Using Drone Based LiDAR to Reconstruct Forests

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Daniel K. Spiwak
Bachelor of Science
George Mason University, 2019

Director: Konrad Wessels, Associate Professor
Geography and Geoinformation Science

Summer Semester 2021
George Mason University
Fairfax, VA

DEDICATION

This work which I have pursued for the past 4 years is dedicated to my loving parents and family John, Piper, Emily Spiwak, and Evan Gass; who nurtured my love for learning and exploring the world through adventures and scholarship. My grandparents Kenneth and Peggy Ward; who sparked my love for the environment and taught me to appreciate family and the lesser beauties in life. My relatives Dale and Laurie Ward; who helped me find my path in life and inspired me to always learn more from the smaller majority which surrounds us every day.

ACKNOWLEDGEMENTS

I would like to thank the Smithsonian Conservation Biology Institute for their gracious support throughout my undergraduate and graduate studies. I would also like to acknowledge my friends and colleagues Justin Elarde, Colin Flynn, Shawn Li, and James Valentine who have supported me by going over programming logic and providing insights. Finally, I would like to thank my mentors Dr. Konrad Wessels, Dr. Qiongyu Huang, and Patricia Dietly for encouraging me to pursue academia and to always challenge myself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND SYMBOLS

Above Ground Biomass ..... AGB
Above Ground Level ..... AGL
Aerial Observation Platform ..... AOP
Airborne Laser Scanning ..... ALS
Basal Area ..... BA
Canopy Base Height ..... CBH
Canopy Height Distribution ..... CHD
Canopy Height Model ..... CHM
Canopy Volume Profile ..... CVP
Centimeter ..... cm
Cloud Compare CC
Cursory Woody Debris ..... CWD
Density-Based Spatial Clustering of Applications with Noise ..... DBSCAN
Diameter at Breast Height ..... DBH
Differential GPS ..... DGPS
Digital Elevation Model ..... DEM
Digital Surface Model ..... DSM
Digital Terrain Model ..... DTM
Epsilon ..... ε
Federal Aviation Administration ..... FAA
ForestGEO ..... FGEO
George Mason University ..... GMU
Global Ecosystem Dynamics Investigation ..... GEDI
Global Positioning System ..... GPS
Goddard LiDAR, Hyperspectral and Thermal Imager ..... G-LiHT
Inertial Measurement Unit ..... IMU
International Space Station ..... ISS
K-Nearest Neighbor ..... KNN
Land, Vegetation, and Ice Sensor ..... LVIS
Leaf Area Index ..... LAI
Light Detection and Ranging ..... LiDAR
Meter ..... m
Millimeter ..... mm
Nanometer ..... nm
National Aeronautics and Space Administration ..... NASA
National Ecological Observatory Network ..... NEON
Post-Processing Kinematic ..... PPK
Random Sample Consensus ..... RANSAC
Root Mean Squared Error ..... RMSE
Smithsonian Conservation Biology Institute ..... SCBI

Terrestrial Laser Scanning ..... TLS
Unmanned Aircraft System.....UAS
Unmanned Laser Scanning ..... ULS
Visual-line-of-sight ..... VLOS
Mobile Laser Scanning ..... MLS

ABSTRACT

USING DRONE BASED LIDAR TO RECONSTRUCT FORESTS

Daniel K. Spiwak, M.S.

George Mason University, 2021

Thesis Director: Dr. Konrad Wessels

Forests are an essential ecosystem for the sequestration of CO_2, the dominant greenhouse gas driving climate change, in the environment. The ability to accurately determine the amount of carbon stock and sequestration within this system through biomass estimation is crucial to informing carbon budgets, carbon offset projects, and commercial forestry. However, national, and regional biomass models rely heavily on laborious stand-level, typically field-derived, metrics such as Diameter at Breast Height (DBH) of individual trees, that are then scaled up via models with satellite imagery.

To facilitate easier biomass estimation, this study employed the use of a small-footprint Light Detection and Ranging (LiDAR) sensor, a Small Unmanned Aircraft System (SUAS), and advanced LiDAR point cloud processing to extract and estimate the DBH of individual stems at a well-studied ForestGEO site (12.5ha.) in Virginia. Unmanned aerial vehicle borne Laser Scanning (ULS), as performed here, can significantly improve stand-

level biomass estimates which can then be used to develop empirical models that predict regional biomass using satellite imagery.

Our specific objectives were to (i) assess the ability to automatically detect and extract individual tree stems using Density Based Spatial Clustering of Applications with Noise (DBSCAN), and more prominently (ii) test the accuracy of four DBH estimation methods adopted from Terrestrial-Laser-Scanning (TLS) and Airborne-Laser-Scanning (ALS) at stand-level scales. The DBH estimation methods assessed were (i) Convex Hull approach (CH), (ii) Pratt (Pt) and (iii) Levenberg-Marquardt (LM) circle fitting, and (iv) Random Sample Consensus (RANSAC).

We demonstrated that through DBSCAN, individual stems larger than 18cm DBH could be detected across the full study area with an accuracy of 65%. Estimation bias was the lowest in small stems ranging from 10-50cm (67% of the known stems); where all DBH estimation methods displayed a relationship of increasing negative bias (underestimation) for progressively larger stems. For stems approximately 10-20cm DBH, LM and RANSAC had a positive bias of 1.6 and 3.8cm, which turned negative and increased to -10.7 and -9.3cm for stems 40-50cm DBH. Pt failed to reconstruct small stems 10-20cm DBH with an initial bias of 14.2cm which then decreased to -0.2cm at 40-50cm DBH. CH similarly failed to reconstruct small stems but had the smallest overall range in bias across the 10-50cm DBH interval of 10.8-4cm.

Due to errors in co-location between the ForestGEO data and the ULS point cloud, initial R^2 values were low for the full study area with the highest being .17 for LM, followed by .16, .06, and .04 for CH, RANSAC, and Pt, respectively. Limiting the analysis to only

high-confidence matches between in-situ and ULS clusters drastically improved R^2 to .69, .71, .23, and .13 for LM, CH, RANSAC, and Pt correspondingly. This underscores the importance of reliably aligning the two datasets before analyses.

With these findings, this study hopes to pave the way for ULS DBH estimation for individual stems and provide a significant contribution towards the improvement of non-destructive biomass estimation. Through this study and its successors, rapid stand-level metrics will be attainable from UAS LiDAR and could supplement regional satellite and ALS-based biomass estimates.

CHAPTER ONE: INTRODUCTION

In 2020, 31% of the Earth's surface was covered by forests and contributed to over 45% of terrestrial carbon stocks [1]. Holistically, forests are centers for all variations of ecosystem services. Forests not only regulate regional climates, sequester large amounts of carbon as biomass, and act as ecological support interdependent ecosystems, they also serve as provisional, and cultural centers for humans [2]–[4].

To estimate the effectiveness and relative health of these ecosystems, forestry relies upon biophysical metrics such as Diameter at Breast Height (DBH), Stem Density, Basal Area (BA), Above Ground Biomass (AGB), and the distribution of species and canopy composition [2], [5], [6]. Traditionally, the computation of these metrics has been performed by collecting laborious in situ measurements, which frequently require years to collate data into reliable and publishable forms [2], [6]–[8]. However LiDAR, when integrated with different Earth Observation (EO) platforms such as satellites or aircraft, can collect spatially expansive data, not logistically feasible through traditional methods [6], [9]–[11]. Previous studies have demonstrated the ability to extract forest structure metrics using Terrestrial Laser Scanning (TLS) [9], [10], [12]–[17], Airborne Laser Scanning (ALS) [18]–[22], Satellite-based remote sensing [4], [23], and more recently ULS [10], [17], [24]–[30]. Therefore, we propose to collect and extract plot-level forest

metrics using a LiDAR capable SUAS and compare these findings with field derived measurements.

Due to weather, terrain, and operational costs, ALS cannot be flown low enough, or with the frequency needed to achieve the point density and temporal resolutions obtainable through terrestrial LiDAR [18], [31]. Similarly, TLS is hindered by the logistics, terrain navigability, and time requirements needed to sufficiently survey, non-occluded, mergeable scans [9], [32], [33]. With these limiting factors in mind, UAS fill a unique niche between ALS, and TLS by being able to operate low enough to collect dense data, while simultaneously being mobile enough to cover local landscape and forest variability. However, major challenges, which this study hopes to address, are related to the complexity of automatic extraction of forest structure metrics from the very dense ULS point cloud data.

Aim

The aim of this study is to extend and compare methods previously applied to TLS and ALS data to extract forest metrics from dense ULS point cloud data and assess results using a well-documented forest plot at the Smithsonian Conservation Biology Institute's (SCBI) Forest Global Earth Observatory (ForestGEO) site in Front Royal, Virginia. The primary objective of this study being the evaluation of various methods for extracting "diameter at breast height" (DBH) from leaf-off ULS data and compare this to the DBH field data from the ForestGEO plot. DBH is routinely used in allometric equations to calculate tree and plot-level the Above Ground Biomass (AGB) to estimate

carbon stock and monitor forest structure [2], [8], [34]. While most studies in this field use LiDAR data collected during leaf-on period to estimate canopy structure [13], [16], [24], [27], [35], [36], we opted for using leaf-off ULS data to facilitate effective stem structure detection. We are ultimately striving to develop methods to accurately estimate DBH from which ABG can be calculated for an entire forest plot surveyed by means of ULS.

Specific Objectives

1. Assess the accuracy of DBSCAN clustering on ULS point cloud data to detect individual trees stems with >18cm DBH, by comparing it to matched field data.
2. Compare the following DBH extraction techniques to the DBH field data to determine the most accurate approach:
    a. Convex Hull Estimation
    b. Circle Fitting
        i. Algebraic Approach
        ii. Geometric Approach
    c. RANSAC Primitive Fitting

CHAPTER TWO: LITERATURE REVIEW

LiDAR data Collection Techniques

Terrestrial Based LiDAR

One of the most basic form of LiDAR scanning, Terrestrial Laser Scanning (TLS), is a non-mobile form of remote sensing. Ecologists and field surveyors have previously adopted TLS to capture 3D subcanopy datasets which facilitate the extraction of important biophysical metrics [9], [12], [32].

Higher scan resolutions are achievable through TLS when compared to non-stationary techniques, but the lack of movement and extended time requirements during scanning results in a smaller area of coverage. Moreover, objects scanned using TLS are highly susceptible to occlusion, and require multiple scans and "set-ups" to be performed to fully capture them [9], [32]. To mitigate the shadowing affect introduced by TLS, Stovall et al. adopted the multi-scan methodology and performed five scans around a "core" point with great success. However, the issue of coregistration between multiple scans is another limiting factor hindering forestry applications due to the amount of effort involved in setting up visible "backsights" in dense understory. Without quantifiable offsets, accurately aligning point clouds with minimal error becomes a significant challenge.

A depiction of the occlusion phenomena which occurs during single scans is shown in Figure 1. Note, the specific dataset depicted in Figure 1 was produced by displaying data along the periphery of a ULS flight line. The occlusion depicted is

resolved through subsequent overlapping flight lines, or in the case of TLS, by performing multiple scans throughout the target area.

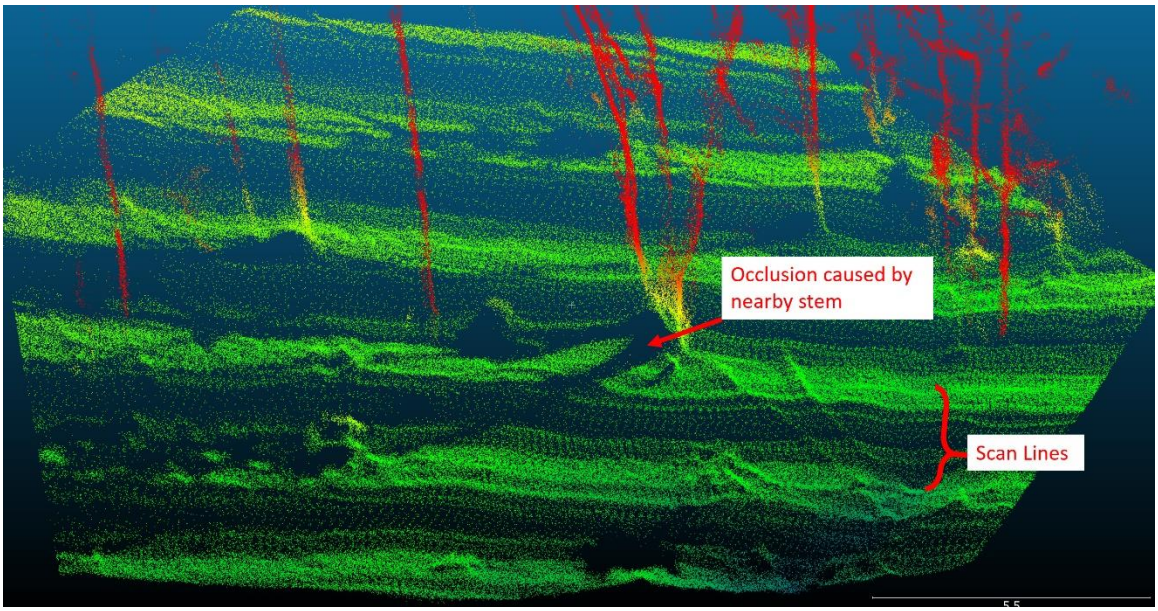

Figure 1: Terrain and stem occlusion which would occur in single-scan-based TLS data. Note, the data shown here was acquired using a ULS system, however a similar occlusion phenomenon can be shown with a singular flight line as depicted on its' periphery.

As noted by GOFC-GOLD, monetary and time requirements should be considered when selecting a forest inventorying regime [34]. TLS is the cheapest form of laser scanning, monetarily, but the immobility of the method scales time requirements drastically when the area of interest increases.

As an extension of TLS, Mobile Laser Scanning (MLS) has been to circumnavigate the need to for "re-setups" of the TLS sensor between scanning positions [37]–[39]. MLS addresses the lack of sensor mobility, whilst simultaneously reducing operational costs [13], [39]. MLS, similar to ULS and ALS, requires the use of an

integrated GPS and Inertial Measurement Unit (IMU) to determine the relative position and orientation of the system during data collection. In forest scenes, if canopy closure is a significant issue and occludes the underlying terrain or subcanopy, TLS or MLS can be used in conjunction with ALS or other non-ground-based EO to extract middle to upper canopy features that TLS could not observe [19].

The revisit interval or temporality of TLS regimes are also important to consider as it allows investigators to track the ecological succession of an area of interest [3], [31]. TLS requires comparatively less effort to collect than a detailed survey by hand, and as a result, can have higher temporal resolutions. Primarily limited by weather and the navigability of the terrain, TLS is applicable in most forestry scenarios. Compared to other EO platforms that are at altitude, TLS is not significantly impacted by atmospheric conditions (clouds), other than precipitation, thus potentially daily revisits can be achieved.

Drone Based LiDAR

The growth of commercially available and compact, small-footprint LiDAR sensors, as well as remotely operated vehicles, has increased the prevalence of these systems in forestry, and other ecology-related fields [40], [41]. ULS provide a “middle ground” in EO with resolutions comparable to those achievable through TLS, but with enough mobility to investigate local areas of interest. In addition to the increased mobility over TLS, drones allow for a significant reduction in the scene occlusion by allowing greater amount of “side shots” to be collected through wide scan angles (similar to the

Field-of-View) of $\pm 30°$ without significantly hindering LiDAR return accuracy [25], [27].

Sensors, such as the Ibeo LUX, Velodyne Puck 16, Quanergy M8 Ultra, and the Riegel series of scanners are lightweight and robust enough to withstand the vibrations encountered during SUAS flight. These sensors are adept at detecting vegetation and typically scan within the IR wavelengths [26], [42]–[44]. Due to the relatively low wattage of small footprint sensors, unlike the more powerful airborne and spaceborne sensors, ULS may struggle to penetrate the canopy of densely vegetated forests, thus hindering the collection of dense ground points for generation of terrain models under dense forest canopies. This prompted our decision to collect data for sub-canopy metrics during leaf-off conditions. ULS pulse footprint sizes are similar to distant TLS pulses and are comparable to high resolution airborne systems at ~10cm, but can be upwards of 25cm depending on the flight altitude or sensor specifications [42]. Similar to TLS, SUAS-borne LiDAR is cheaper to operate than their airborne and spaceborne counterparts. An additional cost which should be considered when selecting a ULS system, is the vehicle/ platform itself, as larger platforms must adhere to more stringent regulations, and may not provide larger ranges [26], [29], [45].

Airborne Based LiDAR

Compared to ULS, airborne laser scanning (ALS) can cover a much larger area of interest that are limited only by operational cost, aircraft's range and data storage limits. Although ALS can cover a greater area, the point density throughout the scan is typically

reduced due to the minimum flight altitude and speed of the aircraft. The point density of ALS of 10-50 points per m^2, is lower than ULS, reducing its ability to accurately detect tree trunk compared to TLS or ULS [18], [46]. Current examples of ALS sensors include the National Aeronautics and Space Administration's (NASA) Goddard LiDAR, Hyperspectral and Thermal Imager (G-LiHT) and Land, Vegetation, and Ice Sensor (LVIS), and the National Ecological Observatory Network (NEON) Aerial Observation Platform (AOP) [21], [47].

The footprint size of most airborne systems' pulse at altitude varies greatly between sensor but is comparable to the less powerful ULS sensors. G-LiHT for example has a nominal footprint of ~10cm, while operating at an altitude of 600m AGL [46], while LVIS has a footprint of ~10m when operating at altitude of 28,000 feet AGL [47], [48].

Operation costs for ALS sensors is proportional to the duration and frequency of the flights. NEON currently intends to bi-annually collect ALS data over ~60 sites for a 30-year duration. Due to the high operating costs, ALS sensors typically undergo more extensive calibration compared to commercially available TLS and ULS sensors. Unlike spaceborne platforms, ALS can have frequent maintenance routines to mitigate erroneous datasets from occurring and can allow for upgrades in the sensors systems should technology advance or faults occur.

Feature Extraction Techniques

This section introduces the various methods used to extract objects and metrics from LiDAR point cloud data.

Hyperplane 'slicing'

Although multiple definitions and types of hyperplane extraction exist, for the purpose of this thesis, hyperplane extraction shall be defined as any method which extracts a subset of points in a point cloud at a fixed height above ground for the purpose of extracting a metric from the subset. Colloquially known as slicing, this technique has been adopted by many algorithms to detect connected components[9], [18], [49], perform general segmentation and object identification [43], and yield metrics for the characterization of vertically gridded structures [42]. Examples of canopy related metrics derivable from a slicing regime include Canopy Base Height (CBH), Canopy Height Distribution (CHD), and Canopy Volume Profile (CVP).

At its core, slicing techniques rely on the creation of a canopy height model (CHM) to normalize the height values for each of the points throughout the point cloud to a nominal above ground height. After the generation of the CHM, slice(s) of variable or set widths can be created to extract all points present within the slice extents. Although all scanning techniques such as TLS, ALS, and SLS, can potentially employ slicing, TLS and SLS are limited in the amount of "side-shots" being collected within the point cloud. Whereas SLS is primarily limited by the resolution of the sensor to capture enough datapoints to warrant a slicing regime to extract subcanopy metrics [23], TLS is primarily

limited by the significant occlusion which takes place throughout the scanned area[9], [32].

Compared to other data extraction techniques, slicing can be considered the simplest to employ as it does not require overly complicated algorithms to extract the subsets of points. Furthermore, there is a wide variety of summarization methods for the extracted slice and can be employed using different scanning methods. Slicing is also invariant to the types of forest structure with which it can be employed, although the density of the sliced cloud are highly dependent on the quality of the scan [43].

Voxelization

For brevity, voxelization serves as a method of data reduction whereby the coordinate precision is rounded to fulfill a desired cell size [31], [50]. Voxels exist as various kernels, typically described as cuboid, rook, or queen in shape, and can be used to address noise or fill gaps in data [50]. In practice voxelization has been used as a precursory step for connected component analysis or density based metric extraction [22], [31], [50].

Density-based metrics such as the aforementioned CVP, or CBH, can readily adopt voxelization as a computationally efficient means of expansive data reduction, permitting the calculation of cell density [31]. Although performing data compression, this method is lossy and cannot be easily reversed.

Cylinder Fitting

Considered to be a form of full scene reconstruction, cylinder fitting is computationally more expensive than averaging coordinate values using voxelization or

performing dimensionality reduction to assess a general trend. Cylinder fitting is extremely powerful for accurately estimating biomass based on volumetric assessments of the trees within a point cloud, with ~5-15% error [9], [32], [51]. As with voxelization, cylinder fitting struggles with occluded tree stems as the diameter of an individual tree stem typically tapers nearing the crown, requiring the fitted primitive to also taper accordingly [51], [32]. Kelbe et al. note that the cylinder fitting process overestimates and underestimates the trees curvature when the height of the cylinder is increasingly extruded. Interestingly, voxelization is typically one of the initial steps during cylinder fitting as this allows for indexing of the objects, and performs data reduction to aid in computation [32]. After voxelization, or initial data reduction, a feature extraction algorithm such as RANSAC, or the Hough transform is used to convert the points along the stem into boundary lines, within which the cylinders are then fitted [18], [32], [51].

Cylinder fitting has been limited to TLS as it requires a significant number of points along the stem to accurately model the individual tree. Recently, a form of cylinder fitting algorithm (named aTrunk) was adapted to ALS by means of separating the base of the canopy from the relatively isolated stem in the trunk segmentation algorithm [18]. Both the classical form of cylinder fitting, and aTrunk rely on Dijkstra's shortest path method to discern individual trunks from one another [18], [32], [51].

General Review

As previously discussed, cheaper, more accessible, and increasingly accurate LiDAR systems have led to an increase in the use of LiDAR in forestry studies [32], [38],

[52], [53]. Among others, Kelbe et al.,have shown that forest reconstructions can be performed using lower quality, ‘off-the-shelf’, sensors which have a beam divergence of ~15mrad, or approximately 1.5m at 100m distance from the scan location [32]. Moreover, inventories and metric extraction can be performed at relatively low point densities [25].

Liu et al. recently performed a study using MLS to extract stem DBH that adopted a similar methodology performed herein. RANSAC was used as the method of DBH estimation to determine the best fit circle within point cloud sub spaces. Unique among studies was that Liu et al. extracted multiple “sub-hyperplanes” within the greater “parent” slice and recalculated a new fitted circle within each sub-plane to mitigate potential outliers [38]. As noise was a prime concern in our study, the approach proposed by Liu et al. served as a basis in our study which used less dense ULS data.

Liu et al. also proposed a method of “relative density segmentation”, which was capable of detecting 140/148 (94%) of trees in a natural stand, whereas the “traditional” hierarchical method used as a basis for control only detected 94/148 (63%). Furthermore, Liu et al.’s implementation of iterative heights of RANSAC circles, using their clustering method, had a precision of 0.967, whilst the hierarchical clustering method had a precision of 0.633. The DBH RMSE for these methods were 3.17cm, and 3.88cm, respectively.

Until recently [16], [17], [24], studies incorporating UAS LiDAR primarily focused on assessing the ability to model easily predictable, well organized stems, such those in plantations, or other agricultural landscapes [25], [29], [30], [33]. A limited

number of the studies within this field assess extents larger than 5 hectares, at times surveying fewer than 100 trees, which may limit the studies' applicability to natural, variable forest landscapes [9], [33], [54]. Stovall et al. noted that upwards of 200 samples are needed to sufficiently generate an allometric relationship between structural parameters such as DBH or height and biomass. However, by combining information derived from multiple metrics such as CBH,CW, or DBH, similar to ensemble model simulation, the number of necessary samples needed to generate a relationship can be significantly reduced [9].

Regarding the effectiveness of the aforementioned DBH extraction techniques, they are foremost limited by a study's ability to successfully discretize individual stems from a raw LiDAR point cloud. Stem segmentation methods can be broadly dependent upon the collection platform, data quality, or even seasonality of the data. Currently, common segmentation approaches include Random Sample Consensus (RANSAC), MeanShift, Density-based Spatial Clustering of Applications with Noise (DBSCAN), and Marker Controlled Watershed segmentation [18]–[20], [55]. The present study selected DBSCAN as the preferred clustering approach due to its' widespread use in clustering problems with significant noise, and the ease through which it could be implemented or adapted to new sets of data.

When considering the commonly used DBH extraction techniques for this study, Mean Gravity Distance, a simple yet effective approach was an initial consideration. Mean Gravity Distance relies upon the calculation of a stem's gravity center, commonly referred to as the Centroid, and the average distance to all LiDAR points [13], [33].

However, Mean Gravity Distance approaches are highly prone to errors when the data is noisy, or a stem's face is partially occluded due to objects blocking the incident LiDAR pulse, or if the collection plan did not sufficiently cover this area. Although this method may function for TLS where the density of the data may compensate for these errors, it may not be sufficient for relatively sparser, noisier ULS data and therefore additional DHB extraction methods were tested in this study.

CHAPTER THREE: DATA COLLECTION AND PREPROCESSING

Study Area and field data

A subset of the Smithsonian Conservation Biology Institute's (SCBI) Forest Global Earth Observatory's (ForestGEO) 25.6 ha. site was surveyed due to the availability of field data, and the relative complexity of the overall canopy structure [2]. As of 2018, ForestGEO's census cataloged over 40,000 stems $\geq 1cm$ DBH. In part fulfillment of the ForestGEO protocols, stems $\geq 1cm$ were assigned a unique tag ID and located within a localized grid coordinate system (GCS) [52], [56], [57].

To facilitate the comparison of this study's results with the previous work done at this location, the census data was filtered to only include stems from 10 dominant species known to exist within the forest plot [9]. These ten species, predominately consisting of regional variations in *Quercus* (Oak), *Carya* (Hickory), and other hardwoods, reportedly made up 80% of the total forest biomass [9]. Of the aforementioned 10 dominant species, approximately 70% of the total biomass was determined to have DBH $\geq 50cm$ [9]. In addition to proportionally contributing more to the biomass in this study area, filtering the data to only include these 10 species facilitated better ULS-field data matching.

Figure 2 depicts the distribution of the 10 dominant species of stems across the southern half of the ForestGEO site bound by the ULS scan extents. To further facilitate proper data alignment, the census data was further filtered to only include stems of the 10 dominant species with DBH $\geq 18cm$. Removal of stems under this DBH would exclude the abundant, small understory plants that would hinder proper ULS-field data matching.

Although stems <18cm DBH would be more difficult to detect in the UAS LiDAR, the foliage itself would potentially be mistaken as stems during the clustering process and erroneously matched to field data entries. Ultimately, a minimum DBH of $18cm$ was selected through recently published literature which defined the maximum acceptable inventorying size ($\leq$30cm) [34].

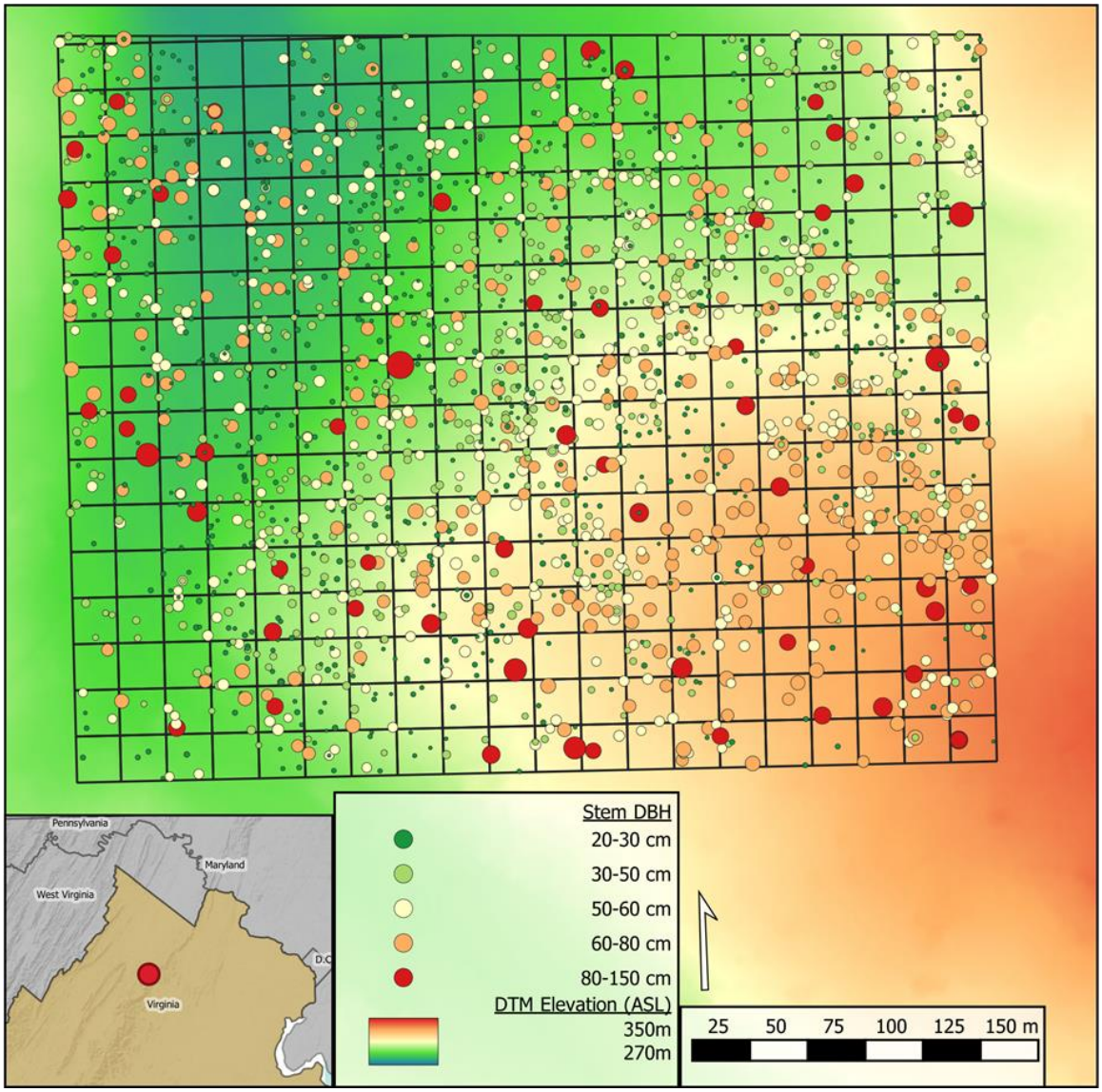

Figure 2: Stem locations and DBH derived from ForestGEO data (circles) in relation to the extents of the ULS data (grid). This figure was developed to mimic a style produced by Stovall et al. whilst highlighting this study's coverage [9].

Flight Plan

A total of five flights were undertaken on 03/19/2020 over SCBI's ForestGEO site. Limitations introduced by the maximum flight time of the UAS, lack of terrain following capability, site elevation fluctuations, and the desired data quality all

influenced the flight line orientation, and overall flight parameters. Furthermore, for sufficient ULS data collection, insights regarding proper flight altitude, speed, and overlap were drawn from Lu et al., Liu et al., and Wallace et al [25], [33], [58].

A target flight speed of 4m/s was entered into the flight control software, however, intermittent gusts of winds upwards of 4m/s caused sporadic speed fluctuations. 4m/s has been shown to collect dense enough returns using ULS at 45-75m AGL[16], [24]. General flight parameters are outlined in Table 1.

Table 1: Flight parameters used to survey the ForestGEO site.

Flight Parameters
Altitude: 45-75m AGL (Relative to Takeoff Site)
Speed: 4 m/s
Flight line spacing: 30m
Flight line Length: 300m
Number of flight lines: 14

Figure 3 depicts the flight plan with respect to the covered area of the ForestGEO site. A total of 12ha of the ForestGEO site was surveyed with an average point density of ~600 points / m^2. Flight lines were oriented in a manner to reduce the above canopy altitude variation within one flight. Due to a minimal altitude difference between the tops of the canopy, and the takeoff site, lower altitudes were flown on the more western flight lines. The increase in terrain elevation, and subsequent increase in relative canopy height required a higher altitude AGL respective of the takeoff site being flown. Nominal height

above canopy was 15-20m, while altitude of individual flights varied from 45-75m to compensate for variability in terrain.

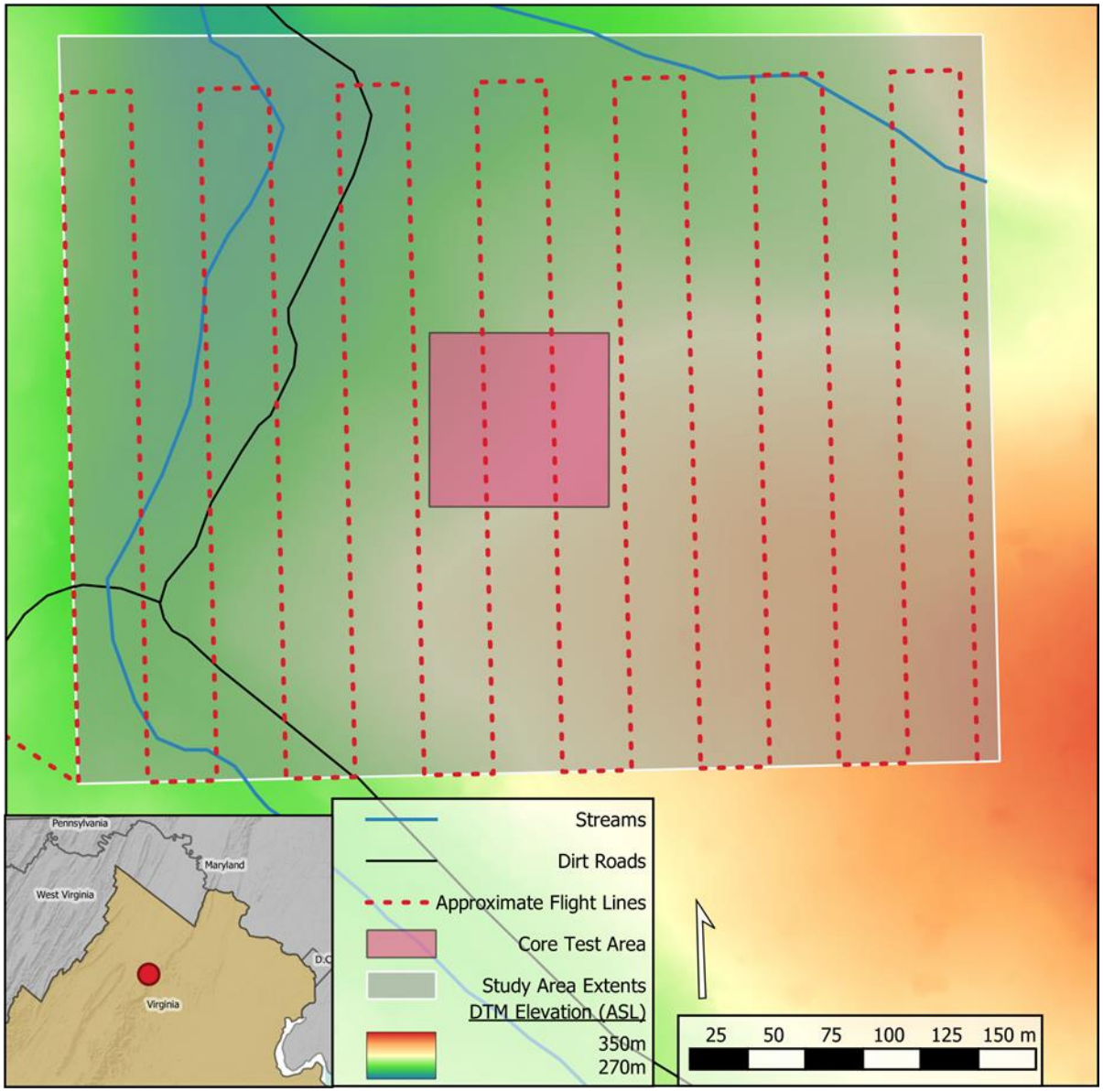

Figure 3: Programmed ULS flight lines flown over the ForestGEO study area selected (grey). A core test area (0.5 ha) is also depicted (pink) within which all methods were initially optimized.

UAS and Sensor Specifications

A DJI Matrice M200 was selected as the UAS to perform the 5 flights over the ForestGEO site. Compared to relatively outdated DJI Matrice M600pro, which weighs roughly 33kg when cased, the cased weight of the M200 is only 11kg, allowing for greater mobility and access to different takeoff sites around the area of interest [59], [60]. However, one drawback of the M200 is the reduced flight time, approximately 12 minutes, when operating the LiDAR sensor.

The LiDAR sensor selected for the 5 flights was the small footprint Quanergy M8-Ultra. This LiDAR emitted focused pulses of infrared light (λ=905nm) at a rate of 5-20Hz during operation [61]. A maximum of 3 returns was capable in a singular pulse, resulting in 1.3 million points being collected a second.

The sensor was oriented such that the top was parallel to the direction of UAS travel, which permitted collection of data a full 360º perpendicular to the direction of travel, however this was limited to only $\pm 55°$ to prevent "sky shots" propagating when the UAS turned at the end of a flight line. The M8-Ultra also scanned $+3°$ forward, and $-17°$ aft of the UAS, reducing the chances of occlusion along a flight line. The angular resolution was dependent upon the update frequency of the sensor but could achieve resolutions of $0.033° - 0.132°$.

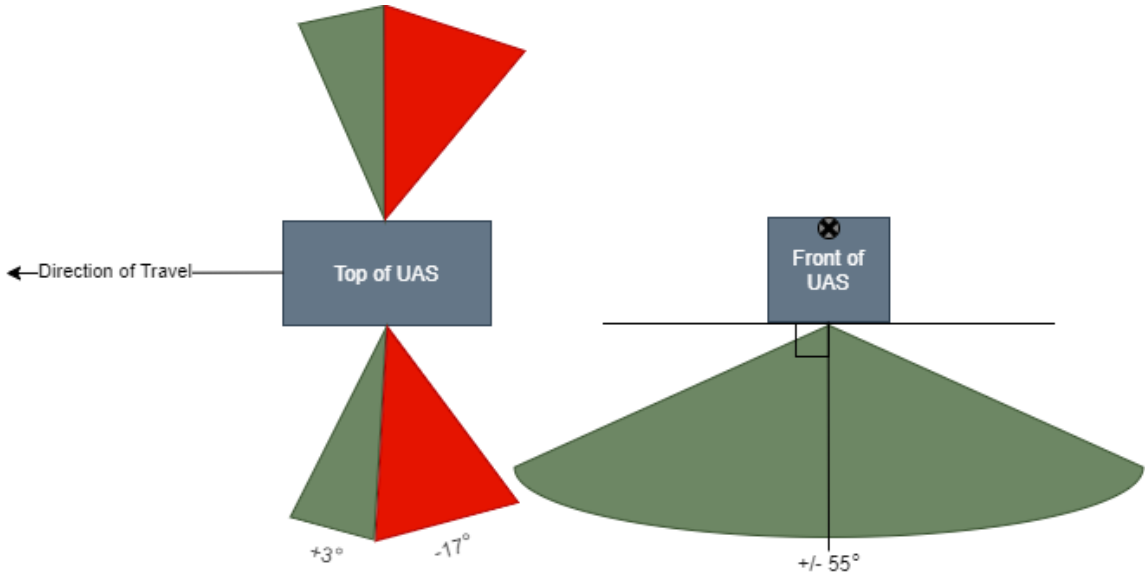

Figure 4: Fields of view of the UAS LiDAR, note the direction of travel is towards the reader in the "Front of UAS" depiction.

This sensor was selected primarily due to its reduced weight (900g), which permitted longer flight times over the target area, as well as the maximum range of 200m. The accuracy of the unit was moderate, <3cm at 50m (1σ). More information regarding the size and power requirements for this sensor can be found on Quanergy's website [62].

Data Preprocessing

Following initial data collection, boresight and Post-Processing Kinematic (PPK) corrections were applied in accordance with manufacturer recommendations using Inertial Explorer and ScanlookPC [63]–[65]. Cloud Compare, an open-source point cloud processing software was then used to perform cursory denoising with the Statistical Outlier Removal (SOR) tool to remove potential noise satisfying the criteria of $\geq 2\sigma$ away from a given point's 6 nearest neighbors [66]. In addition to the use of the SOR filter, the scan angle was also restricted such that each flight line's data consisted of

returns $\pm 55°$ perpendicular to the direction of travel. Restriction of the scanning angle has been shown to reduce the impact of stem occlusion which occurs at increasing ranges from the flight path [67].

The denoised point cloud was then spatially clipped to the area of interest depicted in Figure 3 and was normalized using R's lidR package [68], [69]. A normalized point cloud is representative of a CHM, whereby all point elevations represent their AGL height, rather than MSL height. Point cloud normalization allowed for the extraction of hyperplanes, as discussed on page 9 and depicted later in Figure 6, to provide uniform data extraction at a specified height AGL across the study area.

Software & Standards Used

Table 2 depicts the programs used throughout this analysis with respect to their version number.

Table 2: Programs used throughout this study and their respective versions.

Programs Used	
Program Name	Version Number
Cloud Compare	v.2.12 alpha
Inertial Explorer	v.8.80
R	v.4.0.3
Scanlook PC	v1.0.135

CHAPTER FOUR: METHODOLOGY

This chapter provides an overview and details of data processing, feature extraction be various methods and the statistical evaluation of the methods during comparisons with field data (Figure 4).

Clustering and Stem Segmentation

Once the point cloud was cleaned and preprocessed, the ambiguous LiDAR points underwent stem segmentation, discretization, and reconstruction using commonly used TLS and ALS modelling approaches. Regarding the former two methodologies, stem segmentation performed data reduction with the aim of removing the bulk of the data's noise due to undergrowth, whereas the latter aims to individualize stems whilst simultaneously removing spurious noise. Using the ULS derived stems, this study applied four methods to extract the DBH per stem: Convex Hull Estimation, Algebraic Circle fitting using Pratt's approach, Geometric Circle fitting using the Levenberg-Marquardt approach, and RANSAC Primitive Fitting [13], [15], [18], [70].

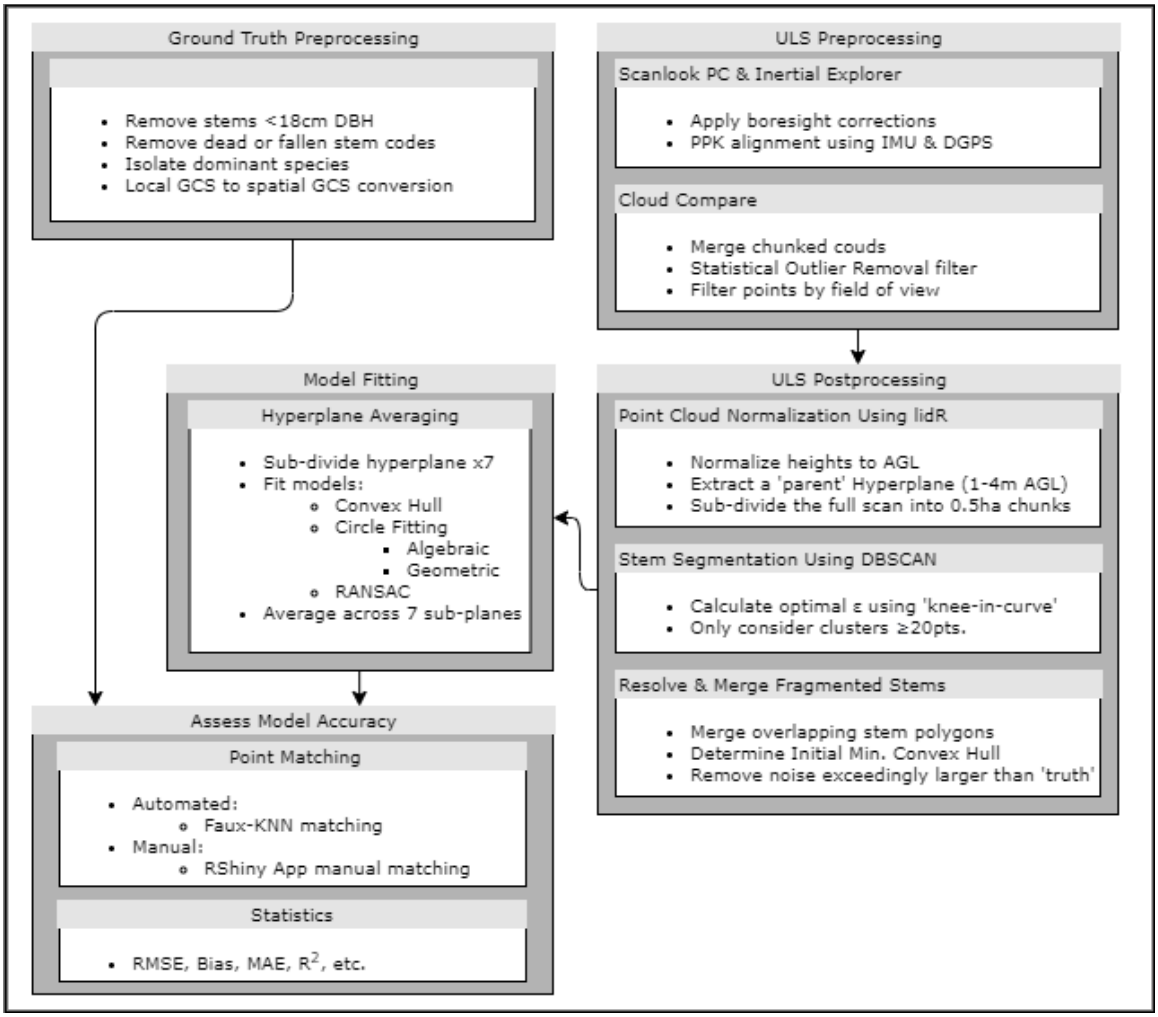

Figure 5: Generalized workflow highlighting steps implemented to extract stem DBH through various methods of model fitting.

Stem Segmentation

Prior to reconstruction, Hyperplane Slicing was implemented. This method, commonly referred to as “slicing” extracts data from the normalized point cloud through the use of a horizontally extruded plane of a defined thickness [18], [31], [33], [43]. Since the primary objective of this study was to determine stem diameter, location, and count, a slice ranging from 1.2m-4m AGL was extracted (Figure 5). The 2.8m thick slice at this

height was determined to include sufficient data for the implemented reconstruction techniques, in addition to aligning within the range of in-situ DBH collection heights.

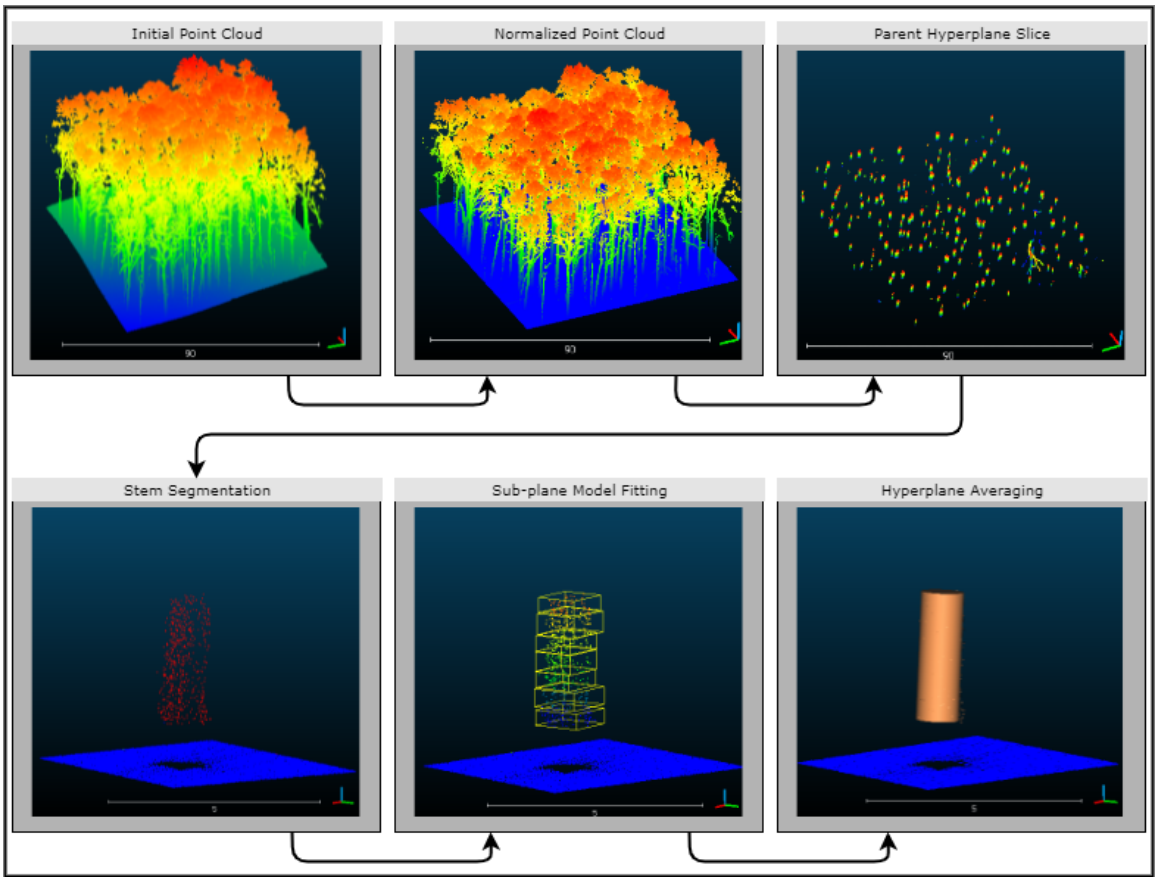

Figure 6: Major steps of the hyperplane fitting and stem segmentation workflow adopted.

Stem Discretization: DBSCAN Clustering

DBSCAN, unlike k-means clustering, is capable of clustering points when significant noise is present and has been shown to be robust enough to delineate LiDAR derived tree stems [13], [71]. Furthermore, compared to other categories of clustering, density-based clustering approaches such as DBSCAN or OPTICS show improved

performance when the shape of the cluster is ‘non-normally distributed’ [72]. Assuming the number of stems within the forest plot is not known, DBSCAN only requires the knowledge of the minimum number of points ($minPTS$) within a given cluster, and a neighborhood distance to be defined (ε). As the minimum points per cluster is dependent upon the sensor used, flight parameters flown, and the density of the point cloud, a $minPTS = 20$ was experimentally determined for this study. To reduce potentially noisy clusters that would bridge the spatial gap between multiple stems, a criterion was also implemented which limited the maximum cross sectional and cluster edge length to the maximum known DBH+20cm. This restriction had an intended secondary effect of removing prostrate or leaning stems that may not have been removed previously.

The package ‘dbscan’ in R was used in conjunction with an automated approach to determine the optimal ε per spatially relevant chunk. By removing the need for the user to manually define an optimal ε, the potential bias is thusly placed upon the specific flight parameters being flown, rather than through experimental trial and error by an analyst. This automated approach, referred to as detecting the ‘knee’ in a curve will be discussed in later sections.

Comparing this method of clustering to Liu et al.’s methodology, the method proposed and implemented here should produce clusters that are more adaptable to the point distribution which varies throughout the point cloud. Liu et al. assumed that the threshold value of the point cloud density remains constant for their clustering approach to be successful [38]. However, non-uniform data collection is highly likely to occur when, in the case of Liu et al.’s study, a “walker” (the operator of the instrument during

data collection), or in this study, the drone, travels at different speeds. Thus, DBSCAN should be augmented using subsections of the ULS point cloud (chunks) to create multiple adaptable neighborhood distances (ε) throughout the data.

Automated determination of Optimal Epsilon

Due to the expansive nature of our full dataset, 0.5ha data chunks were created to act as manageable subregions with which localized ε values could be determined. The k-nearest neighbor (kNN) is the i^{th} closest point to a given feature based on a predefined distance function. To determine the knee in a curve, the function "kNNdist" from the dbscan R package was leveraged to calculate all nearest neighbor distances between all points within 0.5ha chunks [73]. Points of criticality, or knees, were then calculated along this curve using an experimentally determined threshold. Given the density of the LiDAR data collected, a threshold of 0.0067 was determined to be sufficient for stem segmentation. A depiction of the knee-in-curve approach for a given 0.5ha subplot is shown in Figure 7.

As implemented, a locally adapted ε would be more robust to non-uniform point densities, as opposed to a fixed scan-wide value adopted in similar studies [13]. However, the implementation used to locally cluster stems resulted in the potential for stems to become fragmented across subplot boundaries. To prevent these fragmented stems from occurring, a spatial "chunking" process was adopted which duplicated and spatially shifted the subplot boundaries such that they would not interfere with stem clustering. Further justification for, and a more detailed description of the chunking process is discussed in the following sections on page 28.

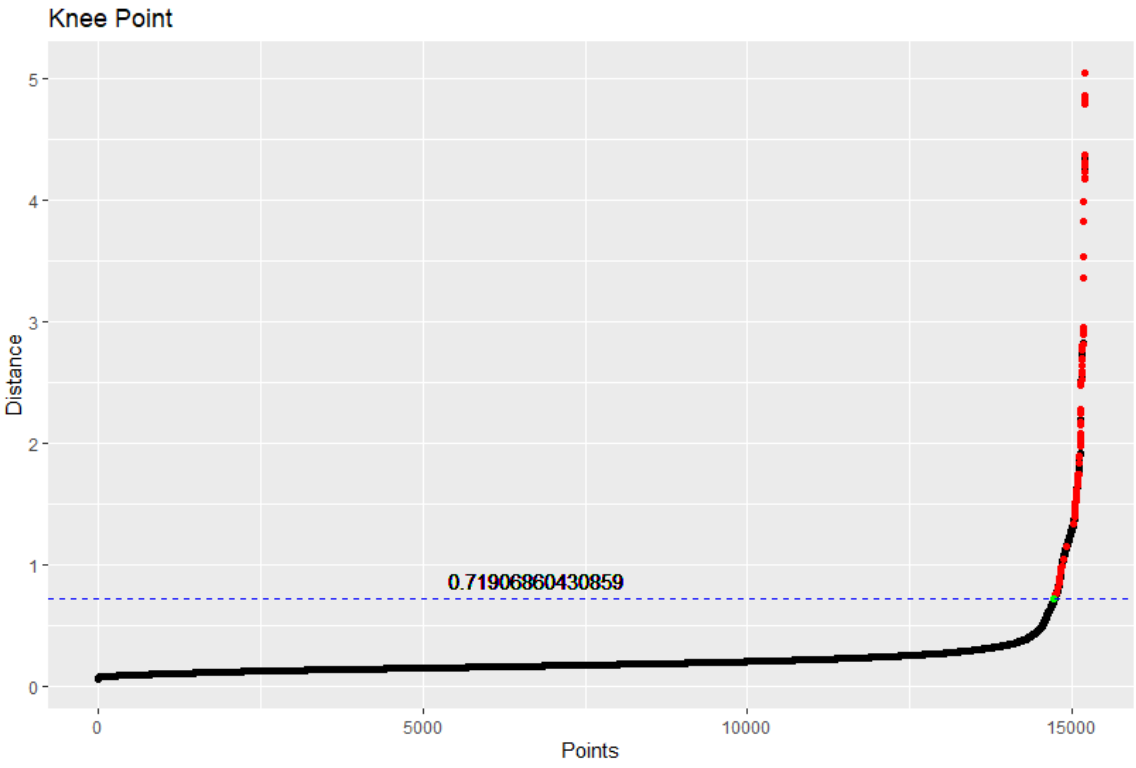

Figure 7: Locally fitted ε (green) determined using the minimum of the points of criticality (red). The black curve represents the k-NN distances across a randomly selected 0.5ha subset of the point cloud.

Necessity of "Chunking" Procedure

As shown in Figure 8, stems which occurred along the chunk boundaries required merging and reassessment to accurately estimate the DBH. Fragmented stems, although rare, occurred for ~68 (5%) of the detected stems.

To address this issue, additional, supplementary chunks were generated and spatially shifted such that corners would be located on the previous iteration's centroid [Figure 8[C]]. The adopted method assumed that if a stem should exist which crossed a boundary Figure 8[A], shifting the conflicting chunk boundaries 50m would resolve the potential fragmentation. When iterating through the shifted chunks, re-clustering would

generate slightly different chunks due to the presence of data which existed beyond the prior chunk's extents. Spatially unifying clusters which intersected with one another, due to successive chunks and re-clustering applications, allowed for recalculation of the DBH for the previously fragmented stem Figure 8[B].

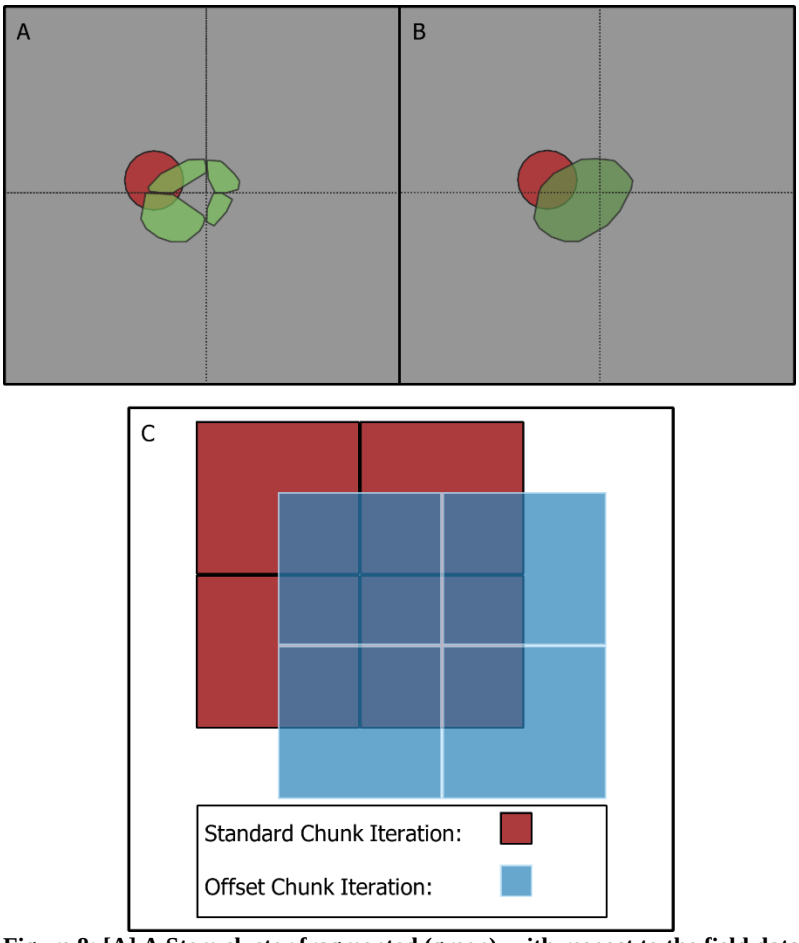

Figure 8: [A] A Stem cluster fragmented (green), with respect to the field data (red), at the boundary of 4 0.5ha chunks. [B] Resulting convex hull of the previously fragmented stem following chunk shift and merge process. [C] Generalized depiction of the chunk shift process whereby additional chunks are created and offset to resolve boundary issues throughout study area.

Field and ULS Data Matching Approach

Convex hulls of the clustered point cloud were manually matched and aligned with the ForestGEO census dataset using an RShiny application that the author developed [Figure 9]. The creation of this app was necessitated by the ubiquitous non-systematic spatial shifts of stem locations within the ForestGEO census data. During testing, and in practice, it was determined that the true geospatial accuracy of a stem's location was questionable and precluded the traditional methods of spatial data alignment such as kNN or topological overlap.

Figure 9: The RShiny app user interface developed and used to manually match the DBSCAN produced clusteres (red) with the ForestGEO census data (blue). Note, the current stem to be matched is displayed using a more vibrant shade of red, whereas the lighter red polygons provide the user with the context needed to select the most likely match.

It was determined that only the foremost grid corners of the ForestGEO site were geolocated by using a technique known as “pulling line” [7]. This method employs a theodolite, or surveyor’s telescope, to optically determine the azimuth and distance to an object in reference to a known location, or backsight. Stakes had been placed in situ to create 20x20m grids, each being further sub-divided into 5x5m sub-quadrats[7]. Stems were geolocated “entirely by eye” within these sub-quadrats [7]. The location error for a given stem would subsequently propagate as the distance increased from the previously surveyed backsight. Error was also introduced into the mapping protocol through the lack of standardization in estimating a stems relative center, or face [7].

With this knowledge, the decision was made to visually match the ULS derived stems with the available census data. To gauge the accuracy of these user generated matches, a confidence value was also reported by the operator which ranged from 1-5 (low-high). The confidence value determined using the probability of a potential match, as well as if there was any discernable spatial pattern within the data. The specific criteria used to justify each confidence value was as follows:

(1) A minor chance of a match, potentially noise but otherwise missing alternative matches, roughly a 5-20% chance of match.

(2) Some chance of a match, the relative hull size visually matches the field data, but no spatially related pattern could be determined, roughly a 20-40% chance of a match.

(3) A likely match, the relative hull size visually matches, and there is a weak discernable spatial pattern or offset that may have occurred between the field and ULS data, roughly a 40-70% chance of a match.

(4) Most likely a match, there is some noise or missing stems which could be disrupting the discernable spatial pattern of related field and ULS data, roughly a 70-90% chance of a match.

(5) Almost certainly a match, there is a strong spatial pattern present, >90% chance of a match.

Methods of Stem Reconstruction for The Extraction of DBH

Convex Hull Estimation

The minimum bounding polygon, or convex hull (CH), is a closed shape which will topologically contain an object [Figure 10]. Recent TLS studies have adopted CH and highlighted the practicality of this method and ease of implementation [74]. Outside of DBH estimation, CH has served as a basis for other methods of biophysical metric extraction such as Canopy Width (CW) [27].

Estimation of DBH using CH, is inherently naïve since it does not use any density-oriented relationships in the data, but only uses LiDAR returns along the periphery of a given surface. Moreover, by using points on the periphery of a surface, the relative error of the LiDAR sensor is not accounted for, thus further introducing error into the DBH estimation. Therefore, it is expected that this method will primarily overestimate DBH.

As implemented, to estimate a stem's DBH, the Euclidean distances between the vertices of a CH and its centroid were averaged. These radii, when averaged, were relatively robust to the possible issue of partial occlusion, as previously discussed, as no vertices would exist along the occluded faces.

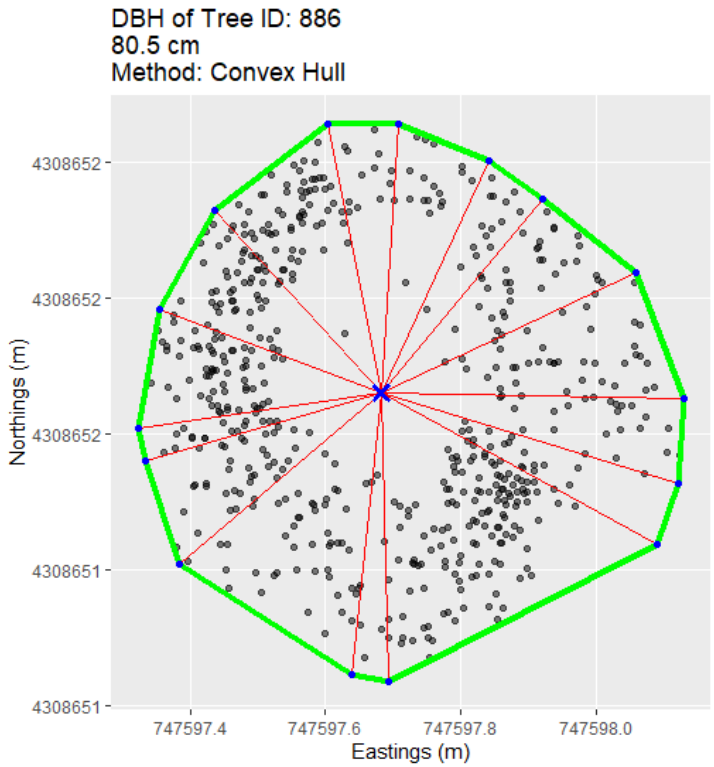

Figure 10: CH reconstruction of a stem where the outer segments (green) represent the minimum bounding surface of the ULS points (grey). DBH estimation is determined by averaging the distance (red segments) between the centroid of the CH and the bounding vertices.

Circle Fitting

Algebraic Approach: Pratt

Algebraic fitting techniques, such as Pratt's, can be adapted to a variety of surfaces such as lines, circles, 3D primitives, and others if they are capable of being

parameterized [75]. A core objective of algebraic methods is to improve the computational efficiency of the process whilst attaining similar results achievable through iterative geometric techniques [75], [76]. Using the Pratt approach (Pt hereafter), circles were fitted to clusters using the method proposed by Pratt in Equation 1 where $A, B, C,$ and D represent the parameterized circle, or when written as an objective function to be minimized in [10], [75], [76].

Equation 1: Pratt's Algebraic Circle Fit [10], [75]

$$A(x^2 + y^2) + B_x + C_y + D = 0$$
Constraints: $B^2 + C^2 - 4AD = 1$

Equation 2: Pratt's Objective Function [75], [76]

$$\mathcal{F}_P = \sum \frac{\left[A_{z_i} + B_{x_i} + C_{y_i} + D\right]^2}{B^2 + C^2 + 4AD}$$

Chernov et al. and Pueschel et al. highlighted that Pt was developed as an improvement to the traditionally used Kåsa and Simple Fit least squares [10], [76]. Pt is more efficient than Kåsa at fitting to circles that only contain partial segments of the arc. Compared to the previously mentioned CH method, Pt will fit a circle within the bounds of the LiDAR points, which could address the maximum accuracy of the UAS LiDAR (<3cm). One concern however is that Pt may perform worse when the samples are sparse,

which would occur in small stems, as the minimum orthogonal distances between very few points may create an erroneously large stem.

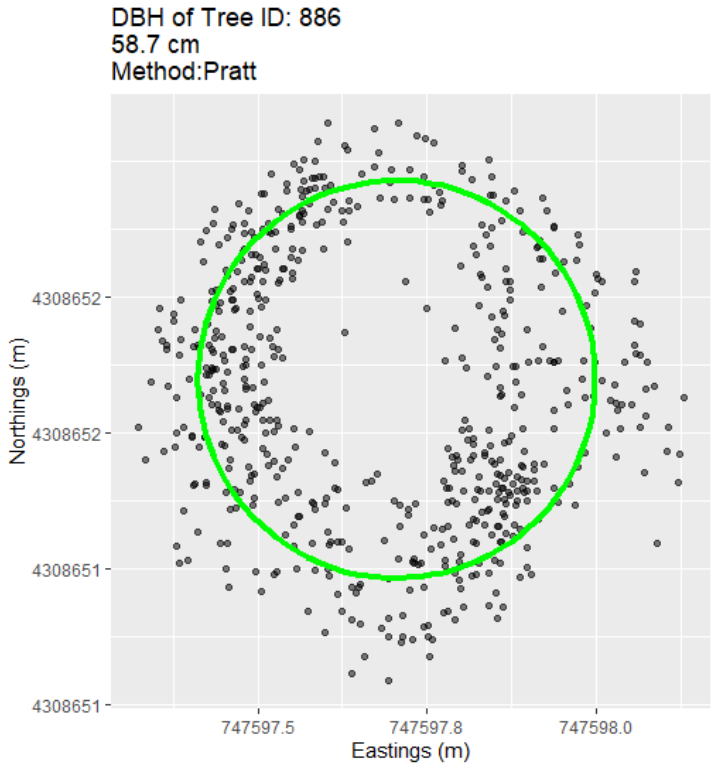

Figure 11: An example of DBH estimation using PT across the 2.8m parent hyperplane. Note the semi-inclusion of the extraneous data on the eastern face of the tree which is thought to be part of a branch. The green circle depicts the resulting optimally fitted PT circle for the ULS derived stem (grey points)

Geometric Approach: Levenberg-Marquardt

Geometric approaches of circle fitting, such as the Levenberg-Marquardt algorithm (LM), are considered capable of generating results superior to those achievable through algebraic fitting [76], [77]. Compared to the algebraic techniques, these require successive iterations to re-evaluate estimated parameters until the algorithm converges on an optimal solution [76], [78]. The specific instance of LM employed by this study is

referred to as the Reduced Levenberg-Marquardt, and is provided by R's 'conicfit' package [79]. As implemented here, LM requires the declaration of the maximum number of iterations per stem, which was limited to 20 as suggested for similar applications [79], and an initial guess for the potential center of the circle. Chernov and Lesort note that a potential, although statistically unlikely, flaw in geometric fits is their ability to become trapped in local maxima and minima during fitting [78]. To address this flaw, and mitigate the possibility of becoming trapped, an initial dampening factor (λ), and an initial guess for the circle's center can be used to guide movement of the model upon instantiation [78], [79]. For this study, λ was set to a value of 1, and the initial center was found using the gravity center of the stem's point cloud. The epsilon tolerance used to validate acceptable models was also set to 1, although this potentially could have been reduced [79]. These parameters, upon testing, were shown to produce sufficient results [Figure 12].

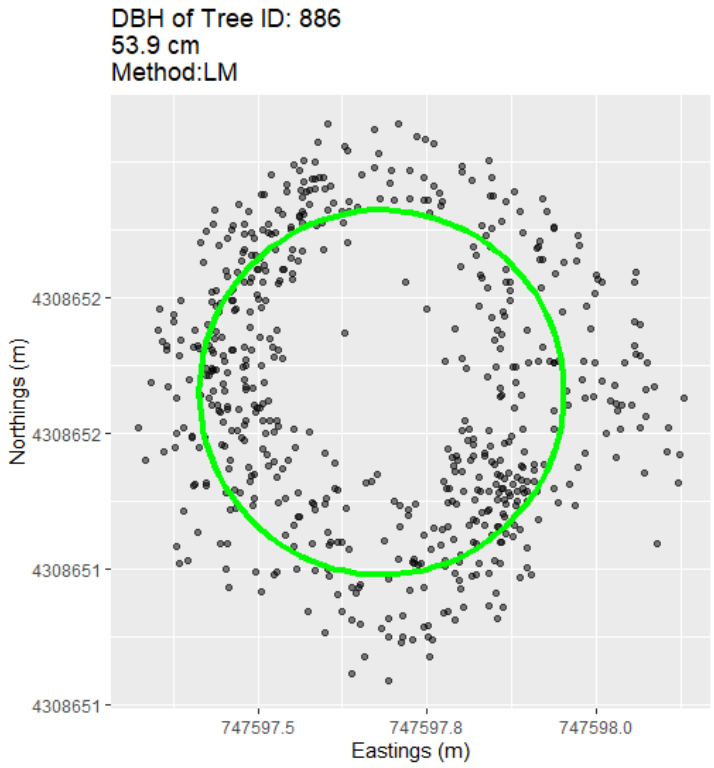

Figure 12: LM fitting example depicting the final optimal estimation of the stem's circular parameters (green). Note the relationship of the fitted circle's perimeter to the density of the LiDAR returns, which unlike PT does not include much of the extraneous branch on the eastern face of the stem.

RANSAC Primitive Fitting

Random Sample Consensus (RANSAC), is a more robust approach than LM and is more computationally intensive [78], [80]. Whereas LM relies on the gradient of data and a dampening factor to fit between iterations, RANSAC uses subsets of randomly sampled data to generate an optimal model. Additionally, RANSAC classifies the sampled points into sets of inliers and outliers, which has been shown to assist in noise removal [18], [81]. Compared to geometric and algebraic techniques which handle noise by "smoothing", RANSAC sampling assumes noise obeys a normal distribution [81]. RANSAC, and it's derivations have been used to reconstruct stem skeletons using ALS

data and fitted lines [18], estimate DBH by fitting RANSAC circles [82], [83], and to act as an initial seeding function for other models [80].

RANSAC requires the declaration of the maximum number of iterations, and a threshold distance that will be used to classify outliers between iterations. Through Equation 3, the number of iterations can be estimated using the probability of sampling an inlier and represents the number of iterations required to generate at least 1 optimal model. For our purposes, the probability of finding an outlier free sample was $z = .95$, the proportion of outliers was based upon Nurunnnabi et al.'s approach $\varphi = .5$, and the number of subset points was $h_0 = 3$. Only 3 points were sampled during each iteration as this is the minimum number of points required to define a circle, and improves the overall efficiency of this method [80].

Equation 3: Probabilistic approach for estimating RANSAC iterations [81], [83]

$$k = \frac{[\log(1 - z)]}{[\log(1 - (1 - \varphi)^{h_0})]}$$

Where: z = Probability of an outlier free sample
φ = Proportion of outliers
h_0 = Number of points in dataset

At the time of this study, no known R package contained a ported version of RANSAC circle fitting usable for our purposes. The authors transliterated core functions from Python's pyRANSAC-3D library [84], and implemented R's inherent optimization strategies. An example highlighting the rejected and accepted RANSAC fitted circles is shown in Figure 13.

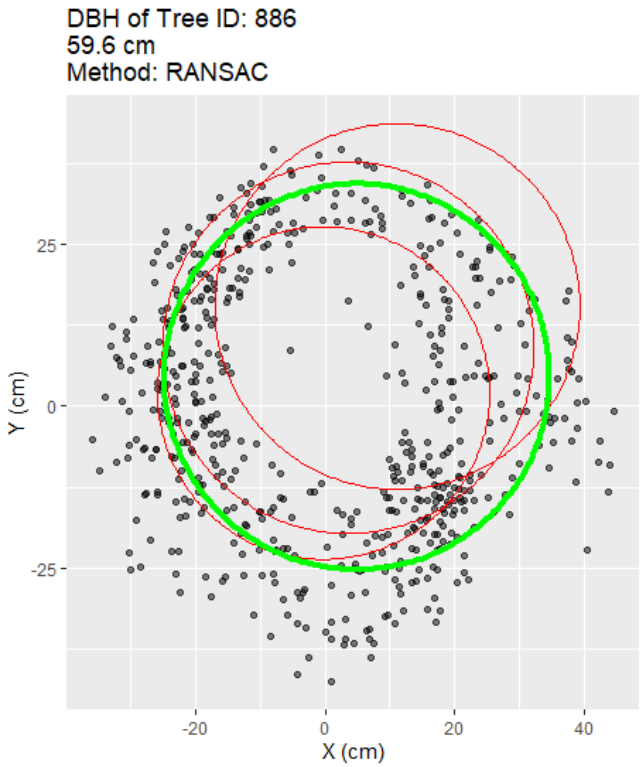

Figure 13: Iterative fitting of RANSAC circles where initial iterations (red) were optimized until the final solution (green) was determined upon reaching the maximum iteration or a convergence threshold was met.

CHAPTER FIVE: RESULTS

Results Summary

Two sets of primary results were generated for (i) the core test area (0.5 ha.), (ii) the full study area (12.5ha.), and (iii) an additional supplementary dataset using an assured confidence threshold of 5 (very high). To briefly reiterate, the adopted methods were initially tested and optimized using the 0.5ha core test area and subsequently scaled-up and applied to the full study area. Ultimately, when the analysis was scaled to the full extent, a marked decrease in performance was noted for both stem detection and DBH estimation as outlined below. In response to the decreased metrics, we investigated possible influencing factors and leveraged the confidence value defined during the manual matching to assess the theoretically best possible results.

Results:

Core Test Area

Of the 148 stems detected by the ULS, 121 could be visually matched to a potential field entry of the 10 dominant species. Within the 121 matched stems, 4 did not have unique tag IDs, which indicated that double matching, or false positives, occurred during the visual alignment process. The 27 stems that were not matched represent possible detections, or false negatives, of non-dominant species, or spurious foliage. As shown in Figure 14, detection rates varied with respect to the DBH bins. On average, stems <60cm DBH were detected at a rate of 85%, whereas the detection of the 12 remaining larger stems >60cm DBH showed a marked decrease to 58% in detection.

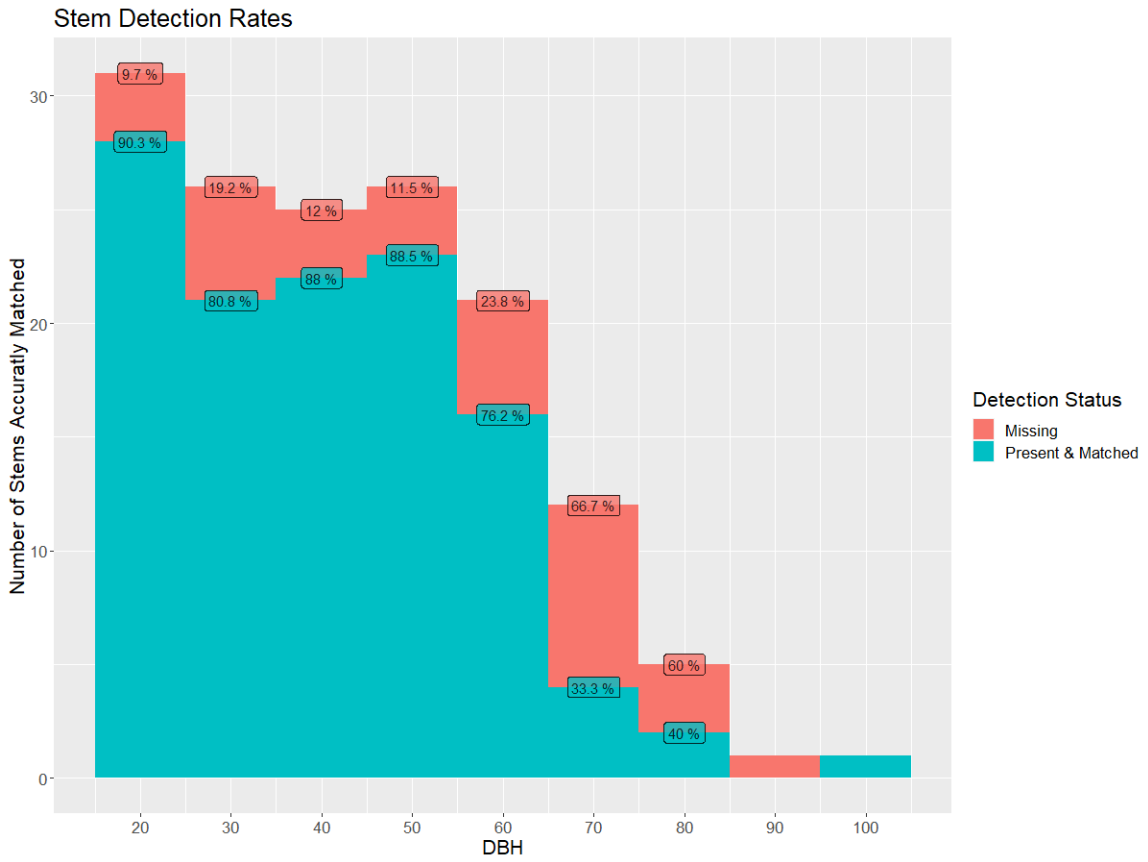

Figure 14: Detected and missing tree tags within the core test area depicted across DBH bins.

We investigated the initial accuracy of each of the DBH estimation methodologies (Figure 15). Although the R^2 was relatively low, it is important to note the increased concentration of samples along the 1:1 reference line, signifying a potential underlying relationship. Significant Pearson correlation coefficients between the True DBH and each of the methods CH, Pt, LM, and RANSAC were 0.61, 0.19, 0.6, and 0.48, respectively. Of the four methods, PT, had the lowest Pearson's correlation of 0.19, and the only method to show significant overestimation of stems at lower stem sizes. CH provided the highest correlation and regression values of .61, and .37, correspondingly.

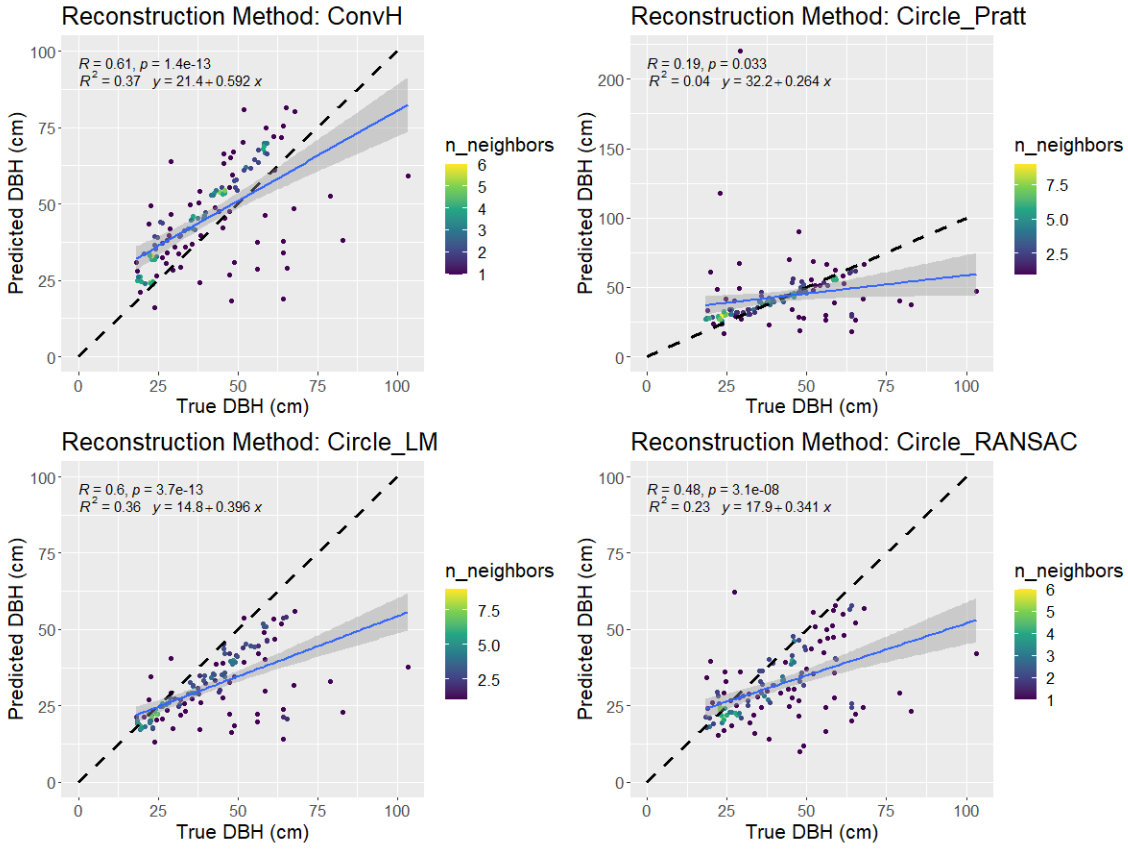

Figure 15 : Core test area DBH estimation regressions. Note the inclusion of a 1:1 reference line to highlight the relationship of the density of the well correlated samples. Each method's respective R^2, Pearson's correlation coefficient, and linear regression equation is depicted in the upper left corner.

Figure 16 displays the bias each method produced within various DBH ranges. A relationship of increasing negative bias for larger stem sizes occurred in all of the reviewed methods of DBH estimation.

Representing roughly 27% of the detected trees, CH and Pt overestimated stems 20-30cm in diameter by 10.4cm and 6.1cm, respectively. The iterative methods, LM and RANSAC fared better for smaller stems and underestimated by -1.9cm and -1.5cm,

correspondingly. Encompassing 58% of the total samples (68 stems) ranging from 30-60cm DBH, the CH overestimation increased from 7.3cm to 9.5cm, whereas all other methods had an increasing underestimation. Within the same range of 30-60cm DBH, Pt's bias was very low ranging from +1.2cm to -2.3cm, and LM's negative bias increased from -6.3cm to -11.4cm (Figure 15).

Stems >60cm DBH, which accounted for approximately 10% of the detected stems, were reconstructed with a negative bias across all methods. CH had a nominal bias of -0.5cm, whereas Pt, LM, and RANSAC were -8.6cm, -14.5cm, and -13.5cm, respectively. An increased quartile spread was also seen in the 25^{th} and 75^{th} intervals from a range of roughly 10cm to over 30cm for stems >60cm DBH.

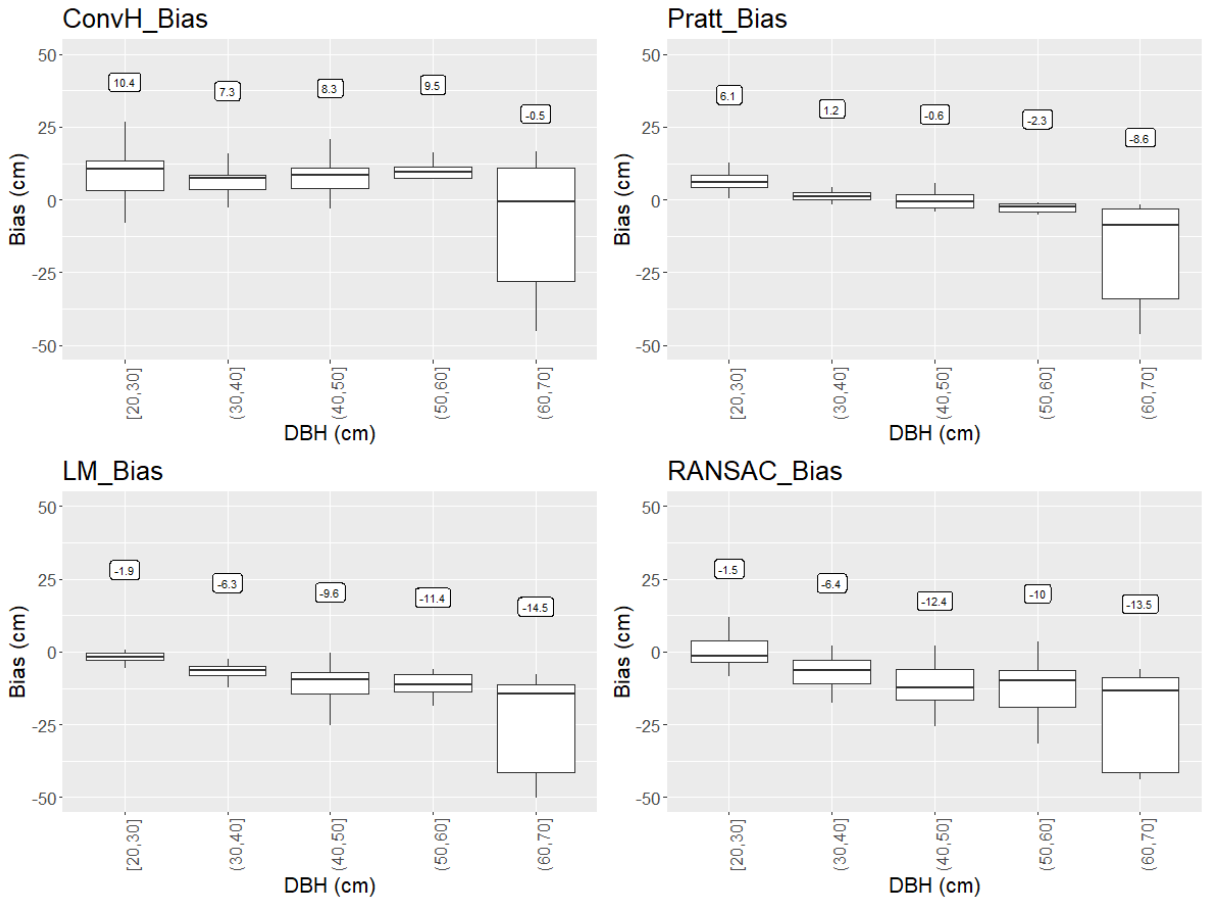

Figure 16: Core test area DBH biases for each respective method. The label above each boxplot denotes the median value per bin. Stems >70cm DBH were removed from these figures as they accounted for <3% of the total number of stems detected.

With regards to the MAE for each method within the core test area, larger amounts of error were present at larger DBH. The Pt method, having the highest initial error of 14cm, showed an initial decrease to 8cm in error across the range of 15-45cm DBH [Figure 17]. Comparatively, the three other methods increased in MAE across this range from 7-11cm for CH, 2-11cm for LM, and 7-13cm for RANSAC. From 45-105cm in DBH, all methods showed in increase in MAE up to 44cm, 56cm, 66cm, and 61cm for CH, Pt, LM, and RANSAC, respectively. One relationship to note is that LM and

RANSAC had the greatest overall variance in MAE across all DBHs. CH, although increasing in error, and having the second highest initial error, retained the lowest MAE from ~60-105cm DBH. MAE increase significantly above 60cm DBH for all methods. As previously discussed, DBH ranges >60cm encompassed less than 3% of the data, however extending the axis to include up the maximum DBH in the core test area emphasizes relationships among the upper DBH sizes. Of these relationships, the two iterative approaches, LM and RANSAC, behaved most similarly to one another.

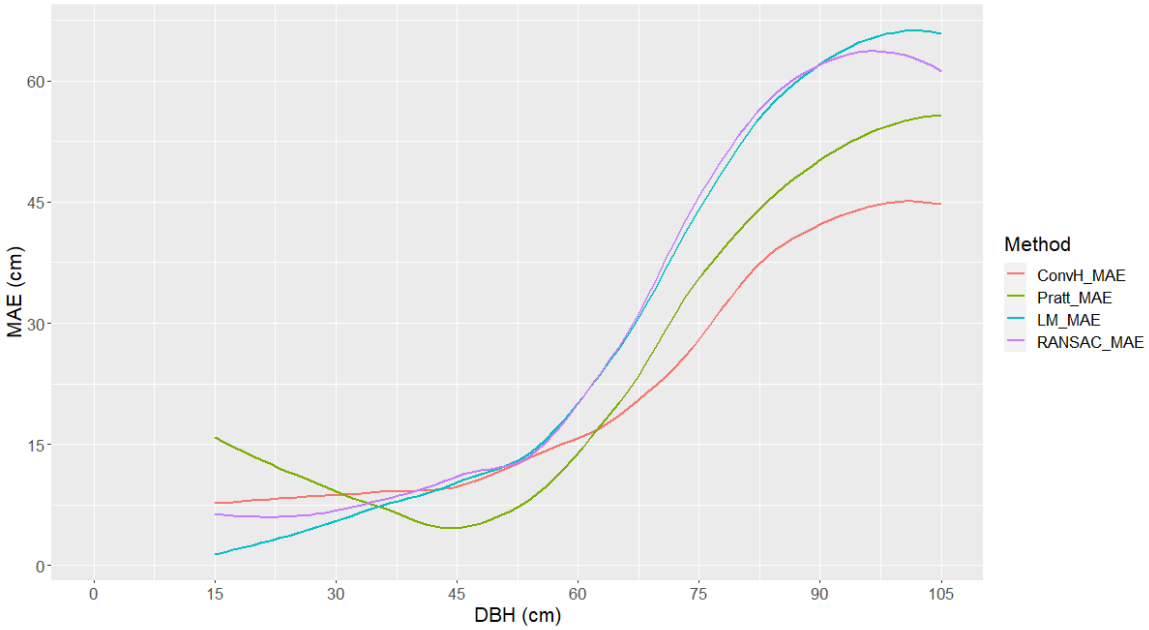

Figure 17: Core test area MAEs for each of the methods of DBH estimation. Note that <3% of the detected stems occurred >60cm DBH.

Normalizing the error values as a percentage of DBH or Mean Absolute Percentage Error (MAPE), enables comparison between DBH bins [Figure 18]. For stems 15-25cm DBH, the highest MAPE was 73% for Pt, followed by 39%, 11%, and 28% for CH, LM, and RANSAC respectively. Pt method continued to decrease in error from 25-

50cm DBH at which point Pt's MAPE has fallen to ~8%, the lowest error of all methods at these stem sizes. Comparatively, CH and RANSAC decreased in error from their initial values of 39% and 28%, to an error of 23% and ~26% respectively. LM continued to increase in error to an error of 24% at 45cm DBH.

For stems larger >60cm DBH, the relationships behaved similar to those depicted earlier in Figure 17 with MAPE values steadily increasing from 25-40% to 50-70%. CH and Pt provided less error than LM and RANSAC with values of 34%, 49%, 58%, and 63% respectively for stems of 75cm DBH.

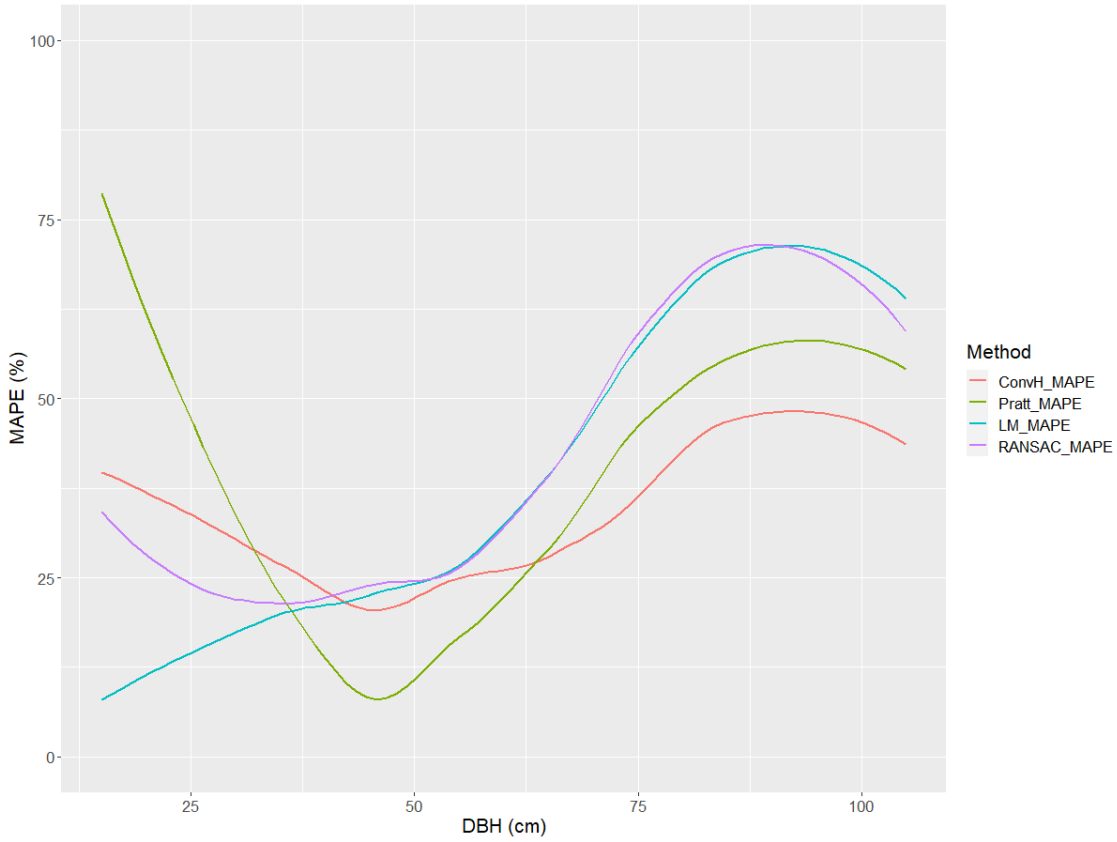

Figure 18: Core test area Mean Absolute Percentage Error within DBH bins.

Full Study Area

The area of investigation was expanded to the full 12.5ha surveyed. Of the 1892 stems detected, 1383 of them were successfully matched, and 1266 represented unique tag IDs [Figure 19]. Overall, this detection rate represents 63% of the total detectable tag IDs present in the full study area, a noted decrease from the 79% detected within the core test area. The 626 stems which were not matched or represented unique tags could be considered detections of non-dominant species, or small amounts of ground noise being detected.

Similar to the core test area, the full study most readily detected stems below approximately 60cm DBH. For the 731 stems detected <40cm DBH, representing 53% of the samples (detected stems), an average detection rate of 69% was found. Comparatively, the average detection rate for stems <40cm DBH was 85% in the core test area.

Within the higher DBH range of 40-60cm, 408 stems were detected which represented 30% of the total detected stems. The average detection rate for stems 40-60cm dropped from the 84% seen in the core test area to 69% in the full study area. A similar decrease in detection was also seen for stems >60cm from the 58% in the core test area to 39% in the full study area. These stems accounted for 17% of the samples, or 241 detected stems in the scope of the full study area.

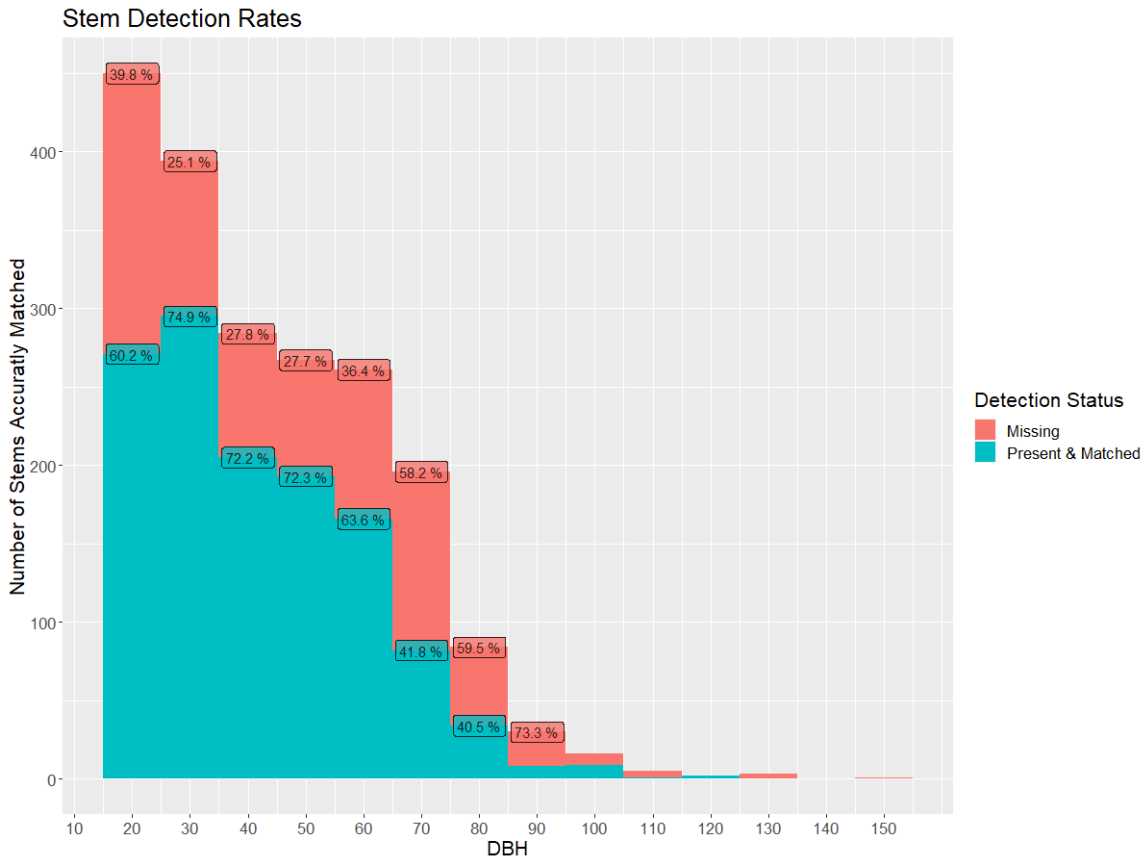

Figure 19: Detected and missing tree tags within the full study area depicted across DBH bins

The previously discussed relationship along the 1:1 reference line in the core test area is more evident in the full study area. This relationship highlights that the DBH estimation methods performed well for stems 15-45cm DBH.

Figure 20 was produced using samples below a noise threshold set to 1.5 times the known maximum DBH. If any method produced an estimation for a stem above this threshold, it was considered as noise and subsequently removed before the generation of Figure 20. A total of 35 stems, or 3% of the total matched stems, were removed using this method.

Little to no discernable visual change was seen in CH and LM, however removal of the noise from RANSAC and Pt facilitated better results interpretation. Initial R^2, prior to removal of noise, was 0.15, 0.01, 0.16, and 0.04, for CH, Pt, LM, and RANSAC, respectively. Following noise removal, these correspondingly increased by 0.01, 0.03, 0.01, and 0.02, reaffirming the amount of noise present in Pt.

Focusing on the pre-cleaned regressions depicted in Figure 20, Pearson's correlation coefficients were 0.4, 0.2, 0.41, and 0.24 for the CH, Pt, LM, and RANSAC methods, respectively. These results showed a decrease in correlation when compared to the core test area but behaved similarly overall. Moreover, the Pt method still performed the worst, when only considering the density plots in Figure 20.

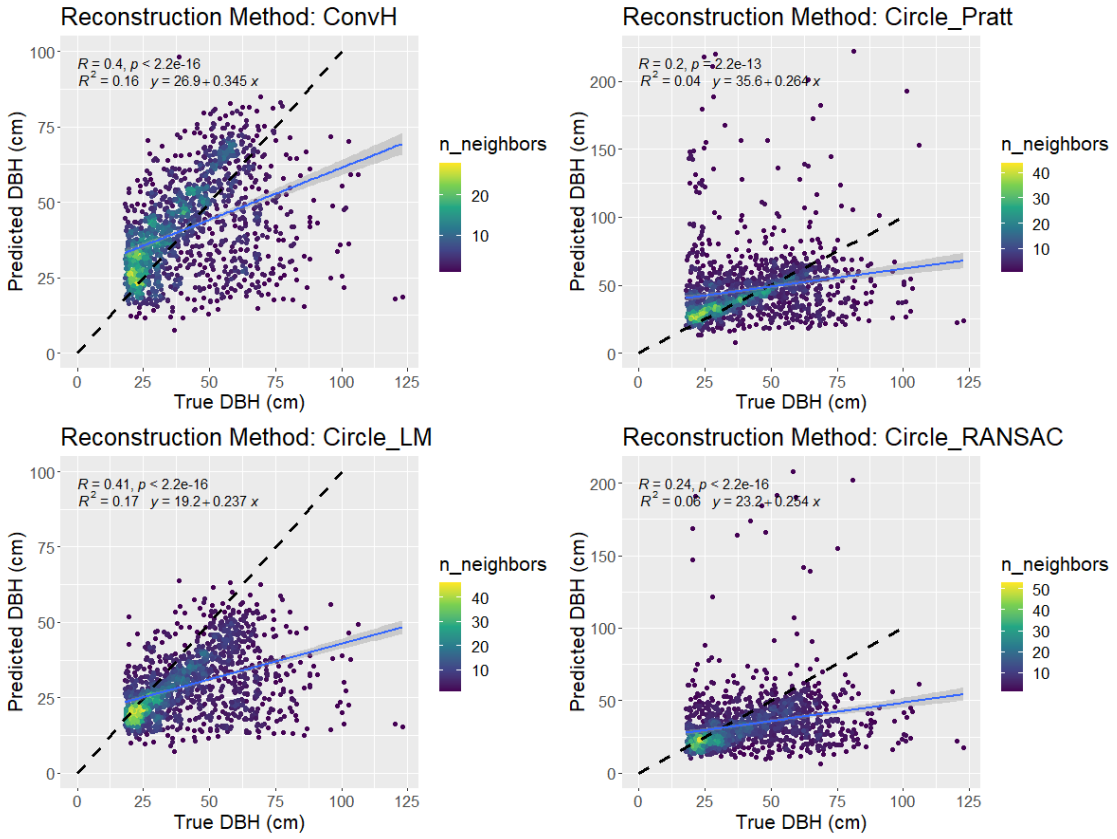

Figure 20: Full study area DBH estimation regressions. Note that the relationship along the 1:1 reference line is more evident than that of the core test area. Each method's respective R^2, Pearson's correlation coefficient, and linear regression equation is depicted in the upper left corner of each respective sub plot.

With regards to the biases introduced by these four methods, similar relationships as seen in the core test area were emphasized when the results were scaled to account for the entire study area [Figure 21]. In short, CH continued to overestimate stems 10-60cm in DBH, Pt initially overestimated for stems 10-40cm in DBH then began to slightly underestimate at larger stem sizes, whereas LM and RANSAC underestimated most all DBH bins.

Accounting for approximately 6% of the detected stems, CH overestimated stems in the range of 10-20cm DBH by10.8cm, Pt overestimated by 14.2cm, and LM and

RANSAC slightly overestimated by 1.6cm, and 3.8cm, respectively. For the 383 stems 20-30cm size, or 28% of the detected stems, CH continued to overestimate by 6.5cm, Pt overestimated by 7.7cm, and LM and RANSAC began to underestimate by -2.6cm and -0.3cm, respectively.

The next two bins include stems 30-50cm in DBH, the 461 stems which represented approximately 33% of the total detected stems continued to show a decrease in bias by 1.5cm, 3.3cm, 3.6cm, and 3cm for CH, Pt, LM, and RANSAC respectively. At 50cm, 3 of the 4 methods, Pt, LM, and RANSAC started underestimating stem DBH. LM had the highest underestimation, followed by RANSAC with -10.7cm and -9.3cm bias at 50cm DBH. Comparatively, at 50cm DBH, CH overestimated by 4cm, and Pt underestimated by a mere -0.2cm.

These trends of increasing underestimation continued from 50-80cm DBH, which accounted for 399, or roughly 28% of samples (detected stems). CH initially overestimated by 2.5cm from 50-60cm DBH but began to underestimate for larger stems by -29.7cm for stems 70-80cm DBH. Pt bias behaved somewhat similarly, by underestimating initially with -2.5cm, and increased to -23.6cm at 70-80cm in DBH. LM and RANSAC showed continued increase in negative bias as seen in previous bins from -15.7 to -43cm and -14.4 to -39.8cm, respectively. The remaining 44 stems >90cm DBH, representing 3% of the stems detected, saw continued increase in underestimation with -44.8cm, -34.3cm, -56.5cm, and -54cm for CH, Pt, LM, and RANSAC.

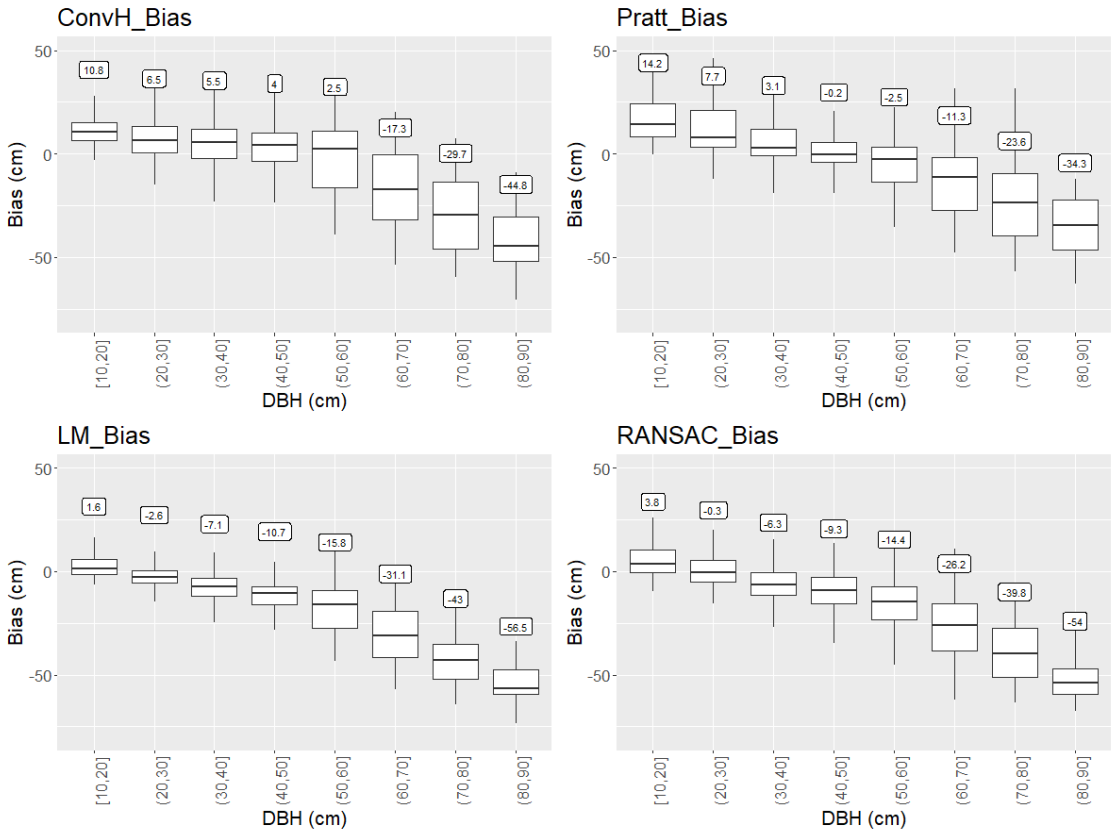

Figure 21: Full test area DBH biases for each respective method. The bin's median value is depicted above each upper limit. Stems >90cm DBH were removed from these figures as there were less than 10 stems per bin.

MAEs for the full study area show significant error propagation as the DBH of the stem increases. Initially, at 15cm, Pt had 22.1cm of mean error, followed CH with 11.6cm, RANSAC with 8.1cm, and LM with 4.74cm. As the DBH of the stem increased, each method behaved identical to results attained within the core test area: Pt decreased showed a marked decrease in MAE, CH showed a negligible amount of increase or decrease in error, and LM and RANSAC had increasing amounts of error. At 35cm DBH and approximately 8-10cm of error, a convergence occurred between all of the methods, causing an inversion in the order of the comparatively poor and well performing methods.

One relationship depicted in the full study area extents, which is not readily apparent at the core test area, is the consistent offset between the two iterative and the two non-iterative methods. Most notably to occur between 60-100cm, this relationship showed that LM and RANSAC had an additional 10-20cm of consistent mean error when compared to CH and Pt.

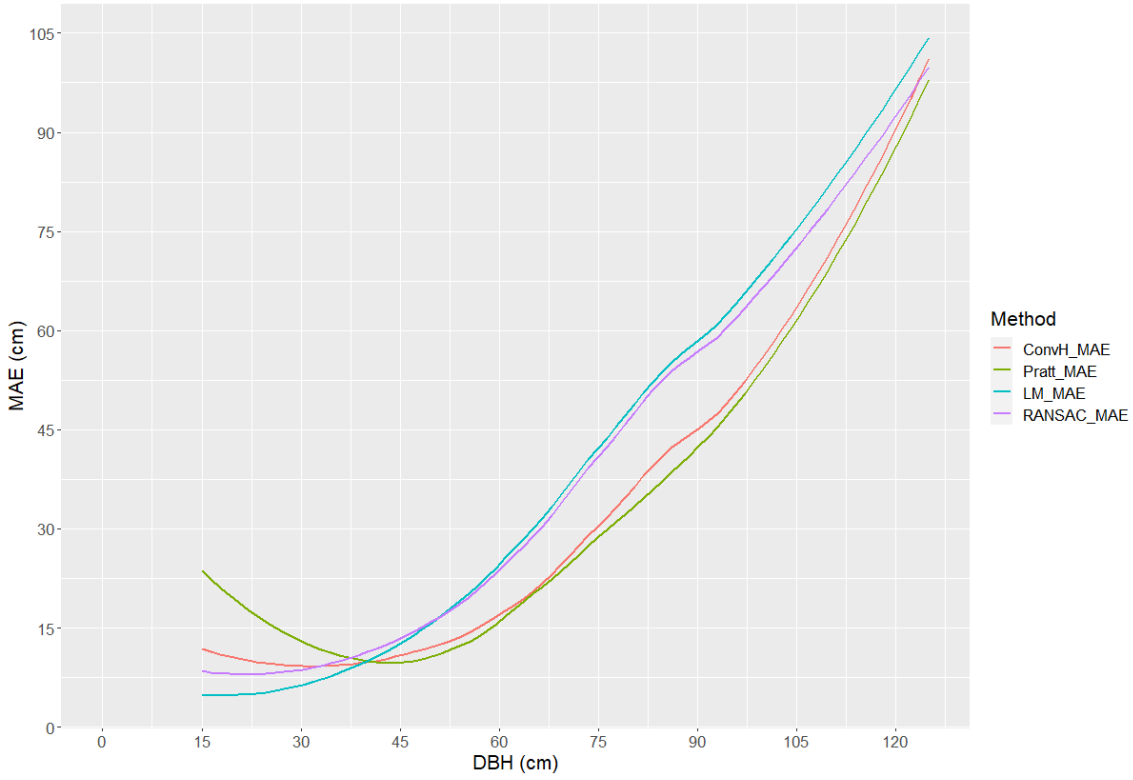

Figure 22: Full study area DBH estimation MAEs. Note the linear relationship of all four method >45cm DBH, and the significant gap which occurred between the iterative and non-iterative methods approximately 60-100cm DBH.

As with the core test area, normalizing the MAE as a percentage facilitated better understanding of where the error was occurring proportionally across the assessed DBH range. The highest amount of error present in all methods occurred for Pt at 15cm DBH with 115% error. Pt's error dropped to a minimum of 20% at 45cm DBH where it again began to increase, although less drastically, and behaved similarly to CH. However, as is evident in Figure 23, estimation using Pt is significantly more unstable than the other methods, as is discernable from approximately 45-75cm where multiple inflections occurred in the curve.

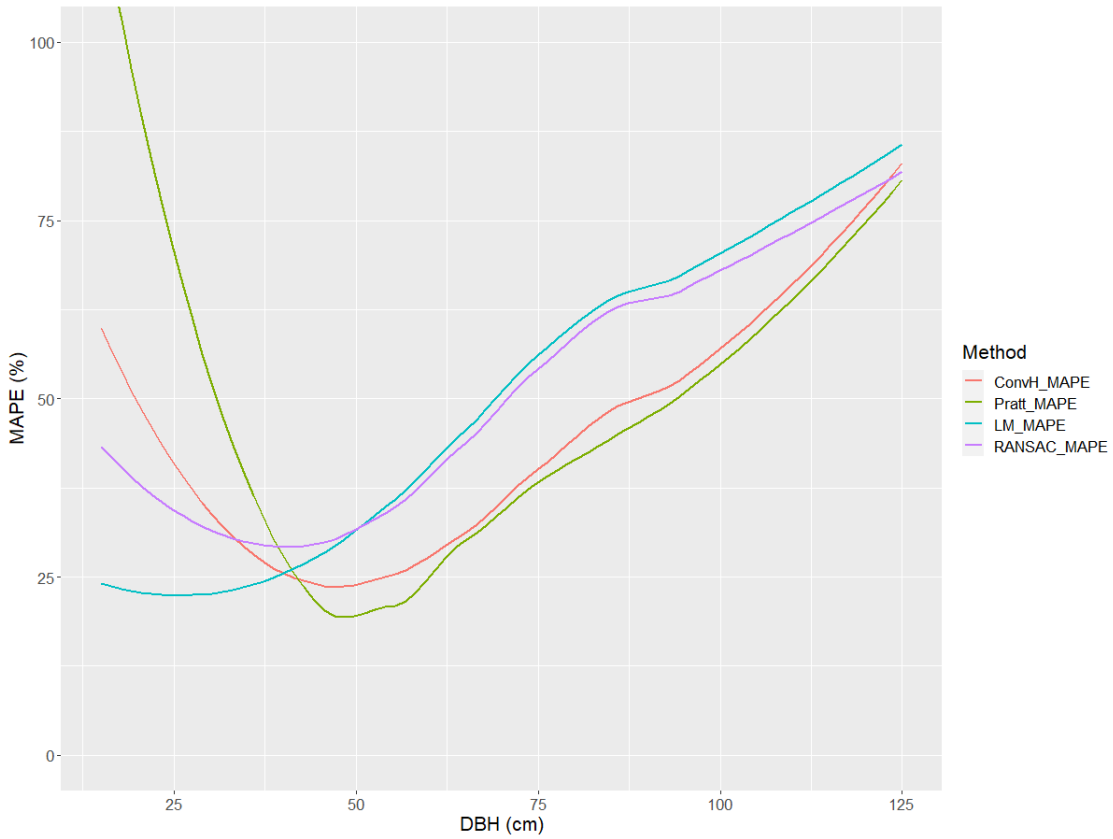

Figure 23: Full study area MAE normalization as percentages. Note the similar, but emphasized, error present in Pt which also occurred in the test area.

Erroneous Detections and DBH Estimates of Large Stems due to Mismatches

To further investigate the increased negative bias at higher DBH ranges, an assessment of large stems >90cm DBH in the full study area was performed. A total of 36 large stems were cataloged in the field data, of which 17 were found to have a corresponding match in the ULS data. Visual inspection of the large stems in the ULS point cloud later revealed this to be incorrect as all 36 large stems were discernable within the ULS data but were incorrectly considered noise during the cluster filtering process [Figure 24].

To be concise, it is believed that the threshold which removed clusters based on their size was set too low to allow large stems to be sufficiently represented in the matching process. Originally, the size-based filter was created to remove clusters that merged closely located stems with one another and prostrate or extraneous foliage; however, the maximum cluster size appears to have been too stringent.

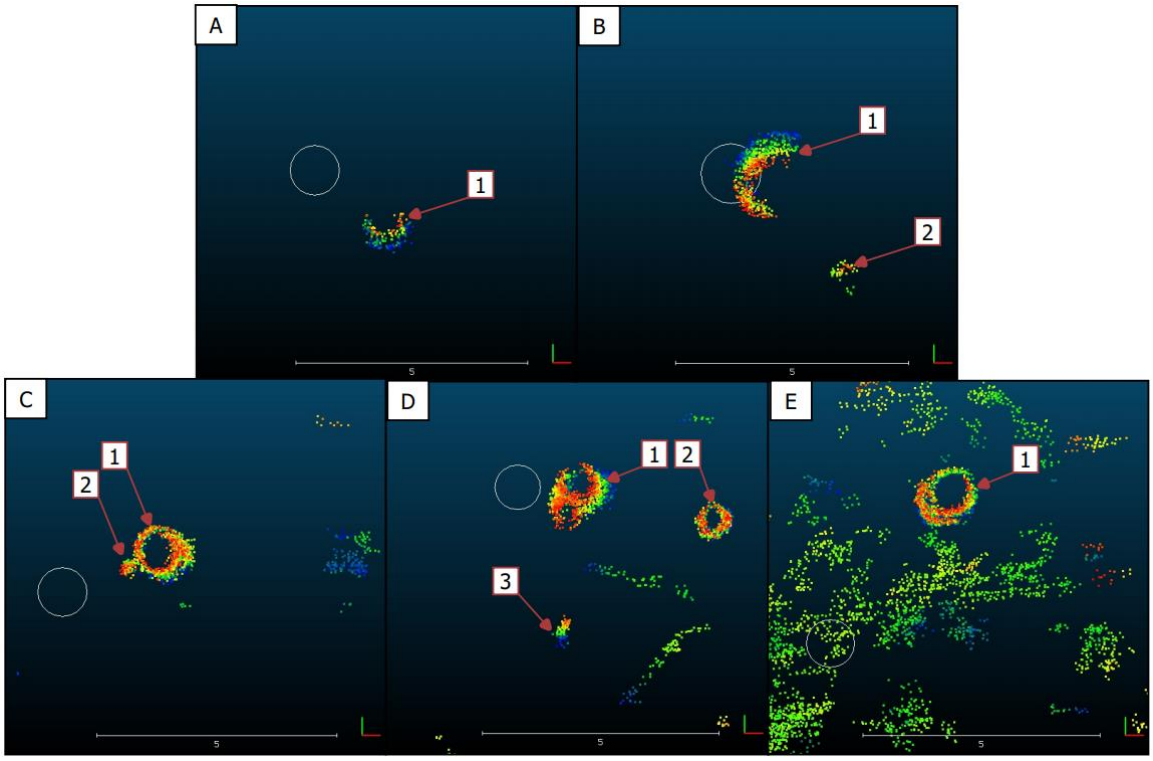

Figure 24: 5 example stems >90cm DBH. (A) Depicts the correctly matched field data (white circle) and the properly clustered, but occluded, large stem (1). (B) Depicts the large stem (1) which should have been clustered, but was not due to occlusion, and the actual stem, incorrectly matched (2) to the field data (white circle). (C) Shows two stems in close proximity, and noise, which were clustered together and subsequently removed due to excessive cluster size. (D) Depicts a large double stemmed tree (1), two smaller stems (2 and 3), and several smaller prostrate stems. (E) Shows a large stem (1) which was surrounded by excessive low-lying foliage, and a significantly offset field data.

For example, A in Figure 24 represents a 106.1cm DBH stem that accurately clustered using DBSCAN and matched; however, as this stem was located along the edge of the surveyed plot, the northern face was occluded, thus causing erroneous DBH estimations. The methods CH, LM, and RANSAC underestimated by approximately 50cm, whereas Pt overestimated by 50cm. Example A is therefore an example where the cluster was not removed, but DBH estimation failed due to lack of sufficient coverage by the ULS.

Alternatively, what was more common when attempting to cluster large stems was the excessive size of the cluster, resulting in removal by the threshold of 20cm greater than the maximum known DBH. Stem 1in example B was not clustered due to the occluded eastern face and the relative slant of the stem due to the terrain. During DBSCAN, the initial cluster had to have been generated for stem 1 but was incorrectly reclassified as noise due to the diameter of the total cluster exceeding the threshold outlined in Stem Discretization: DBSCAN Clustering. Following the removal of stem 1 from the potential stems to be matched in the Stem Matching App, a mismatch occurred with the only other visible cluster, stem 2. The erroneous mismatch would then negatively affect the resulting DBH estimation analysis.

A mismatch also took place in sub-figure C which depicts two stems in close proximity of one another. Inevitably, DBSCAN would have struggled to separate these two stems, resulting in an abnormally large cluster size. Further complicating this situation, noise was also present in the surrounding area which may have accidently been clustered with stems 1 and 2. Ultimately, the foliage, and stems 1 and 2 were removed due to their excessive size, thus a comparably sized stem was erroneously selected as the match (not depicted).

Similar to the previous two examples, a mismatch also occurred in example D, but a double match using stems 2 and 3 was also occurred. The double match resulted from similarly sized stems being weighted equally as potential matches to one ground truth. Within the Stem Matching App, no spatial pattern would have been apparent to the user as stem 1, and the prostrate stems would have been removed due to excessive size.

This example highlights a fundamental flaw in the visually aided matching where these double matches would inevitably occur and introduce twice the amount of error into the analysis.

Lastly, sub-figure E highlights the potential impact of low understory foliage. As with previous examples, stem 1 was removed from the potential clusters in the matching process due to the excessive size of the cluster. Similar to the double stems shown in sub-figure C and D, the noise present in example E would prevent DBSCAN from discriminating proper stem clusters. Ultimately, a mismatch occurred with a small cluster of low-lying foliage (not shown).

Given the examples highlighted in Figure 24, it is evident that the incorrect matches between the field data and DBSCAN clusters significantly contributed to large biases and MAEs shown in the previous sections.

High Confidence Matching and DBH Estimation

In consideration of the erroneous clustering of large stems, and the error compounded by mismatches, the final set of data presented consists of results using only data with a high confidence in the visual matched, i.e. confidence threshold of 5 (very high). These results span the entire study area and focus on the accurately matched field and ULS derived data.

199 stems, or roughly 10% of the matched and present stems met the “very high” confidence threshold. Linear regressions and their related statistics drastically improved for methods CH, Pt, and LM; however, RANSAC returned identical results to what was

obtained across the full study area. R^2 values were 0.71, 0.13, 0.69, and 0.23, respectively [Figure 25]. Using only data with high confidence matches resulted in decreased noise at higher DBH ranges for Pt and RANSAC but did not resolve all of the noise present at the lower DBH ranges of approximately 15-35cm DBH.

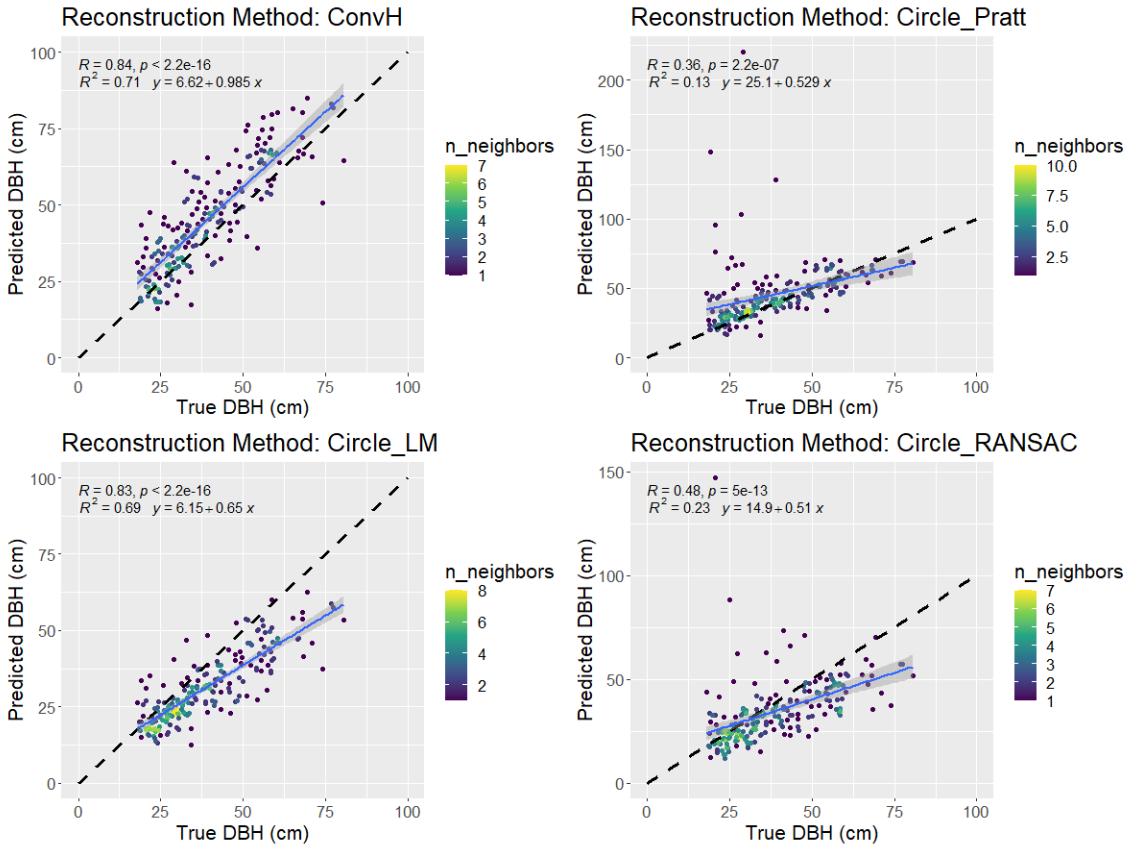

Figure 25: DBH estimation regressions using a matching confidence threshold of 5. Note the improved correlation compared to the previously presented results. Each method's respective R^2, Pearson's correlation coefficient, and linear regression equation is depicted in the upper left corner of each respective sub plot.

When assessing the bias estimates using the confidence threshold, similar improvements were evident Figure 26. CH had a positive bias across all DBH ranges 20-

60cm in DBH, with a minimum positive bias of 3.4cm at 20-30cm, and a maximum of 9.7cm at 50-60cm DBH. Pt initially had a positive bias of 5.1cm at 20-30cm DBH but decreased to 0.2cm for 40-50cm, and -0.6cm at 50cm-60cm. Pt's results using the confidence thresholding showed significant improvement from previous non-threshold areas of study and suggests failure of this method at small DBH sizes, but relative effectiveness at larger stem sizes.

Regarding the two iterative methods, LM and RANSAC consistently showed negative bias, rather than the initial positive bias seen without confidence thresholding. With the confidence threshold, LM had a negative bias of -4.3cm for stems 20-30cm, which subsequently increased to -13.1cm at 50-60cm. RANSAC behaved similarly to LM but had slightly less negative bias of -1.4cm at 20-30cm, and -10.9 at 50-60cm.

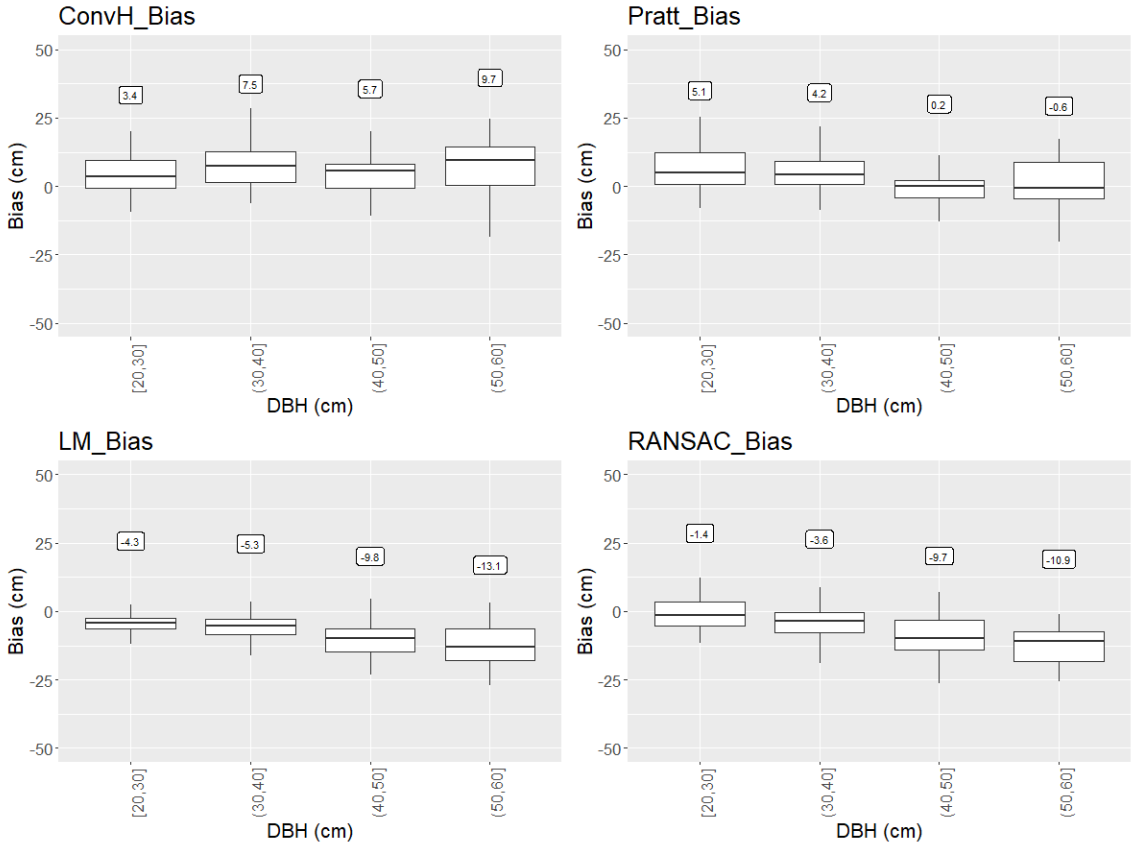

Figure 26: DBH biases for each respective method using a confidence threshold of 5 (very good). The bin's median value is depicted above each upper limit. Less than 10 stems >60cm were present in this assessment and were subsequently removed from the figure.

By selecting the high confidence matched data only, MAPE highlighted the drastic improvement in estimation of higher DBH ranges previously noted for stems >60cm DBH [Figure 27]. Previously, stems approximately 65cm in diameter across the full study area had MAPEs of 30.4%, 32.9%, 46.4%, and 42.3% for CH, Pt, LM, and RANSAC respectively; however, using the confidence threshold this decreased to 12.7%, 7%, 23.2%, and 20.2% correspondingly. This represented a significant decrease of 22% on average for this DBH range.

At approximately 15cm DBH, CH decreased in MAPE from 61% to 46%, which was significant. Pt showed an increase from 115% to 135% for stems approximately 15cm in diameter, further reaffirming the inability of this method to assess small stems. LM and RANSAC behaved in parallel to one another at the smaller stem sizes and changed minimally from previous results. At roughly 15cm DBH, LM decreased by <1% when using the confidence threshold, and RANSAC increased by 3.7%.

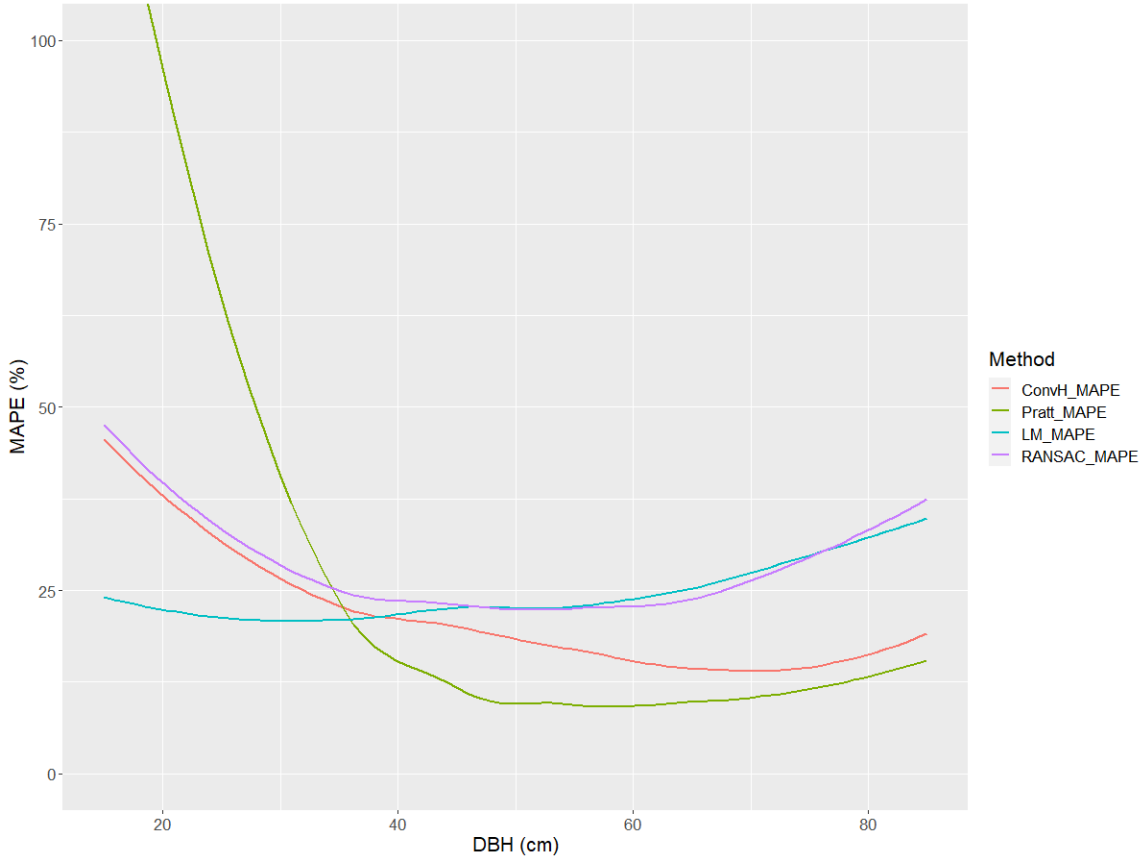

Figure 27: MAE normalization as percentages whilst using only stems that fulfilled the confidence threshold of 5. Note the significant reduction in MAPE at larger DBH ranges compared to the full study area. Also note the minimal change of MAPE at smaller DBHs.

CHAPTER SIX: DISCUSSION

We have shown through three sets of results the underlying relationships which support our ability to discern individual stems from ULS derived data and extract their DBH. We will now provide the context of these results within the greater scientific community and discuss our ability to address this study's specific objectives of (i) assess the ability to automatically detect and extract individual tree stems using DBSCAN, and more prominently (ii) test the accuracy of four DBH estimation methods adopted from TLS and ALS at stand-level scales.

Assessment of Stem Segmentation and Clustering Quality

Within the test area, DBSCAN performed as intended and clustered roughly 85% of the known stems <60cm in diameter. Rather unsurprisingly, the average clustering accuracy then decreased from 85% to 70% when assessing the full study area. We have shown that stem occlusion, inappropriate noise filters, and low-lying foliage are key factors which have adversely affected efforts taken to optimize DBSCAN and its related parameters.

Figure 28 further emphasizes that many stems, such as those shown in examples #1 and #2, were present in the ULS point cloud but were not sufficiently segmented by DBSCAN or represented in the matching program.

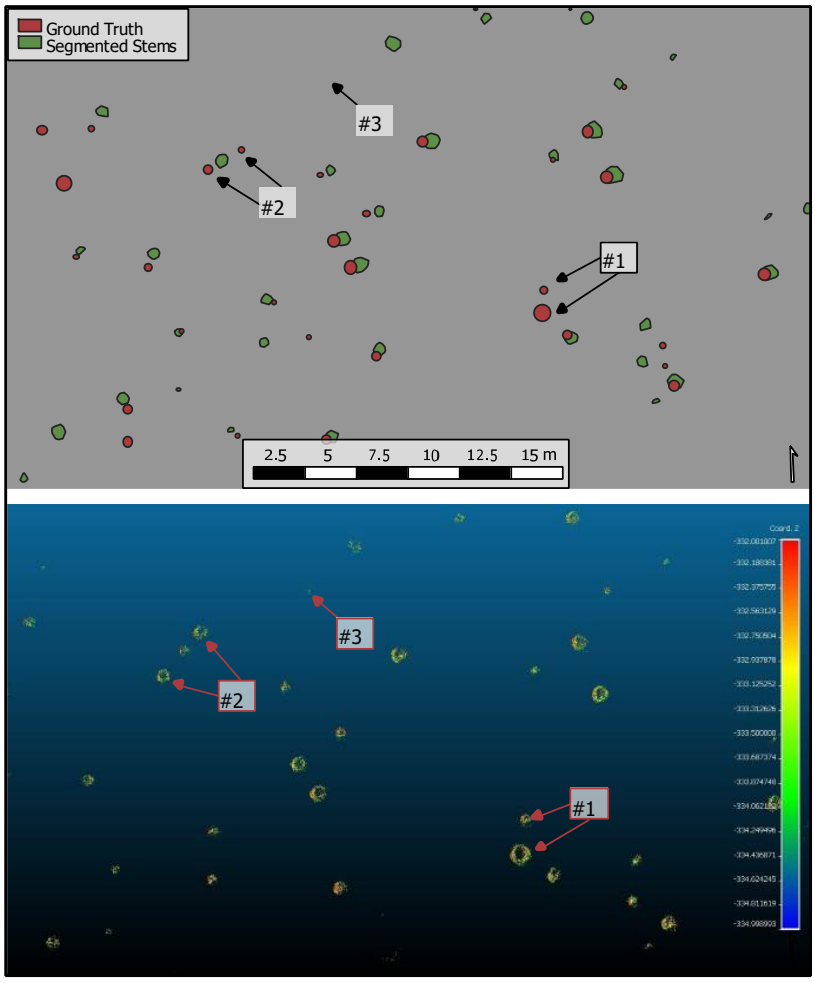

Figure 28: Stem segmentation error examples. Examples #1 and #2 depict missing stems which were detected in the ULS point cloud but were erroneously considered noise. Example #3 depicts a small cluster in the ULS point cloud which was correctly removed from consideration as a stem cluster.

For those examples presented in Figure 24, very large stems were considered noise and consequently removed due to their size being greater than the maximum allowable edge length (max known DBH + 20cm). Since few adjustable parameters inherently exist in DBSCAN, more stems could be adequately clustered and not incorrectly classified as noise if a better method for noise filtering is determined.

Similar studies, including ULS and TLS based studies using DBSCAN, have been capable of detecting >90% of stems present within point clouds [13], [45]. However, a greater assemblage of stem sizes was present in this data than most other studies reviewed [13], [38], [45], [85]. These studies typically surveyed plantations, forests of a homogenous species, or similarly aged stems, rather than the diverse natural forest of our study. Moreover, this study assessed a far greater number of stems in total, where most studies accounted for fewer than a hundred stems. To our knowledge, only two studies, using similar methodologies, had comparably extensive stem counts: Lu et al. with 648 trees [45], Wang et al. with 3986 trees [71], and this study with 1993 trees.

The examples highlighted in Figure 28, in addition to those shown in Figure 24, suggest that a reevaluation of the cluster filtering process is needed to make our detection rates on par with those seen in studies such as Liu et al [38]. Optimization of the detection process using the test area, which lacked large stems, is a major contributing factor to the failure to detecting large stems across the full study area. This however partially reaffirms our initial intentions of developing a method to generate an automated locally scaled DBSCAN ε.

Assessment of Matching Quality

Highlighted by Figure 24, mismatches between field data and detected stems were known to exist within the assessed dataset. Although efforts were taken to address these mismatches, it is evident that the mismatched data is a significant contributing factor that

would partially explain the excessive MAE, and lack of a clear relationship in the presented linear regressions across the full study area.

Ideally, census protocols should use a defined reference point from which stem coordinates are derived. Using a reference location, or backsight, a correctable offset is created. If a backsight using a theodolite is not a viable option, the northern face of a stem could be used, rather than visually estimating locations [7] . Comparable studies did not encounter, or make note of similar issues in data alignment, as many studies adopted censusing strategies utilizing theodolites or differential GPS [13], [20], [24], [45]. Through these studies it has be shown that field data with <1m offset facilitate better methods of data alignment which were not feasible here. The lack of accurate locational information in the ForestGEO dataset therefore caused major mismatches with the ULS data which undermined the correct assessment of DBH estimates.

Assessment of DBH Estimation Results

Although the above results show high levels of MAE, and low levels of correlation, these results are still quite promising. Each method behaved as initially expected, but with more error than desired. Recognizing that the census data used was roughly 2 years old by the time the UAS LiDAR was acquired, this suggests that more current field data could significantly improve these results.

Similar to our findings using the high confidence matches (high confidence threshold), Corte et al. reported a similar correlation coefficient of $r = 0.77$ which only assessed 63 stems using TLS [85]. One source of error in both approaches is the

averaging of sub-slices which are skewed towards fitted outliers. Instead of averaging across the sub-slices, the median may prove to be a better statistic to use in the future in instances of low point density. Liu et al. adopted a method of removing these outliers across the sub-slices using the PauTa criterion [38]. Liu et al. produced a MAPE of 11.54% for stems 11-40cm DBH using RANSAC which was lower than our most stable method LM, which produced an MAPE of 22.08% for stems 18-40cm DBH. Comparatively, our implementation of RANSAC produced an MAPE of 28.06% for stems 18-40cm DBH.

For large stems, our results were difficult to compare to the related literature because very few studies had few, if any, stems >50cm in diameter [36], [38], [77]. No currently known study investigated the potential for any related effect of stem size on DBH estimating methods. Liu et al.'s results, which contained very few stems >50cm DBH (<5 stems), showed some similarity to our results using the high confidence threshold; whereby MAPE is highest for small stems and decreases as DBH increases.

In addition to the error inherently introduced by each DBH estimation method, the maximum accuracy of the LiDAR itself adversely affected our ability to estimate stem size. As noted in the UAS and Sensor Specifications, the Quanergy M8 Ultra had a maximum accuracy of $\pm\ 3cm$, which could contribute to 30% of the error seen in the smallest of stem sizes assessed. Corte et al. achieved similar results to ours using different sensor which was also limited to a geolocation accuracy of $\pm\ 3cm$.

Determination of Optimal DBH Estimation Method

Since these methods have been mostly adapted from studies using much denser TLS point clouds, continued refinement of the DBH estimation techniques is ultimately required to accommodate the less dense ULS data. Pueschel et al. and Liu et al., both of which used TLS and MLS, are the only known, comparable studies to have assessed multiple methods of DBH estimation [38], [77]. Given that the majority of the related literature has adopted iterative methods such as RANSAC and derivatives of LM, it was initially expected that these would perform the best compared to the non-iterative and relatively naïve approaches such as Pt and CH [15], [38], [77]. This hypothesis was only partly supported by our findings which reviewed both iterative and non-iterative approaches. Results attained using LM were the most stable, and certainly performed the best overall in each study area; however, our results also highlighted the potential for a relationship between error and the relative stem-size.

For smaller stems <40cm DBH, which comprised the vast majority of DBH estimation related literature, our results corroborated conclusions made by Liu et al., but not Pueschel et al. [77], [82]. Both Liu et al.'s findings, and ours, found significant relative error, upwards of 50% using RANSAC for stems <20cm. Thus, through our findings, and those made by Liu et al., the least effective methods of measuring stems <40cm in DBH are CH, RANSAC, and Pt. The last method, LM, a method which Pueschel et al. implemented a derivative of, was found to be the best of the four methods reviewed for stems <40cm in diameter and accounted for <25% MAPE.

For large stems >50cm DBH, it has been shown that non-iterative approaches such as Pt and CH produced results with the least amount of MAPE. These methods reliably achieved 5-15% lower MAPE than the iterative approaches. As previously discussed, it is difficult to frame these results in the context of other studies as few comparable works assessed large stems other than Liu et al [38].

Using the confident matches, we can support that non-iterative approaches functioned as the most optimal methods for size estimation until approximately 80cm in diameter. Recognizing concerns presented in “Erroneous Detections and DBH Estimates of Large Stems due to Mismatches” it has been shown that the erroneous removal and mismatching of very large stems >90cm DBH has prohibited our ability to determine which DBH estimation method, or class (iterative or non-iterative) is optimal for the largest of stem sizes present in the study area.

CHAPTER SEVEN: IMPLICATIONS & FUTURE RESEARCH

Future Research and Improvements

Improvements Upon Data Alignment and Field Data

The reliable matching of field data and the ULS point cloud is the foremost issue that must be resolved before any future studies can be performed. As outlined, the geolocation error of the field data precluded traditional methods of data alignment, resulting in the adoption of a manual matching procedure.

One investigator matched the 1892 stems initially detected in 7 hours, a drastic improvement from the multiple months required to generate such data using traditional census data in the field. Since accurate data alignment is paramount to assess the accuracy of any methods adopted, we recommend future researchers invest time in developing more automated approaches of data alignment. Further automation of this process, beyond visual alignment using an RShiny app will facilitate further scaling of related studies beyond individual plot levels.

Improvements to the RShiny app itself should also be implemented if future studies attempt to replicate these findings. In its current state, DBSCAN lacked the ability to sufficiently cluster stems >90cm in diameter. Without these clusters and the situational context of the raw point cloud, mismatches were frequently introduced at larger stem sizes. It is therefore recommended that the point cloud's parent hyperplane should at a minimum be included in any visual matching applications to ensure accurate matches. If

integrated, the context provided by the point cloud could also be used to generate seed points to aid in clustering algorithms [86].

To further address the uncertainty of data alignment present in this study, we also recommend the generation of new field data which specifically aligns with the study requirements. The ForestGEO census data was never intended for validating ULS stem extraction, and the stem locations were only reliable to within 5m of a stem's true location[7]. A future study will employ the same DBH estimation techniques employed here on our own field data using refined census protocols. In fulfillment of the proposed study, a manageably sized, and accessible study site should be created in forested area on George Mason University's Fairfax Campus. The adopted censusing protocol should be developed to specifically target the intended study's geolocation requirements. In our discussion we have suggested that a more accurate approach to determining a stems location could be to use a fixed backsight or point of reference to ensure correctable offsets for stem locations.

Additionally, we can increase the applicability of this study and more accurately report detection rates if all species are considered rather than just the dominant species. Stovall et al. used the ten dominant species because they were known to contain the most biomass in the ForestGEO site, but we rather naively used this filtering process as a means to increase the probability of accurate matches [52]. Artificially filtering the field data for dominant species performed as intended, but also fundamentally introduced bias towards these species. A future study should investigate and quantify the potential bias in biomass estimation or detection rate introduced by the dominant species filtering process.

Improvements for Metric Extraction

Given that our study covered a wider array of species, sizes, and total number of samples than most studies, our findings using these approaches of DBH estimation are encouraging even with the aforementioned flaws in data alignment [38], [77]. Contrary to initial predictions, RANSAC had relatively low performance, although it has already been widely adopted as a core method of DBH estimation in the related literature [16], [32]. Both LM and RANSAC were expected to provide the best results, and behaved quite similarly, but ultimately failed to sufficiently estimate DBH at larger stem sizes. Implementing more robust forms of RANSAC such as LO-RANSAC [87] and assessing if an apparent size dependency is still present (as discussed in Determination of Optimal DBH Estimation Method). Similar to optimizing ε for DBSCAN, LO-RANSAC seeks to provide better model fitting in areas of highly variable in point density (ULS point clouds).

In addition to improved reconstruction techniques, and matching, robust methods of clustering should also be investigated. The current implementation of DBSCAN and ε optimization generated acceptable results across much of the study area, however better noise detection should be pursued. One of the reasons DBSCAN was adopted is that it can serve as an easily comparable baseline for future improvements such as CHAMELEON[88], Mean-Shift [89], OPTICS [90]. Of these, OPTICS would be the most readily implementable as it has already been written into the same ‘dbscan’ packaged used this study.

Concluding Remarks

This study, rather naively, attempted to address a significant research gap by not only using novel methods of handling DBSCAN clustering and edge effects, but also by using the ForestGeo field data beyond its potential accuracy capabilities of 5m. Although we developed a novel method for manual matching of ULS point cloud clusters and the ForestGEO field data, mismatches occurred which affected all subsequent results.

Aside from the study's shortcomings, the research framework and workflows created here will facilitate further research in ULS Lidar data processing and object extraction. Compared to naïve methods such as CH, or algebraic models such as Pt, iterative approaches have been shown to produce more stable results across all stem sizes. Using accurate matches, the optimal method of DBH estimation, LM, could estimate stem size with an R^2 of 0.69.

This study bolsters current literature by comparing methods of individual stem DBH estimation in mixed deciduous broadleaf forests. Ultimately, our insights in individual stem reconstruction will serve to improve future stand-level, allometry-based biomass estimates. Successors to this study will provide rapidly attainable plot level UAS Lidar-derived metrics to supplement regional level satellite-based biomass estimates.

REFERENCES

[1] FAO and UNEP, *The State of the World's Forests 2020*. Rome: FAO and UNEP, 2020. doi: 10.4060/ca8642en.

[2] K. J. Anderson-Teixeira *et al.*, "CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change," *Glob. Change Biol.*, vol. 21, no. 2, pp. 528–549, 2015, doi: 10.1111/gcb.12712.

[3] H. H. Shugart, S. Saatchi, and F. G. Hall, "Importance of structure and its measurement in quantifying function of forest ecosystems," *J. Geophys. Res. Biogeosciences*, vol. 115, no. G2, 2010, doi: 10.1029/2009JG000993.

[4] H. Tang and J. Armston, "GEDI L2B Footprint Canopy Cover and Vertical Profile Metrics," *Univ. Md. NASA*, vol. 1.0, p. 39, Dec. 2019.

[5] N. Camarretta *et al.*, "Monitoring forest structure to guide adaptive management of forest restoration: a review of remote sensing approaches," *New For.*, vol. 51, no. 4, pp. 573–596, Jul. 2020, doi: 10.1007/s11056-019-09754-5.

[6] M. A. Lefsky, W. B. Cohen, G. G. Parker, and D. J. Harding, "Lidar Remote Sensing for Ecosystem StudiesLidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists," *BioScience*, vol. 52, no. 1, pp. 19–30, Jan. 2002, doi: 10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2.

[7] R. Condit, *Tropical Forest Census Plots*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998. doi: 10.1007/978-3-662-03664-8.

[8] K. J. Anderson-Teixeira *et al.*, "Size-related scaling of tree form and function in a mixed-age forest," *Funct. Ecol.*, vol. 29, no. 12, pp. 1587–1602, 2015, doi: 10.1111/1365-2435.12470.

[9] A. Stovall, K. Anderson-Teixeira, and H. Shugart, "Assessing terrestrial laser scanning for developing non-destructive biomass allometry - ScienceDirect," *For. Ecol. Manag.*, vol. 427, pp. 217–229, Jan. 2018, doi: https://doi.org/10.1016/j.foreco.2018.06.004.

[10] P. Pueschel, G. Newnham, G. Rock, T. Udelhoven, W. Werner, and J. Hill, "The influence of scan mode and circle fitting on tree stem detection, stem diameter and volume extraction from terrestrial laser scans," *ISPRS J. Photogramm. Remote Sens.*, vol. 77, pp. 44–56, Mar. 2013, doi: 10.1016/j.isprsjprs.2012.12.001.

[11] Z. Zhen, L. J. Quackenbush, and L. Zhang, "Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data," *Remote Sens.*, vol. 8, no. 4, p. 333, Apr. 2016, doi: 10.3390/rs8040333.

[12] E. Grau, S. Durrieu, R. Fournier, J.-P. Gastellu-Etchegorry, and T. Yin, "Estimation of 3D vegetation density with Terrestrial Laser Scanning data using voxels. A sensitivity analysis of influencing parameters," *Remote Sens. Environ.*, vol. 191, pp. 373–388, Mar. 2017, doi: 10.1016/j.rse.2017.01.032.

[13] S. Tao *et al.*, “Segmenting tree crowns from terrestrial and mobile LiDAR data by exploring ecological theories,” *ISPRS J. Photogramm. Remote Sens.*, vol. 110, pp. 66–76, Dec. 2015, doi: 10.1016/j.isprsjprs.2015.10.007.
[14] X. Zhu *et al.*, “Foliar and woody materials discriminated using terrestrial LiDAR in a mixed natural forest,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 64, pp. 43–50, Feb. 2018, doi: 10.1016/j.jag.2017.09.004.
[15] A. Nurunnabi, Y. Sadahiro, and R. Lindenbergh, “ROBUST CYLINDER FITTING IN THREE-DIMENSIONAL POINT CLOUD DATA,” *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLII-1/W1, pp. 63–70, May 2017, doi: 10.5194/isprs-archives-XLII-1-W1-63-2017.
[16] L. Liu *et al.*, “Single Tree Segmentation and Diameter at Breast Height Estimation With Mobile LiDAR,” *IEEE Access*, vol. 9, pp. 24314–24325, 2021, doi: 10.1109/ACCESS.2021.3056877.
[17] S. Shokirov, S. R. Levick, T. Jucker, P. Yeoh, and K. Youngentob, “Comparison of TLS and ULS Data for Wildlife Habitat Assessments in Temperate Woodlands,” in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, Sep. 2020, pp. 6097–6100. doi: 10.1109/IGARSS39084.2020.9323451.
[18] S. Lamprecht, J. Stoffels, S. Dotzler, E. Haß, and T. Udelhoven, “aTrunk—An ALS-Based Trunk Detection Algorithm,” *Remote Sens.*, vol. 7, no. 8, pp. 9975–9997, Aug. 2015, doi: 10.3390/rs70809975.
[19] M. N. Bazezew, Y. A. Hussin, and E. H. Kloosterman, “Integrating Airborne LiDAR and Terrestrial Laser Scanner forest parameters for accurate above-ground biomass/carbon estimation in Ayer Hitam tropical forest, Malaysia,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 73, pp. 638–652, Dec. 2018, doi: 10.1016/j.jag.2018.07.026.
[20] W. Chen, X. Hu, W. Chen, Y. Hong, and M. Yang, “Airborne LiDAR Remote Sensing for Individual Tree Forest Inventory Using Trunk Detection-Aided Mean Shift Clustering Techniques,” *Remote Sens.*, vol. 10, no. 7, Art. no. 7, Jul. 2018, doi: 10.3390/rs10071078.
[21] A. G. Kamoske, K. M. Dahlin, S. C. Stark, and S. P. Serbin, “Leaf area density from airborne LiDAR: Comparing sensors and resolutions in a temperate broadleaf forest ecosystem,” *For. Ecol. Manag.*, vol. 433, pp. 364–375, Feb. 2019, doi: 10.1016/j.foreco.2018.11.017.
[22] A. Harikumar, F. Bovolo, and L. Bruzzone, “A Local Projection-Based Approach to Individual Tree Detection and 3-D Crown Delineation in Multistoried Coniferous Forests Using High-Density Airborne LiDAR Data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1168–1182, Feb. 2019, doi: 10.1109/TGRS.2018.2865014.
[23] A. L. Neuenschwander and L. A. Magruder, “Canopy and Terrain Height Retrievals with ICESat-2: A First Look,” *Remote Sens.*, vol. 11, no. 14, Art. no. 14, Jan. 2019, doi: 10.3390/rs11141721.
[24] M. L. M. Rudge, S. R. Levick, R. E. Bartolo, and P. D. Erskine, “Modelling the Diameter Distribution of Savanna Trees with Drone-Based LiDAR,” *Remote Sens.*, vol. 13, no. 7, Art. no. 7, Jan. 2021, doi: 10.3390/rs13071266.

[25] K. Liu, X. Shen, L. Cao, G. Wang, and F. Cao, “Estimating forest structural attributes using UAV-LiDAR data in Ginkgo plantations,” *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 465–482, Dec. 2018, doi: 10.1016/j.isprsjprs.2018.11.001.
[26] L. Wallace, A. Lucieer, Z. Malenovský, D. Turner, and P. Vopěnka, “Assessment of Forest Structure Using Two UAV Techniques: A Comparison of Airborne Laser Scanning and Structure from Motion (SfM) Point Clouds,” *For. 19994907*, vol. 7, no. 3, p. 62, Mar. 2016, doi: 10.3390/f7030062.
[27] L. Wallace, A. Lucieer, and C. S. Watson, “Evaluating Tree Detection and Segmentation Routines on Very High Resolution UAV LiDAR Data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7619–7628, Dec. 2014, doi: 10.1109/TGRS.2014.2315649.
[28] L. Wallace, A. Lucieer, and C. S. Watson, “Evaluating Tree Detection and Segmentation Routines on Very High Resolution UAV LiDAR Data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7619–7628, Dec. 2014, doi: 10.1109/TGRS.2014.2315649.
[29] P. Polewski, W. Yao, L. Cao, and S. Gao, “Marker-free coregistration of UAV and backpack LiDAR point clouds in forested areas,” *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 307–318, Jan. 2019, doi: 10.1016/j.isprsjprs.2018.11.020.
[30] J. Picos, G. Bastos, D. Míguez, L. Alonso, and J. Armesto, “Individual Tree Detection in a Eucalyptus Plantation Using Unmanned Aerial Vehicle (UAV)-LiDAR,” *Remote Sens.*, vol. 12, no. 5, p. 885, 2020, doi: 10.3390/rs12050885.
[31] B. Lecigne, S. Delagrange, and C. Messier, “Exploring trees in three dimensions: VoxR, a novel voxel-based R package dedicated to analysing the complex arrangement of tree crowns,” *Ann. Bot.*, vol. 121, no. 4, pp. 589–601, Mar. 2018, doi: 10.1093/aob/mcx095.
[32] D. Kelbe, J. van Aardt, P. Romanczyk, M. van Leeuwen, and K. Cawse-Nicholson, “Single-Scan Stem Reconstruction Using Low-Resolution Terrestrial Laser Scanner Data,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 7, pp. 3414–3427, Jul. 2015, doi: 10.1109/JSTARS.2015.2416001.
[33] J. Lu *et al.*, “Estimation of aboveground biomass of Robinia pseudoacacia forest in the Yellow River Delta based on UAV and Backpack LiDAR point clouds,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 86, p. 102014, Apr. 2020, doi: 10.1016/j.jag.2019.102014.
[34] GOFC-GOLD, *A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gase emissions and removals associated with deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation.* Wageningen University, The Netherlands: GOFC-GOLD Land Cover Project Office, 2016. Accessed: Jun. 03, 2021. [Online]. Available: http://www.gofcgold.wur.nl/redd/sourcebook/GOFC-GOLD_Sourcebook.pdf
[35] H. Zhou, J. Zhang, L. Ge, X. Yu, Y. Wang, and C. Zhang, “Research on volume prediction of single tree canopy based on three-dimensional (3D) LiDAR and clustering segmentation,” *Int. J. Remote Sens.*, vol. 42, no. 2, pp. 738–755, Jan. 2021, doi: 10.1080/01431161.2020.1811917.

[36] D. Wang, V. Kankare, E. Puttonen, M. Hollaus, and N. Pfeifer, “Reconstructing Stem Cross Section Shapes From Terrestrial Laser Scanning,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 272–276, Feb. 2017, doi: 10.1109/LGRS.2016.2638738.
[37] E. Che, J. Jung, and M. J. Olsen, “Object Recognition, Segmentation, and Classification of Mobile Laser Scanning Point Clouds: A State of the Art Review,” *Sens. 14248220*, vol. 19, no. 4, pp. 810–1, Feb. 2019, doi: 10.3390/s19040810.
[38] L. Liu *et al.*, “Single Tree Segmentation and Diameter at Breast Height Estimation With Mobile LiDAR,” *IEEE Access*, vol. 9, pp. 24314–24325, 2021, doi: 10.1109/ACCESS.2021.3056877.
[39] G. G. PARKER, D. J. HARDING, and M. L. BERGER, “A portable LIDAR system for rapid determination of forest canopy structure,” *J. Appl. Ecol.*, vol. 41, no. 4, pp. 755–767, Aug. 2004, doi: 10.1111/j.0021-8901.2004.00925.x.
[40] C. H. Hugenholtz, B. J. Moorman, K. Riddell, and K. Whitehead, “Small unmanned aircraft systems for remote sensing and Earth science research,” *Eos Trans. Am. Geophys. Union*, vol. 93, no. 25, pp. 236–236, 2012, doi: 10.1029/2012EO250005.
[41] M. P. McClelland, J. van Aardt, and D. Hale, “Manned aircraft versus small unmanned aerial system—forestry remote sensing comparison utilizing lidar and structure-from-motion for forest carbon modeling and disturbance detection,” *J. Appl. Remote Sens.*, vol. 14, no. 2, p. 022202, Aug. 2019, doi: 10.1117/1.JRS.14.022202.
[42] L. Cao, H. Liu, X. Fu, Z. Zhang, X. Shen, and H. Ruan, “Comparison of UAV LiDAR and Digital Aerial Photogrammetry Point Clouds for Estimating Forest Structural Attributes in Subtropical Planted Forests,” *Forests*, vol. 10, no. 2, Art. no. 2, Feb. 2019, doi: 10.3390/f10020145.
[43] R. A. Chisholm, J. Cui, S. K. Y. Lum, and B. M. Chen, “UAV LiDAR for below-canopy forest surveys,” *J. Unmanned Veh. Syst.*, vol. 01, no. 01, pp. 61–68, Nov. 2013, doi: 10.1139/juvs-2013-0017.
[44] Y. Lin, J. Hyyppä, and A. Jaakkola, “Mini-UAV-Borne LIDAR for Fine-Scale Mapping,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 426–430, May 2011, doi: 10.1109/LGRS.2010.2079913.
[45] J. Lu *et al.*, “Estimation of aboveground biomass of Robinia pseudoacacia forest in the Yellow River Delta based on UAV and Backpack LiDAR point clouds,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 86, p. 102014, Apr. 2020, doi: 10.1016/j.jag.2019.102014.
[46] L. I. Duncanson, B. D. Cook, G. C. Hurtt, and R. O. Dubayah, “An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems,” *Remote Sens. Environ.*, vol. 154, pp. 378–386, Nov. 2014, doi: 10.1016/j.rse.2013.07.044.
[47] R. Dubayah *et al.*, “The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth’s forests and topography,” *Sci. Remote Sens.*, vol. 1, p. 100002, Jun. 2020, doi: 10.1016/j.srs.2020.100002.
[48] B. Blair, D. Rabine, and M. Hofton, “The Laser Vegetation Imaging Sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation

and topography," *Photogramm. Remote Sens.*, vol. 54, pp. 115–122, 1999, doi: 10.1016/S0924-2716(99)00002-7.
[49] G. Vosselman, B. G. H. Gorte, G. Sithole, and T. Rabbani, "RECOGNISING STRUCTURE IN LASER SCANNER POINT CLOUDS," p. 6.
[50] B. Gorte and N. Pfeifer, "STRUCTURING LASER-SCANNED TREES USING 3D MATHEMATICAL MORPHOLOGY," p. 5.
[51] J. Hackenberg, H. Spiecker, K. Calders, M. Disney, and P. Raumonen, "Forests | Free Full-Text | SimpleTree —An Efficient Open Source Tool to Build Tree Models from TLS Clouds | HTML," *MDPI Forests*, Nov. 23, 2015. https://www.mdpi.com/1999-4907/6/11/4245/html#sec2dot2-forests-06-04245 (accessed Jan. 28, 2020).
[52] A. E. L. Stovall, K. J. Anderson-Teixeira, and H. H. Shugart, "Assessing terrestrial laser scanning for developing non-destructive biomass allometry," *For. Ecol. Manag.*, vol. 427, pp. 217–229, Nov. 2018, doi: 10.1016/j.foreco.2018.06.004.
[53] J. R. Kellner *et al.*, "New Opportunities for Forest Remote Sensing Through Ultra-High-Density Drone Lidar," *Surv. Geophys. Dordr.*, vol. 40, no. 4, pp. 959–977, Jul. 2019, doi: http://dx.doi.org.mutex.gmu.edu/10.1007/s10712-019-09529-9.
[54] L. Duncanson, O. Rourke, and R. Dubayah, "Small Sample Sizes Yield Biased Allometric Equations in Temperate Forests," *Sci. Rep.*, vol. 5, p. 17153, Nov. 2015, doi: 10.1038/srep17153.
[55] C. Wang, M. Ji, J. Wang, W. Wen, T. Li, and Y. Sun, "An Improved DBSCAN Method for LiDAR Data Segmentation with Automatic Eps Estimation," *Sensors*, vol. 19, no. 1, Jan. 2019, doi: 10.3390/s19010172.
[56] N. A. Bourg, W. J. McShea, J. R. Thompson, J. C. McGarvey, and X. Shen, "Initial census, woody seedling, seed rain, and stand structure data for the SCBI SIGEO Large Forest Dynamics Plot," *Ecology*, vol. 94, no. 9, pp. 2111–2112, 2013, doi: 10.1890/13-0010.1.
[57] K. J. Anderson-Teixeira *et al.*, "CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change," *Glob. Change Biol.*, vol. 21, no. 2, Art. no. 2, 2015, doi: 10.1111/gcb.12712.
[58] L. Wallace, C. Watson, and A. Lucieer, "Detecting pruning of individual stems using Airborne Laser Scanning data captured from an Unmanned Aerial Vehicle," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 30, pp. 76–85, Aug. 2014, doi: 10.1016/j.jag.2014.01.010.
[59] DJI, "Matrice_600_Pro_User_Manual_v1.0_EN_1208.pdf." Da-Jiang Innovations, Apr. 17, 2018. Accessed: Jun. 27, 2021. [Online]. Available: https://dl.djicdn.com/downloads/m600%20pro/1208EN/Matrice_600_Pro_User_Manual_v1.0_EN_1208.pdf
[60] DJI, "M200_User_Manual_EN_20201120.pdf." Da-Jiang Innovations, Nov. 20, 2018. Accessed: Jun. 27, 2021. [Online]. Available: https://dl.djicdn.com/downloads/M200/20201120/M200_User_Manual_EN_20201120.pdf
[61] Quanergy Systems Inc., "M8 LiDAR Sensor." Jan. 28, 2020.

[62] “Quanergy | M Series High Performance 360 Degree LiDAR Sensor,” *Quanergy*. https://quanergy.com/products/m8/ (accessed Jun. 27, 2021).
[63] Fagerman Technologies INC., *Scanlook PC*. Somerville, AL, USA: Fagerman Technologies INC.
[64] LiDAR USA, “Operation of Scanlook PC.” 2017. Accessed: Nov. 10, 2020. [Online]. Available: http://wiki.lidarusa.com/lib/exe/fetch.php?media=operation_of_scanlook_pc.pdf
[65] LiDAR USA, “Operation of Inertial Explorer.” 2017. Accessed: Jun. 02, 2021. [Online]. Available: http://wiki.lidarusa.com/lib/exe/fetch.php?media=operation_of_inertial_explorer.pdf
[66] “CloudCompare - Open Source project.” https://www.danielgm.net/cc/ (accessed Nov. 13, 2020).
[67] L. Wallace, A. Lucieer, C. Watson, and D. Turner, “Development of a UAV-LiDAR System with Application to Forest Inventory,” *Remote Sens.*, vol. 4, no. 6, Art. no. 6, Jun. 2012, doi: 10.3390/rs4061519.
[68] R. Jean-Romain and A. David, *Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*. 2021. [Online]. Available: https://cran.r-project.org/package=lidR
[69] J.-R. Roussel *et al.*, “lidR: An R package for analysis of Airborne Laser Scanning (ALS) data,” *Remote Sens. Environ.*, vol. 251, p. 112061, Dec. 2020, doi: 10.1016/j.rse.2020.112061.
[70] G. Fan *et al.*, “A New Quantitative Approach to Tree Attributes Estimation Based on LiDAR Point Clouds,” *Remote Sens.*, vol. 1779, no. 12, p. 20, Jun. 2020, doi: https://doi.org/10.3390/rs12111779.
[71] X. Wang, Y. Zhang, and Z. Luo, “Combining Trunk Detection With Canopy Segmentation to Delineate Single Deciduous Trees Using Airborne LiDAR Data,” *IEEE Access*, vol. 8, pp. 99783–99796, 2020, doi: 10.1109/ACCESS.2020.2995389.
[72] M. Z. Rodriguez *et al.*, “Clustering algorithms: A comparative approach,” *PLoS ONE*, vol. 14, no. 1, Jan. 2019, doi: 10.1371/journal.pone.0210236.
[73] M. Hahsler, M. Piekenbrok, S. Arya, and D. Mount, *dbscan*. 2021. [Online]. Available: https://github.com/mhahsler/dbscan
[74] G. Fan *et al.*, “A New Quantitative Approach to Tree Attributes Estimation Based on LiDAR Point Clouds,” *Remote Sens.*, vol. 12, no. 11, Art. no. 11, Jan. 2020, doi: 10.3390/rs12111779.
[75] V. Pratt, “Direct least-squares fitting of algebraic surfaces,” in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’87*, Not Known, 1987, pp. 145–152. doi: 10.1145/37401.37420.
[76] Ali Al-Sharadqah and Nikolai Chernov, “Error analysis for circle fitting algorithms,” *Electron. J. Stat.*, vol. 3, no. none, pp. 886–911, Jan. 2009, doi: 10.1214/09-EJS419.
[77] P. Pueschel, G. Newnham, G. Rock, T. Udelhoven, W. Werner, and J. Hill, “The influence of scan mode and circle fitting on tree stem detection, stem diameter and volume extraction from terrestrial laser scans,” *ISPRS J. Photogramm. Remote Sens.*, vol. 77, pp. 44–56, Mar. 2013, doi: 10.1016/j.isprsjprs.2012.12.001.

[78] N. Chernov and C. Lesort, “Least Squares Fitting of Circles,” *J. Math. Imaging Vis.*, vol. 23, no. 3, pp. 239–252, Nov. 2005, doi: 10.1007/s10851-005-0482-8.
[79] J. Gama and N. Chernov, *conicfit: Algorithms for Fitting Circles, Ellipses and Conics Based on the Work by Prof. Nikolai Chernov*. 2015. Accessed: Jun. 17, 2021. [Online]. Available: https://CRAN.R-project.org/package=conicfit
[80] I. Ladrón de Guevara, J. Muñoz, O. D. de Cózar, and E. B. Blázquez, “Robust Fitting of Circle Arcs,” *J. Math. Imaging Vis.*, vol. 40, no. 2, pp. 147–161, Jun. 2011, doi: 10.1007/s10851-010-0249-8.
[81] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981, doi: 10.1145/358669.358692.
[82] K. Liu, X. Shen, L. Cao, G. Wang, and F. Cao, “Estimating forest structural attributes using UAV-LiDAR data in Ginkgo plantations,” *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 465–482, Dec. 2018, doi: 10.1016/j.isprsjprs.2018.11.001.
[83] A. Nurunnabi, Y. Sadahiro, and R. Lindenbergh, “ROBUST CYLINDER FITTING IN THREE-DIMENSIONAL POINT CLOUD DATA,” *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLII-1/W1, pp. 63–70, May 2017, doi: 10.5194/isprs-archives-XLII-1-W1-63-2017.
[84] L. Mariga, *leomariga/pyRANSAC-3D*. 2021. Accessed: Jul. 15, 2021. [Online]. Available: https://github.com/leomariga/pyRANSAC-3D/blob/34f2e7cb5561b45ec8083940ffc5f3d959da9804/pyransac3d/circle.py
[85] A. P. D. Corte, “Measuring Individual Tree Diameter and Height Using GatorEye High-Density UAV-LiDAR in an Integrated Crop-Livestock-Forest System,” Jul. 2020.
[86] J. Lu *et al.*, “Estimation of aboveground biomass of Robinia pseudoacacia forest in the Yellow River Delta based on UAV and Backpack LiDAR point clouds,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 86, p. 102014, Apr. 2020, doi: 10.1016/j.jag.2019.102014.
[87] O. Chum, J. Matas, and J. Kittler, “Locally Optimized RANSAC,” in *Pattern Recognition*, Berlin, Heidelberg, 2003, pp. 236–243. doi: 10.1007/978-3-540-45243-0_31.
[88] G. Karypis and V. Kumar, “CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling,” p. 22.
[89] W. Chen, X. Hu, W. Chen, Y. Hong, and M. Yang, “Airborne LiDAR Remote Sensing for Individual Tree Forest Inventory Using Trunk Detection-Aided Mean Shift Clustering Techniques,” *Remote Sens.*, vol. 10, no. 7, Art. no. 7, Jul. 2018, doi: 10.3390/rs10071078.
[90] M. Daszykowski, B. Walczak, and D. L. Massart, “Looking for Natural Patterns in Analytical Data. 2. Tracing Local Density with OPTICS,” *J. Chem. Inf. Comput. Sci.*, vol. 42, no. 3, Art. no. 3, May 2002, doi: 10.1021/ci010384s.

BIOGRAPHY

During his undergraduate degree, Daniel Spiwak briefly studied endangered species conservation at the Smithsonian Conservation Biology Institute (SCBI). During this time, he leveraged his prior experience using drones and cursory R knowledge to begin working on forestry related research with the GIS Lab at SCBI. Following his brief stay at SCBI, and an undergraduate remote sensing course, his mentors: Dr. Konrad Wessels and Dr. Qiongyu Huang, encouraged him to pursue a graduate degree which would further pursue this research. The research documented here served as an ideological successor to the work which began in the spring of 2018.