ESTIMATING DIAMETER AT BREAST HEIGHT OF DOUGLAS FIR TREES USING DRONE IMAGERY

By

Bryan Mohlman

A thesis
submitted in partial fulfillment
of the requirements for the degree of
Master of Environmental Studies
The Evergreen State College
June 2023

This Thesis for the Master of Environmental Studies Degree

by

Bryan Mohlman

has been approved for

The Evergreen State College

by

Kevin Francis, Ph. D.

Member of Faculty

Date

ABSTRACT

Estimating Diameter at Breast Height of Douglas Fir Trees Using Drone Imagery

Bryan Mohlman

Forest health, composition, and density are key data that help researchers and governments build models for future change. Collecting inventory data has high costs in terms of manpower and time, so using new tools to work more efficiently will be crucial to future restoration and management goals. Remote sensed data collected with unmanned aerial vehicles (UAVs) such as satellites and drones show potential in aiding future analysis efforts; however, much research still needs to be done on the reliability and accuracy of these methods. This thesis provides a foundation on which future projects can be built. In this thesis, fifteen plots were randomly selected, and the trees within them were marked with a GPS device, tree height was measured, and diameter at breast height was collected. Regression models were compared and analyzed as to their ability to predict the diameter at breast height of Douglas fir trees using both ground-collected data and remotely sensed data from a Mavic 2 Pro drone. Support vector regression was the top-performing model across all configurations of data. This indicates that reduced consistency and accuracy of height can cause these models to underperform. Future missions should include methods to improve image quality. With improvements in mission plan parameters, software processing, and adoption rates, UAVs have a bright future in ecosystem analysis and management.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my sincere gratitude to Professor Mike Ruth for his invaluable expertise, guidance, and unwavering support throughout the execution of the drone missions. His extensive knowledge and dedication were instrumental in the successful implementation of this research. Without Mike’s support this project would not have been possible for me. I am also grateful to my thesis reader, Kevin Francis, for his meticulous editing and insightful suggestions that greatly improved the quality of my thesis. His feedback and expertise were invaluable in shaping the final product.

I would also like to thank my family for the support they provided for me over the last two years as none of this is possible without them. They afforded me a strong foundation that ensured the least amount of external stress possible. They constant support and encouragement through not only the last two years but my entire life has meant the world to me.

Lastly, I want to thank my wife for her consistent support, understanding, and patience during the past two years of my master's degree. Her encouragement and belief in my abilities have been a source of strength and motivation throughout this challenging journey. I am truly grateful for her love and support.

Introduction

1.1 Summary of Problem

In the face of anthropogenic climate change, forests and urban green spaces are increasingly important areas we need to better monitor and understand. Forest biomes can sequester carbon, clean the air we breathe, filter rainwater, reduce erosion and flooding, mitigate the heat island effect, and much more that has yet to be fully understood (U.S. Department of Agriculture Forest Service, 2019). These benefits hold true across rural and urban areas as well as entirely undeveloped spaces (Fryd, 2012). The effects of climate change on the Pacific Northwest endanger ecosystem services that are used by all. Dryer hotter Summers along with wetter Winters shift forest growth patterns and habitability.

One asset that is vital to informed action is data collection and analysis. To prepare for major changes, scientists have developed models which estimate global and regional changes. While built on solid foundations, these models should be tested and rigorously verified. The quickly changing nature of forest ecosystems in the face of climate change is a challenge for adapting new recovery plans, and one of the most important things we can do in the face of this issue is to gather accurate information quickly.

Historically in the United States the most up to date data come from the Forest Service's Forest Inventory and Analysis (FIA), conducted on federal lands with a slowly increasing frequency. The first full forest inventory took thirty years to cover the entire United States (Shaw, 2008). Since then, the time to complete a comprehensive forest inventory has reduced significantly to a 5-year cycle (Shaw, 2008). Even this, however, is not fast enough to properly advise how we should focus conservation and restoration efforts. Monitoring the composition, health, and size of forests is key to proper management and restoration.

The necessity of increased frequency of recording and monitoring is greater than ever. In fact, in 2020, the demand for consultations increased by over 50% for the FIA (Alvarez, 2022). This shows not only the interest of researchers and state/federal governmental bodies but also the desire from private interests to better understand our forests. Increasing the frequency of these inventories is not an easy task. As described above, forest inventories and analyses take large amounts of time and labor to complete making it difficult to increase the “temporal resolution” or update frequency. One possible solution that could aid in this effort is the use of unmanned aerial vehicles and remotely sensed data. With proper development, this technology could help us gather the up-to-date information that is needed to better inform restoration plans, silviculture practices, and even update climate models for increased accuracy.

1.2 Situating My Research in the Current Landscape

Drones and the technology surrounding them are becoming more affordable and accessible, and every day new information is revealed about their utility in remote sensing analysis (Dainelli et al., 2021). With this increased accessibility researchers can build foundational tools for future remote sensing endeavors. The validation and testing that we do now will enable workers in the future to run missions and efficiently collect data that will inform management decisions on a variety of sectors and potentially different biomes. With further development in time, remote sensing will only get more powerful in its ability to aid in our understanding of forest growth amid a fluctuating climate.

The focus of this project lies with the estimation of the diameter at breast height of Douglas fir trees. Recent studies have attempted to tackle this very issue with Japanese Cypress tree as the sample species. For the estimation of DBH, linear regression has been shown to have varying degrees of accuracy depending on the species of tree and variables used in the analysis (Iizuka et al., 2018, Machado et al., 2019). The primary variables used in this estimation that have been shown to increase predictive ability are the tree's height, crown width, and crown area (Iizuka et al., 2018). Iizuka et al. also compared different methods of estimation between a linear model, logarithmic model, sigmoidal model, and a type of machine learning called support vector regression based on the support vector machine learning algorithm (Iizuka et al., 2021). This methodology is promising and should be tested on other species to determine extrapolation potential.

In addition, machine learning and its' subset, deep learning, show great potential in DBH estimation/prediction (Iizuka et al., 2021, James et al., 2022). Machine learning has the added benefit of typically stronger predictive ability when dealing with non-linear data. This gives machine learning an edge when compared to linear regression. Deep learning is often used for image and audio analysis; however, it is still strong with regression problems, both linear and non-linear alike (Asiltürk, & Çunkaş, 2011).

For restoration and forest recovery, understanding the species composition, health, and age of the forest is very important. Iizuka posits that monitoring forest compositional changes can give us a better understanding of the health of a forest stand (Iizuka et al., 2021). Increasingly researchers are seeking more granular temporal resolution of forest growth, which drone imagery can support.

1.3 Project Goals and Methods

In this thesis I assess different methods of tree metric estimation using drone imagery and prediction models. The goal is to determine whether drone imagery can be used to predict the DBH of Douglas fir trees with tree height and crown area data that has been calculated with remote sensing. DBH is an important tree metric that is heavily tied in with age, biomass, carbon sequestration, and more. Throughout this thesis I identify knowledge gaps and research needs, alongside the current body of research related to forest inventories, drone image capture, and estimation. While this is a small piece of the overall puzzle of forest measurement and assessment, it is a crucial starting point that can be used for future research and analysis.

Analysis will follow to determine if drone imagery can be used to accurately assess the dbh of the dense population of Douglas Fir trees (*Pseudotsuga menziesii*) found within LBA park wooded area in Olympia, WA. I completed this thesis project by focusing on two major questions.

- How do estimation models such as linear regression, support vector regression, or deep learning differ from ground truthing in terms of the accuracy of DBH prediction?
- Do the Iizuka et al. 2018 and 2021 methods of multivariate linear regression and support vector regression for DBH estimation on Japanese Cypress maintain their validity when used on Douglas Fir trees?

As mentioned above, studies have been conducted showing very promising results in the estimation of DBH using height and crown area with Japanese Cypress trees as the sample species. For this project I used a similar methodology to estimate the DBH of Douglas fir and Western Red Cedar (*Thuja plicata*) and Big Leaf Maple (*Acer macrophyllum*) found within my sample plots even though these two species did not have enough individual trees for robust statistical analysis.

To accomplish these research goals, three main tasks were undertaken. The first area of data collection begins with ground truthing. This is a process by which “true data” will be hand measured on-site to set a standard that will help evaluate predictive models’ strength. I produced an equal area grid of LBA park and categorized the cells by their 2018 aerial imagery. The categories were “primarily coniferous”, “primarily deciduous”, and “mixed forest”. From that classification I randomly selected plots from a proportional stratified sample. In each plot every tree of the three significant species mentioned above were measured for their height and DBH along with their GPS location data.

Next drone imagery was collected over three separate mission days to form ortho mosaics, digital surface models, and digital terrain models used in elevation estimation. Using this drone imagery, I segmented the canopy by manually drawing crown polygons in ArcGIS pro. Automatic watershed segmentation in R was rendered unusable due to weather conditions affecting elevation models. With geoprocessing tools in ArcGIS pro I was able to create a table of ground-truthed height, remotely sensed height, crown area, and species.

All of the data gathered is then input into various models for training and prediction. In this last stage, I compare the linear analysis to Support Vector Regression modelling to predict the diameter of a tree at breast height. In theory, if DBH estimations come close to reflecting the ground-truthed data, then we can use this inferred information to assess measured trees’ environmental benefits using a program such as I-tree and use this to help the community better understand the environmental services forests provide.

2.0 Literature Review

2.1 Introduction / Road Map

The United States uses tools for federal lands like the Forest Inventory and Analysis or FIA to assess the state of our federal forest lands. These assessments have changed in intent over the years from fully resource-focused to also incorporating environmental sustainability. The main issue with these FIAs is the time it takes to complete, up to years depending on the scale of the project. Field professionals say that one way to reduce time in the field for ecosystem analysis is the use of aerial imagery and remote sensing. Personal drones, aircrafts, and satellites have come a long way in the last 40 years when it comes to affordability, spatial resolution, and image quality. This growth has set the stage for remote sensing technologies to be a vital tool going forward for forest analysis.

Along with this advancement in the devices we use to capture imagery is the growth of new ways to analyze and estimate highly valuable data. While these estimation tools are still in the early phase of development and testing, they show great promise in analyzing forest stands quickly and accurately. First, I will broadly discuss the origins of a Forest Inventory and Analysis report (FIA) conducted by the Forest Service to understand where we started as a country with forest analysis. Next, I will examine the current state of drone use and recent advances in in this technology, certain advancements that have taken place, and allometric relationships. Following a discussion of allometric relationships, specific techniques used to segment canopies and estimate tree metrics using drone footage will be compared to demonstrate their pros and cons. In the final section I show remote sensing techniques using drone images can provide useful estimates of DBH, and discuss areas of further testing to be done in the world of deep learning for image interpretation.

2.2 Forest Inventory and Analysis

Brief History of FIA

Until the early 1900s our forests were not a priority for the federal government; however, as cities grew along with the demand for building materials, it was essential to have a general count for our stockpiles (Shaw, 2008). To provide the inventory, the U.S. government required the Forest Service to perform what was previously called a Forest Survey, later renamed Forest Inventory and Analysis or FIA (Shaw, 2008).

In a broad sense, forest inventories are a way that we, as researchers, can quantify various aspects of a forest. From these inventories, estimates are calculated, such as species composition, density, age, etc. These different estimates can inform decisions for managing restoration and natural resource extraction.

Methods / Shareholders

Forest inventories and analyses adhere to standardized and structured protocols. In the US Forest Services' typical protocol the first step is a double sampling design is built with plots on a grid, after which a 3-phase design is implemented (Shaw, 2008). Regarding the size of inventory sample space, there is a general rule of thumb whereby 20% of an area should be sampled if the total land is under 300 acres and 10% of the total area if larger than 300 acres (Hassani, 2018). In the first phase, photography is acquired, and forest stratification takes place (Shaw, 2008). This stratification classifies land in the imagery into two categories forest and non-forset. Phase 2 designates subplots based on stratification, and these "P2 plots" are further subdivided into four plots (Shaw, 2008). Finally, phase 3 plots or forest health monitoring plots are sampled from the P2 plots to analyze health indicators (Shaw, 2008). These health indicators can include the following: “woody material, soils, lichens, crown conditions, ozone, and more” (Shaw, 2008). A small percentage of these stratified plots is directly measured, and then results are generalized based on the strata. This is the process for a comprehensive forest analysis for state resource assessment, but data gathering does not need to be on such a grand scale to have value. Individual counties, cities, and even personal properties may seek scaled-down versions of this process to quantify their land resources and benefits (U.S. Department of Agriculture Forest Service, 2019).

Benefits/summary of FIA

When the first Forest Survey was conducted it was a time when the focus on inventory for the USA was centered on trees as a resource for extraction, more so than ecosystem benefits such as carbon storage and habitat (Shaw, 2008). With a changing climate and loss of ecosystems, the benefits provided by forest stands have become increasingly desirable to grow and enhance. There are a multitude of benefits involved in conducting forest inventory. These benefits in part account for our national desire to invest time, money, and manpower into inventory projects.

The Forest Service describes the dichotomy in need as scientific versus societal however they note that science is a subset of society and thus not entirely separable (Shaw, 2008). These reports are typically used in the scientific sphere to further other research goals. These goals could include producing density management programs, risk assessment, temporal analysis of forest health, species distribution, and more. On the local level, tree assessments focus on the needs of the nearby populace, whether that be mitigating heat island effects, increasing equitable distribution of green space in highly populated areas, increasing air quality, or mitigating weather events through bolstering stormwater management systems (U.S. Department of Agriculture

Forest Service, 2019). The question now is how to collect data on a local level quickly and cost-efficiently.

The first Forest Survey, which began in 1932, took 30 years to complete (Shaw, 2008). Since then, the time to complete a Forest Survey/FIA has decreased, initially dropping to about 13-15 years by the 1980s, then to a 10-year cycle in the 90s, followed by a 5-year cycle as we neared the 2000s (Shaw, 2008). Even with this rapid decrease in completion time, a more frequent analysis is needed. The United States' urban advancement and general societal goals require up-to-date information for intelligent land use decisions, forestry conservation goals, and more. FIAs only focus on permanent forested land, which can leave many data out. Smaller-scale operations for counties, cities and properties still require significant labor. However, as they are situated on land that is potentially easier to consume with urbanization, updated information on these spaces is vital. This is where drones can be an invaluable asset.

2.3 Drones in Forestry

Drone technology is more advanced and affordable than at any time in the past, and the remote sensing data they can provide is emerging as a vital tool for forest and land assessment (Dainelli et al., 2021a, Surový, & Kuželka, 2019). Remote sensing is the technique by which aerial vehicles and satellites can detect and monitor the characteristics of an area using calculations based on reflected light (USGS, n.d.). The remote sensing data can be used in various monitoring tasks such as tracking temperatures globally and noting changes, mapping forest fire activity, tracking urban growth, and more (USGS, n.d.). With specialized lenses that specify recorded wavelengths of light, we can "see" imagery on the light spectrum that would typically be outside of our human ability (Surový, & Kuželka, 2019). With near-infrared light waves, forestry workers can estimate how healthy a chosen forest stand is. One can quickly expand this process to agriculture, where quick snapshots of crop health and nitrogen absorption give farmers valuable information for crop management. For that information to be useful for farmers or smaller property owners, the spatial and temporal “resolution” would need to be quite high, and that is where personal drones come in. The next two sections will focus on first the pros of unmanned aerial vehicles followed by the cons.

Pros of Unmanned Aerial Vehicle use in Forestry.

Small unmanned aerial vehicles (sUAVs) have benefits not yet seen in satellite and aircraft imagery. Personal drones have many benefits one cannot easily attain from the other options. Satellite imagery produces data on a global scale and does so very well; however, the resolution can be quite low, around 10+ meters per pixel. Resolution typically refers to the number of pixels on a screen, and while the resolution for spatial imagery is still dealing with detail, it’s a little different in derivation. As it relates to drone photogrammetry, the resolution is connected to a unit called a ground sample distance (GSD) (Joyce, 2022). Since this imagery is connected to spatial data, each pixel is related to a specific spatial dimension. Screen resolution becomes more detailed as it is increased, so to does spatial resolution. Depending on the system used, higher-resolution drone imagery can get down as low as multi-centimeter levels of accuracy. Satellite imagery that comes anywhere close to that becomes very expensive. Temporal resolution is an additional consideration. Luckily with drones, barring poor weather conditions, we can effectively use whatever frequency of image capture is needed for the temporal analysis (Xiang et al., 2019). Because of this flexibility, updating data from drones can be thought of as real-time. Larger aircrafts are much harder to fly on demand, and satellites rely on a specific orbit schedule reducing their availability.

Cons of Unmanned Aerial Vehicle use in Forestry.

Even with their benefits, small unmanned aerial vehicles are imperfect and bring specific limitations. sUAVs suffer from the constraints of their hardware. This is seen in three main areas: battery life, weather restrictions, and inability to see past the canopy into the under forest (Dainielli et al. 2021a). According to the DJI Mavic battery specifications, one can expect around 46 minutes for a full charge flight. Time constraints and the need for additional batteries limit the total area that can be covered with a personal drone, making them less effective for larger areas. At the level of 5 hectares studies show that in 2015 costs in euro for satellite imagery are around 2650, for aircraft imagery it is 2450, and the cost for sUAV imagery is closer to 2200 (Matese et al. 2015). Compare this to a 50-hectare scale and we see the numbers reverse in cost efficiency where we see satellites cost around 2650 still, aircraft cost increases to 3800, and UAV imagery jumps the most to 5300 Euros (Figure 1) (Matese et al. 2015).

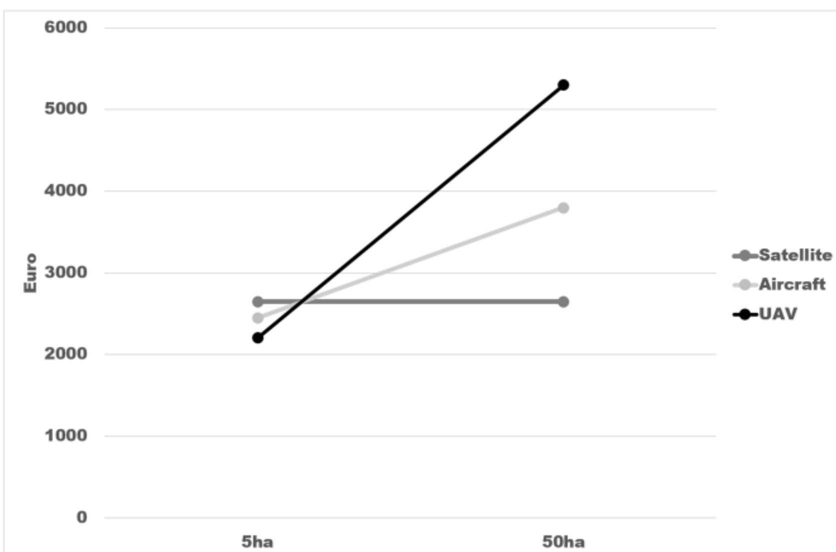


***Figure 1*:** Comparison plot of imagery costs in Euro between Satellite, Aircraft, and UAV missions. On the left we see the 5-hectare cost with the 50-hectare cost on the right. (Figure 7 in Matese et al. 2015).

All but high-end waterproof personal drones are prone to damage if water is taken into the motors; thus, inclement weather limits many drones' flight readiness. Satellites can be limited by weather constraints as well with cloud clover being a limiting visual factor, however, drones can fly under cloud cover and are therefore more adaptable to certain situations. The last major downside of drone imagery is the inability to penetrate the under canopy where significant amounts of data lie (Dainielli et al. 2021a).

The increased affordability and availability of personal drones are relatively new in the research space providing challenges that must be overcome in the near future. As with many emergent technologies, the distribution globally is inequitable, and many areas that could benefit from their research need more resources, such as South America and Africa, comprising less than 14% of studies conducted on remote sensing (Dainielli et al., 2021a). Much research must be undertaken to test the accuracy of the remote sensing ability of drones under different conditions and with other models. In the next section, we will take a step back and examine the foundations for tree metric estimation, and the basis for some of the remote sensing predictions this project will attempt to employ.

2.4 Allometric relationships

Thus far we have taken a detailed look at forest inventories and the pros and cons of drone use within forestry. The main goal of this project is to predict DBH of Douglas fir trees using its relationship to tree height and crown area. This section will delve into allometric formula fundamentals and the basis of proportional relationships in organisms.

Allometry is the growth of body parts at different rates, resulting in a change of proportions. The main idea of this concept being that there are generally formulas that can, with reliability, use body parts as input criteria for a formula to predict other body parts size. The idea of using measurements of the different sections of an organism to make inferences has been around for a long time having been coined “allometry” in the mid 1930’s (Gayon, 2015). Even though the moniker was finalized in 1936 the concept had been percolating through scientific circles for decades prior (Gayon, 2015). While the early research was focused on insects and animals, forestry has come to use allometric relationships to make educated estimations of many tree metrics such as tree volume, biomass, and carbon storage. In 1996 a study from the Netherlands used linear regression analysis to test strength of correlation between stem, crown dimensions, biomass, and needle area (Bartelink, 1996). The authors found very strong relationships between DBH and crown length, crown area, and tree height.

It should be noted that there are factors that affect these relationships (Bartelink, 1996). The researchers collected data from samples that were in a mono-species forest stand, and the soil was relatively uniform across all collection sites. If, however, factors like species, or climate were different, we could not make the same assumptions of correlation (Mencuccini and Grace, 1995). Scots pine trees that were measured in two sites, one in England and one in Scotland, were found to be significantly different even with similar ages (Mencuccini and Grace, 1995). Broad patterns in allometric relationships are clearly present, but each species has a unique profile. Even without a universal model, a specially tailored model for individual species and growth zone can show effective predictive ability. In the next section, we will spotlight key segmentation and metric estimation examples using drone data.

2.5 Estimating Tree Attributes

Drone utility in forestry is highly tied to the ability to measure forest attributes quickly and accurately. Researchers can estimate tree metrics using UAV photogrammetry with varying levels of reliability as compared to the ground truthed data. The reliability of the drone image assessment determines how valid the extrapolations of this data can be. While the assessment methods can be sound, the final product will only be successful if the data going in is accurate. Being able to reproduce the results of drone analysis is paramount. Thus, I will start by detailing two systems for automatic segmentation and classification, describing how they aid in identifying trees within images. Finally, I will illustrate and compare methods of estimating the diameter at breast height of various tree species using linear regression and machine learning.

R –Packages for Computing Canopy Segmentation

There are two main automatic segmentation methods in development using unmanned aerial vehicle (UAV) imagery. The first method is the use of spatial packages in R and with the second method being deep learning mask regional convolutional neural networks (Mask R-CNN). First it is important to understand the R infrastructure more thoroughly and how it has been adapted to work with spatial data analysis. R was initially designed based on the S programming language that conveniently ran code without a compiler for statistical analysis (Ihaka and Gentleman, 1996). When coding in R, one manages data using code commands and packages. Since the inception of R, many coding packages have been developed to address specific tasks, including spatial analysis, which can be applied in forestry. The primary packages used in interpreting geospatial data include, but are most certainly not limited to, lidR (Roussel and Auty, 2022; Roussel et al., 2020), raster

(Hijmans, 2022), rgdal (Bivend et al., 2022), ForestTools (Plowright and Roussel, 2021), and many more.

Many R packages can be used to analyze spatial data. Two packages will be highlighted in this section tied to aerial imagery and analysis LidR and Forest Tools. The LidR package was built to be used in conjunction with LIDAR imagery data or, as the authors describe it, Airborne Laser Scanning (ALS) Data (Roussel et al., 2020). ALS is a potent tool used to capture large amounts of geospatial data, including terrain, canopy, and vegetation structure (Roussel et al., 2020). This package seeks to harness the data of ALS imagery using "state-of-the-art algorithms" for interpretation (Roussel et al., 2020). LidR can create a canopy height model in conjunction with a. las 3d point cloud. The basic formula for a canopy height model is CHM = DSM-DTM, where DSM is the digital surface model, and DTM is the digital terrain model (figure 2). DSM captures both the land and objects on the land, including trees and structures, whereas the DTM attempts only to reflect the ground. LidR is a versatile package with many more uses, including tree detection and segmentation.

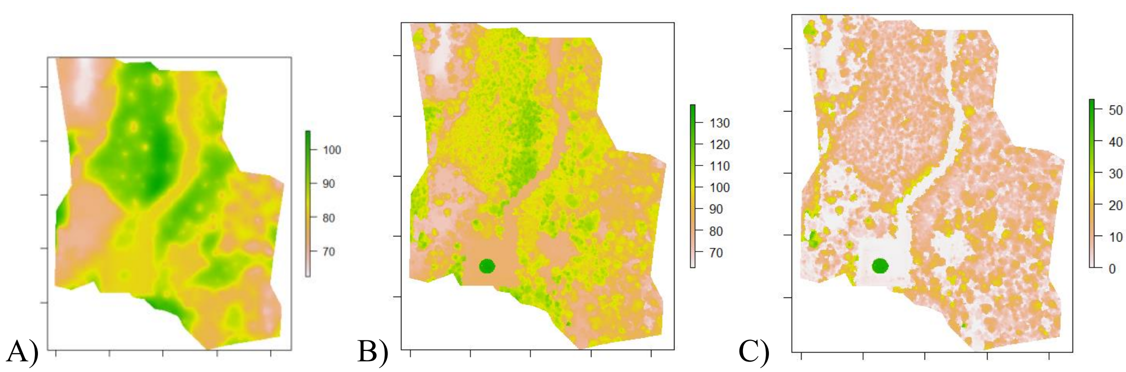


Figure 2: Visual representations of spatial data plotted in R: A) point cloud converted to DTM, B) point cloud converted to DSM, C) using R, the values of the DTM have been subtracted from the DSM, producing a basic canopy height model (CHM).

The second package to mention is ForestTools which was developed and maintained by Andrew Plowright. This package, similar to lidR, helps us analyze drone imagery. While the lidR package will process the las point cloud into usable DSM and DTM maps, we need the further ability to manipulate this data to help segment and classify the forest, which is where ForestTools steps in (Plowright, 2021). ForestTools can aid in detecting treetops using a variable window filter (VWF) algorithm. This algorithm uses a dynamically sized window to "scan" the canopy and assign points on the highest value within the window (figure 2) (Plowright, 2021). The next major function of ForestTools is to infer crown shape with a "marker-controlled watershed segmentation" algorithm employed using that newly gathered tree top data (Plowright, 2021). The watershed algorithm can have over-segmentation issues, increasing the misclassification of trees due to branches and abnormalities (Plowright, 2021). The treetops can be used as "markers" by which to classify individual trees and avoid this over-segmentation problem by only allowing a single "crown" polygon to be connected to each detected "treetop." The slight over-segmentation that can be found with the watershed algorithm counterintuitively has proven to have higher detection accuracy in more densely populated forest canopies when compared to other tree detection algorithms (Gu, 2020). This segmentation method has been shown to work well for further canopy analysis (Iizuka et al., 2018).

A) B)

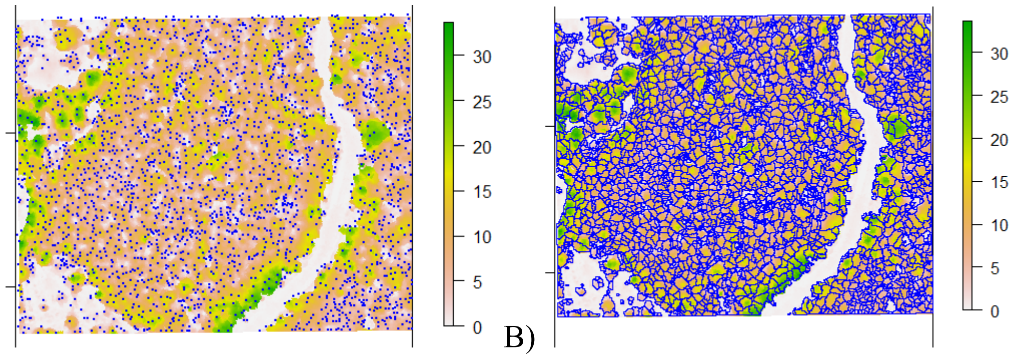


Figure 3: Using ForestTools, trees are identified in a subplot of LBA Park Olympia, WA. A) Treetops are mapped using the VWF, B) crowns are segmented using the MCWS

Mask R-CNN for Tree Segmentation and Classification

Mask regional convolutional neural networks have the potential to take what was seen with R further, but they are not without their drawbacks. First let's focus on what a Mask R-CNN is comprised of and how it can be utilized in image classification. A convolutional neural network is a type of deep machine learning primarily used to analyze imagery and detect objects (Odemakinde, 2022). The network is designed similarly to a standard artificial neural network; however, now there are convolutional layers, pooling layers, and flattening layers to effectively translate greater dimensionality to a model for classification (Odemakinde, 2022). Highly useful in spatial recognition, a CNN will analyze windows of image data and "learn" from the context of that window, I.E., the network takes the pattern around and within the specified window to tell the model something about the section of an image (Odemakinde, 2022). This identification process can be used in several ways, primarily semantic and instance segmentation. These two types of segmentation differ in a couple of important ways. Firstly, semantic segmentation attempts to classify every pixel within an image (Figure 4) (Wang et al., 2018). Instanced segmentation goes one step further by not only classifying the pixels but also defining specific instances of the classification. The value of this has been shown in forestry already, with studies showing relatively high accuracy of the model to predict tree density and classification (figure 5) (Sun et al. 2022).

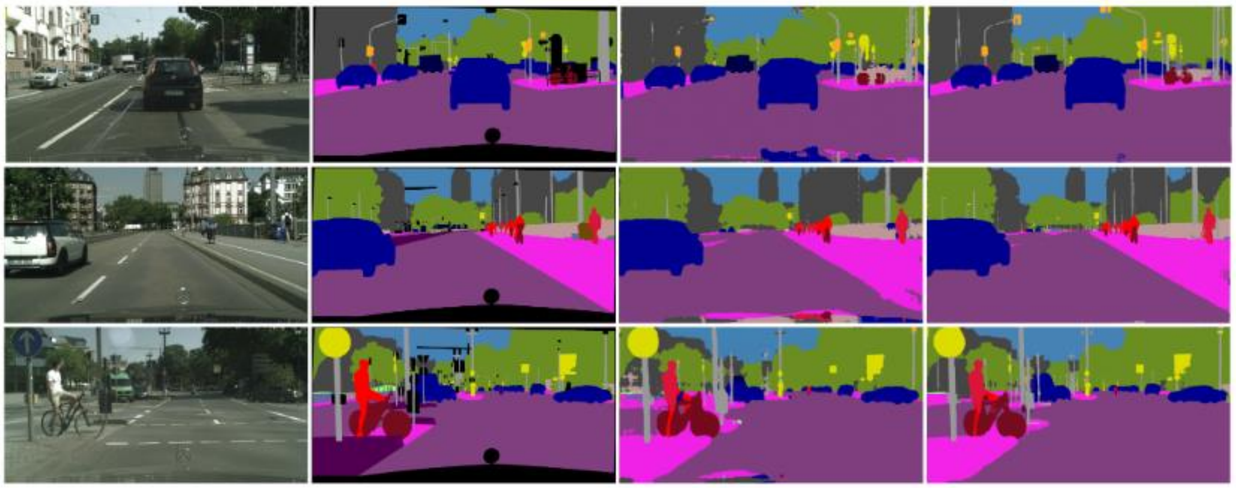


Figure 4: Example of semantic segmentation using a fully convolutional neural network by Wang et al., 2018. Notice how each type of object has its unique color associated with it.

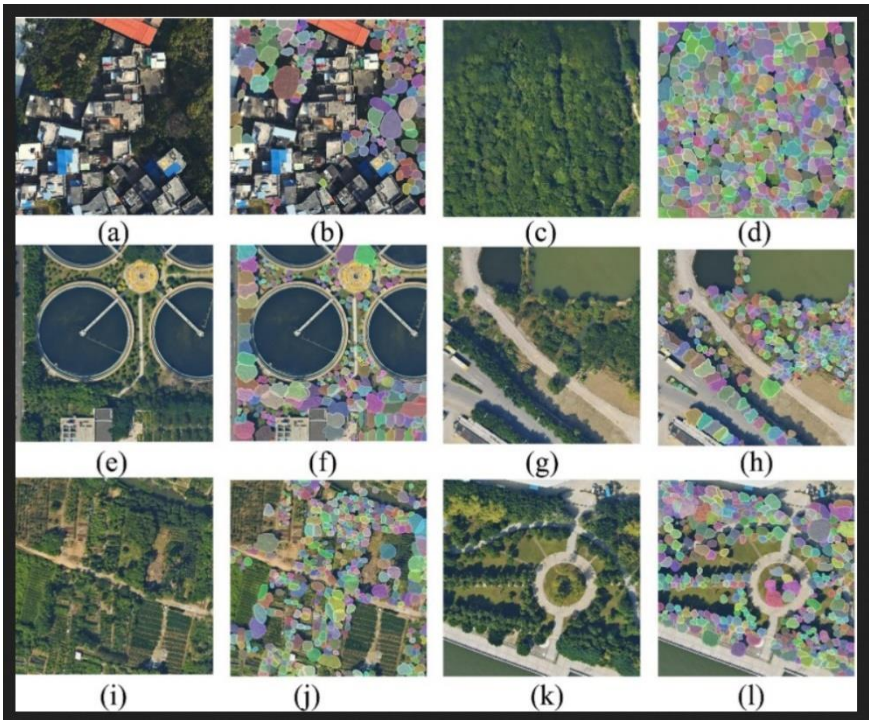


***Figure 5*: Example of instance segmentation employed on various images to detect individual trees as described by the source. Each pair of letters represents a different land use type. I.e., (a) and (b) represent urban land, (c) and (d) forests, and so on (Sun et al. 2022)

Watershed segmentation or mask R-CNN

Automatic segmentation is being developed for remote sensed imagery, this section will detail multiple variations along with their pros and cons. Watershed segmentation processes quickly once you have the canopy height model, whereas Mask R-CNN is very resource intensive for broad training. When a canopy is densely populated, watershed segmentation has been found to have higher accuracy than comparable algorithms (Gu, 2020). The resulting polygons contain both areas and height data gathered from CHM. The main downside to using only that system for analysis is the lack of ability to classify the individual trees. There is no real ability to classify polygons in this way. Mask R-CNN has been utilized in combination with segmentation recently.

In fact, studies have shown high levels of classification accuracy when a Mask R-CNN model was applied to the polygons produced by segmentation, with accuracy rates of over 90% (figure 6) (Onishi and Ise, 2021). Combing these methods or using a method with uncertain results is not without its drawbacks, and the added complexity may be out of the scope of a research project unless that is the focus of the project (Iizuka et al., 2020)

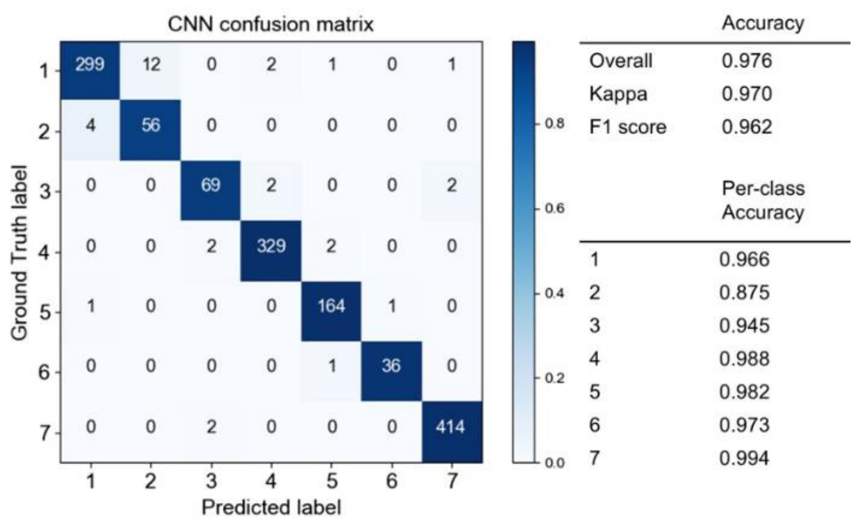


***Figure 6*:** This confusion matrix shows the classification accuracy of the Mask R-CNN model employed on seven different category labels (Onishi and Ise, 2021). With an overall accuracy score of 0.976 this was shown to be a very reliable method of classification.

2.6 Estimating Diameter at Breast Height (DBH)

Being able to accurately estimate diameter at breast height of trees has a multitude of benefits. DBH is not strictly valued for its quantification alone but rather its relationship to other aspects of the tree's health. From DBH, we can infer beneficial metrics such as biomass, carbon storage, age, and more (Snyder, 2006). When estimating the DBH, researchers focus on how to connect the best accurate estimation of DBH's relationship to other tree measurements. As discussed in the previous section, we gather height and crown area information from drone imagery, so these will be the primary data used in the inference process. The primary method used to predict DBH up until recently has been linear regression (Iizuka, 2021). In recent years with the advancement of computing ability on a mass scale, the use of machine learning techniques for prediction has begun to be tested for viability.

Linear Regression

Linear regression has been demonstrated to be a robust tool with which to estimate the diameter at breast height of trees (Iizuka et al., 2018, Bartelink, 1996). As discussed earlier in section 2.4, allometric relationships have used linear regression in tree metrics. However, in this case we are using remote sensed data. Since we are using a different data collection method it's worth questioning the strength of the relationship in this new use. In this case, the y variable or dependent variable would be the DBH. There are two possible X variables that will be used comprised of height and canopy area. Studies have shown that DBH strongly correlates with the crown's canopy area; however, this did not hold true for the height value (Iizuka et al. 2018, Iizuka et al. 2021). When looking at Japanese Cypress trees, Iizuka et al. found an R^2 value of 0.7786 and 0.7923 for canopy width and canopy area, respectively (Iizuka et al., 2018). If we compare this to the R^2 value 0.1784 for tree height as it corresponds with DBH, we see a starkly different picture (Iizuka, 2018). However, this doesn't tell the whole story, as there are times when certain variables are connected indirectly and can still aid in prediction. Tree height and crown diameter can be employed together in a multivariable regression model. When combined, their accuracy for the prediction can increase; one study showed an R^2 value of 0.9492 adjusted (Machado et al., 2019). Conversely, the same researchers stated that there can be no single linear regression model as different classes of trees do not always share the same connections in their parameters (Machado et al., 2019). This would partly explain the difference in R^2 value between Machado and Iizuka's studies, given that they evaluated different species. While part of the answer, this does not fully explain the variance. Machado et al. sampled a much smaller space and used a laser scanning device attached to their drone to produce a point cloud with increased accuracy. When assessing different forest compositions, it is apparent that one must consider species class, drone technology, and drone mission plan details such as the height of flight and speed.

Machine learning and Deep Learning for Prediction

Drone technology is advancing rapidly and becoming more accessible while simultaneously, our tools to estimate and analyze data mirror this evolution. For example, machine learning is an excellent example of this and has strong potential to aid in accurately predicting tree metrics like DBH (Iizuka et al. 2021). Machine learning is a subset of AI designed to tackle specific predetermined tasks and "learn" from experience (Zhou, 2021). Using machine learning, researchers can solve problems using data and algorithms with a robust ability to scale to the problem (Zhou, 2021). Machine learning presents a unique solution to the failures of more fixed human-designed models like linear regression (Deep learning, 2022). Deep learning is a further subset of AI within the machine learning category and is often used for less straightforward data processing like image analysis and natural language processing (Deep learning, 2022). These systems accomplish this processing by attempting to mimic the way a human brain creates patterns by transferring dynamically weighted data through "neurons." When there are multiple "hidden layers" within a network, it is considered deep learning. With a data set of varying species and growth conditions, linearity in the relationship of variables should not be automatically assumed. These AI tools can engage with data in ways that can better fit nonlinear models, as I will show below.

Support Vector Regression

Support vector regression or SVR machine learning shows promise in predicting DBH (Iizuka, 2021). SVR makes inferences about relationships in bivariate and multivariate modeling by effectively segmenting data into different regions through a “kernel trick” (James et al., 2022). A kernel trick is a function to analyze transformed data without actually transforming the original data. This mathematical simulation adds dimensionality to your data and produces more definable groups to aid in classification boundaries. Depending on the task you are trying to accomplish, there are a few different kernel types, such as linear, polynomial, and radial, also known as radial bias function (RBF), gaussian and more (James et al., 2022).

The authors used the SVR method due to its flexibility in handling linear as well as non-linear data. SVR produced a slightly more accurate model for prediction with an R^2 score of 0.758 than a linear model ($R^2 = 0.717$) and just barely edged out the sigmoidal model ($R^2 = 0.756$) (figure 7) (Iizuka et al., 2021). Not surprisingly, the R^2 score of the SVR model increased when it employed multiple variables. Even with the increase in performance from the linear model with both tree height and crown diameter, the SVR model with those same added variables surpassed it with an R^2 score of 0.787 and 0.834, respectively (figure8) (Iizuka et al. 2021). This consistent increase in results begs the question of whether SVR could be used in other species to gain improved model fit.

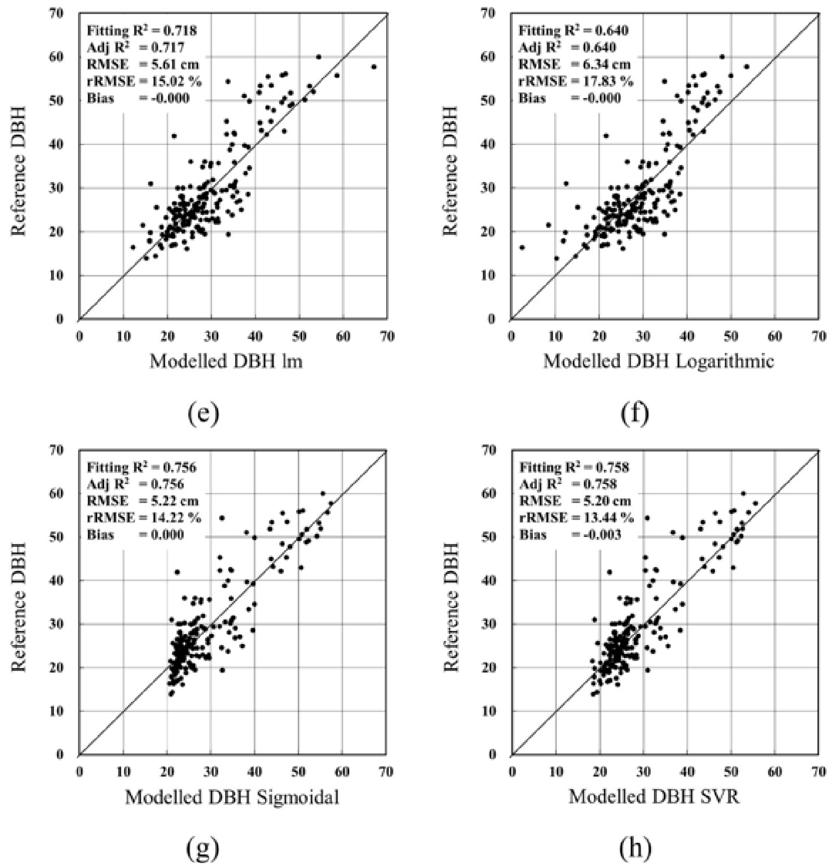


Figure 7: Scatter plot demonstrating fit of various model types to reference data (e)linear, (f)logarithmic, (g) sigmoidal, and (h) SVR (figure 6 in Iizuka et al., 2021).

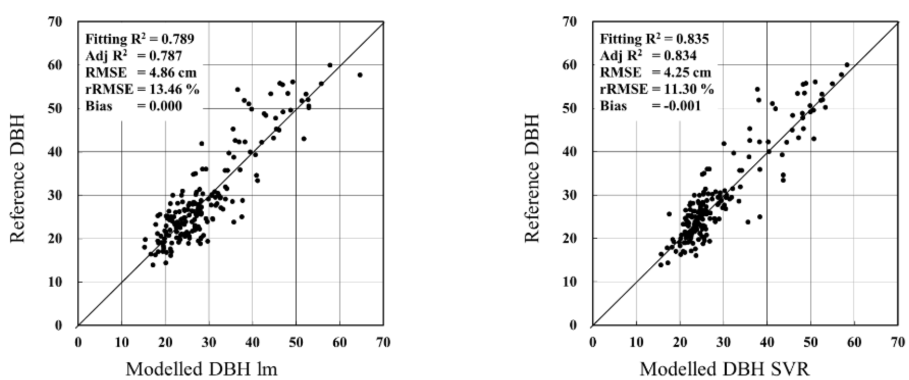


Figure 8: On the left is a scatter plot of the relationship between the reference and modeled DBH, with a line bisecting the points showing the fit of the model in question. On the left is the

multivariate linear model and on the right is the multivariate SVR model (Fig.7 in Iizuka et al. 2021). The SVR model appears to have an advantage when compared to the linear models.

Deep Learning AI

Deep learning, as described above, is a subset of machine learning often used to analyze complex data and infer relationships without the programmer explicitly dictating those relationships. These networks can be used on many types of data, for example, predicting stock market returns or, as we discussed earlier, using Mask R-CNN as a tool to segment forest canopies/object detection in images. While commonly used for image processing, object detection, and audio recognition, deep learning artificial neural networks or ANNs work well within regression-based problems. Because ANNs are relatively new tools for prediction and classification, research in specific fields, such as forestry, is limited or nonexistent.

ANNs have been shown to be helpful in regression-based problems, even if they have not been implemented within forestry for prediction. For example, in a test to predict the roughness of a surface based on cutting parameters like cutting speed, feed rate, and depth of cut, the ANN performed slightly better than the other regression models (R^2 = 98.9%) (Asiltürk, & Çunkaş, 2011). One note, though, about the prediction results is that the ANN performed better during the training stage (R^2 = 99.8%) but not as well in the testing phase (R^2 = 99.4%), which indicates that the generalizations that it develops may not always translate as well in the real world (Asiltürk, & Çunkaş, 2011).

Deep learning systems aren't always the correct choice for the data set being predicted. Deep learning has some drawbacks; as a researcher, one cannot assume its superiority. For example, when there are limited data sets, the training, and testing systems for deep learning struggle, often overfitting or underfitting the model to the data (Jiao et al., 2020). Additionally, deep learning models suffer from the complexity of their design. It is far simpler to demonstrate the reasoning behind a linear model compared to the "black box" mechanisms of a deep neural network (Jiao et al., 2020). The "black box" mechanisms refer to the stages between the input and the output because while we can have power over certain weights and activations used, the way that the AI system determines relationships is not directly controlled.

2.7 Summary/Conclusion

This review discussed many examples of drone technology and analysis that can and will be vital to future forestry efforts. Forest inventories are currently extensive laborious projects that require significant labor and time. Unmanned aerial vehicle technology has advanced over the last thirty years to the point that they can contribute to FIAs with spatial resolutions of a few centimeters and temporal resolutions of whatever the project requires, not to mention better pricing. In many ways, sUAVs show benefits over satellite and airplane imaging missions; however, personal drones possess downsides as well. Battery power limitations and weather constraints are chief among these concerns. Additionally, a common issue with any aerial photogrammetry is the inability to pierce the overstory canopy reliably.

Remote sensing data pulled from drones has been found to be reliable and can be used in further research projects. From drone imagery canopy height models can be produced allowing the segmentation of canopies through multiple possible methods. The methods explored in this review are the use of a watershed segmentation algorithm in R and object detection with Mask R-CNN AI. While both options have advantages and disadvantages and can even be used together, the reliability of reproducing results needs further research. Authors suggest that the watershed algorithm is reliable to employ and has improved results in dense canopies when compared to other tree top segmentation algorithms. This data can now be used to further estimate tree metrics.

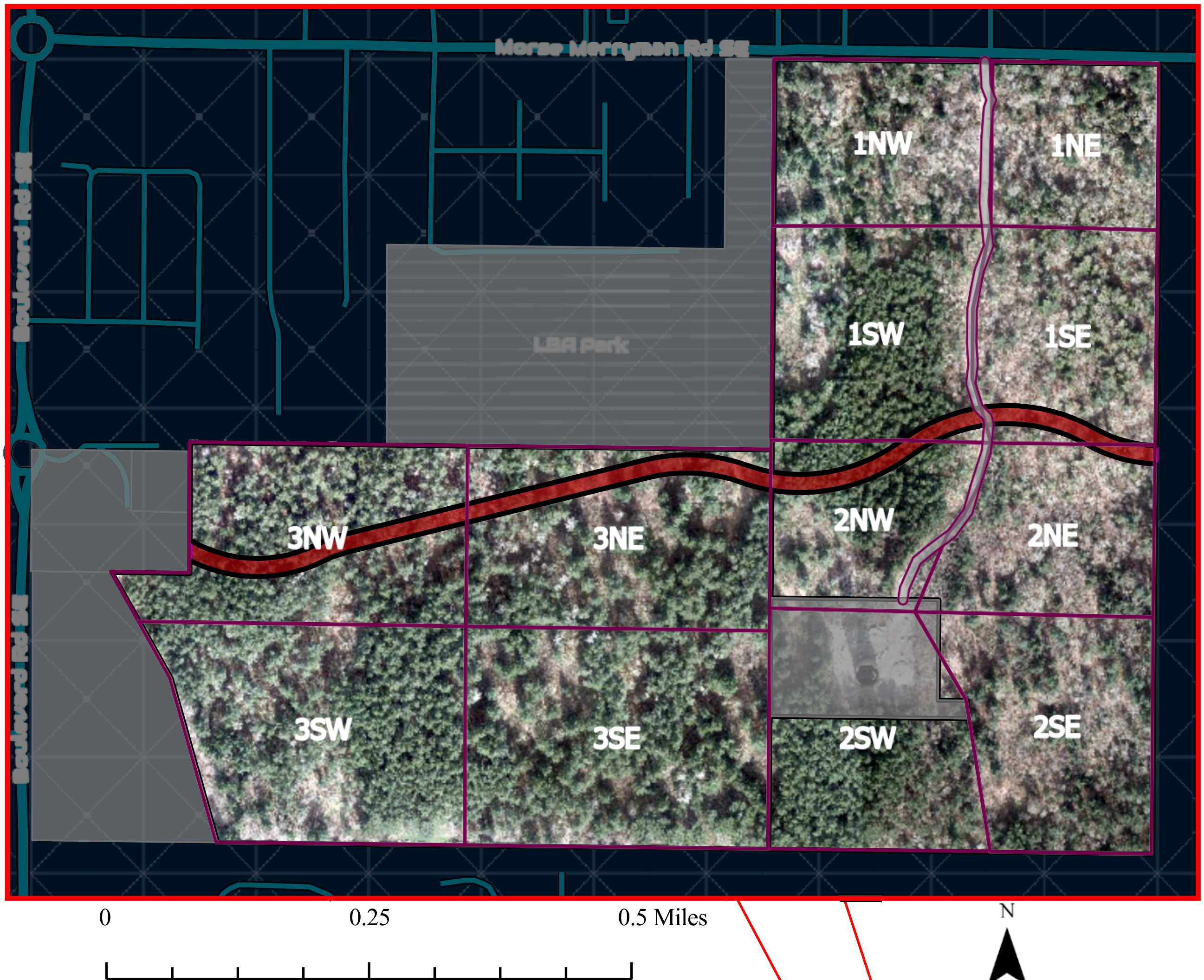
There are many methods to predict a tree's diameter at breast height which require further testing and analysis to determine viability. Due to different region's growth potential, species growth patterns, and climactic differences a general model for DBH estimation is nearly impossible to produce. This does not mean, however, that DBH can't be predicted, but rather a refined approach to species and area may be necessary to increase the validity of predictions. In previous studies methods using linear, sigmoidal, logarithmic, and support vector regression models have been used for the prediction of DBH. These models require multiple variables (tree height and canopy width/area) to achieve the highest R^2 scores. The relationships between different tree metrics are not always linear and thus the SVR model handily outperformed the competing models. The primary focus of these tests has been on Japanese Cypress trees and needs further research on other species for validation. Lastly, deep learning artificial neural networks can be used for regression-based problems even though typically they are used for image detection/recognition, audio processing, and other non-linear data. One of the strengths of Deep learning models is their ability to analyze non-linear data like the SVR models and thus could be suited to analyzing tree data.

Continued advancement and research are much needed in remote sensing technologies, to improve efficiency and accuracy. Better understanding the general application of prediction models, species and compositional differences, as well as weather conditions are key to future growth of this vital tool.

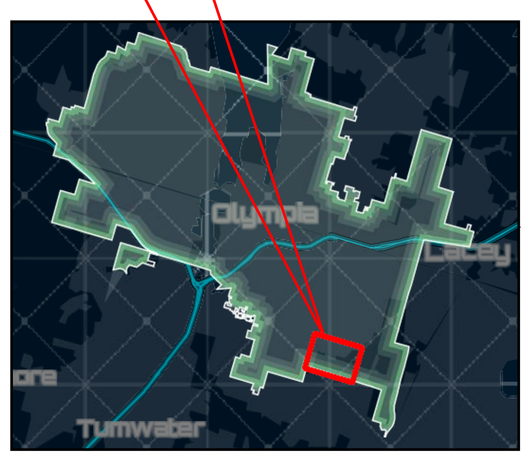
3.0 Study Design and Methodology

3.1 Study Location

Little Baseball Association (LBA) Park, Olympia Washington



This map of LBA park shows an overview of the park and its' boundaries, including the proposed road addition, water tower site boundary plus access road, and survey zones produced by habitat stewards.



LBA park has been selected to be the location of this experiment. Located between Olympia, Tumwater, and Lacey, this park was officially established in 1974 through the partnership of the Little Baseball Association (LBA) and the City of Olympia (Figure 1) (Olympia, n.d). Through this effort, the primary ball fields and playgrounds were established. At the time, however, plots of land surrounded the park that, as of then, were undeveloped. Due to the prime location and relation to the urban surroundings, housing development became a real threat to this forested space. Only through an unyielding effort from the local populace was this threat diverted. Eventually the City of Olympia purchased approximately 155 acres of forested land as an addition to the park. In 2016 Olympia purchased the Ashton Woods area, followed by the Bentridge parcel only a year later in 2017. Since then, the park's stewards have made a concerted effort to increase local engagement and ownership of the park through trail additions and community events.

LBA woods is an ideal site for my experiment. The reasoning behind this is threefold. Given the city and community desire to increase the park's long-term viability, it has become somewhat of a sandbox for experimentation, with easy access to resources and park space. Second, the woods have a variety of compositional characteristics that will help test the bounds of modeling choices. For example, there are areas within the woods that have very dense, large clusters of Douglas fir without disruption, similar to a tree farm. In addition, there are large areas of mixed forest with a good representation of both deciduous and evergreen trees. Trees grow differently based on many factors, and the forest composition is just one factor. A variety of available compositional structures for further analysis will provide a deeper insight into potential model generalization. Finally, the site is well-suited for drone imagery, with the water tower and main ball fields for launch zones and no major obstructions to mission planning. Given these three major criteria, accessibility, the composition of the forest, and drone use feasibility LBA has proven to be an excellent experimentation site.

3.2 Experimental Design Summary

This study examined the predictive ability of multiple models to estimate the diameter at breast height (DBH) of various trees. Primarily the focus was centered on the selected model's application in reference to Douglas fir trees (*Pseudotsuga menziesii*). Additional data was gathered to test the models' applicability on big leaf maples (*Acer macrophyllum*) and Western red cedar (*Thuja plicata*). The Independent variables in this experiment, tree height and crown area, were used to predict the dependent variable, which is the diameter at breast height.

DBH and tree height were collected via ground truthing. Past LBA park aerial imagery was segmented into multiple compositional categories. From these categories, a stratified sampling area was produced. Using simple random sampling within the strata, the sample sites were determined. In these sites, every tree of the specific significant species group mentioned above was measured. The original ground truthing gave me the dependent variable (DBH) on site measurements and the first independent variable (tree height). Next aerial images were captured of the test sites. This imagery was used to segment the canopy giving us our final independent variable (canopy area). The data gathered above was then input into a spreadsheet to use in building and testing the models.

To measure the various model comparative strengths, this project employed a regression analysis design. The metrics used to compare the strength of the individual models were the coefficient of determination (R^2, root mean square error (RMSE), and mean absolute error (MAE). These metrics have been selected so that the models can be analyzed from a few different angles.

R^2 is strictly focused on the question of the relationship between variables and how well the results can be explained by the model. Strictly speaking, it will not tell us how the model performs on novel information rather it focuses on the training data. MAE takes the mean of errors in our prediction sets and ensures that we receive the absolute value of the result. RMSE is similar to MAE in that it focuses on the mean error, positive or negative, in predictions; however, it is more sensitive to large outliers and will reflect them more acutely in the resulting score. These results were compared to see if outliers affected the prediction results greatly.

3.3 Experimental Methodology Detailed Explanation

Sample Site Selection and Stratification

LBA park has been previously subdivided into habitat survey zones by the park stewards. This produced 15 separate zones in total. I used this as a base for further segmentation; within ArcGIS pro, I used the subdivide polygon tool to produce a grid of even size. While these spaces are equal parts of each of the survey zones, they vary in size from a low of 20,118 square feet to a high of 20,985 square feet. The final count of spaces in the grid is 300.

I next categorized each of these zones by their compositional characteristics. The first option is “Primarily Coniferous”. This is defined for the sake of this study as being more than 75% populated by evergreen conifers. The second category I selected is “Mixed Forest,” comprised of any combination of coniferous and deciduous trees under the 75% threshold. Third, I used the category “Primarily Deciduous” as an opposite to the first category comprised of 75% or greater Deciduous trees. Lastly, I added a category for unusable space within the test that I called “Grass / Structure”. As the title suggests, these areas are found to be mostly comprised of grassland or built structures. With little in the way of usable data collection space, these areas were intentionally excluded from consideration.

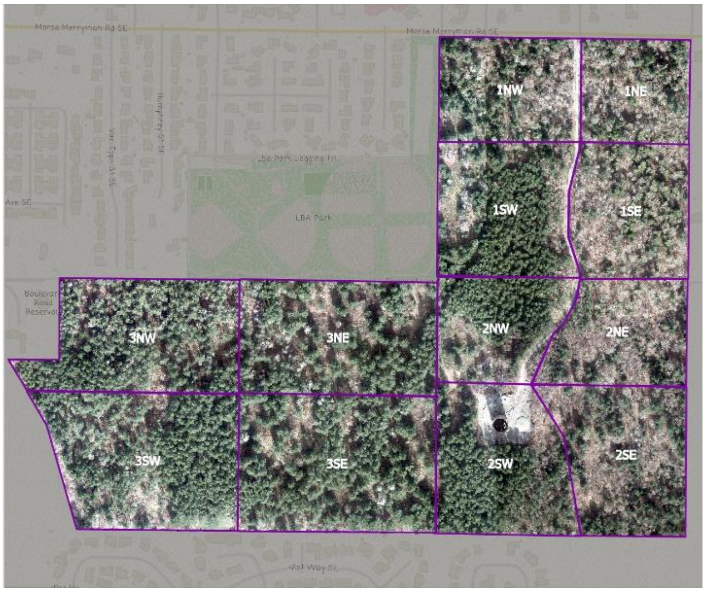


Figure 9A: LBA woods separated into survey zones by park stewards.

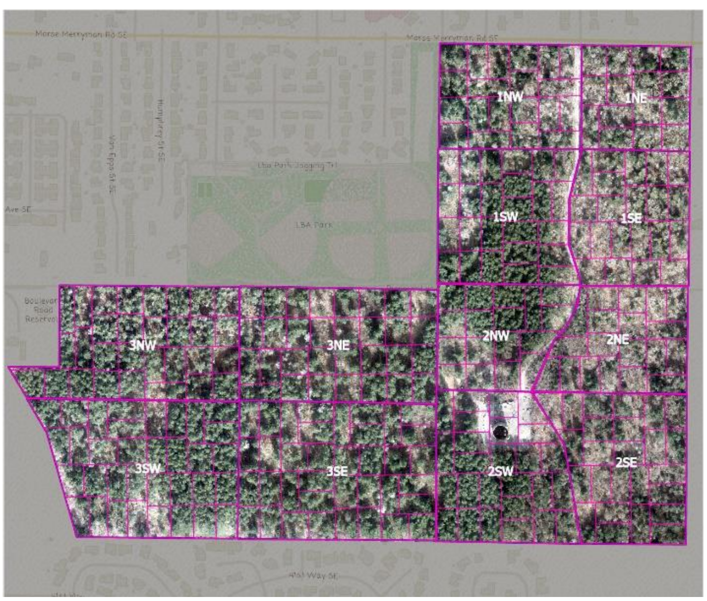


Figure 9B: LBA woods survey zones were further segmented for sample site selection using the equal area polygon tool in ArcGIS pro.

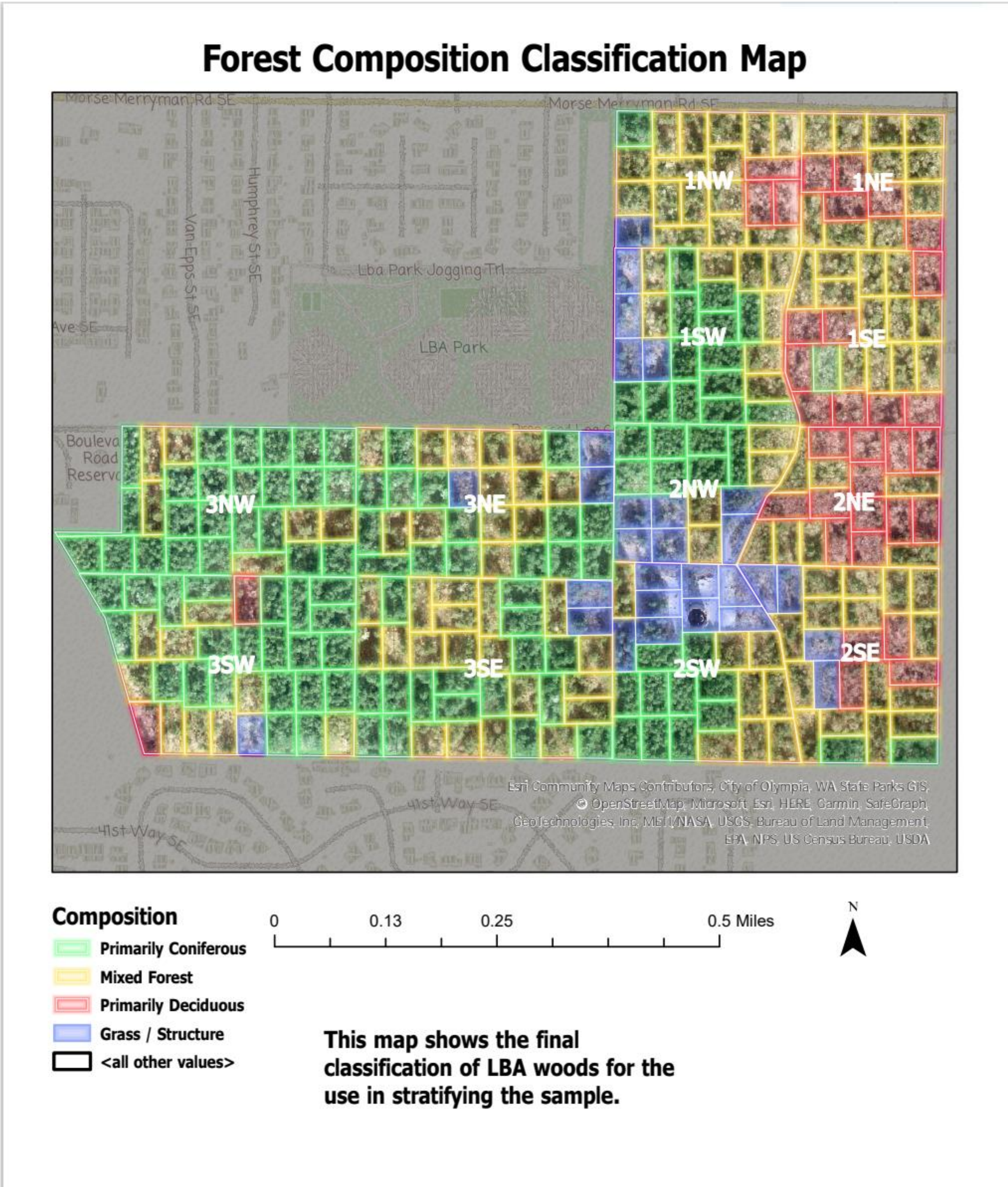


Figure 10: Full LBA woods grid post manual classification.

Based on this stratification the composition percentages are as follows: Mixed Forest – 42%, "Primarily Coniferous" - 37% of the total, "Primarily Deciduous" – 11 %, and finally, the "Grass / Structure" – 10 %; all percentages were rounded for simplicity. Next, to select the specific sites on this grid, simple random sampling was performed to attain a number proportional to the category specified (excluding Grass / Structure entirely). Each of these plots has an object ID automatically assigned by ArcGIS Pro.

As can be seen above in the stratified map, there is quite a variety of densities found within the grid. There is an approximate rule of thumb for the number of samples needed depending on the desired confidence levels (Figure 2) (Syuukaku, 2019). If we are also using this data for future research on other remote sensed metrics, it would be ideal to measure around 333 trees. This approximate sample necessity states that we can assume a 5% error rate for DBH, Tree Height, Single Wood Volume, Forest Stand Volume, and Stand Density with a sample size of 26, 11, 138, 225, 333, respectively (Syuukaku, 2019).

To assume Error rate	DBH	Tree height	Single wood product	Forest volume	Forest stand density
5 %	26	11	138	225	333
10 %	7	3	35	57	84
15 %	3	2	16	25	37

Figure 11: Table from Syuukaku, 2019. detailing the number of samples needed to assume error rate based on the metric desired.

Simple random selection was now applied to the stratified sample to attain fifteen plots proportionally distributed between the compositional categories (Figure 12). For the primarily coniferous category, there were six sample plots selected. For the mixed forest category, seven plots were selected. Lastly, the primarily deciduous category comprised two samples. The total number of plots starts at 15 as an achievable goal, but adding future plots will not be a problem if needed to reach the sample count requirement of greater than 200 Douglas fir trees counted.

LBA Woods Sample Plot Map

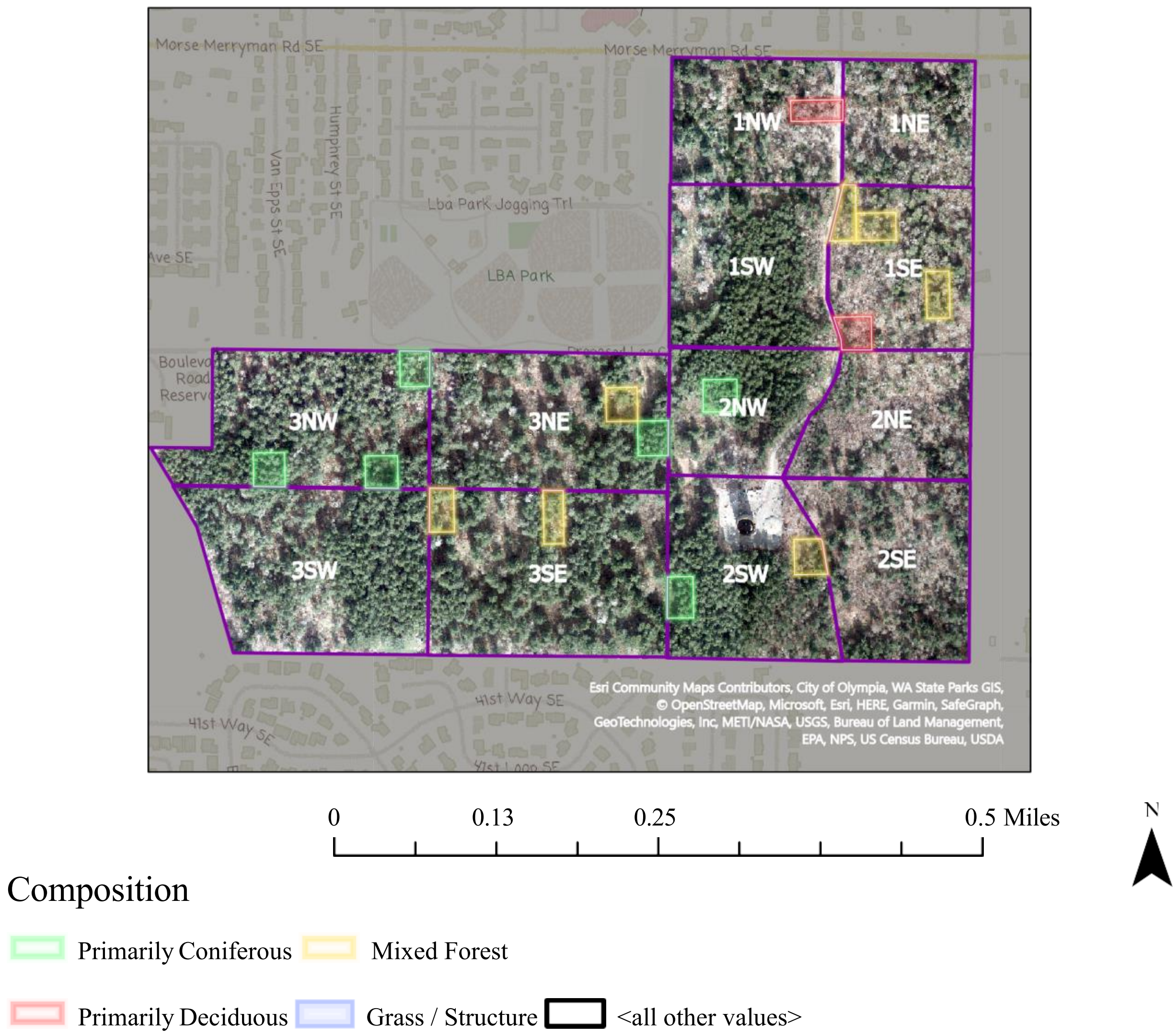


Figure 12: 15 proportionally selected sample plots in LBA park for ground truthing and testing

Ground Truthing

Marking site

Using the Juniper Geode GPS for high-precision geolocation in conjunction with Esri's Field Maps application, the boundaries of each sample plot were marked using highly visible stake flags.

Marking trees

Trees that were counted and measured were identified with flagging tape secured to the base of the tree. After the sample plot had been fully inventoried, the tape was removed and, depending on the condition, was reused on further sample plots.

GPS measurement

The Juniper Geode differential GPS device was utilized to aid in mapping consistency and geolocation accuracy. This device allows for sub-meter accuracy. Additionally, this data will help identify tree crowns for future delineation and segmentation once the orthomosaics have been tied to the ground control points.

DBH measurement

The following steps were based on US Forest Service guidelines to measure the diameter at breast height (DBH) (Powell, 2016).

- Single Trunk
    - Standing at the base of the measuring tape was used to note a height of 4.5' or 1.37 meters.

- At the measured height, diameter tape was wrapped around the trunk in line with the growth angle of the trunk. See below figure 4.

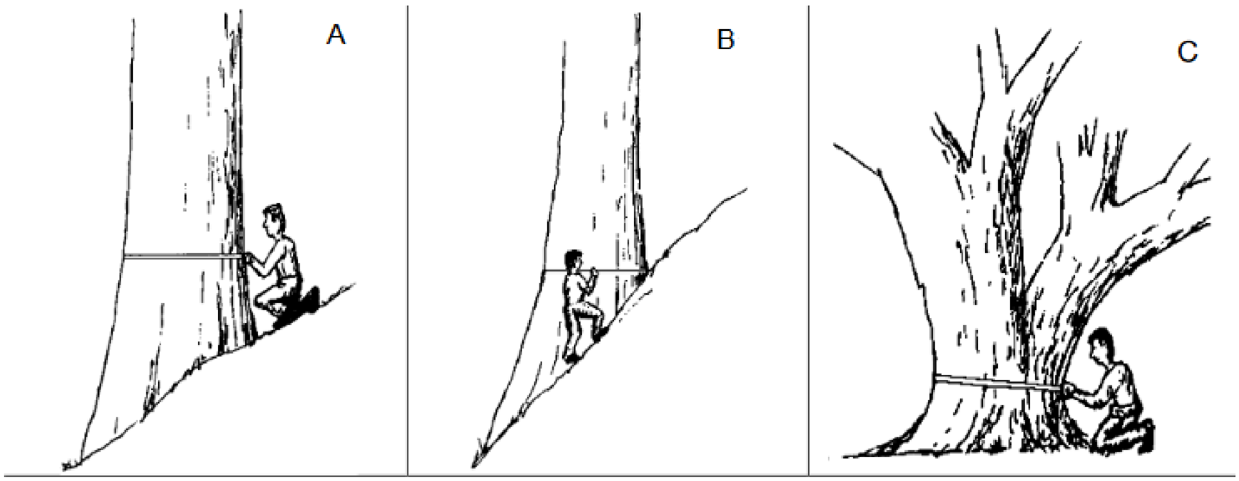


Figure 13: The US Forest Service document demonstrates three variations of tree growth that can affect the way trees are measured. Example A uses a technique where 4.5' was measured from the low end of the slope and the high end, after which the difference between the two was used as the breast height. Example B demonstrates measuring the DBH at the ground level on the high end of the slope due to the steep angle of the slope. Finally, Example C shows that when a tree is forked, it may be better to measure just below the fork. Even though 4.5' is preferred it may not be the most accurate trunk measurement height for forked trees.

- Multi- Trunk (Process adapted from New York State Department of Environmental Conservation) (NYDEC, n.d.)
  - When a tree has multiple distinct trunks that split from a single base (different from forked trunks), a slight variation in the above steps must occur.
  - Measure each of the divergent trunks at 4.5' or 1.37 meters
  - Using the diameter tape, measure each trunk individually. See below figure 5.
  - The square root of the sum of the squared trunk measurements will be used as the DBH of the multi-stemmed trees.

- For Example – if the DBH of three stems is 15, 17, and 20 the formula will look like this $\sqrt{(15^2+17^2+20^2)} = \sqrt{(225+289+400)} = \sqrt{(914)} = 30.23$”

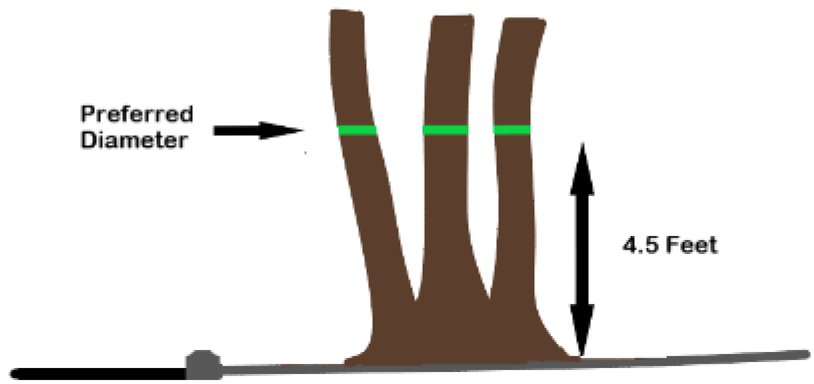


Figure 14: DBH measurement is taken from the multi-stemmed tree at 4.5’. These measurements are used in the formula described above. (NYDEC, n.d.)

Height measurement

Tree height measurements were taken using the methodology described by the US Forest Service (Powell, 2016). The tools used in this measurement were a clinometer (Suunto) and a laser range finder/laser hypsometer (Insight 400LH). Ideally, the distance from the measured tree was 100 feet. While primary height measurements were gathered with the clinometer, the values were verified with the laser hypsometer height measuring functionality to ensure consistency. In the rare circumstances when 100 feet was not attainable due to natural restrictions of the area the laser hypsometer was used as the primary height measurement tool at max visible distance from the target.

- Flat ground (see figure 15)
  - Standing 100’ away from the tree find with percent clinometer.

- Pointed directly at the tree on a level line the percent above or A is added to the percent below or B which gives the tree height.

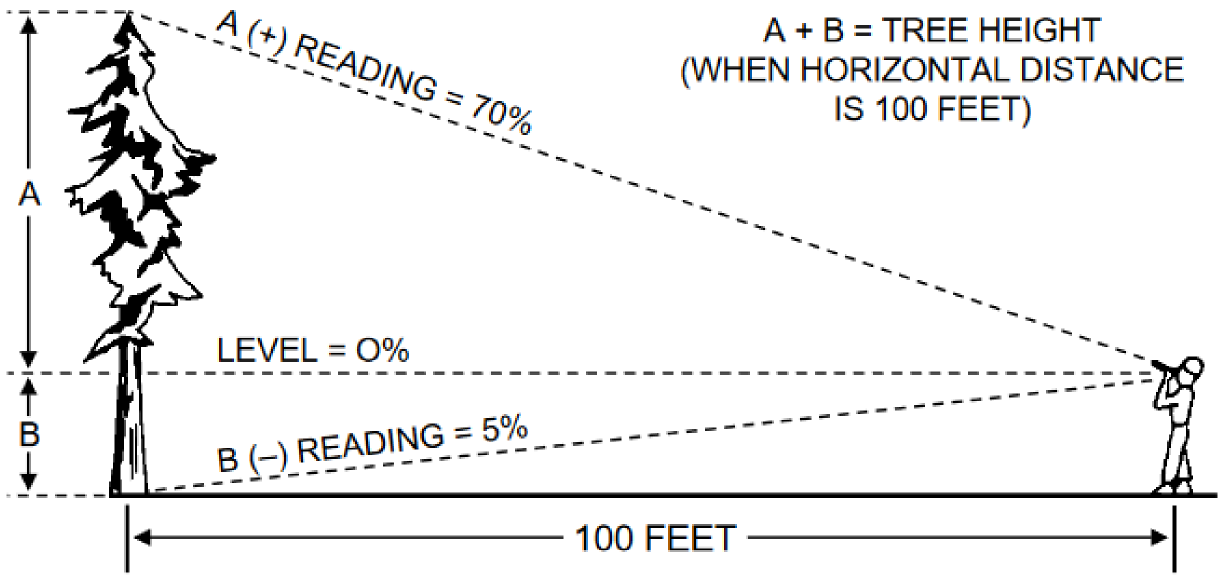


Figure 15: tree height measured using percent clinometer. Angle A is added to angle B to attain the tree height on flat ground (Powell, 2016).

- Sloped ground (see figure 16)
  - Find the point on the tree that is level with your current elevation in relation to the tree as your level 0.
  - Aim the clinometer parallel with the ground to find the percentage reading of the eye height of the tree.
  - This percentage is applied to the slop correction factor table.
  - Multiply the slope correction factor by the 100’ distance from the targe to attain the needed distance for the slope.
  - Take readings for A and B as described above for flat ground.

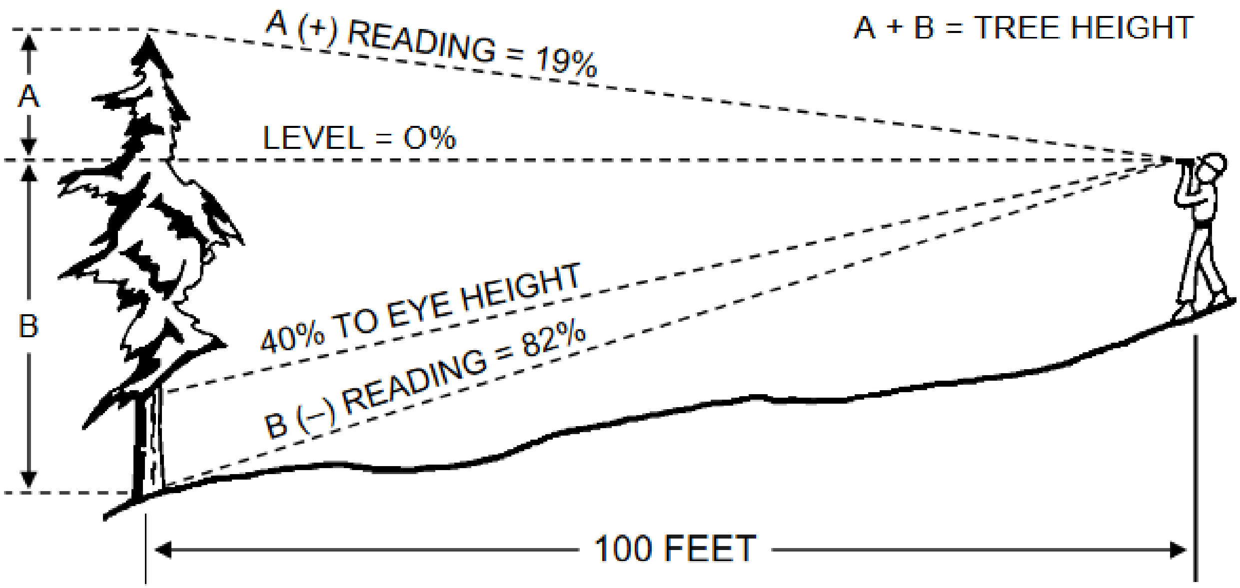
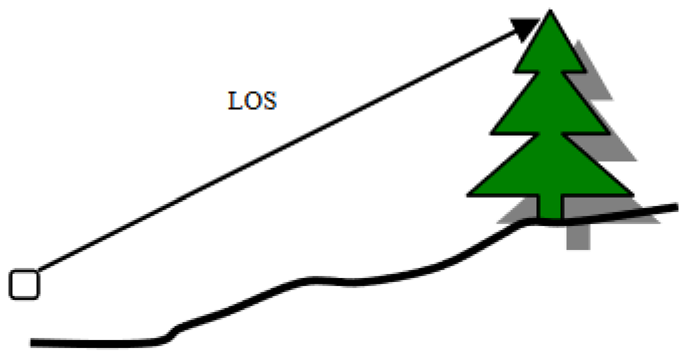


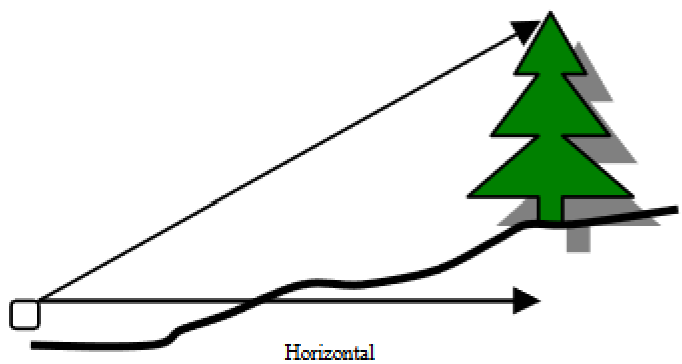
Figure 16: Tree height measured on a slope using a slope correction factor as described above (Powell, 2016).

- Sub 100’ distance height measurement
    - Use the above steps for flat or sloped calculations to start
    - Since the measurement has been taken at a smaller than 100’ distance a percentage calculation must be made
    - Divide the impromptu baseline distance by 100
    - Using the calculated sum of A and B multiply it by the product from the previous step.
        - For example, if you are standing 60 feet from the tree, first divide by 100, giving you 0.6.
        - Multiply 0.6 * (A+B), and tree height will be found.

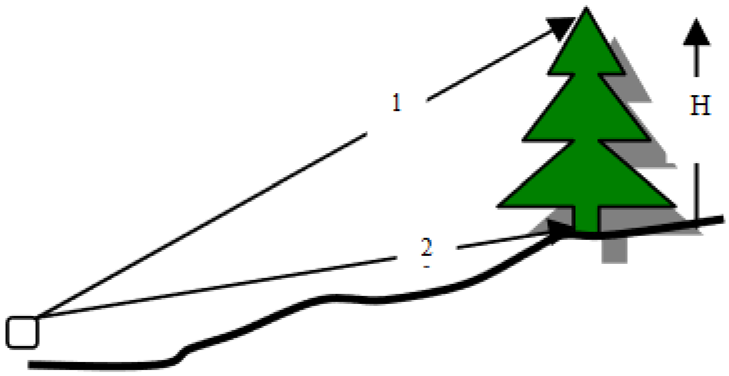
<u>Laser Range Finder/Hypsometer</u>

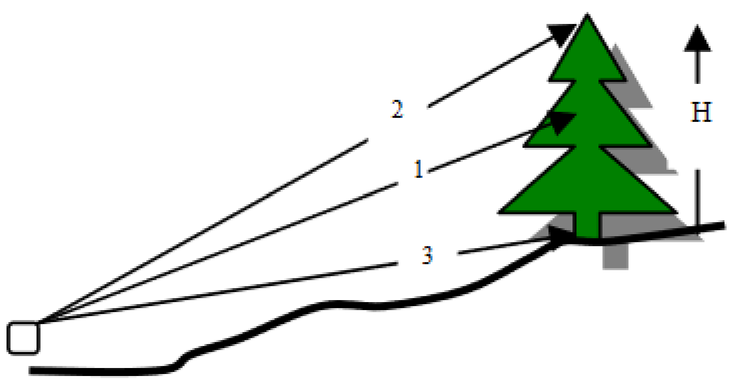
The Insight 400LH has four functional modes that can be easily swapped between. The following details are adapted from the operation manual.

- Mode 0 – Line of Sight Distance
    - Simple distance measurement between the device and the point that the laser reaches.
    - 

- Mode 1 – Horizontal Distance
    - The device uses a vertical angle and the onboard computer to process horizontal distance.
    - 

- Mode 2 - Two-point Height Measurement

- Using a top point and a bottom point measurement, the device calculates the approximate height of the object.
- This works best when visibility is unobstructed.
- 

- Mode 3 – Three-point Height Measurement
  - A central point measurement is first captured which is used as horizontal distance and starting angle.
  - A top angle and bottom angle measurement are taken after the initial central point measure to then finish the calculations.
  - 

- Mode 4 – Angle Measurement
  - Simple single angle measurement.

- Point device at desired angle and press the capture button to see displayed the appropriate value.
- This can be used in lieu of a clinometer.

Drone imagery

The drone imagery that has been captured was processed using a DJI Mavic Pro 2 drone. For mission planning, the proprietary DJI application was employed. The first mission flew 380 feet in altitude and had 80 % overlap. Ground control points were placed in visible flight path areas to ensure high-quality geolocation correlation and measured with the Geode GPS device. The resulting imagery was processed to make 2d imagery products using Esri's "Drone2map" software. The 2d products produced were the orthomosaic, digital terrain, and digital surface models. These products helped save processing time in R for the automatic segmentation process.

Segmentation

Segmentation was accomplished in two ways. The first method used was manually drawn crown segmentation in ArcGIS pro. I simply created a feature layer and sketched the approximate crown area. The secondary method attempted was automatic canopy segmentation through R using the ForestTools package to implement the watershed algorithm. The method for this segmentation has been adapted from the ForestTools use instructions developed by Andrew Plowright (Plowright, 2021). The first step in automatic segmentation is to produce a canopy height model. To do this, simply import the DSM and the DTM into R. The CHM = DSM-DTM. CHMs can be easily produced in ArcGIS Pro using the DSM and DTM rasters and the "minus" geoprocessing tool. It is vitally important when exporting the raster that the x and y cell sizes are the same to produce "square cells." If this is not done, ForestTools can't process the treetops. In R, I then imported the CHM tif file using the raster package. From here on, the workflow described by Andrew Plowright on canopy analysis is used, starting with variable window filtering to detect individual treetops (Plowright, 2021). This filter is added to the marker-controlled watershed algorithm to produce tree crown polygons. In this case, the markers are the treetops defined with the filtering function. Simple adjustments to the equation in R will allow the production of a polygon layer to be exported as a shape file. This allows us to add a shape file to ArcGIS Pro that automatically segments crown areas for our predictive models.

Joining Data

At this point ground truthed location, height, and DBH have been collected, ortho images and elevation models have been produced, and all that remains is connecting this data together through ArcGIS Pro processing tools.

- The first thing that needed to occur was a new map with all the necessary layers added together. This includes the CHMs, the crown polygon layer that was manually drawn, the sample plot grids, and the manually measured GPS locations along with their ground truthed data.
- With that done the next step is to use the Zonal Statistics as Table (Spatial Analyst) tool in ArcGIS Pro. For this tool I entered boundaries used to define zones, a field that will be used for the join, the value layer with, in this case, height data, and finally what type of statistics desired. See Figure 17 below as an example of the setup.

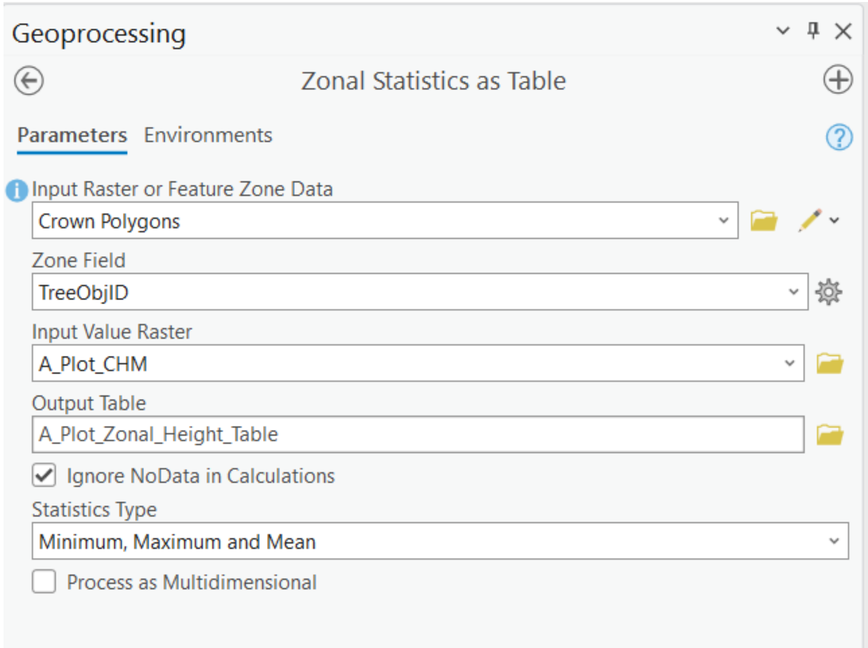


Figure 17: This figure shows the inputs used for the A Plot zonal statistics table.

Once this table has been completed the next step is to join the crown polygons to the individual GPS points that were manually gathered on site. This process is relatively standard for joining procedures in ArcGIS Pro. For this portion I used the Join Field geoprocessing tool with the settings shown in Figure 18 below. With this unified data layer the real processing and model building can begin.

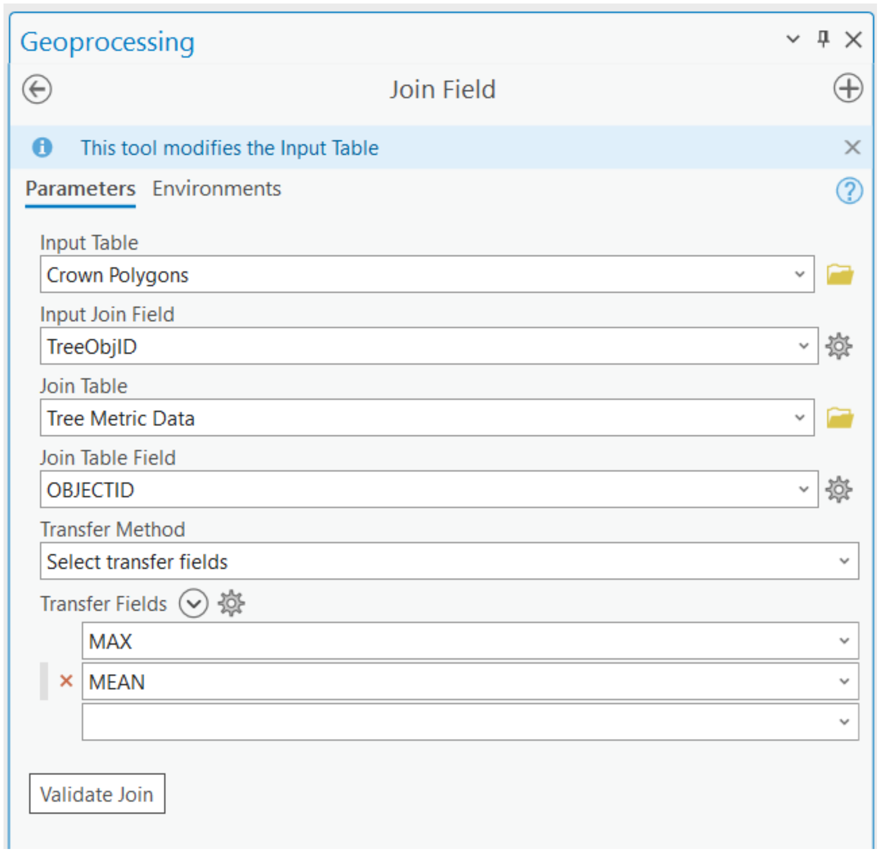


Figure 18: This figure shows the inputs used to join the crown polygon layer to the manually collected GPS point layer.

Prediction Models

Linear regression

One of the main premises of the strength of my models being tested is that a single variable linear analysis is less effective than multiple variables. Since data has been collected for multiple variables, testing single variables, and comparing is a simple additional step.

- Single Variable Linear Regression
  - Y = Diameter at Breast Height
  - X = Tree Height or X = Crown Area
  - β = slope
  - ε = error
  - β_0 = y-intercept
  - Formula used:
    - $Y = \beta_0 + \beta X + \varepsilon$
- Multiple Variable Linear Regression
  - Y = Diameter at Breast Height
  - X_1 = Tree Height
  - X_2 = Crown Area
  - β = slope
  - ε = error
  - β_0 = y-intercept
  - Formula used:
    - $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i$

Support Vector Regression

Support vector regression was used on the data collected. The Same variables were used as above in the linear regression models. To make the SVR model in R, the package "e1071" was used. After the SVR model was built and implemented, I performed fine-tuning on the model to ensure the best fit in two ways. This process is called hyperparameter optimization, and it focuses on the correct Epsilon selection as well as the kernel. As performed by Iizuka et al. (2022), for this project, I used a grid search method to optimize the Epsilon value in the model. This works by simply training dozens of models and homing in on the hyperparameter selections that bring us closer to a 0 RMSE. The second area of fine-tuning is kernel selection. There are many kernels that can be used in SVR, and they fit best depending on the data that they are being trained with. Remember that a kernel trick within our SVR allows us to engage with a transformed version of our data without truly transforming it, which allows our machine-learning model to make connections that linear models may fail to achieve. For this stage of fine-tuning, I tested the results of model prediction using the following popular kernel options: Linear, Polynomial, Gaussian Radial Basis Function (RBF), and Sigmoid kernel.

4.0 Results and Discussion

4.1 Overview

Data collection and analysis for this project had the goal of answering the two research questions.

- How do estimation models such as linear regression and support vector regression differ from ground truthing in terms of the accuracy of DBH prediction?
- Do the Iizuka et al. 2018 and 2021 methods of multivariate linear regression and support vector regression for DBH estimation on Japanese Cypress maintain their validity when used on Douglas Fir trees?

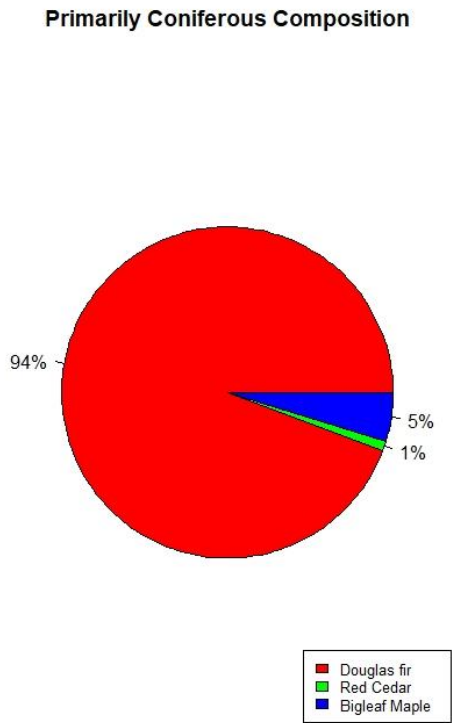
A total of 15 test plots were selected for the purpose of this research project. Three common species - Douglas fir, Western Red Cedar, and Big Leaf Maple - were counted and measured for DBH, height (feet), and GPS location. In section 4.2 I show the overall proportional representation of each of the significant tree species measured in each of the composition groups. Additionally, further context is added to better understand growing conditions effect on tree measurements through topology and hydrology of LBA park woods. In 4.3 - 4.5 drone imagery is collected and processed for height estimation and crown polygon segmentation. The last set of steps in 4.6 involved regression analysis performed in ten different model variations. Modeling was performed on the results using the manually collected tree height data from the ground truthing and compared to remote-sensed tree height data models for single variable, multiple variable linear regression, and support vector regression. The plan was to show which variables had the strongest relationships and whether support vector machine learning could outperform the more traditional linear regression.

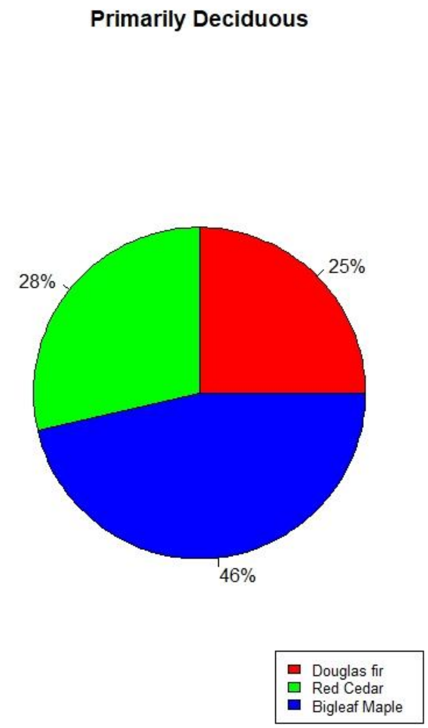
4.2 Ground Truthing

Based on the aerial imagery from 2018, three major compositional types were identified. These categories were formed by estimating the percentage of canopy cover that appeared to be coniferous or deciduous in terms of species visible. Each of these plots were categorized with specific criteria in mind for example: Primarily Coniferous plots should have a conifer canopy cover within the plot of $\geq$ 75%, Primarily Deciduous plots should have a deciduous canopy cover within the plot of $\geq$ 75%, and Mixed Forest should have neither conifer or deciduous canopy cover in excess of 75%. Within each plot there were species that I did not measure as I was focused on only Douglas fir, Bigleaf Maple, and Western Red Cedar. Keep note of that going forward as the proportions may not seem to match the category criteria, but what is shown below only accounts for the three species of significance in this study. For this project I set a threshold of at least 200 trees needed to have statistical significance. Douglas fir was the only one to meet this criteria, however, the preliminary results from Bigleaf Maple and Western Red Cedar are worth examining as an indicator for future research.

Through ground truthing procedure a final total of 517 trees were measured, with Douglas fir comprising the majority at 83% (430), followed by Bigleaf maple at 11% (57), and Western Red Cedar at 6% (30). While the goal was to achieve a statistically relevant sample size of 200 trees for all species, only Douglas fir met that criterion. Although Bigleaf Maple and Red Cedar did not meet the necessary sample size bar, I have included their results even though they are not as statistically relevant as the Douglas fir samples.

Each of the plots were classified prior to selection to inform a stratified random sampling using 2018 aerial imagery to make assumptions about the composition. This produced seven mixed forest plots with 136 trees sampled, six primarily coniferous plots with 353 trees sampled, and two primarily deciduous plots with only 28 trees sampled. Upon gathering the onsite tree information these compositions were close to the goals set by each category even without the non-significant trees and shrubs included (Figure 19). The primarily deciduous group seems to indicate an issue with the categorization criteria however, this was in large part due to there being deciduous species not identified by this study such as Red Alder, Cottonwood, and underbrush species like Vine Maple.





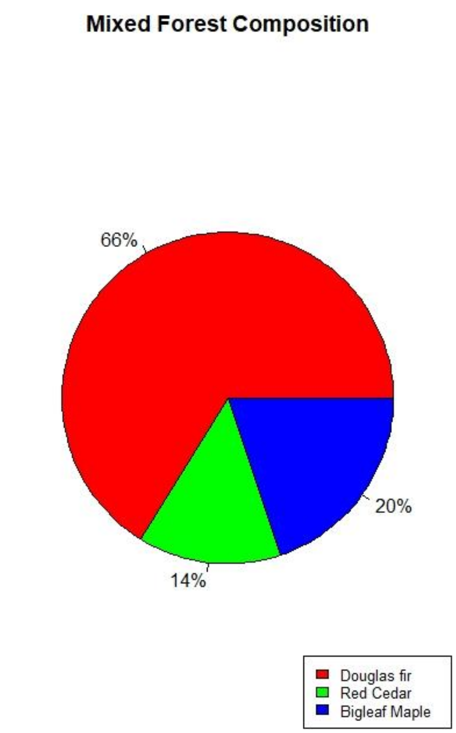


Figure 19: Three pie charts labeled with the compositions that were used in stratification. Primarily Coniferous (Top Left), Primarily Deciduous (Top Right), and Mixed Forest (Bottom Left). Keep in mind these plots contain trees and underbrush not listed in these pie charts. These only demonstrate the proportional concentration of the three species I focused on for analysis.

Ground-truthed Height and Diameter at Breast Height (DBH)

Height measurements were gathered for each of the trees within the fifteen sample plots using a combination of a clinometer and a laser hypsometer for confirmation. For Douglas fir trees heights ranged from 15 feet up to 158 feet high with a median height of ninety-five feet. The DBH gathered ranged from 5.21 inches to 49.21 inches. Bigleaf maple saw heights ranging between 30 feet to 119 feet, and DBH values between 7.48 inches to 47.24 inches. Lastly the Red Cedar had height values ranging between 28 feet and 120 feet along with DBH values ranging between 7.09 inches to 49.61 inches.

Differences in stratified grid plots based on composition

When the different composition classes were separated for each of the tree species, we see differences in growing patterns (Table 1,2,3). For the Douglas fir samples we see generally larger diameters and increased heights found in mixed forest when compared to either of the other compositions, however, the tallest tree was found within one of the primarily coniferous zones. Bigleaf Maple showed largely the same trends, oddly, with mixed forest composition having the trees that were on average taller and larger in diameter. The tallest Maples were found in the primarily coniferous plots even though the largest diameter trees were found in mixed forest plots. Red Cedar was slightly different in that it appeared to be underdeveloped in the primarily coniferous spaces with low height and DBH values. Mixed forest proved to be more suited for Red Cedar with primarily deciduous not far behind. The tallest cedars were found in mixed forest along with the largest diameters. Red Cedar struggled to gain a foot hold in the primarily coniferous zones due to the competition found within. In the denser coniferous areas, trees seem to show a pattern of quick growth to compete for sunlight. This is indicated with the lower DBH and outliers in tree height.

Douglas fir stats	**Mean**	**Median**	**Max**
Primarily Coniferous			
DBH (inches)	**15.81**	**14.17**	**39.76**
Height (feet)	**93.15**	**92**	**158**
Mixed Forest			
DBH (inches)	**22.14**	**22.84**	**49.21**
Height (feet)	**103**	**106**	**143**
Primarily Deciduous			
DBH (inches)	**12.54**	**11.42**	**27.56**
Height (feet)	**46**	**37**	**90**

Table 1: Mean, median, and max values collected via ground-truthing for Douglas fir trees in various composition plots.

Bigleaf Maple	**Mean**	**Median**	**Max**
Primarily Coniferous			
DBH (inches)	**18.39**	**16.93**	**33.07**
Height (feet)	**83.29**	**80**	**119**
Mixed Forest			
DBH (inches)	**26.01**	**24.41**	**47.24**
Height(feet)	**87.67**	**90**	**112**

Primarily Deciduous			
DBH (inches)	**17.90**	**18.11**	**38.19**
Height (feet)	**65.46**	**67**	**96**

Table 2**:** Mean, median, and max values collected via ground-truthing for Bigleaf Maple trees within the various composition plots.

Red Cedar	**Mean**	**Median**	**Max**
Primarily Coniferous			
DBH (inches)	**13.65**	**15.75**	**18.11**
Height (feet)	**48**	**57**	**59**
Mixed Forest			
DBH (inches)	**27.87**	**25.20**	**49.61**
Height (feet)	**85.53**	**79**	**120**
Primarily Deciduous			
DBH (inches)	**25.15**	**24.02**	**32.28**
Height (feet)	**72.5**	**68.5**	**90**

Table 3: Mean, median, and max values collected via ground-truthing for Western Red Cedar in the three different compositional plot types.

To give a small amount of hydrologic and topographic context I requested lidar elevation data from the city of Olympia to produce water flow accumulation estimates and contour maps below (Figure 20 and 21). The primarily coniferous plots were found on areas with little to no slope as well as some plots with mild slope of around 10-15 feet of elevation change. The primarily coniferous sample plots are located in the central and Western portions of the woods, additionally, these plots do not seem to have water flow in and around them to any notable degree. The mixed forest plots appear more commonly in the central and Eastern portions of the woods on sloped terrain with anywhere from 5 – 20 feet of elevation change across the plot. Water accumulation is present in some of the plots with the East most plot showing significant flow. Finally, when looking at the two deciduous plots in the central and Norther sites, there is a very small elevation change of around 5 feet in each and little to no water accumulation present.

The differences in topography and water flow patterns inform species composition and growth rates. The plots on the Eastern most portion see increased morning sunshine but less mid-day and evening sun. The Central and Western portions see the lion's share of the sunshine with midday and evening light. Douglas fir trees desire full sun thus the areas with the most sunlight through the day are more likely to have that species present. Bigleaf Maples thrive in soils with high levels of moisture but also need good drainage. While the two deciduous plots are found centrally, almost all of the deciduous plots are on the Eastern side of the hill slopping away from the park (Figure10). This would allow for less intense sunlight, high moisture flow and good drainage. Lastly the Red Cedars are known to succeed in moist soil with good drainage. Red Cedars do not handle drought or lack of water well, and thus it follows that they would appear more commonly on the Eastern portion of the park which was found in the data.

Topographic Contour Map of LBA Park and Woods

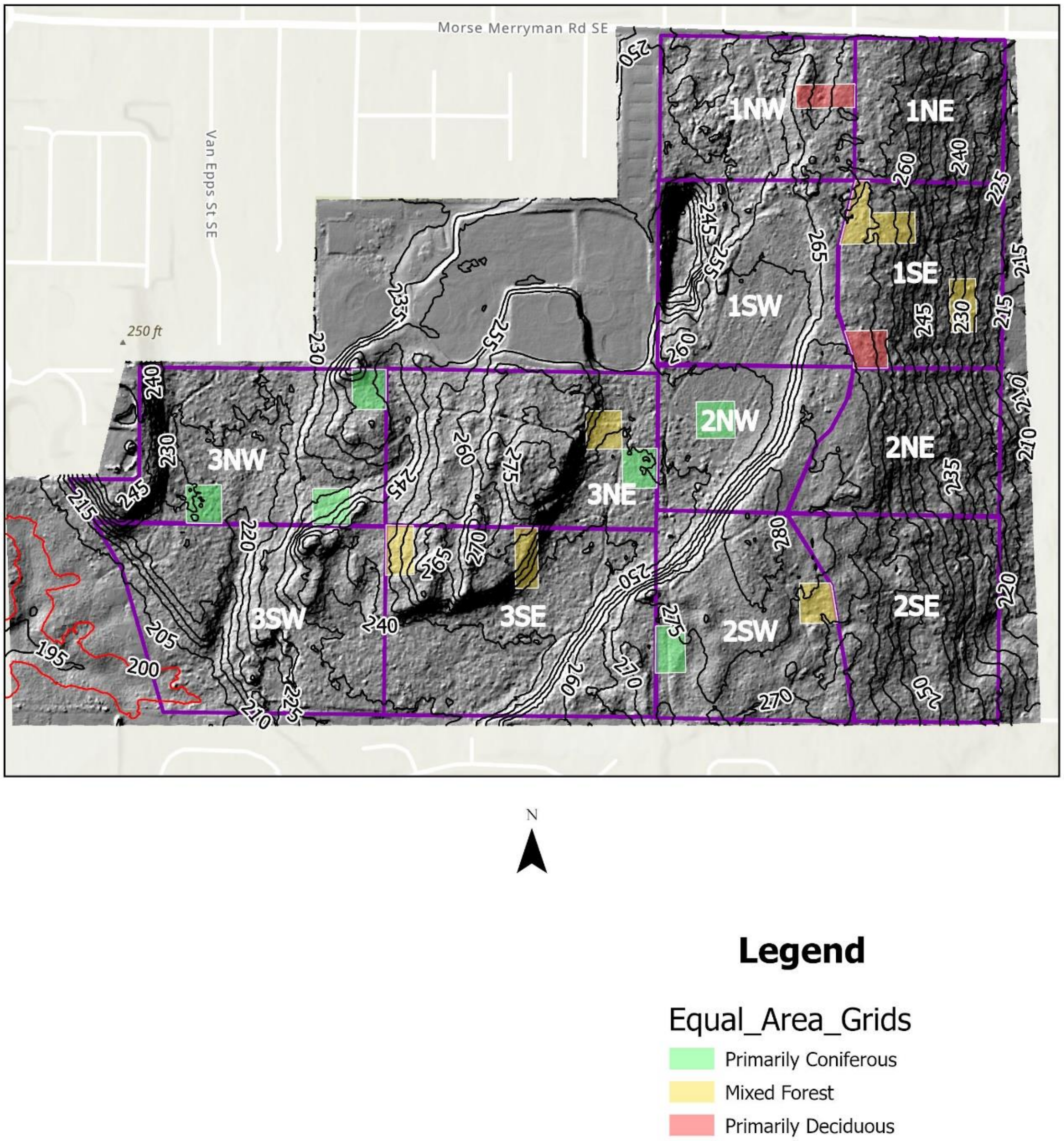


Figure 20: This contour map shows LBA park and woods area with a hillshade filter to better visualize slope of the terrain. The Maximum value can be found toward the center of the map just south of 2NW at a value of 280 feet with the lowest points to the Southwest and East side of the property at a value of 195 and 210 respectively.

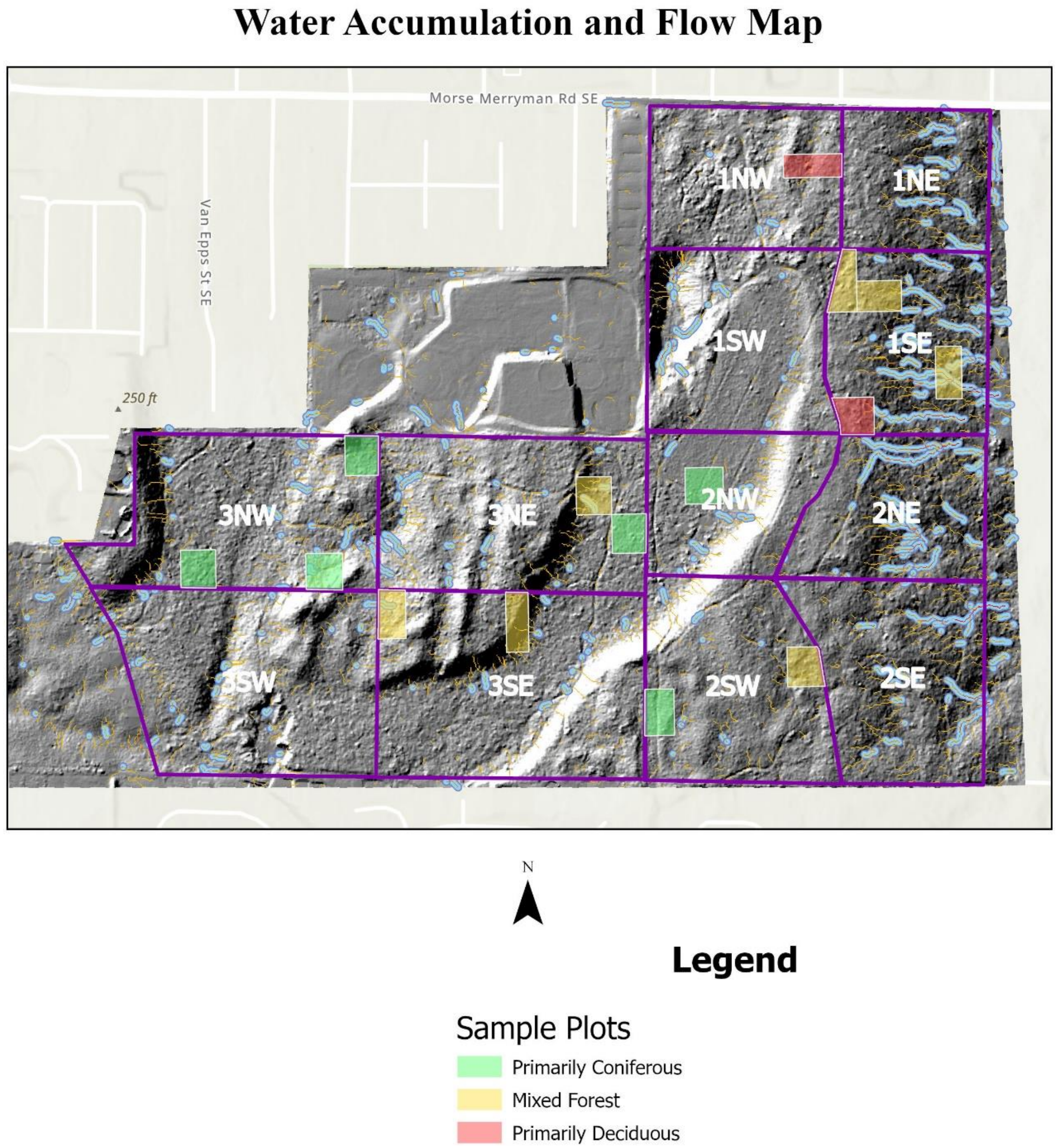


Figure 21: This terrain map of the LBA park and woods area shows the sample plots with flow accumulation overlayed as blue lines. The areas with concentrated blue lines depict areas with higher water accumulation.

4.3 Drone Imagery

Due to the scale of the project drone imagery capture was split into 3 mission dates: March 15th, 2023, March 16th, 2023, and March 27th, 2023. These missions together lasted around 15 hours of flight time, and 1,838 images were captured. The flight paths covered nearly all of the 155 acres of the park woods, even though the focus was to capture the sample plots. Each flight started around 11 am to ensure consistency and peak light quality. We flew a Mavic Pro 2 drone 380 feet above the ground, with 80% image overlap to ensure high-quality stitching. After collecting images in the field, Drone2Map (ESRI drone image processing software) was used to stitch the images together into the orthomosaics and elevation models Figure 22 and Figure 23.

Repeated missions were necessary due to a few key issues. The time of year affected the quality of drone imagery and the ability to capture good-quality photos. Issues with weather and the solar azimuth (the sun's angle in relation to the Earth during this time of year) proved challenging to work through. The solar azimuth in March produces longer shadows in Western Washington, decreasing the processing software's ability to correlate images causing blurriness and muddiness.

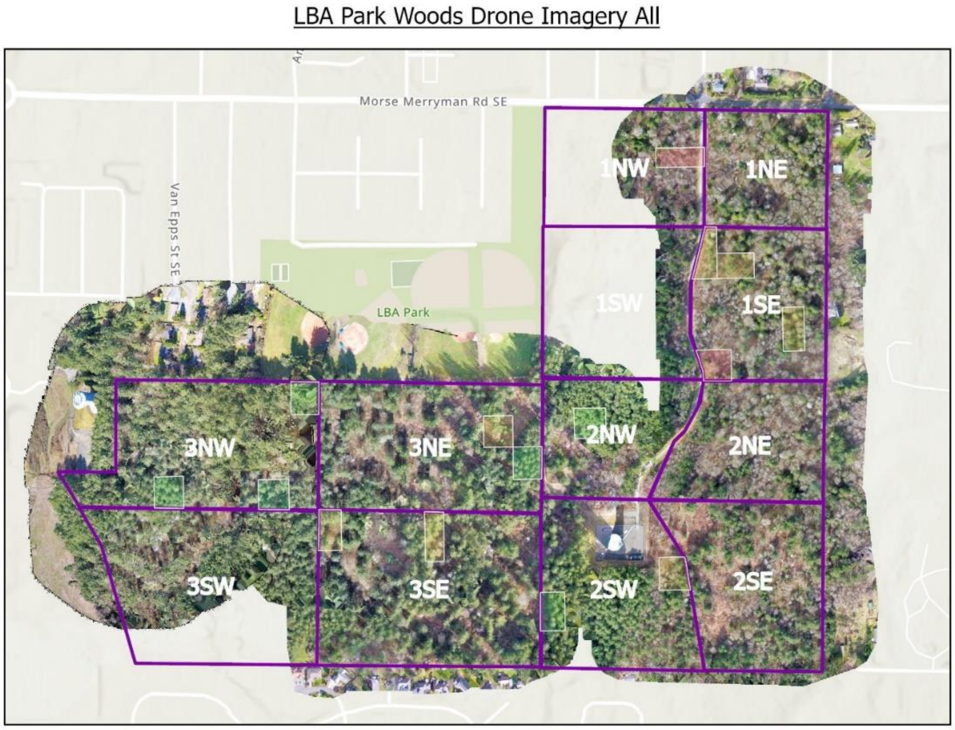


Figure 22: Map of LBA Woods drone imagery captured for this project processed as an orthomosaic.

4.4 Segmentation

Tree crowns were segmented in each of the sample plots to provide crown area. ArcGIS pro version 3.1.0 was the program used to draw the polygons and estimate spatial area. A feature layer was created with the preselected species as an option which allowed easy delineation of crown. Where images were less clear the use of derived canopy height models helped provide clarity between some trees.

Sample Plot Crown Area Segmentation

Figure 23: Example of the sample plots that have been manually segmented. Each tree crown has a polygon drawn around it and been coded by species.

4.5 Height Estimation

In this study height was estimated through remote sensing as well as the ground-truthed tree height data. To determine the reliability of remote sensed height within the simple mission plan, the data were compared to the ground truthed height data. For the height estimation a digital terrain model and a digital surface model were derived in the Drone2Map processing software for each of the missions flown and used to produce the canopy height model picture below (Figure 24). Using Esri's "zonal statistics as a table" geoprocessing tool, with the canopy area set as the extent, height statistics such as Max, Min, Mean, etc., were gathered. For tree height, I assumed that the Max point or tallest point measured on the tree should be considered the height (feet).

LBA Park Woods Crown Height Model All

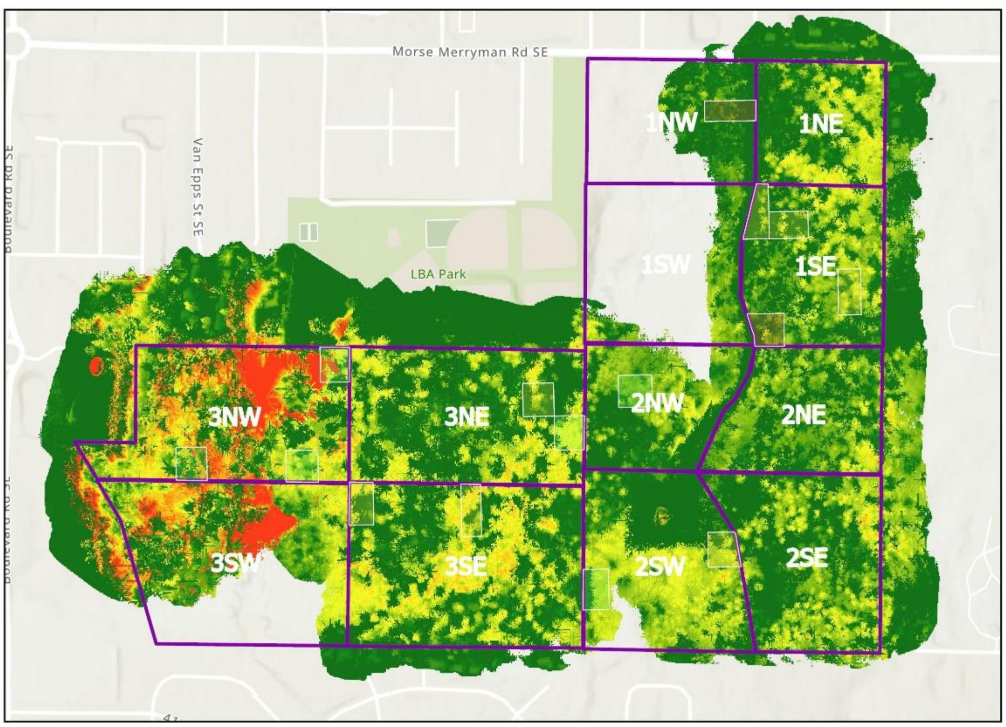


Figure 24: Map of LBA Woods drone imagery processed in ArcGIS Pro to show a canopy height model. The individual missions have been added to a mosaic, so they show values in relation to

each other. Generally taller trees appear to be focused in the 3NW and 3SW zones in red and were primarily composed of Douglas fir trees.

Remote-sensed Height vs. Ground-truthed Height and Possible Explanations

As expected, there is a difference between the ground-truthed results and the remotely sensed height measurements. The root mean square error (RMSE) of the height estimate was 20.1, mean absolute error (MAE) of 15.49, and R^2 score of 0.56 (Figure 24). The RMSE and MAE are relatively high and will impact the accuracy of the models. This project in part seeks to find the difference in predictive accuracy between ground-truthed data and remote-sensed data. The other aspect being tested is whether different variations of model structure and input will increase remote sensed accuracy to be more in line with what was measured in field.

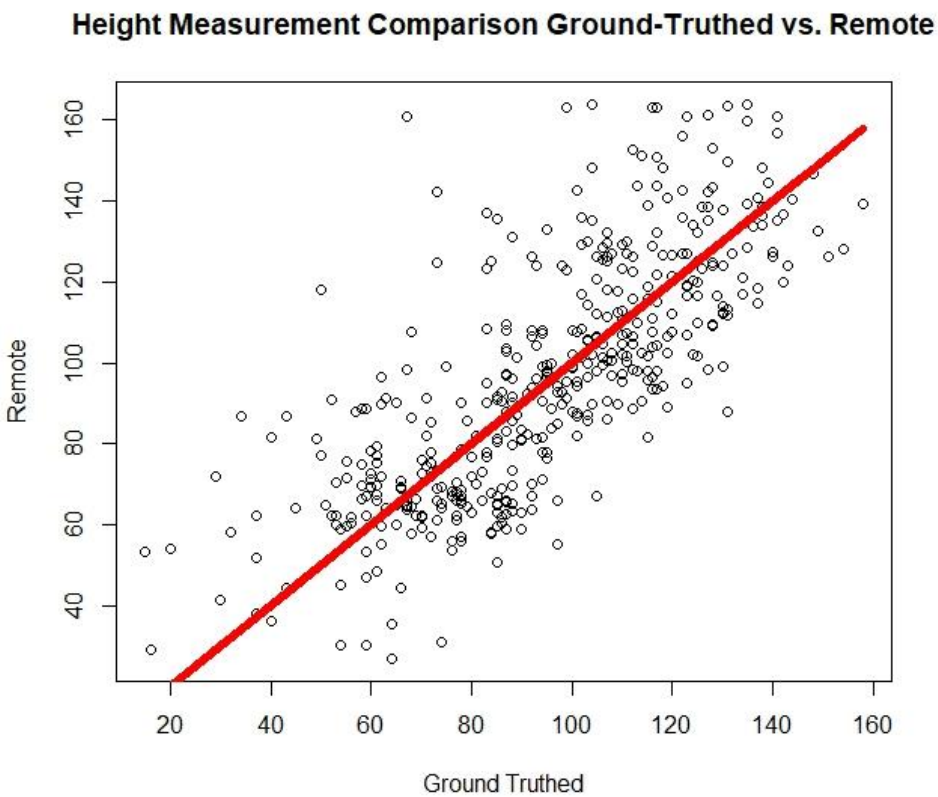


Figure 25: The graph above shows how closely the remote sensed tree height of Douglas fir trees in feet (Y axis) relates to the ground truthed manually gathered tree height of Douglas fir trees in

feet (X axis). The Red line shows what a 1:1 ratio would look like if the two methods produced identical values.

There are a few causes for potential issues with height sensing. The quality of height estimation is directly connected to the generation of elevation products like the DTM and DSM (Iizuka et al., 2018). Weather can impact drone imagery significantly. Trees shift when overly windy weather is present, making establishing tie points between images difficult for the stitching software (Figure 25). Another potential element that has affected the quality of the elevation models is the time of year. The imagery was captured in mid-March, increasing the shadow length of trees and structures even if the sun is at its peak height. These long shadows reduce stitching quality (Figure 25). The Mavic Pro 2 drone used with this project relied entirely on the RGB visible light spectrums as it did not have a multispectral camera or a LIDAR attachment. This means that the derived elevation models used what was visible to the camera, causing the artifacting seen below (Figure 25). Through different combinations of image processing settings, the images used for the sample plots were functional. Discrepancies remained even with the fine-tuned processing settings, considering the difference between height values.

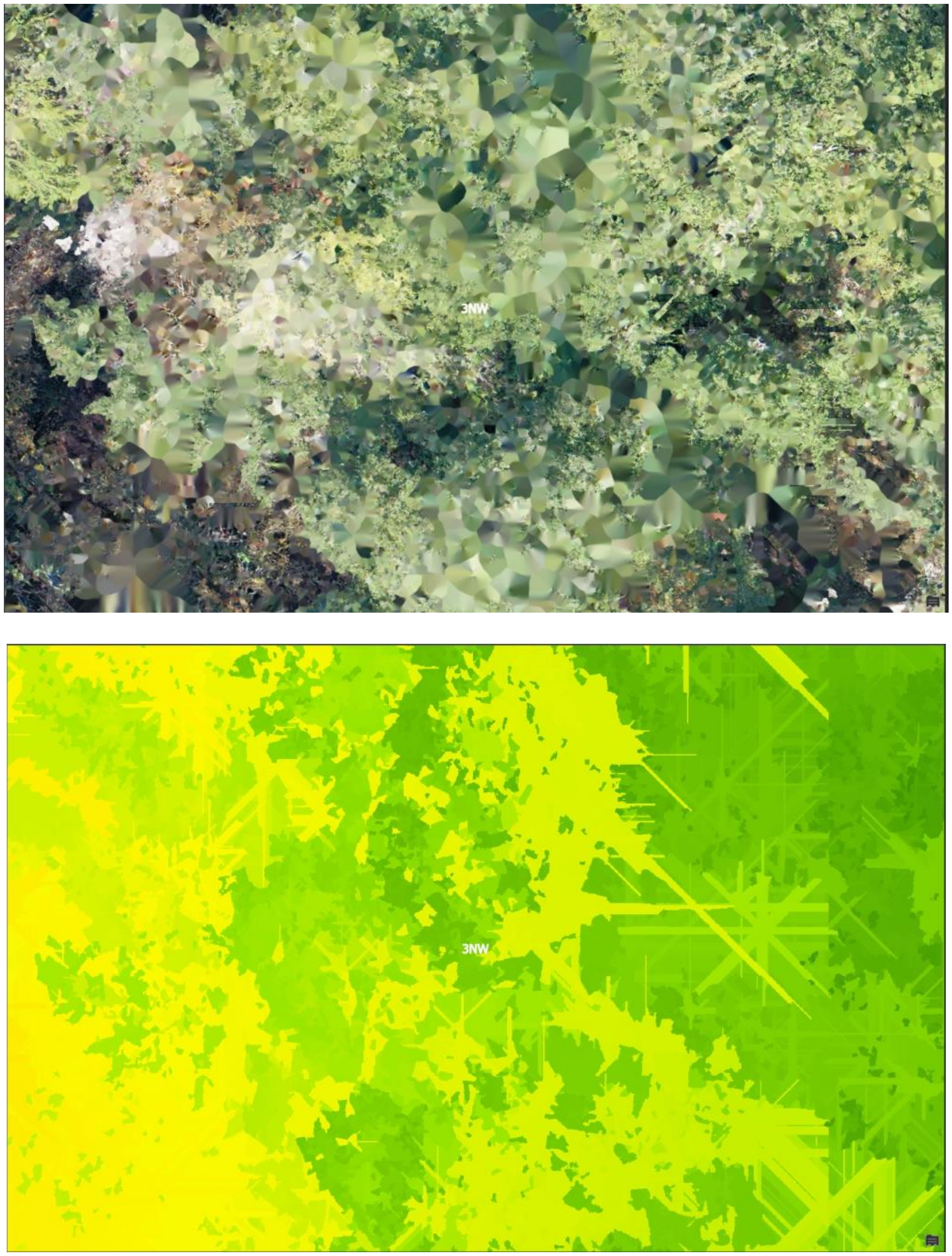


Figure 26: These images show the potential issues found with poor stitching of ortho images in the Drone2map software. Top: RGB Orthomosaic of particularly badly stitched images. Bottom: DSM of the ortho above. Note the sharp crystal-like lines that are spread throughout.

At this point all of the raw data have been collected including ground-truthed DBH and height, along with remote-sensed crown area and height. These will now be used as variables in the modelling and regression analysis for Douglas fir, Bigleaf Maple, and Western Red Cedar to predict DBH. This will start with single variable analysis for both linear and support vector and then move on to multi-variable analysis. Using the single variable models, we can see which variables individually have the strongest relationships with DBH. Afterwards I used both height and crown area for multivariable analysis to attempt to improve upon the predictive capability of the single variable models.

4.6 Regression Analysis

Single Variable Regression Overview

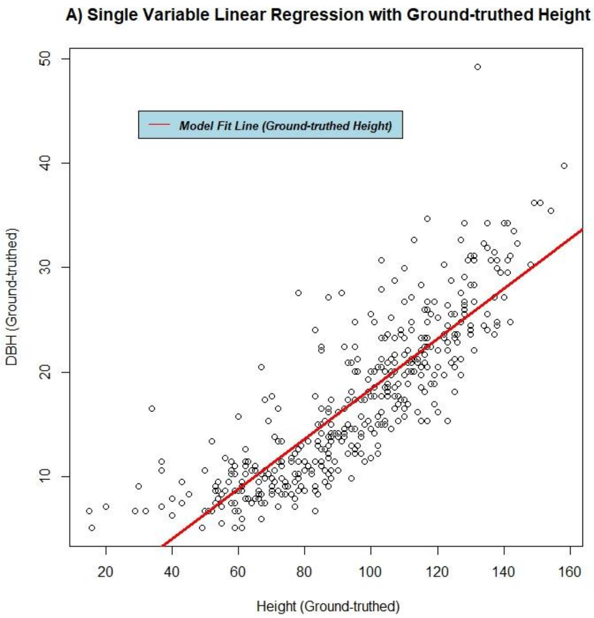
Now that the ground-truthed data as well as drone imagery derived remote-sensed data have been collected, I built and tested the various regression models in R. All of the models are attempting to predict the value of Y (DBH) using different X variables and values. The first modeling attempt used the ground-truthed height as the X variable giving a "best-case scenario" with which to compare the remote sensed data. The second model uses remote sensed height as the X variable. The third model built used remote-sensed crown area as the X variable.

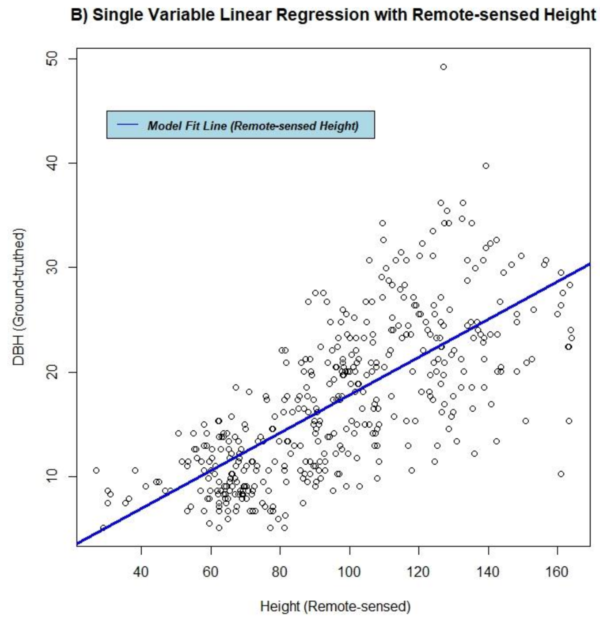
One quick note before the in-depth regression analysis. There is heavy emphasis on the R^2 score going forward and it's important to know a little more about it and how it relates to this analysis. As mentioned in the methods R^2 is an evaluation tool used commonly in regression analysis to describe how well a model explains the variance of our target variable in this case diameter at breast height (DBH). Based on the studies referenced in this paper an R^2 value of at least 0.7 is desired for an indication of the strength of a model's explanatory ability. For example, Iizuka et al. 2018 examined the relationship between DBH and tree height as well as crown area in Japanese Cypress trees. Iizuka et al. found that for this species using remote measurement there was a weak R^2 relationship between DBH and tree height (0.20) when compared to the R^2 found between DBH and crown area (0.7923). When we compare the results for Japanese Cypress to what is seen with the Douglas fir models created below, and we find similar if not slightly stronger values in the ground truthed models. With the lowered height reliability of the CHM found in the remote sensed data set we see a drop in the R^2 across all model types, linear regression, and support vector regression alike. These variances show the readily apparent fact that accuracy of the height model impacts results of remote sensed data dramatically and high-quality capture and processing should be prioritized.

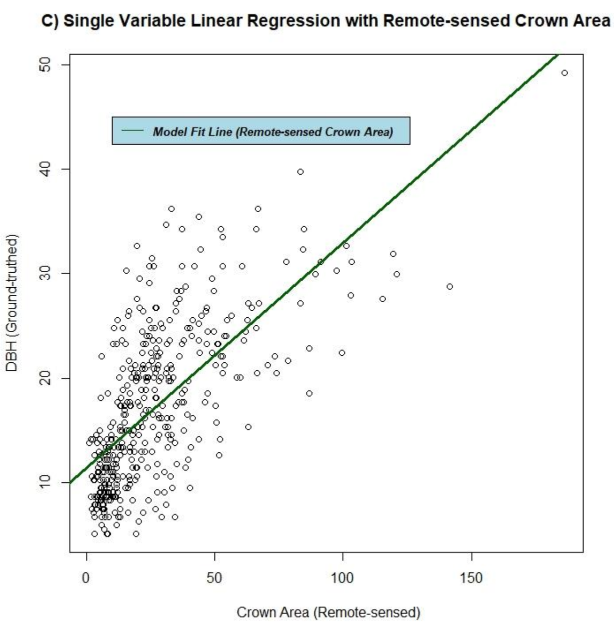
Single Variable Linear Regression Results

Douglas fir Single Variable Linear Results

Ground-truthed height used as the explanatory variable produces predicted values closer to the ground-truthed DBH of Douglas fir trees ($R^2 = 0.71$) when compared to remote-sensed height ($R^2 = 0.50$), and crown area ($R^2 = 0.45$) (Figure 27). This is most likely due to the consistent accuracy of the data being used to inform the model. The remote-sensed data models show a drop in predictive ability. Even with this drop, however, we see the Crown area scoring the lowest. These trends continue through the other values with ground-truthed data having the lowest root mean square error (RMSE = 4.03) inches with remote sensed height (RMSE = 5.31) and crown area (RMSE = 5.58). The RMSE score is more directly affected by outliers and higher values are to be expected in a sample population where outliers can be expected. Mean absolute error can offer smoother results as it helps reduce the impact of outliers. In this case the same trend of results is seen with ground-truthed height having a clear advantage.





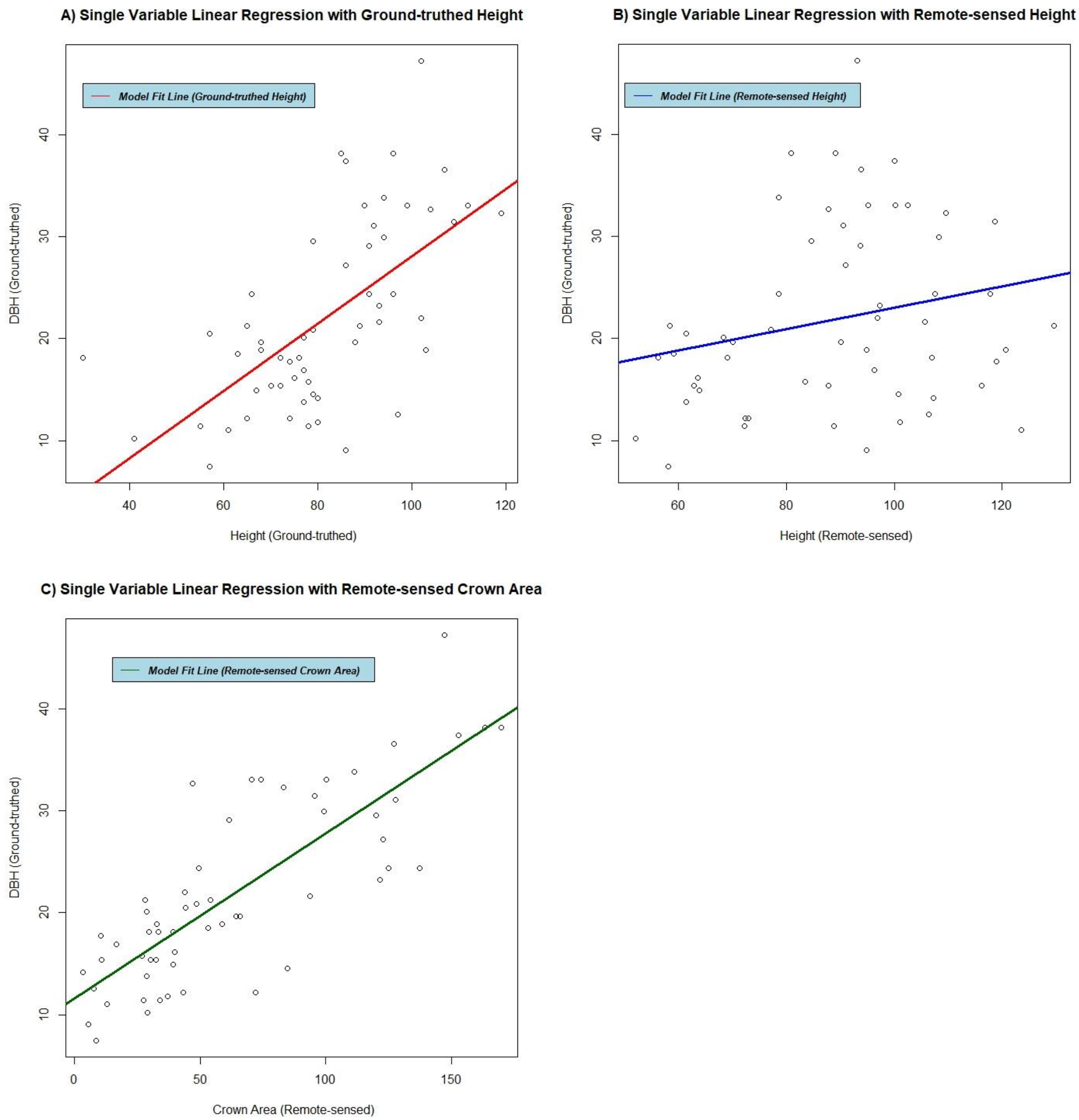


Douglas Fir Single Variable Linear Regression Values	**R^2 value**	**RMSE of DBH (inches)**	**MAE of DBH (inches)**
A) X = Height (ground-truthed)	0.71	4.03	3.08
B) X = Height (remotely sensed)	0.50	5.31	4.19
C) X = Crown Area	0.45	5.58	4.45

Figure 27: Single variable linear regression model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Douglas fir DBH. A) X = Height (ground-truthed), B) X = Height (remote-sensed), C) X = Crown Area

Bigleaf Maple Single Variable Linear Results

Ground-truthed height used as the explanatory variable produces predicted values closer to the ground-truthed DBH of Bigleaf Maple trees (R^2 = 0.38) when compared to remote-sensed height (R^2 = 0.04), however the model using crown area scored better than both height-based models (R^2 = 0.65) (Figure 28). Compared to the Douglas fir results the Bigleaf Maple models show much stronger predictive accuracy using the crown area. The RMSE values see similar trends with crown area having the lowest score (RMSE = 5.28) followed by the ground-truthed height (RMSE = 6.93) and the remote-sensed height (RMSE = 8.70). Maples are deciduous and thus have different growth patterns, so it follows that different metrics would be connected to size and health when compared to coniferous trees. In this case a larger crown area means more energy gathered from the large leaves which seems to translate to increased diameter size of the trunk.

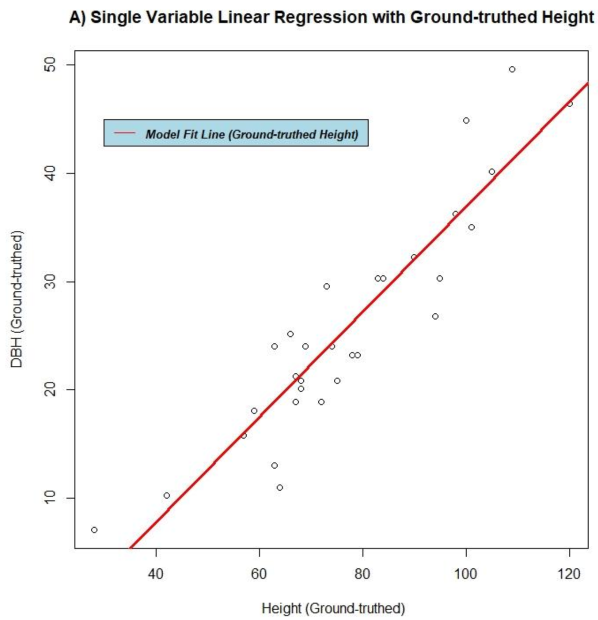


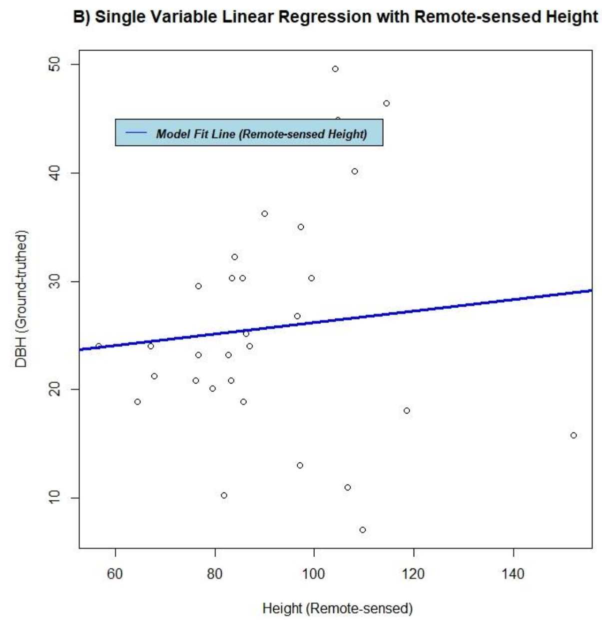
Bigleaf Maple Single Variable Linear Regression Values	**R^2 value**	**RMSE of DBH (inches)**	**MAE of DBH (inches)**
A) X = Height (ground-truthed)	0.38	6.93	5.51
B) X = Height (remotely sensed)	0.04	8.70	7.05
C) X = Crown Area	0.65	5.28	4.17

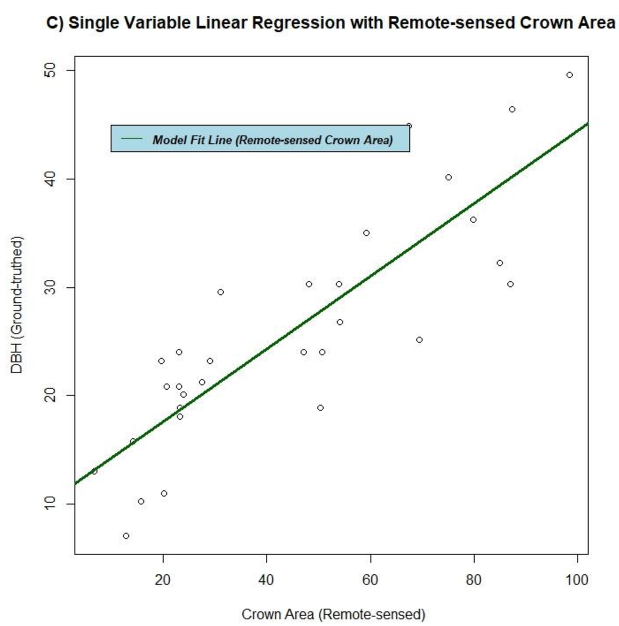
Figure 28: Single variable linear regression model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Bigleaf Maple DBH. A) X = Height (ground-truthed), B) X = Height (remote-sensed), C) X = Crown Area.

Western Red Cedar Single Variable Linear Results

Ground-truthed height used as the explanatory variable produces predicted values closer to the ground-truthed DBH of Red Cedar trees ($R^2 = 0.84$) when compared to remote-sensed height ($R^2 = -0.03$) and crown area ($R^2 = 0.72$) (Figure 29). The remote-sensed height model failed likely due to reasons outlined in section 4.5 such as weather caused misreading of height values. Additionally, it must be reiterated that the sample size for Red Cedar was much smaller than the desired 200 samples. RMSE values fall in a similar range compared to Douglas fir, remote-sensed height being the exception. Ground-truthed height model (RMSE = 4.07) outperformed the remote-sensed height model (RMSE = 10.33) as well as the crown area model(RMSE = 5.42) (Figure 29).







Red Cedar Single Variable Linear Regression Values	**R^2 value**	**RMSE of DBH (inches)**	**MAE of DBH (inches)**
A) X = Height (ground-truthed)	0.84	4.07	3.12
B) X = Height (remotely sensed)	-0.03	10.33	8.03
C) X = Crown Area	0.72	5.42	4.42

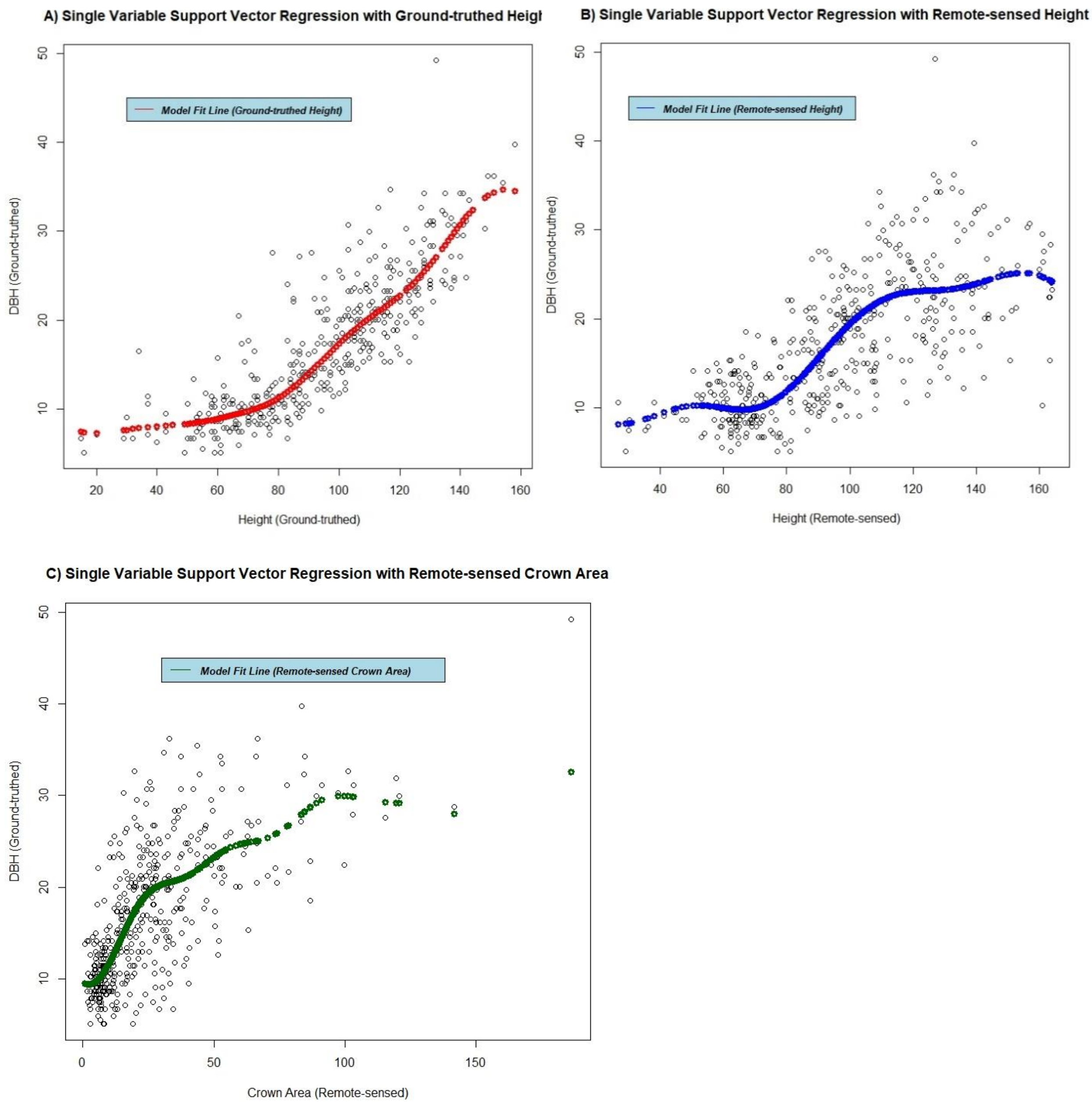
Figure 29: Single variable linear regression model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Red Cedar DBH. A) X = Height (ground-truthed), B) X = Height (remote-sensed), C) X = Crown Area.

Single Variable Support Vector Regression Results

Single variable support vector regression attempts to increase model fit when compared to single variable linear regression using machine learning to explain the relationship between two variables. It is less dependent upon linear relationship and thus can fit more variety of relational structures. As described in the methodology the variables are the same as what we would find in single variable linear regression with the Y variable being DBH, X being ground-truthed height, remote sensed -height and crown area individually based on the model variant being tested.

Douglas fir Single Variable Support Vector Results

Ground-truthed height used as the explanatory variable produces predicted values closer to the ground-truthed DBH of Douglas fir trees ($R^2 = 0.76$) when compared to remote-sensed height ($R^2 = 0.55$), and crown area ($R^2 = 0.51$) (Figure 30).The trends that were seen in single variable linear regression are present as well in the support vector regression models, however minor improvements have been achieved. In each of the variable’s cases there is a rise in the R^2 value of 0.05 meaning that these models explain the data just slightly better than the single variable linear models. RMSE’s saw an average reduction of 0.28 meaning that they were effectively about a quarter inch more accurate. MAE saw a larger average reduction of 0.41 inches reducing the effect of outliers even further since the models could fit better to them and not hurt the prediction downstream of the large or small outliers.

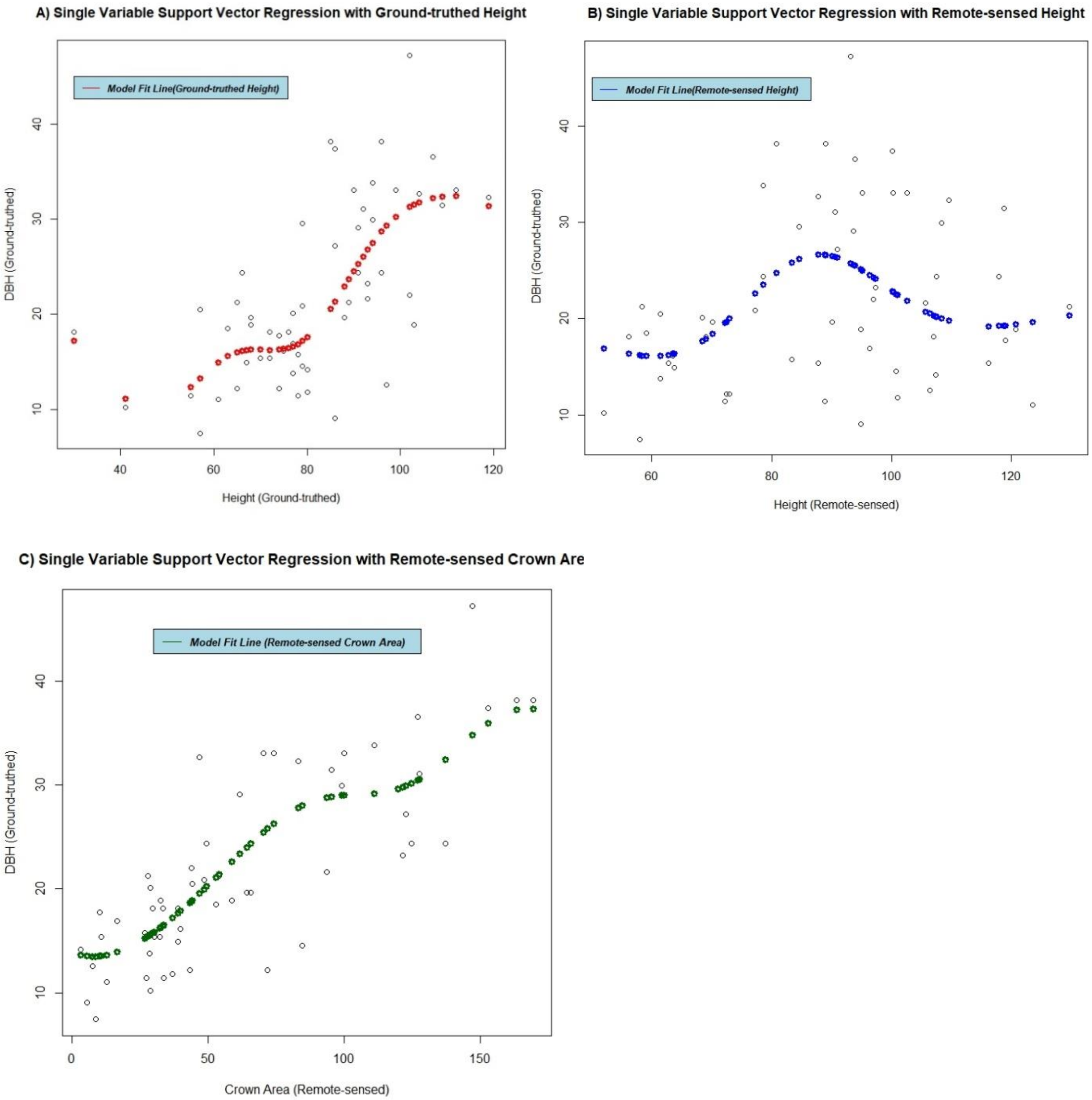


Douglas Fir Support Vector Regression Values	**R^2 value**	**RMSE of DBH (inches)**	**MAE of DBH (inches)**
A) SVR X = Height (ground-truthed)	0.76	3.73	2.61
B) SVR X = Height (remotely sensed)	0.55	5.06	3.85
C) SVR X = Crown Area	0.51	5.29	4.02

Figure 30: Single variable support vector model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Douglas fir DBH. A) X = Height (ground-truthed), B) X = Height (remote-sensed), C) X = Crown Area.

Bigleaf Maple Single Variable Support Vector Results

Ground-truthed height used as the explanatory variable produces predicted values closer to the ground-truthed DBH of Bigleaf Maple trees (R^2 = 0.47) when compared to remote-sensed height (R^2 = 0.19), and both height measurements underperform the model using crown area as the X variable (R^2 = 0.66) (Figure 31). On average across the three models there was an R^2 score increase of 0.08 with the remote sensed height seeing a larger gain than the other two models. RMSEs lowered on average by 0.39 inches. Again, the main benefactor was the remote sensed height model, but keep in mind it was still a much lower performing model than the other two. MAEs dropped for the height values by 0.73 and 0.63 inches for ground-truthed and remote-sensed respectively seeing large value from this modelling method. With a dearth of total samples, and lower than desired elevation quality some of the height models may have overfit to the data with large swings toward outliers.

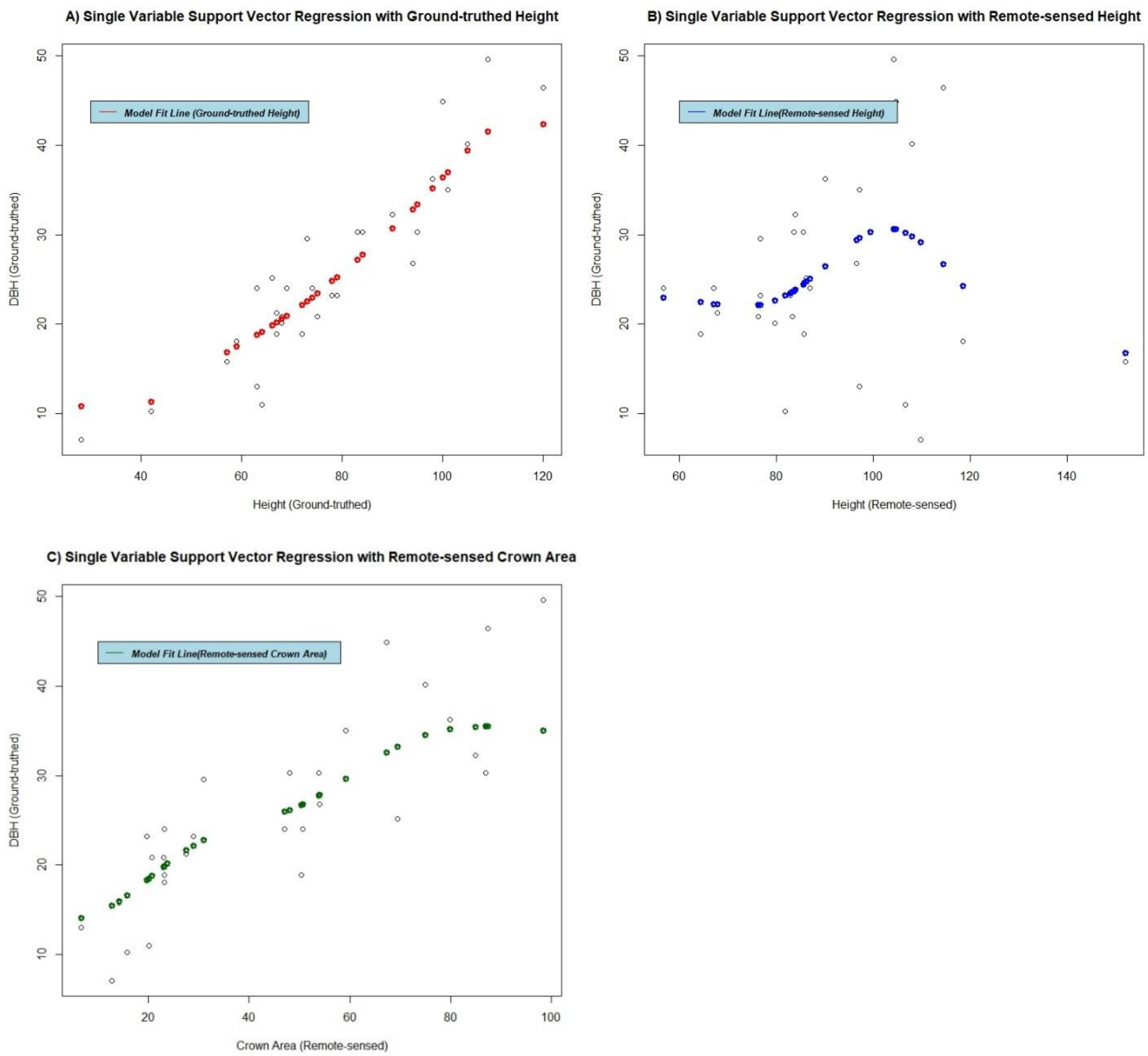


Bigleaf Maple Support Vector Regression Values	R^2 value	RMSE of DBH (inches)	MAE of DBH (inches)
A) SVR X = Height (ground-truthed)	0.47	6.52	4.78
B) SVR X = Height (remotely sensed)	0.19	8.02	6.42
C) SVR X = Crown Area	0.66	5.21	4.04

Figure 31: Single variable support vector model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Bigleaf Maple DBH. A) X = Height (ground-truthed), B) X = Height (remote-sensed), C) X = Crown Area.

Western Red Cedar Single Variable Support Vector Results

Ground-truthed height used as the explanatory variable produces predicted values closer to the ground-truthed DBH of Red Cedar trees (R^2 = 0.87) when compared to remote-sensed height (R^2 = 0.14). The model using crown area as the X variable (R^2 = 0.73) outperformed the remote-sensed height model but trailed behind the ground-truthed height model. (Figure 32). Red Cedar showed a similar level of improvement as Douglas fir from support vector regression modelling with a 0.07 average rise in R^2 value. RMSE values show an interesting change from the standard across the board reductions. While the height-based models saw reductions in RMSE the crown area model increased both in RMSE and MAE. It's possible that the models over fit to inadequate samples, or that it just underperforms the linear model for crown area. In this case even with a model that showed improvement, like the remote-sensed height data group, that does not necessarily mean that it fit well (Figure 32). In this case I believe that more samples needed to be collected and trained to alleviate some of this issue.



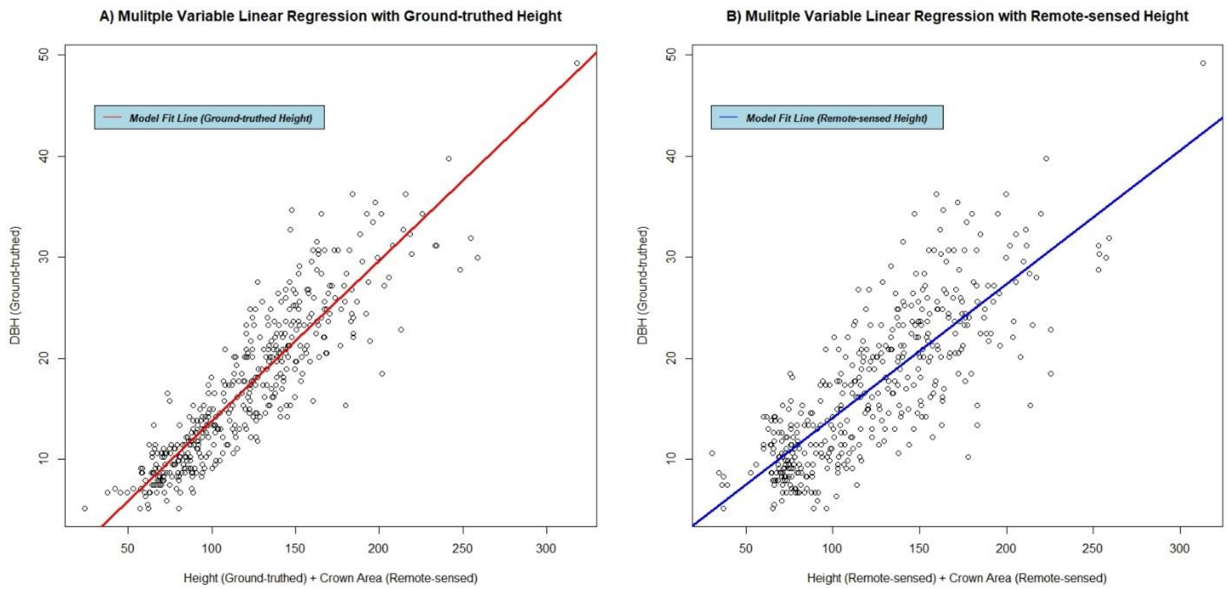
Red Cedar Support Vector Regression Values	R^2 value	RMSE of DBH (inches)	MAE of DBH (inches)
A) SVR X = Height (ground-truthed)	0.87	4.02	3.18
B) SVR X = Height (remotely sensed)	0.14	9.66	6.99
C) SVR X = Crown Area	0.73	5.79	4.44

Figure 32: Single variable support vector model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Red Cedar DBH. A) X = Height (ground-truthed), B) X = Height (remote-sensed), C) X = Crown Are

Multiple Variable Linear Regression

Douglas fir Multiple Variable Linear Regression Results

When using multiple variable linear regression, the model with ground-truthed height used as the X_1 and crown area as the X_2 variable predicted values closer to the ground-truthed DBH of Douglas fir trees ($R^2 = 0.82$) when compared to the model using remote-sensed height and crown area ($R^2 = 0.65$) (Figure 33). There is an expectation of an increase in R^2 value due to the use of multiple variables; it shifts the focus toward reductions in error. RMSEs dropped by over half an inch in both cases. This shows that more accurate results are achieved when incorporating both variables into the model. MAE only dropped by around a quarter inch on average. Using multiple variable regression must reduce the impact of the outliers affecting the RMSE which closes the gap slightly with the MAE. Even with improvements in the modeling for remote-sensed data it seems to still struggle with the higher end data.

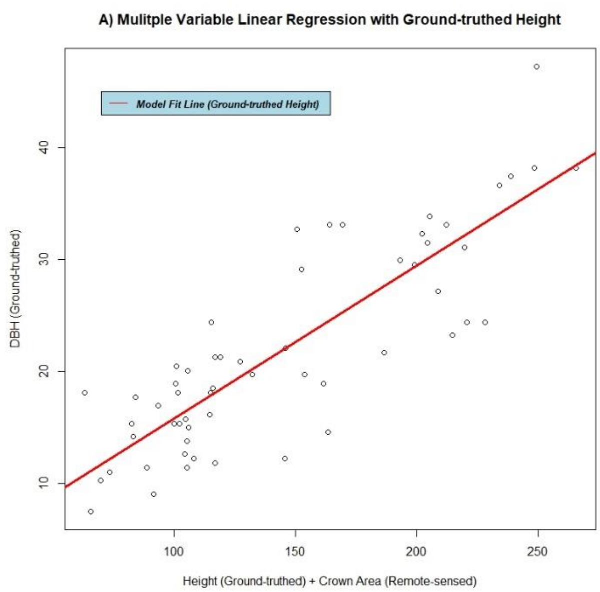


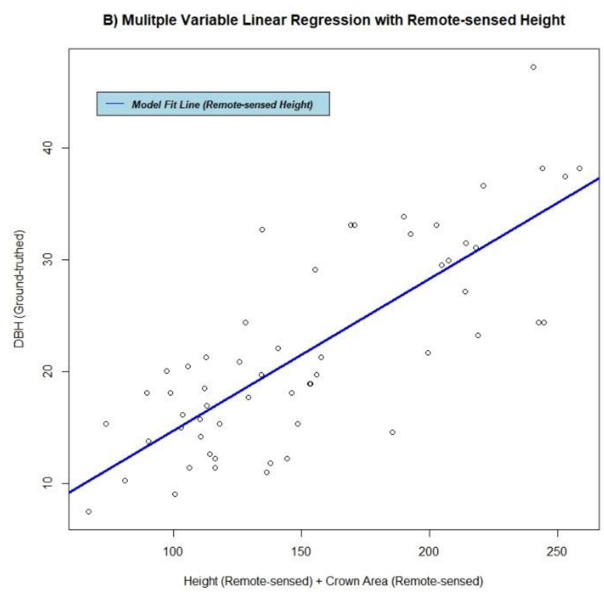
Douglas Fir Multiple Variable Linear Regression Values	R^2 value	RMSE of DBH (inches)	MAE of DBH (inches)
A) MLR (ground-truthed height)	0.82	3.20	2.48
B) MLR (remotely sensed height)	0.65	4.48	3.56

Figure 33: Multiple variable linear regression model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Douglas fir DBH. A) X_1 = Height (ground-truthed) and X_2 = Crown Area, B) X_1 = Height (remote-sensed) and X_2 = Crown Area.

Bigleaf Maple Multiple Variable Linear Regression Results

When using multiple variable linear regression, the model with ground-truthed height used as the X_1 and crown area as the X_2 variable predicted values closer to the ground-truthed DBH of Bigleaf Maple trees (R^2 = 0.71) when compared to the model using remote-sensed height and crown area (R^2 = 0.64) (Figure 34). The area of remote sensed data sees the most drastic increase in quality. Recall that in the best case prior the value of remote-sensed height as a single x variable was 0.13 showing very little model explanation of the variances in the data to a now strong explanatory ability. Again, we see a ~0.50-inch reduction in RMSE of the ground-truthed data. The remote sensed data is brought significantly down. When looking at the remote-sensed height as a single variable in the support vector regression the RMSE was 8.02. Now with the crown area added the RMSE drops to 5.21 inches. The largest improvements appear to be in the remote-sensed data since the crown area is helping to correct some of the inaccuracies. Comparing the remote-sensed data model to the support vector model that only used crown area shows a decrease in R2 value and increase in MAE. This suggests that the accuracy of the crown area relationship was hampered by the reduced height consistency.



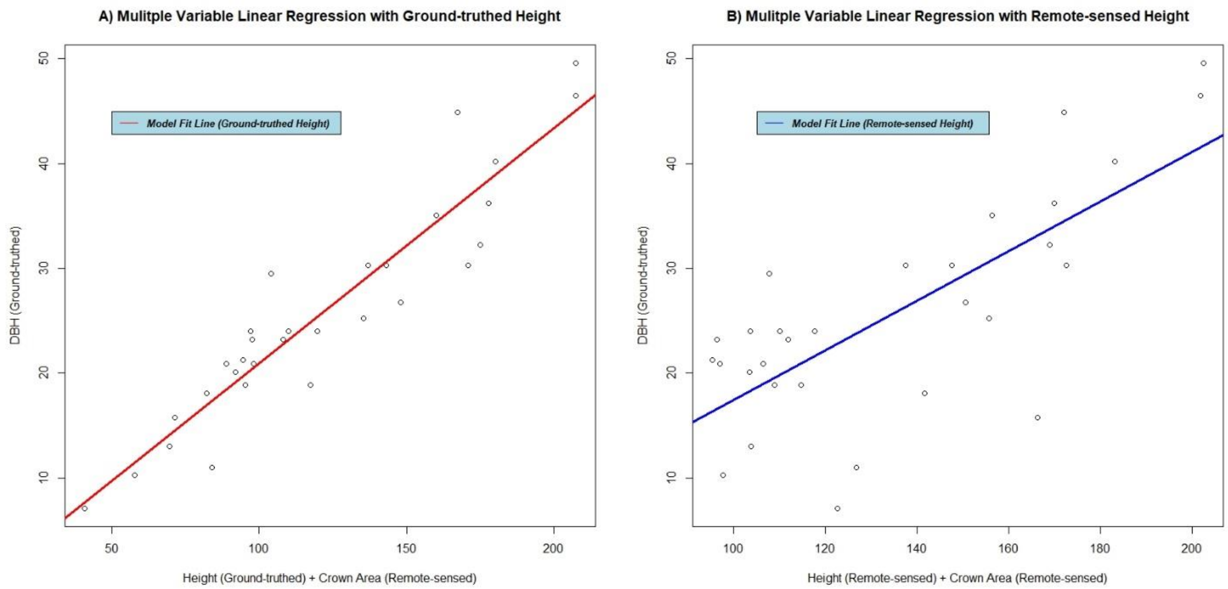


Bigleaf Maple Multiple Variable Linear Regression Values	R^2 value	RMSE of DBH (inches)	MAE of DBH (inches)
A) MLR (ground-truthed height)	0.71	4.75	3.72
B) MLR (remotely sensed height)	0.64	5.21	4.08

Figure 34: Multiple variable linear regression model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Bigleaf Maple DBH. A) X_1 = Height (ground-truthed) and X_2 = Crown Area, B) X_1 = Height (remote-sensed) and X_2 = Crown Area.

Western Red Cedar Multiple Variable Linear Regression Results

When using multiple variable linear regression, the model with ground-truthed height used as the X_1 and crown area as the X_2 variable predicted values closer to the ground-truthed DBH of Red Cedar trees (R^2 = 0.89) when compared to the model using remote-sensed height and crown area (R^2 = 0.71) (Figure 35). This set of models shows that the R2 scores aren't a perfect indicator of the quality of prediction. In the support vector single variable models performed above the peak R2 score for remote-sensed height data was 0.14 and crown area 0.73. We see here that with the two variables used in combination the R2 score drops to 0.71. The RMSE was reduced to 5.35, however, the MAE saw an increase from the 4.44 inches of single variable support vector regression using crown area to 4.49. This seems to be a continuation of the trend seen in the Bigleaf Maple models with multiple variables. The effect of confounding data and outliers is reduced which improves the RMSE and MAE which was largely unaffected by those issues sees reduced improvements. Similar to the Maple both fit lines are very similar (Figure 35). Slowly increasing accuracy of remote-sensed data will close that gap.



Red Cedar Variable Linear Regression Values	R^2 value	RMSE of DBH (inches)	MAE of DBH (inches)
A) MLR (ground-truthed height)	0.89	3.29	2.45
B) MLR (remotely sensed height)	0.71	5.35	4.49

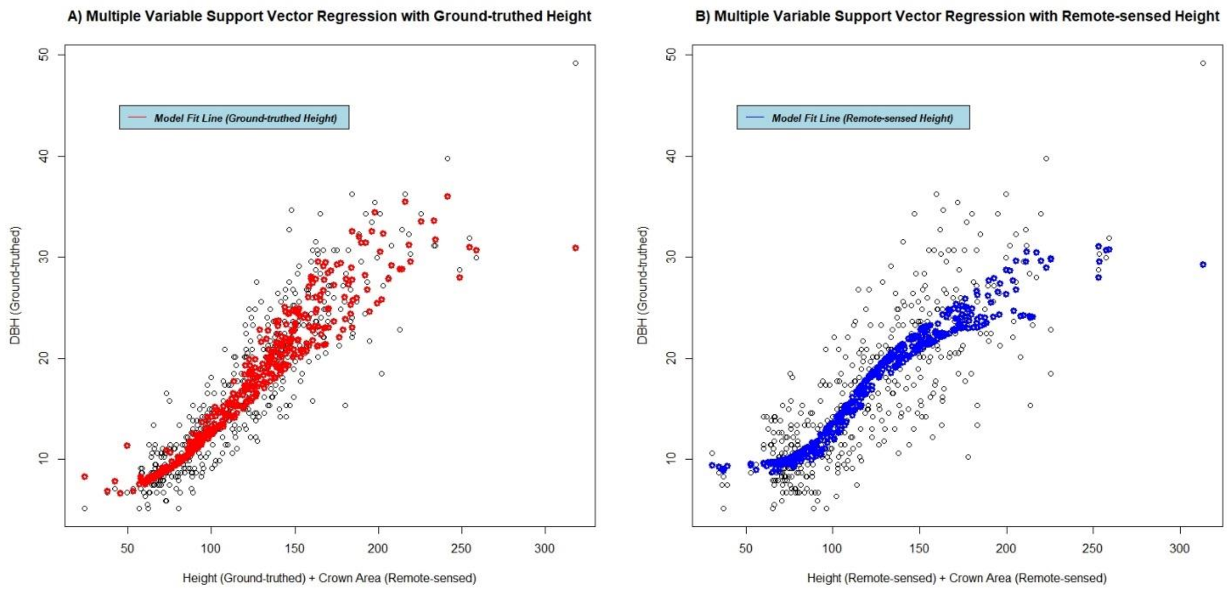
Figure 35: Multiple variable linear regression model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Red Cedar DBH. A) X_1 = Height (ground-truthed) and X_2 = Crown Area, B) X_1 = Height (remote-sensed) and X_2 = Crown Area.

Multiple Variable Support Vector Regression

In the same way that single variable attempts to improve upon single variable linear regression, multiple variable support vector regression attempts to increase model fit when compared to multiple variable linear regression using machine learning to explain the relationship between the Y variable and one or more X variables. As described in the methodology the variables are the same as what we would find in multiple variable linear regression with the Y variable being DBH, X_1 height and X_2 is the same crown area across both models.

Douglas fir Multiple Variable Support Vector Regression Results

When using multiple variable support vector regression, the model with ground-truthed height used as the X_1 and crown area as the X_2 variable predicted values closer to the ground-truthed DBH of Douglas fir trees ($R^2 = 0.84$) when compared to the model using remote-sensed height and crown area ($R^2 = 0.67$) (Figure36). At this point returns from model iterations are diminishing. It is in this group of models that the most accurate values have been attained, but the leaps are not nearly as drastic as in previous iterations. The final R^2 value for the remote-sensed height and crown area model ($R^2 = 0.67$) is very close to the 0.7 R^2 score that I had hoped to attain. Outliers seem to affect these models less, and the bulk of the fit conforms to the true data better. With a final RMSE of 3.0 inches and 4.35 inches for ground-truthed and remote-sensed respectively we see the most accurate values across all models when accounting for the specific height data used. MAE benefited similarly, and depending on future projects consideration of outliers this is quite impressive to have dropped to 2.23 inches.

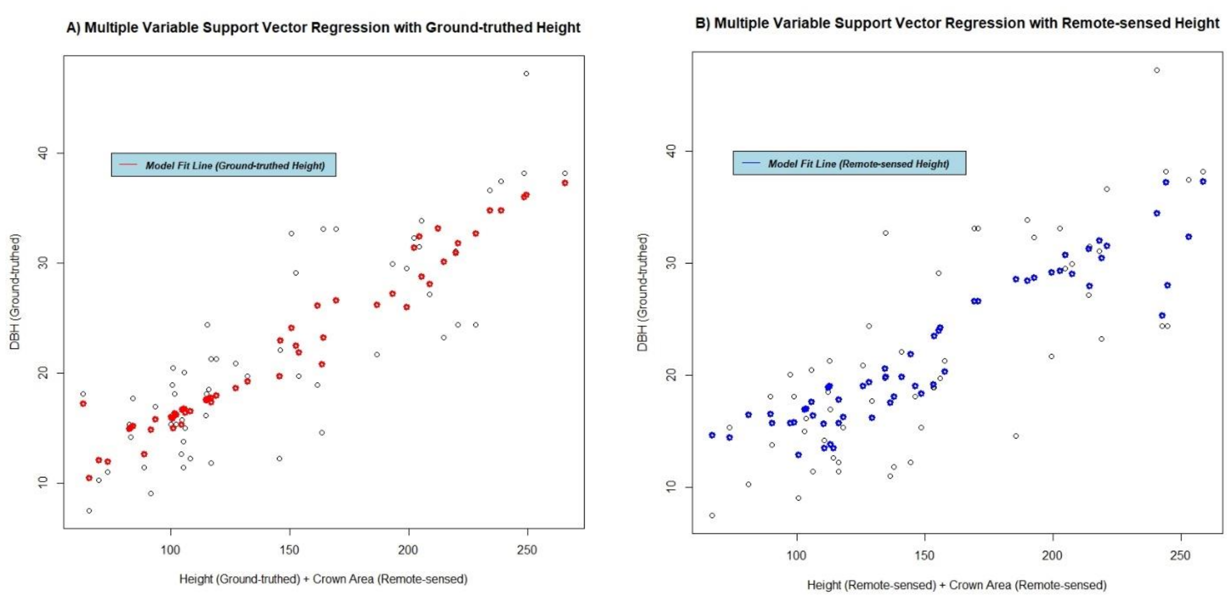


Douglas Fir Multiple Variable Support Vector Regression Values	**R^2 value**	**RMSE of DBH (inches)**	**MAE of DBH (inches)**
A) MSVR (ground-truthed height)	0.84	3.0	2.23
B) MSVR (remote-sensed height)	0.67	4.35	3.33

Figure 36: Multiple variable support vector regression model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Douglas fir DBH. A) X_1 = Height (ground-truthed) and X_2 = Crown Area, B) X_1 = Height (remote-sensed) and X_2 = Crown Area.

Bigleaf Maple Multiple Variable Support Vector Regression Results

When using multiple variable support vector regression, the model with ground-truthed height used as the X_1 and crown area as the X_2 variable predicted values closer to the ground-truthed DBH of Bigleaf Maple trees ($R^2 = 0.77$) when compared to the model using remote-sensed height and crown area ($R^2 = 0.70$) (Figure37). R^2 values both increased by 0.06 putting both models at or above the 0.7 goal threshold. RMSE as well as MAE dropped to the lowest values across all model types. Bigleaf Maple trees' relationship between height and crown area clearly benefits from the machine learning techniques used to attain these values and the lowered reliance on a linear relationship (Figure 37).

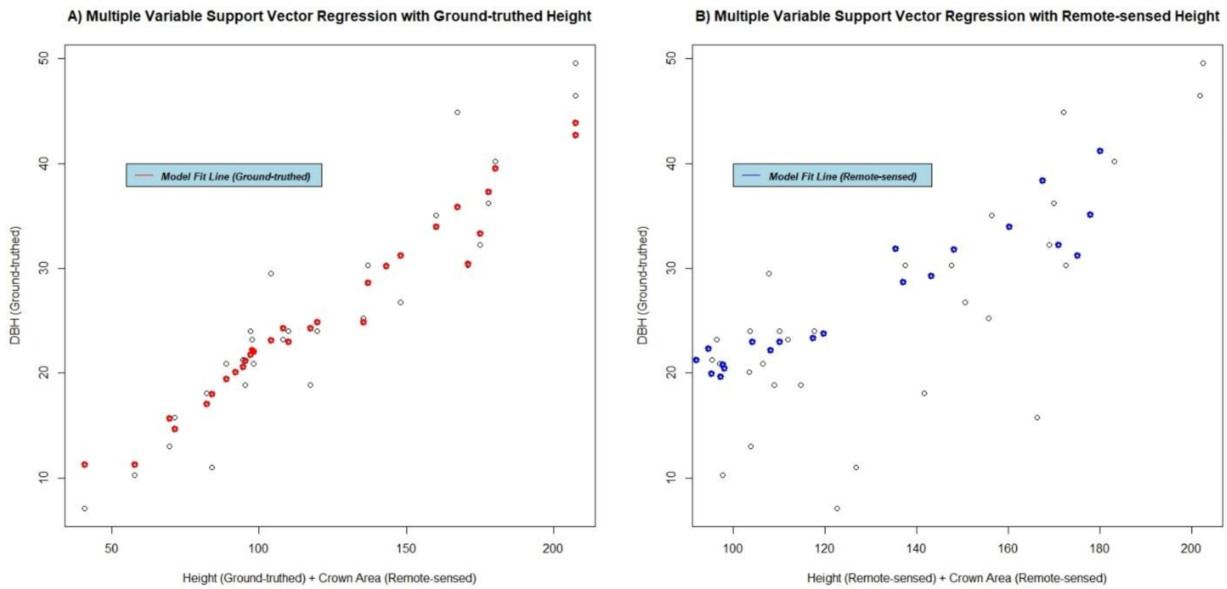


Bigleaf Maple Multiple Variable Support Vector Regression Values	R^2 value	RMSE of DBH (inches)	MAE of DBH (inches)
A) MSVR (ground-truthed height)	0.77	4.36	3.38
B) MSVR (remotely sensed height)	0.70	4.97	3.82

Figure 37 Multiple variable support vector regression model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Bigleaf Maple DBH. A) X_1 = Height (ground-truthed) and X_2 = Crown Area, B) X_1 = Height (remote-sensed) and X_2 = Crown Area.

Red Cedar Multiple Variable Support Vector Regression Results

When using multiple variable support vector regression, the model with ground-truthed height used as the X_1 and crown area as the X_2 variable predicted values closer to the ground-truthed DBH of Red Cedar trees ($R^2 = 0.92$) when compared to the model using remote-sensed height and crown area ($R^2 = 0.88$) (Figure38). The data seem to indicate that we are running into a bit of an accuracy ceiling with the ground-truthed data the remote-sensed data experienced dramatic improvements. The RMSE and MAE show little to no improvement when compared to the multiple variable linear models. The remote sensed data had the largest change in evaluation metrics when compared to the multiple variable linear regression. The R^2 value increased from 0.71 to 0.88, the RMSE dropped from 5.35 down to 3.73inches and the MAE dropped from 4.49 to 2.78 inches. Even with these dramatic improvements the model still struggles with smaller trees (Figure 38). Again, it should be noted that the results should be taken critically in this section as the number of samples collected was relatively small and under the threshold originally desired.



Red Cedar Multiple Variable Support Vector Regression Values	R^2 value	RMSE of DBH (inches)	MAE of DBH (inches)
A) MSVR (ground-truthed height)	0.92	3.30	2.32
B) MSVR (remotely sensed height)	0.88	3.73	2.78

Figure 38: Multiple variable support vector regression model fit lines graphed above with evaluation metrics in the table for three model variations used to predict Red Cedar DBH. A) X_1 = Height (ground-truthed) and X_2 = Crown Area, B) X_1 = Height (remote-sensed) and X_2 = Crown Area.

5.0 Conclusion

This research project contributes to the growing body of knowledge dealing with remote sensing estimates of diameter at breast height (DBH) measurement. Fast, accurate, and repeatable data collection is needed to respond to our quickly changing environment. This thesis builds on previous research using aerial vehicles to aid in the timeliness of data collection. While research with unmanned aerial vehicles (UAVs) is becoming more accessible and affordable, further research is needed to tease out more accurate forest composition data.

My first research question focused on how well remote-sensed estimation models predict ground-truthed DBH. The results showed robust predictive ability, especially with the manually measured tree heights. The results varied significantly when remotely sensed height was used as one of the variables. Using the ground-truthed height data, the multiple variable support vector regression (SVR) model was the top performer with a root mean square error (RMSE) of only approximately 3 inches. When using entirely remote data, the multiple variable SVR model again outperformed the others with an RMSE of 4.35 inches. An increase of 1.35 inches in the RMSE is significant, and whether that can be accepted depends on the parameters of future projects.

The second research question I focused on was whether methods of multivariate linear regression and support vector regression for Japanese Cypress DBH (Iizuka et al.2018, Iizuka et al. 2021) can be extrapolated to a regionally prevalent species Douglas fir. These two studies were foundational in terms of the design of this experiment. Both studies used remotely sensed data to test the allometric relationship found in Japanese Cypress trees. The author's found that the crown area for these trees had a strong correlation and that height was significantly less correlated with DBH size. Interestingly this project, when focusing on Douglas fir trees, found that height, both remote-sensed and ground-truthed, was a stronger predictor of DBH than remote-sensed crown area. This speaks to the fact that a universal formula is likely impossible across all species. This paper has shown that while relationships between tree metrics may differ between species, modeling techniques like linear regression, and especially support vector regression, is an effective way of estimating DBH. This project provides answers to those two questions; however, many remain to be researched more thoroughly. Problems arose over the course of the experiment that demonstrated some of the limitations with this method of data collection and analysis.

The predictive ability of the various models relies heavily on the quality of the images gathered, which in turn depends on local climatic factors. Both favorable weather conditions and lighting are crucial factors for obtaining high-quality UAV imagery. It was determined through this experiment that the solar azimuth had a distinct impact on image quality and, by proxy, the quality of any elevation or height models derived therein. On the mission dates, the conditions were mostly cloudy with mixed points of sunshine and relatively stable temperatures (46-48ºF) and moderate wind (9-13MPH). For future missions it would be beneficial to move the image capture dates further in the year between late May to early September. This would give the highest likely hood of clear, well lit, low wind conditions which in turn would aid in producing the highest quality image products.

This research has provided some of the building block foundational experimentation to further ecosystem monitoring and analysis by demonstrating the ability of regression models to predict the DBH of Douglas fir trees using remotely sensed data. Refinements in modeling and stitching software can and should be pursued for increased reliability. The main factors that should be adjusted and investigated in future research are mission plan variables such as height of flight, speed, overlap, ground control point (GCP) placement, etc., and potential automatic segmentation methods.

The pre-planned mission parameters for the drone flight have a significant impact on the resulting image quality. Future missions should keep this in mind and test alternative flight variables. When flying over forests one must weigh the pros and cons between flying higher for an increased likelihood of accurate image stitching, and lower altitude for increased image clarity and spatial resolution. Often the flight height varies between 100 – 400 feet (30- 120 meters). For this experiment we used an altitude of 380 feet (~115 meters) above the launch point. The goal was to achieve the best possible stitch quality given the weather constraints. Ideally a lower flight plan should be attempted to increase image clarity. This would only currently be feasible as recommended above, later in the year.

Proper placement and density of ground control points (GCPs) are vital to ensuring high-quality and spatially valid drone imagery stitching. These GCPs function as fixed, accurate, GPS coordinate points that tie images to the measured coordinates in aerial imagery and mapping. For this project there were limitations on our ability to place GCPs due to dense forest canopy and undergrowth. Ideally it would be best to have around twelve GCPs for a smaller drone mission, however a maximum of 2-5 points were established per flight session during the experiment (Yu et al., 2020, Tahar, 2013). In future attempts the use of an increased number of GCPs should also increase the accuracy and quality of both the image stitching and elevation products.

Automatic segmentation of tree crowns, while not within this project's scope, used in further research will be helpful for consistency (Iizuka et al. 2021). One possible option for segmentation would be through marker-controlled watershed segmentation (MCWS) which uses the canopy height model (CHM) to predict treetops and use those as limiting markers to reduce overlap and overprediction. If the CHM, however, is not clear or contains artifacting, this process is almost unusable. Additionally, there is no species differentiation using MCWS so further analysis would need to be performed. Alternatively, the use of deep learning models especially specifically trained convolutional neural network models could be the best of both worlds. These models could provide both crown delineation as well as species identification given enough training samples and hyper parameter optimization.

With the increased adoption of UAVs, we will soon see an increase in our ability to quickly assess forests for health, compositional change, and forest volume. On varying levels of scale groups will be able to respond more effectively to future challenges with the more detailed understanding of our environment that remote sensing will provide. I am confident that with enough research and adoption we will see UAV analysis change the way researchers, land managers, and governments interact with the environment. Key areas of opportunity for future use include forest inventory and analysis, fire management, and conservation and restoration.

When managing land from national inventories down to city parks, quickly updated data of the different dimensions of the forest are vital to informed decision making. One of the main tools used to assess the national size and density of forests in the United States is the FIA. This process, while thorough, can take years to complete. One way to reduce the time needed for these inventories would be to incorporate UAVs. LIDAR equipped UAVs are especially capable when it comes to measuring dense forest stands. Looking to the future, satellite imagery is increasing in quality and resolution and will continue to advance in its ability to help monitor large scale even global trends. Multispectral drones that can capture different wavelengths of light can be used in normalized difference vegetation index (NDVI) analysis. NDVI analysis allows researchers to better understand the health of vegetation within the area of interest. Since small UAVs are relatively cheap and convenient to launch this opens the door to temporal analysis and increased temporal resolution. With rapidly changing climate conditions, monitoring, tracking, and responding to altered forest composition and health is crucial to long term restoration and conservation.

Over the course of this study, it has been demonstrated that DBH can be estimated using remote sensed data of Douglas fir trees with accuracy that is heavily tied to the quality of the elevation and image products. This process is appealing due to the reduced manpower needs when compared to traditional inventory measurements. The potential benefit of incorporating UAVs into forestry and restoration is quite promising. However, research and replication is needed to fully capitalize on the opportunity before us. It is imperative that we continue to learn more about the ways we can support and strengthen our natural environment and all the tools that can make this process easier.

References

Alvarez, M. (2022). *Forest Inventory and Analysis, Fiscal Year 2020 Business Report*. https://www.fs.usda.gov/sites/default/files/fs_media/fs_document/FS%20FIA%20Fiscal%20Year%202020%20BusinessReport_508.pdf

Asiltürk, İ., & Çunkaş, M. (2011). Modeling and prediction of surface roughness in turning operations using artificial neural network and multiple regression method. *Expert Systems with Applications*, *38*(5), 5826–5832. https://doi.org/10.1016/j.eswa.2010.11.041

Cabán, A. G. (2007). Wildland fire management policy and fire management economic efficiency in the USDA Forest Service. In *keynote presented at the IV International Wildland Fire Conference* (pp. 13-17).

Bartelink, H. H. (1996). Allometric relationships on biomass and needle area of Douglas-fir. *Forest Ecology and Management*, *86*(1), 193–203. https://doi.org/10.1016/S0378-1127(96)03783-8

Bivand, Roger, Keitt, Tim, Rowlingson, Barry (2022). rgdal: Bindings for the 'geospatial' data abstraction library. R package version 1.5-32. https://CRAN.R project.org/package=rgdal

Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983). 1—AN OVERVIEW OF MACHINE LEARNING. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine Learning* (pp. 3–23). Morgan Kaufmann. https://doi.org/10.1016/B978-0-08-051054-5.50005-4

Ceballos, G., Ehrlich, P. R., & Raven, P. H. (2020). Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences*, *117*(24), 13596–13602. https://doi.org/10.1073/pnas.1922686117

Cervantes, J., Garcia-Lamont, G., Rodriguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Dainelli, R., Toscano, P., Di Gennaro, S. F., & Matese, A. (2021). Recent Advances in Unmanned Aerial Vehicle Forest Remote Sensing—A Systematic Review. Part I: A General Framework. *Forests*, *12*(3), Article 3. https://doi.org/10.3390/f12030327

Dainelli, R., Toscano, P., Di Gennaro, S. F., & Matese, A. (2021b). Recent Advances in Unmanned Aerial Vehicles Forest Remote Sensing—A Systematic Review. Part II: Research Applications. *Forests*, *12*(4), Article 4. https://doi.org/10.3390/f12040397

Deep Learning vs. Machine Learning: Beginner's Guide. (2022). Coursera. Retrieved October 25, 2022, from https://www.coursera.org/articles/ai-vs-deep-learning-vs-machine-learning-beginners-guide

Edson, C., & Wing, M. G. (2011). Airborne Light Detection and Ranging (LiDAR) for Individual Tree Stem Location, Height, and Biomass Measurements. *Remote Sensing*, *3*(11), Article 11. https://doi.org/10.3390/rs3112494

Fryd, O., Pauleit, S., & Bühler, O. (2012). The role of urban green space and trees in relation to climate change. *CABI Reviews*, *2011*, 1–18. https://doi.org/10.1079/PAVSNNR20116053

Gallardo-Salazar, J. L., & Pompa-García, M. (2020). Detecting Individual Tree Attributes and Multispectral Indices Using Unmanned Aerial Vehicles: Applications in a Pine Clonal Orchard. *Remote Sensing*, *12*(24), Article 24. https://doi.org/10.3390/rs12244144

Gayon, J. (2000). History of the Concept of Allometry1. *American Zoologist*, *40*(5), 748–758. https://doi.org/10.1093/icb/40.5.748

Gu, J., Grybas, H., & Congalton, R. G. (2020). A Comparison of Forest Tree Crown Delineation from Unmanned Aerial Imagery Using Canopy Height Models vs. Spectral Lightness. *Forests*, *11*(6), Article 6. https://doi.org/10.3390/f11060605

Hassani, N. J. M. (2018, May 31). What is a Forest Inventory? *Forestrypedia*. https://forestrypedia.com/what-is-forest-inventory/

Hijmans, Robert J. (2022). raster: geographic data analysis and modeling. R package version 3.5-29. https://CRAN.R-project.org/package=raster

Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314. https://doi.org/10.1080/10618600.1996.10474713

Iizuka, K., Kosugi, Y., Noguchi, S., & Iwagami, S. (2022). Toward a comprehensive model for estimating diameter at breast height of Japanese cypress (Chamaecyparis obtusa) using crown size derived from unmanned aerial systems. *Computers and Electronics in Agriculture*, *192*, 106579. https://doi.org/10.1016/j.compag.2021.106579

Iizuka, K., Yonehara, T., Itoh, M., & Kosugi, Y. (2018). Estimating tree height and diameter at breast height (DBH) from digital surface models and orthophotos obtained with an unmanned aerial system for a japanese cypress (Chamaecyparis Obtusa) forest. *Remote Sensing*, *10*(1), Article 1. https://doi.org/10.3390/rs10010013

IPCC, 2022: *Climate Change 2022: Impacts, Adaptation, and Vulnerability.* Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:10.1017/9781009325844.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2022). *An introduction to statistical learning with applications in R:* (Vol. 6). https://link-springer-com.evergreen.idm.oclc.org/book/10.1007/978-1-4614-7138-7

Jiao, S., Gao, Y., Feng, J., Lei, T., Lei, T., Yuan, X., & Yuan, X. (2020). Does deep learning always outperform simple linear regression in optical imaging? *Optics Express*, *28*(3), 3717–3731. https://doi.org/10.1364/OE.382319

Joyce, D. K. (2022, February 9). What is the resolution of drone mapping? *GeoNadir*. https://geonadir.com/drone-resolution/

Komárek, J., Klápště, P., Hrach, K., & Klouček, T. (2022). The Potential of Widespread UAV Cameras in the Identification of Conifers and the Delineation of Their Crowns. *Forests*, *13*(5), Article 5. https://doi.org/10.3390/f13050710

Lim, Y., La, H., Park, J., Lee, M., Pyeon, M., & Kim, J.-I. (2015). Calculation of Tree Height and Canopy Crown from Drone Images Using Segmentation. *Journal of the Korean Society of*

Surveying, Geodesy, Photogrammetry and Cartography, *33*, 605–613. https://doi.org/10.7848/ksgpc.2015.33.6.605

Machado, M., Tommaselli, A., Tachibana, V., Pereira Martins-Neto, R., & Campos, M. (2019). *Evaluation of multiple linear regression model to obtain DBH of trees using data from a lightweight laser scanning system on-board UAV*.

Matese, A., Toscano, P., Di Gennaro, S. F., Genesio, L., Vaccari, F. P., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., & Gioli, B. (2015). Intercomparison of UAV, Aircraft and Satellite Remote Sensing Platforms for Precision Viticulture. *Remote Sensing*, *7*(3), Article 3. https://doi.org/10.3390/rs70302971

Mencuccini, M., & Grace, J. (1995). Climate influences the leaf area/sapwood area ratio in Scots pine. *Tree Physiology*, *15*(1), 1–10. https://doi.org/10.1093/treephys/15.1.1

New York State Department of Environmental Conservation. (n.d.). *DBHguidelines.pdf*. Retrieved December 29, 2022, from https://www.dec.ny.gov/docs/lands_forests_pdf/DBHguidelines.pdf

Odemakinde, E. (2022, March 19). *Everything about Mask R-CNN: A Beginner's Guide*. Viso.Ai. https://viso.ai/deep-learning/mask-R-CNN/

Olympia Parks and Rec (n.d.). *LBA Park and Woods* https://www.olympiawa.gov/services/parks___recreation/parks___trails/lba_park_and_woods.php

Onishi, M., & Ise, T. (2021). Explainable identification and mapping of trees using UAV RGB image and deep learning. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-020-79653-9

Plowright, Andrew. (2021). *Introduction*. https://cran.r-project.org/web/packages/ForestTools/vignettes/treetop_analysis.html

Plowright, Andrew, and Roussel, Jean-Romain. (2021). ForestTools: Analyzing Remotely Sensed Forest Data. R package version 0.2.5.

https://CRAN.R-project.org/package=ForestTools

Popescu, S. C., & Wynne, R. H. (2004). Seeing the Trees in the Forest. *Photogrammetric Engineering & Remote Sensing*, *70*(5), 589–604. https://doi.org/10.14358/PERS.70.5.589

Powell, D. C., & Silviculturist, F. (2016). *WHITE PAPER F14-SO-WP-SILV-24*.

Roussel, J.-R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R. H., Meador, A. S., Bourdon, J.-F., de Boissieu, F., & Achim, A. (2020). LidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment*, *251*, 112061. https://doi.org/10.1016/j.rse.2020.112061

Roussel J, Auty D (2022). *Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*. R package version 4.0.1, https://cran.r-project.org/package=lidR.

Shaw, J. (2008). Benefits of a strategic national forest inventory to science and society: The USDA Forest Service Forest Inventory and Analysis program. *IForest - Biogeosciences and Forestry*, *1*(1), 81–85. https://doi.org/10.3832/ifor0345-0010081

Shaw, J. D. (2017). Introduction to the Special Section on Forest Inventory and Analysis. *Journal of Forestry*, *115*(4), 246–248. https://doi.org/10.5849/JOF-2017-002R1

Snyder, M. (2006). *What is DBH? | Summer 2006 | Articles | W*. https://northernwoodlands.org/articles/article/what_is_DBH

Syuukaku, K. (2019). (rep.). *Fiscal 2019 Commissioned Project for Studying Efficiency Methods for Harvest Surveys Utilizing Remote Sensing Technology* (pp. 79–85).

Sun, Y., Li, Z., He, H., Guo, L., Zhang, X., & Xin, Q. (2022). Counting trees in a subtropical mega city using the instance segmentation method. *International Journal of Applied Earth Observation and Geoinformation*, *106*, 102662. https://doi.org/10.1016/j.jag.2021.102662

Surový, P., & Kuželka, K. (2019). Acquisition of Forest Attributes for Decision Support at the Forest Enterprise Level Using Remote-Sensing Techniques—A Review. *Forests*, *10*(3), Article 3. https://doi.org/10.3390/f10030273

Tang, L., & Shao, G. (2015). Drone remote sensing for forestry research and practices. *Journal of Forestry Research*, *26*, 791–797. https://doi.org/10.1007/s11676-015-0088-y

Tahar, K. N. (2013). An evaluation on different number of ground control points in unmanned aerial vehicle photogrammetric block. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-2/W2*, 93–98. https://doi.org/10.5194/isprsarchives-XL-2-W2-93-2013

Town, C. (2018, July). *USDA Forest Inventory Methods*. Forestry Technical Note No. FOR-1. https://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=42554.wba

USDA Forest Service. (1992). *Fs_fia-overview.pdf*. Forest Service Resource Inventories an Overview. https://www.srs.fs.usda.gov/pubs/misc/fs_fia-overview.pdf

U.S. Department of Agriculture, Forest Service. 2019. Urban tree canopy assessment: a community's path to understanding and managing the urban forest. FS–1121. Washington, DC. 16 p https://www.fs.usda.gov/sites/default/files/fs_media/fs_document/Urban%20Tree%20Canopy%20paper.pdf

USGS. (n.d.). *What is remote sensing and what is it used for? | U.S. Geological Survey*. Retrieved November 13, 2022, from https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used

Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., & Cottrell, G. (2018). *Understanding Convolution for Semantic Segmentation* (arXiv:1702.08502). arXiv. http://arxiv.org/abs/1702.08502

Xiang, T.-Z., Xia, G.-S., & Zhang, L. (2019). Mini-Unmanned Aerial Vehicle-Based Remote Sensing: Techniques, Applications, and Prospects. *IEEE Geoscience and Remote Sensing Magazine*, *7*(3), 29–63. https://doi.org/10.1109/MGRS.2019.2918840

Younglfesh, C., Socolar, J., Amaral, B., Arab, A., Guralnick, R., Hurlbert, A., LaFrance, R., Mayer, S., Miller, D., Tingley, M. (2021). Migratory strategy drives species-level variation in bird sensitivity to vegetation green-up. *Nature Ecology and Evolution,* DOI:10.1038/s41559-021-01442-y.

Yu, J. J., Kim, D. W., Lee, E. J., & Son, S. W. (2020). Determining the optimal number of ground control points for varying study sites through accuracy evaluation of unmanned aerial system-based 3D point clouds and digital surface models. *Drones*, *4*(3), 49. MDPI AG. Retrieved from http://dx.doi.org/10.3390/drones4030049

Zhou, Z.-H. (2021). *Machine Learning*. Springer Singapore. https://doi.org/10.1007/978-981-15-1967-3