from low-density lidar data collected across large areas.

Directions for future research

In the second chapter, I used several different methods to predict surface fuel models that are used as inputs for wildfire behavior models such as FARSITE. Despite these efforts, I found that predicting such specific models using lidar data and remotely sensed imagery is challenging. The subjectivity of these fuel models, which are classified by experts and do not necessarily correspond to the field data they are based on, likely played a significant role. Based on past studies and personal communication with fire experts, it is clear that this issue of subjectivity will need to be addressed by the wildfire behavior modeling community if more automated methods that rely on remotely sensed data are to be embraced. This is especially important as the availability of remotely sensed data (lidar in particular) increases, and could be used to model fire behavior across large landscapes. Two approaches are possible to address these issues. The first is to eliminate or significantly reduce the significance of

expert-derived fuel models within FARSITE, and replace them with empirically-derived input that is solely based on the ground truth data. A second approach is to fundamentally change FARSITE-like models in the direction of physics process-based fire models similar to FIRETEC.

In my dissertation, I use lidar point cloud segmentation to delineate individual trees. While this method worked, algorithm development in this area is still in its infancy. The current algorithms need to be further developed and tested in other ecological areas, such as broadleaf forests. Future research should focus on how to improve the algorithm to better differentiate individual trees, especially in densely vegetated forests. At the same time, while there is always room for improvement in algorithm development, the current algorithm *does* perform well in mixed-conifer forests, for which it was designed. As such, efforts should be dedicated to transform the algorithm into a more deliverable product so that it can be disseminated to land managers and other researchers for more involved testing.

Classifying individual tree species in a Sierra Nevada mixed-conifer forest based on fusion of multispectral and lidar data has proven to be challenging. Hyperspectral imagery has been used in many past studies to successfully extract spectral signatures and differentiate individual tree species. Fusing the two types of technologies into a single sensor, or at least a single-platform, would likely significantly improve differentiation of individual tree species. In an ideal scenario, these two technologies would be merged at the sensor level, such that the optical path out of the sensor is the same, and the data output is either a hyperspectral waveform data, or a point cloud where each point has a spectral signature associated with it (i.e. a hyperspectral point cloud). Such data could then be segmented using not only the points' relative positions as was the case in this work, but it could then further be classified into individual material and/or tree species using a machine-learning classifier.

Chapter 4 concentrated on determining pulse density that would enable accurate extraction of forest parameters and, at the same time, optimize the pulse density required to obtain these results. An operational version of the sensor described above may not be developed for a number years, but waveform lidar is already being used in forest studies, and its use is on a steep rise. It may be beneficial for the forthcoming research to evaluate similar thresholds for a waveform lidar system. More importantly, one key element that should be tested is the detection of individual trees at various pulse densities. Many previous studies have speculated that a dense lidar point cloud is required to extract individual trees; however, to date, no rigorous studies have been conducted on this topic.

This dissertation addresses a number of key questions regarding the utility of remote sensing, and in particular lidar, in forest mapping and management. Specifically, I found that many important forest metrics can be predicted reliably with lidar, and that relatively course-density lidar was sufficient. I also investigated the use of lidar in forest behavior models, and found that although lidar can be used to feed some of these models, there is still a need for more research, both in the remote sensing and in the fire behavior communities. Collectively, these results will be useful for the fire and forest management community, and they are key to the larger SNAMP program.

Cited Literature

Agee, J.K., Bahro, B., Finney, M.A., Omi, P.N., Sapsis, D.B., Skinner, C.N., Wagtendonk, J.W.v., & Weatherspoon, C.P. (2000). The use of shaded fuelbreaks in landscape fire management. *Forest Ecology and Management, 127*, 55-66.

Agee, J.K., & Skinner, C.N. (2005). Basic principles of forest fuel reduction treatments. *Forest Ecology and Management, 211*(1-2), 83-96.

Andersen, H.E., McGaughey, R.J., & Reutebuch, S.E. (2005). Estimating forest canopy fuel parameters using LIDAR data. *Remote Sensing of Environment, 94*(4), 441-449.

Anderson, E.S., Thompson, J.A., Crouse, D.A., & Austin, R.E. (2006). Horizontal resolution and data density effects on remotely sensed LIDAR-based DEM. *Geoderma, 132*(3-4), 406-415.

Anderson, H.E. (1982). *Aids to determining fuel models for estimating fire behavior,* General Technical Report INT-122. Ogden, Utah: U.F. Service, Intermountain Forest and Range Experiment Station.

Antonarakis, A.S., Richards, K.S., & Brasington, J. (2008). Object-based land cover classification using airborne LiDAR. *Remote Sensing of Environment, 112*, 2988-2998.

Baatz, M., Hoffmann, C., & Willhauck, G. (2008). Progressing from object-based to object-oriented image analysis. In T. Blaschke, S. Lang & G.J. Hay (Eds.), *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications* (pp. 29-42): Springer.

Baltsavias, E.P. (1999a). Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry & Remote Sensing, 54*, 199-214.

Baltsavias, E.P. (1999b). Airborne laser scanning: existing systems and firms and other resources. *ISPRS Journal of Photogrammetry & Remote Sensing, 54*, 164-198.

Barilotti, A., Crosilla, F., & Sepic, F. (2009). *Curvature analysis of LiDAR data for single tree species classification in alpine latitude forests*. Paper presented at the Laser scanning 2009, IAPRS, Voi. XXXVIII, Partì/WS, Paris, France.

Bertolette, D.R., & Spotskey, D.B. (1999). *Fuel model and forest type mapping for FARSITE input.* Paper presented at the The joint fire science conference and workshop, Boise, Idaho.

Blaschke, T. (2010). Object based image analysis for remote sensing *ISPRS Journal of Photogrammetry and Remote Sensing, 65*, 2-16.

Brandtberg, T., & Walter, F. (1998). Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple-scale analysis. *Machine Vision and Applications, 11*(2), 64-73.

Brandtberg, T., Warner, T.A., Landenberger, R.E., & McGraw, J.B. (2003). Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment, 85*(3), 290-303.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., & Solberg, S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment, 114*(4), 911-924.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32.

Brennan, R., & Webster, T. (2006). Object-oriented land cover classification of lidar-derived surfaces. *Canadian Journal of Remote Sensing, 32*(2), 162-172.

Brokaw, N., & Thompson, J. (2000). The H for DBH. *Forest Ecology and Management, 129*(1/3), 89-91.

Brown, J.K., & Roussopoulos, P.J. (1974). Eliminating Biases in Planar Intersect Method for Estimating Volumes of Small Fuels. *Forest Science, 20*(4), 350-356.

Buhmann, M.D. (2000). Radial basis functions. *Acta numerica, 9*(1), 1-38.

Burnett, C., Aaviksoo, K., Lang, S., Langanke, T., & Blaschke, T. (2003). *An object-based methodology for mapping mires using high resolution imagery.* Paper presented at the Ecohydrological Processes in Northern Wetlands, Tallinn.

Carlton, D. (2005). *Fuel Management Analyst Plus v 3.0.* Estacada, OR.: Fire Program Solutions, LLC.

Chance, K. (2005). Ultraviolet and visible spectroscopy and spaceborne remote sensing of the Earth's atmosphere. *Comptes Rendus Physique, 6*(8), 836-847.

Chang, Y.-C., Habib, A.F., Lee, D.C., & Yom, J.-H. (2008). *Automatic classification of lidar data into ground and non-ground points.* Paper presented at the ISPRS Congress Beijing 2008, Beijing, China.

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating individual trees in a savanna woodland using small footprint LIDAR data. *Photogrammetric Engineering & Remote Sensing, 72*(8), 923-932.

Chen, Q., Gong, P., Baldocchi, D., & Tian, Y.Q. (2007). Estimating basal area and stem volume for individual trees from lidar data. *Photogrammetric Engineering & Remote Sensing, 73*(12), 1355-1365.

Chubey, M.S., Franklin, S.E., & Wulder, M.A. (2006). Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric Engineering & Remote Sensing, 72*(4), 383-394.

Clark, M.L., Clark, D.B., & Roberts, D.A. (2004). Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sensing of Environment, 91*(1), 68-89.

Cleve, C., Kelly, M., Kearns, F.R., & Moritz, M. (2008). Classification of the wildland–urban interface: A comparison of pixel-and object-based classifications using high-resolution aerial photography. *Computers, Environment and Urban Systems, 32*(4), 317-326.

Clinton, N., Holt, A., Scarborough, J., Yan, L., & Gong, P. (2010). Accuracy assessment measures for object-based image segmentation goodness. *Photogrammetric Engineering & Remote Sensing, 76*(3), 289-299.

Collins, B., Miller, J., Kelly, M., Van Wagtendonk, J.W., & Stephens, S.L. (2008). Interactions among wildland fires in a long-established Sierra Nevada natural fire area. *Ecosystems, 12*(1), 15.

Collins, B.M., Everett, R.G., & Stephens, S.L. (2011a). Impacts of fire exclusion and recent managed fire on forest structure in old growth Sierra Nevada mixed-conifer forests. *Ecosphere, 2*(4), 14.

Collins, B.M., Stephens, S.L., Moghaddas, J.J., & Battles, J. (2010). Challenges and Approaches in Planning Fuel Treatments across Fire-Excluded Forested Landscapes. *Journal of Forestry, January/February 2010*, 24-31.

Collins, B.M., Stephens, S.L., Roller, G.B., & Battles, J.J. (2011b). Simulating fire and forest dynamics for a landscape fuel treatment project in the Sierra Nevada. *Forest Science, 57*(2), 77-88.

Cruz, M.G., & Alexander, M.E. (2010). Assessing crown fire potential in coniferous forests of western North America: a critique of current approaches and recent simulation studies. *International Journal of Wildland Fire, 19*(4), 377-398.

Cunningham, R., Gisclair, D., & Craig, J. (2004). *The Louisiana statewide lidar project*. Baton Rouge, Louisiana: Louisiana State University.

Drăguţ, L., Tiede, D., & Levick, S.R. (2010). ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science, 24*(6), 859-871.

Dubayah, R., & Drake, J. (2000). Lidar remote sensing for forestry. *Journal of Forestry, 98*(6), 44-46.

Duda, R., Hart, P., & Stork, D. (2001). *Pattern classification* (2 ed. Vol. 2). New York: John Wiley & Sons, Inc.

Erdody, T.L., & Moskal, M.L. (2010). Fusion of LiDAR and imagery for estimating forest canopy fuels. *Remote Sensing of Environment, 114*(4), 725-737.

EXELIS. (2012). ENVI software. McLean, VA. Retrieved from www.ittvis.com/ENVI.

Falkowski, M.J., Smith, A.M.S., Gessler, P.E., Hudak, A.T., Vierling, L.A., & Evans, J.S. (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*(S2), 338-350.

Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., & Lin, C.J. (2008). LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research, 9*, 1871-1874.

Feldesman, M.R. (2002). Classification trees as an alternative to Linear Discriminant Analysis. *American Journal of Physical Anthropology, 119*, 257-275.

Finney, M.A. (1998). *FARSITE: Fire Area Simulator-Model Development and Evaluation,* USDA Forest Service, Research Paper RMRS-RP-4. Missoula, MT: Rocky Mountain Forest and Range Experiment Station.

Finney, M.A. (2001). Design of regular landscape fuel treatment patterns for modifying fire growth and behavior. *Forest Science, 47*, 2.

Finney, M.A. (2003). Calculation of fire spread rates across random landscapes. *International Journal of Wildland Fire, 12*(2), 167-174.

Finney, M.A. (2006, 28-30 March 2006). *An overview of FlamMap fire modeling capabilities.* Paper presented at the Fuels Management-How to Measure Success, Portland, OR.

Forman, R.T.T., & Godron, M. (1986). *Landscape ecology*. New York: JohnWiley and Sons.

Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I.H., & Trigg, L. (2010). Weka-a machine learning workbench for data mining. *Data Mining and Knowledge Discovery Handbook*, 1269-1277.

Friedman, J. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis, 38*(4), 367-378.

Garcia-Feced, C., Temple, D., & Kelly, M. (2011a). Characterizing California Spotted Owl nest sites and their associated forest stands using Lidar data. *Journal of Forestry, 108*(8), 436-443.

Garcia-Feced, C., Temple, D., & Kelly, M. (2011b). LiDAR as a Tool to Characterize Wildlife Habitat: California Spotted Owl Nesting Habitat as an Example. *Journal of Forestry, 108*, 436-443.

Gatziolis, D. (2011). Dynamic Range-based Intensity Normalization for Airborne, Discrete Return Lidar Data of Forest Canopies. *Photogrammetric Engineering & Remote Sensing, 77*(3), 251.

Gleason, C.J., & Im, J. (2012). A fusion approach for tree crown delineation from LiDAR data. *Photogrammetric Engineering & Remote Sensing, 78*(7), 679-692.

Gong, P., Pu, R., & Yu, B. (1997). Conifer species recognition: an exploratory analysis of *in situ* hyperspectral data. *Remote Sensing of Environment, 62*, 189-200.

Goodwin, N.R., Coops, N.C., & Culvenor, D.S. (2006). Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment, 103*(2), 140-152.

Gougeon, F.A., & Leckie, D.G. (2006). The individual tree crown approach applied to Ikonos images of a coniferous plantation area. *Photogrammetric Engineering & Remote Sensing, 72*(11), 1287-1297.

Guo, Q., Li, W., Yu, H., & Alvarez, O. (2010). Effects of topographic variability and lidar sampling density on several DEM interpolation methods. *Photogrammetric Engineering & Remote Sensing, 76*(6), 701-712.

Guo, Q.C., Kelly, M., Gong, P., & Liu, D. (2007). An object-based classification approach in mapping tree mortality using high spatial resolution imagery. *GIScience and Remote Sensing, 44*(1), 24-47.

Haralick, R.M., Sternberg, S.R., & Zhuang, X. (1987). Image analysis using mathematical morphology. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(4), 532-550.

He, Q., & Li, N. (2012). Estimation of Individual Tree Parameters Using Small-Footprint LiDAR with Different Density in a Coniferous Forest. *Advanced Materials Research, 518-523*, 5320-5323.

Hodgson, M.E., & Bresnahan, P. (2004). Accuracy of airborne lidar-derived elevation: empirical assessment and error budget. *Photogrammetric Engineering & Remote Sensing, 70*(3), 331-339.

Hollaus, M., Wagner, W., Eberhöfer, C., & Karel, W. (2006). Accuracy of large-scale canopy heights derived from LiDAR data under operational constraints in a complex alpine environment. *ISPRS Journal of Photogrammetry and Remote Sensing, 60*(5), 323-338.

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

Holt, A.C., Seto, E.Y.W., Rivard, T., & Gong, P. (2009). Object-based detection and classification of vehicles from high-resolution aerial photography. *Photogrammetric Engineering & Remote Sensing, 75*(7), 871-880.

Hsu, C.W., Chang, C.C., & Lin, C.J. (2003). *A practical guide to support vector classification*. Taipei, Taiwan: N.T.U. Department of Computer Science, National Taiwan University.

Huang, H., Gong, P., Cheng, X., Clinton, N., & Li, Z. (2009). Improving Measurement of Forest Structural Parameters by Co-Registering of High Resolution Aerial Imagery and Low Density LiDAR Data. *Sensors, 9*(3), 1541-1558.

Huang, H., Li, Z., Gong, P., Cheng, X., Clinton, N., Cao, C., Ni, W., & Wang, L. (2011). Automated methods for measuring DBH and tree heights with a commercial scanning lidar. *Photgrammetric Engineering & Remote Sensing, 77*(3), 219-227.

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., & Falkowski, M.J. (2008). Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment, 112*(5), 2232-2245.

Hyde, P., Dubayah, R., Peterson, B., Blair, J.B., Hofton, M., Hunsaker, C., Knox, R., & Walker, W. (2005). Mapping forest structure for wildlife habitat analysis using waveform lidar: Validation of montane ecosystems. *Remote Sensing of Environment, 96*, 427-437.

Hyyppä, J., Hyyppä, H., Yu, X., Kaartinen, H., Kukko, A., & Holopainen, M. (2009). Forest inventory using small-footprint airborne lidar. In J. Shan & C.K. Toth (Eds.), *Topographic Laser Ranging and Scanning: Principles and Processing* (pp. 335-370). New York: Taylor & Francis Group.

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

Im, J., Jensen, J.R., & Hodgson, M.E. (2008). Object-based land cover classification using high-posting-density LiDAR data. *GIScience and Remote Sensing, 45*(2), 209-228.

Isenburg, M. (2011). LAStools - efficient tools for LiDAR processing (Version 111216). Retrieved from http://lastools.org.

ITT Visual Information Solutions. (2009). ENVI software. Boulder, CO. Retrieved from www.ittvis.com/ENVI.

Jakubowski, M.K., Guo, Q., Collins, B., Stephens, S., & Kelly, M. (2013). Predicting surface fuel models and fuel metrics using lidar and imagery in dense, mountainous forest. In press in *Photogrammetric Engineering & Remote Sensing.*

Jakubowski, M.K., Guo, Q., & Kelly, M. Tradeoffs between lidar pulse density and forest measurement accuracy. Accepted in *Remote Sensing and Environment*.

Jensen, J.R. (2007). *Remote Sensing of the Environment: An Earth Resource Perspective* (2 ed.). Upper Saddle River, NJ: Prentice Hall.

Jia, G.J., Burke, I.C., Goetz, A.F.H., Kaufmann, M.R., & Kindel, B.C. (2006). Assessing spatial patterns of forest fuel using AVIRIS data. *Remote Sensing of Environment, 102*(3-4), 318-327.

Jiao, L., & Liua, Y. (2012). *Analyzing the shape characteristics of land use classes in remote sensing imagery.* Paper presented at the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Melbourne, Australia.

Jing, L., Hu, B., Li, J., & Nolan, T. (2012). Automated delineation of individual tree crowns from lidar data by multi-scale analysis and segmentation. *Photgrammetric Engineering & Remote Sensing, 78*(12), 1261-1284.

Johansen, K., Phinn, S., & Witte, C. (2010). Mapping of riparian zone attributes using discrete return LiDAR, QuickBird and SPOT-5 imagery: Assessing accuracy and costs. *Remote Sensing of Environment, 114*(11), 2679-2691.

Ke, Y., Quackenbush, L.J., & Im, J. (2010). Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification. *Remote Sensing of Environment, 114*(6), 1141-1154.

Keane, R.E., Garner, J.L., Schmidt, K.M., Long, D.G., Menakis, J.P., & Finney, M.A. (1998). *Development of input data layers for the FARSITE fire growth model for the Selway-Bitterroot Wilderness Complex, USA,* General Technical Report RMRS-GTR-3. Ogden, UT: R.M.R.S. USDA Forest Service.

Keerthi, S.S., Shevade, S.K., Bhattacharyya, C., & Murthy, K.R.K. (2001). Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Computation, 13*(3), 637-649.

Kelly, M., Blanchard, S., Kersten, E., & Koy, K. (2011). Object-based analysis of imagery in support of public health: new avenues of research. *Remote Sensing, 3*(11), 2321-2345.

Kilgore, B.M., & Sando, R.W. (1975). Crown-fire potential in a sequoia forest after prescribed burning. *Forest Science, 21*(1), 83-87.

Koch, B., Heyder, U., & Weinacker, H. (2006). Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering & Remote Sensing, 72*(4), 357-363.

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection.* Paper presented at the International Joint Conference on Artificial Intelligence, Stanford University.

Koukoulas, S., & Blackburn, G.A. (2005). Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data. *International Journal of Remote Sensing, 26*(3), 431-455.

Kramer, H.J. (2001). *Observation of the Earth and its Environment: Survey of Missions and Sensors* (4 ed.): Springer.

Kwak, D.A., Lee, W.K., Lee, J.H., Biging, G.S., & Gong, P. (2007). Detection of individual trees and estimation of tree height using LiDAR data. *Journal of Forest Research, 12*(6), 425-434.

Laben, C.A., & Brower, B.V. (2000). USA Patent No.

Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., & Shreenan, R. (2003). Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Canadian Journal of Remote Sensing, 29*(5), 633-649.

Lee, H., Slatton, K.C., Roth, B.E., & Cropper, W.P. (2010). Adaptive clustering of airborne LiDAR data to segment individual tree crowns in managed pine forests. *International Journal Of Remote Sensing, 31*(1), 117-139.

Lee, J.-S. (1980). Digital Image Enhancement and Noise Filtering by Use of Local Statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-2*(2), 165-168.

Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., & Gower, S.T. (2002). Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography, 11*(5), 393-399.

Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., & Shugart, H.H. (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment, 67*(1), 83-98.

Leonard, J. (2005). *Technical Approach for LIDAR Acquisition and Processing.* Frederick, MD: EarthData Inc.

Li, W., Guo, Q., Jakubowski, M., & Kelly, M. (2012). A new method for segmenting individual trees from the lidar point cloud. *Photgrammetric Engineering & Remote Sensing, 78*(1), 75-84.

Lim, K., Treitz, P., Wulder, M.A., St-Onge, B., & Floodd, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography, 27*(1), 88-106.

Lin, C., Thomson, G., Lo, C.-S., & Yang, M.-S. (2011). A Multi-level Morphological Active Contmy Algorithm for Delineating Tree Crowns in Mountainous Forest. *Photogrammetric Engineering & Remote Sensing, 77*(3), 9.

Liu, X., & Zhang, Z. (2008). *LIDAR data reduction for efficient and high quality DEM generation*. Paper presented at the ISPRS Congress Beijing 2008, Beijing, China. http://eprints.usq.edu.au/4569/

Lopes, A., Touzi, R., & Nezry, E. (1990). Adaptive speckle filters and scene heterogeneity. *Geoscience and Remote Sensing, IEEE Transactions, 28*(6), 992-1000.

Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., & Coops, N.C. (2003). Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing, 29*(5), 607-622.

Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Coops, N.C., & Culvenor, D.S. (2005). Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. *Forest Ecology and Management, 214*(1-3), 398-412.

Magnusson, M., Fransson, J.E.S., & Holmgren, J. (2007). Effects on estimation accuracy of forest variables using different pulse density of laser data. *Forest Science, 53*(6), 619-626.

Mallet, C., & Bretar, F. (2009). Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*(1), 1-16.

Maltamo, M., Mustonen, K., Hyyppä, J., Pitkänen, J., & Yu, X. (2004). The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. *Canadian Journal of Forest Research, 34*(9), 1791-1801.

MathWorks. (2011). MATLAB (Version R2011b). Natick, Massachusetts.

McGarigal, K., Cushman, S.A., & E., E. (2012). FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. University of Massachusetts, Amherst.

McKelvey, K.S., Skinner, C.N., Chang, C., Erman, D.C., Husari, S., Parsons, D.J., van Wagtendonk, J.W., & Weatherspoon, C.P. (1996). *An overview of fire in the Sierra Nevada,* Sierra Nevada Ecosystem Project: Final Report to Congress: Assessments and scientific basis for management option 37. Davis, California: University of California, Centers for Water and Wildland Resources.

Melesse, A.M., Weng, Q., Thenkabail, P.S., & Senay, G.B. (2007). Remote sensing sensors and applications in environmental resources mapping and modelling. *Sensors, 7*(12), 3209-3241.

Menning, K.M., & Stephens, S.L. (2007). Fire Climbing in the Forest: A Semiqualitative, Semiquantitative Approach to Assessing Ladder Fuel Hazards. *Western Journal of Applied Forestry, 22*(2), 88-93.

Miller, D.R., Quine, C.P., & Hadley, W. (2000). An investigation of the potential of digital photogrammetry to provide measurements of forest characteristics and abiotic damage. *Forest Ecology and Management, 135*(1), 279-288.

Miller, J.D., Safford, H.D., Crimmins, M., & Thode, A.E. (2009). Quantitative evidence for increasing forest fire severity in the Sierra Nevada and southern Cascade Mountains, California and Nevada, USA. *Ecosystems, 12*(1), 16-32.

Mitchell, J.J., Glenn, N.F., Sankey, T.T., Derryberry, D.R., Anderson, M.O., & Hruska, R.C. (2011). Small-footprint Lidar Estimations of Sagebrush Canopy Characteristics. *Photogrammetric Engineering & Remote Sensing, 77*(5), 10.

Moghaddas, J.J., Collins, B.M., Menning, K., Moghaddas, E.E.Y., & Stephens, S.L. (2010). Fuel treatment effects on modeled landscape-level fire behavior in the northern Sierra Nevada. *Canadian Journal of Forest Research, 40*, 14.

Moghaddas, J.J., & Craggs, L. (2007). A fuel treatment reduces fire severity and increases suppression efficiency in a mixed conifer forest. *International Journal of Wildland Fire, 16*(6), 673-678.

Morsdorf, F., Meier, E., Kötz, B., Itten, K.I., Dobbertin, M., & Allgöwer, B. (2004). LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment, 92*(3), 353-362.

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

Næsset, E. (2004). Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment, 91*(2), 243-255.

Næsset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment, 112*(6), 3079-3090.

Nebiker, S., Annen, A., Scherrer, M., & Oesch, D. (2008). *A light-weight multispectral sensor for micro UAV–opportunities for very high resolution airborne remote sensing.* Paper presented at the The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

North, M.N.M., Innes, J.I.J., & Zald, H.Z.H. (2007). Comparison of thinning and prescribed fire restoration treatments to Sierran mixed-conifer historic conditions. *Canadian Journal of Forest Research, 37*(2), 331-342.

Okin, G.S., Roberts, D.A., Murray, B., & Okin, W.J. (2001). Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sensing of Environment, 77*(2), 212-225.

Olsen, R.C., Puetz, A.M., & Anderson, B. (2009). *Effects of lidar point density on bare earth extraction and DEM creation.* Paper presented at the American Society for Photogrammetry and Remote Sensing Baltimore 2009, Baltimore, Maryland.

Padwick, C., Deskevich, M., Pacifici, F., & Smallwood, S. (2010, April 26-30, 2010). *WorldView-2 pan-sharpening.* Paper presented at the American Society for Photogrammetry and Remote Sensing, San Diego, CA.

Persson, A., Holmgren, J., & Söderman, U. (2002). Detecting and Measuring Individual Trees Using an Airborne Laser Scanner. *Photogrammetric Engineering & Remote Sensing, 68*(9), 925-932.

Persson, Å., Holmgren, J., Söderman, U., & Olsson, H. (2004). Tree species classification of individual trees in Sweden by combining high resolution laser data with high resolution near infrared digital images. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI*(VIII/W2), 204-207.

Peterson, B., Dubayah, R., Hyde, P., Hofton, M., Blair, J.B., & Fites-Kaufman, J.A. (2005). *Use of LIDAR for forest inventory and forest management application.* Paper presented at the Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium.

Peterson, B.E. (2005). *Canopy fuels inventory and mapping using large-footprint lidar.* Ph.D., University of Maryland, College Park.

Pirotti, F., & Tarolli, P. (2010). Suitability of LiDAR point density and derived landform curvature maps for channel network extraction. *Hydrological Processes, 24*(9), 1187-1197.

Pitkänen, J. (2001). Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima methods. *Canadian Journal of Forest Research, 31*(5), 832-844.

Pitkänen, J., Maltamo, M., Hyyppä, J., & Yu, X. (2004). *Adaptive methods for individual tree detection on airborne laser based canopy height model*. Paper presented at the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI(8/W2), Freiburg, Germany.

Popescu, S.C., & Wynne, R.H. (2004). Seeing the trees in the forest: using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering & Remote Sensing, 70*(5), 589-604.

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

Popescu, S.C., Wynne, R.H., & Scrivani, J.A. (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.

Quattrochi, D.A., & Luvall, J.C. (1999). Thermal infrared remote sensing for analysis of landscape ecological processes: methods and applications. *Landscape Ecology, 14*(6), 577-598.

Rasmussen, C. (2004). Gaussian Processes in Machine Learning. In O. Bousquet, U. von Luxburg & G. Rätsch (Eds.), *Advanced Lectures on Machine Learning* (Vol. 3176, pp. 63-71). Berlin: Springer.

Reinhardt, E., Lutes, D., & Scott, J. (2006a). *FuelCalc: a method for estimating fuel characteristics.* Paper presented at the Fuels management—how to measure success, Fort Collins, CO.

Reinhardt, E., Scott, J., Gray, K., & Keane, R. (2006b). Estimating canopy fuel characteristics in five conifer stands in the western United States using tree and stand measurements.

Canadian Journal of Forest Research, 36(11), 2803-2814.

Riaño, D., Chuvieco, E., Salas, J., Palacios-Orueta, A., & Bastarrika, A. (2002). Generation of fuel type maps from Landsat TM images and ancillary data in Mediterranean ecosystems. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere, 32*(8), 1301-1315.

Riaño, D., Chuvieco, E., Ustin, S., Salas, J., Rodríguez-Pérez, J., Ribeiro, L., Viegas, D., Moreno, J., & Fernández, H. (2007). Estimation of shrub height for fuel-type mapping combining airborne LiDAR and simultaneous color infrared ortho imaging. *International Journal of Wildland Fire, 16*(3), 341-348.

Riaño, D., Meier, E., Allgöwer, B., Chuvieco, E., & Ustin, S.L. (2003). Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment, 86*(2), 177–186.

Riaño, D., Valladares, F., Condes, S., & Chuvieco, E. (2004). Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. *Agricultural and Forest Meteorology, 124*, 269-275.

Roberts, D., Gardner, M., Church, R., Ustin, S., Scheer, G., & Green, R. (1998). Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment, 65*(3), 267-279.

Rocadenbosch, F. (2003). Lidar-aerosol sensing *Encyclopedia of Optical Engineering* (Vol. 1, pp. 1090-1102).

Rothermel, R.C. (1972). *A mathematical model for predicting fire spread in wildland fuels,* USDA Forest Service, Research Paper INT-115. Ogden, UT: Intermountain Forest and Range Experiment Station.

Saatchi, S., Halligan, K., Despain, D.G., & Crabtree, R.L. (2007). Estimation of forest fuel load from radar remote sensing. *IEEE Transactions on Geoscience and Remote Sensing, 45*(6), 1726-1740.

Sanii, S. (2008). *Assessing the effect of point density and terrain complexity on the quality of LiDAR-derived DEMs in multiple resolutions.* Masters of Geographic Information Systems Masters, University of Calgary, Calgary, Alberta, Canada.

Schott, J.R. (2007). *Remote sensing: the image chain approach*. New York: Oxford University Press, USA.

Scott, J.H., & Burgan, R.E. (2005). *Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model,* General Technical Report RMRS-GTR-153. Fort Collins, Colorado: F.S. USDA, Rocky Mountain Research Station.

Scott, J.H., & Reinhardt, E.D. (2001). *Assessing crown fire potential by linking models of surface and crown fire behavior,* Research Paper RMRS-RP-29 RMRS-RP-29. Fort Collins, CO: USDA Forest Service, Rocky Mountain Research Station.

Sheng, Y., Gong, P., & Biging, G.S. (2001). Model-based conifer-crown surface reconstruction from high-resolution aerial images. *Photogrammetric Engineering & Remote Sensing, 67*(8), 957-965.

Skinner, C.N., & Chang, C.R. (1996). *Fire Regimes, Past and Present,* Sierra Nevada Ecosystem Project: Final Report to Congress: Assessments and scientific basis for management option 38. Davis, California: University of California, Centers for Water and Wildland Resources.

Skowronski, N., Clark, K., Nelson, R., Hom, J., & Patterson, M. (2007). Remotely sensed measurements of forest structure and fuel loads in the Pinelands of New Jersey. *Remote Sensing of Environment, 108*(2), 123-129.

Smola, A., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199-222.

Soininen, A. (2012). Terra Scan for MicroStation, User's Guide (Version 132). Jyvaskyla, Finland: Terrasolid Ltd.

Solberg, S., Naesset, E., Hanssen, K.H., & Christiansen, E. (2006). Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sensing of Environment, 102*(3-4), 364-376.

Stephens, S.L. (1998). Evaluation of the effects of silvicultural and fuels treatments on potential fire behavimy in Sierra Nevada mixed-conifer forests. *Forest Ecology and Management, 105*(1-3), 21-35.

Stephens, S.L. (2001). Fire history differences in adjacent Jeffrey pine and upper montane forests in the eastern Sierra Nevada. *International Journal of Wildland Fire, 10*(2), 161-167.

Stephens, S.L., & Collins, B.M. (2004). Fire regimes of mixed conifer forests in the north-central Sierra Nevada at multiple spatial scales. *Northwest Science, 78*(1), 12-23.

Stephens, S.L., Martin, R.E., & Clinton, N.E. (2007). Prehistoric fire area and emissions from California's forests, woodlands, shrublands and grasslands. *Forest Ecology and Management, 251*(3), 11.

Stephens, S.L., McIver, J.D., Boerner, R.E.J., Fettig, C.J., Fontaine, J.B., Hartsough, B.R., Kennedy, P.L., & Schwilk, D.W. (2012). The effects of forest fuel-reduction treatments in the United States. *BioScience, 62*(6), 549-560.

Stephens, S.L., & Ruth, L. (2005). Federal forest-fire policy in the United States. *Ecological Applications, 15*(2), 532-542.

Su, J., & Bork, E. (2006). Influence of vegetation, slope and LiDAR sampling angle on DEM accuracy. *Photogrammetric Engineering & Remote Sensing, 72*(11), 1265-1274.

Suarez, J.C., Ontiveros, C., Smith, S., & Snape, S. (2005). Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry. *Computers & Geosciences, 31*, 253–262.

Sugihara, N., Wagtendonk, J., Shaffer, K., Fites-Kaufman, J., & Thode, A. (2006). *Fire in California's ecosystems*. Berkeley: Univ of California Press.

Takahashi, T., Awaya, Y., Hirata, Y., Furuya, N., Sakai, T., & Sakai, A. (2008). Effects of flight altitude on LiDAR-derived tree heights in mountainous forests with poor laser penetration rates. *Photogrammetric Journal of Finland, 21*(1), 86-96.

Takahashi, T., Awaya, Y., Hirata, Y., Furuya, N., Sakai, T., & Sakai, A. (2010). Stand volume estimation by combining low laser-sampling density LiDAR data with QuickBird panchromatic imagery in closed-canopy Japanese cedar (*Cryptomeria japonica*) plantations. *International Journal of Remote Sensing, 31*(5), 1281-1301.

Taylor, A.H. (2001). Fire regimes and forest changes in mid and upper montane forests of the southern Cascades, Lassen Volcanic National Park, California, USA. *Journal of Biogeography, 27*(1), 87-104.

Tesfamichael, S.G., van Aardt, J.A.N., & Ahmed, F. (2010). Estimating plot-level tree height and volume of *Eucalyptus grandis* plantations using small-footprint, discrete return lidar data. *Progress in Physical Geography, 34*(4), 515-540.

Thomas, V., Treitz, P., McCaughey, J.H., & Morrison, I. (2006). Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: an examination of scanning density. *Canadian Journal of Forest Research, 36*(1), 34-47.

Tilley, B.K., Munn, I.A., Evans, D.L., Parker, R.C., & Roberts, S.D. (2004, March 14-16). *Cost considerations of using LiDAR for timber inventory.* Paper presented at the Southern Forest Economics Workshop, St. Augustine, Florida.

Trani, M.K., & Giles Jr, R.H. (1999). An analysis of deforestation: metrics used to describe pattern change. *Forest Ecology and Management, 114*(2), 459-470.

Trimble. (2012). *eCognition Developer 8.8*. Munich, Germany.

van Wagtendonk, J.W. (1996). *Use of a Deterministic Fire Growth Model to Test Fuel Treatments,* Sierra Nevada Ecosystem Project: Final Report to Congress: Assessments and scientific basis for management option 43. Davis: University of California, Centers for Water and Wildland Resources.

van Wagtendonk, J.W., Benedict, J.M., & Sydoriak, W.M. (1998). Fuel bed characteristics of Sierra Nevada conifers. *Western Journal of Applied Forestry, 13*(3), 73-84.

van Wagtendonk, J.W., & Moore, P.E. (2010). Fuel deposition rates of montane and subalpine conifers in the central Sierra Nevada, California, USA. *Forest Ecology and Management, 259*(10), 2122-2132.

van Wagtendonk, J.W., & Root, R.R. (2003). The use of multi-temporal Landsat Normalized Difference Vegetation Index (NDVI) data for mapping fuel models in Yosemite National Park, USA. *International Journal of Remote Sensing, 24*(8), 1639 - 1651.

Varner, J.M., & Keyes, C.R. (2009). Fuels treatments and fire models: errors and corrections. *Fire Management Today, 69*(3), 47–50.

Veneziano, D., Hallmark, S., & Souleyrette, R. (2002). *Comparison of lidar and conventional mapping methods for highway corridor studies,* Final Report. Ames, Iowa: Center for Transportation Research and Education, Iowa State University.

Vepakomma, U., St-Onge, B., & Kneeshaw, D. (2011). Response of a boreal forest to canopy opening: assessing vertical and lateral tree growth with multi-temporal lidar data. *Ecological applications: a publication of the Ecological Society of America, 21*(1), 99-121.

Wang, L. (2010). A multi-scale approach for delineating individual tree crowns with very high resolution imagery. *Photogrammetric Engineering & Remote Sensing, 76*(4), 371-378.

Wang, L., Gong, P., & Biging, G.S. (2004). Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. *Photogrammetric Engineering & Remote Sensing, 70*(3), 351-358.

Weatherspoon, C.P., & Skinner, C.N. (1996). *Landscape-level strategies for forest fuel management.* Paper presented at the Sierra Nevada ecosystem project: final report to congress.

Weng, Q. (2009). Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*(4), 335-344.

Whiteside, T.G., Boggs, G.S., & Maier, S.W. (2011). Extraction of tree crowns from high resolution imagery over eucalypt dominant tropical savannas. *Photogrammetric Engineering & Remote Sensing, 77*(8), 813-824.

Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., & White, J.C. (2008a). The role of LiDAR in sustainable forest management. *The Forestry Chronicle, 84*(6), 807-826.

Wulder, M.A., White, J.C., Coops, N.C., & Butson, C.R. (2008b). Multi-temporal analysis of high spatial resolution imagery for disturbance monitoring. *Remote Sensing of Environment, 112*(6), 2729-2740.

Yu, X., Hyyppä, H., Kaartinen, H., Hyyppä, J., Ahokas, E., & Kaasalainen, S. (2005). *Applicability of first pulse derived digital terrain models for boreal forest studies.* Paper presented at the ISPRS WG III/3, III/4 "Laser scanning 2005", Enschede, the Netherlands.

Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., & Viitala, R. (2011). Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(1), 28-37.

Zhang, C., & Qiu, F. (2012). Mapping individual tree species in an urban forest using airborne lidar data and hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing, 78*(10), 1079-1087.

Zhao, F., Guo, Q., & Kelly, M. (2012a). Allometric equation choice impacts lidar-based forest biomass estimates: A case study from the Sierra National Forest, CA. *Agricultural and Forest Meteorology, 165*, 64-72.

Zhao, F., Sweitzer, R.A., Guo, Q., & Kelly, M. (2012b). Characterizing habitats associated with fisher den structures in the Southern Sierra Nevada, California using discrete return lidar. *Forest Ecology & Management, 280*, 112-119.

Zhao, K., Popescu, S., Meng, X., Pang, Y., & Agca, M. (2011). Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sensing of Environment, 115*(8), 1978-1996.