

My Final College Paper

A Thesis
Presented to
The Division of Mathematics and Natural Sciences
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Ingrid Tallulah Loraine Zoll

May 2022

Approved for the Division
(Biology)

Aaron Ramirez

Anna Ritz

Acknowledgements

I want to thank some people...

Table of Contents

List of Abbreviations	1
Chapter 1: Introduction	3
1.1 The Accessability of Science	3
1.2 Urban Forests and Urbanization	3
1.3 The Benefits of Urban Forests	3
1.4 Environmental Challenges for Urban Trees	6
1.5 Urban Forestry in Portland, Oregon	7
1.6 Remote Sensing	8
1.7 Previous work	9
1.8 This Thesis	10
Chapter 2: Data Resources and Methods	13
2.1 Overview	13
2.2 Previously Collected Field Data	14
2.3 Portland Tree Inventory	15
2.4 Canopy width and tree height model	17
2.5 Satellite Imagery	17
2.6 NDVI calculation	19
2.7 Canopy Height Model	19
2.8 Tree Crown Delineation and Pixel Selection	20
2.8.1 Point method	21
2.8.2 Radius method	22
2.8.3 LiDAR method	23
2.9 Modeling Tree Health	26
Chapter 3: Results	29
3.1 Canopy Width and Tree Height Model	29

3.2 Point Method	31
3.3 Radius Method	31
3.4 LiDAR Method	32
3.5 Method Comparison	33
3.6 Predictive Model	37
Chapter 4: Discussion	43
Appendix A: Additional Data	45
A.1 Tree Height and Crown Width Predictive Model Equations	45
extra modeling graphics	47
Radius and LiDAR model tests	49
Appendix B: Code	51
B.1 In Chapter 2:	51
Portland Tree Inventory counts and calculations	51
Calculating NDVI from satellite imagery	52
LiDAR canopy delineation	53
B.2 In Chapter 3:	54
Model for Tree Height and Canopy Width	54
Point Method Modeling	55
Radius Method Modeling	55
LiDAR Method Modeling	56
Radius and LiDAR Replications	56
References	59

List of Tables

2.1	Final CNH Tree counts	15
2.2	Species counts in Street Trees Database	16
2.3	Species counts in Park Trees Database	16
2.4	Satellite product specifications	18
2.5	4-band satellite wavelength ranges	18

List of Figures

2.1	Sampled CNH Trees	15
2.2	Distribution of Health Ratings in the Portland Tree Inventories	17
2.3	Product areas for 2016, 2019, and 2021 satellite scenes	19
2.4	Berkeley Park and CNH trees	21
2.5	CNH trees sampled in Berkeley Park	22
2.6	CNH trees with crown width buffer	23
2.7	LiDAR data clipped to 30m tree buffer	24
2.8	All inventoried trees within the 30m buffer	24
2.9	LiDAR tree crown delineations with selected CNH crowns	25
2.10	Geographic extent and random sampling	25
2.11	LiDAR tree crown delineations with selected CNH crowns	26
3.1	Tree height predictive model	30
3.2	Crown width predictive model	30
3.3	Point Method NDVI and Health Rating, by Species	31
3.4	Radius method mean NDVI and health condition by species	32
3.5	NDVI and health condition for LiDAR method	33
3.6	NDVI and health rating comparison across methods	34
3.7	Average NDVI comparison between species and methods.	35
3.8	Confusion matrixes for Radius data model	39

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here. TESTING TESTING TESTING

Dedication

DEDICATION TESTING

test

List of Abbreviations

ACMA	Bigleaf Maple (<i>Acer macrophyllum</i> , <i>Sapindaceae</i>)
ACPL	Norway Maple (<i>Acer platanoides</i> , <i>Sapindaceae</i>)
CHM	Canopy Height Model
CNH	Coupled Natural-Human (in reference to the CNH2 Project data)
DBH	Diamater at Breast Height
LiDAR	Light Detection and Ranging
NDVI	Normalized Difference Vegetation Index
PSME	Douglas Fir (<i>Pseudotsuga menziesii</i> , <i>Pinaceae</i>)
THPL	Western Redcedar (<i>Thuja plicata</i> , <i>Cupressaceae</i>)

Chapter 1

Introduction

1.1 The Accessability of Science

1.2 Urban Forests and Urbanization

An urban forest is the total population of trees in an urban area. Urban forests are comprised of parks, street trees, landscaped boulevards, green spaces, and any other location where trees can be found in urban spaces. Urban forests are in close proximity to large or dense human populations, have a relatively high diversity of species and forest patch structures as well as both public and private ownership, and their management is often geared toward sustaining tree health and maximizing the potential benefits that trees provide (Robertson & Mason, 2016). In 2011, urban forests in the United States contained around 74.4 billion trees, which is about a quarter of the total tree population (USDA Forest Service, 2011). The US Census bureau reports that, in 2010, nearly 81% of Americans lived in urban areas, up from 79% 10 years earlier . The United Nations predicts that, by 2050, 68% of the world population will live in urban areas (United Nations & Affairs, 2018). As urbanization continues, it becomes increasingly important for both those working to manage urban forests and residents of urban areas to understand the dynamics and health of urban forests in order to retain and protect the numerous benefits they provide.

1.3 The Benefits of Urban Forests

Urban forests and urban trees have numerous benefits, which range from environmental to economic. The environmental benefits of urban trees include numerous forms

of pollution removal from both water and air. In undeveloped areas, most of the precipitated water is absorbed into the earth. However, due to the high amount of impervious surfaces such as sidewalks, streets, and parking lots in urban areas, rain and snowmelt are unable to soak back into the earth and become stormwater runoff instead. This runoff flows over developed impervious surfaces and picks up trash, yard waste, dirt, and many other potentially harmful chemicals and pollutants. It is then deposited in streams, rivers, wetlands, and other bodies of water that are damaged by polluted runoff. Green infrastructure like urban trees help in reducing the volume and rate of runoff by allowing more precipitation to be soaked into the earth. Research has shown that the presence of street trees also has a positive impact in reducing stormwater runoff volume (US EPA, 2020). A 2021 study conducted in Fond du Lac, Wisconsin, showed that the removal of street trees increased the volume of stormwater runoff by 4% (Selbig et al., 2021). The study calculated that on a per-tree basis for each square meter of canopy that was removed, 66 liters of rainfall could have been intercepted and stored by the street trees. This results in an annual runoff volume reduction estimated at 6,376 liters per tree.

Urban trees also remove pollutants from the air. Numerous studies have shown that trees can remove many different pollutants (O₃, PM10, NO₂, SO₂, CO) by uptake via leaf stomata. Pollutant particles can also be collected and stored on a tree's surface. Urban trees provide a total annual air pollution removal of 711,000 tons, which is valued at \$3.8 billion (David J. Nowak, Crane, & Stevens, 2006). Cities with higher levels of tree canopy cover have higher rates of pollution removal by trees, and longer on-leaf growing periods of trees lead to more pollution removal as well. While the removal of air pollution by urban trees results in the improvement of air quality, trees also help mitigate climate change, improve atmospheric conditions and air quality through carbon sequestration.

The increase of atmospheric carbon dioxide from human sources is one of the primary drivers of global climate change. In 2019, U.S. greenhouse gas emissions totaled 6,558 million metric tons of carbon dioxide equivalents. In the same year, the city of Portland's carbon emissions totaled around 55,000 metric tons of carbon dioxide equivalents (City of Portland, 2020). Rural and urban forests, as well as other natural and nature based carbon sinks have been suggested as a method of mitigating greenhouse gas emissions of cities in order to reduce the impacts of global climate change (Lazarus, Chandler, & Erickson, 2013). These natural carbon sinks capture atmospheric carbon dioxide during photosynthesis and store the carbon as biomass, releasing oxygen back into the atmosphere.

Multiple scholars estimate that urban trees in the United States currently store over 708 million tons of carbon, and capture another 28.2 million tons of carbon per year, which is approximately 0.05% of annual carbon dioxide emissions in the United States (David J. Nowak & Crane, 2002; David J. Nowak, Greenfield, Hoehn, & Lapoint, 2013; Safford, Larry, McPherson, Nowak, & Westphal, 2013). The current carbon storage of urban trees is valued at more than \$50 billion, with carbon sequestration valued at an additional \$2 billion per year (David J. Nowak et al., 2013). The environmental benefits of urban trees can result in economic advantages under future carbon trading schemas, but there are other unique economic and social benefits that urban trees provide.

Residents of urban areas experience the benefits of urban trees most immediately through the beauty and visual stimulation they provide. Aesthetics alone are a large driver in the plantings of urban trees. Studies have found that trees are one of the main contributors to positive visual aesthetic quality of residential areas, and that large trees contribute more to perceived beauty than smaller trees (Herbert W. Schroeder, 2011; Herbert W. Schroeder & Cannon, 1987; H. Schroeder & Cannon, 1983).

The positive impacts of urban trees goes far beyond their visual contributions. Numerous studies have shown that people living near urban forests live longer, experience better mental health, and self-rate their overall health higher than people who do not live near urban forests (James, Banay, Hart, & Laden, 2015). Residents of two different towns in Germany visited urban forests and green spaces more frequently after the beginning of the COVID-19 pandemic, which contributed significantly to the residents' well-being (Beckmann-Wübbelt et al., 2021). Research into the psychological benefits of urban trees shows that teenage girls who spend more time around trees and other sources of nature and vegetation have higher levels of self-discipline, and children with diagnosed attention deficit disorder show improved focus and ability to learn after spending time outside (Taylor, Kuo, & Sullivan, 2002; Taylor, Kuo, & Sullivan, 2001).

Additionally, urban trees provide numerous economic benefits on both a nation wide scale, and an individual scale. Urban tree canopy cover positively impacts residential property values. In Athens, Georgia, landscaping with trees results in a 3.5%–4.5% increase in home sale price (Anderson & Cordell, 1988). In Ramsey and Dakota Counties, Minnesota, researchers found that a 10% increase in tree cover within 100m of a house increases the average home price by 0.5% (Sander, Polasky, & Haight, 2010). Both summer cooling and winter heating demands can be reduced

through shading and wind speed reduction by urban trees, which lowers energy costs (David J. Nowak et al., 2010). A 2009 study of 460 single-family homes in Sacramento, California showed that tree cover on the south and west sides of houses reduced summer electricity use by 5.2%, whereas trees on the north side of a house increased electricity use by 1.5% (Donovan & Butry, 2009). A more recent study conducted in the city of Thessaloniki in northern Greece found that the cooling potential of street trees is directly related to the foliage density and the shade provided can lower temperatures up to 5 degrees C, leading to energy savings of up to 54% (Tsoka, Leduc, & Rodler, 2021).

1.4 Environmental Challenges for Urban Trees

Human activities have impacted and altered the Earth's climate and land surface at a fundamental level. These changes have led to the creation of a new geologic epoch, known as the Anthropocene (Huang, Anderegg, & Asner, 2019). The Anthropocene has seen elevated temperatures and prolonged droughts. These extreme droughts can trigger extensive forest die-off as well as increased tree and shrub mortality rates, which has impacted forests and woodlands on all vegetated continents. Remote sensing research on the impacts of droughts has shown that the impacts of a drought can produce a suppression of forest canopy greenness, which relates to a failing of plant ecophysiological processes and a reduction of chlorophyll. Droughts are predicted to occur more frequently and have higher intensities as we progress through the Anthropocene.

The impacts of climate change and the environmental challenges that all trees will face is even more extreme in urban areas. The multiple economic, environmental and quality of life, health benefits that are provided by urban trees are dependent on tree health. Urban trees are generally more stressed than those in rural areas due to the adverse growing conditions they face. This includes higher temperatures, additional soil compaction, root zone restrictions, and extreme variations in environmental conditions such as wind speed and sunlight level (Flint, 1985; Ward & Johnson, 2007). Monitoring and tracking tree health over time is an essential component for the ability to model and predict the future changes that will occur. Research into the remote sensing of health indicators of forest die-off has shown that vegetation greenness metrics, such as NDVI, reflect the changes that occur with tree die-off (Breshears et al., 2005; Byer & Jin, 2017). In order to work towards understanding these changes and how to mitigate the impacts of forest die-off, it is extremely important to understand

the current health dynamics of urban forests at a city-wide level, and carefully track changes in urban forest health over time.

1.5 Urban Forestry in Portland, Oregon

Portland is one of many cities to create tree inventories in the last 25 years, with the goal of better understanding the urban forest. Over a period of 9 years, 2,000 volunteers along with members of Portland's Urban Forestry team inventoried Portland's 245,000 park and street trees (DiSalvo, Fukuda, Ramsey, & Parks, 2017; Portland Urban Forestry, 2019). In addition to basic physical and environmental variables such as tree height, diameter, and location, volunteers also visually assessed the health condition of each tree, and categorized it as good, fair, poor, or dead. Portland inventoried the 218,602 street trees between 2010 and 2016, and 25,740 park trees between 2017 and 2019. Portland Urban Forestry estimates that Portland's parks contain upwards of 1.2 million trees, but the tree inventory project only inventoried trees in developed portions of parks. Portland Urban Forestry estimates that Portland's street trees produce an estimated \$28.6 million annually in environmental and aesthetic benefits, with a full replacement value of \$753 million (DiSalvo et al., 2017). Portland's inventoried park trees have an estimated worth of \$128 million (Portland Urban Forestry, 2019). While Portland's tree inventories are spectacular resources which help shed light on the urban tree population in Portland and the benefits it can provide, there are limitations to the information available through the inventories.

 One drawback of the inventories is that volunteers were only able to inventory trees on public land, which excluded any trees growing in yards or other privately owned areas. Since only developed portions of parks were inventoried, roughly 98% of Portland's park trees were not inventoried. Additionally, since the inventories were collected over a 9 year period, the collected health assessments for many trees are no longer reliable or representative of present day conditions. They represent data from a range of 9 years and in the 11 years since inventory collection began much has changed.

Measuring urban tree health through field surveys can be extremely time and labor intensive. It requires the collection of detailed data on numerous environmental variables, as well as extensive groundwork to conduct the field surveys. To collect Portland's tree inventories, more than 2,000 volunteers collectively spend upwards of 25,000 hours in the field. Additionally, in order to get a good picture of changes in tree health over time, trees need to be revisited numerous times over the study

period. Remote sensing data from satellite imagery can be used to locate and map trees in both urban and rural areas, as well as monitor tree health.

1.6 Remote Sensing

While the term “remote sensing” was first used in the 1960s, the first aerial images were taken in the 1850s from hot air balloons. Later, small cameras were attached to kites and even pigeons to capture aerial images. With the development of airplanes in the early 20th century, images were able to be taken from higher altitudes, providing aerial views of larger surface areas (Moore, 1979). Aerial images taken from planes provided essential military reconnaissance during both World War I and II. The first environmental applications of aerial imagery began in the 1930s, when the Agriculture Department began to use aerial photography to map and catalog farmland in the United States. Soon after, aerial imagery became a tool for conservation and land planning purposes. Capturing aerial images from planes was the primary method of capturing images of the earth’s surface until the early 1960s. Since the first satellite was launched in 1957, the technical capabilities of satellites has greatly increased, along with the types and applications of satellite-collected data (Khorram, 2012).

Within the field of forestry and ecology, remote sensing has numerous applications, from measuring the cover and structure of vegetation, to examining biodiversity and soil characteristics of specific areas. Additionally, remote sensing measurements can be used to calculate and monitor changes in forest density, which is critical for determining the fuel load and forest health in regards to fire risk. One of the most commonly used remote sensing metrics used to measure forest health is the Normalized Difference Vegetation Index (NDVI), which is calculated from the red and near infrared (NIR) bands from remote sensing imagery. Vegetation that is photosynthetically active absorbs most of the red light and reflects much of the near infrared light. Conversely, vegetation that is dead or stressed reflects more red light and absorbs more near infrared light. Biologically, NDVI can be interpreted as the fraction of absorbed photosynthetically active radiation. A value closer to 1 indicates vegetation that is more photosynthetically active and greener, which can be used as a proxy for vegetation health. NDVI is not a diagnostic tool of vegetation health, but can be used as an indicator of health for further analysis. Studies have shown that NDVI is highly correlated with chlorophyll content. Specifically, NDVI has a near-linear relationship to the chlorophyll content of soybean crops (Myneni, Hall, Sellers, & Marshak, 1995; Tucker, 1979).

Especially when the goal is monitoring tree health changes over time, remote sensing data eliminates the need for repeated sampling over long time periods, since satellite images are taken at regular intervals as the satellites continually orbit the earth. With the availability and accuracy of aerial imagery increasing as these technologies continue to advance, remote sensing is becoming an important and effective method for mapping, monitoring, and analyzing tree health on an individual tree scale (Xiao & McPherson, 2005).

1.7 Previous work

Xiao & McPherson (2005) used multispectral remote sensing data paired with field collected tree health data with the goal of mapping tree health on the University of California Davis campus. Field data on 81 campus trees was collected in the summer of 2004, and the health of the trees was classified as “healthy” or “unhealthy.” Additionally, a second dataset of 1,186 trees was collected which included randomly selected trees to check the accuracy of the resulting tree health mapping. With high resolution multispectral remote sensing data collected in the summers of 2003 and 2004, NDVI was calculated and used to classify each pixel as vegetation or non-vegetation. The pixels representing trees and shrubs were manually selected and extracted, resulting in NDVI data just for trees and shrubs. The remaining data was split into 5 separate layers based on physiognomic tree type (broadleaf deciduous, broadleaf evergreen, conifer, palm, and mixed). Tree health was evaluated at both a pixel scale, and a tree scale, with the pixels or trees being mapped as either healthy or unhealthy. A tree was labeled as unhealthy if 30% or more of the pixels within the manually delineated tree crown were mapped as unhealthy, and if the average NDVI of the pixels were less than the NDVI threshold for healthy trees. The accuracy of the tree health assessment was checked against the validation dataset of 1,186 trees. The field health assessment agreed with the remotely sensed health classification for 88% of the trees.

Fang, McNeil, Warner, Dahle, & Eutsler (2020) used a similar approach to evaluate the health of street trees in Washington D.C. using multispectral remote sensing data and D.C.’s street tree inventory. The tree inventory contained 18,434 trees, each with a tree health classification of excellent, good, fair, poor, or dead. The researchers purchased remote sensing images for June 11, July 30, and August 30, 2017, to compare the sensitivity of tree health at different points in the trees’ on-leaf period. To extract pixels belonging to tree crowns, a radial buffer based on tree

crown diameter was used, and pixels with low NDVI values were masked. This paper tested 5 different vegetation indices (VIs), which included three different variations of NDVI. The different VIs were calculated for each pixel and averaged for each tree. They found that the VI values of trees in good, fair, and poor health conditions were highly statistically different, and traditional NDVI was the most sensitive VI for detecting tree health conditions. Additionally, it was determined that remote sensing imagery taken in the middle of the on-leaf period had the best potential to assess the health condition of trees.

These two studies form the basis of the approach and methods for this thesis, but there are two main places where I believe I can add new insights and improve methodology. First, a frustrating drawback of both Fang et al. (2020) and Xiao & McPherson (2005) is that a true replication of their process is inaccessible due to the sources of their data. Xiao & McPherson (2005) used multispectral data that was specially collected just for the UC Davis campus, and Fang et al. (2020) purchased the high resolution multispectral data that was used in their study. Second, there is little consistency in the methods used to select image pixels for NDVI evaluation. Xiao & McPherson (2005) manually detected and delineated tree crowns for health assessment, and any trees with overlapping crowns were removed from the analysis. Manual crown delineation is extremely time intensive and is an unrealistic method for large sample sizes. In Fang et al. (2020), a standardized radius based on the average DBH of all trees was used to select the tree crown area. However, this method will only use the center pixels of large trees eliminating the edges, and it is unclear how overlapping tree crowns were dealt with. The impact of these different methods on health analysis is unknown.

1.8 This Thesis

Based on the assumptions that tree health can be approximated using satellite spectral data, specifically NDVI, this thesis aims to replicate the process of Fang et al. (2020) and Xiao & McPherson (2005). Additionally, I will investigate the questions of the impact of species differentiation when using NDVI to calculate tree health, and if different methods of tree crown delineation impact the health rating outcome. Based on these questions, I predict that differentiating health rating predictions by tree species will improve the accuracy of the model, and differentiating by tree type (coniferous evergreen vs broadleaf deciduous trees) will also improve the accuracy, but less than species. Additionally, I predict that using more in-depth tree crown

delineation approaches will allow for more trees to be analyzed, and will also improve the accuracy of the health rating.

Finally, this thesis only uses data sources, tools, and processes that are either publicly accessible or free, in order to keep this type of research accessible to all.

Chapter 2

Data Resources and Methods

2.1 Overview

For my thesis, the data resources and processing components can be split into four main sections.

First, there is the tree data. This includes the Portland park and street tree databases, and the subset of those trees that I sampled health data on prior to this thesis. Primarily, the tree data is used for tree location points, as well as basic tree metrics such as species, height, and crown width. In general, I will refer to these data sources as park trees (the Portland park tree dataset), street trees (the Portland street tree dataset), and CNH trees, which are the trees we collected data on in the summer of 2021. A majority of the processing with the tree points involved tidying and wrangling the data, subsetting it and transforming it into usable data products. Additionally, due to the inconsistency in the types of variables measured for each tree data product, I created a model to predict tree crown width and tree height based on DBH and species.

The second aspect of my data and processing is the retrieval and processing of Planetscope satellite imagery products to produce usable NDVI data files, which was done in Python. Third, are the different tree delineation methods I examine in this thesis, which include point value, radius, and LiDAR methods. This involved the processing of LiDAR and tree crown radius data, as well as obtaining and computing associated NDVI values.

Lastly, with the combined data from the previous three parts, I created a predictive model for predicting tree health rating based on NDVI values.

2.2 Previously Collected Field Data

We collected data tree health during the summer of 2021 as part of a larger collaborative interdisciplinary project examining the relationship between various socioeconomic variables and urban tree health. For the purposes of the larger project, we selected trees from eight different neighborhoods in Portland, with four different categorizations of historic and current investment or disinvestment. The four chosen species are *Pseudotsuga menziesii*, *Pinaceae* (PSME), *Thuja plicata*, *Cupressaceae* (THPL), *Acer macrophyllum*, *Sapindaceae* (ACMA), and *Acer platanoides*, *Sapindaceae* (ACPL). They are some of the most abundant tree species in Portland, and make up a large proportion of Portland's urban forest. PSME, THPL, and ACMA are all native to the area, whereas Norway Maple is a nonnative tree species that was frequently planted in residential areas, and is now the most common street tree in Portland.

Individual trees for sampling were selected from the Portland tree inventories, with the goal of sampling an equal proportion of street trees and park trees of each species in each neighborhood. Sampling was also focused on mature trees, so the inventories were filtered to only include individuals above 25 feet in height. 4 trees of each species were randomly selected per neighborhood, with an attempt to maintain equal proportions of park and street trees. However, due to varying field conditions, some trees were not able to be sampled, so the nearest tree of the same species that met all criteria was substituted. In total, 128 trees were surveyed. Fieldwork was conducted between July 7 and August 19, 2021 (Figure 2.1). Each tree was sampled in a single visit between 10am and 3pm. Biotic and abiotic variables were measured, as well as health attributes and physiology of the trees. For each tree, a GPS point was collected to mark its location using ArcGIS Explorer on IOS. Due to the uncertainty of the GPS points, during post-processing, each collected tree was matched with the tree location point in the Portland Park and Street Tree inventories. Any sampled tree points that did not match up with an inventoried tree point in a margin of 20 feet, it was removed from further analysis. Any tree individual that did not contain a health categorization was also removed. The final CNH dataset for my analysis contained 112 trees (Table 2.1).

Table 2.1: Founts of species for final CNH tree dataset

Species code	Number collected
ACMA	29
ACPL	30
PSME	25
THPL	28

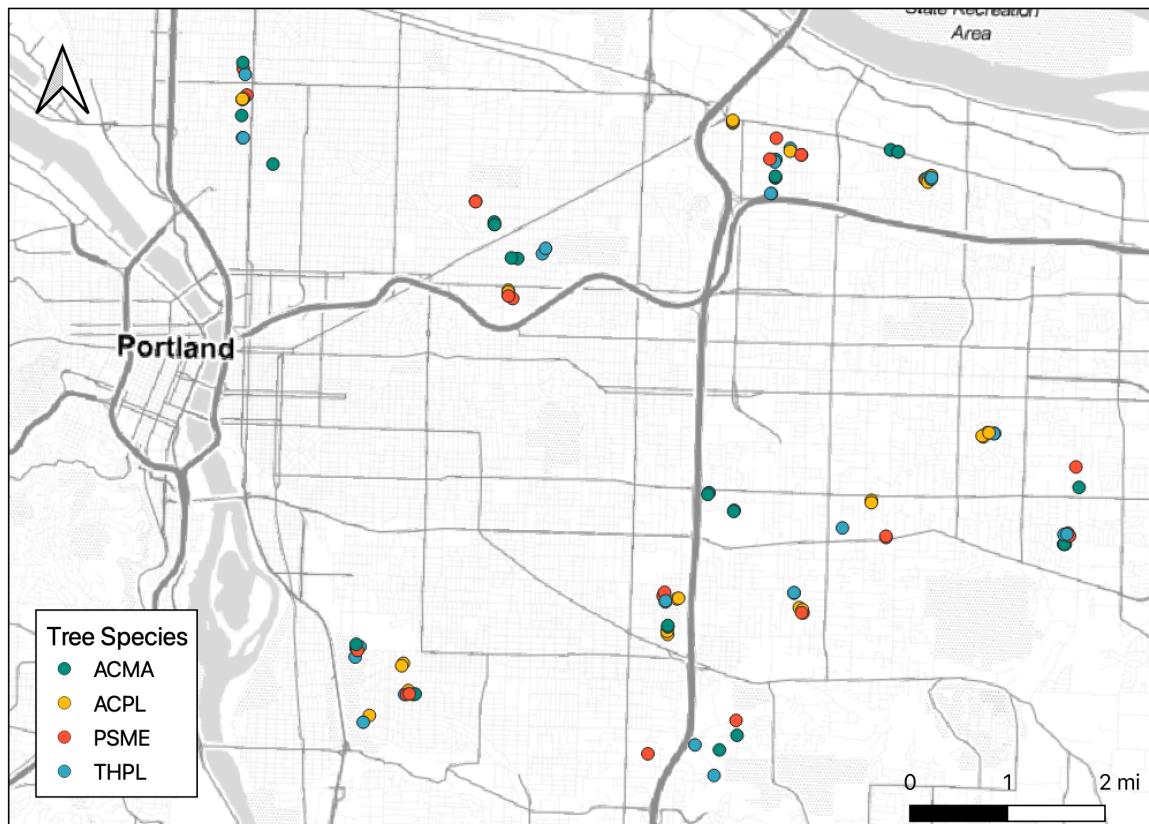


Figure 2.1: Sampled CNH Trees

2.3 Portland Tree Inventory

The Portland Tree Inventory Project, managed by the City of Portland's Parks & Recreation Urban Forestry Department, cataloged nearly 245,000 street and park trees in Portland between 2010 to 2019. The street tree inventory, which was collected from 2010 to 2016, contains information on 216,750 street trees of 145 genera. The park tree inventory, collected from 2017 to 2019, contains data on 25,740 park trees of 116 genera. While many of the collected variables differ between the two inventories, they both include data on location, tree identification, DBH, and a visual assessment

Table 2.2: Species counts in the Portland Street Trees Database

Species code	Scientific name	Collected in 2016	Total in inventory
ACPL	Acer platanoides	4373	19209
THPL	Thuja plicata	578	1341
ACMA	Acer macrophyllum	1306	2609
PSME	Pseudotsuga menziesii	1254	3141

Table 2.3: Species counts in the Portland Park Trees Database

Species code	Scientific name	Collected in 2019	Total in inventory
ACPL	Acer platanoides	411	1502
THPL	Thuja plicata	331	964
PSME	Pseudotsuga menziesii	3237	6783
ACMA	Acer macrophyllum	256	490

of the trees health which was rated as good, fair, poor, or dead.

To reduce the variability in data due to the time span of data collection, I filtered the street and park tree inventories to trees that were sampled in 2016 and 2019, respectively. For each dataset, the selected year was the year with the highest count of trees sampled (Tables 2.2 and 2.3). The Portland park and street tree inventories include a health categorization variable of Good, Fair, Poor, or Dead. A drawback that can come with qualitative health ratings such as this one is that if a tree does not appear close to death, or perfectly healthy, it is very easy to categorize its health as “Fair”, which is what we see in both the street and park tree databases. The proportion of trees rated “Fair” is extremely high, especially in the park trees inventory (Figure 2.2).

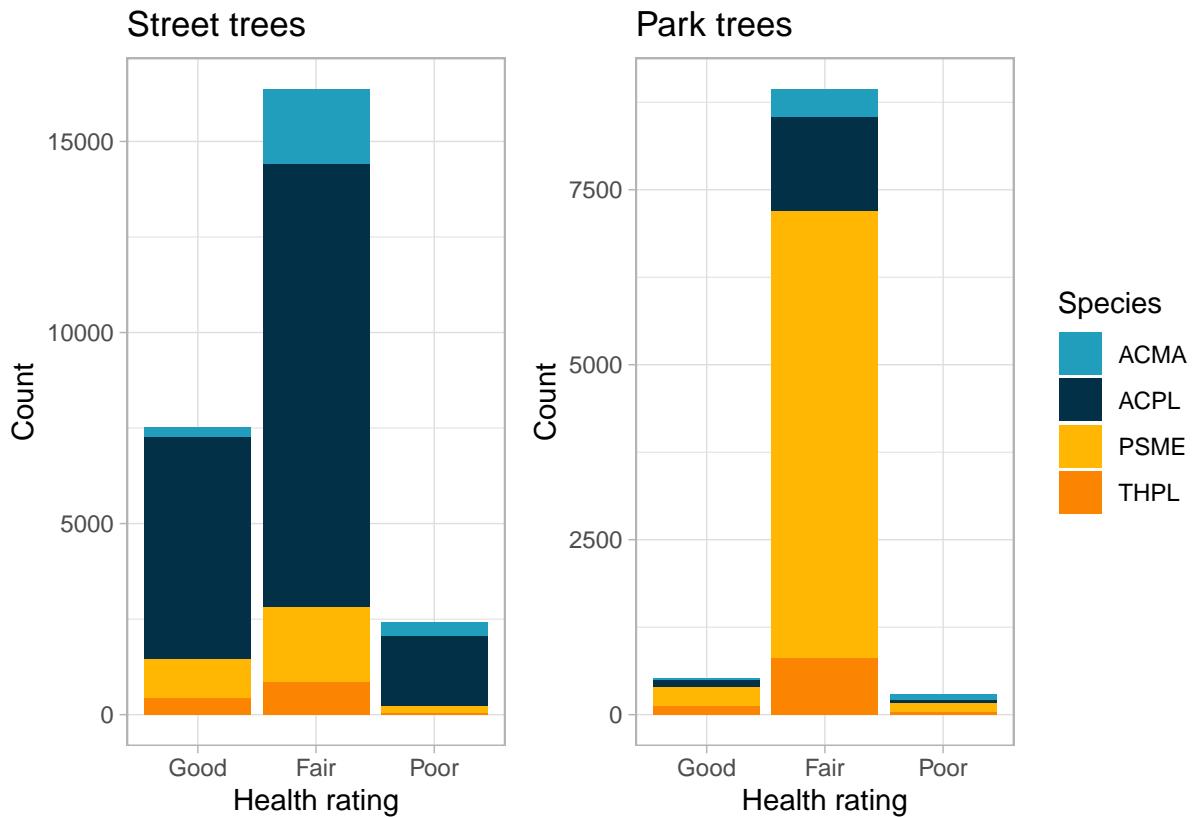


Figure 2.2: Distribution of Health Ratings in the Portland Tree Inventories

2.4 Canopy width and tree height model

The park tree dataset contains measurements for tree height, canopy width, and DBH, but the street tree dataset only contains measurements for DBH. Canopy width and tree height are both important factors for tree selection and canopy delineation. Using RStudio, I created a statistical model to predict tree height and crown width based on tree DBH and species, in order to be able to use the same pixel selection methods for both the park and street tree inventories.

2.5 Satellite Imagery

PlanetScope satellite imagery (Planet Labs, 3m resolution, 4-band RGB-NIR) was used for vegetation index calculation. PlanetScope, also known as the Flock, is a constellation of approximately 130 satellites . The first 28 PlanetScope satellites

were launched in July 2014, and the newest batch of satellites were launched in January 2022. PlanetScope, operated by Planet, is a constellation of approximately 130 satellites, able to image the entire land surface of the Earth every day (a daily collection capacity of 200 million km²/day). PlanetScope images are approximately 3 meters per pixel resolution (“Planet imagery product specifications,” 2022).

I downloaded satellite images corresponding to the summers of when the data was collected (2016, 2019, and 2021). Images were first filtered for those captured in July of each year, since images taken during the middle of the on-leaf period have been shown to be most sensitive in detecting a statistical difference in tree health measured by NDVI (Fang et al., 2020). The remaining images were then filtered for minimal cloud cover and maximum coverage of the area of study. The final selected images were taken between 6:30 and 7pm on 6 July 2016, 31 July 2019, and 26 July 2021 (Table 2.4).

Table 2.4: Satellite product specifications

Collection date	Collection time	Number of scene products	Satellite ID
2016-07-06	18:55	5	0c22
2019-07-31	18:42	4	0f42
2021-07-26	18:40	4	1003

For each of the chosen dates, I downloaded the available analytic product from Planet. The analytic multispectral imagery products are orthorectified, calibrated, corrected for terrain distortions, and transformed to Top of Atmosphere radiance to ensure accurate geolocation and cartographic projection.

The bands in a saptellite product refer to the reflectance wavelength that is picked up by the satellite. For a 4-band satellite product, the wavelengths are split into red, green, blue, and near-infrared (NIR) bands (Table 2.5). In the raw PlanetScope product, these 4 bands come compressed in one .tif image. In order to extract NDVI data from the imagery products, the data requires further processing.

Table 2.5: Wavelength ranges for 4-band Planetscope satellite bands

Band name	Wavelength range
Blue	455 - 515 nm
Green	500 - 590 nm
Red	590 - 670 nm

NIR 780 - 860 nm

2.6 NDVI calculation

The NDVI calculations were done using Python in PyCharm. Besides minor alterations, the processing code came from Planet's instructions on calculating NDVI from 4-band PlanetScope data (Planet, n.d.). NDVI is calculated from the red and NIR satellite bands, as shown in Equation (2.1).

$$\text{NDVI} = \frac{(NIR - Red)}{(NIR + Red)} \quad (2.1)$$

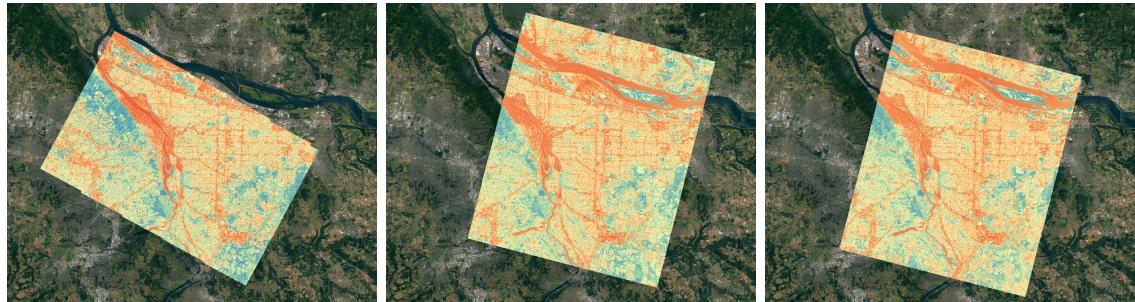


Figure 2.3: Product areas for 2016, 2019, and 2021 satellite scenes

2.7 Canopy Height Model

A canopy height model for the Portland metro area was developed using LiDAR and satellite spectral imagery collected in the summer of 2014 (“Canopy 2014,” 2016). The purpose of this data is to monitor natural areas in the Portland metro area, specifically change over time analysis and the examination of the potential loss of habitat in riparian areas. The canopy was detected using both NDVI values and LiDAR feature heights. Errors and noise in the data, such as electrical lines above tree tops, were cleaned using geometric post-processing. The canopy height model was clipped to remove anything below ten feet to eliminate any understory shrubs or grasses that were included in the raw data (“Canopy 2014,” 2016).

2.8 Tree Crown Delineation and Pixel Selection

The goal of tree crown delineation is to understand where the foliage of a tree is located. This is extremely important because we want to ensure that the satellite pixels used for health analysis have measurements that belong to the given tree. Previous papers such as have used manual tree crown delineation (Xiao & McPherson, 2005), or chosen a standardized radius for all trees in their sample (Fang et al., 2020). Manual tree crown delineation would be extremely time consuming, especially when trying to examine data on a city-wide level. Additionally, with a standardized radius, there will be many trees that have crowns either larger or smaller than the standardized radius. If the true crown is smaller than that of the radius, pixels that correspond to things like grass or pavement will be included in the measurement analysis. Conversely, if the true crown is larger than the chosen radius, the edges of the tree will be ignored, and valuable data will be lost. With both of these methods, any overlapping tree crowns were removed from the final analysis, even further reducing the sample size.

In this thesis, I test three different methods of pixel selection for NDVI analysis. First, I test a “point method” which uses the NDVI value directly below an inventoried trees location point. Second, I use crown width measurements and predictions to create a variable radius method, and average the NDVI pixels within the created circle. Lastly, I use a LiDAR created canopy height model for Portland’s urban canopy with `ForestTools` canopy delineation algorithm to create canopy polygons for NDVI analysis.

For the following examples, I use Berkeley Park, located in SE Portland, to demonstrate the different delineation techniques. We sampled 6 CNH trees in Berkeley Park: 2 Bigleaf Maple, 1 Norway Maple, 2 Douglas Fir, and 1 Western Redcedar (Figure 2.4).



Figure 2.4: Berkeley Park and CNH trees

2.8.1 Point method

Point Method: For the point method, I extracted the NDVI value from the pixel directly underneath the tree location point. This is the simplest of the three methods, since it obtains one NDVI value for one singular pixel. The tree location points were sourced from the street and park tree inventories and processed in QGIS (Figure 2.5). The main source of potential error with this technique is with the tree point location. If the point was placed incorrectly, the NDVI value will not be representative of the tree's greenness, or health.



Figure 2.5: CNH trees sampled in Berkeley Park

2.8.2 Radius method

The first method of tree crown delineation I used is based on the individual crown width for each selected tree. In the Portland park tree inventory, crown width is measured as an east to west diameter, and a north to south diameter. To get the radius for the buffer, I took the average of both measurements and divided it by 2. For the Portland street tree inventory, crown width was not collected. I used the tree height and crown width predictive model to create crown width measurements. For each selected tree point, I created a buffer with the radius of the measured or predicted tree canopy using QGIS (Figure 2.6). To get NDVI for each tree in this method, I averaged the value of all pixels in the buffer circle for each tree using QGIS.

With the radius method, it involves the same risk as the point method that if the central point is incorrectly located, the values will not be representative. Additionally, average NDVI ratings obtained with this method for trees with overlapping crowns will contain NDVI pixels that belong to other trees. Previous studies have removed all overlapping tree crowns when using a radius technique, but that can severely limit sample size of available data for analysis, since especially in urban areas and on streets, most of the trees will have overlapping crowns.

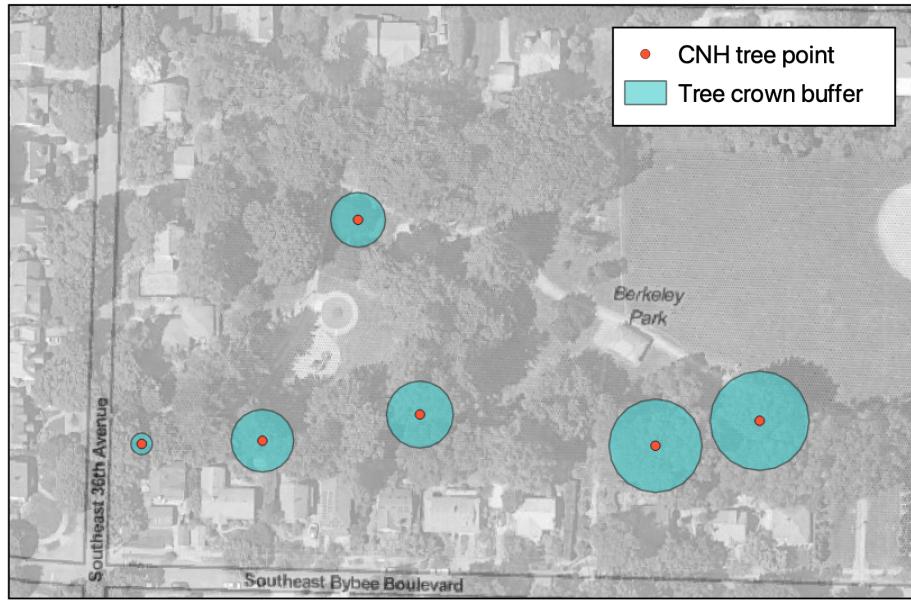


Figure 2.6: CNH trees with crown width buffer

2.8.3 LiDAR method

The LiDAR method is the most complex and involved of the three tree crown delineation methods but has the potential to be the most precise.

The **ForestTools** crown delineation algorithm is a modified watershed delineation algorithm that takes a LiDAR file input, as well as treetop location point file and outputs a polygon file with predicted tree crowns. **ForestTools** also has the ability to predict treetop location points based on LiDAR data, but that functionality introduced a lot of error in my processing because it became difficult to re-associate predicted treetop points with the actual tree location points that I was analyzing.

The LiDAR canopy height model was clipped to a 30 meter buffer around each selected tree point to reduce file size in the delineation processing (Figure 2.7). A 30m buffer was chosen to ensure that no part of the tree canopy would be omitted from the processing, and to include any near neighboring trees that may have canopy that overlaps with the tree of interest. In order for the algorithm to delineate tree crowns, it needs both LiDAR canopy data and location points for the potential trees for delineation. With a smaller buffer radius, the locations of neighboring trees would be omitted and the canopy delineation algorithm would compute a tree crown as much bigger than it actually is.

With the 30m buffer, all other inventoried trees that are contained within the buffer were selected for inclusion as treetop location points in addition to the trees of

interest (Figure 2.8). A minimum tree height of 20 feet was included in the algorithm options but since all sampled trees were already filtered for height requirements, this may not be necessary.

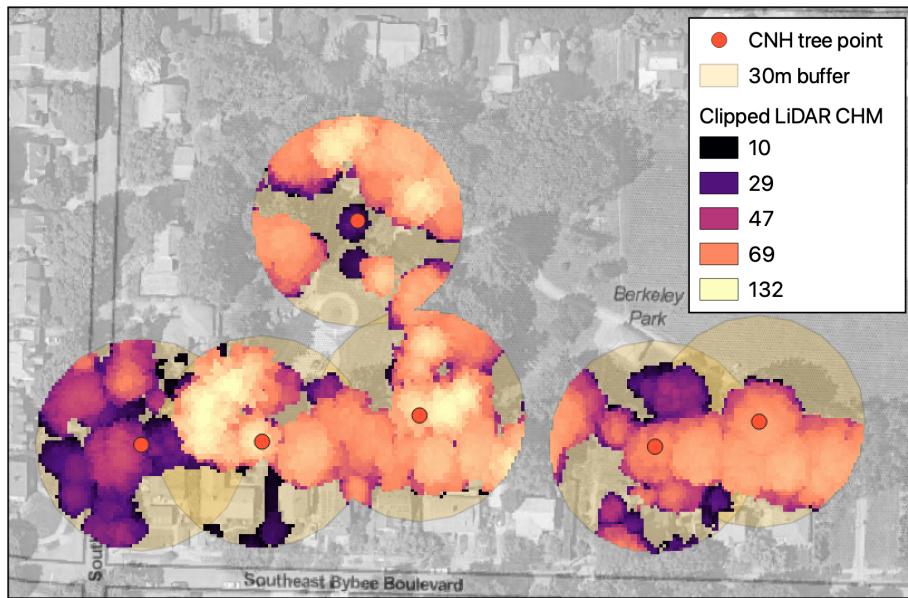


Figure 2.7: LiDAR data clipped to 30m tree buffer

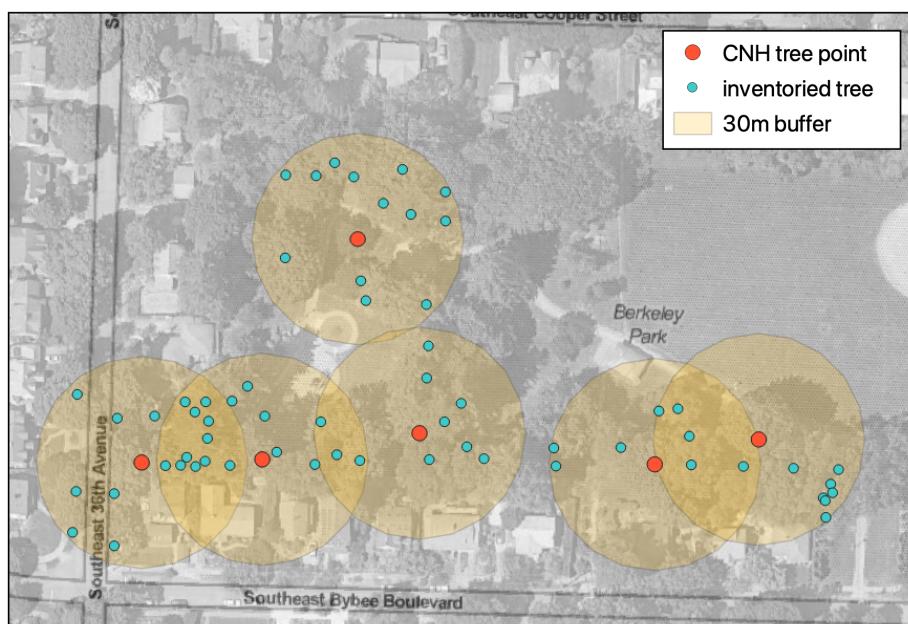


Figure 2.8: All inventoried trees within the 30m buffer

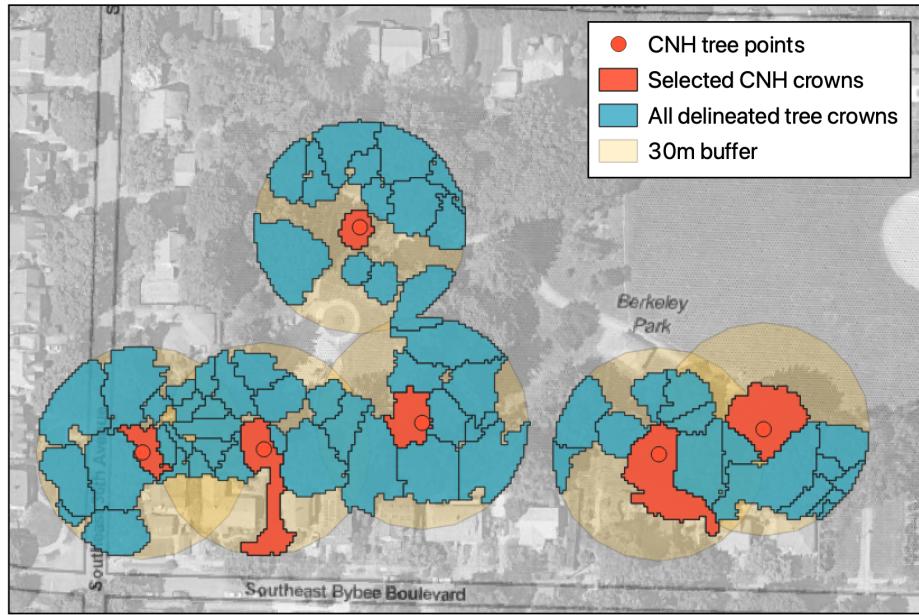


Figure 2.9: LiDAR tree crown delineations with selected CNH crowns

These three processes were repeated for random samples of 100 park trees and 100 street trees, though due to restrictions in file size for LiDAR processing, the sampling had to be constrained to an area of East Portland (Figure 2.10). This extent was chosen to align with the sampling region of the CNH data, though it had to be shrunk slightly.

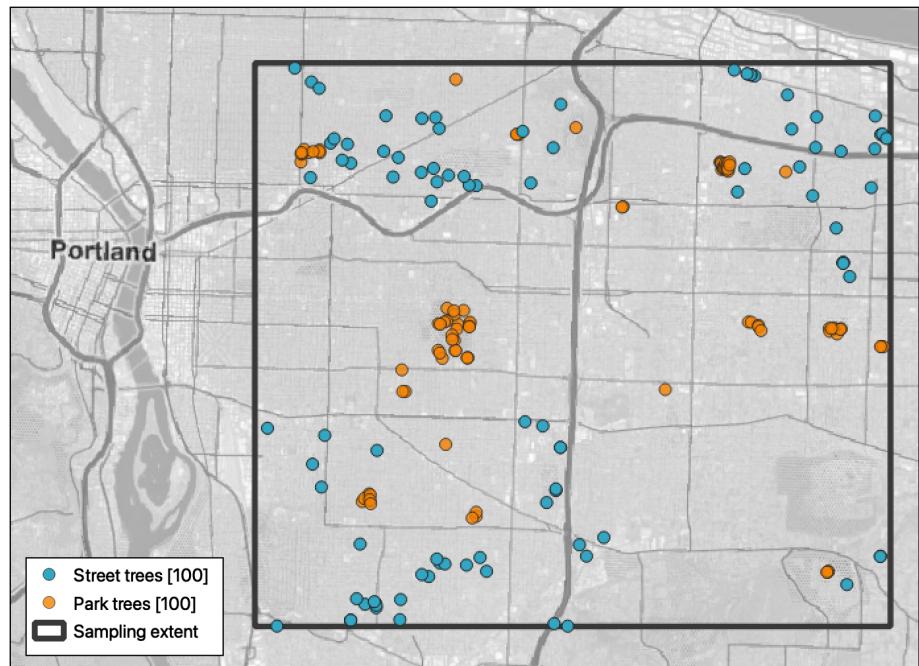


Figure 2.10: Geographic extent and random sampling

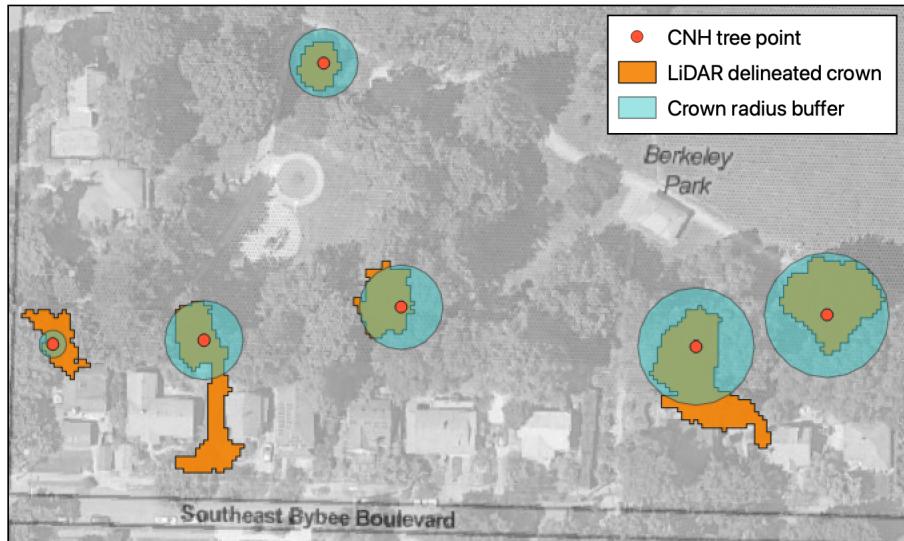


Figure 2.11: LiDAR tree crown delineations with selected CNH crowns

2.9 Modeling Tree Health

To tie all of my smaller pieces of data analysis together, I created a model to predict tree health categorization from NDVI values. I chose an ordinal logistic regression, which is used to model and predict the relationship between an ordinal categorical response variable and one or more explanatory variables, which can be either categorical or continuous. In order to use this type of model, I tested each model for the Proportional Odds Assumption, and

I used the CNH data as both my training and testing data. I chose to do this because of the limited sample size of CNH trees, and the uneven size of the health category ratings. Because of this, it is important to note that the model tests are an idealized version of a true model.

For each model, I created a confusion matrix to examine the validity of the model. A confusion matrix is a calculated cross-tabulation of observed and predicted classes, which in this case is health categorization, along with associated statistics. The two main statistics I use in my data analysis are overall accuracy and an un-weighted Kappa statistic. “Accuracy” refers to the sensitivity [Sensitivity = $\frac{\text{True positive}}{\text{True positive} + \text{False negative}}$] and specificity [Specificity = $\frac{\text{True negative}}{\text{True negative} + \text{False positive}}$] of the prediction.

$$\text{Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

 Kappa (κ) refers to Cohen's Kappa Statistic which is used to measure inter and intra-rater reliability for categorical items.

$$\kappa = \frac{p_o - p_e}{1 - p_e} = \frac{1 - p_o}{1 - p_e}$$

 where p_o is the relative observed agreement between the reference and prediction, and p_e is the hypothetical probability of agreement by chance. If the two raters (in this case the reference and prediction categorizations) are in complete agreement, then $\kappa = 1$.

Chapter 3

Results

3.1 Canopy Width and Tree Height Model

I used a third order polynomial regression to predict both tree height and crown width from tree DBH. In order to account for expected differences between species, I included species as an interaction term. I chose a third order polynomial because it not only encapsulated the relationship between DBH and tree height as well as DBH and canopy width, but it also was the most accurate when modeling the individual species. While the tree height model has an adjusted R-squared value of 0.78, and a P-value of < 2.2e-16 (Figure 3.1). The crown width model has an adjusted R-squared value of 0.72, and a P-value of < 2.2e-16 (Figure 3.2, A.1).

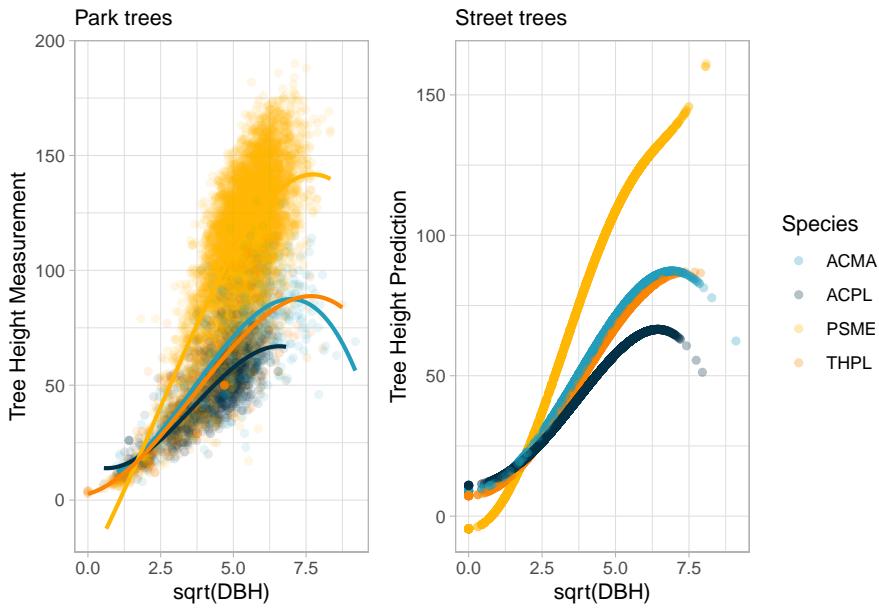


Figure 3.1: Predictive model for tree height from measured DBH. A third order polynomial regression was used in order to account for the variation between species. (Adjusted R-squared = 0.72, P <2.2e-16)

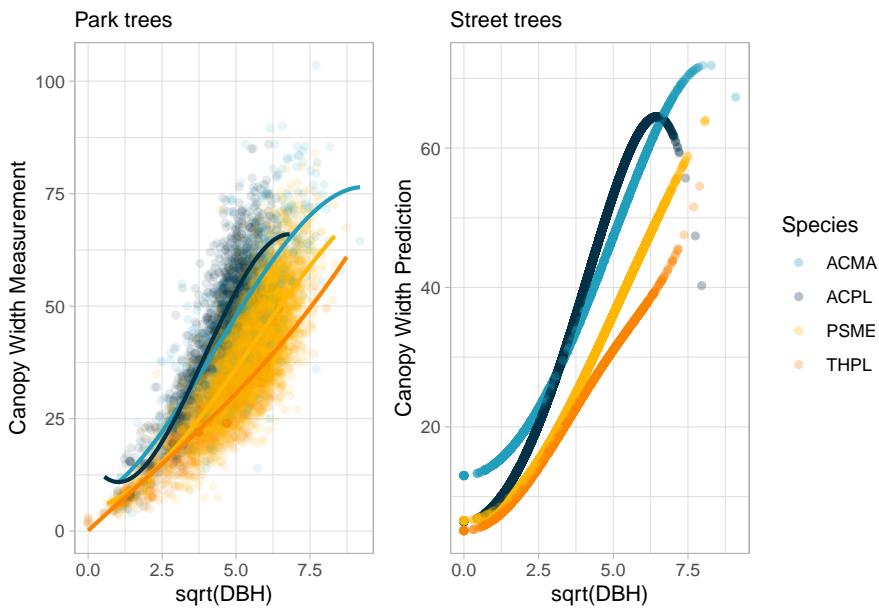


Figure 3.2: Third order polynomial model for predicting tree crown width based on measured DBH and species, with predictions for street trees. (R-squared = 0.72, P <2.2e-16)

3.2 Point Method

With the selection of CNH trees and 2021 NDVI data, a statistical analysis of the point value method shows that there is a statistically significant difference in the average NDVI values between health categorizations of “Fair” and “Good” (ANOVA, $F_{2,109} = 3.892$, $P = 0.023$, TukeyHSD). There is a general positive correlation between NDVI and health category, specifically Fair and Good. The relationship between health and NDVI is more apparent in the two maple species, but still holds in the “fair” and “good” categories in the two coniferous species (Figure 3.3).

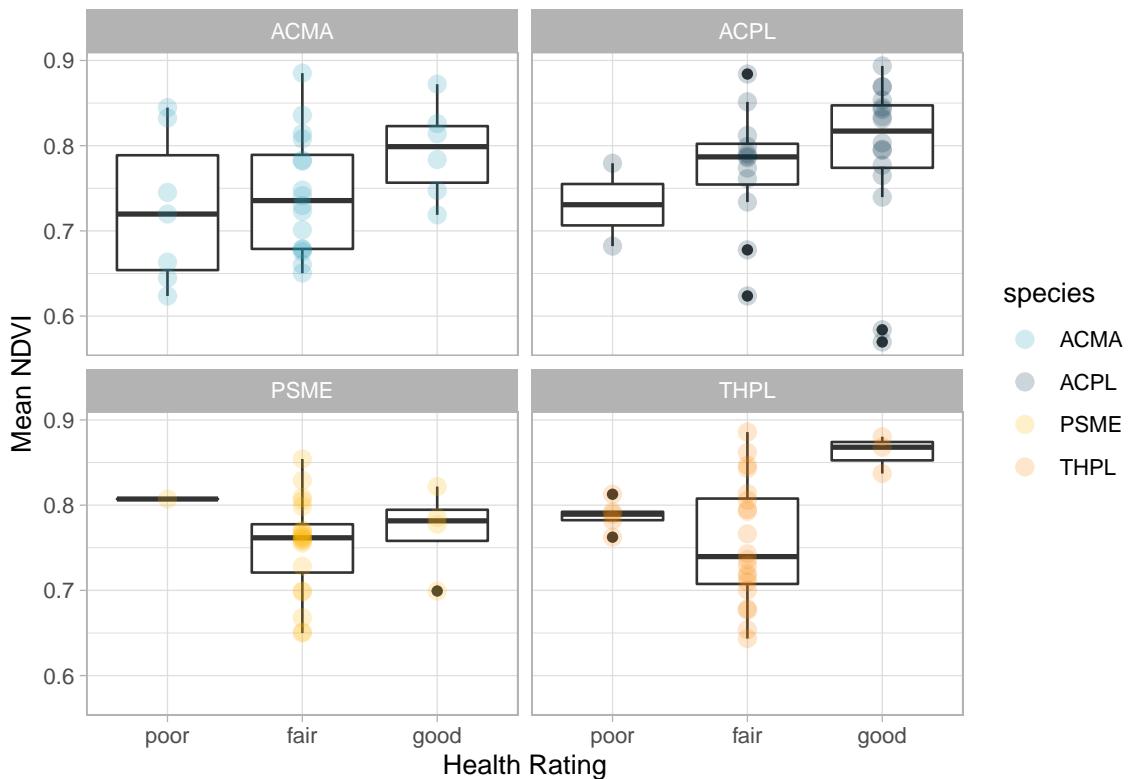


Figure 3.3: NDVI extracted for CNH trees using the point method compared to CNH tree health categorization. There is a statistically significant difference between Fair and Good (ANOVA, $F_{2,109} = 3.892$, $P = 0.023$, TukeyHSD)

3.3 Radius Method

With the radius method of pixel selection and crown delineation, there is a statistically significant difference in the mean NDVI values for health categories (ANOVA, $F_{2,109} = 4.923$, $P = 0.0089$; TukeyHSD: good-poor $P = 0.033$, good-fair $P = 0.014$). Similar

to the point method, there is a larger difference in average NDVI between the “fair” and “good” categories than “poor” and “fair”, and with the radius method, this is now a statistically significant difference (Figure 3.4).

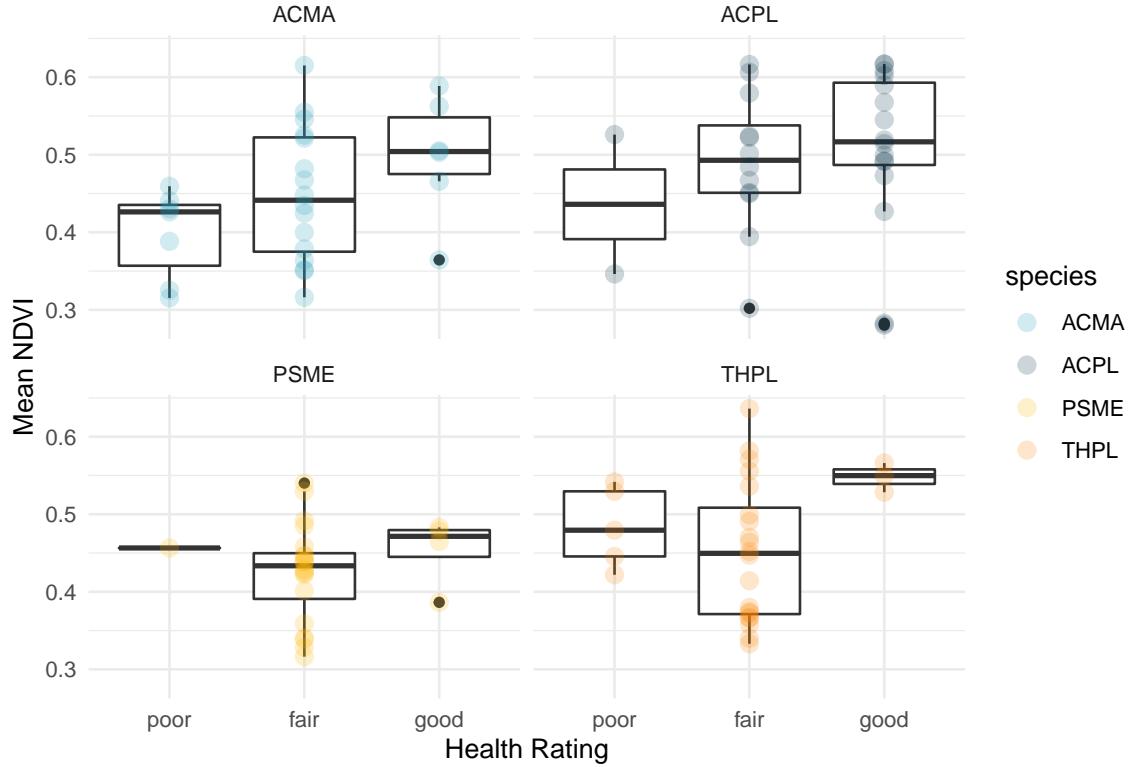


Figure 3.4: Boxplot of average NDVI from radius method and health condition, by species. The relationship between NDVI and health category is strongest in the maples, but still somewhat apparent in the conifers. (ANOVA, $F_{2,109} = 4.923$, $P = 0.0089$; TukeyHSD: good-poor $P = 0.033$, good-fair $P = 0.014$)

3.4 LiDAR Method

There is a statistically significant difference in NDVI between health categories for the LiDAR method, specifically between good and fair (ANOVA, $F_{2,101} = 5.405$, $P = 0.00589$, TukeyHSD). The same relationship between NDVI and health condition for the maples that was seen in the point and radius method is seen here (Figure 3.5).

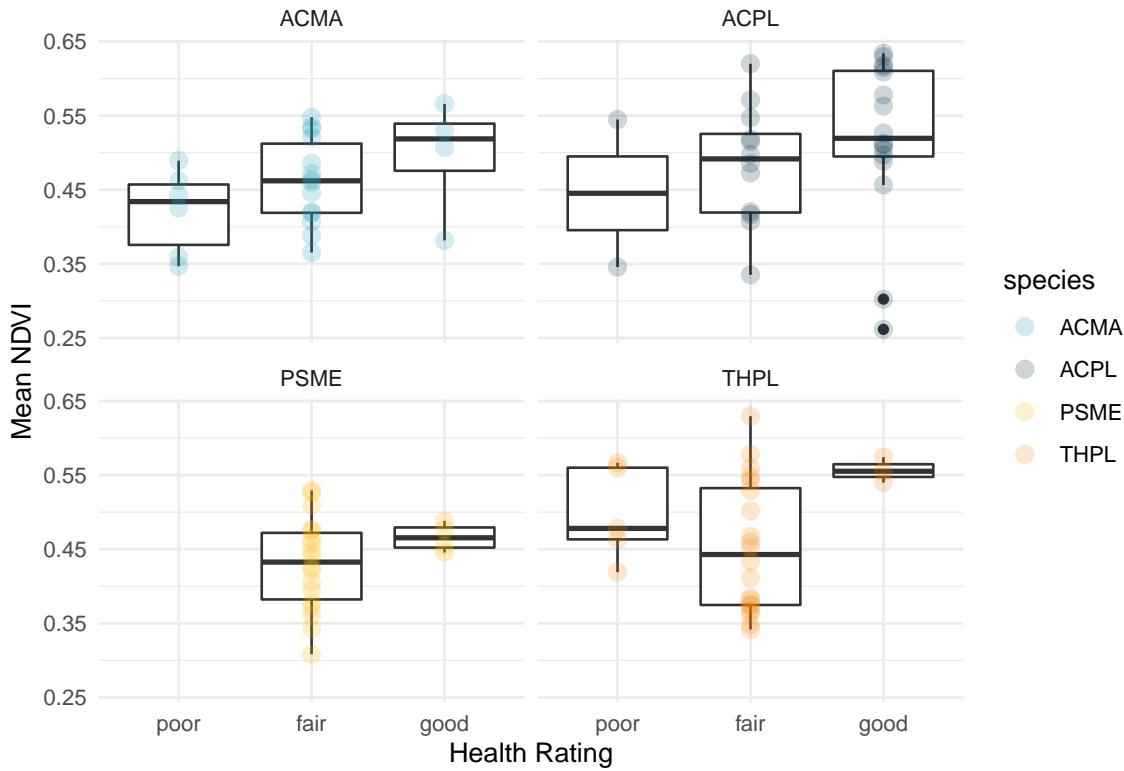


Figure 3.5: Boxplot of average NDVI from LiDAR method and health condition for each species. Due to the nature of the LiDAR delineation algorithm, no crown was detected for the PSME individual with a poor health rating. The difference between good and fair average NDVIs is statistically significant (ANOVA, $F_{2,101} = 5.405$, $P = 0.00589$, TukeyHSD)

Another aspect of the LiDAR delineation method is the potential for data loss. Five ACMA and three PSME individuals were lost in the LiDAR crown delineation process. If this was with a very large sample size the data loss would likely not be detrimental but especially with a sample size such as mine, a loss of eight trees is a loss of 5% of the total data, 17% loss for ACMA, and 12% loss for THPL.

3.5 Method Comparison

There is a statistically significant difference in NDVI values for the three different pixel selection methods (ANOVA, $F_{2,325} = 517.8$, $P < 2e - 16$; TukeyHSD, radius-point $P = 0.00$, lidar-point $P = 0.00$). When the species are clumped, the three methods show very similar patterns, though the NDVI values are higher for the point method than either of the other two methods (Figure 3.6). When all three

methods are compared along with species, we can see that the general trend of health rating and NDVI is consistent within each species, but varies across methods (Figure 3.7). Additionally, the pattern between health rating and NDVI is also very similar between trees of the same functional type. Because of the similarities in health patter, I included additional testing examining the difference in species specificity addition versus functional type separation.

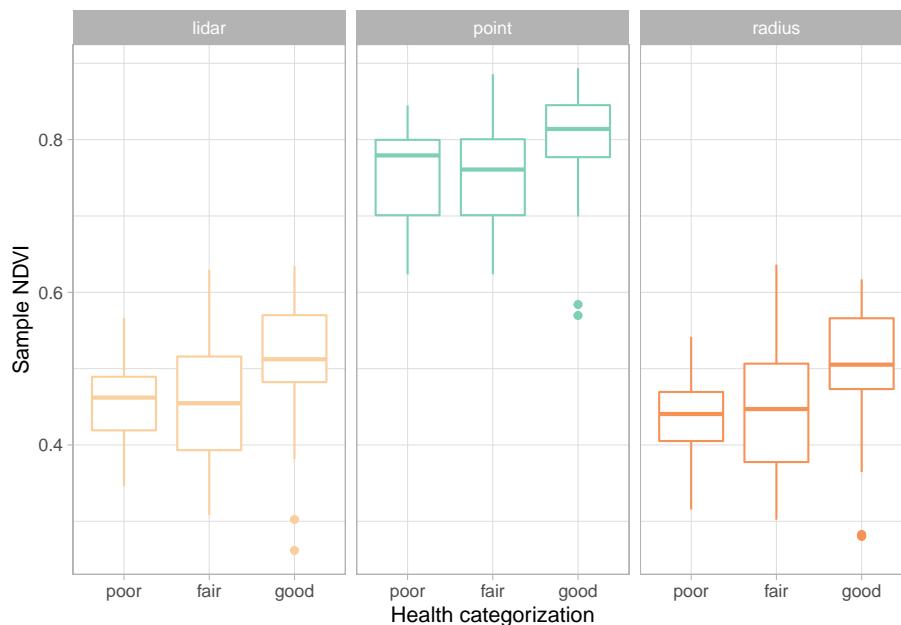


Figure 3.6: Average NDVI values from all three pixel selection methods with CNH health rating. There is a statistically significant difference in mean NDVI between point and both radius and lidar methods. (ANOVA, $F_{2,325} = 517.8$, $P = < 2e - 16$, TukeyHSD: radius-point $P = 0.00$, lidar-point $P = 0.00$)

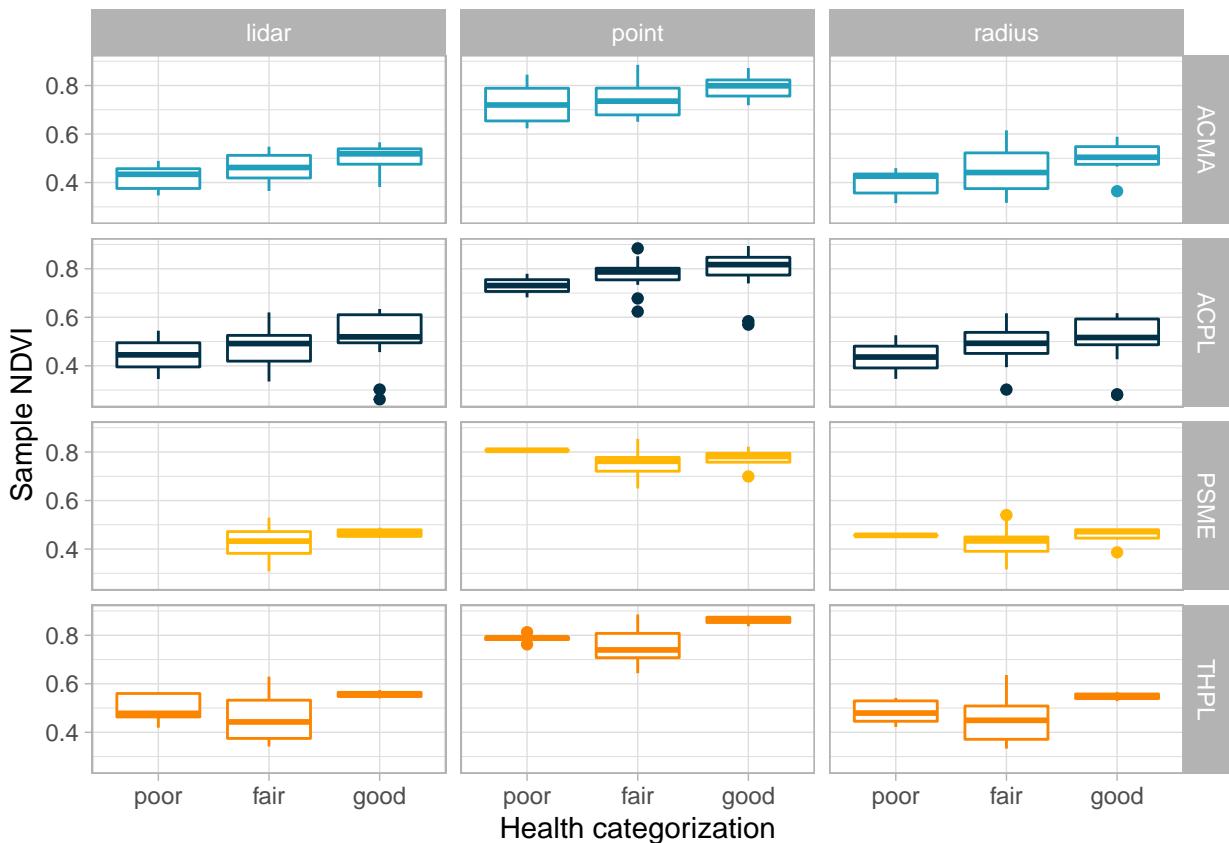


Figure 3.7: Average NDVI for each health category, split by tree species and pixel selection method. The general trends remain the same within species, but vary among pixel selection methods.

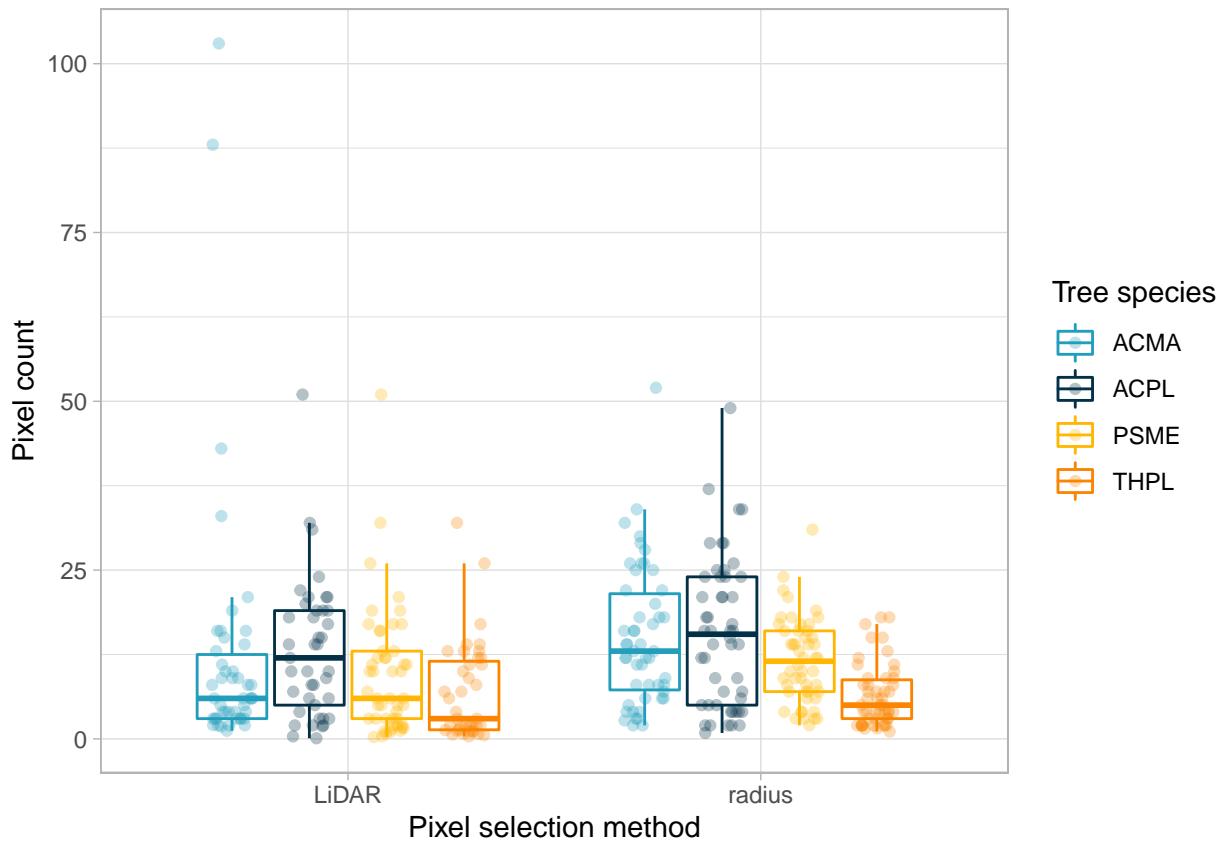
Another important factor to consider when comparing the three pixel selection methods is the amount of pixels used in the NDVI analysis. For the point method, each NDVI value is based off one singular point.

```

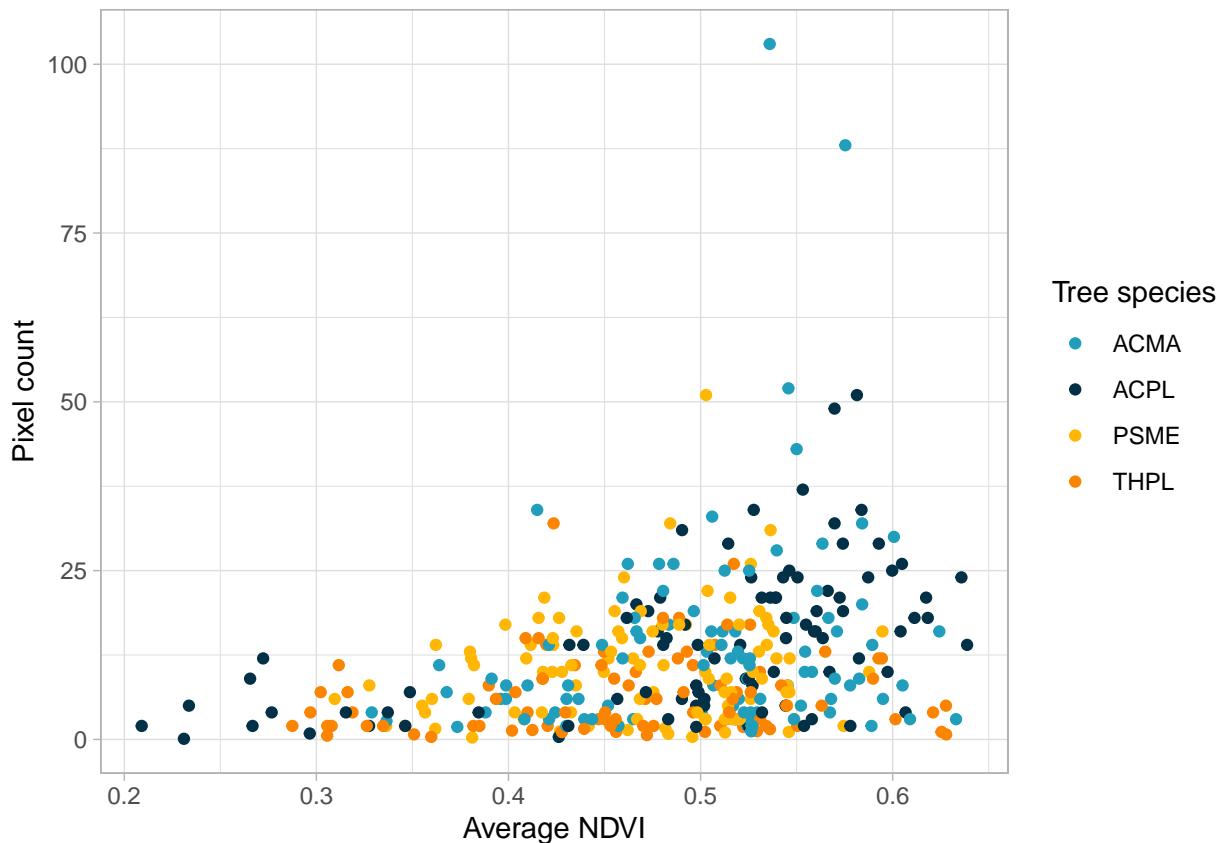
radius_sample_data <- read_csv("data/radius_sample_data.csv") %>%
  dplyr::select(-4, -5) %>%
  mutate(method = "radius")
lidar_sample_data <- read_csv("data/lidar_sample_data.csv") %>%
  dplyr::select(-4) %>%
  mutate(method = "LiDAR")
count_sample_data <- lidar_sample_data %>%
  bind_rows(radius_sample_data)
ggplot(data = count_sample_data, aes(x = method, y = ndvi_count,
  color = species)) + geom_boxplot(outlier.shape = NA) + geom_point(position =

```

```
alpha = 0.3) + scale_color_manual(values = species_pal) +
  labs(x = "Pixel selection method", y = "Pixel count", color = "Tree species")
```



```
ggplot(data = count_sample_data, aes(x = ndvi_mean, y = ndvi_count,
  color = species)) + geom_point() + scale_color_manual(values = species_pal) +
  labs(x = "Average NDVI", y = "Pixel count", color = "Tree species")
```

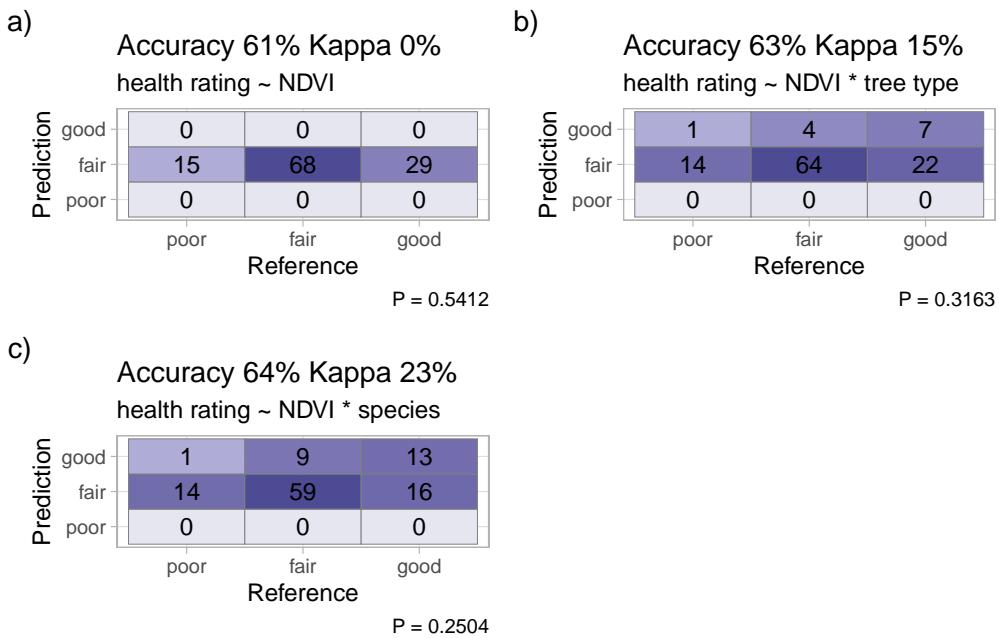


3.6 Predictive Model

In order to determine the impact of differentiating by species in predictive models, I ran three different models for each point selection method, moving from least to most specific: a model that only uses NDVI to predict health rating, a model that includes functional tree type as an interaction term, and a third model that includes species as an interaction term. Due to the limited sample size of the data I am working with, the following models are trained and tested on the same data. If it were to work perfectly, the model would correctly predict the health condition of a CNH tree 100% of the time. The data I use is not perfect, and the models are not perfect either. However, they are likely an idealized version of how model testing and training would work in reality. With the point method data, the first model with no interaction term led to all tree points being rated as “Fair”. The addition of tree type as an interaction term led to an increase in predictions of “good” health trees, and increased kappa from 0% to 15%. The overall accuracy only increased by 2%, but kappa is more informative in terms of model validity than the accuracy score. Using tree species instead of tree

type further improved the kappa of the model. The number of “good” trees that were correctly predicted increased from 7 to 13, but additionally the number of “fair” trees that were correctly predicted dropped from 64 to 59, with the new predictions as “good”. (Figure ??). None of these models with the point method were able to correctly predict any trees with poor health categorization.

Point method



The radius method was slightly more successful than the point method for the model, with all three method models performing better than those for the point method. From just NDVI to NDVI * tree type, there was an increase in 9 trees to good ratings. Including no species or tree type specification in the predictive model still seems to perform the worst of the three models. The main takeaway from the radius method models is that with the inclusion of species as an interaction term with the radius data, we see our first predictions of the “poor” category. The radius method does involve more pixels than the point method, and with the range in NDVI values being lower than that of the point method it makes sense that there is an increase in “poor” rated trees. However, out of the 15 reference trees with health ratings of “poor”, only one was correctly predicted as such (Figure 3.8).

Radius method

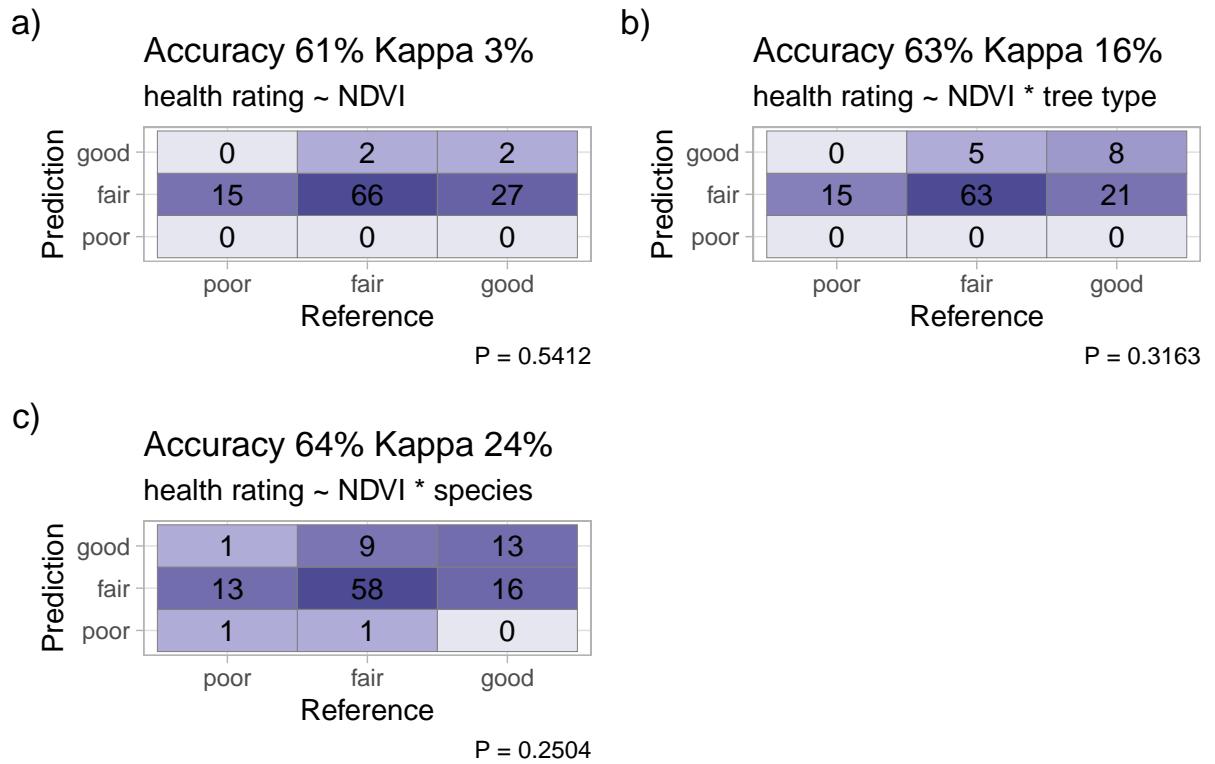
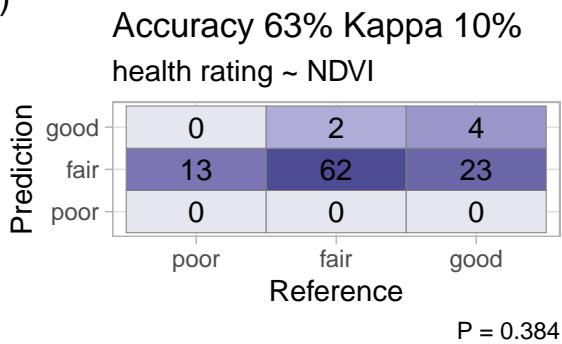


Figure 3.8: Confusion matrixes for each species predictions with radius method data

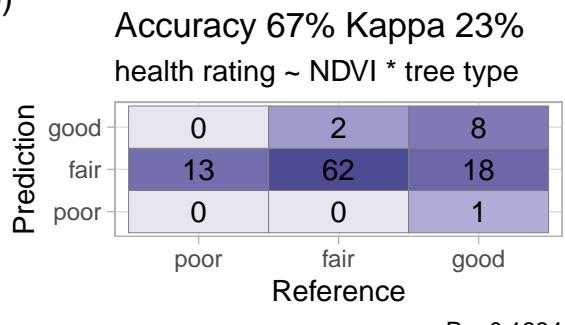
The LiDAR data was the most effective out of the three methods at predicting tree health, but also had the only occurrence of a “good” tree being predicted as “poor”.

LiDAR method

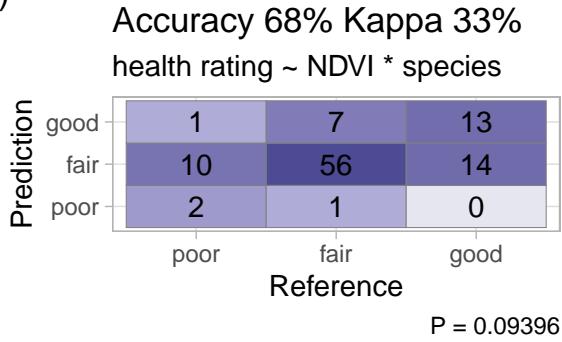
a)



b)



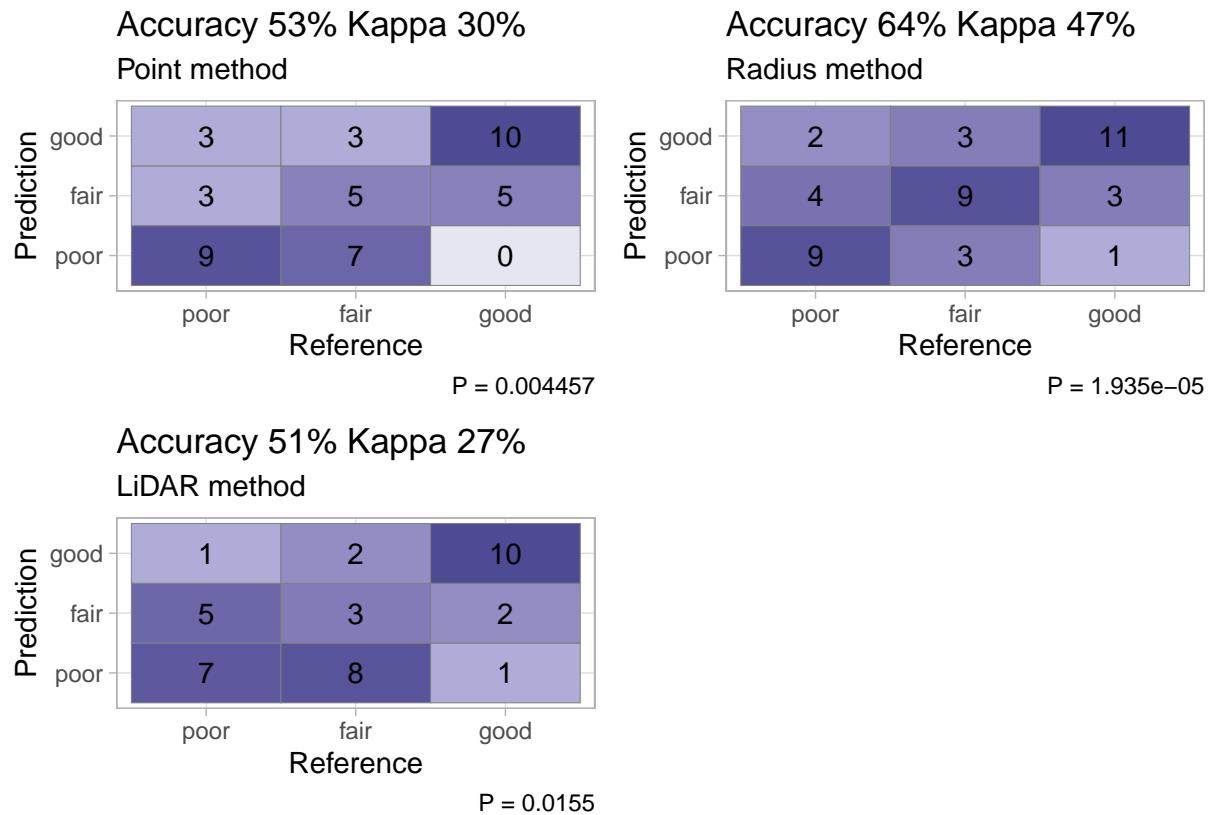
c)



A predictive model was also run with all species together, but with species as an interaction term in the formula. This showed better results than when the species were split, but it seems that most of that is due to increased overall sample size.

Due to the variation in sample sizes for each health condition category, I sampled down the data so each health category had the same number of points. This improved the results of the model by a lot.

Given the uneven number of data points available for the varying health conditions, I down-sampled each health category to the smallest number of points in a given category (point = 15, radius = 15, LiDAR = 13). When the predictive models were ran with the even category sizes, the results were statistically significant for all three pixel selection models. For the poi



Chapter 4

Discussion

testing



Appendix A

Additional Data

A.1 Tree Height and Crown Width Predictive Model Equations

ACMA equations: ($x = \text{DBH}$)

$$\text{Tree Height} = 8.977 + 3.617x - 0.049x^2 + 0.001x^3 \quad (\text{A.1})$$

$$\text{Crown Width} = 13.01 + 1.593x - 0.007x^2 + 0.0004x^3 \quad (\text{A.2})$$

ACPL equations: ($x = \text{DBH}$)

$$\text{Tree Height} = 10.955 + 2.67x - 0.032x^2 + 0.001x^3 \quad (\text{A.3})$$

$$\text{Crown Width} = 6.445 + 2.29x - 0.009x^2 - 0.0002x^3 \quad (\text{A.4})$$

PSME equations: ($x = \text{DBH}$)

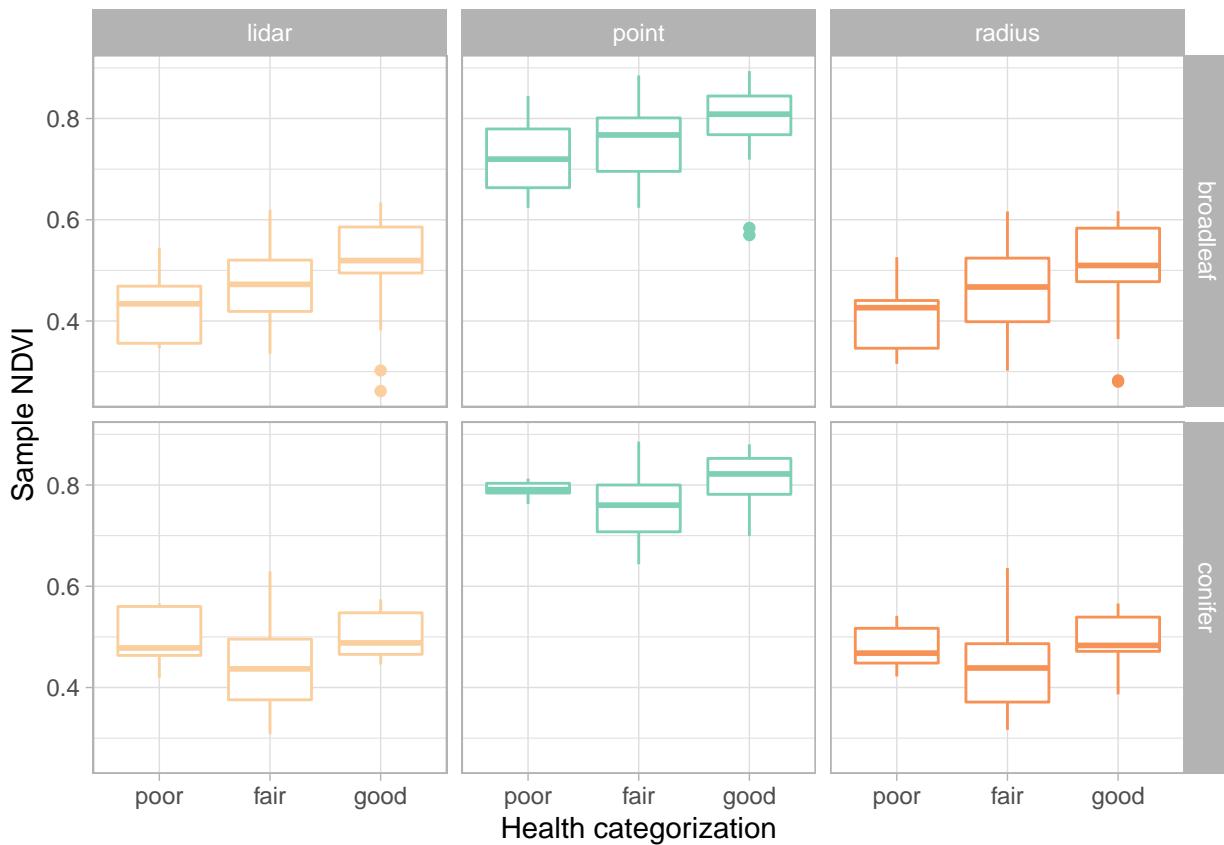
$$\text{Tree Height} = -4.497 + 7.516x - 0.146x^2 + 0.001x^3 \quad (\text{A.5})$$

$$\text{Crown Width} = 6.536 + 1.471x - 0.013x^2 + 0.0005x^3 \quad (\text{A.6})$$

THPL equations: ($x = \text{DBH}$)

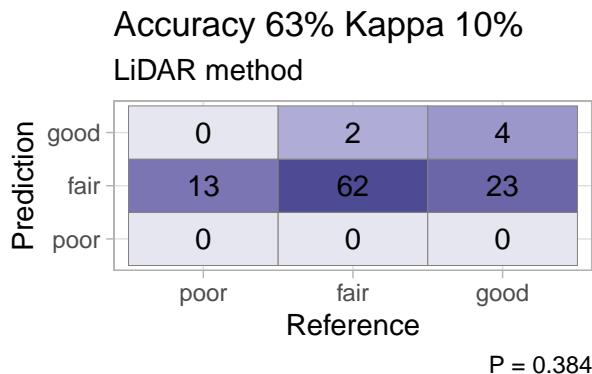
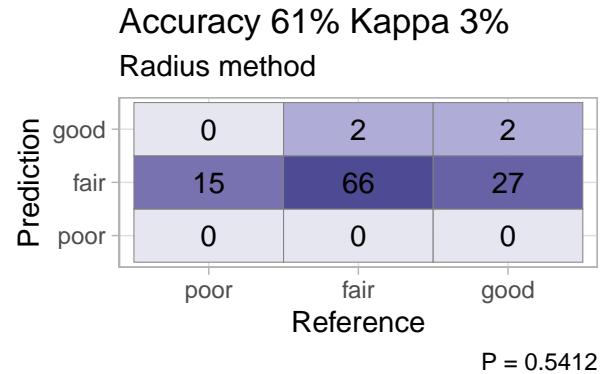
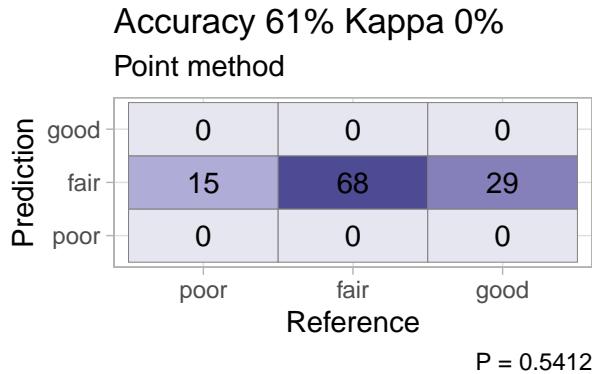
$$\text{Tree Height} = 7.311 + 3.457x - 0.0473x^2 + 0.001x^3 \quad (\text{A.7})$$

$$\text{Crown Width} = 5.113 + 1.622x - 0.032x^2 + 0.0002x^3 \quad (\text{A.8})$$

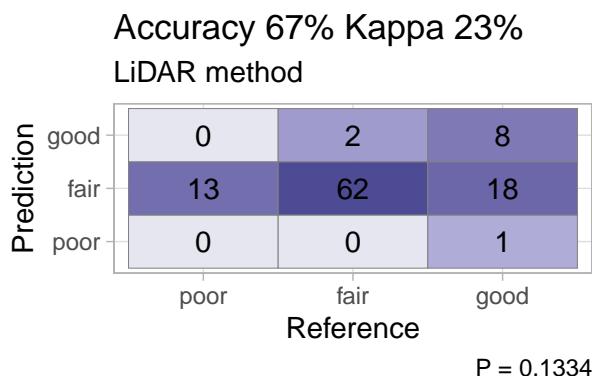
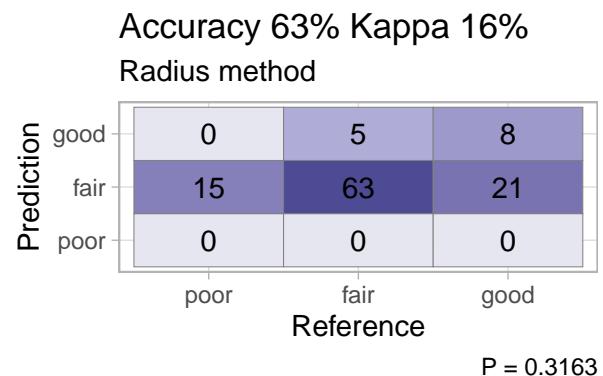
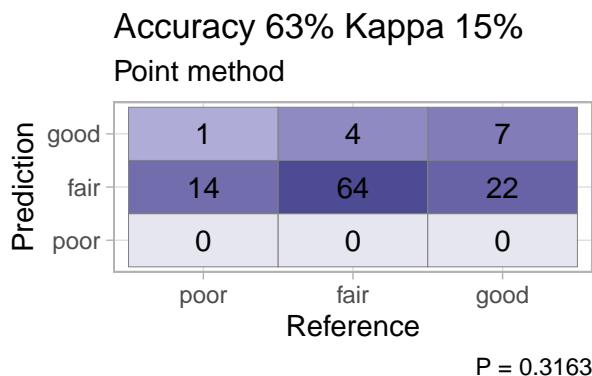


extra modeling graphics

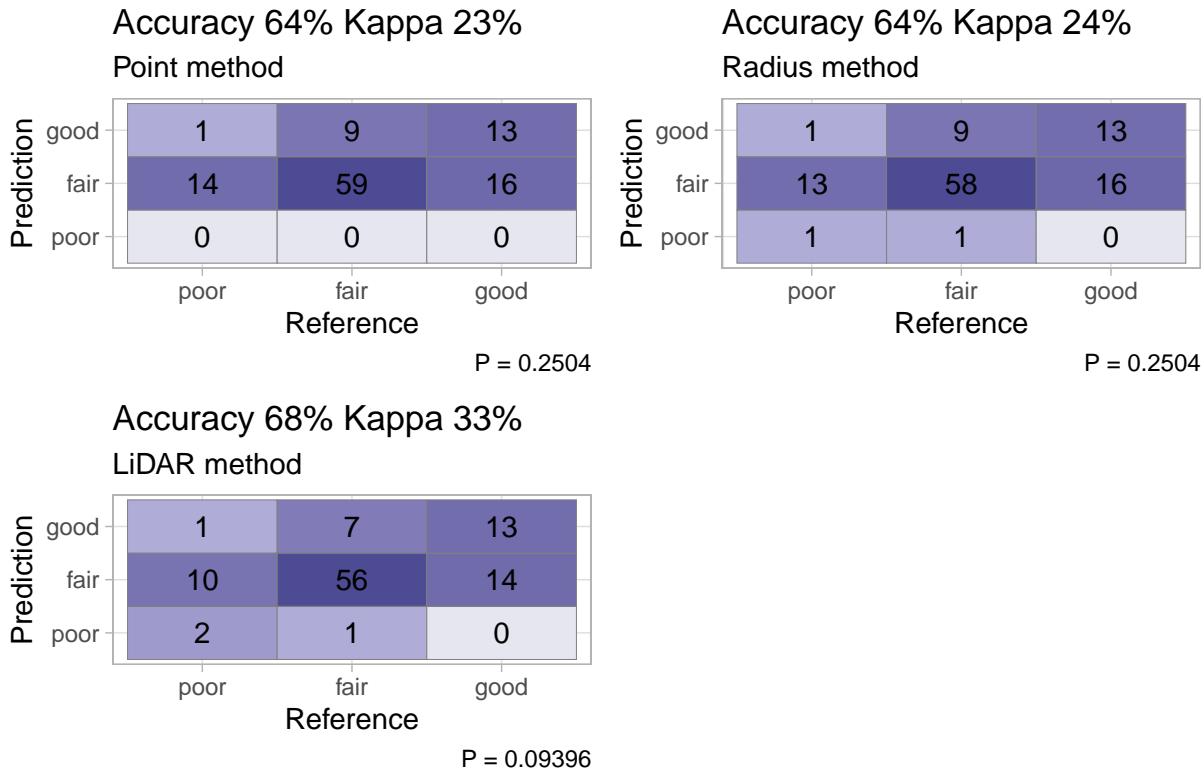
formula = health rating ~ NDVI



formula = health rating ~ NDVI * tree type

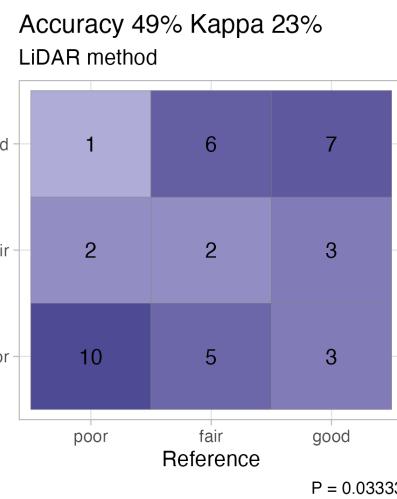
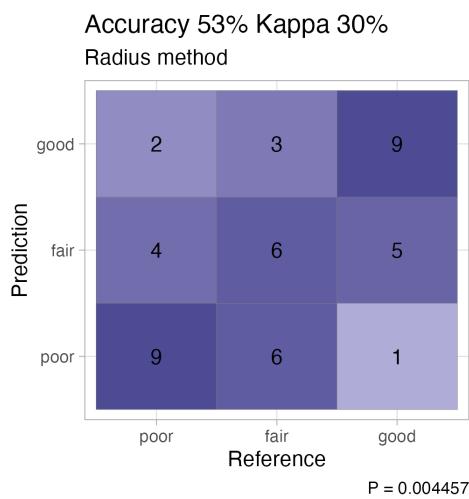


formula = health rating ~ NDVI * species

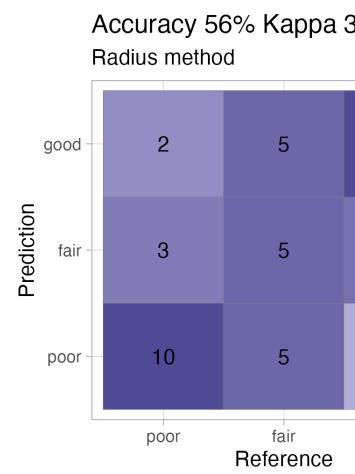


Radius and LiDAR model tests

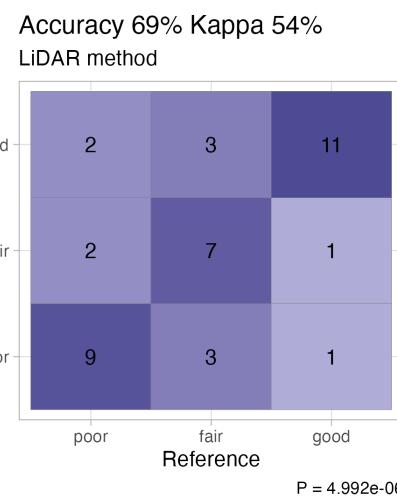
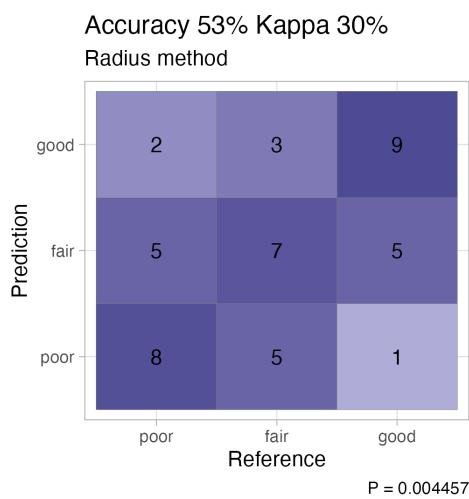
test 1



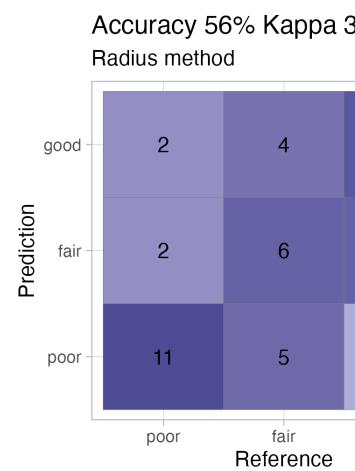
test 2



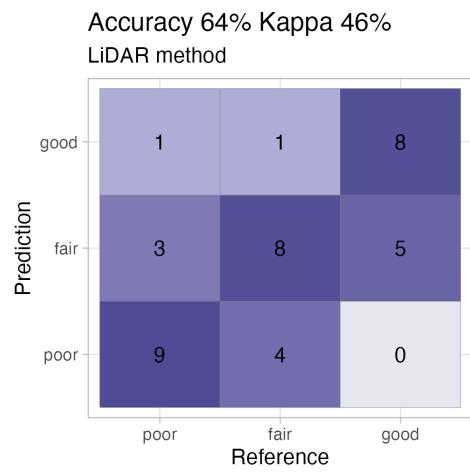
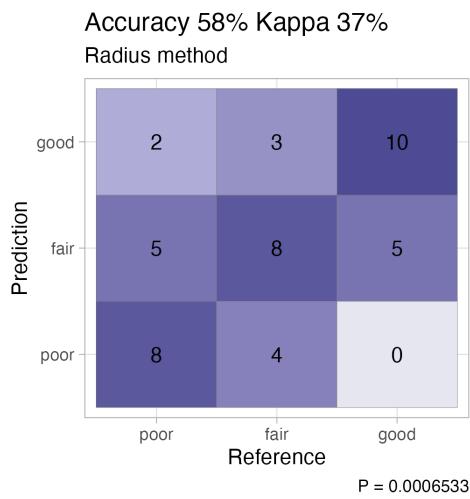
test 3



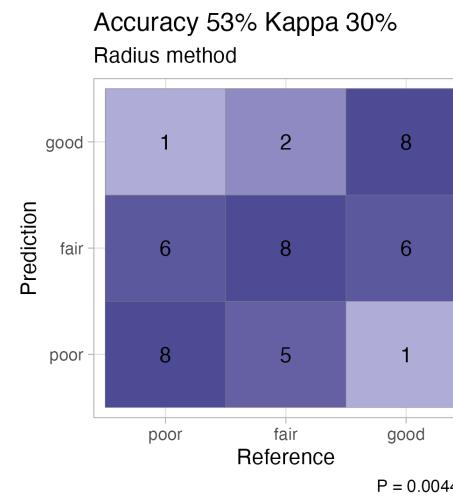
test 4



test 5



test 6



Appendix B

Code

This second appendix includes all of the R chunks of code that were hidden throughout the document to help with readability and/or setup.

B.1 In Chapter 2:

Portland Tree Inventory counts and calculations

```
street_counts <- get_pdxTrees_streets() %>%
  filter(Species %in% c("ACMA", "ACPL", "PSME", "THPL")) %>%
  mutate(year = year(Inventory_Date)) %>%
  dplyr::select(Species, Scientific, year) %>%
  add_count(Species, year) %>%
  distinct() %>%
  group_by(Species) %>%
  mutate(Total = sum(n)) %>%
  rename(Count_2016 = n) %>%
  filter(year == 2016) %>%
  dplyr::select(-year)

park_counts <- get_pdxTrees_parks() %>%
  filter(Species %in% c("ACMA", "ACPL", "PSME", "THPL")) %>%
  mutate(year = year(Inventory_Date)) %>%
  dplyr::select(Species, Scientific_Name, year) %>%
  add_count(Species, year) %>%
  distinct() %>%
  group_by(Species) %>%
  mutate(Total = sum(n)) %>%
  rename(Count_2019 = n) %>%
  filter(year == 2019) %>%
```

```
dplyr::select(-year)
```

Calculating NDVI from satellite imagery

```
import rasterio
import numpy

files = ["1", "2", "3", "4"] # unique file identifiers
folder = "ndvi_calc_thesis/files" # folder with the raw files
date = "20210726" # product date

for i in range(len(files)):
    filenum = files[i]
    image_file = folder+"/"+date+"_"+filenum+"_AnalyticMS.tif"

    metafile = folder+"/"+date+"_"+filenum+"_AnalyticMS_metadata.xml"
    print(image_file)
    print(metafile)

    # Load red and NIR bands -
    # note all PlanetScope 4-band images have band order BGRN

    with rasterio.open(image_file) as src:
        band_red = src.read(3)

    with rasterio.open(image_file) as src:
        band_nir = src.read(4)

    from xml.dom import minidom

    xmldoc = minidom.parse(metafile)
    nodes = xmldoc.getElementsByTagName("ps:bandSpecificMetadata")

    # XML parser refers to bands by numbers 1-4
    coeffs = {}
    for node in nodes:
        bn = node.getElementsByTagName("ps:bandNumber")[0].firstChild.data
        if bn in ['1', '2', '3', '4']:
            i = int(bn)
            value = node.getElementsByTagName("ps:reflectanceCoefficient")[0].firstChild.data
            coeffs[i] = float(value)
```

```

# Multiply by corresponding coefficients
band_red = band_red * coeffs[3]
band_nir = band_nir * coeffs[4]

# Allow division by zero
numpy.seterr(divide='ignore', invalid='ignore')

# Calculate NDVI
ndvi = (band_nir.astype(float) - band_red.astype(float)) / (band_nir + band_red)

# Set spatial characteristics of the output object to mirror the input
kwargs = src.meta
kwargs.update(
    dtype=rasterio.float32,
    count=1)

# Set name for new NDVI file
newfile = folder + '/ndvi_' + date + "_" + filenum + '.tif'

# Create the file
with rasterio.open(newfile, 'w', **kwargs) as dst:
    dst.write_band(1, ndvi.astype(rasterio.float32))

```

LiDAR canopy delineation

```

library(rgdal)
library(imager)
library(raster)
library(ForestTools)
options(rgdal_show_exportToProj4_warnings = "none")
## Loading raster data
EML_CHM <- raster("~/Desktop/thesis_data/lidar_crown/clipped_lidar.tif")
## Loading tree points
all_ttops <- shapefile("~/Desktop/thesis_data/lidar_crown/all_tree_points_for_delin.tif")
# delineate tree crowns inputs: my tree points, LiDAR data
crownsPoly <- mcws(treetops = all_ttops, CHM = EML_CHM, format = "polygons",
                     minHeight = 20, verbose = TRUE)
# export file
writeOGR(crownsPoly, "~/Desktop/thesis_data/lidar_crown", "canopy_delin_polygons_poly",
          driver = "ESRI Shapefile")

```

B.2 In Chapter 3:

Model for Tree Height and Canopy Width

```

# load park data, filter to relevant species, and get
# average crown width.
my_park <- pdxTrees::get_pdxTrees_parks() %>%
  filter(Species %in% c("ACPL", "THPL", "PSME", "ACMA")) %>%
  mutate(crown_width = (Crown_Width_EW + Crown_Width_NS)/2)
# create model for heights
heights <- lm(Tree_Height ~ poly(DBH, degree = 3, raw = T) *
  Species, data = my_park)
# create model for crown width
cr_width <- lm(crown_width ~ poly(DBH, degree = 3, raw = T) *
  Species, data = my_park)
# load street data, filter to relevant species
my_street <- pdxTrees::get_pdxTrees_streets() %>%
  filter(Species %in% c("ACPL", "THPL", "PSME", "ACMA"))
# run street trees through both models
my_street <- my_street %>%
  mutate(width_preds = predict(cr_width, my_street, se.fit = FALSE),
  height_preds = predict(heights, my_street, se.fit = F))
# export data file write_csv(my_street, 'my_street.csv')
# graphing
p_height_model <- my_park %>%
  ggplot(aes(x = sqrt(DBH), y = Tree_Height, color = Species)) +
  geom_point(alpha = 0.1) + scale_color_manual(values = species_pal) +
  geom_smooth(method = "lm", se = F, formula = y ~ poly(x,
  degree = 3, raw = T)) + labs(subtitle = "Park trees",
  y = "Tree Height Measurement") + guides(color = "none")
p_width_model <- my_park %>%
  ggplot(aes(x = sqrt(DBH), y = crown_width, color = Species)) +
  geom_point(alpha = 0.1) + scale_color_manual(values = species_pal) +
  geom_smooth(method = "lm", se = F, formula = y ~ poly(x,
  degree = 3, raw = T)) + labs(subtitle = "Park trees",
  y = "Canopy Width Measurement") + guides(color = "none")
s_height_preds <- my_street %>%
  ggplot(aes(x = sqrt(DBH), y = height_preds, color = Species)) +
  geom_point(alpha = 0.3) + scale_color_manual(values = species_pal) +
  labs(subtitle = "Street trees", y = "Tree Height Prediction")
s_width_preds <- my_street %>%
  ggplot(aes(x = sqrt(DBH), y = width_preds, color = Species)) +
  geom_point(alpha = 0.3) + scale_color_manual(values = species_pal) +

```

```
labs(subtitle = "Street trees", y = "Canopy Width Prediction")
```

Point Method Modeling

```
point_all <- polr(health_rat ~ sample_ndvi, Hess = TRUE, data = cnh_point)
point_all_preds <- predict(point_all, cnh_point)
cfm_p_all <- confusionMatrix(point_all_preds, cnh_point$health_rat)
p_all <- ggplotConfusionMatrix(cfm_p_all) + labs(subtitle = "health rating ~ NDVI")
brant::brant(point_all)
point_type <- polr(health_rat ~ sample_ndvi * tree_type, Hess = TRUE,
                    data = cnh_point)
point_type_preds <- predict(point_type, cnh_point)
cfm_p_type <- confusionMatrix(point_type_preds, cnh_point$health_rat)
p_type <- ggplotConfusionMatrix(cfm_p_type) + labs(subtitle = "health rating ~ NDVI * tree type")
brant::brant(point_type)
point_species <- polr(health_rat ~ sample_ndvi * species, Hess = TRUE,
                      data = cnh_point)
point_species_preds <- predict(point_species, cnh_point)
cfm_p_species <- confusionMatrix(point_species_preds, cnh_point$health_rat)
p_species <- ggplotConfusionMatrix(cfm_p_species) + labs(subtitle = "health rating ~ NDVI * species")
brant::brant(point_species)
```

Radius Method Modeling

```
radius_all <- polr(health_rat ~ sample_ndvi, Hess = TRUE, data = cnh_radius)
radius_all_preds <- predict(radius_all, cnh_radius)
cfm_r_all <- confusionMatrix(radius_all_preds, cnh_radius$health_rat)
r_all <- ggplotConfusionMatrix(cfm_r_all) + labs(subtitle = "health rating ~ NDVI")
radius_type <- polr(health_rat ~ sample_ndvi * tree_type, Hess = TRUE,
                     data = cnh_radius)
radius_type_preds <- predict(radius_type, cnh_radius)
cfm_r_type <- confusionMatrix(radius_type_preds, cnh_radius$health_rat)
r_type <- ggplotConfusionMatrix(cfm_r_type) + labs(subtitle = "health rating ~ NDVI * tree type")
radius_species <- polr(health_rat ~ sample_ndvi * species, Hess = TRUE,
                      data = cnh_radius)
radius_species_preds <- predict(radius_species, cnh_radius)
cfm_r_species <- confusionMatrix(radius_species_preds, cnh_radius$health_rat)
r_species <- ggplotConfusionMatrix(cfm_r_species) + labs(subtitle = "health rating ~ NDVI * species")
```

LiDAR Method Modeling

```

lidar_all <- polr(health_rat ~ sample_ndvi, Hess = TRUE, data = cnh_lidar)
lidar_all_preds <- predict(lidar_all, cnh_lidar)
cfm_l_all <- confusionMatrix(lidar_all_preds, cnh_lidar$health_rat)
l_all <- ggplotConfusionMatrix(cfm_l_all) + labs(subtitle = "health rating ~ NDVI")
lidar_type <- polr(health_rat ~ sample_ndvi * tree_type, Hess = TRUE,
                     data = cnh_lidar)
lidar_type_preds <- predict(lidar_type, cnh_lidar)
cfm_l_type <- confusionMatrix(lidar_type_preds, cnh_lidar$health_rat)
l_type <- ggplotConfusionMatrix(cfm_l_type) + labs(subtitle = "health rating ~ NDVI * tree type")
lidar_species <- polr(health_rat ~ sample_ndvi * species, Hess = TRUE,
                      data = cnh_lidar)
lidar_species_preds <- predict(lidar_species, cnh_lidar)
cfm_l_species <- confusionMatrix(lidar_species_preds, cnh_lidar$health_rat)
l_species <- ggplotConfusionMatrix(cfm_l_species) + labs(subtitle = "health rating ~ NDVI * species")

```

Radius and LiDAR Replications

```

test_data_point <- cnh_long %>%
  filter(method == "point") %>%
  group_by(health_rat) %>%
  sample_n(15)
test_point <- polr(health_rat ~ sample_ndvi * species, Hess = TRUE,
                   data = test_data_point)
preds_point <- predict(test_point, test_data_point)
cfm_point <- confusionMatrix(preds_point, test_data_point$health_rat)
gg_point <- ggplotConfusionMatrix(cfm_point) + labs(subtitle = "Point method")
test_data_radius <- cnh_long %>%
  filter(method == "radius") %>%
  group_by(health_rat) %>%
  sample_n(15)
test_radius <- polr(health_rat ~ sample_ndvi * species, Hess = TRUE,
                     data = test_data_radius)
preds_radius <- predict(test_radius, test_data_radius)
cfm_radius <- confusionMatrix(preds_radius, test_data_radius$health_rat)
gg_radius <- ggplotConfusionMatrix(cfm_radius) + labs(subtitle = "Radius method")
test_data_lidar <- cnh_long %>%
  filter(method == "lidar") %>%
  group_by(health_rat) %>%
  sample_n(13)
test_lidar <- polr(health_rat ~ sample_ndvi * species, Hess = TRUE,

```

```
    data = test_data_lidar)
preds_lidar <- predict(test_lidar, test_data_lidar)
cfm_lidar <- confusionMatrix(preds_lidar, test_data_lidar$health_rat)
gg_lidar <- ggplotConfusionMatrix(cfm_lidar) + labs(subtitle = "LiDAR method")
```


References

- Anderson, L. M., & Cordell, H. K. (1988). Influence of trees on residential property values in Athens, Georgia (U.S.A.): A survey based on actual sales prices. *Landscape and Urban Planning*, 15(1), 153–164. [http://doi.org/10.1016/0169-2046\(88\)90023-0](http://doi.org/10.1016/0169-2046(88)90023-0)
- Beckmann-Wübbelt, A., Fricke, A., Sebesvari, Z., Yakouchenkova, I. A., Fröhlich, K., & Saha, S. (2021). High public appreciation for the cultural ecosystem services of urban and peri-urban forests during the COVID-19 pandemic. *Sustainable Cities and Society*, 74, 103240. <http://doi.org/10.1016/j.scs.2021.103240>
- Breshears, D. D., Cobb, N. S., Rich, P. M., Price, K. P., Allen, C. D., Balice, R. G., ... Meyer, C. W. (2005). Regional vegetation die-off in response to global-change-type drought. *Proceedings of the National Academy of Sciences*, 102(42), 15144–15148. <http://doi.org/10.1073/pnas.0505734102>
- Byer, S., & Jin, Y. (2017). Detecting Drought-Induced Tree Mortality in Sierra Nevada Forests with Time Series of Satellite Data. *Remote Sensing*, 9(9), 929. <http://doi.org/10.3390/rs9090929>
- Canopy 2014. (2016). Retrieved from <https://www.arcgis.com/sharing/rest/content/items/05cbf1630c2f4904a076298b888de9bc/info/metadata/metadata.xml?format=default&output=html>
- City of Portland. (2020). Carbon emissions. Retrieved from <https://www.portland.gov/bps/scg/scg-dashboard/carbon-emissions>
- DiSalvo, A., Fukuda, J., Ramsey, J., & Parks, P. (2017). *Street Tree Inventory Report: City of Portland* (p. 48).
- Donovan, G., & Butry, D. (2009). The value of shade: Estimating the effect of urban trees on summertime electricity use. *Energy and Buildings*, 41, 662–668. <http://doi.org/10.1016/j.enbuild.2009.01.002>
- Fang, F., McNeil, B., Warner, T., Dahle, G., & Eutsler, E. (2020). Street tree health from space? An evaluation using WorldView-3 data and the Washington D.C. Street Tree Spatial Database. *Urban forestry & urban greening*, 49, 126634–.

- http://doi.org/10.1016/j.ufug.2020.126634
- Flint, H. L. (1985). Plants showing tolerance of urban stress. *Journal of Environmental Horticulture*, 3(2), 85–89. http://doi.org/10.24266/0738-2898-3.2.85
- Huang, C., Anderegg, W. R. L., & Asner, G. P. (2019). Remote sensing of forest die-off in the Anthropocene: From plant ecophysiology to canopy structure. *Remote Sensing of Environment*, 231, 111233. http://doi.org/10.1016/j.rse.2019.111233
- James, P., Banay, R. F., Hart, J. E., & Laden, F. (2015). A Review of the Health Benefits of Greenness. *Current Epidemiology Reports*, 2(2), 131–142. http://doi.org/10.1007/s40471-015-0043-7
- Khorram, S. (2012). *Remote Sensing* (1st ed. 2012.). New York, NY: Springer New York : Imprint: Springer.
- Lazarus, M., Chandler, C., & Erickson, P. (2013). A core framework and scenario for deep GHG reductions at the city scale. *Energy Policy*, 57, 563–574. http://doi.org/10.1016/j.enpol.2013.02.031
- Moore, G. K. (1979). What is a picture worth? A history of remote sensing / quelle est la valeur d'une image? Un tour d'horizon de télédétection. *Hydrological Sciences Bulletin*, 24(4), 477–485. http://doi.org/10.1080/02626667909491887
- Myneni, R., Hall, F., Sellers, P., & Marshak, A. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*. http://doi.org/10.1109/36.377948
- Nowak, David J., & Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the USA. *Environmental Pollution*, 116(3), 381–389. http://doi.org/10.1016/S0269-7491(01)00214-7
- Nowak, David J., Crane, D. E., & Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry & Urban Greening*, 4(3), 115–123. http://doi.org/10.1016/j.ufug.2006.01.007
- Nowak, David J., Greenfield, E. J., Hoehn, R. E., & Lapoint, E. (2013). Carbon storage and sequestration by trees in urban and community areas of the United States. *Environmental Pollution*, 178, 229–236. http://doi.org/10.1016/j.envpol.2013.03.019
- Nowak, David J., Stein, S. M., Randler, P. B., Comas, S. J., Carr, M. A., & Alig, R. J. (2010). *Sustaining America's Urban Trees and Forests* (p. 28).
- Planet. (n.d.). Calculate an NDVI in Python: Learn to calculate a vegetation index to assess the quality of vegetation within an area of interest. Retrieved from <https://developers.planet.com/planetschool/calculate-an-ndvi-in-python/>
- Planet imagery product specifications. (2022). Retrieved from https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_

- screen.pdf
- Portland Urban Forestry. (2019). Inventory findings 2019. Retrieved from <https://www.portland.gov/trees/get-involved/tree-summit>
- Robertson, G., & Mason, A. (Eds.). (2016). Assessing the Sustainability of Agricultural and Urban Forests in the United States. *USDA Forest Service FS-1067*, 75 Pp., (FS-1067). Retrieved from <http://www.fs.usda.gov/treesearch/pubs/52278>
- Safford, H., Larry, E., McPherson, E. G., Nowak, D. J., & Westphal, L. M. (2013, August). Urban forests and climate change. Retrieved from <https://www.fs.usda.gov/ccrc/topics/urban-forests>
- Sander, H., Polasky, S., & Haight, R. G. (2010). The value of urban tree cover: A hedonic property price model in Ramsey and Dakota Counties, Minnesota, USA. *Ecological Economics*, 69(8), 1646–1656. <http://doi.org/10.1016/j.ecolecon.2010.03.011>
- Schroeder, Herbert W. (2011). Does beauty still matter? Experiential and utilitarian values of urban trees. In: *Trees, People and the Built Environment. Proceedings of the Urban Trees Research Conference; 2011 April 13-14; Edgbaston, Birmingham, UK. Institute of Chartered Foresters*: 159-165., 159–165. Retrieved from <http://www.fs.usda.gov/treesearch/pubs/41349>
- Schroeder, Herbert W., & Cannon, W. N. (1987). Visual quality of residential streets: Both street and yard trees make a difference, 5.
- Schroeder, H., & Cannon, W. N. (1983). The esthetic contribution of trees to residential streets in Ohio towns. *Undefined*. Retrieved from <https://www.semanticscholar.org/paper/The-esthetic-contribution-of-trees-to-residential-Schroeder-Cannon/76ab66281fa734ae7106cc7794310b5463cb8f77>
- Selbig, W. R., Loheide, S. P., Shuster, W., Scharenbroch, B. C., Coville, R. C., Kruegler, J., ... Nowak, D. (2021). Quantifying the stormwater runoff volume reduction benefits of urban street tree canopy. *Science of The Total Environment*, 151296. <http://doi.org/10.1016/j.scitotenv.2021.151296>
- Taylor, A. F., Kuo, F. E., & Sullivan, W. (2002). Views of nature and self discipline: Evidence from inner city children. *Journal of Environmental Psychology*, 22(1), 49–63. <http://doi.org/10.1006/jenv.2001.0241>
- Taylor, A. F., Kuo, F. E., & Sullivan, W. C. (2001). Coping with ADD: The Surprising Connection to Green Play Settings. *Environment and Behavior*, 33(1), 54–77. <http://doi.org/10.1177/00139160121972864>
- Tsoka, S., Leduc, T., & Rodler, A. (2021). Assessing the effects of urban street trees on building cooling energy needs: The role of foliage den-

- sity and planting pattern. *Sustainable Cities and Society*, 65, 102633. <http://doi.org/10.1016/j.scs.2020.102633>
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [http://doi.org/10.1016/0034-4257\(79\)90013-0](http://doi.org/10.1016/0034-4257(79)90013-0)
- United Nations, D. of E., & Affairs, S. (2018, May 16). 68. Retrieved from <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>
- US EPA. (2020, January 27). Urbanization and Storm Water Runoff. Retrieved from <https://www.epa.gov/sourcewaterprotection/urbanization-and-storm-water-runoff>
- USDA Forest Service, N. R. S. (2011). Urban Tree Canopy Analysis Helps Urban Planners With Tree Planting Campaigns. *NRS Research Review*, 13, 6.
- Ward, K. T., & Johnson, G. R. (2007). Geospatial methods provide timely and comprehensive urban forest information. *Urban Forestry & Urban Greening*, 6(1), 15–22. <http://doi.org/10.1016/j.ufug.2006.11.002>
- Xiao, Q., & McPherson, E. G. (2005). Tree health mapping with multispectral remote sensing data at UC Davis, California. *Urban Ecosystems*, 8(3-4), 349–361. <http://doi.org/10.1007/s11252-005-4867-7>