SCHOOL FOR ENVIRONMENT AND SUSTAINABILITY
UNIVERSITY OF MICHIGAN

# Characterizing the Urban Tree Canopy (UTC) to Elevate its Role in Mitigating Climate Change and Creating a Healthy and Vibrant Community in Ann Arbor, MI

by

Thomas Estabrook, Christian Schluter, Alyssa Sklar, Lyndsay Zemanek

A project submitted
in partial fulfillment of the requirements
for the degree of
Master of Science in Environment & Sustainability
at the University of Michigan
April 2022

Faculty Advisor: *Shannon Brines*
Client: *City of Ann Arbor Office of Sustainability and Innovations*

## Executive Summary

Ann Arbor, Michigan is affectionately called "Tree Town" due to its distinguishable and valued urban tree canopy (UTC). The city consists of roughly 18,605 acres with 6,015 acres identified as UTC in 2010 (Hanou, 2010). The City of Ann Arbor recognizes the importance and potential benefits the UTC can provide in its "Urban and Community Forest Management Plan" and also mentions them in the "A$^2$Zero Climate Action Plan". While the forest remains a priority, the management plan was implemented eight years ago, and the last canopy assessment occurred twelve years ago. Furthermore, this assessment delineates where the canopy is but does not provide details about canopy composition. Additionally, data is available regarding the City's street trees, but certain locations, such as parks and private land, have little available canopy data. Like many city governments working to elevate positive UTC contributions to the community at large, Ann Arbor is in need of descriptive analyses to enhance current data.

Our goals were to identify turfgrass, delineate native forest fragments, and classify trees by genus. By emphasizing these three areas, we sought to elevate the role of UTC as an ecosystem service, mitigator of climate change, and guiding factor in stewardship actions.

Turfgrass was identified to inform residents and decision-makers about its spatial extent. This would highlight locations that could receive greater ecosystem service benefits by reducing turfgrass to expand the urban tree canopy. To complete this analysis, aerial imagery was clipped using a Canopy Height Model, Normalized Difference Vegetation Index (NDVI) thresholds indicating live green vegetation, and an Unsupervised Classification was performed. From our assessment, turfgrass accounts for 50% of the total canopy.

Native forest fragments were delineated both manually and with an unsupervised classification. Manual delineation was based on historic imagery dating back to the 1940s and refined using a Canopy Height Model. The unsupervised classification aided in identifying smaller forest fragments. Our manual delineation accounts for 14% of the total canopy, with two-thirds located outside of City-owned property. The classification results indicated native forest fragments comprising 28% of the total canopy, with three-fourths located outside of City-owned property.

Genus classification was attempted to allow one to easily characterize the types of forest communities at a large scale and to aid in identifying invasive species. Tree data was collected to utilize as training/testing data, various predictor layers were associated with tree points, and machine learning models were assessed. Overall, LiDAR segmentation and the Random Forest model performed the best with the highest accuracy genus classification at 54%.

While we were able to provide descriptive analyses of the UTC to enhance the City's current dataset, the accuracy and reproducibility of our methods should be improved with future work. All three areas of focus could benefit from a better spatio-temporal alignment of imagery and tree data along with collecting ground truth data to test accuracy.

## Acknowledgements

We would like to acknowledge and offer our sincere gratitude to the many people who made this research possible. First and foremost, we would like to thank our advisor, Shannon Brines, for sharing his expert knowledge and providing moral support throughout the duration of this project.

This project would not be possible with the guidance from Jason Tallant, Data Manager and Research Scientist at the University of Michigan Biological Station.

At the City of Ann Arbor, we would like to thank the Office of Sustainability and Innovations, Natural Features Working Group, and Environmental Commission – especially our client representative, Christopher Graham, Chair of Natural Features Working Group and Environmental Commissioner. We also want to thank Tiffany Giacobazzi, Urban Forestry and Natural Resources Planning Coordinator; Sean Reynolds, Community Sustainability Coordinator; Dr. Missy Stults, Sustainability and Innovations Manager; Rita Mitchell, Member of the Environmental Commission; and Anthony Bedogne, Senior Application Specialists GIS.

We also extend our appreciation to faculty members at the School for Environment and Sustainability (SEAS), family, and friends who supported us through this process.

## Table of Contents

# Chapter 1: Urban Tree Canopy of Ann Arbor, Michigan

## Introduction

The City of Ann Arbor is home to a rich canopy of nearly 1.5 million trees that provide an estimated $4.6 million in annual benefits (i-Tree Ecosystem Analysis: Ann Arbor, 2013; *The Urban Forest,* n.d.). These urban tree canopy (UTC) benefits include removing air pollution, lowering energy costs, reducing carbon dioxide levels, increasing property values, and reducing stormwater run-off. Much of the information the city has available comes from a 2014 management plan and 2010 canopy analysis. And the latest digitized natural features inventory only includes street trees, a relatively small part of the overall urban canopy. Moreover, there exists no publicly available data distinguishing undisturbed native forest fragments from newer stands on recently disturbed land. This lack of descriptive analyses is the main problem the City of Ann Arbor faces.

To address this problem with a comprehensive census of all of Ann Arbor's million-plus trees would be prohibitively time-consuming and expensive. However, recent advances in machine learning make it possible to study UTC using widely available remote sensing data. In broad terms, leveraging machine learning algorithms will automate the task of analyzing specific aspects of the UTC.

In our project, we focused on identifying turfgrass, delineating native forest fragments, and classifying trees by genus. By emphasizing these three areas, we sought to elevate the role of UTC as an ecosystem service, a mitigator of climate change, a guiding factor in stewardship actions, and an ensurer of environmental equity.

### *Client and Site*

The City of Ann Arbor Office of Sustainability and Innovations (OSI) is tasked with developing environmental policies to improve the natural and built environment while protecting and enhancing air, water, land, and public health. They are guided by the "$A^2$Zero Climate Action Plan" to ensure sustainability and equity throughout the city based on four main areas: 1) climate and energy, 2) community, 3) land use and access; and 4) resource management (Office of Sustainability and Innovations). The city recognizes the potential benefits the UTC can provide in its "Urban and Community Forest Management Plan" and also mentions them in the "$A^2$Zero Climate Action Plan". In the "$A^2$Zero Climate Action Plan", the main goal is to "preserve and enhance the local tree canopy" with the goal of planting 10,000 new trees on private property and 10,000 new trees on city property by 2030 (City of Ann Arbor, 2020). OSI emphasized understanding and learning more about the canopy details of privately owned land. This was directly taken into consideration in the analyses we performed. Our study area, Ann Arbor, Michigan (Figure 1.1), is roughly 18,605 acres with 6,015 acres identified as UTC in 2010 (Hanou, 2010). Additionally, 15% of property within Ann Arbor's city boundary is City-owned.

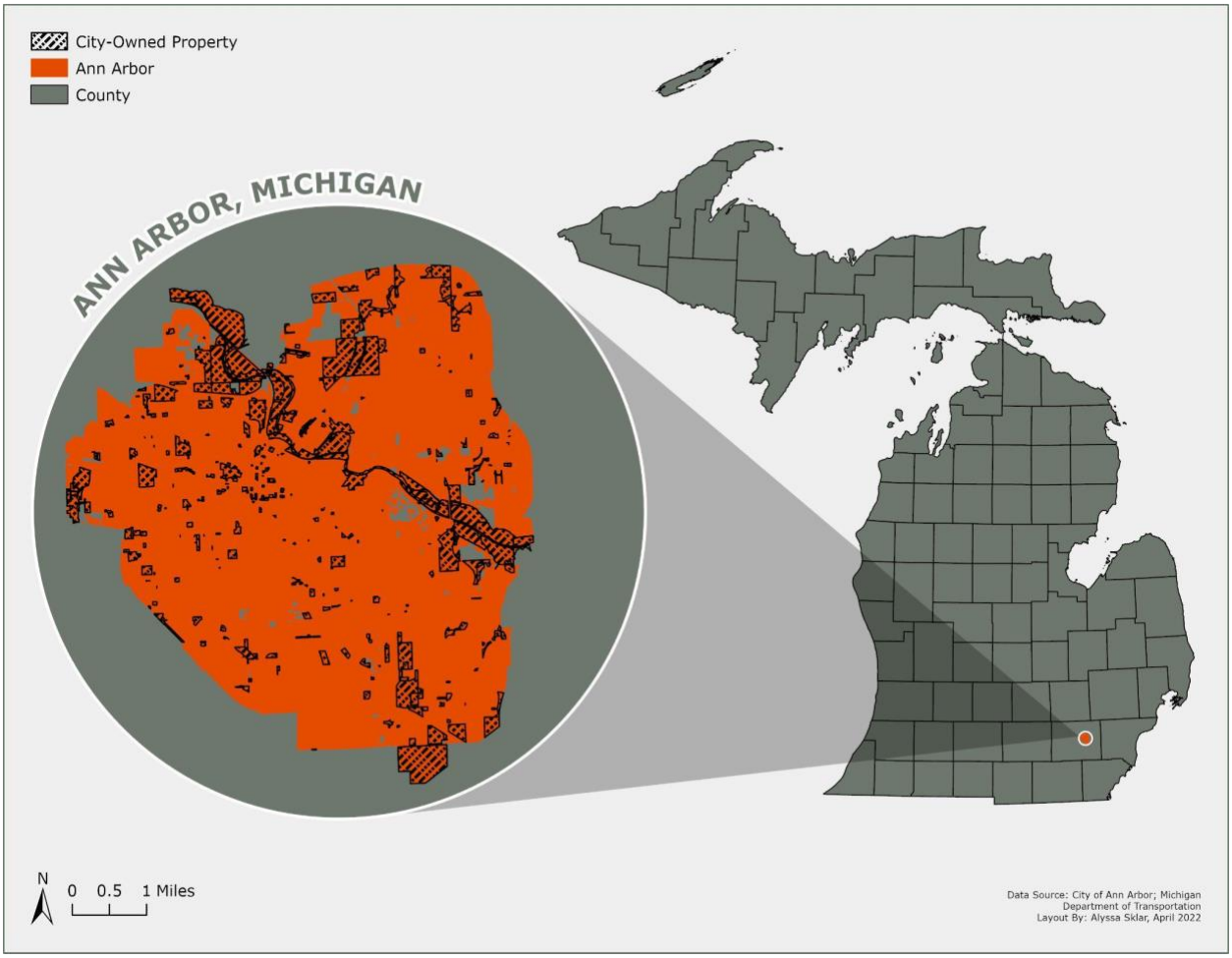



**Figure 1.1.**: Study area of Ann Arbor, Michigan located in Washtenaw County with City-owned property overlaid.

## Background

Urban Tree Canopy (UTC) cover contributes to environmental, economic, social, and health benefits to communities worldwide. It provides ecosystem services of carbon storage and sequestration (Nowak & Crane, 2002), energy conservation (McPherson & Simpson, 2003), air quality improvement (Nowak et al., 2006; Livesley c., 2016), stormwater runoff and watershed health (Berland et al., 2017; Decina et al., 2020), habitat preservation (Alvey, 2006; Ossola et al., 2019), and the amelioration of microclimate (Georgi & Zafiriadis, 2006). Economic benefits can include increasing property value and reducing building heating and cooling costs (Jones et al., 2012; Berry et al., 2013). Furthermore, UTC mitigates urban heat island effect (Loughner et al., 2012), offers aesthetic beauty and psychological benefits (Smardon, 1988), reduces stress (Hartig et al., 2014), creates a connection to nature or "biophilia" (Beatley, 2011), and increases a sense of safety and place (Kuo et al., 1998; Hausmann et al., 2015). While benefits may vary across time and space, overall, they persist.

Recognizing that UTC is a crucial aspect of urban landscapes and ecosystems, it is vital that UTC distribution is equitable and not a disservice. Several studies have found that low-income neighborhoods and metropolitan areas with higher percentages of Black, Indigenous, and People of Color tend to have low tree canopy cover compared to white neighborhoods (Landry &

Chakraborty, 2009; Schwarz et al., 2015; Locke et al., 2021). Furthermore, tree planting efforts can create community disservices due to fear of economic strain for the city to plant and maintain, and increased water consumption (Roy et al., 2012).

## Project Aims

In this project we characterized the urban tree canopy of Ann Arbor, MI to elevate its role in mitigating climate change and creating a healthy and vibrant community. Identifying, describing, and quantifying the UTC was carried out with particular interest in fragments of old-growth forest and turfgrass on privately-owned land where it is difficult for conservation planners to document.

We developed three data layers to support decision-makers in Ann Arbor's Environmental Commission and Office of Sustainability and Innovations. First, we created a map of native forest fragment locations. Second, we developed a map of turfgrass to aid in identifying tree planting sites and focus areas for incentivizing sustainable lawns. Third, we developed a map of Ann Arbor's urban canopy classified by tree genus.

As final deliverables, we created quality-tested data layers from native forest fragments and turfgrass analysis along with a detailed writeup of the genus classification process for future research to build on. We also assembled an ArcGIS StoryMap to convey our findings to stakeholders.

## Chapter Organization

This report is organized in three main sections: Native Forest Fragment Delineation, Turfgrass Estimation, and Genus Classification. Within each section, an introduction/background, methods, results, and discussion will describe the objective and relevant research pertaining to the topic, including a detailed workflow of parameters and spatial analyses explored, ending with the results, main takeaways, and potential future directions.

# Chapter 2: Native Forest Fragment Delineation

## Introduction

One of the objectives we were tasked with was the identification of native forest fragments within the city limits. Native forest land, also known as primary, old-growth, or virgin forest, is an important ecological resource. The difference between old-growth and second-growth forests is primarily found in the amount of disturbance the woodland has experienced, which greatly affects forest composition (White & Lloyd, 1994).

## Background

The City of Ann Arbor utilizes three chief categories to classify woodlands: pioneer woodlands, urban woodlands, and native forest fragment woodlands. These three categories have different values assigned to them, and their level of protection varies accordingly. Native forest fragment woodlands, or primary forest fragments, are awarded the greatest value for several reasons. Native forest woodlands are locations in which the original native trees have largely avoided disturbance, surviving Michigan's logging era. This means that native forest fragments are composed chiefly of mature, indigenous trees, fostering richer biodiversity and providing greater ecosystem services than that of second-growth woodlands. These services include the provision of habitat for native species, regulation and filtration of stormwater runoff, and a greater degree of carbon sequestration.

Due to the value represented by these native forest fragments, they are considered a priority for conservation efforts. While the City of Ann Arbor's Natural Features Working Group is aware of many native fragments on public land within the city, they did not have a consolidated inventory of these fragments and requested that we identify these fragments within the city. Of particular interest were those located on private property, as we were informed of the Natural Features Working Group's objective to reach out to private landowners about conserving the native forest fragments on their properties.

## Data

For this portion of the project, we utilized historic aerial imagery hosted by Washtenaw County, NAIP imagery, and LiDAR data.

| Name | Year | Format or Resolution |
|---|---|---|
| National Agriculture Imagery Program (NAIP) | 2018 | 2 feet |
| Historical Aerial Imagery | 1940 | Webmap layer (unknown resolution) |
| Light Detection and Ranging (LiDAR) | 2018 | 1,101,733,871 points, 1.246-1.9 feet point-spacing |

**Table 2.1.** Data used to delineate and classify native forest fragments.

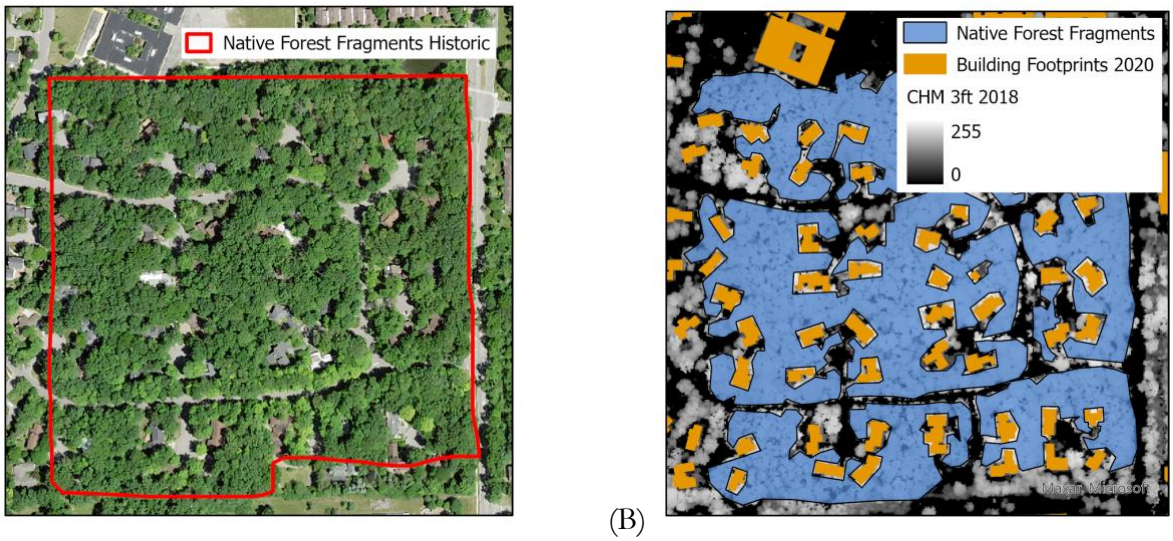
## Methods

The process we undertook to identify Ann Arbor's native forest fragments can be divided into two separate major processes: manual delineation and unsupervised classification. The first step we worked on was the manual delineation of native forest fragments. We began the identification process using Esri's basemap hybrid as a guide. Local native forest fragments are composed in large part of oak and hickory trees, which seem to change color slightly faster in the fall than many other species. This slight difference in coloration made it easier to spot possible fragment locations, which would be classified as either possible or likely fragments based on whether we could find any information supporting their existence as fragments, primarily based on information posted by the city's parks and recreation department. If a forested location was described as having seen little disturbance in the department's description, we would create a polygon representing that location and label it as a likely native fragment location.

After meeting and discussing with two of our clients, Jason Tallant and Christopher Graham, we altered and refined this methodology. Jason recommended we utilize historic aerial imagery from 1940 hosted by Washtenaw County. We agreed that any mature forest fragments displayed within this historic imagery could be considered a native fragment. We created a shapefile outlining the general location and area of mature forest fragments from 1940, and used this data, along with our canopy height model, to revise our native fragment location layer (Figure 2.1)

The second major process we undertook to delineate Ann Arbor's native forest fragments was an unsupervised classification. This was necessary due to the presence of many smaller fragments within the city – it had been expressed to us that it would be highly beneficial to be able to locate even the smallest of fragments, even down to the scale of a cluster of several trees. It was not feasible to manually locate and delineate every minor native fragment within the city, but with clustering we hoped to automate fragment identification.

We opted to perform an unsupervised classification rather than supervised. The three major land cover classes of interest (native forest, secondary forest, and shrubs) each encompass a diverse range of forest communities. We had no a priori reason to assume that these classes could be cleanly delineated based on the data we had access to and did not have the time or expertise to assemble adequate samples of each community type to train a classifier. Unsupervised classification elided these issues by identifying easily distinguished variations in the data, allowing us to manually consolidate and define the resulting clusters.




(A) (B)

**Figure 2.1.** (A) Native forest fragment location, based on historic imagery; (B) Native forest fragment location, refined using our Canopy Height Model. Note the careful exclusion of developed features (buildings, roads)

The predictor layers we used for the classification were: canopy height model, LiDAR PCA classes 1-4 (Appendix E), an NDVI layer calculated from NAIP imagery, a texture raster calculated from our CHM, and 4-band NAIP imagery. All layers were limited to a mask area calculated from canopy height and NDVI in order to only perform clustering on areas of tree canopy. For more information on how these layers were compiled, see Appendix A. After some preliminary testing, we settled on the following parameters: 20 individual land cover classes, with a resolution of five-by-five feet.

## Results

After examining the resulting land cover raster, we consolidated the original twenty land cover classes into twelve. As previously mentioned, there is a high degree of variance found within Ann Arbor's forested land, even between native fragments. This prevented us from being able to assign a single class to native forest fragments, although we identified two main classes. Class 4: mature emergent trees, and class 6: medium trees/under canopy, were found to be strong indicators for native fragments, especially when found together. We compared these results to the manually delineated fragments and observed that the majority of manually-delineated native fragment polygons were composed chiefly of these two land cover classes.

To label the pixel classes resulting from our unsupervised classification, several inputs were used for reference. Our manually delineated fragment data was the chief reference layer, as it assisted in determining what pixel values were more prevalent in locations we had previously determined were native forest fragments, second-growth forest, and shrubland. We also used our CHM layer to determine the relative height of features represented by different pixel values, and our ground truth data to approximate species composition within the classes. We found that native fragments were chiefly composed of mature oak, hickory, and maple, while some of the more common species in our manually-delineated second-growth forest polygons were black walnut, buckthorn, and black cherry.

Some of the pixel classes resulting from our unsupervised classification received labels involving forest canopy structure. The LiDAR data utilized during our classification played a major role in

making this possible, as spectral signature alone was not a valid option for determining the elevation of features and feature elevation is an important determinant for classification of forest canopy structure.

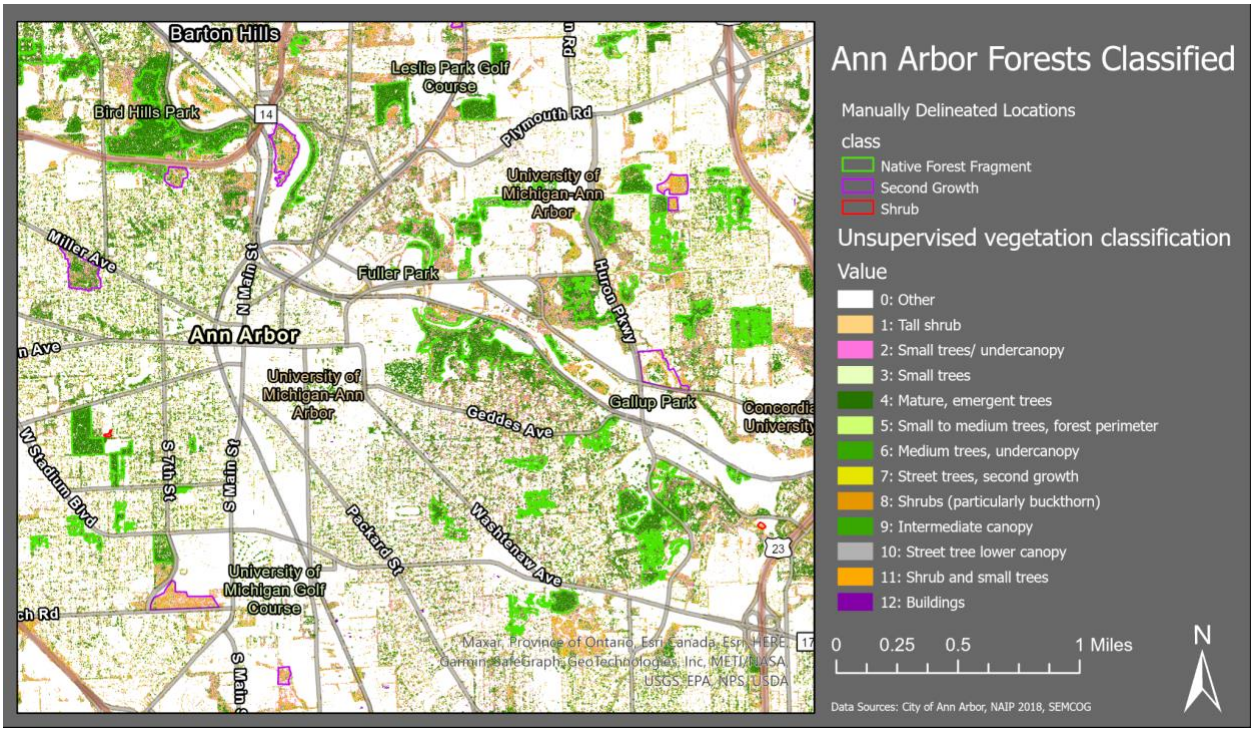



**Figure 2.2.** Map of Ann Arbor Forests Unsupervised Classification results, compared to our manually delineated fragments. Parts of the city have a higher density of native fragments, frequently interspersed with larger, delineated fragments. For a larger version of this map, see Appendix B.

## Discussion

The two major processes we ran to identify native fragments served to identify different aspects of Ann Arbor's native forest land. The manual delineation served to identify and confirm the largest native fragment locations, most of which our clients were already aware of. The unsupervised classification served to find much smaller native fragments, which is especially useful in areas that have seen a higher degree of development and fragmentation. Both methods had their benefits and faults- the manual delineation process was verified by Christopher and Jason, our CHM layer, and often by information posted by Ann Arbor's Parks and Recreation Department, but was highly time-consuming. It was also an important source when consolidating and labeling the land cover classes created by our unsupervised assessment. For example, class five pixels were often found near the perimeter of delineated native fragments and mixed within secondary forest land. This allowed us to tentatively determine that class five pixels represented small to medium trees, as well as forest perimeter. The unsupervised classification served to identify much smaller native fragments that were more frequently located on private property but was not as readily verifiable. Utilizing these two outputs together serves to mitigate much of these faults.

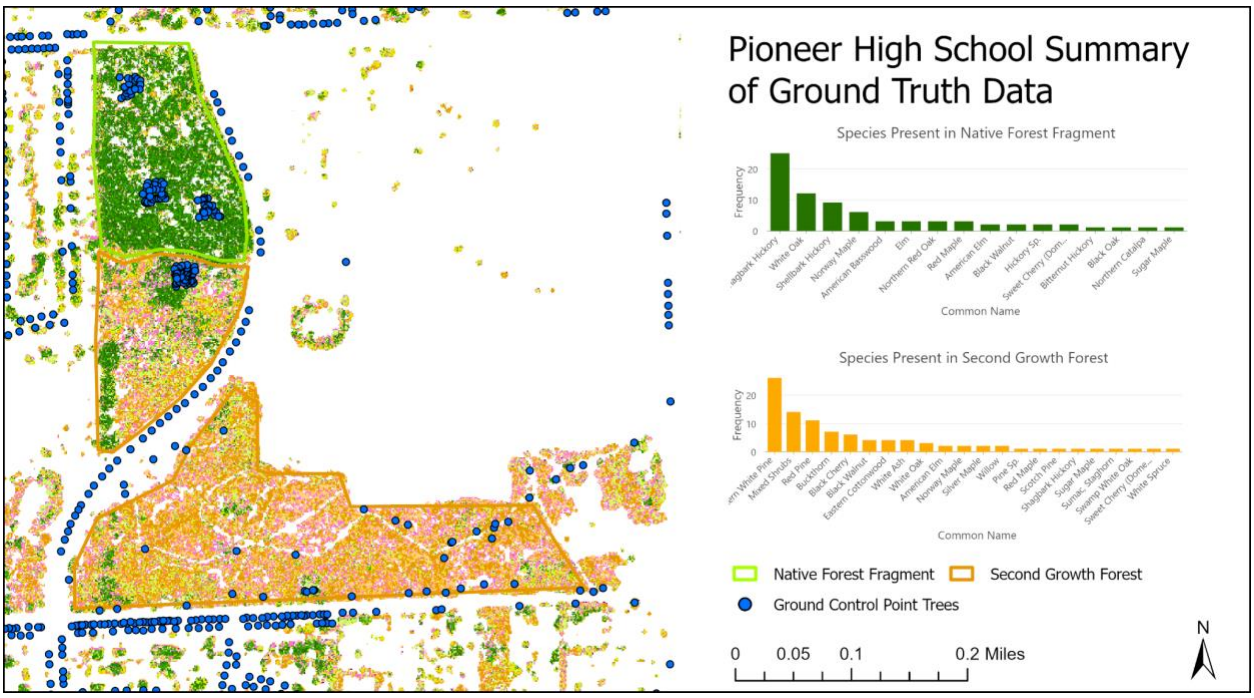



**Figure 2.3.** The forest land around Pioneer High School is composed of a mix of native and second-growth forest. The most common species found during ground truth data collection in this area were Shagbark Hickory, White Oak, and Shellbark Hickory within the native forest fragment, and Eastern White Pine, mixed shrubs, and Red Pine within the second-growth forest areas. Additional summaries of ground truth trees by cluster can be found in Appendix I.

Overall, our efforts have shown that the LiDAR and aerial imagery available to the city of Ann Arbor is able to capture subtle differences in canopy composition and structure. Future classification processes could benefit from the inclusion of a biologist to aid in interpreting and verifying those differences. Further refinement of urban canopy classes could be attained by collecting additional field data within manually delineated native and second-growth fragments, to aid in predicting species composition depending on forest type.

# Chapter 3: Turfgrass Estimation

## Introduction

We aimed to identify areas of turfgrass within the City of Ann Arbor as a way to inform residents and decision-makers about its extent on both public and private property. The hope is that citizens and the City can use this to further investigate and evaluate if there are areas that would receive greater benefits from vegetation that provides more diverse ecosystem services. This could include vegetation that is more beneficial to pollinators or trees that could expand the urban canopy of Ann Arbor.

## Background

Turfgrass is common in many urban and suburban communities, typically present in residential yards, parks, and other green spaces. It provides important space for outdoor recreation and contributes to aesthetics on private and public property alike. In addition, turf provides ecosystem services, and is in fact better suited than most vegetation for filtering pollution, protecting land from erosion, and decreasing runoff (Monteiro, 2017). Despite these benefits, urban forests in conjunction with turf provide stronger air cooling than turf alone, which would reduce urban heat island effects (Fung & Jim, 2019). There are also environmental risks associated with turfgrass connected to care and management. Lawn maintenance can result in increased water usage, emissions from motorized tools, and overuse of fertilizers (Runfola et al., 2013). Given the services and disservices for turfgrass and lawns, it is important for cities and communities to be aware of its extent and assess its role while planning for the future.

## Data

This analysis relied on multispectral imagery from the National Agriculture Imagery Program (NAIP) captured in July 2018 as well as light detection and ranging (LiDAR) with complete coverage of the city collected in 2017. We created a canopy height model (CHM) with a pixel resolution of 3 square feet and resampled the NAIP imagery so the resolutions would match, and the grids aligned.

## Methods

Using the red and near-infrared bands of the imagery, we calculated the normalized difference vegetation index (NDVI) and identified a lower bound for vegetation in this image where, generally, values lower than it would represent non-photosynthesizing pixels, and higher would capture pixels representing photosynthesizing vegetation. Then, three different binary layers were created from the CHM to identify areas above and below 1, 2, and 3 feet. There was not much difference between them, so we proceeded using only the 3-foot threshold on the CHM. Multiplying the binary NDVI and CHM layers resulted in another binary layer that distinguished between vegetation below 3-feet and all other pixels, seen in Figure 3.1. This was used as a mask to extract NAIP pixels that ideally depicted vegetation below 3 feet, which largely included areas of turfgrass.

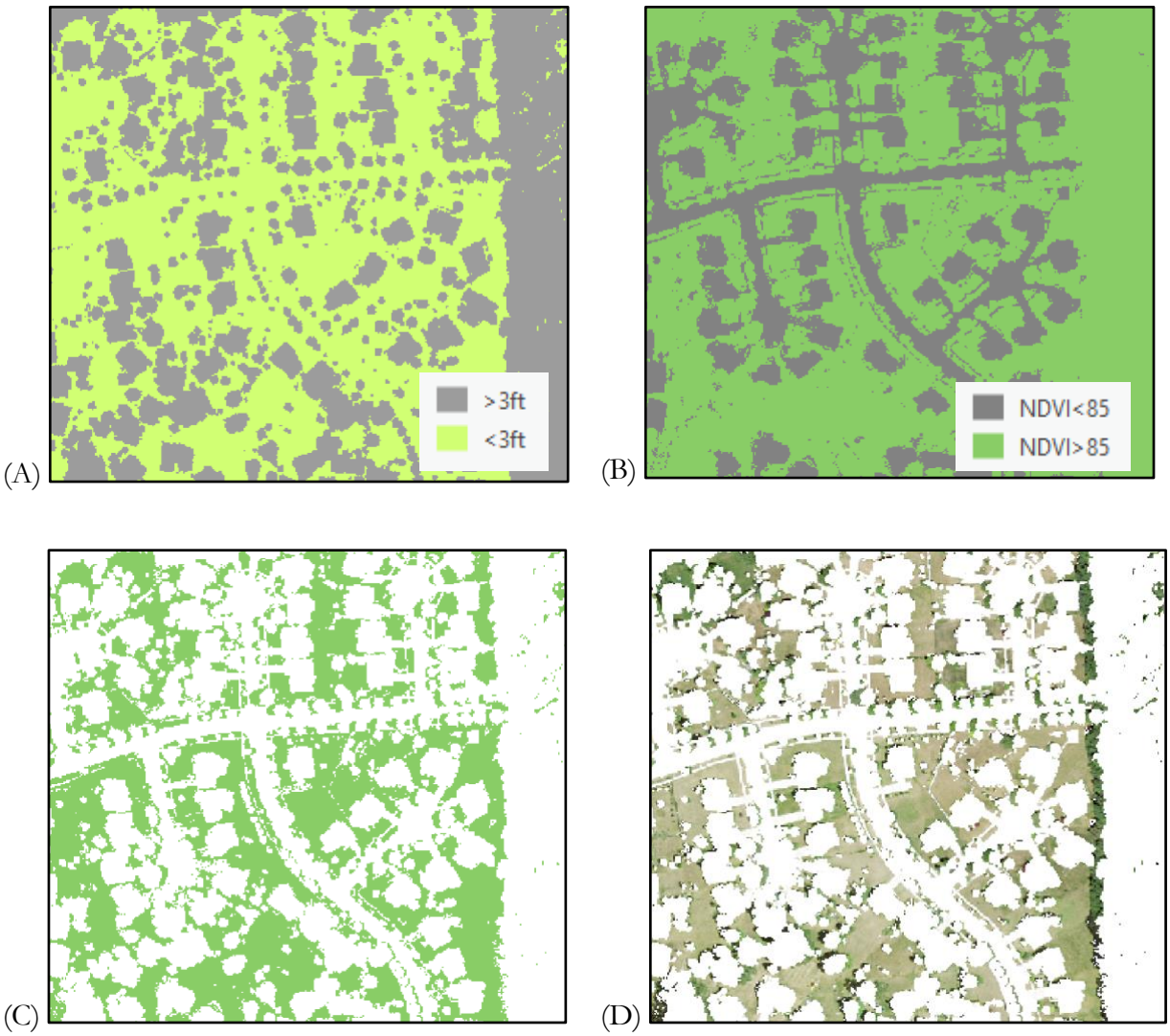



**Figure 3.1.** (A) Binary height layer, with a 3ft threshold; (B) Binary NDVI layer; (C) Product of the binary NDVI and height layers, depicting pixels that represent vegetation below 3ft in green; (D) Extracted NAIP imagery using (C) as the mask feature.

After visual inspection and image interpretation, we deemed that there were still areas included within this extracted imagery that did not represent turfgrass. These non-turf areas included bunkers on golf courses, bare earth, and edges of canopy. These slivers of canopy were likely included due to off-nadir LiDAR points and/or temporal differences between the 2017 LiDAR and 2018 NAIP imagery failing to capture a year's worth of canopy growth. An unsupervised classification with 6 classes managed to capture that variation in the extracted imagery, and we interpreted the six classes as the following: (1) shadows, (2) canopy, (3) very dry grass, (4) moderately dry grass, (5) healthy grass, and (6) dirt/sand. The three classes representing grass were merged resulting in a final classification with four classes. The merged grass class was extracted to create the final turfgrass layer within the extent of the city.

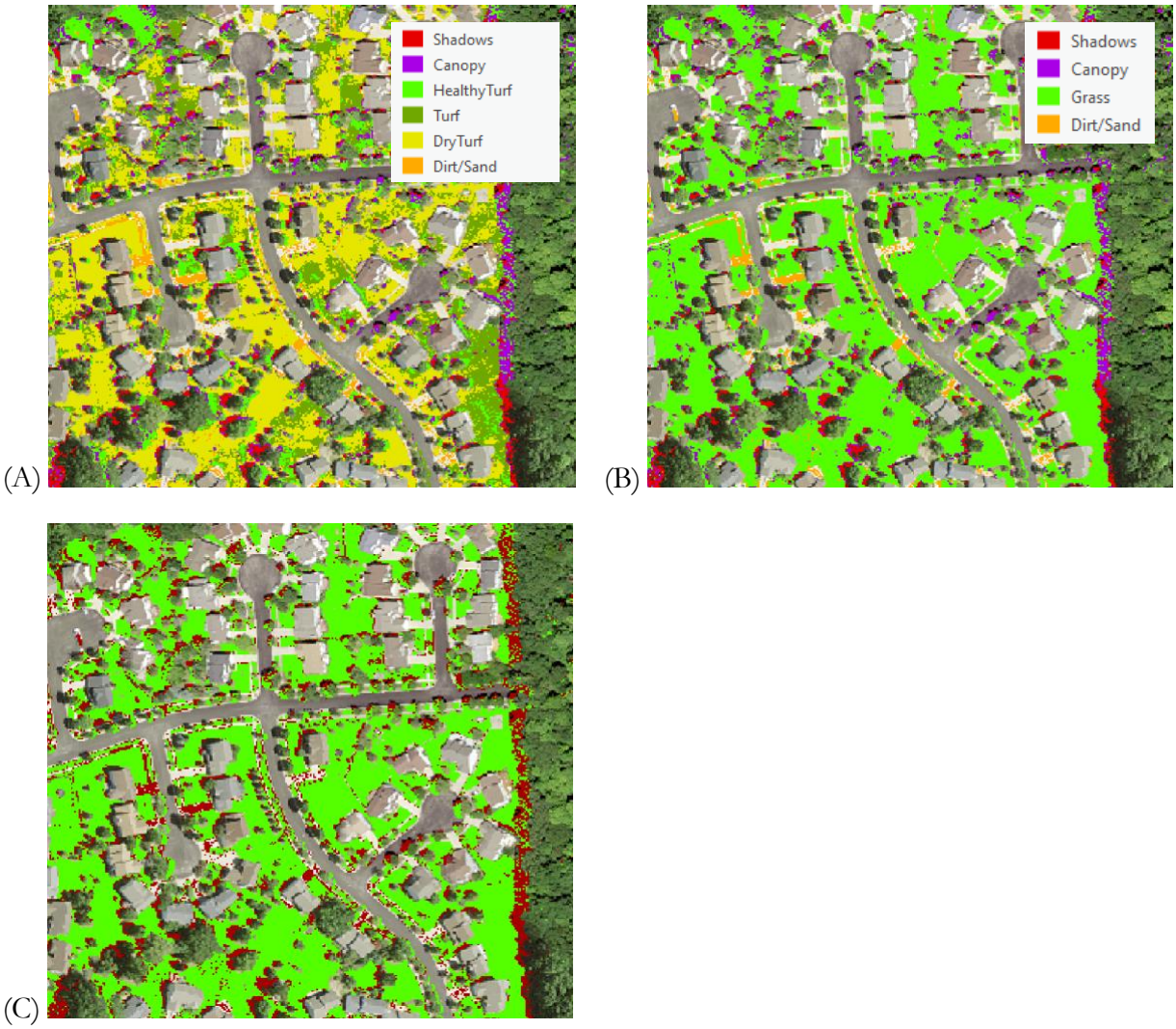



**Figure 3.2.** (A) Subset of the city showing the 6-classes of the first classification; (B) Resulting classification after merging the three grass classes, which is our final turfgrass layer; (C) The final turfgrass layer shown in green, which is an improvement from the initial layer shown in red.

## Results

To assess the accuracy of the unsupervised classification, we took a stratified random sample across the four classes and classified each point based on interpretations of the NAIP imagery. The resulting confusion matrix is shown in Figure 3.3 below. The turfgrass class has the second highest user's accuracy at 79 percent however it also has the lowest producer's accuracy of only 49 percent. The classification confuses turfgrass for many of the other classes, but mostly canopy.

| | | Reference Data | | | | User's Accuracy |
|---|---|---|---|---|---|---|
| | | Shadows | Canopy | Turfgrass | Dirt/Sand | |
| Classification | Shadows | 76 | 18 | 6 | 0 | 76% |
| | Canopy | 6 | 28 | 61 | 5 | 28% |
| | Turfgrass | 0 | 6 | 79 | 15 | 79% |
| | Dirt/Sand | 0 | 0 | 14 | 86 | 86% |
| Producer's Accuracy | | 92.68% | 53.85% | 49.38% | 81.13% | |
| Overall Accuracy = 67.25%<br>Kappa = 43.67% | | | | | | |

**Figure 3.3.** Confusion matrix that shows how well our classification performed on a set of reference data - a stratified random sample across all 4 classes, as well as the overall accuracy of the classification (67.25%) and kappa statistic (43.67%).

## Discussion

Ultimately, we found that the turfgrass class extracted from the classification was an improvement compared to the first attempt using only the NDVI and height thresholds product. The classification did help to filter out non-turf features such as some dirt paths, sand, driveways, and most canopy. However, it still includes areas of water vegetation, wetlands, and some old fields which could be further classified and filtered out.

Though the turfgrass extracted from the final classification has high user's accuracy, the overall performance of the classification is not very accurate or precise. Finer resolutions of both the imagery and CHM could improve our turfgrass identification. Initially, the NAIP imagery had a finer resolution of 0.66 meters but was resampled to 3-feet pixels in order to align with the CHM grid. Additionally, the reference data was collected based on interpretations of the NAIP imagery. The improved resolutions could lead to higher precision in the reference data if interpreted from imagery, as well as more accurate classifications performed to filter out features that are not turfgrass.

Accuracy of reference data could be increased by gathering ground truth data to include or use exclusively to test the accuracy of the classification. This could, however, lead to greater confusion if the ground truth data is not collected around the same time as the data used since there could be changes in the land between the time that imagery or LiDAR was collected and ground truth collection. Better temporal alignment between imagery and LiDAR could also improve our turfgrass identification. In our case, there is only one year between when the LiDAR was collected and the

NAIP imagery. However, either dataset–whichever was taken later–could fail to capture events that result in a change of land cover like tree removal, canopy growth, or development. Additionally, performing this analysis over many years could be more helpful to managers so they have an idea of the rate that turfgrass is expanding in the city.

# Chapter 4: Genus Classification

## Introduction and Background

Classifying an urban tree canopy to the genus or species level from remotely sensed data would offer a variety of benefits to decision-makers tasked with preserving and enhancing the canopy. Such a classification would, for instance, allow one to easily characterize types of forest communities at a large scale. In the case of Ann Arbor, this would aid in definitively separating undisturbed forest fragments from "pioneer" woodlands. It could also aid in the control of invasive shrubs such as honeysuckle and buckthorn. Furthermore, in combination with allometric equations published by the U.S. Forestry Service (Chojnacky et al., 2014), an accurately classified canopy can aid in the estimation of above-ground biomass and the quantification of associated ecosystem services.

In general, the process of classifying trees from remotely sensed data entails three steps, each with a variety of possible approaches. First, training and testing data must be collected, typically in the form of geolocated and manually identified tree stems or crowns. Second, remotely sensed data must be assembled for use as predictor variables and a value corresponding to each layer must be calculated for each tree. Finally, the resulting dataset pairing tree genus or species labels with predictor variables must be fed into a classification algorithm.

In terms of input predictors, several studies have utilized both LiDAR and aerial imagery to identify tree crowns (Weinstein et al., 2019; Katz et al., 2020). While others utilize just LiDAR data to detect and segment individual trees based on crown shapes, height, and/or diameter at breast height (DBH) (Lu et al., 2014; Matasci et al., 2018; Oono & Tsuyuki, 2018; Hastings et al., 2020). Furthermore, Onishi & Ise (2021) used aerial imagery and machine vision systems to identify tree crowns and species from color, three-dimensional information, and a slope model. Breidenbach et al. (2009) highlight the use of ground-truth data and airborne laser scanning data to identify individual tree crowns.

Machine learning algorithms exhibit a similar variety of approaches, Katz et. al. (2020) used random forest, a decision tree algorithm available as a free R package, to classify the urban canopy in Detroit. Other researchers have used methods like support vector regression and k-nearest neighbor clustering (Zhang & Qiu, 2012). Finally, some researchers have successfully used neural networks for tree identification (Onishi et al. 2021, Weinstein et al., 2019).

## Methods

We used a combination of field-collected and available data of the urban tree canopy to leverage as training/testing data in our machine learning algorithms. Both object-based and pixel-based classification methods were implemented.

### *Tree Data Description*

In order to ensure a robust sample of trees growing in native forest fragments, our team field-collected training data from natural areas. We established plots within Ann Arbor focusing on City-owned or University of Michigan-owned land to sample and survey the UTC. With guidance from our advisor and clients, we chose areas that gave us as close to a representative sample of the city's different tree and forest types as possible. Figure 4.1 depicts the sixteen locations we surveyed.

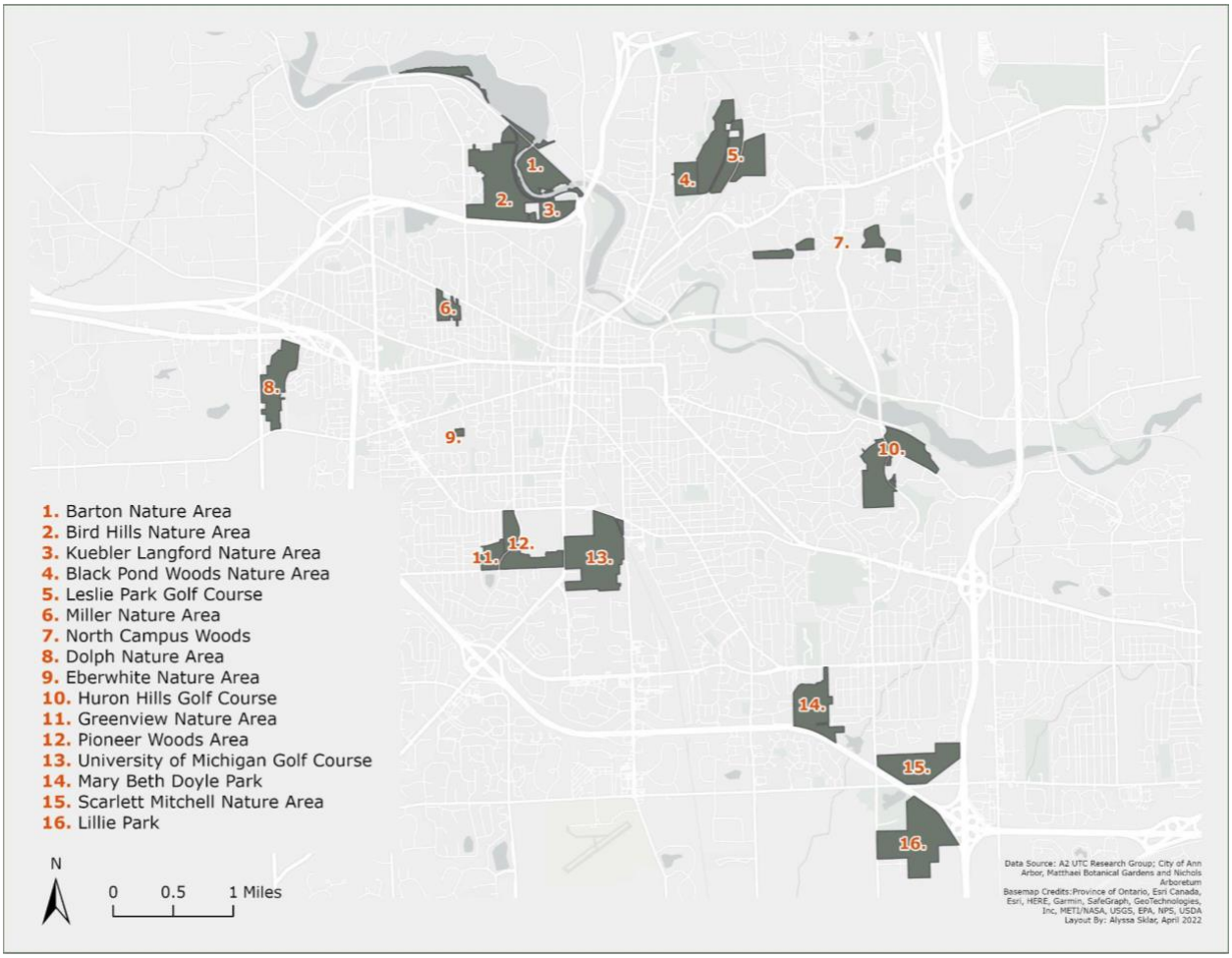



**Figure 4.1.** Natural areas and golf courses surveyed.

Between all potential survey locations, we created 1,000 random points using ArcGIS *Create Random Points* tool to randomly select possible plot locations. In the field, a random point was selected and assessed for accessibility in regards to terrain and land features. Once at a feasible location, we used a Trimble R1 to locate and monument the southwest corner of the plot. Then a measuring tape was laid out to the east and north from the initial plot corner generating a 100 x 100 feet plot. Living canopy trees were recorded using ArcGIS *Collector*. We identified the tree to genus and species, measured the diameter at breast height (DBH), and recorded an X, Y value along with the measuring tape of the tree location (Figure 4.2). Tree identification was carried out using Barnes & Wagner (2004), iNaturalist app, PictureThis app, and direct assistance from our advisor and clients.

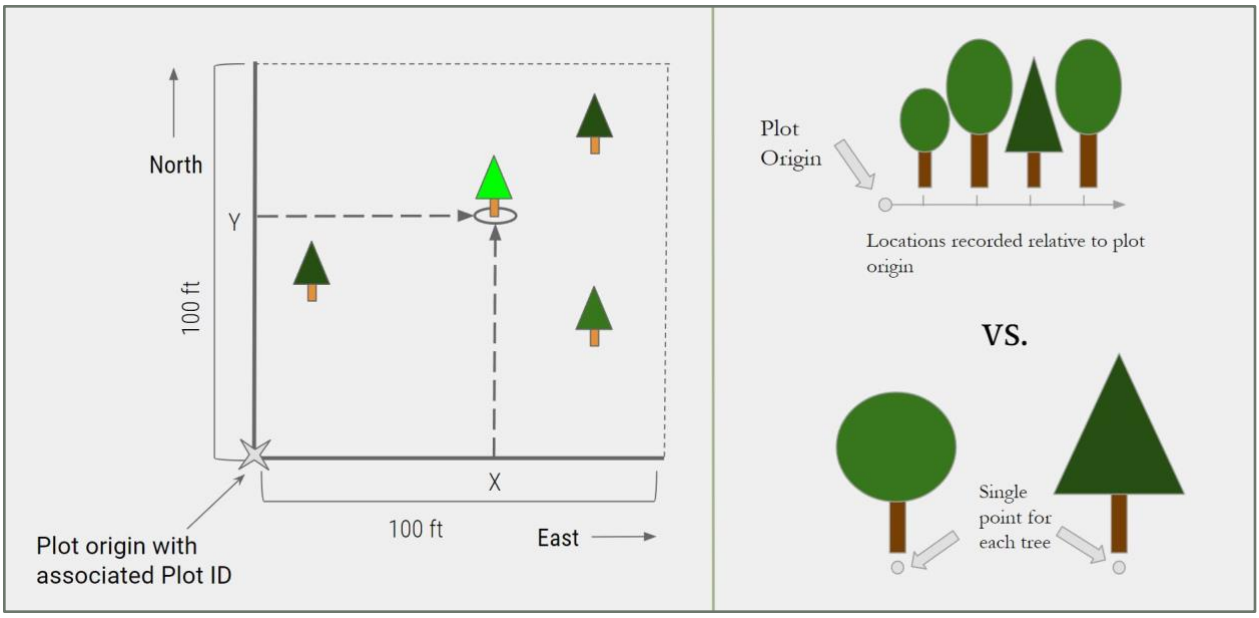



**Figure 4.2.** Comparing plot collection methods to single-point data collection.

At non-forested locations such as parks and golf courses, or areas where dense undergrowth made laying out a measuring tape difficult, we used the Trimble R1 unit to record single points for each tree stem in ArcGIS *FieldMaps*. As with plot collection, we recorded genus, species, and DBH for each tree.

Figure 4.3 compares the output of field collection methods at three locations: Bird Hills Nature Area, Huron Hills Golf Course, and Lillie Park. At Bird Hills, we used the 100 x 100 feet plot collection method. At Huron Hills Golf Course, we only used the single-point method since trees were typically open-grown. Lastly, at Lillie Park, we used both the plot and single-point collection method to allow us to capture the variety of tree composition.

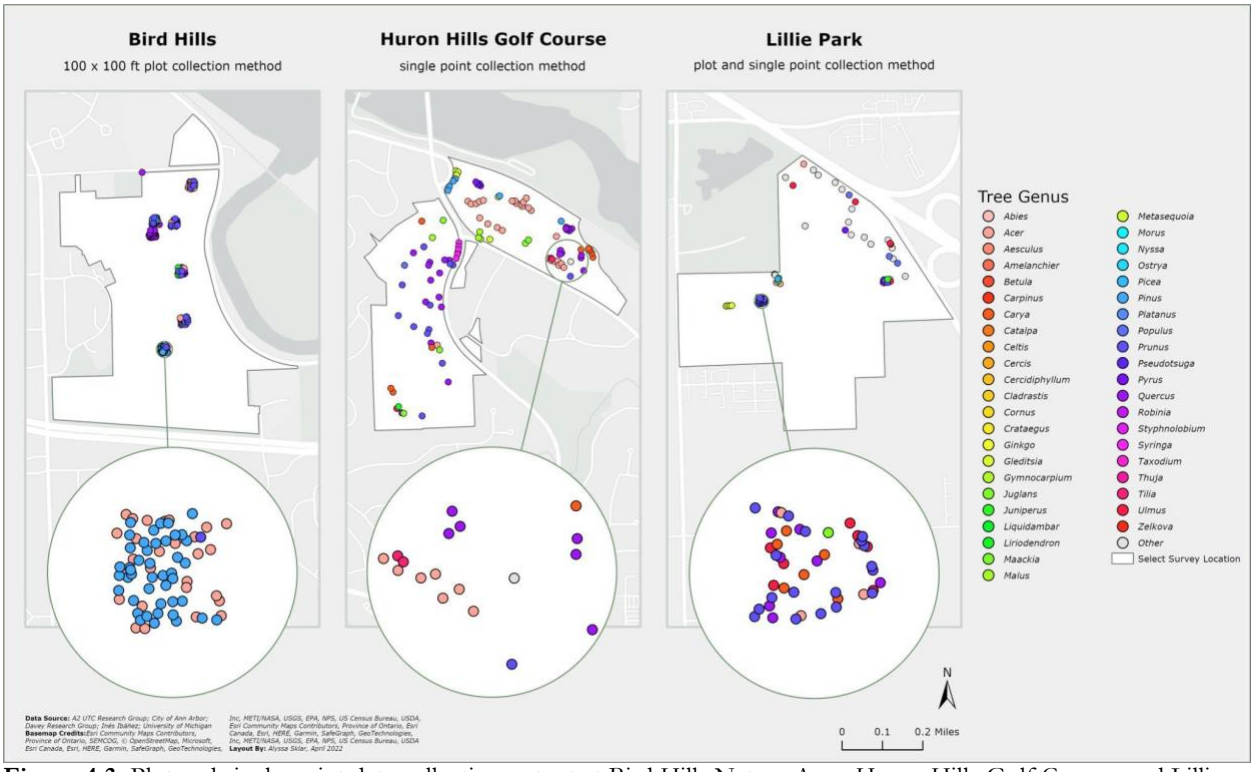



**Figure 4.3.** Plot and single-point data collection output at Bird Hills Nature Area, Huron Hills Golf Course, and Lillie Park.

Data was collected May 05, 2021 - August 17, 2021 at thirteen natural areas and three golf courses. 1,542 trees and/or 2 x 2-meter shrubs were surveyed. Table 4.1 depicts the thirty-one genera recorded and the respective DBH.

| Genus | Count of Genus | Mean DBH (cm) | Minimum DBH (cm) | Maximum DBH (cm) |
|---|---|---|---|---|
| *Acer* | 220 | 32.07 | 8 | 200 |
| *Aesculus* | 2 | 65.5 | 58 | 73 |
| *Carya* | 172 | 37.25 | 11.5 | 91 |
| *Castanea* | 3 | 14.5 | 11.5 | 18.5 |
| *Catalpa* | 1 | 19.5 | 19.5 | 19.5 |
| *Celtis* | 7 | 21.57 | 9.5 | 34 |
| *Cornus* | 1 | 10 | 10 | 10 |
| *Crataegus* | 15 | 23.53 | 12 | 34.5 |
| *Elaeagnus* | 3 | N/A | N/A | N/A |

| | | | | |
|---|---|---|---|---|
| *Fraxinus* | 13 | 20.08 | 6 | 44.5 |
| *Gleditsia* | 16 | 55.56 | 23 | 131 |
| *Juglans* | 63 | 41.46 | 11 | 124 |
| *Juniperus* | 1 | 16 | 16 | 16 |
| *Liriodendron* | 13 | 34.12 | 16.5 | 68 |
| *Malus* | 8 | 34 | 18 | 104 |
| Mixed-Shrub | 23 | N/A | N/A | N/A |
| *Morus* | 3 | 50.67 | 23 | 74 |
| *Picea* | 27 | 52.57 | 17.5 | 94.5 |
| *Pinus* | 170 | 31.67 | 15 | 64 |
| *Platanus* | 2 | 20.5 | 17 | 24 |
| *Populus* | 53 | 44.37 | 5 | 200 |
| *Prunus* | 145 | 36.45 | 9.5 | 132 |
| *Pyrus* | 7 | 42.86 | 25 | 72 |
| *Quercus* | 344 | 49.49 | 9 | 137 |
| *Rhamnus* | 27 | N/A | N/A | N/A |
| *Rhus* | 6 | 151.33 | 38 | 200 |
| *Salix* | 4 | 60.25 | 32 | 118 |
| *Sassafras* | 4 | 31.63 | 20.5 | 39.5 |
| *Taxodium* | 10 | 105.7 | 74 | 139 |
| *Tilia* | 116 | 27.66 | 10 | 114 |
| *Ulmus* | 57 | 26.98 | 7 | 131 |
| *Zelkova* | 1 | 32 | 32 | 32 |
| Unknown Living Tree | 5 | 46.6 | 36 | 53.5 |

**Table 4.1.** Summary of genera collected in the field with respective DBH values. Note: values of N/A represent shrubs with many stems for which measuring DBH would be prohibitively difficult.

After all plots were collected, an R script was used to adjust for magnetic declination and transform the X and Y plot locations into Easting and Northing coordinates in the Michigan State Plane South system (U.S. Feet) (EPSG 2253).

Additional sources of tree data came from the City of Ann Arbor's Street Trees, University of Michigan's Campus Trees, Nichols Arboretum, and a study at Radrick Forest by Dr. Inés Ibáñez. All sources were cleaned as described below (Table 4.2). To ensure fully consistent coding by species, a table was manually compiled to act as a join key for relating tree common names with species codes. The datasets were merged into a single master dataset with columns for common name, DBH, species code, genus code (USDA), and source dataset.

| Name | Location | Date of Creation | Cleaning Steps |
|---|---|---|---|
| City of Ann Arbor Street Trees | Ann Arbor, Michigan | February 10, 2021 | Elimination of trees with no recorded DBH, conversion of DBH range into single median value, elimination of trees listed as "Stump" or "Vacant" |
| University of Michigan Campus Trees | Ann Arbor, Michigan | 2014 | Elimination of trees with unknown species, elimination of trees marked as removed |
| Radrick Forest Stem Map | Radrick Farms Golf Course, Ann Arbor, Michigan | June 1, 2016 to July 31, 2018 | None |
| Nichols Arboretum | Nichols Arboretum, Ann Arbor, Michigan | 2018 | Elimination of trees listed as "Dead" or "Remove", elimination of all trees not specified as canopy trees |
| Survey | Ann Arbor, Michigan | May 05, 2021 to August 17, 2021 | Manual typo correction |

**Table 4.2.** Tree data sources and cleaning steps.

Overall, our training data consisted of 69,000+ tree points representing 97 genera. Figure 4.4 displays tree points with a count greater than or equal to 200 per genus. Genera with less than 200 were lumped into an "Other" category.

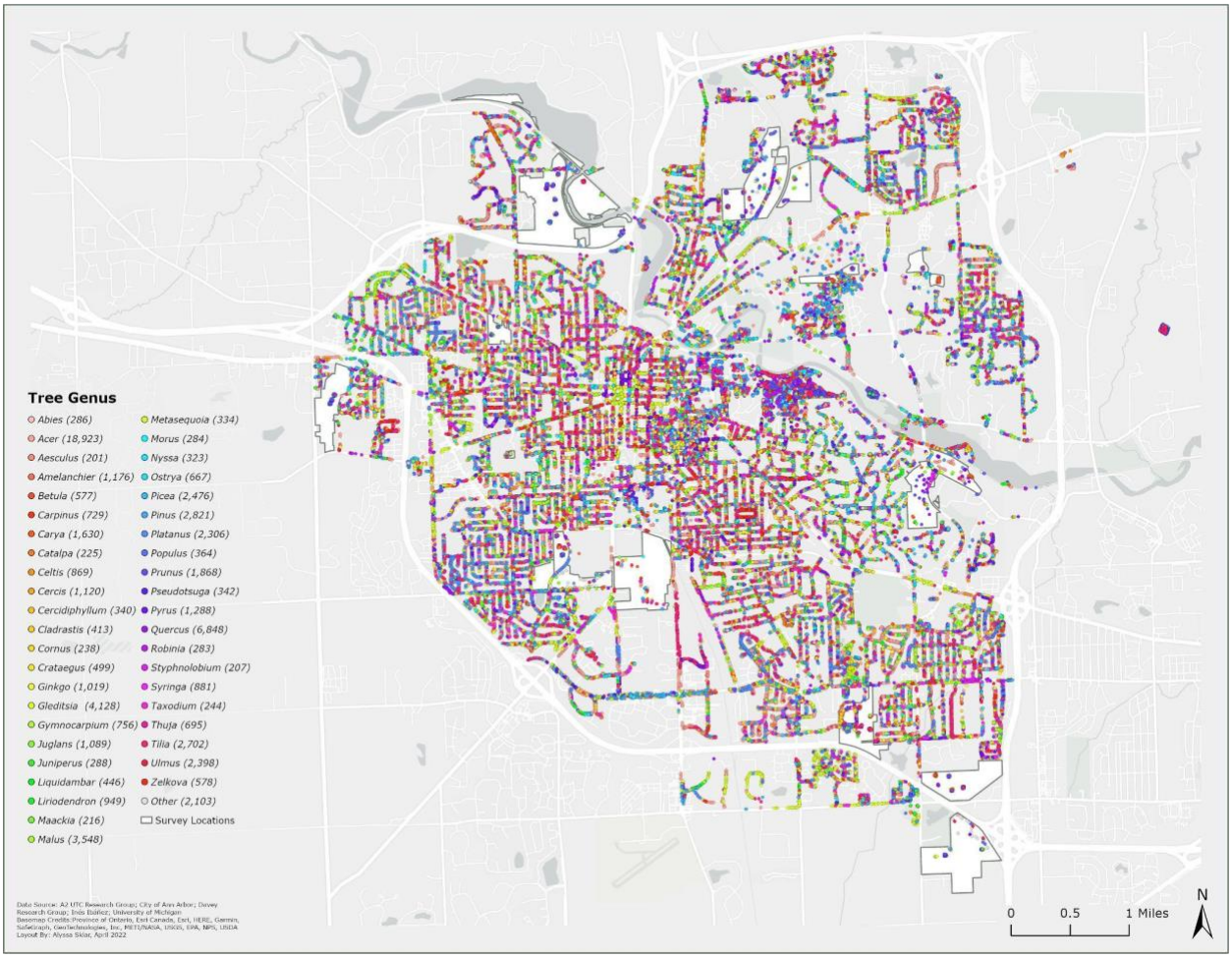



**Figure 4.4.** Genera with more than 200 representative samples in relation to our survey locations.

## *Predictor Rasters*

| Name | Year | Format or Resolution |
|---|---|---|
| Color-Infrared (CIR) | 2020 | 0.5 feet |
| Color-Infrared (CIR) | 2018 | 0.5 feet |
| National Agriculture Imagery Program (NAIP) | 2018 | 2 feet |
| NearMap | July 17, 2018 | 3.918 feet |
| Light Detection and Ranging (LiDAR) | 2018 | 1,101,733,871 points, 1.246-1.9 feet point-spacing |
| Soils | September 16, 2019 | N/A (vector shapefile) |

**Table 4.3.** Predictor Raster datasets.

We gathered six raster datasets (Table 4.3) to use as predictor variables in the genus and native fragment classification processes.

From these initial layers, we created an NDVI layer, texture layers from the near-infrared (NIR) and Green bands of the NAIP imagery, a canopy height model and canopy texture layer, and six principal component layers from the LiDAR data.

For a full summary of how these layers were prepared, see Appendices A and E.

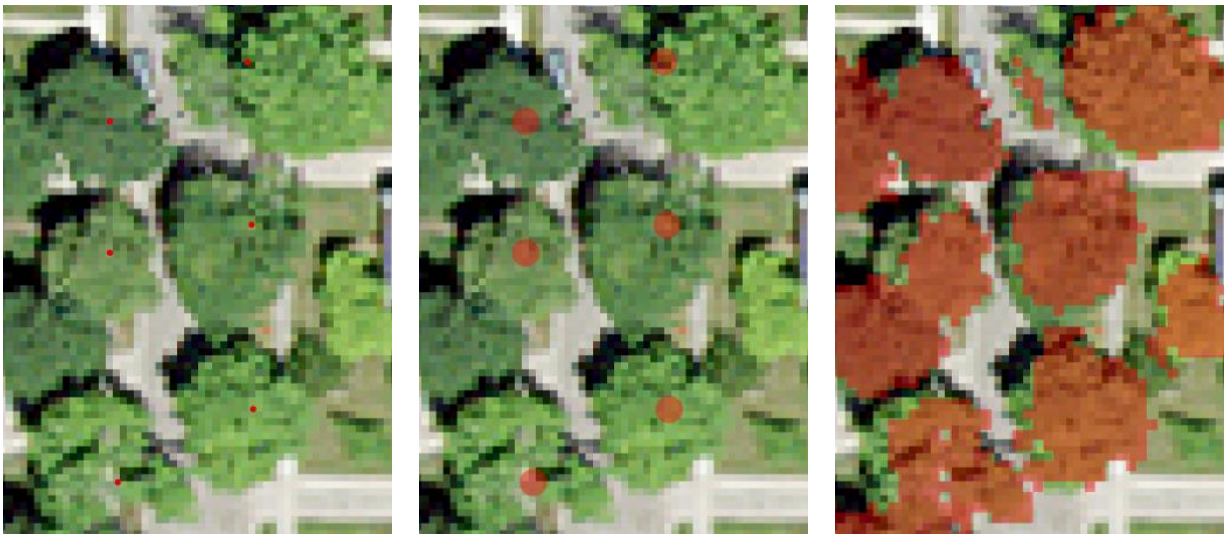
### *Segmentation*

Relating the representation of trees in the training/testing dataset (single points in a vector dataset) with their representation in the predictor layers (amorphous collections of pixels associated with each tree crown) posed a challenging problem.

We explored three ways of associating training data with predictor variables. First, we used single-pixel extraction as an easily implemented baseline. For this method, we used the *Extract Multi-values to Points* tool in ArcGIS Pro to turn the individual pixel values from each predictor layer overlaying each tree stem point into attributes of the tree point feature class. This approach was simple and easy to implement but leveraged relatively little of each dataset.

Second, we implemented a buffered pixel-based classification approach to group pixels based on a tree's diameter at breast height (DBH). A buffer for each tree was created in ArcGIS Pro and pixel values within the buffer were summarized using *Zonal Statistics*. Pixel values were summarized from both multi-temporal aerial imagery and satellite imagery using the individual bands of each image source.

Third, we used the *locate_trees* and *segment_trees* functions in the *lidR* package for R to delineate entire tree crowns. First, to locate trees we used a subset of Ann Arbor's LiDAR to test three moving window parameters for the *locate_trees* function and three variations on the input canopy height model. For the moving window, we tested a fixed 20-foot window, a fixed 30-foot window, and a variable-sized window following the example in Roussel et al. (2021). We ran these on the raw canopy height model and canopy height models smoothed by a 3 x 3 pixel and 5 x 5 pixel median filter. The combination of three moving window parameters and three input CHMs resulted in nine different treetop identifications.



(A) (B) (C)

**Figure 4.5.** Comparison of three different segmentation methods: single-pixel (A), buffered pixel (B), and via the Silva (2016) algorithm as implemented in the *lid*R package for R (C).

Each treetop map was used as a basis for segmentation with the *segment_trees* function. Early testing indicated that of the segmentation algorithms provided in the package, Silva et al. (2016) was by far the most effective at delineating trees in Ann Arbor's mostly deciduous canopy. Comparison between segmentations was evaluated with visual inspection against NAIP and Nearmap imagery, comparison to manually delineated tree crowns, by comparing the number of oversized segments (tree crowns with multiple recorded tree stems within them) to the total number of crowns found in the testing area, and by testing their performance directly in initial machine learning models run on the testing area.

We found that segmentation based on a 30-foot window over a CHM smoothed by a 3 x 3 pixel kernel resulted in the highest performance on our test models and also performed well relative to our other assessments. This segmentation was scaled up to the entire city using the LasCatalog functionality in the *lid*R package. Predictor variables were associated with it by using the *Zonal Statistics as Table* tool in ArcGIS Pro to record the mean value of each input band as well as the minimum, maximum, standard deviation, 70th percentile, 80th percentile, and 90th percentile statistics for the canopy height model. For more details on the calculated variables and some exploratory data analysis, see Appendix F.

Genus labels were assigned to delineated crowns using a modified spatial join operation wherein a crown overlapping multiple stems was assigned the genus of the stem with the highest DBH.

### *Machine Learning*

Each segmentation method (pixel, buffered pixel, and lidR) resulted in a table of labeled tree stems/crowns with associated predictor variables taken from each input band. These tables were fed into a standardized R script that took a random stratified 50% sample from each genus to serve as training data (with the remainder set aside as testing data). For consistency, a random seed of 1 was

used for all sampling. We implemented a "genus threshold" parameter that grouped genera with fewer instances than the threshold into an "Other" category.

Training data were then used to train a Random Forest model, a Support Vector Machine, and a Multinomial Regression model. Model accuracy was assessed by using the *caret* package (Kuhn, 2022) to record confusion matrices, overall error, and kappa statistics.

The random forest model was run using the *randomForest* command from the *randomForest* package (Liaw & Weiner, 2022). Default parameters were used, and *xtest* and *ytest* were provided in order to generate immediate predictions on the testing dataset.

The support vector machine model was run using the *svm* command in the *e1071* package (Meyer et al., 2021). All default settings were preserved.

The multinomial model was run using the *multinom* command from the *nnet* package (Ripley, 2022). The limit on the number of weights (the *maxnwts* parameter) was removed, and the model was allowed to attempt to converge for 2000 iterations (the *maxit* parameter). All other settings were kept at default.

## Results

LidR segmentation achieved a maximum accuracy of 54.3%. Buffered pixel segmentation achieved a maximum accuracy of 38.3%. Single-pixel achieved a maximum of 47%. In each case, these maximal accuracies were achieved with the random forest classifier and the use of all available data layers. Additional model runs and accuracy metrics for the SVM and multinomial models are recorded in Appendix D.

Our highest-performing model was a random forest model run on lidR segmented trees and using all available predictors. (These were the canopy height model, NDVI, all four bands of 2018 NAIP, all four bands of 2018 CIR, all three bands of 2018 Nearmap, texture rasters of NAIP Green and NIR bands, CHM texture, soils, and LiDAR PCs 1-6. For more information on the dataset, see Appendix F). This run achieved an overall accuracy of 54.3% with a Kappa statistic of 0.37. The model used a genus threshold of 100 resulting in 29 genera: *Acer, Amelanchier, Betula, Carya, Catalpa, Celtis, Cercis, Cratagea, Gleditsia, Gymnocarpium, Juglans, Liriodendron, Malus, Morus, Picea, Pinus, Platanus, Populus, Prunus, Pseudotsuga, Pyrus, Quercus, Robinia, Sophora, Thuja, Tilia, Ulmus,* and *Zelkova*, with remaining trees grouped into an "Other" category. The full confusion matrix for this run can be viewed in Appendix C.

With the use of the *randomForestExplainer* package (Paluszynska et al., 2020) for R, metrics of relative importance were calculated for each variable in the highest-performing model. The mean of the texture raster calculated from the NAIP NIR band leads for both the mean decrease in accuracy and mean decrease in GINI metrics. The green NAIP texture raster, the mean of the NAIP NIR band, and the first LiDAR principal component also rank highly for both.

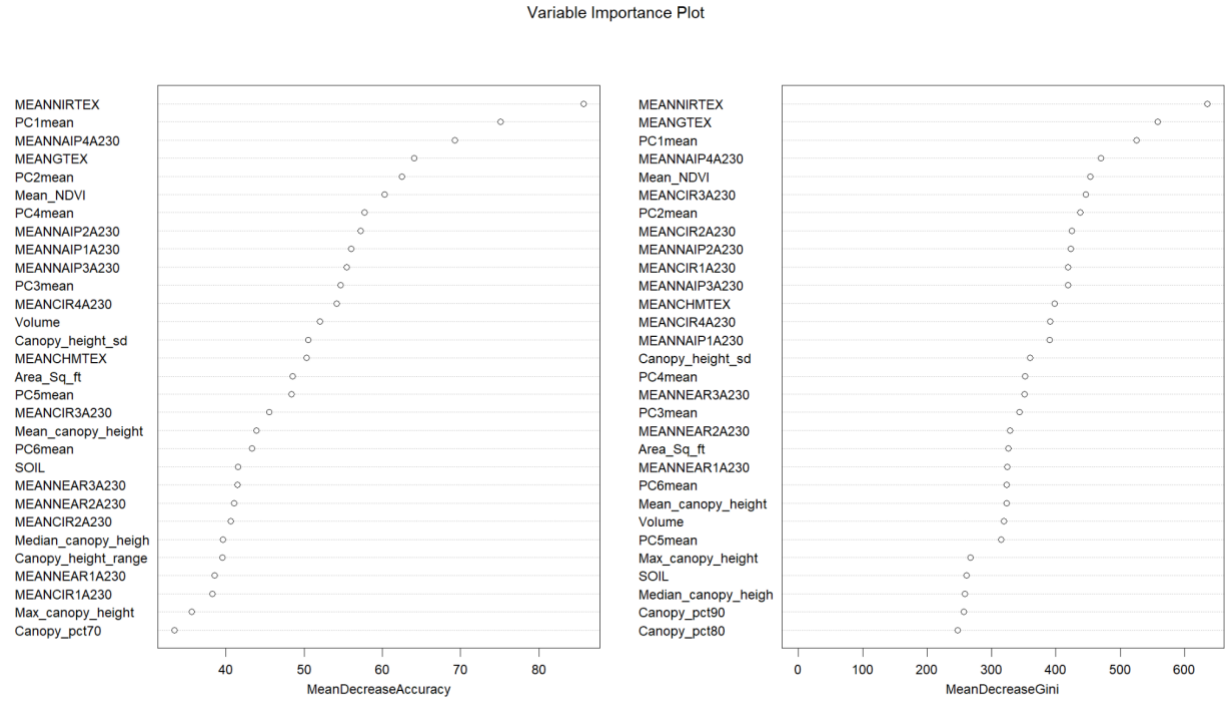



**Figure 4.6.** Predictor variables in the highest-performing random forest model ranked according to mean decrease in accuracy and mean decrease in GINI coefficient.

The *randomForestExplainer* package also computes variable rankings based on combined metrics (Figure 4.7). We can see that green and NIR texture, the first two LiDAR PCs, the green and NIR NAIP bands, NDVI, and the visible bands from the CIR imagery all are top contributors to the random forest model's ability to accurately classify genera.

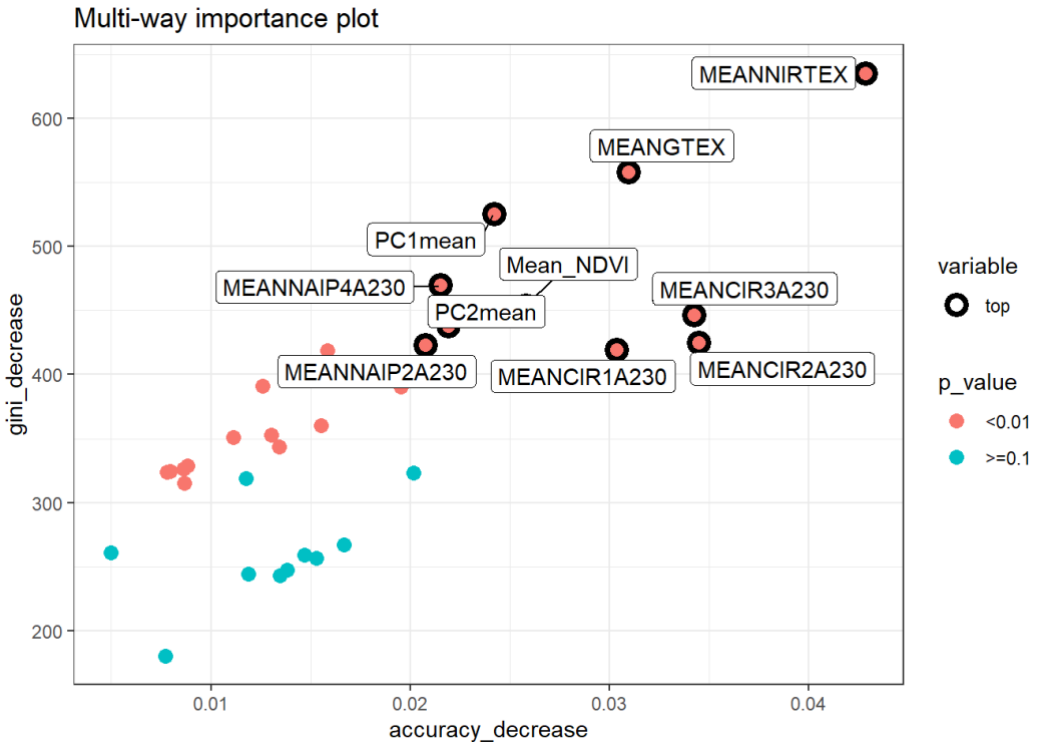



**Figure 4.7.** Multi-way importance plot generated by *randomForestExplainer* for highest-performing random forest model. Variables marked as "top" are ranked into the top ten according to summed rankings of GINI decrease, accuracy decrease, and p-value.

## Discussion

Although our efforts did not result in a level of accuracy that could reasonably be of use to conservation planners, we can draw several important conclusions to aid future efforts at genus classification. For one, the inclusion of our more experimental layers (the texture rasters, LiDAR PCs, and soils layer) did increase overall classification accuracy. Second, segmenting trees using the *lidR* package resulted in higher accuracy across the board than other segmentation methods.

Classification attempts were robust with respect to model type and genus threshold; although random forest models consistently performed better than SVM or multinomial regression, they were rarely more than a few percentage points off. Similarly, adjustments to genus threshold within a range of 50 to 200 seemed to have minimal effect on resulting accuracy and kappa scores.

We also found that maples were very often subject to erroneous classification insofar as trees of other genera were consistently classed as maples; in the summary of our highest performing model generated with the *confusionMatrix* function in the *caret* package for R, the *Acer* class had the highest sensitivity of any class (0.92) and the lowest specificity (0.51). This is likely because *Acer* has the highest prevalence within our training and testing data: 37% for the dataset of *lidR* segmented trees. Because maples are shade-tolerant and often found in the sub-canopy, it is possible that a large number of canopy segments were mislabeled due to the presence of subcanopy maple stems. Furthermore, due to the popularity of maples as a street tree and the prevalence of varied cultivars within the street tree dataset, it may be that *Acer* encompasses a wider variety of species and cultivars than most other genera, making it difficult to separate out effectively.

Maples aside, we can identify more general sources of error. Foremost among these is poor alignment between our training data and delineated tree crowns. Our training data was drawn from five different sources, only one of which was gathered by the project team. These stem locations often did not overlap with delineated crowns, cutting our available training and testing data by half in some cases. Moreover, it is likely that some datasets did not carefully delineate between over- and understory trees. More careful data cleaning may have resulted in higher quality data.

Based on the above, we have the following recommendations for improving the genus classification process:

- Given the consistently higher performance of lidR segmentation, we recommend that greater care and attention be given to tuning the parameters of the *find_trees* and *segment_trees* algorithms. In particular, it may be beneficial to use different parameters for street trees and dense forest canopies by selectively masking out one or the other and then merging the resulting layers. It may also be worth devoting time and computational resources to implementing segmentation methods that operate directly on LiDAR point clouds, such as the deciduous-focused method developed by Lu et al. (2013).
- Important metrics calculated from the random forest model indicate that both the texture rasters and the LiDAR principal components were important contributors to classification accuracy. As with segmentation, both of these variables admit extensive parameter optimization. In particular, we recommend experimenting with moving window sizes for the texture layers and testing different voxel lengths, widths, and heights for the LiDAR PCA. In particular, wider and flatter voxel sizes (similar to those used in Ciuti et al. (2017) may capture less *within-tree* variation and instead characterize broad forest types more reliably.
- As mentioned above, more care should be taken to clean training data. Using one source at a time and comparing resulting accuracy could provide important insight into the quality of the data provided.
- There is room to incorporate a broader temporal range to take advantage of plant phenology beyond the inclusion of leaf-off data.
- As discussed above, the genus *Acer* may encompass a wider range of species and cultivars than most other genera. Models might perform better if samples of *Acer* are grouped at the species level rather than as a single genus.

## Chapter 5: Project Conclusions

One major rationale for this project was to find and preserve mature, native trees located on property not owned by the City. In Table 5.1, we found that 82% of the urban canopy exists on non-city owned land. Our manual fragment delineation covers 14% of the total canopy, with two-thirds located outside of City-owned property. Clustering classification results indicate that roughly twice as much, 28% of the total canopy, should be considered a native forest fragment candidate, with most of that additional area falling outside of City-owned land. Furthermore, turfgrass was identified as 50% of the total canopy.

Given that clustering found roughly twice as much native forest as manual delineation and almost three times as much native forest outside City-owned property, we can tentatively conclude that machine learning is a useful and justified approach to the problem of identifying native forest fragments on private land, though more ground-truthing and quality control is needed.

| **Category** | **Acreage** | **% of Total Canopy** |
|---|---|---|
| Urban canopy | 6,900 | 100% |
| Urban canopy (non-city-owned) | 5,653 | 82% |
| Manually delineated fragments | 994 | 14% |
| Manually delineated fragments (non-city-owned) | 624 | 9% |
| Fragments ID'd by clustering | 1,966 | 28% |
| Fragments ID'd by clustering (non-city-owned) | 1,489 | 22% |
| Turfgrass | 3,442 | *50%* |

**Table 5.1.** Acreage and total canopy cover of manually delineated fragments, fragments identified by clustering, and turfgrass on both city and non-city owned property.

Overall, our efforts have shown that there is a substantial and ecologically relevant amount of fragmented old-growth forest within the city's neighborhoods and outside managed natural areas,

thus demonstrating the utility and importance of leveraging remotely sensed data to aid in stewardship actions.

Finally, we strongly recommend that all three analyses presented here be refined and repeated when possible. In each chapter conclusion we have outlined specific suggestions for improving upon our work. Implementing these could substantially improve accuracy and use value of each layer. Furthermore, recreating these maps with new data as it becomes available would allow our clients to track changes to the city's urban canopy and turf grass coverage over time. The ability to identify trends in forest fragmentation and succession could be an invaluable aid to uplifting the urban canopy's contribution to a healthy, sustainable, and vibrant community.

## References

Alvey, A. A. (2006). Promoting and preserving biodiversity in the urban forest. *Urban Forestry & Urban Greening, 5*(4), 195-201. https://doi.org/10.1016/j.ufug.2006.09.003.

Beatley, T. (2011). *Biophilic Cities: Integrating Nature into Urban Design Planning*. Washington D.C.: Island Press.

Berland, A., Shiflett, S. A., Shuster, W. D., Garmestani, A. S., Goddard, H. C., Herrmann, D. L., & Hopton, M. E. (2017). The role of trees in urban stormwater management. *Landscape and Urban Planning, 162*, 167-177. https://doi.org/10.1016/j.landurbplan.2017.02.017.

Berry, R., Livesley, S. J., & Aye, L. (2013). Tree canopy shade impacts on solar irradiance received by building walls and their surface temperature. *Building and Environment, 69*, 91-100. https://doi.org/10.1016/j.buildenv.2013.07.009.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., & Solberg, S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment, 114,* 911-924. https://doi.org/10.1016/j.rse.2009.12.004.

Chojnacky, D. C., Heath, L.S., & Jenkins, J. C. (2014). Updated generalized biomass equations for North American tree species. *Forestry: an international journal of forest research, 87* (1), 129-151. https://doi.org/10.1093/forestry/cpt053.

City of Ann Arbor. (2021). Trees. [Data file]. Retrieved from https://www.a2gov.org/services/data/Pages/default.aspx.

City of Ann Arbor. (2020, April). *$A^2$Zero Climate Action Plan*. City of Ann Arbor. https://www.a2gov.org/departments/sustainability/Documents/A2Zero%20Climate%20Action%20Plan%20_4.0.pdf.

City of Ann Arbor. (2014). City of Ann Arbor urban & community forest management plan. https://www.a2gov.org/departments/forestry/Documents/UCFMP_FINAL_022515.pdf.

Ciuti, S., Tripke, H., Antkowiak, P., Silveyra Gonzalez, R., Dormann, C. F., & Heurich, M. (2017). An efficient method to exploit LiDAR data in animal ecology. *Methods in Ecology and Evolution 9,* 893-904. https://doi.org/10.1111/2041-210X.12921.

Decina, S. M., Ponette-González, A. G., & Rindy, J. E. (2020). Urban tree canopy effects on water quality via inputs to the urban ground surface. In: D. F. Levia, K. Nanko, D. E. Carlyle-Moses, B.

Michalzik, A. Tischer, & S. Iida (Eds.), *Forest-Water Interactions* (pp. 433-457). Springer Nature. https://doi.org/10.1007/978-3-030-26086-6_18.

Fung, C. K., & Jim, C. Y. (2019). Microclimatic resilience of subtropical woodlands and urban-forest benefits. *Urban Forestry & Urban Greening*, *42*, 100-112. https://doi.org/10.1016/j.ufug.2019.05.014.

Georgi, N. J., & Zafiriadis, K. (2006). The impact of park trees on microclimate in urban areas. *Urban Ecosystems, 9*, 195-206. https://doi.org/10.1007/s11252-006-8590-9.

Hanou, Ian. (2010). Ann Arbor, Michigan urban tree canopy (UTC) assessment. https://www.a2gov.org/departments/forestry/Documents/Ann%20Arbor%20UTC%20Report%20-%20AMEC%20-%20April-10.pdf.

Hartig, T., Mitchell, R., De Vries, S., & Frumkin, H. (2014). Nature and health. *Annual Review of Public Health, 35*, 207-228. https://doi.org/10.1146/annurev-publhealth-032013-182443.

Hastings, J. H., Ollinger, S. V., Ouimette, A. P., Sanders-DeMott, R., Palace, M. W., Ducey, M. J., Sullivan, F. B., Basler, D., & Orwig, D. A. (2020). Tree species traits determine the success of LiDAR-based crown mapping in a mixed temperate forest. *Remote Sensing, 12*(309). https://doi.org/10.3390/rs12020309.

Hausmann, A., Slotow, R., Burns, J. K., Di Minin, E. (2015). The ecosystem service of sense of place: Benefits for human well-being and biodiversity conservation. *Environmental Conservation, 43*(2), 117-127. https://doi.org/10.1017/S0376892915000314.

Ibanez, I. (2021). Stem maps of eight 1 ha forest plots distributed around Ann Arbor, MI and around the University of Michigan Biological Station (UMBS). [Data file]. Retrieved from https://mfield.umich.edu/dataset/stem-maps-eight-1-ha-forest-plots-distributed-around-ann-arbor-mi-and-around-university.

i-Tree ecosystem analysis: Ann Arbor. (2013). Retrieved March 01, 2021, from https://www.a2gov.org/departments/forestry/Documents/AnnArbor_iTreeEcoReport.pdf.

Jones R. E., Davis, K. L., Bradford, J. (2012). The value of trees: Factors influencing homeowner support for protecting local urban trees. *Environment and Behavior, 45*(5), 650-676. https://doi.org/10.1177/0013916512439409.

Katz, D. S., Batterman, S. A., & Brines, S. J. (2020). Improved classification of urban trees using widespread multi-temporal aerial image dataset. *Remote Sensing, 12*(15), 2475. https://doi.org/10.3390/rs12152475.

Kuhn, M. (2022). caret: Classification and Regression Training. R package. Version 60.0-91. https://cran.r-project.org/web/packages/caret/caret.pdf.

Kuo, F. E., Bacaicoa, M., & Sullivan, W. C. (1998). Transforming inner-city landscapes: Trees, sense

of safety, and preference. *Environment and Behavior, 30*(1), 28-59. https://doi.org/10.1177/0013916598301002.

Landry, S. M., & Chakraborty, J. (2009). Street trees and equity: Evaluating the spatial distribution of an urban amenity. *Environment and Planning A: Economy and Space, 41*(11), 2651-2670. https://doi.org/10.1068/a41236.

Liaw, A., Wiener, M. (2022). randomForest: Breiman and Cutler's Random Forests for classification and regression. R Package. Version 4.7-1. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.

Livesley, S. J., McPherson, E.G., & Calfapietra, C. (2016). The urban forest and ecosystem services: Impacts on urban water, heath, and pollution cycles at the tree, street, and city scale. *Journal of Environmental Quality, 45*(1), 119-124. https://doi.org/10.2134/jeq2015.11.0567.

Locke, D. H., Hall, B., Grove, M., Pickett, S. T. A., Ogden, L. A., Aoki, C., Boone, C. G., & O'Neil-Dunne, J. P. M. (2021). Residential housing segregation and urban tree canopy in 37 US cities. *Urban Sustainability, 1*(15). https://doi.org/10.1038/s42949-021-00022-0.

Loughner, C. P., Allen, D. J., Zhang, D., Pickering, K. E., Dickerson, R. R., & Landry, L. (2012). Roles of urban tree canopy and buildings in urban heat island effects: Parameterization and preliminary results. *Journal of Applied Meteorology and Climatology, 51*(10), 1175-1793. https://doi.org/10.1175/JAMC-D-11-0228.1.

Lu, X., Guo, Q., Li, W., & Flanagan, J. (2014). A bottom-up approach to segment individual deciduous trees using leaf-off lidar point cloud data. *ISPRS Journal of Photogrammetry and Remote Sensing, 94*, 1-12. https://doi.org/10.1016/j.isprsjprs.2014.03.014.

Matasci, G., Coops, N. C., Williams, D. A. R., & Page, N. (2018). Mapping tree canopies in urban environments using airborne laser scanning (ALS): a Vancouver case study. *Forest Ecosystems, 5*(31). https://doi.org/10.1186/s40663-018-0146-y.

Matthaei-Nichols Natural Areas Data Resources. (2018). Nichols Arboretum trees. [Data file].

McPherson, E. G., & Simpson, J. R. (2003). Potential energy savings in building by an urban tree planting programme in California. *Urban Forestry & Urban Greening, 2*(2), 73-86. https://doi.org/10.1078/1618-8667-00025.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F. (2021). e1071: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R Package. Version 1.7-9. https://cran.r-project.org/web/packages/e1071/e1071.pdf.

Monteiro, J. A. (2017). Ecosystem services from turfgrass landscapes. *Urban Forestry & Urban Greening, 26,* 151-157. https://doi.org/10.1016/j.ufug.2017.04.001.

Nearmap Limited (July 17, 2018). 3.918-Foot Resolution aerial imagery. [Data file]. Retrieved September 28, 2021.

Nowak, D. J., & Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the USA. *Environmental Pollution, 116*(3), 381-389. https://doi.org/10.1016/S0269-7491(01)00214-7.

Nowak, D. J., Crane, D. E., & Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry & Urban Greening, 4*(3-4), 115-123. https://doi.org/10.1016/j.ufug.2006.01.007.

Office of Sustainability and Innovations. *About Home*. City of Ann Arbor. Retrieved April 14, 2022. https://www.a2gov.org/departments/sustainability/about/Pages/default.aspx.

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

Oono, K., & Tsuyuki, S. (2018). Estimating individual tree diameter and stem volume using airborne LiDAR in Saga Prefecture, Japan. *Open Journal of Forestry, 08*(02), 205-228. https://doi.org/10.4236/ojf.2018.82015.

Ossola, A., Locke, D., Lin, B., & Minor, E. (2019). Yards increase forest connectivity in urban landscapes. *Landscape Ecology, 34*, 2935-2948. https://doi.org/10.1007/s10980-019-00923-7.

Paluszynska, A., Biecek, P., & Jiang, Y. (2020). randomForestExplainer: Explaining and visualizing Random Forests in terms of variable importance. R Package. Version 0.10.1. https://cran.r-project.org/web/packages/randomForestExplainer/randomForestExplainer.pdf.

Ripley, B. (2022). nnet: Feed-forward neural networks and multinomial log-linear models. R package. Version 7.3-17. https://cran.r-project.org/web/packages/nnet/nnet.pdf.

Roussel, J. -R. & Auty, D. (2022). lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R package. Version 4.0.0. https://cran.r-project.org/web/packages/caret/caret.pdf

Roussel, J. -R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R. H., Sánchez Meador, A., 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. -R., Goodbody, T. R. H., & Tompalski, P. (2021). The lidR package. *Github.*

https://r-lidar.github.io/lidRbook/index.html.

Roy, S., Byrne, J., Pickering, C. (2012). A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. *Urban Forestry & Urban Greening, 11*(4), 351-363. https://doi.org/10.1016/j.ufug.2012.06.006.

Runfola, D. M., Polsky, C., Giner, N., Pontius, R. G., & Nicolson, C. (2013). Future suburban development and the environmental implications of lawns: A case study in New England, USA. In D. Malkinson, D. Czamanski, & I. Benenson (Eds.), *Modeling of land-use and ecological dynamics* (pp. 119-141). Cities and Nature. https://doi.org/10.1007/978-3-642-40199-2_7.

Schwarz, K., Fragkias, M., Boone, C. G., Zhou, W., McHale, M., Grove, J. M., O'Neil-Dunne, J., McFadden, J. P., Buckley, G. L., Childers, D., Ogden, L., Pincetl, S., Pataki, D., Whitmer, A., & Cadenasso, M. L. (2015). Trees grow on money: Urban tree canopy cover and environmental justice. *PLoS ONE, 10*(4). https://doi.org/10.1371/journal.pone.0122051.

Silva, C. A., Hudak, A. T., Vierling, L. A., Loudermilk, E. L., O'Brien, J. J., Hiers, J. K., Khosravipour, A. (2016). Imputation of individual longleaf Pine (Pinus palustris Mill.) tree attributes from field and LiDAR data. *Canadian Journal of Remote Sensing, 42*(5), 554–573. https://doi.org/10.1080/07038992.2016.1196582.

Smardon, R. C. (1988). Perception and aesthetics of the urban environment: Review of the role of vegetation. *Landscape and Urban Planning, 15*(1-2), 85-106. https://doi.org/10.1016/0169-2046(88)90018-7.

State of Michigan. (2017). South Michigan - Washtenaw LiDAR - Michigan 2017. [Data file].

University of Michigan. (2014). Campus Trees. [Data file].

The Urban Forest. (n.d.). Retrieved March 01, 2021, from https://www.a2gov.org/departments/forestry/Pages/The-Urban-Forest.aspx.

USDA. *Plant Database*. United States Department of Agriculture. https://plants.usda.gov/home.

USDA. (2016). Web Soil Survey. [Data file]. Retrieved from https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx.

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., & White, E. (2019). Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. *Remote Sensing, 11*(11), 1–13. https://doi.org/10.3390/rs11111309.

White, D. L., & Lloyd. T. (1994, November 1-3). *Defining old growth: Implications for management* [Conference presentation]. Eight Biennial Southern Silvicultural Research Conference, Auburn, Alabama, United States. https://www.fs.usda.gov/treesearch/pubs/741.

Zang, C., & Qiu, F. (2012). Mapping individual tree species in an urban forest using airborne Lidar data and hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing, 10*, 1079-1089. https://doi.org/10.14358/PERS.78.10.1079.

Ziegler, P. (2016). Using NAIP Imagery and a Texture Raster to Model the Urban Forest. *Azavea.* https://www.azavea.com/blog/2016/09/14/multispectral-1m-naip-imagery-model-urban-forest/.

Quantum Spatial, Inc. (2020). Ann Arbor 2020 0.5-Foot Orthoimagery. [Data file].

Quantum Spatial, Inc. (2018). Ann Arbor 2018 0.5-Foot Orthoimagery. [Data file].

# Appendices

## Appendix A: Predictor Layer Preparation

We gathered six raster datasets (Table 4.3) to use as predictor variables in the genus and native fragment classification processes.

| Name | Year | Format or Resolution |
|---|---|---|
| Color-Infrared (CIR) | 2020 | 0.5 feet |
| Color-Infrared (CIR) | 2018 | 0.5 feet |
| National Agriculture Imagery Program (NAIP) | 2018 | 2 feet |
| NearMap | July 17, 2018 | 3.918 feet |
| Light Detection and Ranging (LiDAR) | 2018 | 1,101,733,871 points, 1.246-1.9 feet point-spacing |
| Soils | September 16, 2019 | N/A (vector shapefile) |

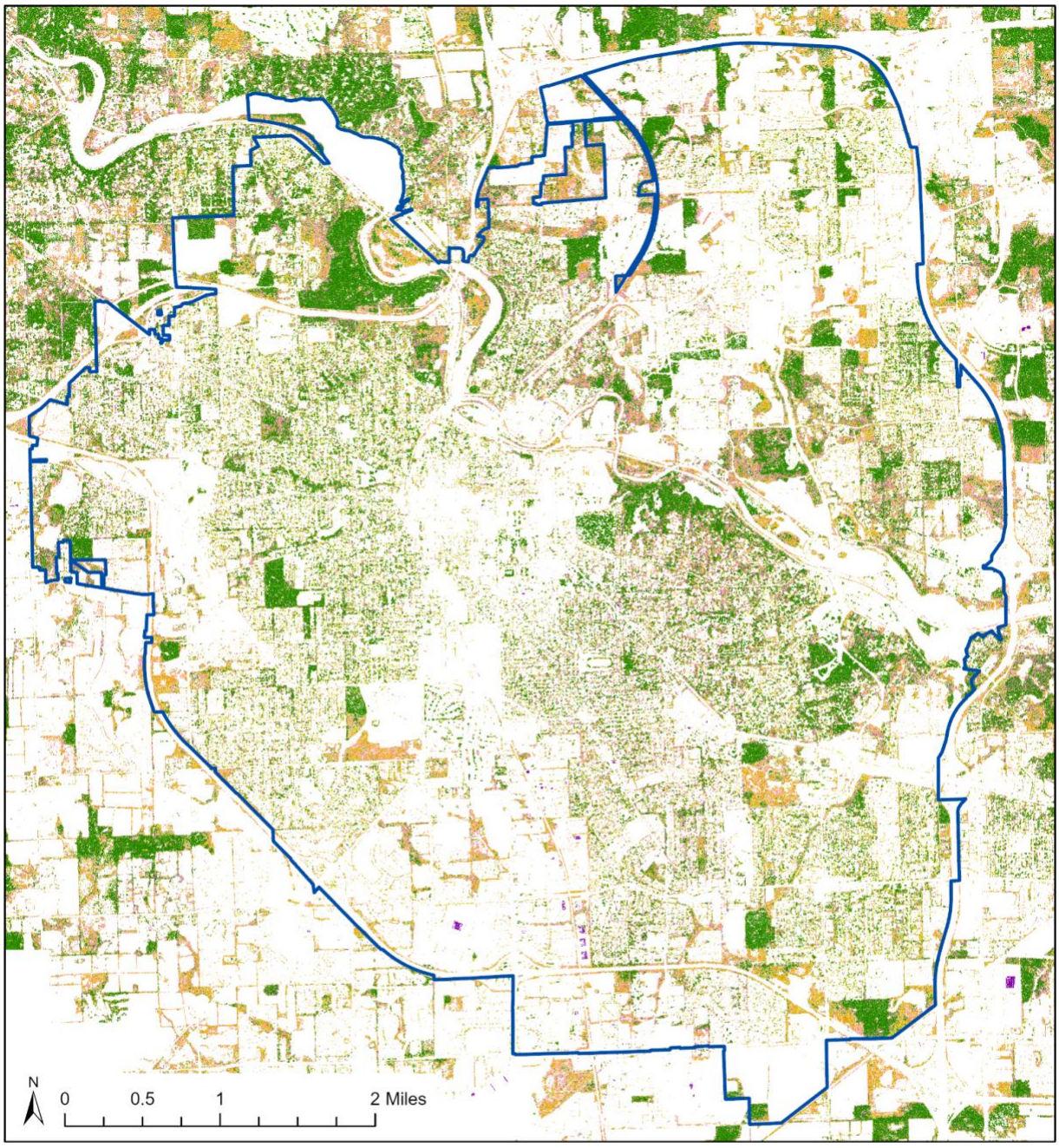
**Table A.1.** Predictor Raster datasets.

Before utilizing the rasters in analyses, georectification was carried out to correct for misalignment between datasets. 115 control points were established based on building chimneys in the 2020 Color-Infrared imagery. This feature was chosen due to its scale and commonality. A first-order polynomial transformation was applied to 2018 CIR and 2018 NAIP imagery.

An NDVI layer was calculated from the NAIP imagery using the built-in NDVI raster function in ArcGIS Pro. This function automatically scales the resulting raster from 0 to 255. Texture rasters were created from the green and near-infrared NAIP bands by using a seven pixel focal window with the range statistic, following Ziegler (2016). Additionally, we created a canopy height model from 2017 LiDAR data using the pit-free algorithm in the *lid*R package (Roussel & Auty, 2022) for R.

As experimental additional predictors aimed at characterizing overall canopy structure, a principal component analysis was performed on the 2017 LiDAR data following the methodology put forward by Ciuti et. al. (2017). This process has been described in full in Appendix E.

## Appendix B: Unsupervised Vegetation Classification




Ann Arbor Forests Full Classification

Legend

City Boundary

Unsupervised Vegetation Classification

0: Other

1: Tall shrub

2: Small trees/ undercanopy

3: Small trees

4: Mature, emergent trees

5: Small to medium trees, forest perimeter

6: Medium trees, undercanopy

7: Street trees, second growth

8: Shrubs (particularly Buckthorn)

9: Intermediate canopy

10: Street tree lower canopy

11: Shrub and small trees

12: Buildings

Data Sources: City of Ann Arbor, NAIP 2018, SEMCOG
Map by Christian Schluter, April 2022

## Appendix C: Confusion Matrix for Top Performing Random Forest Model

| | | Reference | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ACER | AMELA | BETUL | CARYA | CATAL | CELTI | CERCI | CRATA | GLEDI | GYMNO | JUGLA | LIRIO | MALUS | MORUS | Other | PICEA | PINUS | PLATA | POPUL | PRUNU | PSEUD7 | PYRUS | QUERC | ROBIN | SOPHO | THUJA | TILIA | ULMUS | ZELKO |
| Classification | ACER | 5098 | 40 | 39 | 106 | 56 | 51 | 57 | 33 | 530 | 46 | 159 | 86 | 239 | 39 | 373 | 148 | 173 | 429 | 50 | 75 | 21 | 145 | 596 | 44 | 53 | 31 | 496 | 269 | 102 |
| | AMELA | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | BETUL | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | CARYA | 4 | 0 | 0 | 51 | 0 | 0 | 1 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 29 | 0 | 0 | 0 | 1 | 0 | 0 |
| | CATAL | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | CELTI | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | CERCI | 0 | 0 | 1 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| | CRATA | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | GLEDI | 194 | 4 | 4 | 7 | 0 | 4 | 3 | 4 | 899 | 6 | 19 | 0 | 18 | 4 | 29 | 9 | 13 | 18 | 1 | 8 | 6 | 26 | 85 | 2 | 15 | 6 | 96 | 35 | 46 |
| | GYMNO | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | JUGLA | 18 | 1 | 0 | 16 | 1 | 1 | 4 | 0 | 6 | 0 | 84 | 0 | 2 | 2 | 6 | 3 | 3 | 1 | 3 | 4 | 0 | 1 | 25 | 2 | 1 | 1 | 12 | 7 | 0 |
| | LIRIO | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | MALUS | 29 | 3 | 2 | 0 | 0 | 2 | 2 | 16 | 5 | 0 | 0 | 0 | 137 | 0 | 12 | 8 | 14 | 0 | 0 | 8 | 2 | 5 | 10 | 0 | 0 | 1 | 5 | 1 | 3 |
| | MORUS | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | Other | 5 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 2 | 58 | 4 | 8 | 1 | 1 | 1 | 2 | 1 | 3 | 0 | 0 | 0 | 3 | 5 | 0 |
| | PICEA | 18 | 3 | 6 | 1 | 1 | 0 | 4 | 0 | 4 | 3 | 4 | 0 | 16 | 2 | 22 | 168 | 38 | 1 | 0 | 6 | 6 | 5 | 7 | 1 | 0 | 4 | 3 | 4 | 0 |
| | PINUS | 41 | 3 | 3 | 1 | 1 | 0 | 2 | 2 | 9 | 0 | 2 | 0 | 32 | 1 | 35 | 76 | 276 | 2 | 0 | 9 | 9 | 3 | 15 | 0 | 0 | 7 | 10 | 4 | 5 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PLATA | 37 | 2 | 2 | 1 | 0 | 0 | 1 | 1 | 4 | 3 | 3 | 1 | 1 | 0 | 10 | 3 | 2 | 289 | 1 | 3 | 1 | 1 | 14 | 0 | 3 | 0 | 8 | 0 | 0 |
| | POPUL | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 |
| | PRUNU | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 11 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 1 | 0 |
| | PSEUD7 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | PYRUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 1 | 0 | 24 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | QUERC | 101 | 5 | 7 | 127 | 4 | 2 | 13 | 0 | 8 | 5 | 72 | 9 | 12 | 3 | 44 | 13 | 19 | 6 | 35 | 51 | 1 | 2 | 541 | 22 | 0 | 2 | 30 | 29 | 1 |
| | ROBIN | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 1 | 0 |
| | SOPHO | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | THUJA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 0 | 0 | 0 |
| | TILIA | 25 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 16 | 0 | 5 | 0 | 7 | 2 | 4 | 3 | 2 | 1 | 1 | 2 | 1 | 3 | 7 | 2 | 1 | 0 | 139 | 1 | 4 |
| | ULMUS | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 17 | 0 |
| | ZELKO | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |

## Appendix D: Additional Genus Classification Model Tests

All listed model runs were done on the full Ann Arbor dataset with a random seed of 1.

| Segment. method | Genus Threshold | Training Data | Predictor layers | # Trees | # Genera | RF Acc. | RF $\hat{k}$ | SVR Acc. | SVR $\hat{k}$ | Multinom. Acc. | Multinom. $\hat{k}$ | Max Acc. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| lidR, 30 ft. window | 100 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10), texture rasters of NAIP Green and NIR, CHM texture, soils, LiDAR PCs 1-6 | 29399 | 29 | 0.543 | 0.372 | 0.527 | 0.339 | 0.525 | 0.368 | 0.543 |
| lidR, 30 ft. window | 100 | All | (No soil) CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10), texture rasters of NAIP Green and NIR, CHM texture, LiDAR PCs 1-6 | 29399 | 29 | 0.537 | 0.364 | 0.540 | 0.366 | 0.519 | 0.353 | 0.540 |
| lidR, 30 ft. window | 100 | All | (No soil, PCs 1-3 only) CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10), texture rasters of NAIP Green and NIR, CHM texture, soils, LiDAR PCs 1-3 | 29399 | 29 | 0.538 | 0.367 | 0.538 | 0.363 | 0.517 | 0.349 | 0.538 |
| lidR, 30 ft. window | 100 | All | Rescaled [CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10), texture rasters of NAIP Green and NIR, CHM texture, soils] | 33513 | 30 | 0.518 | 0.341 | 0.483 | 0.264 | 0.489 | 0.311 | 0.518 |
| lidR, 30 ft. window | 100 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10), texture rasters of NAIP Green and NIR, CHM texture, soils | 33513 | 30 | 0.518 | 0.340 | 0.483 | 0.264 | 0.488 | 0.311 | 0.518 |
| lidR, 30 ft. window | 20 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10), texture rasters of NAIP Green and NIR, CHM texture, soils | 33513 | 53 | 0.514 | 0.335 | 0.483 | 0.264 | 0.488 | 0.310 | 0.514 |
| lidR, 30 ft. window | 100 | All | (No CIR) CHM, NDVI, NAIP2018, Nearmap2018 (all clipped to chm>10ft NDVI>10), texture rasters of NAIP Green and NIR, CHM texture, soils | 33513 | 30 | 0.505 | 0.319 | 0.458 | 0.211 | 0.464 | 0.263 | 0.505 |
| lidR, 30 ft. window | 400 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10) | 33522 | 15 | 0.488 | 0.294 | 0.479 | 0.257 | 0.423 | 0.223 | 0.488 |
| lidR, 30 ft. window | 200 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10) | 33522 | 18 | 0.488 | 0.291 | 0.476 | 0.249 | 0.411 | 0.197 | 0.488 |
| lidR, 30 ft. window | 40 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10) | 33522 | 21 | 0.481 | 0.279 | 0.472 | 0.243 | 0.417 | 0.213 | 0.481 |
| Pixel | 200 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10), texture rasters of NAIP Green and NIR, CHM texture, soils, LiDAR PCs 1-6 | 35945 | 20 | 0.470 | 0.290 | 0.447 | 0.243 | 0.449 | 0.280 | 0.470 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pixel | 200 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10), texture rasters of NAIP Green and NIR, CHM texture, soils | 42882 | 25 | 0.440 | 0.250 | 0.431 | 0.213 | 0.404 | 0.200 | 0.440 |
| Pixel | 200 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10) | 42882 | 26 | 0.408 | 0.205 | 0.401 | 0.152 | 0.369 | 0.138 | 0.408 |
| Buffered pixel | 40 | Survey | CIR2020, CIR 2018, NAIP2018, Nearmap2018 (July) | 1542 | | 0.383 | 0.268 | X | X | X | X | 0.383 |
| Buffered pixel | 40 | Street Trees | CIR2020, CIR 2018, NAIP2018, Nearmap2018 (July) | 56588 | | 0.338 | 0.137 | X | X | X | X | 0.338 |
| Buffered pixel | 40 | All | CIR2020, CIR 2018, NAIP2018, Nearmap2018 (July) | 69338 | 98 | 0.337 | 0.179 | 0.319 | 0.112 | 0.306 | 0.112 | 0.337 |
| Buffered pixel | 100 | All | CIR2020, CIR 2018, NAIP2018, Nearmap2018 (July) | 69338 | 98 | 0.323 | 0.149 | 0.309 | 0.092 | 0.302 | 0.102 | 0.323 |
| Buffered pixel | 10 | All | CIR2020, CIR 2018, NAIP2018, Nearmap2018 (July) | 69338 | 98 | 0.322 | 0.149 | 0.310 | 0.094 | 0.301 | 0.100 | 0.322 |
| Buffered pixel | 100 | All | CHM, NDVI, NAIP2018, CIR2018, Nearmap2018 (all clipped to chm>10ft NDVI>10) | 69421 | 19 | 0.322 | 0.150 | 0.306 | 0.081 | 0.291 | 0.080 | 0.322 |
| Buffered pixel | 40 | All | CIR2020, CIR2018 | 69338 | 98 | 0.302 | 0.133 | 0.306 | 0.090 | 0.300 | 0.082 | 0.306 |
| Buffered pixel | 40 | All | Nearmap2018(July),NAIP2018 | 69338 | 98 | 0.284 | 0.107 | 0.287 | 0.055 | 0.285 | 0.057 | 0.287 |
| Buffered pixel | 40 | Campus | CIR2020, CIR 2018, NAIP2018, Nearmap2018 (July) | 5787 | | 0.149 | 0.038 | X | X | X | X | 0.149 |

## Appendix E: Summary of LiDAR PCA & R Script

### *Introduction*

As experimental additional predictors aimed at characterizing overall canopy structure, a principal component analysis (PCA) was performed on the 2017 LiDAR data. Our methodology was drawn from that used by Ciuti et. al. (2017), albeit with some simplifications. Ciuti et. al. offer two arguments for using LiDAR principal components. First, PCA allows for a "fit-first, explain later" approach to modeling that eschews the need for *a priori* assumptions about what derived LiDAR metrics (e.g. canopy height) are relevant to a particular problem. Second, PCA is effective as a dimension reducing method, allowing researchers to capture more of a LiDAR point cloud's overall variation with much less computational overhead. Ciuti et. al. applied this method to a deer browsing model, but we hypothesized that this method of quantifying vegetation structure could have some predictive value when assigning genus or native/secondary status.

### *Methods*

Using the *lidR* package for R, we produced a raster stack recording the number of LiDAR returns within 5x5x5 foot voxels. This was done for a vertical range from 0 to 125 feet, resulting in 25 total rasters. In order to ensure that the PCA would only capture variance within the urban canopy, these layers were clipped to only include areas with canopy height over 10 feet and NDVI greater than 100. The clipped rasters were then loaded into GRASS GIS and a PCA was performed using the i.pca function. Rasters for the first six principal components, accounting cumulatively for 75.99% of the variance in the input rasters, were exported as .tif files and used in classifier tests.

| Component | Proportion of Variance | Cumulative Proportion of Variance |
|---|---|---|
| 1 | 35.53% | 35.53% |
| 2 | 11.99% | 47.52% |
| 3 | 9.49% | 57.01% |
| 4 | 7.96% | 64.97% |
| 5 | 6.23% | 71.20% |
| 6 | 4.79% | 75.99% |

**Table E.1.** Principal components and the proportion of variance they account for.

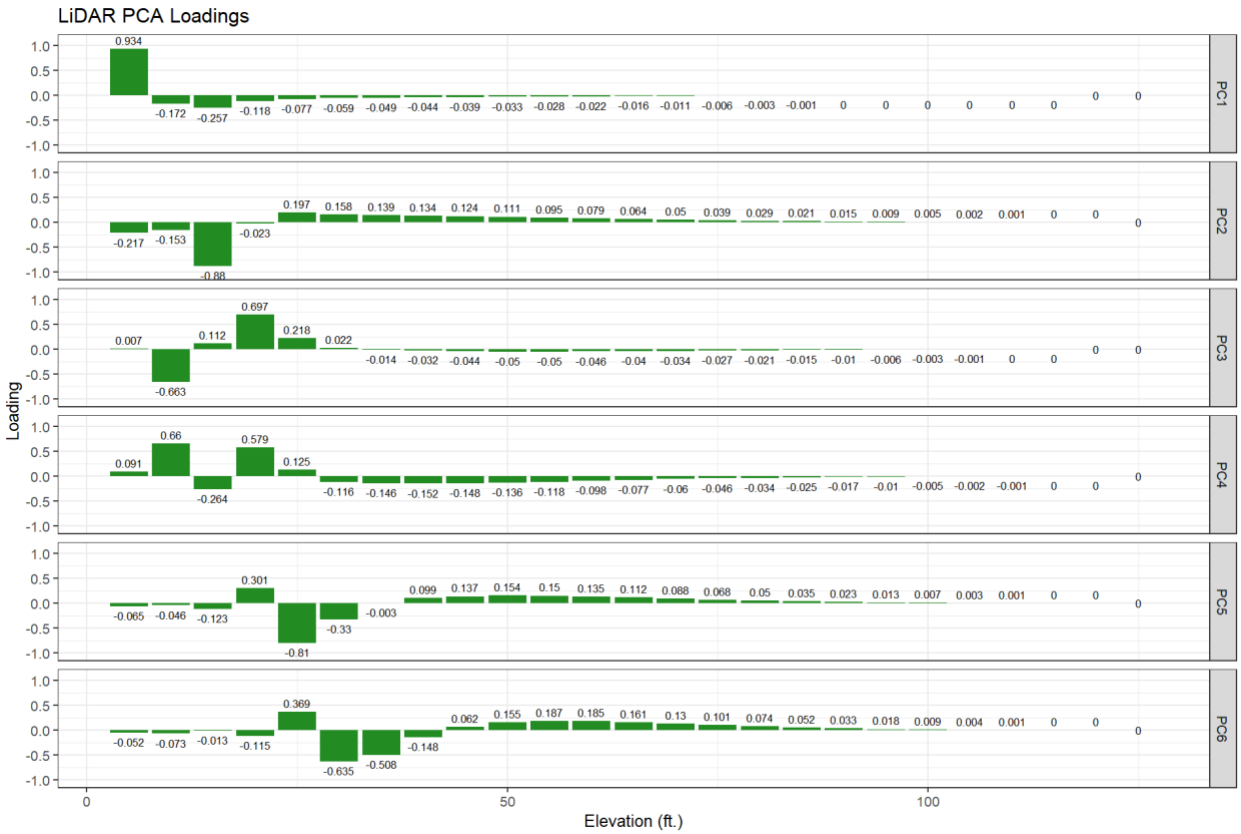



**Figure E.1.** Loadings for the first six components in the principal components analysis.

### *Interpretation*

Based on the loadings for each height interval and visual inspection of the maps produced by displaying component scores for each pixel, we can attempt a preliminary interpretation of the types of canopy structure represented by the first three components.

Component one produces higher values for areas with a high proportion of returns between zero and five feet and lower values for returns between five and fifteen feet.. Accordingly, it tends to result in lower values for areas with dense underbrush and higher values for forests without dense underbrush where pulses that penetrate the canopy tend to result in a ground return.

Component two attaches negative loadings to returns within 0-15 feet, particularly those in the 10-15 foot range. Returns greater than 20 feet are loaded positively. Visual inspection of the resulting map shows that areas known to have dense shrubs have low scores, areas identified as native forest fragments tend to have moderate scores, and maximal values occur for street trees, coniferous stands, and 'pioneer' woodlands.

Component three loads returns from 5-10 feet very negatively and loads returns from 15-20 feet very positively, with slight positive loadings for 10-15 and 20-25 feet. On the map, areas with dense

undergrowth tend to have the highest values, native fragments having middling scores and coniferous areas have the lowest values.

### *R Script*

This R script reads in a folder of LiDAR tiles and outputs a series of tiled .tif files recording return counts for each height interval. These files must be loaded separately into another GIS (we used GRASS) to calculate principal components.

```
##############################
# AAUTC LiDAR PCA Script
# Author: Thomas Estabrook
# 2/9/2022
# Updated: 1/17/2022
##############################

library(raster)
library(lidR)

#################################################################################

cloud_norm <- readLAScatalog("D:/Capstone_Large_Data/Las-
PointCloud/Normalized/")
las_check(cloud_norm)

opt_chunk_buffer <- 0
opt_select(cloud_norm) <- "xyz"

for (i in c(0:24)) {
  lower_height <- i*5
  upper_height <- lower_height + 5
  filter_cmd <- paste("-drop_z_below", lower_height, "-drop_z_above",
upper_height)
  opt_filter(cloud_norm) <- filter_cmd

  outfiles <- paste0('G:/Shared drives/A2_UTC
Drive/R_scripts/LiDAR_PCA/return_counts/h', lower_height, "_", upper_height,
'/{*}_returns', lower_height, "_", upper_height)
  opt_output_files(cloud_norm) <-  outfiles

  return_counts <- grid_metrics(cloud_norm, ~length(Z), 5)
}
```

## Appendix F: Summary of Predictor Variables for Common Tree Genera

*Introduction*

In this appendix, we present some exploratory data analysis and visualization for the LiDAR segmented dataset that was used in our highest performing random forest model in Chapter 4. This dataset contains 29,399 trees with recorded values for thirty three continuous variables, as follows:

| Variable | Description |
|---|---|
| Mean_NDVI | Mean value of NDVI raster calculated from NAIP imagery |
| Min_canopy_height | Minimum value of CHM in tree crown (ft) |
| Max_canopy_height | Maximum value of CHM in tree crown (ft) |
| Canopy_height_range | Range of CHM values in tree crown (ft) |
| Mean_canopy_height | Mean CHM value in tree crown (ft) |
| Canopy_height_sd | Standard deviation of CHM values in tree crown (ft) |
| Volume | Sum of all CHM pixels in tree crown (9 $ft^3$) |
| Median_canopy_heigh | Median of CHM values in tree crown (ft) |
| Canopy_pct90 | Height of the 90th percentile of CHM values in tree crown (ft) |
| Canopy_pct80 | Height of the 80th percentile of CHM values in tree crown (ft) |
| Canopy_pct70 | Height of the 70th percentile of CHM values in tree crown (ft) |
| MEANNAIP1A230 | Mean value of NAIP blue band in tree crown |
| MEANNAIP2A230 | Mean value of NAIP green band in tree crown |
| MEANNAIP3A230 | Mean value of NAIP red band |

| Variable | Description |
|---|---|
| MEANCIR2A230 | Mean value of CIR green band in tree crown |
| MEANCIR3A230 | Mean value of CIR red band in tree crown |
| MEANCIR4A230 | Mean value of CIR NIR band in tree crown |
| MEANNEAR1A230 | Mean value of Nearmap blue band in tree crown |
| MEANNEAR2A230 | Mean value of Nearmap green band in tree crown |
| MEANNEAR3A230 | Mean value of Nearmap red band in tree crown |
| Area_Sq_ft | Area of tree crown ($ft^2$) |
| MEANNIRTEX | Mean value of NIR texture raster calculated from NAIP imagery in tree crown |
| MEANGTEX | Mean value of green texture raster calculated from NAIP imagery in tree crown |
| MEANCHMTEX | Mean value of texture raster calculated from CHM in tree crown |
| PC1mean | Mean value of LiDAR principal component 1 in tree crown |
| PC2mean | Mean value of LiDAR principal component 1 in tree crown |
| PC3mean | Mean value of LiDAR principal component 2 in tree crown |
| PC5mean | Mean value of LiDAR principal |

| | |
|---|---|
| | in tree crown |
| MEANNAIP4A230 | Mean value of NAIP NIR band in tree crown |
| MEANCIR1A230 | Mean value of CIR blue band in tree crown |

| | |
|---|---|
| | component 3 in tree crown |
| PC6mean | Mean value of LiDAR principal component 4 in tree crown |

**Table F.1.** Predictor variables resulting from LiDAR segmentation of tree canopy.

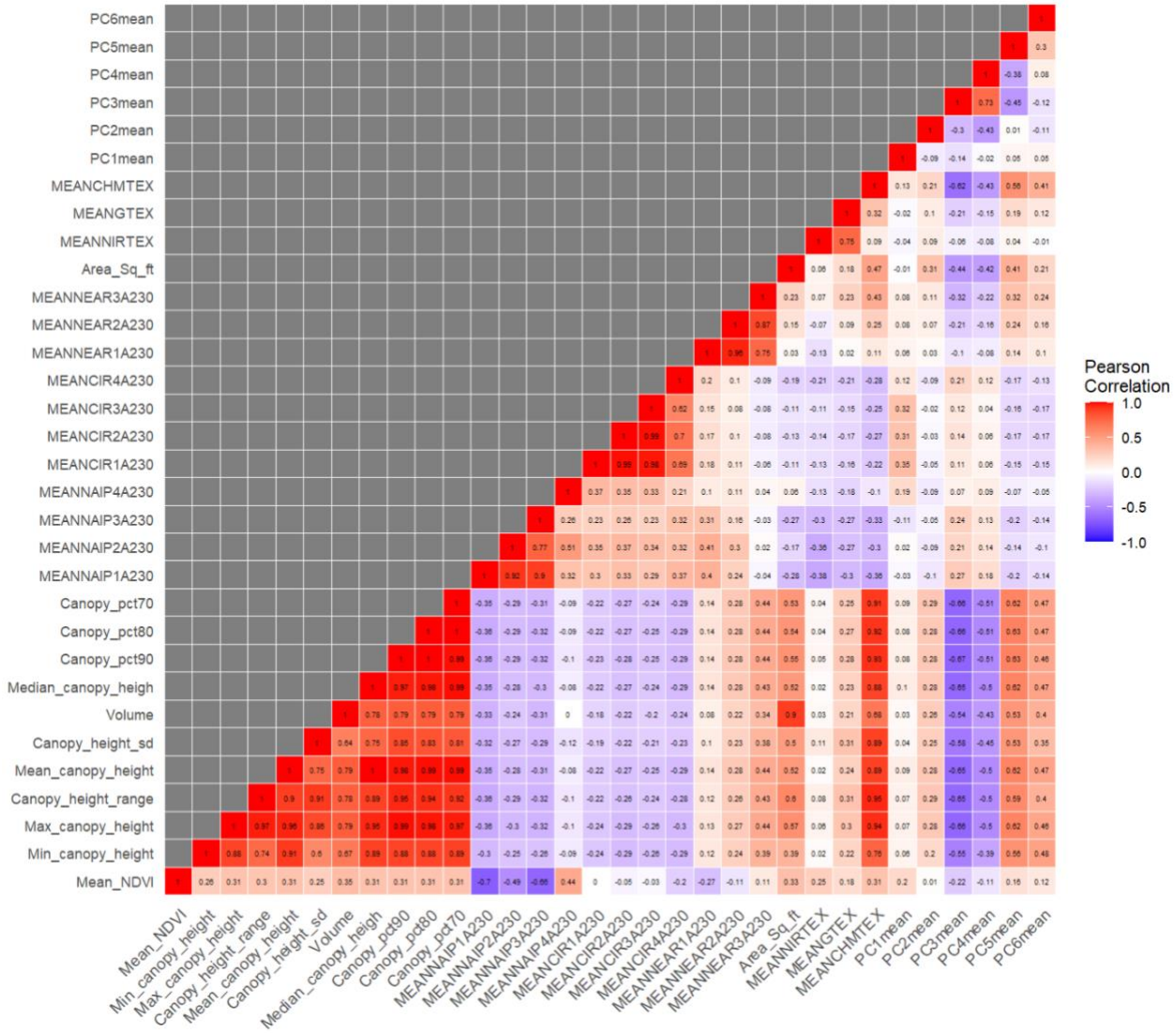



**Figure F.1.** Correlation plot of predictor variables. Strong correlations can be observed between all variables derived from LiDAR and between the bands of each imagery set. .

## *Plots*

For visualization, we selected ten genera based on prevalence within the dataset and importance in native forest communities. These are *Acer, Quercus, Carya, Gleditsa, Tilia, Plata, Pinus, Malus, Ulmus,*

and *Juglans*. Variables displayed have been selected due to their high importance in the random forest model, described in more detail in chapter 4.

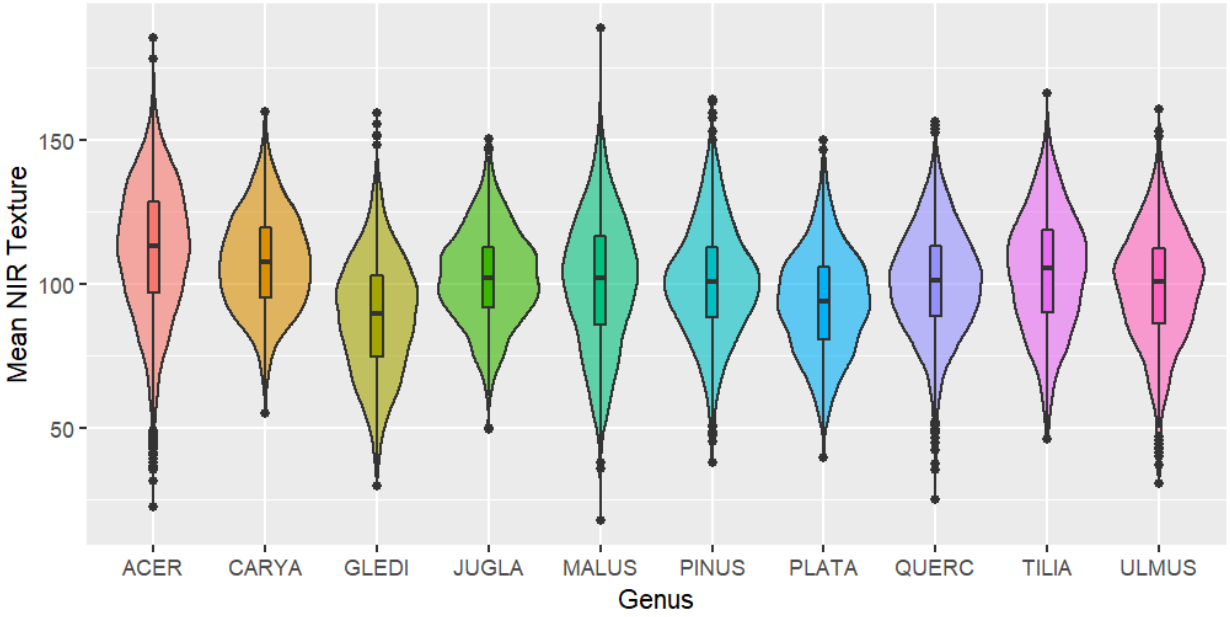



**Figure F.2.** Violin plot of mean NIR texture by genus.

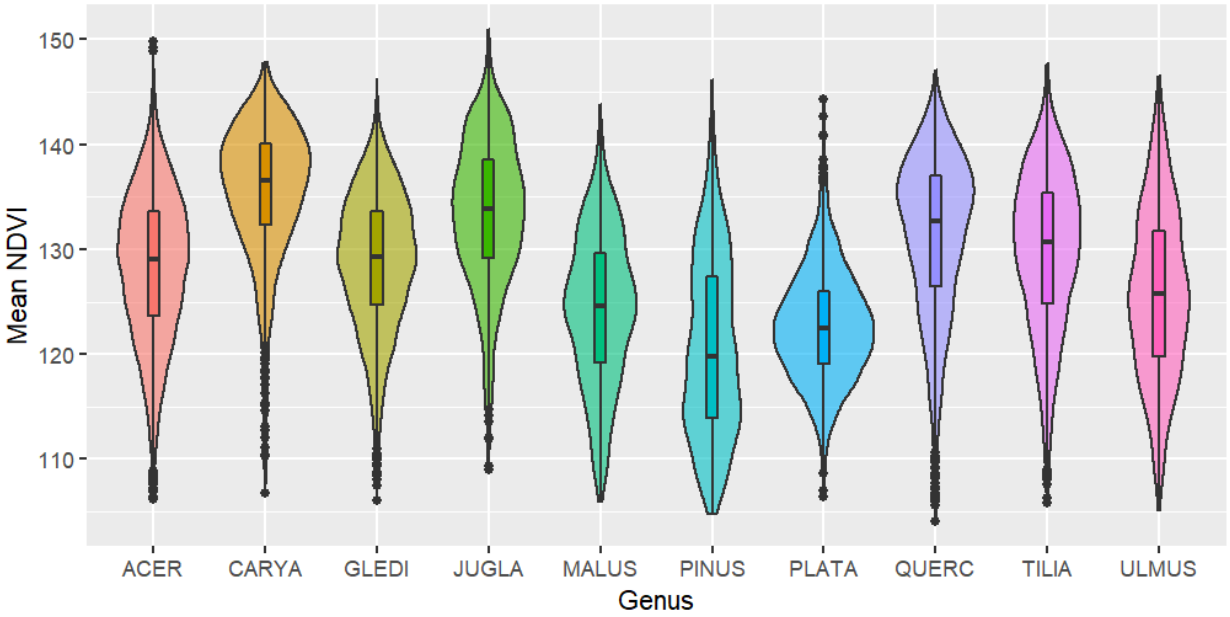



**Figure F.3.** Violin plot of mean NDVI by genus.

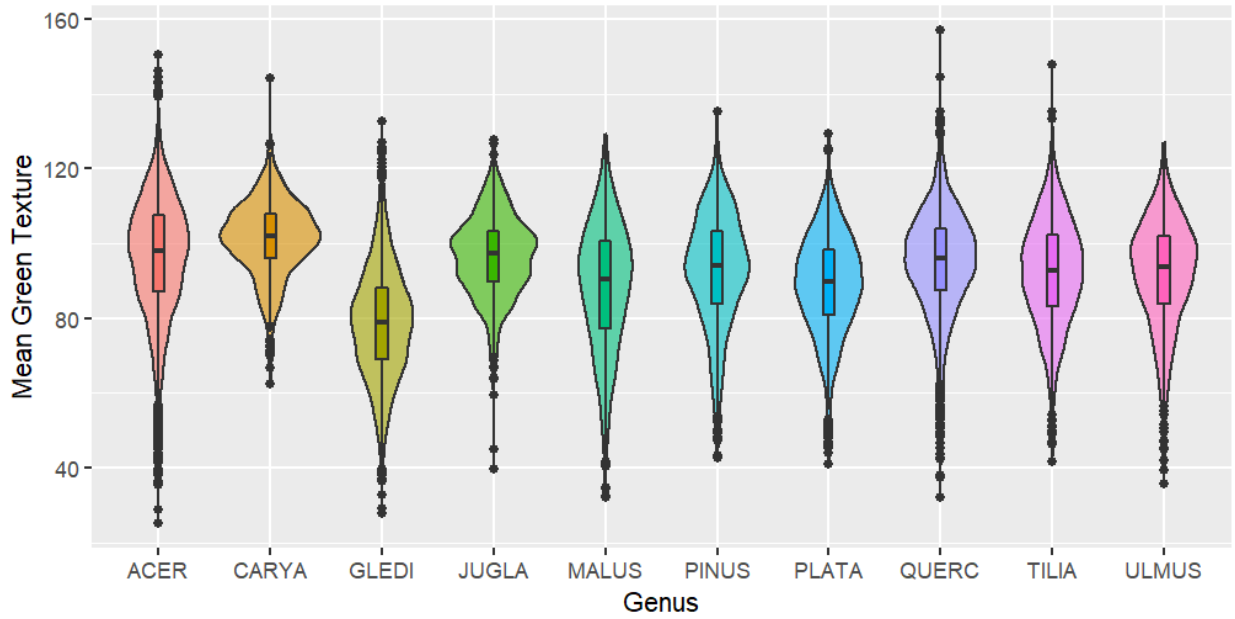



**Figure F.4.** Violin plot of mean NAIP green texture by genus.

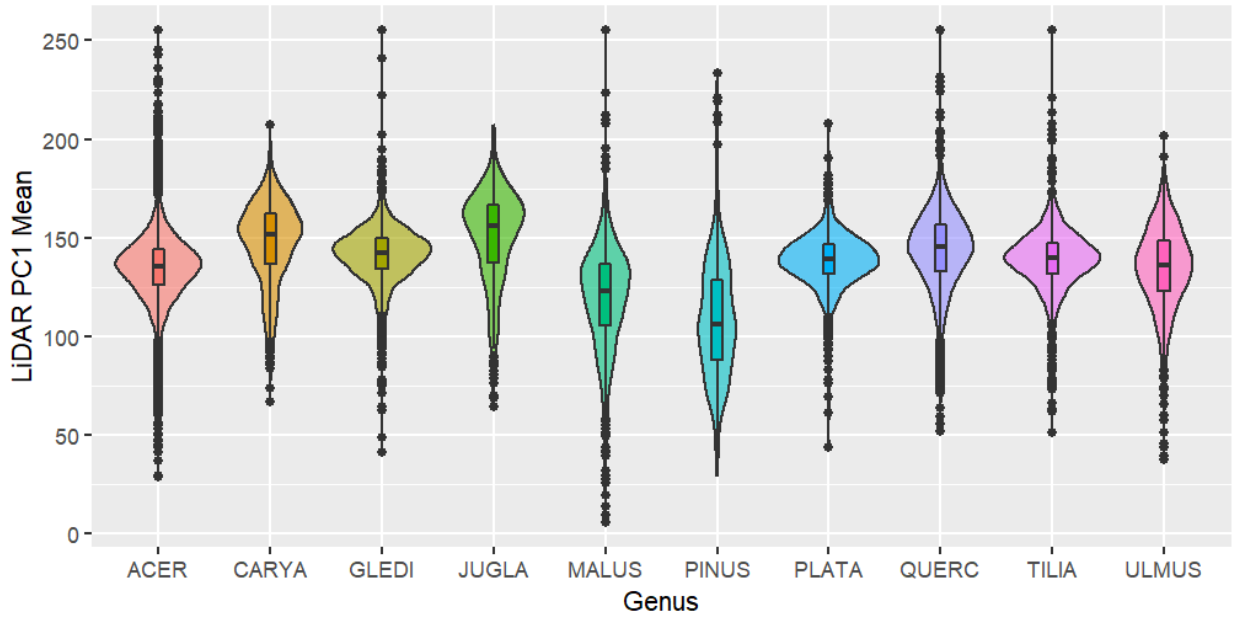



**Figure F.5.** Violin plot of mean value of LiDAR PC1 by genus.

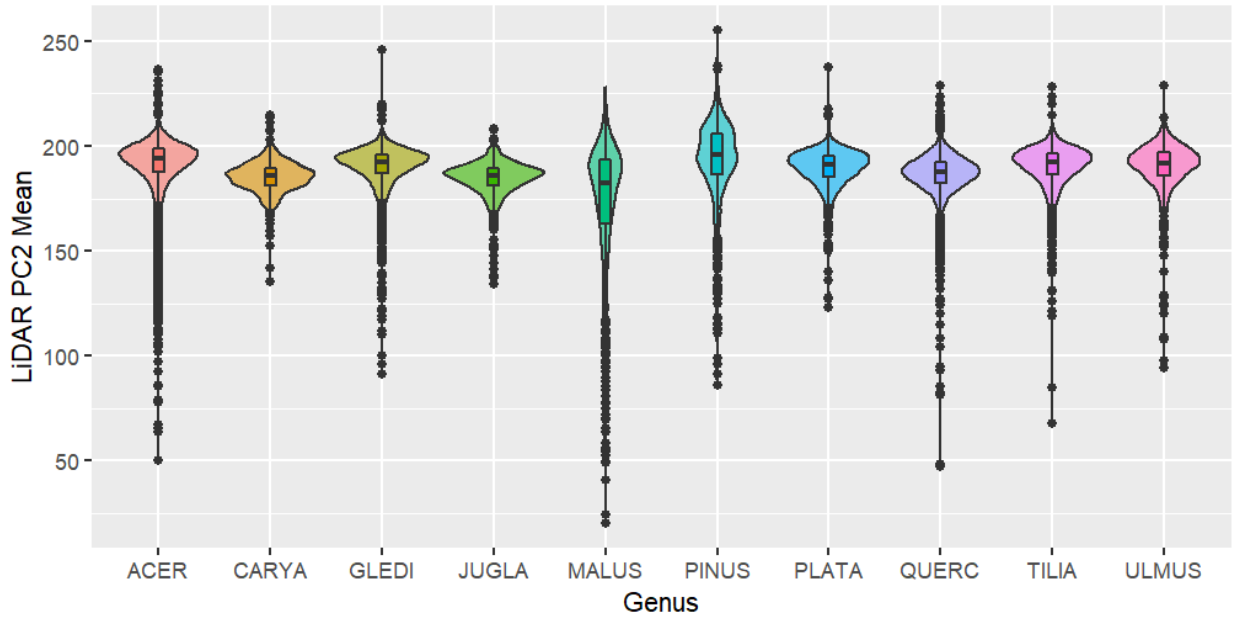



**Figure F.6.** Violin plot of mean LiDAR PC 2 by genus.

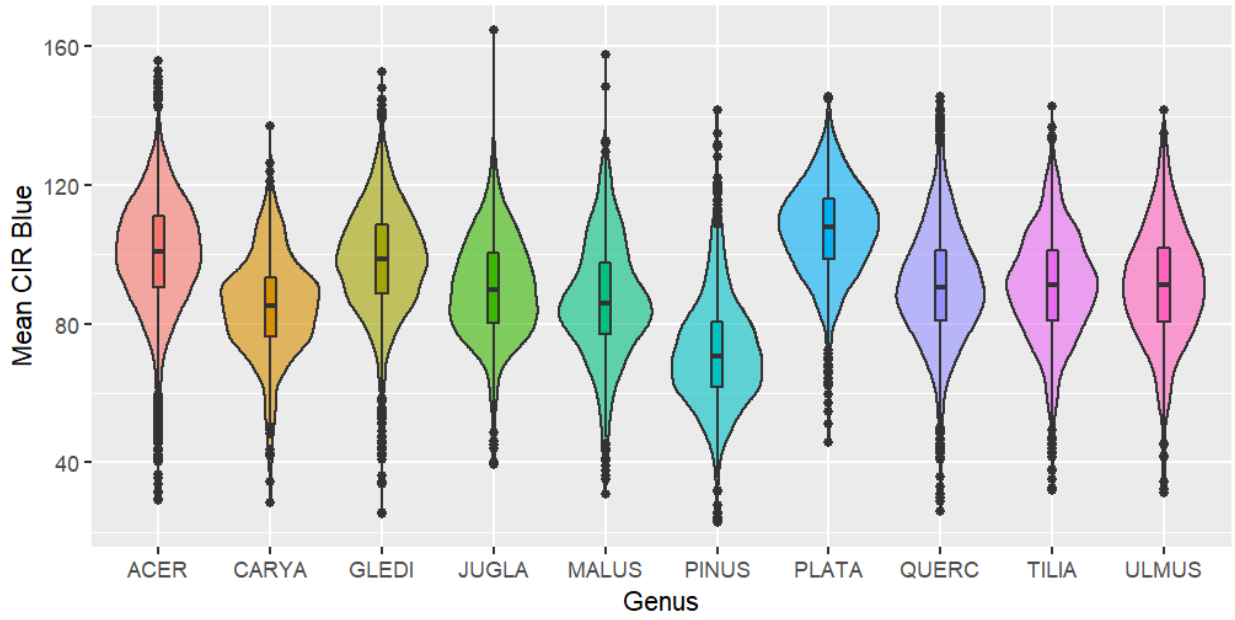



**Figure F.7.** Violin plot of mean blue CIR band by genus.

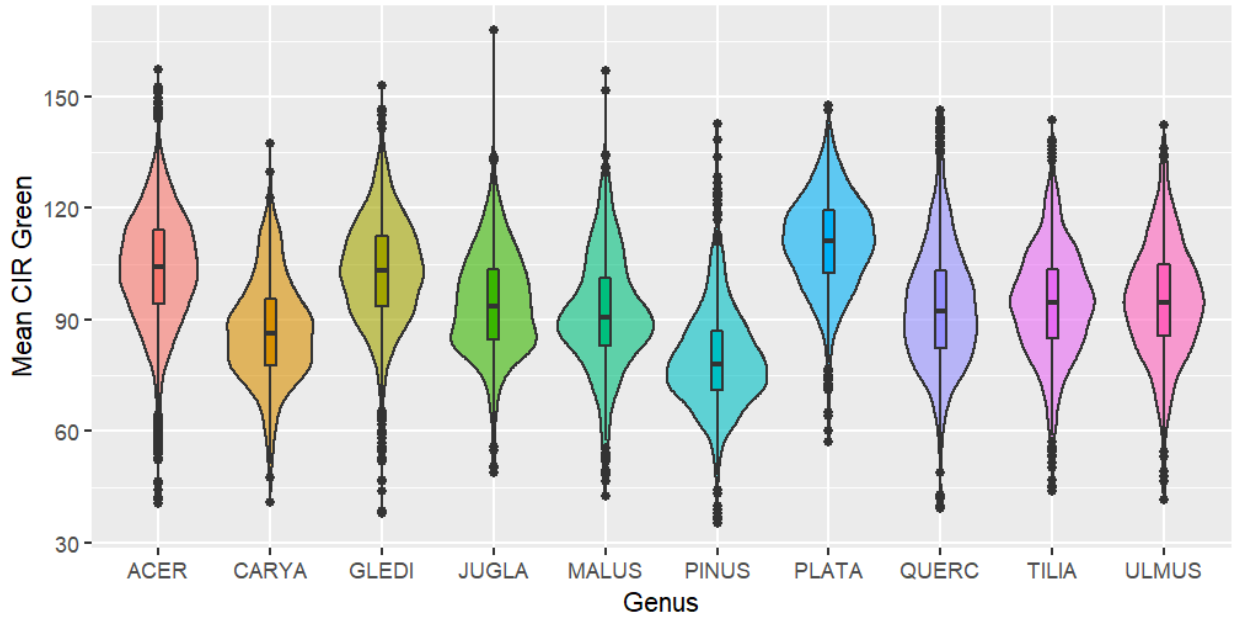



**Figure F.8.** Violin plot of mean CIR green band by genus.

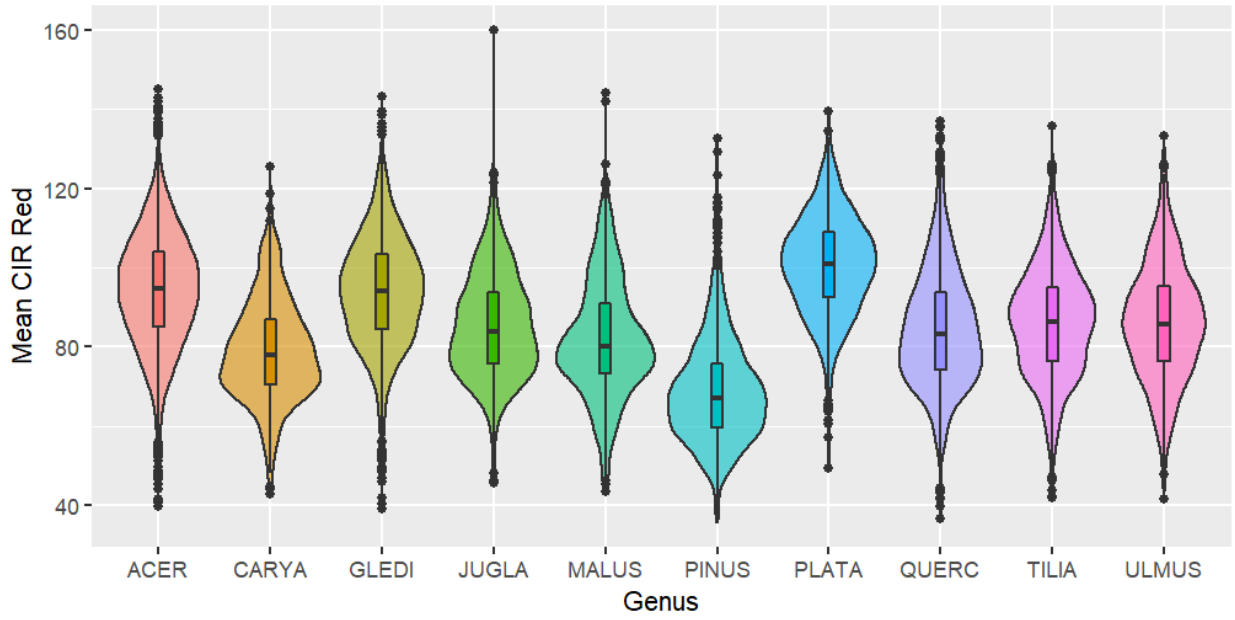



**Figure F.9.** Violin plot of mean CIR red band by genus.

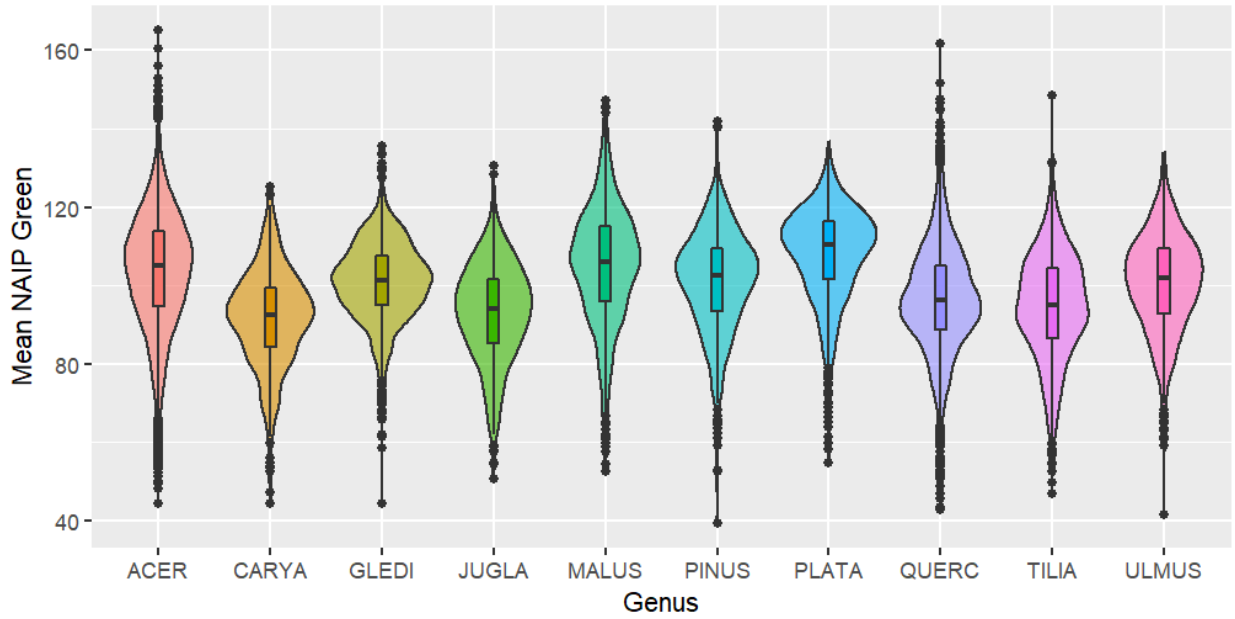



**Figure F.10.** Violin plot of mean NAIP green band by genus.

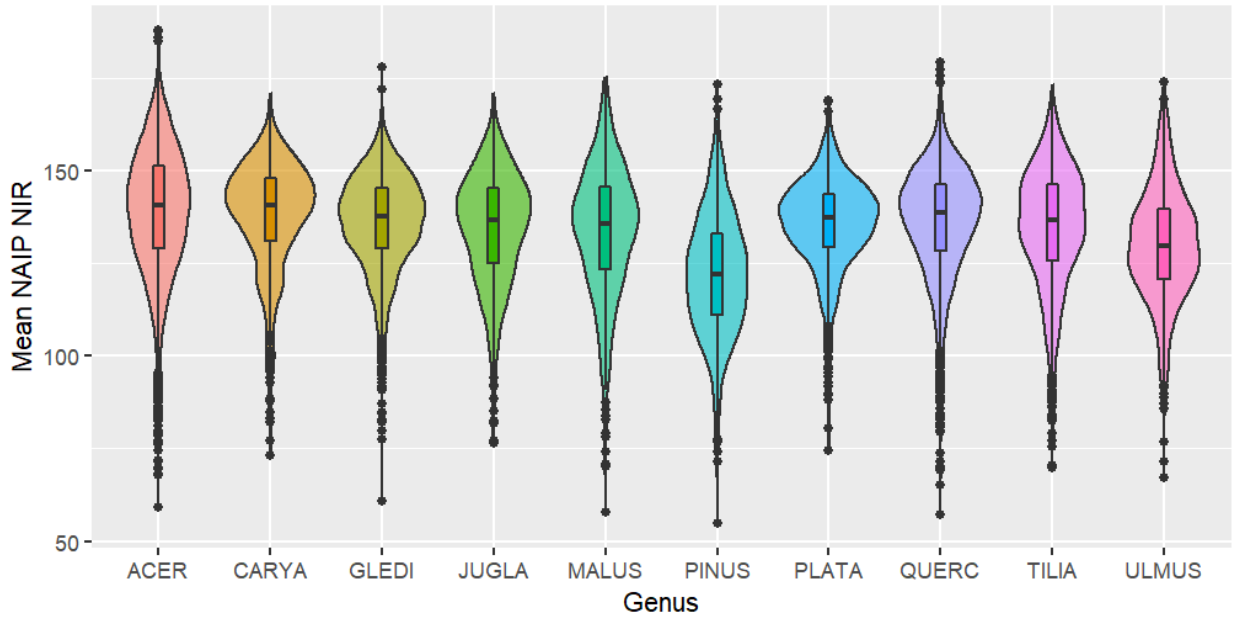



**Figure F.11.** Violin plot of mean NAIP NIR band by genus.

## Appendix G: Machine Learning R Script

This R script takes in a .csv file containing labeled trees and predictor variables. It divides this table into training and testing data, feeds the training data into a random forest, SVM, and multinomial regression model and automatically outputs summary files with confusion matrices for each model.

```
##############################
# AAUTC Machine Learning Script
# Author: Thomas Estabrook
# 1/17/2022
# Updated: 3/21/2022
##############################

library(sf)
library(randomForest)
library(tidyverse)
library(caret)
library(nnet)
library(e1071)

# Next code chunks can be commented/uncommented based on which segmentation
type is used

# seg_method <- "pixel"       # segmentation method (e.g. LiDAR, pixel,
buffered pixel)
# location <- "A2"           # location (e.g. BPW, full_city)
# genus_threshold <- 200       # number below which genera are lumped into
"other" category
# today <- Sys.Date()         # current date
# set.seed(1)                 # set a random seed for reproducibility
# setwd('G:/Shared drives/A2_UTC Drive/R_scripts/Model_Inputs')
# set working directory
#
# # Load in the csv with tree data
# # Should have a column with Genus, and columns for each predictor variable
# trees <- read.csv("A2_trees_single_pixel3_pca.csv") %>%
#   mutate(tree_id = row_number()) %>% # add a new unique ID for the dataset
#   drop_na() %>%
#   filter(Genus != "")
#
# trees$Soils <- as.factor(trees$Soils)
#
# head(trees)
# colnames(trees)
#
# #Use this to list all columns not being used as predictor variables
# dropcols <- c("ï..OID_", "tree_id", "Plnt_Cd", "Cmmn_Nm", "Source",
"Prjct_C", "DBH_ft_cm_2")
```

```
# seg_method <- "Buffpix"       # segmentation method (e.g. LiDAR, pixel,
buffered pixel)
# location <- "A2"           # location (e.g. BPW, full_city)
# genus_threshold <- 100       # number below which genera are lumped into
"other" category
# today <- Sys.Date()         # current date
# set.seed(1)                 # set a random seed for reproducibility
# setwd('G:/Shared drives/A2_UTC Drive/R_scripts/Model_Inputs')
# set working directory
#
# # Load in the csv with tree data
# # Should have a column with Genus, and columns for each predictor variable
#
# trees <-
read.csv("trainingtrees_master_1006_cmDBH_shrubs_Buffer_MeanBand.csv") %>%
#   mutate(tree_id = row_number()) %>% # add a new unique ID for the dataset
#   drop_na() %>%
#   filter(Genus != "") %>%
#   left_join(area)
#
#
# head(trees)
# colnames(trees)
#
# #Use this to list all columns not being used as predictor variables
# dropcols <- c("i..OID_", "Plnt_Cd", "Cmmn_Nm", "Source", "Prjct_C",
"ORIG_FID", "tree_id", "DBH_ft_cm", "BUFF_DIST", "Shape_Length",
"Shape_Area")

seg_method <- "lidR30_pca2"       # segmentation method (e.g. LiDAR, pixel,
buffered pixel)
location <- "A2"           # location (e.g. BPW, full_city)
genus_threshold <- 100       # number below which genera are lumped into
"other" category
today <- Sys.Date()         # current date
set.seed(1)                 # set a random seed for reproducibility
setwd('G:/Shared drives/A2_UTC Drive/R_scripts/Model_Inputs')
# set working directory

# Load in the csv with tree data
# Should have a column with Genus, and columns for each predictor variable
#PCs <- read.csv("G:/Shared drives/A2_UTC
Drive/R_scripts/LiDAR_PCA/A2_LidR_seg_30_PCs.csv")

trees <- read.csv("A2_LidR_seg_30_finalZS_soil_tex_area_PCs.csv") %>%
  mutate(tree_id = row_number()) %>% # add a new unique ID for the dataset
  drop_na() %>%
  filter(Genus != "")
```

```
head(trees)
colnames(trees)

#Use this to list all columns not being used as predictor variables
dropcols <- c("X", "X.1", "X.2", "Id", "MAJORITY", "MINORITY", "VARIETY",
"tree_id")
trees$SOIL <- as.factor(trees$SOIL)

################################################################################
##############################
# Get counts of tree genera
genuscounts <- trees %>%
  group_by(Genus) %>%
  summarise(count = n())

# Filter out underrepresented genuses and recode to "Other"
other_genera <- genuscounts$Genus[genuscounts$count < genus_threshold]
trees$Genus[trees$Genus %in% other_genera] <- "Other"

# Randomly take half of each genus for training and testing
training_trees <- trees %>%
  group_by(Genus) %>%
  slice_sample(prop = 0.5)

testing_trees <- trees[!(trees$tree_id %in% training_trees$tree_id),]

# Remove unneeded columns
training_trees <- training_trees %>%
  select(!all_of(dropcols))

testing_trees <- testing_trees %>%
  select(!all_of(dropcols))

# Turn the Genus column into a factor
training_trees$Genus <- as.factor(training_trees$Genus)
testing_trees$Genus <- as.factor(testing_trees$Genus)

# Run a random forest model
rf_mod <- randomForest(Genus ~ ., data = training_trees,
                       xtest = testing_trees[,!names(testing_trees) ==
"Genus"],
                       ytest = testing_trees$Genus,
                       keep.forest=TRUE)
rf_pred <- rf_mod$test$predicted
rf_conf <- confusionMatrix(data = rf_pred, reference = testing_trees$Genus)

# Run a SVR model
svm_mod <- svm(Genus ~ ., data = training_trees)
```

```
svm_pred <- predict(svm_mod, testing_trees)
svm_conf <- confusionMatrix(data = svm_pred, reference = testing_trees$Genus)

# Run multinomial logistic regression
reg_mod <- multinom(Genus ~ ., data = training_trees, MaxNWts =10000000,
maxit = 2000, na.action = "na.fail")
reg_pred <- predict(reg_mod, testing_trees)
reg_conf <- confusionMatrix(data = reg_pred, reference = testing_trees$Genus)
reg_conf

# Save outputs
setwd('G:/Shared drives/A2_UTC Drive/R_scripts/Model_Outputs')
filename <- paste0(seg_method, "_", location, "_", genus_threshold, "_",
today)
save(rf_mod, file = paste0("rf_mod_", filename))
save(svm_mod, file = paste0("svm_mod_", filename))
save(reg_mod, file = paste0("reg_mod_", filename))

sink(file = paste0("summary_", filename, ".txt"))
print(paste("Number of genera:", length(levels(training_trees$Genus))))
print(paste("Number of trees:", nrow(trees)))

print("Random Forest:")
rf_conf$overall[1:2]
print("Support Vector:")
svm_conf$overall[1:2]
print("Multinomial Regression:")
reg_conf$overall[1:2]

print("Random Forest:")
rf_conf

print("Support Vector:")
svm_conf

print("Multinomial Regression:")
reg_conf
sink(file=NULL)
```

## Appendix H: Tree Crown Segmentation R Script

This R script takes as input a folder with tiled LiDAR pointclouds and outputs a segmented raster with each tree crown assigned a unique ID.

```
library(lidR)
library(raster)

f <- function(x) {  # Taken from https://jean-romain.github.io/lidRbook/itd-
its.html
  y <- 20.1 * (-(exp(-0.08*(x-2)) - 1)) + 10 # altered for feet (state plane)
  y[x < 6] <- 10 #10 feet for less than 6 feet tall
  y[x > 60] <- 30 # 30 feet for greater than 60 feet tall
  return(y)
}

for(i in seq(0:8)){
  setwd("G:/Shared drives/A2_UTC Drive/R_scripts/CHM_full_city_mask/")
  double_name <- paste0("CHM_full_city_mask.", i-1,'_double.tif')
  A2_CHM <- raster(double_name)

  gc()
  ttops30 <- find_trees(A2_CHM, lmf(30))

  # par(mfrow=c(1,1))
  # plot(A2_CHM, col = height.colors(50))
  # plot(ttops30, add = TRUE)


  algo30 <- silva2016(A2_CHM, ttops30)

  setwd("G:/Shared drives/A2_UTC
Drive/R_scripts/CHM_full_city_mask/Unmerged_segs")

  seg30 <- algo30()
  # plot(seg30, col = pastel.colors(5000))
  writeRaster(seg30, paste0("A2_LidR_seg_30_", i-1, ".tif"))
  rm(seg30)
}
```

## Appendix I: Trees Found in Unsupervised Clusters

In an effort to characterize the clusters resulting from the unsupervised clustering algorithm described in Chapter 2, we used the *Extract Multi-values to Points* tool in ArcGIS Pro to acquire the cluster number of the pixel overlapping each of the trees in our training dataset. The following plots display the results for the 25 most common genera in the dataset. It should be noted that because the training dataset is not a truly random sample of the city's canopy trees, the results here may be biased and should be consulted for exploratory purposes only.

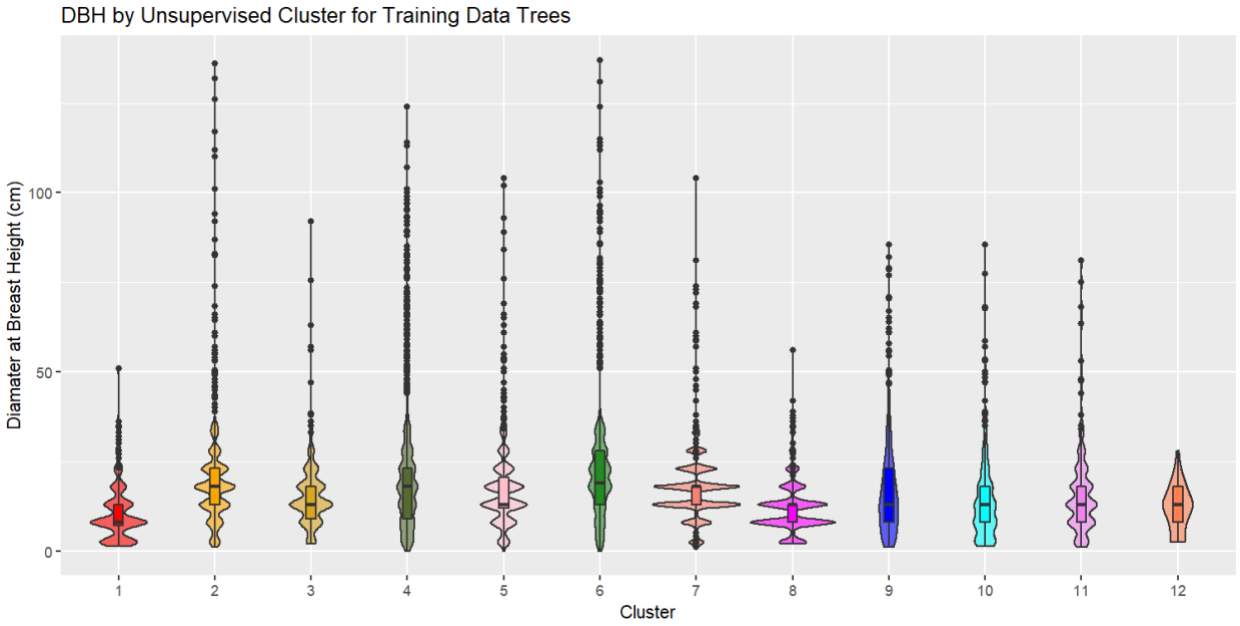



**Figure I.1.** DBH violin plot for training data trees found in each cluster. Clusters 4, 6, and 2 have the highest average DBH. "Wavy" appearance of density curves is due to the way some datasets round DBH in 2.5 cm increments.

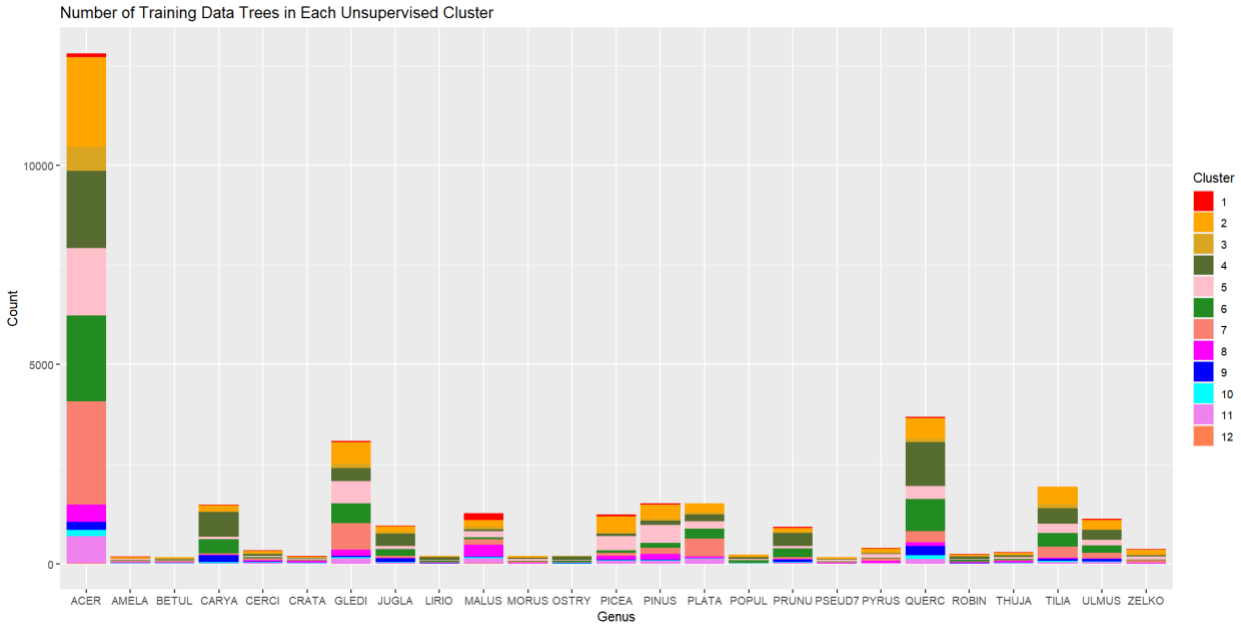



**Figure I.2.** Count of trees by genus found in each cluster.

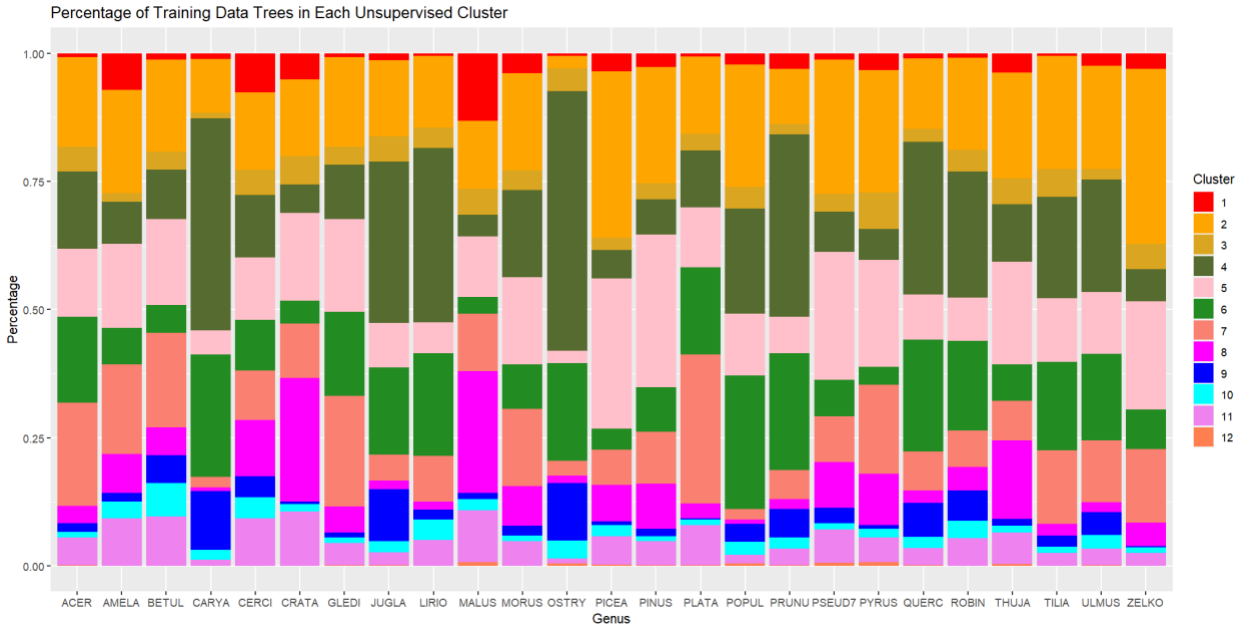



**Figure I.3.** Percentage of each genus falling into each cluster.

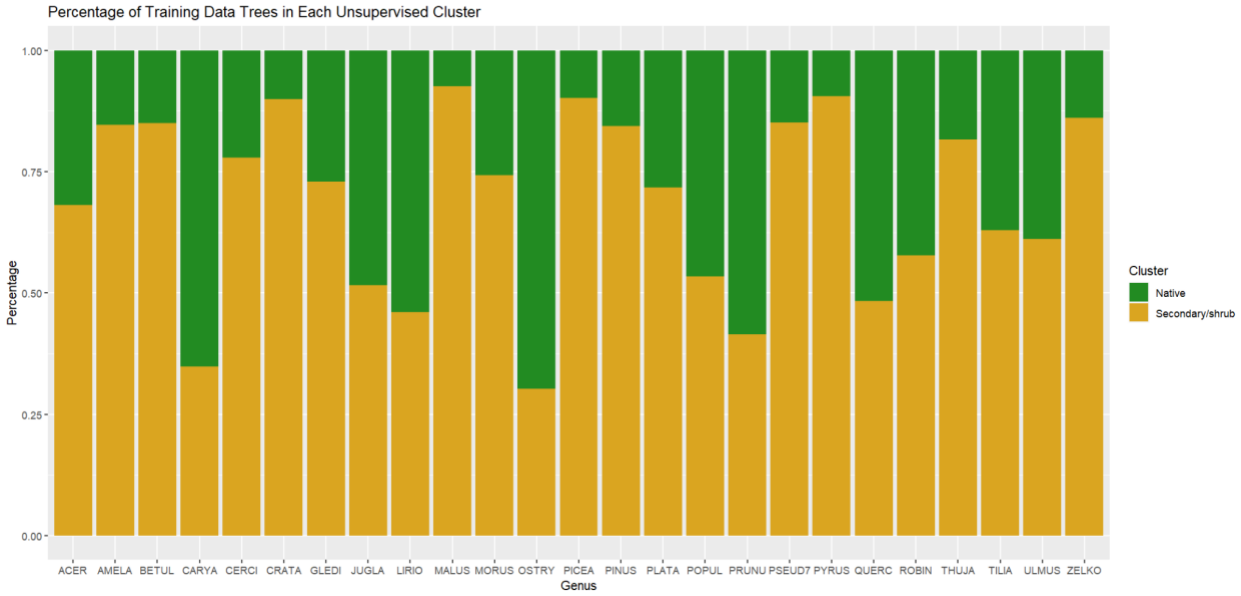



**Figure I.3.** Percentage of each genus falling into clusters classed as native vs. those classed as secondary/shrub. We can observe that hophornbeams (OSTRY), hickories (CARYA), black cherries (PRUNU), oaks (QUERC), walnuts (JUGLA), and poplars (LIRIO) are well-represented in the clusters identified as native.