AN ABSTRACT OF THE THESIS OF

Karin Wolken for the degree of Master of Science in Sustainable Forest Management presented on September 24, 2019.

Title: Analysis of Fire Severity in Southwestern Oregon Using Aerial LiDAR.

Abstract approved:

Temesgen Hailemariam

Acquiring, maintaining, disseminating, and utilizing quality data is key to adequate understanding and management of ecosystems. Modern remote sensing technology provides us an increasingly cost effective, unique opportunity for acquiring highly detailed information across every square meter of a landscape. The plethora of data available to scientists allows for use of high quality data on diverse topics, including the effects of fire through fire severity. However, the abundance of data may not be appropriate for certain modeling challenges and methods. We use aerial Light Detection and Ranging (LiDAR) acquired before local fires in conjunction with Remote Automated Weather Station data to derive fire severity. Observed fire severity was quantified as relative differenced Normalized Burn Ratio (RdNBR), a Landsat-derived parameter indicating basal area mortality. We utilized Pearson's correlation coefficient, random forest importance metrics, and pairs plots to identify and isolate parameters of particular importance in predicting RdNBR. The severity metric was then modeled for prediction through principal component analysis

(PCA) and random forest. Model performance was assessed using root mean square error (RMSE), root mean square prediction error (RMSPE), Bayesian information criterion (BIC), Akaike information criterion (AIC), bias, and adjusted R^2. LiDAR variables representing the six to nine meter height class were generally found to be key, along with weather variables representing duff moisture content. Weather variables were overall found to be most important in modeling RdNBR. PCA-based models were found to perform best for prediction.

Analysis of Fire Severity in Southwestern Oregon Using Aerial LiDAR

by

Karin Wolken

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented September 24, 2019
Commencement June 2020

Master of Science thesis of Karin Wolken presented on September 24, 2019

APPROVED:

Major Professor, representing Sustainable Forest Management

Head of the Department of Forest Engineering, Resources and Management

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Karin Wolken, Author

ACKNOWLEDGEMENTS

I would like to thank my family for their support, love, and understanding; the USDA Forest Service for their partnership and assistance; Karin Kralicek, Liam Gilson, Bryce Frank, Martin Uranga, Ty Nietupski, Al Pancoast, Joonghoon Shin, Jake Putney, and Sukhyun Joo for their conversation, patience, and grasp of statistics; and Kellie Cleaver, Madison Dudley, and Chelsey Durling for being fantastically kind and helpful to the graduate students of FERM.

.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1.1 General Introduction

Fire is a historical procedure, used and cherished in Pacific Northwest ecosystems. Early explorers found areas including the Puget Sound and central Washington to have open, park-like clearings resembling lawns in the middle of pine forests (Boyd, 1999). The clearings hosted berry producing plants, onions, camas, ferns, bunchgrass, and other food crops among widespread oaks, maintained through the frequent application of fire by local tribes (Boyd, 1999). Contemporary thought has driven fire to be a contentious issue, shifting between considering it as a destructive, unfortunate process to be avoided and an important natural process that can sometimes cause excessive damage. However, modern policies have largely lost this variety of well-managed ecosystem (Boyd, 1999).

Prescribed burning has been repeatedly found to be the most effective method in fire-adapted ecosystems to restore fire as a process in many ecosystem types. However, in the past two decades prescribed burning has steadied or decreased in the Western U.S. (Kolden, 2019). Most prescribed fire completed on the landscape is currently enacted by state, municipal agencies, and other groups. Throughout the United States, the Bureau of Indian Affairs uses prescribed fire the most among federal agencies including the Bureau of Land Management, National Park Service, Fish and Wildlife Service, and U.S. Forest Service (Kolden, 2019). Shifting slowly toward fire acceptance, the general population has a strong understanding of the role of fire in ecosystems. A study found that among visitors to three national parks in the Western U.S., one third to one half supported allowing fires from natural ignition to burn in areas that would not damage the ecosystems or human communities (Kneeshaw, Vaske, Bright, & Absher, 2004).

Often, a key limit to understanding and acceptance is data and visual data products. Aerial LiDAR has the potential to provide detailed, quality data across large areas of interest. Landscape-level views and management could help provide information for public education, federal management strategies, and acceptance of local events in the view of large-scale patterns. Aerial LiDAR is particularly adept to provide this information due to its ability to capture details about localized forest structure while still maintaining a large area of analysis (M. A. Lefsky, Cohen, Parker, & Harding, 2002).

1.2 Literature Review

1.2.1 Future of Fire Events and Concerns

Fires are becoming more frequent, larger, and influential on human life in recent history. Fire size, occurrence, intensity, and severity have all increased, commonly attributed to the period of fire exclusion that we still uphold, despite still not reaching the size and occurrence that existed on the landscape before European settlement (Reilly et al., 2017). Ecosystems with high severity fire regimes are believed to be minimally impacted by fire exclusion practices, but areas that typically maintain mixed and low severity fire regimes display higher severity fires than was historically present and larger patches of high severity fire. This has been observed particularly in areas where fire frequently occurred before the fire suppression paradigm was instituted (Reilly et al., 2017). Recent fires with uncharacteristic fire effects have become frequent topics of conversation, discussed by media and taxing communities, including the Boxcar Fire, Carr Fire, Mendocino Complex, Woolsey Fire, and the deadly Camp Fire. These overall trends are likely to continue as climate change further impacts forest and fire dynamics. Current climate change predictions would exacerbate trends we already see (Mote et al., 2019). Predictions for south central Oregon suggest a lengthened growing season by at least one month, exposing the area to hotter temperatures (Mote et al., 2019). This lengthened growing season, along with increased precipitation in the winter and decreased precipitation in the summer, will likely drive high winter growth rates. The dry heat of summer will vegetation for easier ignition and fire that will be challenging to manage (Mote et al., 2019).

An increased wildland urban interface (WUI) further challenges management of active fires. A major challenge WUI provide is an increased interaction between humans and wildland. This leads to a high potential for humans lighting fire, more private property that would need to be protected in the event of a fire, and a higher chance for fatality for both humans living further away from cities and the services they provide and firefighters that would have to enter more dangerous fire conditions (Radeloff et al., 2018). Fires burning in the WUI easily encroach on structures and impact health and recreation in these zones through smoke and access issues (Radeloff et al., 2018). Spread of invasive species through increased human interaction can also lead to damaged native ecology and more continuous fuel beds that can aid the spread of ground fires (Radeloff et al., 2018). The presence of a large and growing WUI also makes historic patterns of fire occurrence and size challenging to implement due to the need to avoid direct and significant indirect impacts on those living in the area.

1.2.2 Fire Management & Ecology

Fire management options essentially fall into two categories: reactive and proactive. Reactive actions are designed around direct suppression after a fire has already begun. Efforts focus on fighting fires as they occur with the focus of containment. The goal is often to minimize impacts on human life, human property, and ecosystems. However, the benefits fire can have for ecosystems are often overlooked due to the need for rapid decisions. Rapid decisions can lead to missed opportunities for fire to let fire continue to burn in safe areas, missing the opportunity to have beneficial fire (Schneider & Breedlove, n.d.). Situations like this can be aided by detailed spatial information of the landscape. However, time-intensive calculations and planning may be required before a fire ignites.

Proactive strategies allow for this effort-intensive planning. Areas believed to be susceptible to high severity, or areas that may not burn hot enough to gain full ecological benefit, can be treated to adjust how a fire would flow. Actively managing fuel in areas of concern, as well as developing a broader view of ecological areas that would be allowed to burn, leads to opportunities to harness the burning potential of a wildland fire to burn prescribed burn areas. This wildland fire use allows an opportunistic view of naturally ignited wildland fires. The focus falls to managing the path of fire to accomplish management objectives in a pre-defined area, enforcing larger boundaries while allowing natural areas to reap the full benefits of natural wildfire ("U.S. Fish & Wildlife Service: Fire Management—Wildland Fire Use," n.d.).

1.2.3 Remotely Sensed Fire Severity & Impact

Remotely sensed fire effects, particularly from aerial LiDAR, must rely on indicators sensed from above. LiDAR goes beyond other technologies by penetrating the canopy and attaining information from lower layers of the forest. While this often places a limit on the breadth of fire effects that can be assessed, it still allows for identification of standing trees that is vital to timber industry supply. Correlating with measurement of fire severity, the relative differenced Normalized Burn Ratio (RdNBR) approximates basal area mortality (Dunn & Bailey, 2016; Reilly et al., 2017; Whitman, Johnston, Schiks, Paugam, & Cantin, 2019). A generalized definition of fire severity falls simply as a measurement of change in an area, reflecting physical change caused by a fire (Keeley, 2009). Detectable from aerial sources, basal area mortality is a strong candidate due to the frequent success researchers have had in modeling basal area from LiDAR predictor variables (Fekety, Falkowski, Hudak, Jain, & Evans, 2018; Hudak et al., 2006; Silva et al., 2017). RdNBR itself strongly correlates with this measure of fire severity (Reilly et al., 2017). RdNBR is an extension of NBR, utilizing both pre- and post-fire Landsat images to calculate pre- and post-fire NBR to account for localized variation in pre-fire spectral responses. Differentiation between fire-caused traces and pre-existing patterns suggests a more accurate relationship between RdNBR and actual fire severity (Miller et al., 2009). RdNBR is a commonly used fire severity metric, easily interpretable by land managers and supported by a growing body of literature (Miller & Thode, 2007). As a continuous variable, RdNBR provides significant opportunity for analysis in a cell-driven framework, allowing for higher detail limited only by the maximum cell size of the inputs.

A key component of RdNBR is that the metric correlates with canopy-related metrics. Its correlation with basal area and other canopy-related variables provides significant modeling opportunities, but the forest floor still eludes this method. Knowledge of the forest floor, particularly the coarse wood debris and its division into 1-hour, 10-hour, and 100-hour fuels, is key in most fire behavior models (Scott & Burgan, 2005). However, this information requires a dense network of plots that are time intensive and rarely available. While canopy structure information does not directly match the information provided by coarse woody debris, it is easily accessible through aerial LiDAR and can provide information that may correlate with key forest floor data.

Despite existing as a continuous representation of severity, RdNBR is often placed into bins to aid interpretability. The bins are typically unburned or low severity, moderate or mixed severity, and high severity. These three categories approximately correspond to RdNBR values of less than 235, between 235 and 649, and greater than 649, respectively (Reilly et al., 2017).

1.2.4 LiDAR Quantification of Forest Structure

Airborne Light Detection and Ranging (LiDAR) is a form of remote sensing that provides detailed information as a collection of returns of laser pulses. Each pulse is distinct, but researchers have created tools that can provide LiDAR-derived high-resolution maps of canopy and ground layers (Lefsky et al., 2002). Canopy metrics are most easily attained and are correlated with other measurements of interest to attain LiDAR-derived variables. LiDAR is often used as a replacement for field plots. This is particularly important, because field plots require considerable labor, time, and cost (Lefsky et al., 2002). However, the potential accuracy of LiDAR increases if a few field plots are used to model attributes that cannot easily be calculated from LiDAR data. Some of these attributes, like basal area, can be related to height from simplified field plots. Using a few field plots to adjust the LiDAR data to the area greatly increases the accuracy, while not overinflating costs (Lefsky et al., 2002). Using a similar approach, a small number of field plots coupled with LiDAR data and an accurate fire behavior model could significantly expand fire risk information at little additional cost. LiDAR data that fully cover the area of interest could pinpoint small area site-specificity of fire risk, allowing identification of isolated high-risk areas (S. Hoe, Dunn, & Temesgen, 2018). While LiDAR is costly, spending less on field plots while receiving better information would improve inventory information and management.

Aerial LiDAR is particularly adept at describing forest structure (M. A. Lefsky et al., 2002). By quantifying point clouds through vertical percentiles, point densities, and variation, we can draw a data-driven picture of the forest (S. Hoe et al., 2018). LiDAR can provide measurements that relate to canopy height, tree density, the degree of variation in forest structure, and approximate canopy sizes. This forest information, while not providing measurements that may be acquired in unique inventory efforts, does cover the entire area of the target region and so avoids missing local variation that could be key in management decisions. This is particularly true when local conditions could greatly influence outcome, such as in active fire management.

Remote sensing strategies can quickly deliver accurate, large-scale data (Lefsky et al., 2002). Several remote sensing strategies are available to the general public. One remote sensing approach employs photogrammetric methods, which employs aerial images to approximate elevations. Methods using Landsat TM have difficulty distinguishing more than two forest structural classes (Means et al., 1999). However, Landsat and similar devices have shown an accuracy and sensitivity in dense forests that is far below what is needed for management of high biomass sites (Lefsky et al., 2002).

LiDAR is becoming increasingly available to non-industrial landowners (Lefsky et al., 2002). Public agencies and universities have completed several flights, and many organizations plan for extensive and repeated flights. This proliferation of LiDAR data allows research in the field to be effective over a large area and allows for continued improvements in LiDAR research. As research makes LiDAR analysis information more available and acquisition costs decrease, large landowners will be able to use this flexible large-scale data type for detailed forest structure information (Kelly & Di Tommaso, 2015).

CHAPTER 2: IDENTIFYING AERIAL LIDAR, TOPOGRAPHIC, AND WEATHER VARIABLES THAT IMPACT FIRE SEVERITY

2.1 Abstract

The availability of quality data has grown dramatically in recent years. Using this data to answer key questions about concerning topics still must be explored. In this chapter, we analyze data from aerial LiDAR and additional weather data to understand and interpret the impact of the variables on fire severity, represented by RdNBR. We examine the importance and relationships of vegetation, topography, and weather data on fire severity through visual analysis of pairs plots, numerical analysis of Pearson's correlation coefficient, and assessment of random forest's importance values. We found that overall relationships between explanatory variables and RdNBR are low. Variables describing the vertical layer of six to nine meters and those describing moisture in the duff layer were found to be key across our methods.

2.2 Introduction

Modeling is a common statistical tool used to identify variables that best describe a chosen response variable. With the plethora of information available in modern forestry, we have an opportunity to employ techniques that require vast amounts of data. By using existing datasets, we reduce sampling costs while still maintaining a high level of potential to discover true relationships through our statistical methods. Employing a predictive framework also allows us to model within fires, where an established fire severity exists, and project the potential fire severity on the remainder of the landscape. If the variables available to us are sufficient, allowing us to minimize error, the prediction on the remainder of the landscape would be an important asset in fire management planning. Practitioners could preemptively target high severity areas with vegetation reduction strategies such as controlled burns or mechanical and chemical removal. Likewise, areas more likely to burn at low severity could be identified to plan holding lines should a wild fire occur.

Using the pillars of extensive data and prediction, we identified two data sources to draw from to model potential fire severity. These two sources target the three sides of the fire behavior triangle: weather, topography, and fuel. Weather is represented by remote automated weather stations (RAWS). Topography is described through LiDAR-derived topographic variables. Fuel is represented by the vegetation structure captured by LiDAR, acquired before our sample fires burned.

However, modeling with too many variables causes the model to be too specific to the sample area. Since our goal is to predict beyond the fires to the remainder of the forests, it is important that the models are not overly keyed into the training information. Overfitting and multicollinearity are key concerns while modeling. It is important to minimize the number of variables, particularly predictor variables that are correlated among each other. Multicollinearity in multiple linear regression can cause significant predictor variables to be excluded, models to be improperly parameterized, and an overall decrease in statistical power (Graham, 2003). Such barriers may lead model parameters to be nonsensical, appear to be more or less important in a model than ecological principles would suggest, or simply make the model much less effective than it could be (Graham, 2003). Variable selection, correlation analyses, and methods that are not affected by multicollinearity are particularly necessary. Selecting a moderately low number of variables will maintain the information available to us while still allowing the model to perform well outside of the training data.

2.3 Methods

2.3.1 Study Area

The study area encompasses the Umpqua and Rogue River National Forests in the Cascade Mountains in southwestern Oregon (Figure 2.1). The forests are species diverse, with complex vegetative structure and dramatic topography. The tree species composition is primarily coniferous, but the forest has high diversity and includes woodlands characterized by an open oak forest structure (“Rogue River-Siskiyou National Forest—Home,” n.d.). The forest as a whole is diverse, including species such as *Pseudotsuga* *menziesii, Bromus orcuttianus, Whipplea modesta, Abies grandis, Holodiscus discolor,* and *Gaultheria shallon* (G. A. Harris, Franklin, & Dyrness, 1990). High elevation areas contain a mix of these conifers, while lower elevations also include the hardwoods ("Umpqua National Forest—About the Forest," n.d.). The area is dominated by moderate to extreme slopes, featuring hot, dry summers and cold, wet winters. Summer temperatures reach around 80-90 degrees Fahrenheit, and winters receive over ten feet of snowfall annually (G. A. Harris et al., 1990). Significant vegetation growth occurs around the wet, temperate winters. Through dry summers, the vegetation is particularly fire prone, driven by low fuel moisture and generally high fuel continuity caused by the productive springs and falls.

The forests house significant diversity in landscape features and uses. The North and South Umpqua rivers have headwaters in the mountains of the forest ("Umpqua National Forest—About the Forest," n.d.). Several species of fish and wildlife, including at risk species, rely on these rivers throughout their lifespans. Through the timber perspective of the Northwest Forest Plan, the forest's use is allocated as 34% to late successional reserve, 10% to riparian reserve, 7% to adaptive management area, and 34% to matrix management ("Umpqua National Forest—About the Forest," n.d.). Fifteen percent of the forest is withdrawn from timber uses for congressional and administrative purposes. The forests are also important sources of revenue. The high productivity has allowed the area to be maintained for active timber management. The Umpqua National Forest alone allocates 41% to adaptive and matrix management, resulting in 63.8 million board feet sold in 2006. Historically, four Native American tribes occupied the Umpqua Basin for more than 10,000 years ("Umpqua National Forest—About the Forest," n.d.). The rich history and value of the Umpqua National Forest provides an intriguing case to examine the impacts of large wildland fires.

The approximate rate at which fire returns to the Umpqua and Rogue National Forests is generally considered jointly, given the similar vegetation, topography, and location. This is typically called a fire return interval, referencing the span of time between two fires in one specified area (Dickmann & Cleland, 2019). The median of the intervals for the forests was calculated in one study to be between 26 and 85 years, depending on the particular plant association for the specific area (White, Briles, & Whitlock, 2015). More specifically, fire return intervals were calculated to be shorter for areas further from riparian areas, generally reaching between 4 to 167 years for riparian areas, while upslope fire return intervals were between 2-110 years (Olson & Agee, 2005). Forests dominated by coastal Douglas-fir had return intervals as low as eight years, depicting a vegetative community that is extremely prone to frequent fire (Olson & Agee, 2005).

2.3.2 LiDAR Dataset

LiDAR collections were acquired for the Umpqua and Rogue River National Forests from the Forest Service, U.S. Department of Agriculture Data Resources Management. Available LiDAR data included collects from 2014, 2015, and 2016, carefully matched with the 2015 National Creek Complex fire, 2017 Umpqua North Complex fires, and 2018 South Umpqua Complex fires. Only data that pre-dated the fire were used in analysis.

LiDAR covers the majority of the Umpqua and Rogue River National Forests. Portions of LiDAR were selected for analysis where LiDAR was available before a major fire. Aerial LiDAR for these forests was predominately collected with a Leica ALS-70 HP system, collecting at least four pulses per square meter. Data were processed through the Forest Service's Data Resources Management and Biometrics groups using FUSION (McGaughey, 2018).

2.3.3 Weather Dataset

Weather data employed in this analysis was acquired from Remote Automated Weather Stations (RAWS). The four RAWS stations used during the fire events were considered for the three fires: Stella 353209, Cinnamon 353031, Grandad 353036, and

Buckeye 353040. The Stella station was used for the National Creek Complex, Grandad for Umpqua North Complex, and Buckeye for South Umpqua Complex.

RAWS data was initially processed through FireFamily Plus 4.2 (Bradshaw & Jolly, 2014), then reduced to the specific burn days for each fire in R (R Core Team, 2019). Fire progression maps were developed and used to identify each day for which the fire progressed. Corresponding summarized daily weather data from RAWS were applied to each of the daily progression areas. Weather values for days without progression were averaged into the first subsequent day that showed fire progression. Fire progression maps were created for each of the three fires in ArcMap (*ArcGIS Desktop*, 2014) from GeoMAC (U.S. Department of the Interior: U.S. Department of Agriculture, n.d.) fire perimeters. Relevant hourly fire variables were extracted and summarized across active burn times each day. These summary statistics from 10:00 to 17:00 were applied to a fire progression maps built from GEOMAC data (Zald & Dunn, 2017).

Weather averages were primarily employed during analysis due to their use in previous research and to improve ease of model interpretation (Zald & Dunn, 2017). Other descriptive statistics were also analyzed including minimum, maximum, and variation measures. In addition to weather being constant for each area burned per day, analysis was also completed averaging topography and vegetation LiDAR variables. This significantly decreases the resolution of the data, but generally results in more visually noticeable trends between explanatory variables and the response, and highly improved model evaluation metrics.

2.3.4 Topographic Dataset

Topography was acquired from LiDAR, similarly processed through FUSION (McGaughey, 2014). Variables representing aspect, curvature, elevation, planform curvature (plancurv), profile curvature (profilecurv), slope, and solar radiation index (SRI) are used in analysis. Topography data follows the same specifications and processing procedure as the vegetation LiDAR data. LiDAR-derived topographic data is considered to be reliable and consistent across land types.

2.3.5 RdNBR

The Relativized differenced Normalized Burn Ratio (RdNBR) was used to quantify fire severity. RdNBR is a satellite-derived metric that employs Landsat images to quantify the change from pre-fire vegetation to post-fire vegetation (Jay D. Miller et al., 2009). The metric is calculated using the following equations:

$$NBR = \frac{Band4 - Band7}{Band4 + Band7}$$

$$RdNBR = \frac{NBR_{prefire} - NBR_{postfire}}{\sqrt{ABS\left(\frac{NBR_{prefire}}{100}\right)}}$$

The metric uses Landsat bands 4 and 7, associated with red color and dry earth, respectively ("Landsat 8 Bands « Landsat Science," n.d.). A sample of the continuous RdNBR values within each of the three final fire perimeters was used as the response variable for analysis.

Cloud-free images from Landsat 8 Operational Land Imager were obtained pre-fire and post-fire for all three fires: National Creek Complex (pre-fire 7/18/2015, post-fire 7/4/2016), Umpqua North Complex (pre-fire 7/7/2017, post-fire 6/24/2018), and South Umpqua Complex (pre-fire 7/17/2018, post-fire 6/11/2019). Post-fire images were selected to be one year after the pre-fire image. The 30-m resolution of Landsat 8 was used for RdNBR, the resolution of which was maintained throughout analysis.

2.3.6 Statistical Analysis

The plots were selected randomly with a minimum distance between plots of 200 meters. This distance minimizes the influence of short distance spatial autocorrelation for all variables while still maintaining a large sample size, as found by Kane et al. (2015). They found that a between plot distance of at least 180 meters reduced residual autocorrelation in their Random Forest models (Kane et al., 2015).

On a 30- by 30-meter grid of the forests, 2000 cells were selected with each cell representing one sample. Of these, 80% were randomly selected for model training, with 20% reserved for cross-validation. Such division of data is required to accurately measure model performance diagnostics for all methods excluding random forest analysis. RdNBR ranged from -5639.3 to 3724.38 across the entire test area, with the majority of the range falling in the low severity and unburned bin as defined by Reilly et al. (2017). The sample data had a minimum of -6.2, a maximum of 328.0, a mean of 15.7, and a standard deviation of 25.5. While this distribution is concerning, a stratified sample could provide a more reasonable sample in further research. By the bins stated above, none of the cells included in the sample burned at high severity. The percentage of high severity burned area is abnormally low, corroborated with local conversation of unexpectedly low severity fires on the Umpqua and Rogue River Forests.

Pairs plots were created using the graphics package in R version 3.6.1 (R Core Team, 2019). Each plot was visually analyzed, comparing all predictor variables against the RdNBR variable. Those with a visually recognizable trend were identified and selected as strong model candidates. Trends between predictor variables were noted as evidence of long distance autocorrelation.

For each variable, Pearson's correlation coefficient was calculated and analyzed. The top several of each variable category (weather, topography, and LiDAR vegetation metrics) were noted and assessed. Variable selection using best subsets was employed to identify groups of variables that may perform optimally. Regsubsets was employed, using the 'leaps' package in R (Lumley, 2017). Due to computational limitations, this automatic variable selection could only be performed on a small percent of the set of predictor variables in each iteration. Variables that expressed high individual Pearson's correlation coefficients were included in this analysis. An exhaustive search for the best subset was used to select candidate models.

Random Forest variable importance values were analyzed to identify a selection of the variables found to be most influential in reducing node impurity of the decision trees, represented in regression as the residual sum of squares (Louppe, Wehenkel, Sutera, & Geurts, 2013). Importance values are produced from the normalized difference between the MSE for prediction through the out-of-bag data and the predictor value of MSE after the predictors are permuted ("Importance function | R Documentation," n.d.). Random forest has a particular ability to handle extremely large datasets. Using this characteristic, we were able to analyze all variables including their summarized forms at once. Thus, variable importance values provide the most complete picture of how all the variables may contribute to modeling fire severity in a single snapshot.

2.4 Results

According to Pearson's correlation coefficient, Keetch-Byram drought index (KBDI) typically has the strongest absolute correlation with RdNBR among weather variables. Herbaceous fuel moisture (FMH) also has a consistent relationship with RdNBR, reaching its strongest point when modeling median RdNBR ($r = 6.154\%$). Aspect and slope feature heavily alongside curvature variables (curvature and planform curvature) when modeling RdNBR, but are overcome by various summaries of the curvature variables, including profile curvature, when modeling a summarized version of RdNBR. First returns feature heavily among the LiDAR variables, along with variables representing the six to nine meter stratum. This stratum is likely catching information about ladder fuels, which would indicate whether or not the fire would jump from a more manageable ground fire to a crown fire (S. Hoe et al., 2018).

Overall, weather and LiDAR variables contribute more to modeling RdNBR than topography variables. Variables relating to first returns were prominent for all three response variables (Table 2.1). The vertical bin of six to nine meters was represented in modeling RdNBR, along with variables describing the entire LiDAR point cloud above two and a half meters (Table 2.1). Weather variables featured KBDI, FMH, and BUI, all of which represent a measure of fuel moisture. Topography variables aspect, curvature measures, and slope, with only curvature, planform curvature, and profile curvature remaining when assessed against mean and median RdNBR (Table 2.1).

Best subset regression was considered to analyze the models selected by an automated variable selection method. Using the variables with Pearson's correlation coefficient values in the top 60% of the range, we found that a five variable equation was selected. Despite allowing for ten variables to enter, the simpler equation was selected, including weather and vegetative LiDAR variables. Topographic variables did not rank in the largest 60% of the range of Pearson's correlation coefficient values. In addition to a joint analysis, weather, topographic, and vegetation variables were assessed separately to ensure that at least one variable from each category remained in the final model. This led to a model with seven significant variables with at least one representative from each of the three categories.

Variables were also analyzed through random forest variable importance values to gain an understanding of the relative value of all variables when explaining RdNBR and mean RdNBR per daily burn area. The variables found to explain the highest percentage of error for RdNBR were KBDI, variables related to the stratum of six to nine meters, and various metrics describing the forest floor reflectance and height (Table 2.2). These variables can be interpreted roughly as corresponding to forest ladder fuel presence, variability and height of wood debris and fuels at the lower limit of LiDAR reach, and the inverse of fine fuel moisture located in the duff and top layer of soil.

Variables found to be of importance in the random forest regression for mean RdNBR per daily burn area focused on vegetative LiDAR variables, particularly those describing the lowest two and a half meters of the LiDAR data's reach. Weather variables representing median solar radiation and wind speed were found to be influential (Table 2.3).

To increase the usefulness of the models suggested by the variable selection methods, we considered variables that are likely to change based on the specific collection, flight parameters, and those that may not be applicable to diverse stands (S. Hoe et al., 2018). Removing intensity values, for example, removes variability between flights. However, due to the particular reliance on reflectance in our response variable, maintaining intensity values in the analysis allows us the opportunity to extract as much information as possible. In this particular setting LiDAR flights are intentionally kept as uniform as possible, providing some flexibility.

Models developed from variables of highest importance in random forest and those with the highest Pearson's correlation coefficient were developed, utilizing best subset regression. The model based solely on the highest Pearson's correlation coefficient values does perform the best according to most metrics, except for RMSPE. For prediction purposes, variables identified to be particularly important should be used (Table 2.4).

Pairs plots between the explanatory variables and the response variable, RdNBR, were difficult to discern. Few relationships had clearly visible patterns that could contribute to modeling RdNBR without excessive variation. Due to this challenge, we relied upon variables selected through high Pearson's correlation coefficient, best subset variable selection, and random forest importance values to suggest variables with stronger visually-recognizable relationships. Despite this dependency, no variables showed visually clear relationships (Figures 2.2, 2.3, 2.4). Transformations were not justified due to the complex nature of the response variable and the lack of visual pattern in the explanatory variables.

2.5 Discussion

Analysis of the weather variables showed that KBDI is most impactful in modeling RdNBR. The index KBDI represents duff and upper soil moisture content, calculated from precipitation and temperature measurements. Due to the apparent negative relationship between KBDI and RdNBR, it is clear when fine fuel on the forest floor has low moisture content, fire severity is typically higher. This matches our ecological expectations, suggesting that in our study area KBDI may truly be a weather variable to watch while making management decisions. The focus of LiDAR variables in the six to nine meter height bin suggests that this vertical space greatly impacts fire severity. While depending on specific highly-localized forest structure, this bin roughly corresponds to ladder fuels that would allow fire to jump from the ground to the crowns, damage to which can cause basal area mortality (S. Hoe et al., 2018).

The clear dominance in importance of weather and vegetation variables over topography variables is clear. Topography should be the most reliable of all three of the categories, as it is a fixed, largely unchanging, well-measured attribute. Vegetation we would expect to be highly variable and perhaps not exactly matched to the vegetation present at the time of the fire, which should reduce its influence. The influence of weather, while highly important in fire management during an event, should likely fluctuate in a model due to the questionable reliability of available weather data sources and our ability to employ the inherent highly variability of weather events effectively in models. Our data suggests that despite the variation and concerns, weather dominates fire severity models with a focus on describing moisture levels in available fuel. The selected vegetation metrics, also very influential, do seem to focus on capturing vegetation near the ground, the intensity of the returns describing the reflectivity of the forest materials, and the ladder fuels. All three of these categories clearly relate to fire dynamics. Understanding of ground fuels allows an understanding of whether or not a fire will succeed in spreading, while the intensity of the returns gives information closely related to aerial imagery similar to our response variable, and ladder fuels provide an opportunity for fire to cause more damage to larger trees that contribute most to basal area mortality as represented in RdNBR.

Overfitting was not evident in this analysis. Variables selected follow ecological and statistical principles. The logical selections suggest that the process did not follow the data more than the underlying biometric patterns. Correlation between the models and the response variable were relatively low, also suggesting that the overall fit of the model did not surpass visible patterns.

2.5.1 Limitations

Aerial LiDAR, by definition, observes and measures objects that are reached from sensors above the canopy. While powerful, lower layers of the forest, particularly the brush and wood debris layers, are often too deep to be reached by LiDAR at a sufficiently fine scale to be useful in analysis. Despite the power of highly detailed information from LiDAR, the lack of the ability to quantify wood debris and brush results in a lack of ground-level fuel information that is important for understanding small-scale fire behavior. Without knowing ground fuel connectivity, we will be unable to truly state whether a fire will be able to spread in the first place, let alone become powerful enough to be impacted by the mid- and upper-canopy level information we do have. However, some information is better than no information. Acquiring an idea of how fire could behave, particularly with the influence of weather and topography, provides another tool for practitioners to use when assessing the activity and danger a fire could present.

Settings made in the LiDAR collection also provide limitations. Height bins only definitively divide portions of the cloud into smaller parcels, providing a more intricate numeric understanding of the mass of points received from a LiDAR collect. While we work to make height bins meaningful, there is no guarantee that a certain bin would equate the same bin in another forest type or even section of the same forest. For example, a six to nine meter height bin could indicate a layer of the forest that contributes to ladder fuel, and so could provide information about the ability fire would have to jump from ground fire to a more destructive crown fire. However, this should relate to varying forest conditions and provide that extra information for the prediction.

RAWS data have a prominent limitation. Quality weather data are required to fully model fire behavior. However, it can be challenging to collect such data, particularly remotely, which led us to rely on the pre-existing RAWS program. While some analysis can be completed to identify whether the data could contribute to a model for RdNBR, it will be impossible to guarantee the same level of information for all fires. RAWS data quality varies immensely between stations. Stations are owned by different agencies and individuals, which provides a baseline of uncertainty, and due to the remotely automated characteristic they are prone to unknown disruptions that can result in missing or incorrect data. As with all data sources that are not collected by analysts, there is a chance the data that is needed is low quality. RAWS remain an unsolved situation. Some researchers face a lack of presence and skill to create fire progression maps, which are used to apply RAWS data spatially. To simplify the issue, researchers may only use summary data of temperature and relative humidity, two of the many potentially value variables (Estes, Knapp, Skinner, Miller, & Preisler, 2017). Specific location of the RAWS are also a challenge. As noted in the Klamath Mountains, just south of our study area and facing similar weather challenges, temperature inversions during large fire events can further complicate the usefulness of RAWS data (Estes et al., 2017). When preparing for fire, RAWS are checked to ensure that they meet minimum standards (Bureau of Land Management & Forest Service, 2019).

Modeling on a daily framework has inherent difficulties in application. While it does increase the performance of predictors, it negates the small-scale specificity available through LiDAR. This is primarily an issue for the usability of prediction. However, use of summary statistics over burned area per day does allow us to see general patterns that may be missed on a smaller scale.

2.6 Conclusion

An understanding of variables is important to any modeling venture, particularly when the predictive power of the models themselves is low under multiple linear regression. Visually, it is important to assess whether patterns, particularly non-linear patterns, are present. However, multicollinear variables must be resolved or assessed individually through pairs plots or Pearson's correlation coefficient, or jointly through methods such as random forest. Through such analysis, we suggest that within the Umpqua and Rogue River National Forests analysts should rely on KBDI, LiDAR intensity metrics, metrics measuring six to nine meters from the ground, aspect, and curvature metrics to model fire severity as RdNBR. These variables capture the most variation and explanatory power while still maintaining the ecological foundation of fire ecology.

Summarized versions of RdNBR may be modeled if analysts require multiple linear regression, do not need small-scale specificity, and have sufficient fires on the landscape to support daily averaged variables. Using a summarized variable dramatically increases correlation between the response variable and the available explanatory variables, leading to stronger apparent relationships. However, the benefit of LiDAR relies on its ability to identify submeter structure. This benefit is already diminished through the need to use 30- by 30-meter pixels to model the Landsat-derived RdNBR. To further diminish the resolution would reduce the amount of information available to the models.

CHAPTER 3: PREDICTIVE MODELING TO CAPTURE VARIABILITY IN FIRE SEVERITY

3.1 Abstract

Variable prediction is crucial in our analyses of ongoing phenomenon. Prediction of effects and landscape characteristics allow us to assess the potential and danger of impactful events. Fire effects are particularly important to capture and predict, due to the number of components that can change the bearing of fire upon the landscape and the difficulty in predicting the entirety of fire behavior. We attempt to model fire severity as RdNBR, a proxy for basal area mortality, from weather and LiDAR-derived vegetation and topography variables. We found that PCA can most easily attain a high adjusted R-squared value, up to 99% with thirty variables. While this indicates overfitting, a more

limited number of PCA variables in regression may be the most effective means we have of capturing the variability in our data that contributes to RdNBR. Random forest was not found to be effective unless RdNBR was summarized over burn day. Our results illustrate the importance of many related variables in modeling RdNBR, as well as the mixed results of methods that eliminate multicollinearity.

3.2 Introduction

Prediction is vital in forest biometrics. Prediction models are created differently than the typical inference models – most assumptions do not need to be followed. The goal is simply to use the most information possible to provide the best estimation of future or unmeasured samples, without overfitting. Because of this, statistics such as adjusted R-squared and MSPE become important. These statistics penalize for adding variables that do not provide comparable information to the prediction and provide an idea of the error that the model does not account for in prediction cases.

Weather is a key driver in fire behavior. It is also extremely variable, and weather shifts cause much of the danger during active fire management situations. While capturing hourly weather shifts is not reasonable in our models, and predicting such shifts and their impacts is beyond the scope of this research, we can attempt to recognize future variability by noting how our predictive models shift in a few different potential weather situations. Much more detailed and accurate fire severity and weather data would be needed to train such fine-tuned models that would be able to recognize hourly changes in weather. As long as vegetation data is fairly frequently updated, the primary unknown variables should be solely related to weather.

Still, prediction models such as these can be used to give an idea to managers about severity patterns in the study area. Managers could use this information to reduce high severity instances around areas or buildings of value, or to create breaks in potential severity that would allow management during fires to be more successful. Essentially, this would create severity breaks under certain weather conditions. An understanding of which variables are important to the specific model could also allow managers to focus on managing those components to adjust the severity of fires to their specific management goals.

3.3 Methods

3.3.1 Study Area

The entirety of the Umpqua and Rogue River National Forests are included in this portion of analysis (Figure 2.1). The models are trained on three fires that occurred within the forests: the National Creek Complex, the Umpqua North Complex, and the South Umpqua Complex. Models are subsequently applied and analyzed on the entire study area.

The study area is as described in section 2.3.1, dominated by diverse vegetation, high growth rates, and dramatic topography. In these forests, slopes are frequently too steep for machinery, causing geared-up ground crews to enact fire containment measures such as fire breaks (O'Connor, Calkin, & Thompson, 2017; Schneider & Breedlove, n.d.). Fire can become too unpredictable or conditions can be too dramatic when slopes are steep, topography is complex, or vegetation is dense. In these instances, fire crews must work from outside the area at a safe distance. In landscapes such as the Umpqua and Rogue River National Forests where slopes are steep and vegetation grows rapidly, this results in large areas unable to be directly managed during a fire. In these situations, fires burn in larger areas and require pre-planning for fire crews to do more than focus on preserving roadways and buildings.

3.3.2 Fire Severity Prediction Models

The models from Chapter 2 section 2.4 are considered in model comparisons, along with models developed in this chapter. However, these models were created based on variable importance. Optimizing prediction power allows use of different methods, completed in this analysis.

Fire severity prediction models are notoriously challenging to model with severity as a continuous variable (Gordon, Price, & Tasker, 2017; L. Harris & Taylor, 2017; McCarley et al., 2017; Parks et al., 2018). Due to the challenges in creating a reliable prediction, these fire severity models should only be used in a prediction setting for managers to gain a general understanding of what may be likely on the landscape. The most value can be found in understanding how variables that can be actively managed by practitioners, such as tree height or density variability, may contribute to fire severity.

3.3.3 Weather Dataset

Weather is a key component of the fire severity model. If wind shifts direction, it could push a fire upslope, allowing it to run quickly to the top of a mountain. If wind becomes lighter and from a consistent direction, it could provide the opening needed for fire crews to contain challenging portions of the perimeter (Schneider & Breedlove, n.d.). While the weather dataset used in the analysis of Chapter 2 is maintained for model development, average monthly weather is used to develop an understanding of model sensitivity to varied weather situations.

3.3.4 Statistical Analysis

Principal component analysis (PCA) is an extremely useful tool in multivariate models that use a large number of variables. PCA is a common statistical method that uses all variables as needed to explain the most variation in the response variable, while orthogonalizing predictors to avoid multicollinearity. Multicollinearity is an inherent in LiDAR datasets, as most of the variables are related to each other, largely due to how they are calculated. A variable representing a vertical bin must be related to the bin above it, through proximity and logic. While multicollinearity can be avoided simply by only including a handful of necessary variables or removing variables that are unlikely to contribute to the response variable, removing variables from analysis could unintentionally reduce the power of the final model. PCA allows all variables to be represented in the final model, denoted through fewer predictors and transformed to eliminate most multicollinearity issues. Analysis of the principle components and further use of the components in regression helps to maximize our use of the variables available in this analysis.

In addition to the individual variable importance values used in Chapter 2, random forests are used to model fire severity as a nonparametric method. Use of random forest decision trees allow us to fully utilize information that the sample data provides, unconstrained to parametric model assumptions (Breiman, 2001b). An expansion of basic decision trees, random forest uses randomized samples of the data to develop multiple Classification and Regression Trees without pruning (Breiman, 2001a). Averaging the decision trees produces a more stable version of simple decision trees, reducing potential of overfitting the training data (Breiman, 2001a). This method is unsupervised, providing an answer that is much less influenced by manual variable selection. Manual variable selection is common in operational LiDAR modeling. However, it is a tool that is applied beyond where it is most useful: inference and producing a fully interpretable model. Other methods, including various unsupervised methods such as random forest allow for models created primarily for prediction without imposing limits of the model's particular mechanisms (Breiman, 2001b).

3.4 Results

The RdNBR variable analyzed in this chapter had a minimum of -6.2, a maximum of 328.0, a mean of 15.7, and a standard deviation of 25.5. The first principal component of the vegetation LiDAR variables showed a reliance on first return variables and intensity of the returns above two and a half meters, heavily employing the summary variables (Table 3.2). The amount of variance explained by each principal component falls rapidly (Figure 3.1). Topography-related principal components spread the majority of the proportion of variance explained across the first few components, dropping rapidly after the third and seventh components (Figure 3.2). Elevation and aspect summary variables dominate the first principal component, suggesting that these variables contain the most variance (Table 3.2). The variation expressed by the first principal component of weather variables is spread evenly across four of the summary variables of the initial spread index (ISI), followed by an even spread across buildup index (BUI) variables and duff moisture code (DMC) variables (Table 3.2). The contribution of the principal components drops the most rapidly for weather variables, suggesting that a fewer number of principal components may be needed in regression (Figure 3.3).

Despite the apparent importance of vegetative LiDAR metrics throughout analysis, regression with principal components states that weather metrics contain the majority of variation present in RdNBR. A linear regression containing the first five principal components from both the vegetation LiDAR data and the topography data with the first twenty-five components of weather data has an adjusted r-squared of 0.9892. A linear regression of only the first twenty-five components of weather data has an adjusted r-squared of 0.9836 (Table 3.1). Reduction of the number of weather variable components leads to a significant drop in adjusted r-squared, suggesting that the twenty-five variable principal component equation may be the most parsimonious basic linear equation. The addition of five more weather variable principal components, in place of adding topography- or vegetation-related components, leads to an adjusted r-squared of 0.9974, increasing the value far more than the addition of ten topography and vegetation components (Table 3.1). Given the similarities of the RMSPE values across the models and the notably lower BIC values, the model containing the thirty weather principal components seems ideal.

However, the adjusted r-squared values indicate a much stronger correlation than what we may expect from the visual and statistical relationships addressed in Chapter 2. It seems likely that overfitting becomes prominent with PCA in this dataset. Despite an abrupt drop in fit among variables with fewer than 25 weather-related principal components, such a model may be needed to combat reliance on our sample over assessment of the relationships overall.

Random forest regression was not promising for modeling RdNBR. Employing all variables, a model for a response variable of RdNBR had a mean of squared residuals of 647.499 and explained -0.43% of the variance, effectively stating that this model is no better than simply applying the mean value across the study site. However, a random forest model for modeling the mean of RdNBR per daily burn area resulted in a mean square of residuals of 0.4836 and explained 98.8% of the variance in mean RdNBR. While this variable has much less flexibility and opportunity for application, it may be able to provide large area daily estimates. In such a variable situation, the dramatic decrease in mean square of residuals could be worth the loss of specificity provided by a 30- by 30-meter raster of individually calculated RdNBR values.

3.5 Discussion

Principal component analysis and regression is particularly useful in our analysis. Capturing even 99.74% of the variability in RdNBR, we can produce reliable models across our dataset. The first principal components generally corroborate the results of the analysis in Chapter 2, but avoids curvature metrics. While KBDI does not contribute most to the first principal component of weather, other metrics contribute to an understanding of duff moisture, fire spread, and fuel buildup. These three variables create a similar understanding as KBDI, an index of duff and upper soil moisture content. Topography and weather variables selected in the first principal components are all easily interpretable. However, the first few LiDAR metrics in the first principal component are less interpretable. They focus on first return and intensity metrics, and do not suggest much captured variability beyond the summary metrics. This is partially a relic of many variables contributing moderately to the first return and the large variability in certain LiDAR metrics, diminishing the usefulness of interpreting the first principal component. For vegetation LiDAR metrics, other methods should be relied upon.

Random forest, while theoretically promising, did not perform well with our data. The lack of clear and dominant relationships between the explanatory variables and RdNBR were captured far better in PCR than random forest, perhaps due to the fact that PCR will use all of the variables while random forest randomly selects a portion for calculations on each node. For response variables with more clear relationships to vegetative LiDAR metrics, random forest may be more effective. However, the method did perform well when modeling summarized forms of RdNBR. The lessened variability likely contributed to the method more successfully predicting the summarized RdNBR.

Our findings suggest that PCR may be an opportune method for modeling RdNBR from vegetative LiDAR, topography, and weather data. This method takes advantage of the many contributing variables to form a cohesive model. Unlike other methods, multicollinearity is fully addressed while information from all variables that contribute to the variability in the data is included. Since principal components are not developed based on a specific response variable, this method seems like it would be effective for models for response variables other than RdNBR.

3.6 Conclusion

Modeling RdNBR well in an easily usable and interpretable fashion is a challenge. Use of PCA allows us to maximize our use of the large number of variables we have available through LiDAR and produces a satisfactory model. However, PCA limits interpretation to just the first principal component, which does not greatly help our understanding of vegetation-related LiDAR variables. Other methods can be used for interpretation of vegetation metrics, while PCA can be relied on for its predictive power.

Random forest performs well when modeling a summary metric of RdNBR. However, it fails to provide an adequate prediction of RdNBR with unique values on the 30- by 30-meter scale. While it can produce interpretation through its importance values, as in Chapter 2, we do not suggest use of the predictive random forest model for fire severity represented by RdNBR in this landscape.

Based on our findings, vegetation management efforts should focus primarily on lessening the vegetation between six and nine meters in height and maximizing the moisture in the duff layer, should managers wish to lessen fire severity. However, these analyses spanning the Umpqua and Rogue River National Forests are unlikely to be this specific at a fine scale. Without inclusion of many variables and further analysis concerning presence of overfitting, it is unlikely that our models will differentiate between each small ecosystem in the study area. Analysis at a stand level will still be required to support specific management plans.

CHAPTER 4

4.1 General Conclusion

The variable information we gathered supports the claim that our models should follow ecological relationships observed relating to fire severity. Variables relating to the moisture of ground fuels in the duff layer and topmost layer of soil form one important category in our models. This category featured KBDI, DMC, FMH, and ISI, all relating to moisture, quantity, and the ability of fire to spread across this layer of the forest. Another category was the layer of vegetation between six and nine meters above the forest floor. This category advises the model about the vegetation ladder, a key section of vegetation that can help a fire rise into tree crowns and convert the event into a more destructive crown fire. The final typical category was one populated by variables describing general LiDAR cloud metrics, or general descriptor variables about the entire vegetation structure. While less specific, this category uses the power of LiDAR data in its ability to capture layers of information, not simply relying on a canopy model or a proxy variable like stand age to feed the model vital fuel information. These findings suggest that practitioners should focus manipulation of fire severity through direct measures or ignition timing on the moisture level of the duff layer, density and quality of vegetation between six and nine meters of vertical height, and overall vegetation control.

Principal component analysis and regression works well for modeling RdNBR as a representation of fire severity. This method produces high adjusted R^2 values, lowers model error, and lowers prediction error from models produced through multiple linear regression of the LiDAR, weather, and topography variables. PCA likely works effectively since it uses many variables at once to capture the most possible variability. Since landscapes, vegetation, weather, and impacts of fire are all highly variable, the variability captured by the principal components seem to be most effective at producing a reliable fire severity model.

Random forest was unable to produce similar results, despite 1600 plots. The nonparametric model was unable to provide a marginally successful result when modeling RdNBR. However, if only a summary metric such as the mean of the value for each day the fire burns is desired, random forest could provide an effective means of prediction similar to that of PCA. When using the mean of RdNBR, the percent of variance explained increases from effectively 0% to 98.8% within the random forest modeling framework. This increase in the R^2 value, while similar to the PCA-based model's adjusted R^2 value of 99.7%, sacrifices the location-specificity of the highly detailed LiDAR data beyond the level of the Landsat resolution. While similar, if not as dramatic, increases were found in equations based on variables selected from random forest's importance values, Pearson's correlation coefficient, and best subsets variable selection, the decrease in resolution and the acceptance of multicollinearity, bias, and error does not support the choice to use a response variable summary metric in analysis. However, if managers or legislation required daily metrics to be considered, such a model would be highly effective.

Despite finding correlation and effective models, LiDAR still may not be the optimal data source for modeling fire severity. Due to its aerial status, LiDAR cannot fully sense or parameterize ground vegetation and wood debris. Without this information, it is unlikely that an optimal model can be developed despite the high adjusted R^2 values found among PCA-based models, simply because the ecological principles cannot be fully described by the variables available to the analysis. However, the focus on RdNBR does align with information available from LiDAR. Highly correlated with basal area mortality, RdNBR is related to a fire response that can be observed aerially. Perhaps including data that can clearly describe whether or not a tree is living would further improve modeling efforts, beyond the general forest structure that LiDAR can provide.

Using these models, users can see and share the impact of variable choice, variability of the data, and modeling method. The models themselves can be used to see the suggested influence of a management choice or natural event. A fire severity model can be developed into many useful maps that can be used by practitioners and teachers to help the populace understand the behavior of modern fires in the area. The fire complexes, while large, are generally low severity and much less harmful to the ecology than many forest users might believe. Visual aids may help increase the belief and trust in the use of fire to accomplish management goals.

While this thesis does not claim to solve fire severity models, we do suggest that some methods may be highly effective at meeting our needs. Principal component analysis, in particular, seems to be a promising method. While using a model with only weather variables simply elucidates the importance of weather in fire severity models, a model that uses components from weather, topography, and vegetation variables seems to provide an ecologically and predictively sound method for similar forests. While this model may suggest overfitting from the high adjusted R^2 value, should that become a concern or if analysts wish to use the model outside of the study area, the number of components used can be adjusted. Particular variables selected in Chapter 1 do suggest an ecological relationship appearing within the chosen models, which suggests an accurate portrayal of variable relationships. Perhaps, given sufficient data and computational power, fire severity could be reliably modeled and predicted across forest types, contributing to fire management planning, fire-wise forest management, and community education.

Research on effective sources of wood debris and vegetation near the ground and the methods to use that information in large area-based analysis would be very beneficial. Currently, much of what is available is data collected from transects and summarized to a population estimate. This summarization does not easily fold into grid-based spatial analyses and requires more time spent on analysis than practitioners may have. A straightforward method that could be used alongside typical LiDAR analyses would be beneficial for researchers and practitioners.

Straightforward, effective methods of using RAWS data could improve the quality of data available to analysts. Daily fire progression maps for all fires would make models with weather data originating from RAWS much simpler to develop. An established method for selecting specific RAWS as well as a database stating which RAWS were used during the fires would allow for more defendable and repeatable use of this resource. The RAWS system is a strong opportunity for even those dabbling in research to access a broad database of weather information. Use should be encouraged through clear procedures and easy access.

To improve understanding of LiDAR variables in modeling fire severity, models should be developed in ecosystems different from the Umpqua and Rogue River National Forests. While our study area does have high variability within its borders, comparing our results to results of a forest dominated in ponderosa pine or a plantation could provide interesting comparisons. Comparison across forest types could drive our understanding of where LiDAR can most contribute to our understanding and prediction of fire severity.

BIBLIOGRAPHY

ArcGIS Desktop (Version 10.3). (2014). Redlands, CA: Environmental Systems Research Institute: ESRI.

Boyd, R. (1999). *Indians, fire, and the land in the Pacific Northwest*. Retrieved from http://archive.org/details/indiansfirelandi00robe

Bradshaw, L., & Jolly, M. (2014). FireFamilyPlus (Version 4.2). Retrieved from https://www.firelab.org/project/firefamilyplus

Breiman, L. (2001b). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Bureau of Land Management, & Forest Service. (2019). *2019 Interagency Standards for Fire and Fire Aviation Operations*. 510.

Dickmann, D., & Cleland, D. (2019). *Fire Return Intervals and Fire Cycles for Historic Fire Regimes in the Great Lakes Region: A Synthesis of the Literature*.

Dunn, C. J., & Bailey, J. D. (2016). Tree mortality and structural change following mixed-severity fire in Pseudotsuga forests of Oregon's western Cascades, USA. *Forest Ecology and Management*, *365*, 107–118. https://doi.org/10.1016/j.foreco.2016.01.031

Estes, B. L., Knapp, E. E., Skinner, C. N., Miller, J. D., & Preisler, H. K. (2017). Factors influencing fire severity under moderate burning conditions in the Klamath Mountains, northern California, USA. *Ecosphere*, *8*(5). https://doi.org/10.1002/ecs2.1794

Fekety, P. A., Falkowski, M. J., Hudak, A. T., Jain, T. B., & Evans, J. S. (2018). Transferability of Lidar-derived Basal Area and Stem Density Models within a Northern Idaho Ecoregion. *Canadian Journal of Remote Sensing*, *44*(2), 131–143. https://doi.org/10.1080/07038992.2018.1461557

FUSION/LDV LIDAR analysis and visualization software. (n.d.). Retrieved September 8, 2019, from http://forsys.cfr.washington.edu/fusion/fusion_overview.html

Galbraith, S. M., Cane, J. H., Moldenke, A. R., & Rivers, J. W. (2019). Wild bee diversity increases with local fire severity in a fire-prone landscape. *Ecosphere*, *10*(4), e02668. https://doi.org/10.1002/ecs2.2668

Gordon, C. E., Price, O. F., & Tasker, E. M. (2017). Mapping and exploring variation in post-fire vegetation recovery following mixed severity wildfire using airborne LiDAR. *Ecological Applications: A Publication of the Ecological Society of America*, *27*(5), 1618–1632. https://doi.org/10.1002/eap.1555

Graham, M. H. (2003). Confronting Multicollinearity in Ecological Multiple Regression. *Ecology*, *84*(11), 2809–2815. https://doi.org/10.1890/02-3114

Harris, G. A., Franklin, J., & Dyrness, C. T. (1990). Natural Vegetation of Oregon and Washington. *Journal of Range Management*, *43*(1), 85. https://doi.org/10.2307/3899130

Harris, L., & Taylor, A. H. (2017). Previous burns and topography limit and reinforce fire severity in a large wildfire. *Ecosphere*, *8*(11), e02019. https://doi.org/10.1002/ecs2.2019

Hudak, A. T., Crookston, N. L., Evans, J. S., Falkowski, M. K., Smith, A. M. S., Gessler, P. E., & Morgan, P. (2006). Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral data. *Canadian Journal of Remote Sensing. 32(2): 126--138.*, 126–138.

Importance function | R Documentation. (n.d.). Retrieved September 8, 2019, from https://www.rdocumentation.org/packages/randomForest/versions/4.6-14/topics/importance

Kane, V. R., Lutz, J. A., Alina Cansler, C., Povak, N. A., Churchill, D. J., Smith, D. F., … North, M. P. (2015). Water balance and topography predict fire and forest structure patterns. *Forest Ecology and Management*, *338*, 1–13. https://doi.org/10.1016/j.foreco.2014.10.038

Keeley, J. E. (2009a). Fire intensity, fire severity and burn severity: A brief review and suggested usage. *International Journal of Wildland Fire*, *18*(1), 116. https://doi.org/10.1071/WF07049

Keeley, J. E. (2009b). Fire intensity, fire severity and burn severity: A brief review and suggested usage. *International Journal of Wildland Fire*, *18*(1), 116. https://doi.org/10.1071/WF07049

Kelly, M., & Di Tommaso, S. (2015). Mapping forests with Lidar provides flexible, accurate data with many uses. *California Agriculture*, *69*(1), 14–20.

Kneeshaw, K., Vaske, J. J., Bright, A. D., & Absher, J. D. (2004). Situational Influences of Acceptable Wildland Fire Management Actions. *Society & Natural Resources*, *17*(6), 477–489. https://doi.org/10.1080/08941920490452427

KNEESHAW, K., VASKE, J. J., BRIGHT, A. D., & ABSHER, J. D. (2004b). Situational Influences of Acceptable Wildland Fire Management Actions. *Society & Natural Resources*, *17*(6), 477–489. https://doi.org/10.1080/08941920490452427

Kolden, C. A. (2019). We’re Not Doing Enough Prescribed Fire in the Western United States to Mitigate Wildfire Risk. *Fire*, *2*(2), 30. https://doi.org/10.3390/fire2020030

Landsat 8 Bands « Landsat Science. (n.d.). Retrieved September 8, 2019, from https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-bands/

Lefsky, M., COHEN, W., Parker, G., & Harding, D. (2009). Lidar Remote Sensing for Ecosystem Studies. *BioScience*, *52*, 19–30. https://doi.org/10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 431–439). Retrieved from http://papers.nips.cc/paper/4928-understanding-variable-importances-in-forests-of-randomized-trees.pdf

McCarley, T. R., Kolden, C. A., Vaillant, N. M., Hudak, A. T., Smith, A. M. S., Wing, B. M., … Kreitler, J. (2017). Multi-temporal LiDAR and Landsat quantification of fire-induced changes to forest structure. *Remote Sensing of Environment*, *191*, 419–432. https://doi.org/10.1016/j.rse.2016.12.022

McGaughey, R. J. (2018). FUSION/LDV: Software for LIDAR Data Analysis and Visualization (Version 3.80). USFS, Pacific Northwest Research Station.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., … McKee, W. A. (1999). Use of Large-Footprint Scanning Airborne Lidar To Estimate Forest Stand Characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment*, *67*(3), 298–308. https://doi.org/10.1016/S0034-4257(98)00091-1

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. https://doi.org/10.1007/s10021-008-9201-9

Miller, J. D., Skinner, C. N., Safford, H. D., Knapp, E. E., & Ramirez, C. M. (2012). Trends and causes of severity, size, and number of fires in northwestern California, USA. *Ecological Applications*, *22*(1), 184–203. https://doi.org/10.1890/10-2108.1

Miller, Jay D., Knapp, E. E., Key, C. H., Skinner, C. N., Isbell, C. J., Creasy, R. M., & Sherlock, J. W. (2009). Calibration and validation of the relative differenced Normalized Burn Ratio (RdNBR) to three measures of fire severity in the Sierra Nevada and Klamath Mountains, California, USA. *Remote Sensing of Environment*, *113*(3), 645–656. https://doi.org/10.1016/j.rse.2008.11.009

Miller, Jay D., & Thode, A. E. (2007). Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR).

Remote Sensing of Environment, *109*(1), 66–80. https://doi.org/10.1016/j.rse.2006.12.006

Mote, P.W., J. Abatzoglou, K.D.Dello, K. Hegewisch, and D.E. Rupp, 2019: Fourth Oregon Climate Assessment Report. Oregon Climate Change Research Institute. occri.net/ocar4.

O'Connor, C. D., Calkin, D. E., & Thompson, M. P. (2017). An empirical machine learning method for predicting potential fire control locations for pre-fire planning and operational fire management. *International Journal of Wildland Fire*, *26*(7), 587. https://doi.org/10.1071/WF16135

Olson, D. L., & Agee, J. K. (2005). Historical fires in Douglas-fir dominated riparian forests of the southern Cascades, Oregon. *Fire Ecology*, *1*(1), 50–74. https://doi.org/10.4996/fireecology.0101050

Parks, S. A., Holsinger, L. M., Panunto, M. H., Jolly, W. M., Dobrowski, S. Z., & Dillon, G. K. (2018a). High-severity fire: Evaluating its key drivers and mapping its probability across western US forests. *Environmental Research Letters*, *13*(4), 044037. https://doi.org/10.1088/1748-9326/aab791

Parks, S. A., Holsinger, L. M., Panunto, M. H., Jolly, W. M., Dobrowski, S. Z., & Dillon, G. K. (2018b). High-severity fire: Evaluating its key drivers and mapping its probability across western US forests. *Environmental Research Letters*, *13*(4), 044037. https://doi.org/10.1088/1748-9326/aab791

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Radeloff, V. C., Helmers, D. P., Kramer, H. A., Mockrin, M. H., Alexandre, P. M., Bar-Massada, A., … Stewart, S. I. (2018). *Rapid growth of the US wildland-urban interface raises wildfire risk*. https://doi.org/10.1073/pnas.1718850115

Reilly, M. J., Dunn, C. J., Meigs, G. W., Spies, T. A., Kennedy, R. E., Bailey, J. D., & Briggs, K. (2017). Contemporary patterns of fire extent and severity in forests of the Pacific Northwest, USA (1985–2010). *Ecosphere*, *8*(3), e01695. https://doi.org/10.1002/ecs2.1695

Rogue River-Siskiyou National Forest—Home. (n.d.). Retrieved September 8, 2019, from https://www.fs.usda.gov/rogue-siskiyou/

S. Hoe, M., Dunn, C., & Temesgen, H. (2018). Multitemporal LiDAR improves estimates of fire severity in forested landscapes. *International Journal of Wildland Fire*. https://doi.org/10.1071/WF17141

Schneider, R. G., & Breedlove, D. (n.d.). *Fire Management Study Unit*. 38.

Scott, J. H., & Burgan, R. E. (2005). *Standard fire behavior fuel models: A comprehensive set for use with Rothermel's surface fire spread model* (No. RMRS-GTR-153; p. RMRS-GTR-153). https://doi.org/10.2737/RMRS-GTR-153

Silva, C. A., Klauberg, C., Hudak, A. T., Vierling, L. A., Fennema, S. J., & Corte, A. P. D. (2017). Modeling and mapping basal area of Pinus taeda L. plantation using airborne LiDAR data. *Anais Da Academia Brasileira de Ciências*, *89*(3), 1895–1905. https://doi.org/10.1590/0001-3765201720160324

Sousa, W. P. (1984). The Role of Disturbance in Natural Communities. *Annual Review of Ecology and Systematics*, *15*, 353–391. Retrieved from JSTOR.

Umpqua National Forest—Home. (n.d.). Retrieved September 8, 2019, from https://www.fs.usda.gov/umpqua/

U.S. Department of the Interior: U.S. Department of Agriculture. (n.d.). GeoMAC Wildfire Application. Retrieved October 7, 2019, from Geospatial Multi-Agency Coordination website: https://www.geomac.gov/

U.S. Fish & Wildlife Service: Fire Management—Wildland Fire Use. (n.d.). Retrieved September 8, 2019, from https://www.fws.gov/fire/what_we_do/wildland_fire_use.shtml

White, A., Briles, C., & Whitlock, C. (2015). Postglacial vegetation and fire history of the southern Cascade Range, Oregon. *Quaternary Research*, *84*(3), 348–357. https://doi.org/10.1016/j.yqres.2015.09.007

Whitman, E., Johnston, J., Schiks, T., Paugam, R., & Cantin, A. (2019). *Imaging Postfire Environments*. https://doi.org/10.1007/978-3-319-51727-8_175-1

Zald, H., & Dunn, C. (2017b). Severe fire weather and intensive forest management increase fire severity in a multi-ownership landscape. *Ecological Applications*, *28*. https://doi.org/10.1002/eap.1710

Table 1.1 Key weather variable names and associated variable abbreviations.

Name	Code
1000-Hour Fuel Moisture	X1000
Herbaceous Fuel Moisture	FMH
Duff Moisture Code	DMC
Buildup Index	BUI
Keetch-Byram Drought Index	KBDI
Initial Spread Index	ISI

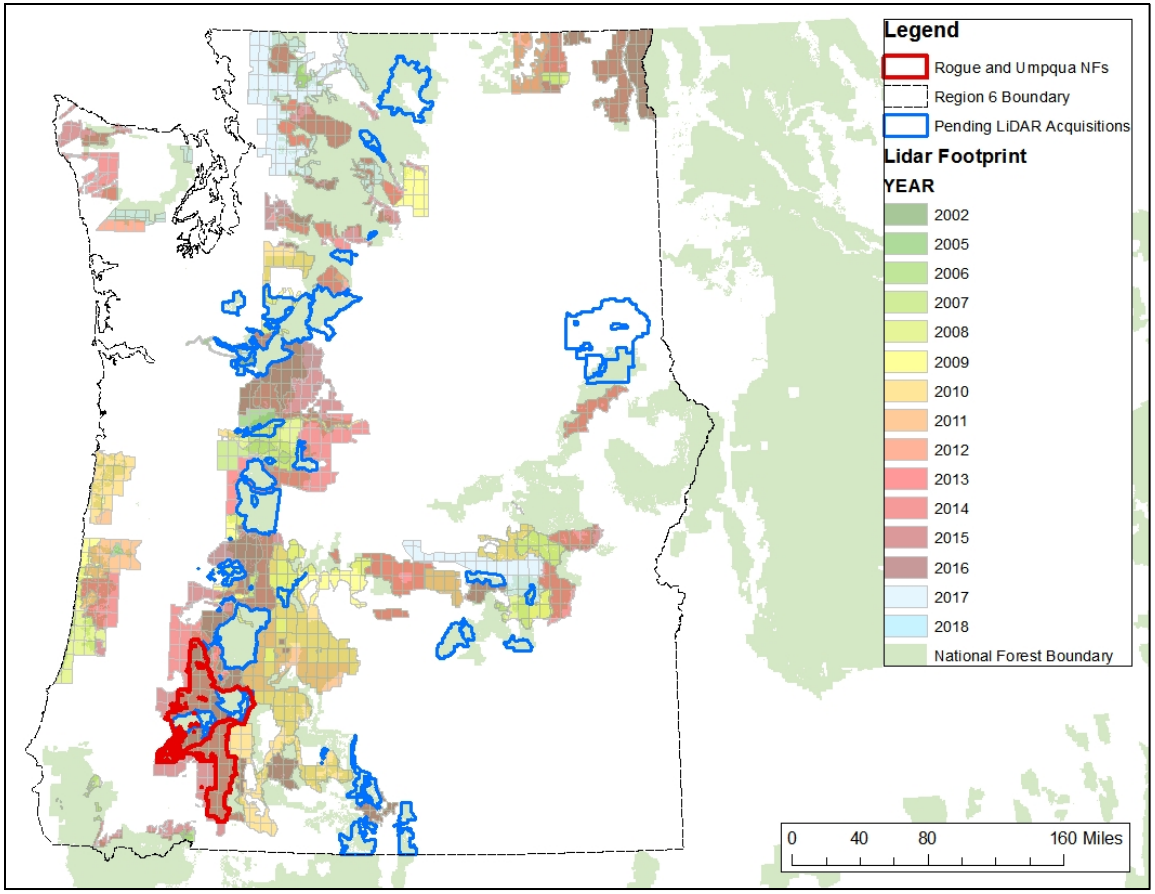

Figure 2.1 Map of LiDAR years available within the Forest Service at the time of the study. Rogue River and Umpqua National Forests indicated within the red outline.

Table 2.1 Largest five absolute values of Pearson's correlation coefficients (rVals) for weather, topography, and LiDAR variables when modeling the response variables RdNBR, RdNBR averaged over burn day, and median RdNBR for each burn day.

Response	RdNBR					
Category	Weather		Topo		LiDAR	
	varNames	rVals	varNames	rVals	varNames	rVals
	X1000	0.1652	med_plancurv	0.0326	strata_6to9M_return_proportion	0.1338
	FMH	0.1664	mean_slope	0.0331	1st_cover_above2p5	0.1354
	DMC	0.1665	med_slope	0.0371	strata_3to6M_total_return_cnt	0.1483
	BUI	0.1678	med_curvature	0.0372	FIRST_RETURNS_all_cover_above2p5	0.1508
	KBDI	0.1801	max_aspect	0.0475	strata_6to9M_total_return_cnt	0.1608
Response	Mean RdNBR					
Category	Weather		Topo		LiDAR	
	varNames	rVals	varNames	rVals	varNames	rVals
	max_FMH	0.6569	max_plancurv	0.4302	max_int_IQ_above2p5	0.6950
	min_KBDI	0.6794	max_profilecurv	0.4494	max_int_P75_above2p5	0.7016
	mean_KBDI	0.6806	min_curvature	0.4834	max_FIRST_RETURNS_int_stddev_above2p5	0.7090
	med_KBDI	0.6808	min_plancurv	0.4986	max_FIRST_RETURNS_int_AAD_above2p5	0.7209
	max_KBDI	0.6809	max_curvature	0.5045	max_FIRST_RETURNS_int_variance_above2p5	0.7465
Response	Median RdNBR					
Category	Weather		Topo		LiDAR	
	varNames	rVals	varNames	rVals	varNames	rVals
	med_KBDI	0.6072	max_profilecurv	0.3916	max_FIRST_RETURNS_int_AAD_above2p5	0.5965
	max_KBDI	0.6073	min_curvature	0.3983	max_FIRST_RETURNS_int_stddev_above2p5	0.6011
	mean_KBDI	0.6074	max_plancurv	0.4033	max_int_variance_above2p5	0.6264
	min_KBDI	0.6078	max_curvature	0.4330	med_strata_1p5to2M_mode	0.6286
	max_FMH	0.6154	min_plancurv	0.4471	max_FIRST_RETURNS_int_variance_above2p5	0.6507

Table 2.2 Largest random forest variable importance values when modeling RdNBR.

Percent Increase in MSE	Variable Name
7.9761	KBDI
5.8384	strata_6to9M_total_return_cnt
5.5674	r6_cnt_above2p5
5.2347	all_1st_cover_above_mean
4.8863	CoverStrata_6to9m
4.7522	strata_0p5to1M_mode
4.5595	r5_cnt_above2p5
4.4082	FIRST_RETURNS_int_P99_above2p5
4.2404	int_P10_above2p5
4.1395	FIRST_RETURNS_1st_cover_above_mode
4.1250	strata_30to36M_mode
4.1213	strata_0to0p15M_median
3.9926	topo_curvature
3.9765	CoverStrata_3to6m

Table 2.3 Largest random forest variable importance values when modeling mean RdNBR per daily burn area.

Percent Increase in MSE	Variable Name
14.6762	max_FIRST_RETURNS_int_AAD_above2p5
12.5885	max_strata_0to0M_mean
10.6412	max_FIRST_RETURNS_int_stddev_above2p5
10.6259	med_elev_L3_plus
10.1939	max_FIRST_RETURNS_int_variance_above2p5
9.8432	max_strata_0to0p15M_min
9.6301	med_strata_1p5to2M_mean
8.4096	min_FIRST_RETURNS_elev_L4_above2p5
8.3065	med_SolR
8.2562	mean_strata_6to9M_kurtosis
7.6840	max_strata_15to18M_skewness
7.3288	med_strata_0to0p15M_CV
6.7068	sd_Wind
6.4158	med_strata_1to1p5M_median
6.2101	mean_strata_1to1p5M_min

Table 2.4 Descriptive metrics for models developed from high variable importance variables and high Pearson's correlation coefficient (r) variables.

Model Description	RMSE	RMSPE	AIC	BIC	Bias	Adjusted R^2
High variable importance from RF	24.61	28.08	14696.94	14734.59	0.1778	0.0746
Highest r values	23.08	28.26	11585.69	11632.05	-0.0306	0.1039

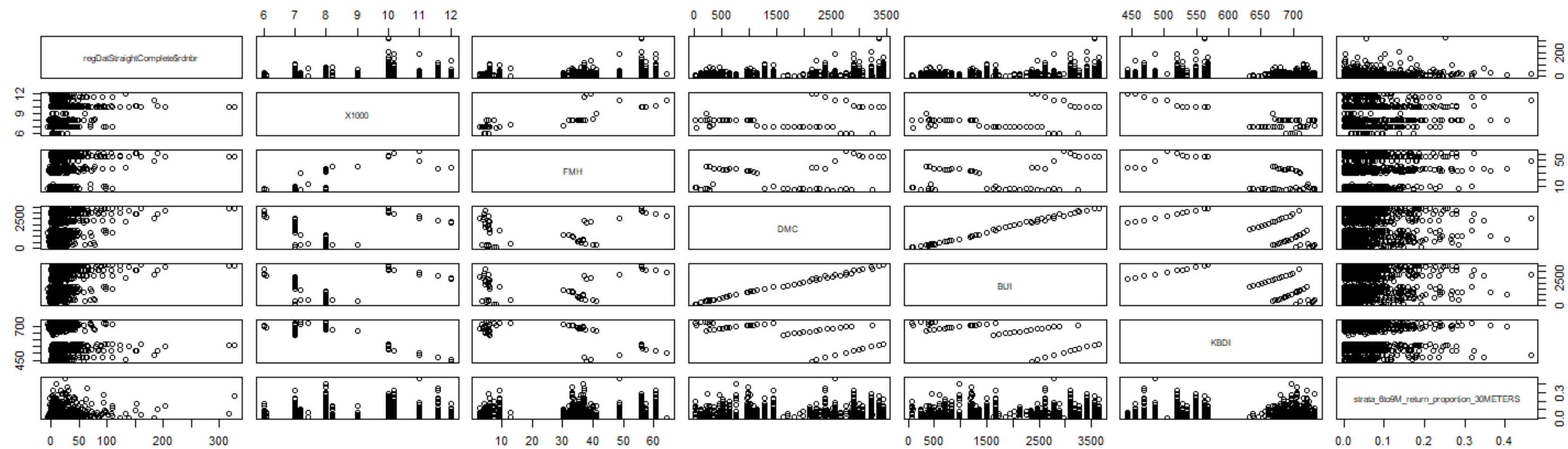

Figure 2.2 Pairs plots of selected weather and LiDAR variables.

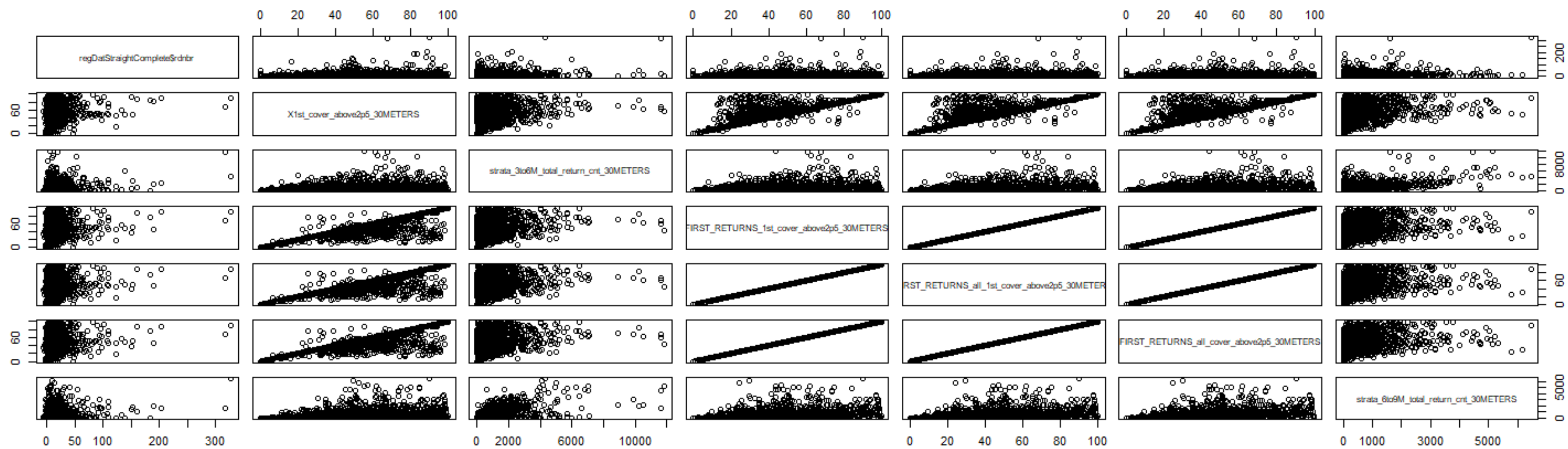

Figure 2.3 Pairs plots of selected vegetation LiDAR variables.

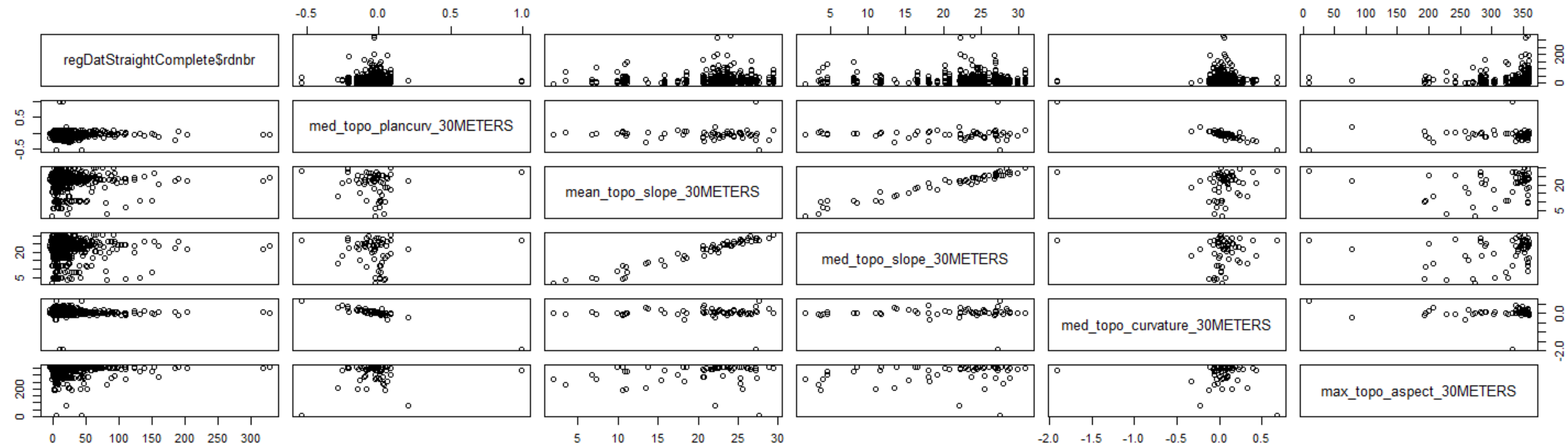

Figure 2.4 Pairs plots of selected topography variables.

Table 3.1 Descriptive metrics for models developed from three sets of principal components, containing all vegetation, all topography, or all weather variables.

Model Description	RMSE	RMSPE	AIC	BIC	Bias	Adjusted R^2
35 vegetation, topography, and weather PCs	2.45	37.18	5942.79	6133.16	0.0000	0.9892
25 weather PCs	3.03	37.21	6460.18	6599.10	0.0000	0.9836
30 weather PCs	1.21	37.27	4146.82	4311.47	0.0000	0.9974

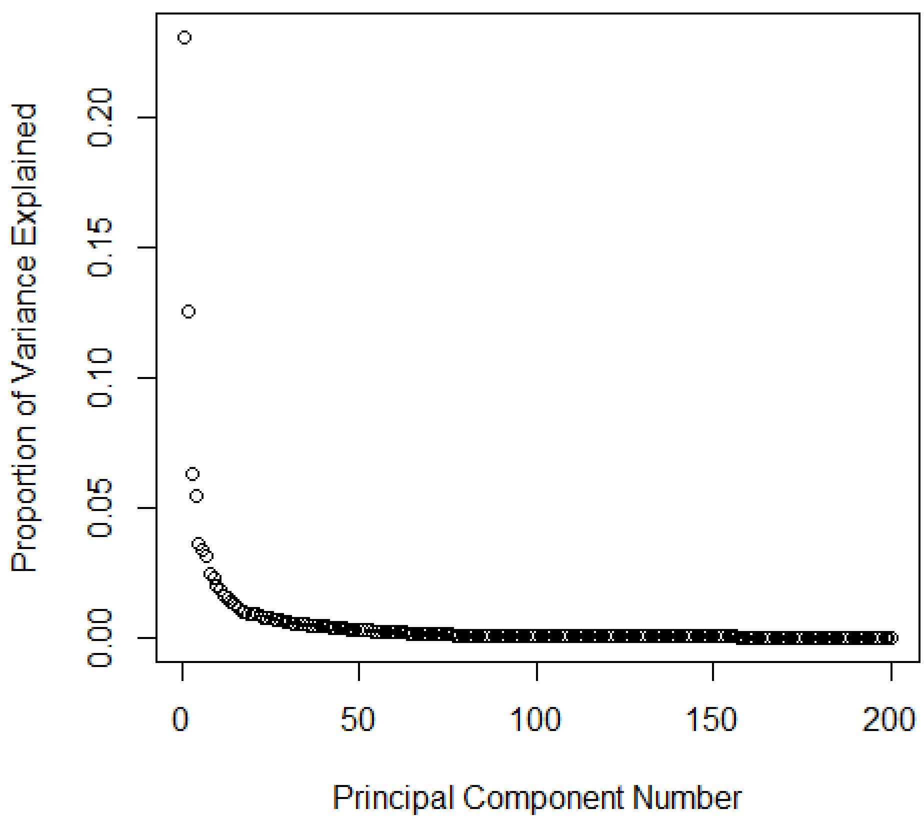

Figure 3.1 Proportion of variance explained by each of the first 200 scaled principal components when using only vegetation-related LiDAR variables.

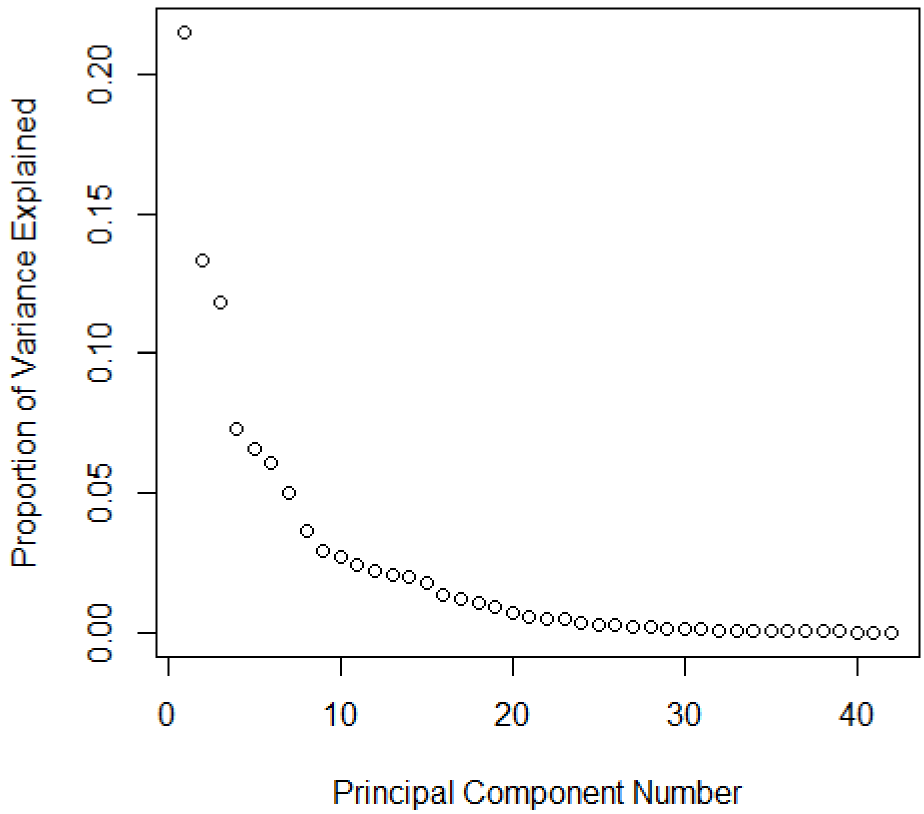

Figure 3.2 Proportion of variance explained by each of the scaled principal components when using only topography-related LiDAR variables.

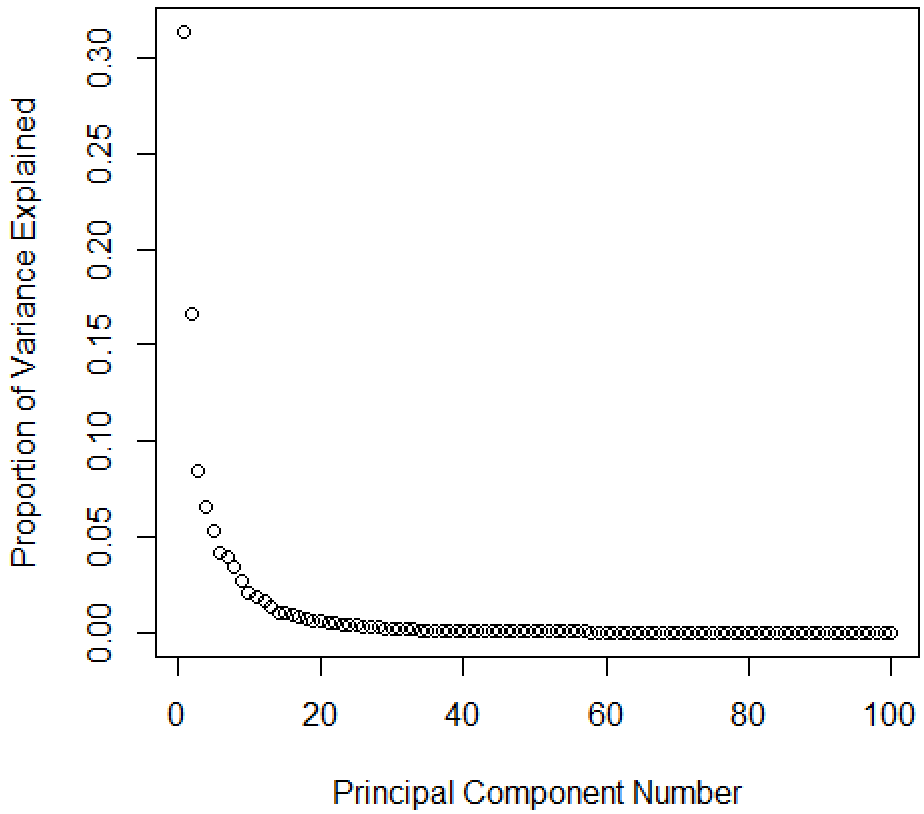

Figure 3.3 Proportion of variance explained by each of the first 100 scaled principal components when using only weather-related variables.

Table 3.2 First 10 contributing variables to the first principal components for three separate groups of variables: vegetation LiDAR, topography, and weather.

Vegetative LiDAR Variables	
Variable Name	Coefficient
max_FIRST_RETURNS_int_variance_above2p5	0.7744
max_int_variance_above2p5	0.5694
sd_FIRST_RETURNS_int_variance_above2p5	0.1273
sd_int_variance_above2p5	0.1117
mean_FIRST_RETURNS_int_variance_above2p5	0.1096
mean_int_variance_above2p5	0.1008
med_FIRST_RETURNS_int_variance_above2p5	0.0965
med_int_variance_above2p5	0.0956
FIRST_RETURNS_int_variance_above2p5	0.0617
int_variance_above2p5	0.046
Topography Variables	
Variable Name	Coefficient
min_elevation	0.6718
med_elevation	0.4978
mean_elevation	0.4581
elevation	0.2628
sd_elevation	0.0999
max_elevation	0.097
max_aspect	0.0325
min_aspect	0.032
sd_aspect	0.0117
max_slope	0.0095
WeatherVariables	
Variable Name	Coefficient
max_ISI	0.4066
mean_ISI	0.4062
med_ISI	0.4062
min_ISI	0.4059
ISI	0.2532
mean_BUI	0.1733
max_BUI	0.1733
med_BUI	0.1733
min_BUI	0.1732
max_DMC	0.171