



University of Kentucky
UKnowledge

Theses and Dissertations--Biosystems and
Agricultural Engineering

Biosystems and Agricultural Engineering


2021

ASSESSING THE USE OF LIDAR AND UAV TECHNOLOGY FOR MONITORING GROWING ALFALFA

Stuart Tucker Sheffield

University of Kentucky, tuckshed@yahoo.com

Author ORCID Identifier:

 <https://orcid.org/0000-0001-9783-4295>

Digital Object Identifier: <https://doi.org/10.13023/etd.2021.340>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Sheffield, Stuart Tucker, "ASSESSING THE USE OF LIDAR AND UAV TECHNOLOGY FOR MONITORING GROWING ALFALFA" (2021). *Theses and Dissertations--Biosystems and Agricultural Engineering*. 80. https://uknowledge.uky.edu/bae_etds/80

This Master's Thesis is brought to you for free and open access by the Biosystems and Agricultural Engineering at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Biosystems and Agricultural Engineering by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Stuart Tucker Sheffield, Student

Dr. Joseph Dvorak, Major Professor

Dr. Donald Colliver, Director of Graduate Studies

ASSESSING THE USE OF LIDAR AND UAV TECHNOLOGY
FOR MONITORING GROWING ALFALFA

THESIS

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Biosystems and Agricultural Engineering in the
Colleges of Agriculture and Engineering
at the University of Kentucky

By

Stuart Tucker Sheffield

Lexington, Kentucky

Director: Dr. Joseph Dvorak, Professor of Biosystems and Agricultural Engineering

Lexington, Kentucky

2021

Copyright © Stuart Tucker Sheffield 2021
<https://orcid.org/0000-0001-9783-4295>

ABSTRACT OF THESIS

ASSESSING THE USE OF LIDAR AND UAV TECHNOLOGY FOR MONITORING GROWING ALFALFA

Alfalfa is a popularly grown crop because of its value as a nutritious feed source for livestock. The efficient production of an alfalfa crop relies on the monitoring of certain parameters, like height, quality, and yield. Traditionally, producers have used manual measurements of alfalfa plant height to estimate the nutritive quality and yield of a growing alfalfa crop. Manual measurements of plant height are often labor intensive and provide low resolution data that is not acceptable for full field scale assessment of growing alfalfa. The two studies presented in this thesis offer detailed insight into the rapid and accurate monitoring of alfalfa with LiDAR and UAV technologies. The first study explores the use of a simple single beam LiDAR sensor to accurately estimate the average canopy height and yield of an alfalfa crop. Predictive models of alfalfa canopy height were developed and evaluated to find the optimal LiDAR derived measurements to use. The resulting measurements were then used to build predictive models of yield, and the best yield model was determined. The best models of canopy height and yield both incorporated the 95th percentile of LiDAR derived canopy height as a single explanatory variable. The second study assesses the field conditions, flight parameters, and general statistical descriptors that should be considered for the stable collection and application of UAV derived canopy height information. Data taken from different alfalfa fields at different flight parameters with different statistical processing were all compared. General canopy height distribution statistics from UAV flights flown at or below 50 m with nadir and oblique camera angles over thick stands of alfalfa were determined to be reliable for the detection and application of the alfalfa canopy surface. Using these determined methods, predictive models of canopy height and yield were generated and compared. The best model of average canopy height used the 50th percentile of UAV derived canopy height from an UAV flight at 30 m in a nadir imaging configuration. The best model of yield used the 95th percentile from an UAV flight at 50 m in an oblique imaging configuration.

KEYWORDS: Alfalfa, Unmanned Aerial Vehicle, LiDAR, Canopy Height Modeling, Photogrammetry, Flight Parameters

Stuart Tucker Sheffield

08/02/2021

Date

ASSESSING THE USE OF LIDAR AND UAV TECHNOLOGY
FOR MONITORING GROWING ALFALFA

By
Stuart Tucker Sheffield

Dr. Joseph Dvorak

Director of Thesis

Dr. Donald Colliver

Director of Graduate Studies

08/02/2021

Date

DEDICATION

To Alexa Sheffield, my beautiful and benevolent wife. Our marriage began just before embarking on the journey to complete this thesis, and our marriage has been a constant source of joy, hope, and comfort throughout this journey.

ACKNOWLEDGMENTS

First and foremost, I would like to thank God for blessing me with life, purpose, and the opportunity to perform this research at the University of Kentucky. Second, I would like to thank my thesis advisor, Dr. Joseph Dvorak. From our very first email correspondence to finalizing this thesis, Dr. Dvorak has been nothing but kind and thoughtful. His guidance, leadership, and friendship have been invaluable to the creation of this thesis and to my success here at the University of Kentucky. I would also like to thank the other members of my advisory committee: Dr. Joshua Jackson and Dr. Chris Teutsch. Both provided expert insight that guided and challenged my thinking, which inevitably improved the quality of this thesis. I thank all of my fellow students that assisted me in the completion of this research: Bo Smith, Cameron Minch, and Cynthia Arnold. I also want to recognize the entire faculty and staff of the Biosystems and Agricultural Engineering Department. Their kindness and professionalism provided a fun, safe, and constructive work environment for me to succeed in my graduate studies. Lastly, I would like to thank my parents, Greg and Mindy Sheffield, and my brother, Tyler Sheffield, for their physical, emotional, and financial support. I love my family with my whole heart, and I know that none of this would be possible without them.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER 1. Introduction	1
CHAPTER 2. Using LiDAR to Measure Alfalfa Canopy Height and Yield.....	4
2.1 Introduction.....	4
2.2 Methods.....	8
2.2.1 Data Collection.....	8
2.2.2 Data Processing.....	13
2.2.3 Modeling	14
2.3 Results.....	16
2.3.1 Modeling Canopy Height	16
2.3.2 Modeling Yield.....	19
2.4 Discussion.....	21
2.4.1 Predictive Models of Canopy Height	21
2.4.2 Predictive Models of Yield.....	22
2.5 Conclusion	23
CHAPTER 3. Determining Stable Methods of Generating and Applying UAV Derived canopy Height Models for Alfalfa Monitoring	25
3.1 Introduction.....	25
3.2 Methods.....	32
3.2.1 Field Conditions	32
3.2.2 Photogrammetry Data Collection	35
3.2.3 Data Processing.....	37
3.2.3.1 Photogrammetry Software Processing	37
3.2.3.2 Point Cloud Processing.....	38
3.2.4 Statistical Analysis	40
3.2.5 Predictive Modeling	41
3.3 Results.....	42
3.3.1 Canopy Height Model Stability.....	42
3.3.2 Predictive Models of Canopy Height and Yield.....	45
3.4 Discussion.....	49
3.4.1 Canopy Height Model Stability.....	50
3.4.2 Predicting Canopy Height and Yield.....	52
3.5 Conclusion	53
CHAPTER 4. Conclusion.....	55

APPENDICES	57
<i>[APPENDIX 1. POINT CLOUD PROCESSING CODE].....</i>	<i>57</i>
<i>[APPENDIX 2. CANOPY HEIGHT PREDICTIVE MODELING CODE]</i>	<i>62</i>
<i>[APPENDIX 3. YIELD PREDICTIVE MODELING CODE]</i>	<i>65</i>
REFERENCES	68
VITA.....	73

LIST OF TABLES

Table 2.1 Weed pressure evaluation scale	11
Table 2.2 Insect and disease damage evaluation scale	11
Table 2.3 Best linear models of average canopy height and their properties	19
Table 2.4 Best models of yield and their properties	21
Table 3.1 Flight parameters tested in each of the two fields.	37
Table 3.2 Experimental variable for creating canopy height models	41
Table 3.3 R ² values between statistical descriptors collected in field 1	43
Table 3.4 R ² values between statistical descriptors collected in field 2	43
Table 3.5 Average R ² matrix for flight parameters.....	44
Table 3.6 Linear regression models of measured average canopy height with different statistical predictors from the 30-90° flight parameter	46
Table 3.7 Linear regression models of measured average canopy height using the 50th percentile at different flight parameters	47
Table 3.8 Different model types using the 50th percentile from the 30-90° flight parameter as a predictor for average canopy height	47
Table 3.9 Highest performing models of yield	49

LIST OF FIGURES

Figure 2.1 This is a satellite image of field 1 at the University of Kentucky's North Farm. The yellow polygon denotes the boundaries of the field.	9
Figure 2.2 This is a satellite image of field 2 at University of Kentucky's North Farm. The yellow shows the boundary of the field.	9
Figure 2.3 This image shows one of the twenty 1 m ² quadrats that were used for data collection. The PVC structure is a square with an area of 1 m ² , raised 1 m above the ground surface.	11
Figure 2.4 This figure shows the orientation of the LiDAR sensor and the plane in which the sensor collects data points. The sensing plane is represented by a circle around the LiDAR sensor. The green portion of the circle represents the relevant data that is used for further processing.	12
Figure 2.5 This image depicts how the 3D scans of the canopy at each quadrat were collected. The LiDAR sensor was attached to a frame and mounted onto the quadrat to acquire each scan.	13
Figure 2.6 The histograms show the distribution of the LiDAR-derived canopy height of quadrat 1 on 05/14/19: one with outliers (a) and one with the outliers filtered out (b).	14
Figure 2.7 This plot shows the relationship between the 95th percentile LiDAR heights and observed average canopy heights.	17
Figure 2.8 This is a goodness of fit plot between the observed average canopy height and the predicted average canopy height from the model described by Equation 2.1.	18
Figure 2.9 This plot shows the relationship between the 95th percentile of LiDAR derived canopy height and manually measured yield. This plot also shows the predictions from the optimal yield model.	20
Figure 2.10 This is a goodness of fit plot showing the relationship between the observed yield and the yield predictions from the optimal yield model	20
Figure 3.1 This flowchart shows the process of creating CHMs from UAV imagery.	26

Figure 3.2 This figure depicts the fields of view from two UAVs as they cross over a field from right to left. The UAV at the top of the figure is in a nadir imaging configuration, and the UAV at the bottom is in an oblique configuration.	31
Figure 3.3 The image depicts field 1. The area highlighted in blue is the alfalfa field, and the green lines represent the path of the UAV.....	33
Figure 3.4 This is an aerial image of field 2. The region highlighted in blue shows the extent of field 2, and the green lines depict the flight path of the UAV.....	33
Figure 3.5 This time series plot shows the average stand density and standard deviation for field 1 and 2 during the summer of 2019.	34
Figure 3.6 This time series plot shows the average yield and standard deviation for field 1 and 2 during the summer of 2019.	35
Figure 3.7 This image shows one of the quadrats that was used in the data collection process. It is a simple PVC structure with a 1 m ² area elevated 1 m from the ground surface.....	36
Figure 3.8 This is a top view of a point cloud depicting a sampling area imaged with the UAV. The points in green represent the annulus of points that were segmented out from each dataset. The red points were disregarded and not used in any further analysis. The dark red square in the middle of the annulus is the top of the quadrat.....	39
Figure 3.9 The Gaussian curve plots show the probability density of UAV-derived canopy height of each flight condition for quadrat 7 in field 1 (a) and quadrat 20 in field 2 (b) on 06/04/2019.	43
Figure 3.10 This figure shows two scatterplots depicting the simple linear regression model described in Equation 3.1 (a) and the goodness of fit for the model (b)....	45
Figure 3.11 This figure shows two scatterplots depicting the KNN yield model using the 95th percentile of canopy height acquired by a UAV at 50-75° (a) and the goodness of fit of that model (b).....	49
Figure 3.12 This scatterplot shows a profile view of the point cloud describing a sampling area in field 2 on 06/04/2019. The quadrat has been left in the point cloud to help visualize the stitching error due to a thin alfalfa stand. Points are color-coded based on height on the z axis.	51

CHAPTER 1. INTRODUCTION

Alfalfa has amassed an international popularity as one of the premier crops for grazing applications and hay/haylage production (Lacefield, Henning, Rasnake, & Collins, 1997). It has gained this popularity because of its reputation for being a nutritious feed for livestock. In the U.S., current estimates place alfalfa among the top 3 field crops in terms of marketing year average prices and among the top 6 field crops in terms of value of production (USDA-NASS, 2021). Major industries, such as dairy and equine, rely on the efficient production of alfalfa (Grev, Wells, Sheaffer, & Martinson, 2017; Martin et al., 2017).

A key factor in alfalfa production systems is harvest scheduling. When planning a harvest, producers are primarily concerned with the yield and nutritive quality of the crop. Often times, producers have a target forage quality, commonly defined by the relative feed value (RFV) of the alfalfa crop, that they are trying to achieve at harvest (Undersander, 2011). This strategy allows for more economical and efficient grazing systems and/or hay production because the quality of the alfalfa can be specifically catered to the nutrient requirements of various types of livestock (Lacefield, 1988). This way of harvesting alfalfa requires thorough assessment and management because missing a scheduled harvest date, even by a few days, can cause the alfalfa harvested to be of a lower quality. As a means to that end, producers have, traditionally, defined a cutting frequency by looking at recommended calendar dates and stage of maturity. Although this has worked in the past, defining a cutting frequency in this way is less likely to account for the current conditions of a specific alfalfa crop and the dynamic environmental factors surrounding that crop.

Certain assessment techniques have been adopted by producers to quantify the yield and quality of a growing alfalfa crop. Producers have employed methods like visual quality inspection, the predictive equations of alfalfa quality (PEAQ) method, and the Robel pole method (Lacefield, Henning, Collins, & Swetnam, 1996; Smith, 2008; Undersander, 2011; Vittetoe & Lang, 2019). These methods depend on manual measurements that can require significant amounts of labor and time to collect. Although alfalfa producers have found success in using these methods, the results of these methods can fluctuate depending on the person performing the measurement (Lacefield, 1988). Even if the measurements are performed perfectly, they only offer low-resolution data that is not suitable for field scale monitoring of the alfalfa.

With the advent of remote sensing technologies and the various platforms that they can be affixed to, autonomous field scale monitoring of alfalfa is a possibility. With technology like LiDAR sensors and unmanned aerial vehicles (UAVs) equipped with RGB sensors, researchers have been able to monitor crop parameters like canopy height and yield in cotton, wheat, bermudagrass, and barley (Bendig et al., 2014; Feng, Zhang, Sudduth, Vories, & Zhou, 2019; Jimenez-Berni et al., 2018; Pittman, Arnall, Interrante, Moffet, & Butler, 2015). Researchers were able to achieve this type of crop monitoring by generating canopy height models (CHMs) from the returns of the remote sensing technologies, such as 3 dimensional point clouds. Since current methods of forage crop assessment, PEAQ and Robel pole, rely at least partially on plant height measurements, sensor derived CHMs stand to benefit alfalfa producers by improving the spatial and temporal resolution in which canopy height measurements can be taken and, in turn, improve the monitoring of production factors, like yield and quality. Before these

technologies can be used for alfalfa applications, research into the optimal parameters to use from the returns of the sensors and the best methodologies to use in collecting the remotely sensed data must be conducted.

The overall goal of this thesis was to explore the application of LiDAR and UAV technologies to monitor growing alfalfa. This research was split into two studies: one focusing on LiDAR technology (Chapter 2) and another focusing on UAV technology (Chapter 3). The primary objective for the LiDAR study was to assess the capability of a simple LiDAR sensor to accurately estimate the average canopy height of an alfalfa crop and perform yield estimations. The primary objective of the UAV study was to determine the field conditions, flight parameters, and descriptive statistics stable enough to use in the generation of robust alfalfa canopy height models and predictive models of alfalfa production variables at the field scale.

CHAPTER 2. USING LiDAR TO MEASURE ALFALFA CANOPY HEIGHT AND YIELD

2.1 Introduction

Alfalfa is a nutrient rich, perennial crop that has commonly been used as feed for various livestock (dairy cows, horses, etc.). Alfalfa also forms a symbiotic relationship with rhizobium bacteria in which nitrogen from the air is converted into a plant available form and fixed into the soil, which is advantageous for producers when rotating crops (Oke & Long, 1999). Due to these attributes, alfalfa is the most widely grown legume worldwide (Cumo, 2013). Being able to timely and accurately monitor quality and yield is essential for profitable production of alfalfa. One popular way of assessing alfalfa is the PEAQ method. This method utilizes measurements of plant height and maturity to estimate relative feed value, which is later used in harvest scheduling (Vittetoe & Lang, 2019). Researchers have also found significant relationships between alfalfa height and yield that can help producers make management decisions (Lyons, Undersander, Welch, & Donnelly, 2016). A common factor in assessing an alfalfa crop is measuring the alfalfa's height.

Traditionally, alfalfa stand height has been measured manually with a marked rod. For PEAQ analysis, growers choose 5 representative plots within a field and measure the tallest alfalfa plant within each plot (Vittetoe & Lang, 2019). Alfalfa stand height has also commonly been defined as the average of multiple rod measurements within a certain area (e.g. 1 m² region). These traditional methods are labor intensive and use limited sampling points that provide height data at a small resolution not suitable for accurate monitoring of alfalfa at the field scale.

Researchers have used remote sensing techniques to improve the acquisition of height data. O. Payero, M. U. Neale, and L. Wright (2004) analyzed the use of vegetation indices, derived from spectral reflectance measurements, to estimate alfalfa plant height. Strong relationships ($R^2 > 0.90$) were found between the vegetation indices and plant height. Utilizing similar techniques, Noland et al. (2018) used spectral reflectance values in conjunction with air temperature and LiDAR canopy height measurements to estimate alfalfa yield and quality. Ultrasonic sensors have also been used to remotely measure plant height for assessing forage yield (Fricke, Richter, & Wachendorf, 2011). One type of remote sensing technology that has not been as thoroughly researched for alfalfa applications is LiDAR. There has been research that mentions LiDAR's potential in the field and uses LiDAR measurements to compliment other remotely sensed data (Noland et al., 2018), but there is a lack of literature investigating LiDAR as a primary means of measuring alfalfa plant height for crop monitoring purposes.

With today's LiDAR sensors and platforms, such as UAVs, LiDAR data can be retrieved from a farm field at previously unattainable spatial and temporal resolutions. With such accurate and timely data, end users can start to make real time farm management decisions. The usual return from a LiDAR scan is a 3D point cloud. This point cloud is made up of multiple data points the sensor collected in and around a plant canopy. Each point represents the measured distance between the LiDAR sensor and the canopy structure (Rosell & Sanz, 2012). This point cloud can then be processed into a digital surface model (DSM). The actual plant canopy is derived from this DSM by crop height modeling (CHM) (Crommelinck & Höfle, 2016). Canopy height modeling requires a digital elevation model (DEM) of the bare soil. Bare soil DEMs can be

obtained from LiDAR scans taken when there is no vegetation. The DEM of the bare soil is then effectively subtracted from the DSM to produce a CHM (Jimenez-Berni et al., 2018).

LiDAR has been used for accurately measuring the plant height of certain crops. The lidar derived plant height of crops such as wheat, hairy vetch, and miscanthus has been shown to be highly correlated to actual plant height measurements (Jimenez-Berni et al., 2018; Wiering, Ehlke, & Sheaffer, 2019; L. Zhang & Grift, 2012). L. Zhang and Grift (2012) compared manual measurements of stem height and average field height to LiDAR derived stem and average field heights. LiDAR measurements were taken statically and dynamically. Statically, the LiDAR's average accuracy was 94.92%. Dynamically, the LiDAR's average accuracy was 96.2%. These findings suggest that using LiDAR to measure plant height is plausible. Jimenez-Berni et al. (2018) used a LiDAR sensor affixed to a ground-based platform to achieve high throughput plant phenotyping of wheat. A R^2 value of 0.99 and a RMSE of 0.017 m was observed between manual canopy height and LiDAR canopy height. Throughout the literature, manual and LiDAR measured plant heights are notably related.

Although numerous use cases for LiDAR derived canopy height have been well defined, the optimum process for converting raw LiDAR outputs into applicable canopy height data is still in question. An important consideration in creating CHMs is the filtering of the point cloud. Proper filtering can be challenging, but if done correctly, it will give a better representation of the canopy height (Song & Wang, 2019). Jimenez-Berni et al. (2018) filtered their LiDAR data by analyzing the frequency distribution of sensor-recorded height. They looked at the associated RMSE between manually

measured canopy height and different quantile values, ranging from 0.8 to 1.0, of LiDAR canopy height of wheat. Within the range of tested quantiles, 0.955 was found to be the optimum quantile to represent the top of the wheat canopy because it exhibited the lowest RMSE (Jimenez-Berni et al., 2018). In another study in wheat, Madec et al. (2017) found that the 99.5 percentile best represents the top of the canopy. By using a robotic rover equipped with a LiDAR sensor, the researchers were able to laser scan microplots of wheat to create CHMs. Using the 99.5 percentile to filter the CHMs, a RMSE value of 3.5 cm was obtained between manually and LiDAR derived plant height. Spatial variability of each microplot was also minimized by the 99.5 percentile filter. Ultimately, the researchers conclude that the 99.5 percentile of the cumulated height distribution is the optimum percentile to use when comparing LiDAR measurements to ground measurements.

Finding an optimal way of collecting alfalfa canopy height data would not only impact plant height acquisition, but it could also impact how accurately the yield of an alfalfa crop can be estimated. LiDAR derived plant height has been used as a proxy for the yield of certain crops. Pittman et al. (2015) found that plant height derived from a laser sensor, similar to LiDAR, was notably correlated to the destructively sampled biomass of bermudagrass ($R = 0.88$). Other researchers have had similar success with LiDAR based yield estimations of crops like wheat and miscanthus (Eitel, Magney, Vierling, Brown, & Huggins, 2014; Mathanker, Maughan, Hansen, Grift, & Ting., 2014). LiDAR derived plant height in conjunction with plant health measurements have also been used to effectively estimate yield in plants like tall fescue (Schaefer & Lamb, 2016). Once canopy height data can be reliably collected from LiDAR sensors, the estimation of

more production focused parameters, like yield, can be performed and analyzed for effectiveness.

Before more advanced implementations of LiDAR sensors, such as a LiDAR equipped UAV, can be routinely used for alfalfa applications, the accuracy of plant height measurements from simpler LiDAR implementations must be tested. Not only should the LiDAR measurements be validated, but an optimal set of parameters representing the top of the alfalfa canopy should be defined to ensure the effective acquisition of plant height data and alfalfa yield estimations. To fill this knowledge gap, alfalfa canopy height models were developed using data from a single beam LiDAR sensor. Second, model performances were compared, and the most efficient LiDAR data index to use for the accurate prediction of alfalfa canopy height was determined. Third, yield models were created from the resulting variables of the canopy height modeling. Fourth, the yield models were compared, and the best model was determined.

2.2 Methods

2.2.1 Data Collection

Data were collected during the 2019 growing season at the University of Kentucky's North Farm. Sampling took place at two alfalfa fields referred to here as field 1 (Figure 2.1), with a 7.03 ha area, and field 2 (Figure 2.2), with a 3.09 ha area. Field 1 was approximately located at (38.128688, -84.509497), while field 2 was approximately located at (38.118587, -84.509612). Field 1 had three different soil types: Armour silt loam (Fine-silty, mixed, active, thermic Ultic Hapludalfs), Egam silt loam (Fine, mixed, active, thermic Cumulic Hapludolls), and Huntington silt loam (Fine-silty,

mixed, active, mesic Fluventic Hapludolls). Field 2 had Armour silt loam, Huntington silt loam, and Bluegrass-Maury silt loam (Fine-silty, mixed, active, mesic Typic Paleudalfs). Field 1 was planted with a reduced lignin and glyphosate resistant variety, Ameristand 400 HVXRR (Forage Genetics International, LLC; Nampa, ID), and field 2 was planted with a different glyphosate resistant variety, Allied 428RR (Allied Seed, LLC; Nampa, ID).



Figure 2.1 This is a satellite image of field 1 at the University of Kentucky's North Farm. The yellow polygon denotes the boundaries of the field.



Figure 2.2 This is a satellite image of field 2 at University of Kentucky's North Farm. The yellow shows the boundary of the field.

Data were collected on five separate dates in the 2019 season: May 14, May 17, May 21, May 28, and June 6. This allowed for data to be acquired throughout one growth cycle, between the first and second cutting of the 2019 season. The first alfalfa cutting was executed with a John Deere 630 Discbine (Moline, Illinois) on May 7, 2019. The alfalfa was then collected as haylage on May 9, 2019. The second cutting took place several days after data collection ended. During the data collection period, herbicide was applied once on May 21, 2019.

On each sampling date, twenty 1 m² quadrats (Figure 2.3) were placed in the fields. Ten quadrats were used in field 1, which equated to a sample density of 1.42 samples per hectare, and ten quadrats were used in field 2, resulting in 3.24 samples per hectare. Data were taken at each quadrat location. All reported data came from within the boundaries of the quadrats. Average plant height within each quadrat were measured using a marked rod. Weed pressure, insect damage, and disease damage were also observed and ranked using evaluation tables (Table 2.1-2.2). Lastly, yield per plot was determined by clipping the alfalfa in each plot to a residual height of approximately 2.5 cm, drying the samples, and weighing the samples.



Figure 2.3 This image shows one of the twenty 1 m² quadrats that were used for data collection. The PVC structure is a square with an area of 1 m², raised 1 m above the ground surface.

Table 2.1 Weed pressure evaluation scale

Value	Weed Pressure
0	Less than 5% weeds present
1	5% - 20% weeds
2	20% - 40% weeds
3	40% - 60% weeds
4	60% - 80% weeds
5	Greater than 80% weeds

Table 2.2 Insect and disease damage evaluation scale

Value	Insect/Disease Pressure
0	Less than 5% insect/disease damage present
1	5% - 20% insect/disease damage
2	20% - 40% insect/disease damage
3	40% - 60% insect/disease damage
4	60% - 80% insect/disease damage
5	Greater than 80% insect/disease damage

Once all the manual field measurements were collected, a Scanse Sweep (Scanse; San Leandro, California, USA) LiDAR sensor was used to acquire single line scans of the canopy. This sensor emits a single beam of light, while rotating 360°, to measure distance and create a point cloud of the surrounding environment (Figure 2.4). The sensor was affixed to a frame that was mounted to each quadrat (Figure 2.5). The frame latched to the quadrat and was designed to place the LiDAR sensor 1 m above the top of the quadrat and directly over the center of the quadrat. Due to this design, the sensor was 2 m above

ground level when taking measurements. Using this ground-based frame rather than a UAV ensured exact placement of the LiDAR system for repeatable and precise comparisons with manually collected samples. During data acquisition, the sensor was allowed to scan the crop canopy for 3 to 4 revolutions, which amounted to approximately 70 data points per quadrat. Flights at 2 m above the ground is near the lowest feasible altitude for safe flights, but it was also near the highest possible height for a ground-based frame to position the LiDAR device precisely, repeatably and reliably. In very flat terrain, UAV flights that scan an entire field may be able to use this 2 m altitude, but it is likely that these flights will ultimately occur at higher altitudes. For flights at these higher altitudes, it may be necessary to use alternative LiDAR units, such as the Velodyne Puck LITE (Velodyne Lidar, San Jose, California, USA), that can record equivalent point clouds while flying at higher altitudes.

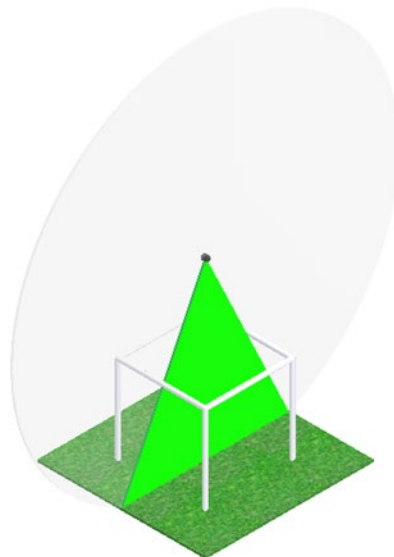


Figure 2.4 This figure shows the orientation of the LiDAR sensor and the plane in which the sensor collects data points. The sensing plane is represented by a circle around the LiDAR sensor. The green portion of the circle represents the relevant data that is used for further processing.



Figure 2.5 This image depicts how the 3D scans of the canopy at each quadrat were collected. The LiDAR sensor was attached to a frame and mounted onto the quadrat to acquire each scan.

2.2.2 Data Processing

All the LiDAR data were processed in MATLAB. Since the sensor captures data in a 360° sweep, the first step in processing each LiDAR scan was to filter out any points that came from outside the quadrat boundary. Once all the points from each quadrat scan were isolated, the data were converted from polar to cartesian coordinates based on the fixed geometric relationships of the quadrat, sensor, and the ground. Before statistical data could be extracted from the processed data, outliers within the canopy height model had to be removed. The Generalized Extreme Studentized Deviate (GESD) test was used to account for outliers by assuming that the number of outliers for each model was no more than 5% of the total data points. As a final step, the distribution of each dataset was visualized on a histogram (Figure 2.6) and manually checked to ensure that all steps of the processing were done successfully. Once proper processing was confirmed, mean canopy height, maximum canopy height, standard deviation, and various percentile values were calculated from each canopy height model.

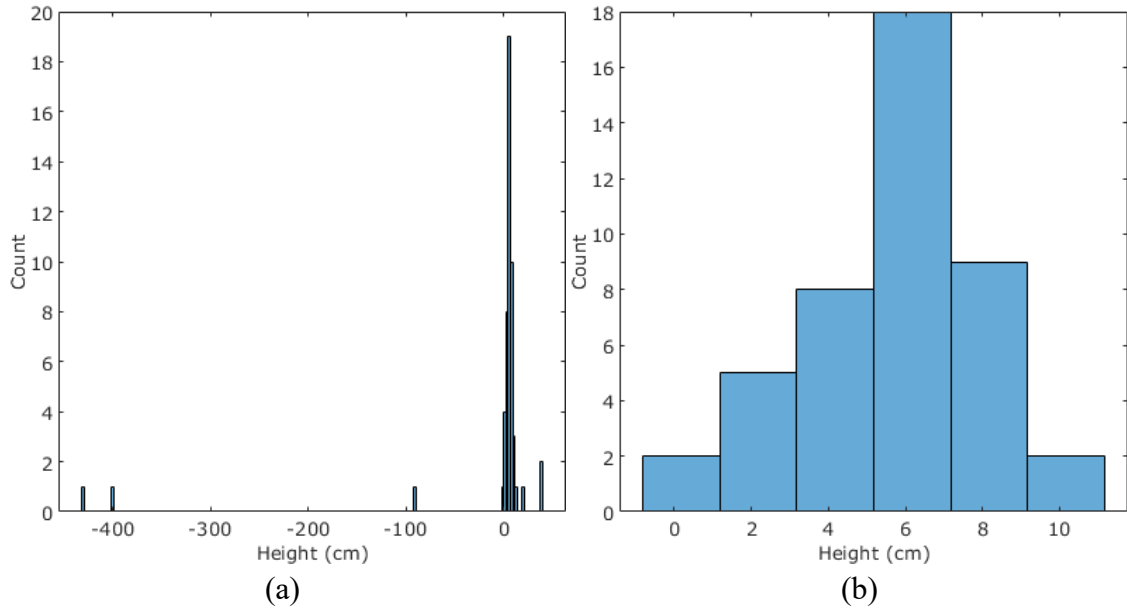


Figure 2.6 The histograms show the distribution of the LiDAR-derived canopy height of quadrat 1 on 05/14/19: one with outliers (a) and one with the outliers filtered out (b).

2.2.3 Modeling

The goal of this modeling step was to find the best model for predicting the average height of the canopy (as measured manually) based on the LiDAR data. The LiDAR data was a distribution of heights, and statistical descriptors for each quadrat sample were created in the data processing step. It was unclear which descriptors or combination of descriptors would be most useful in estimating the actual average height of the canopy. Models were tested based on each single descriptor and on various combinations of descriptors. It was not feasible to consider every possible combination of the descriptors, so testing with more than 2-descriptors was limited to descriptors that appeared promising on their own or in 2-descriptor models.

The modeling process took place within MATLAB. To begin, the manually measured and LiDAR-derived alfalfa data were imported into MATLAB's regression learner application. Simple and multiple linear regression modeling was performed with

the entire dataset. Manually measured average canopy height was chosen as the response variable for all the models. The predictor variables that were used in the modeling process consisted of the nine different LiDAR derived height descriptors: average, maximum, standard deviation, 25th percentile, 50th percentile, 75th percentile, 90th percentile, 95th percentile, and 99th percentile. Certain field observations were also incorporated into the models: weed, insect, and disease pressure. Five-fold cross validation was performed on each model to validate the modeling results. This form of validation is a k-fold technique that divides the entire dataset into $k = 5$ groups where four groups are used for training, and one is used for testing. This process is repeated $k = 5$ times to ensure that each group is used for testing once. The model evaluation metrics, R^2 and RMSE, from all 5 rounds of testing were averaged to create the general model statistics. This validation technique was chosen over the more traditional method of splitting the data into one training and one testing set because of the robustness of cross validation. This cross validation technique allowed the models to be trained and tested on the entire dataset instead of subsets of the dataset.

Each LiDAR derived height descriptor (average, maximum, 95th percentile, etc.) was used as a single predictor to create 9 unique linear models to predict actual average canopy height. Once every LiDAR height variable was tested individually, each pair of LiDAR height descriptors were tested. This resulted in the creation of 36 unique models. It was not feasible to continue checking every combination of descriptors. Only descriptors that were used in models with an R^2 above 0.88 were used in the next stage of modeling. Once models were found that consistently performed well ($R^2 \geq .90$ and $RMSE \leq 4.5$ cm) by using any combination of LiDAR height descriptors, field

observations were added to the models to see if these field observations of weed pressure, pest pressure or disease pressure could improve the model.

Once the best models of average canopy height were determined, the variables utilized in each of those models were used in the modeling of the sampled yield data. This modeling was not limited to linear regression models. Each determined set of explanatory variables were implemented into 19 different modeling techniques which can be classified into 5 categories: linear regression, support vector machine, regression trees, gaussian process, and ensemble of trees. Once all of the models were trained and tested by 5-fold cross validation, the models with the highest R^2 values were chosen and reported in the results section below.

2.3 Results

2.3.1 Modeling Canopy Height

The optimal LiDAR statistical descriptor to use for predicting average canopy height by itself was the 95th percentile value from the LiDAR derived canopy height distribution. Performing linear regression analysis with the 95th percentile and actual average canopy height as the predictor and response variables, respectfully, resulted in a R^2 of 0.90 and RMSE of 4.5 cm. Consequently, the model that was produced from the analysis fit the data points well (Figure 2.7).

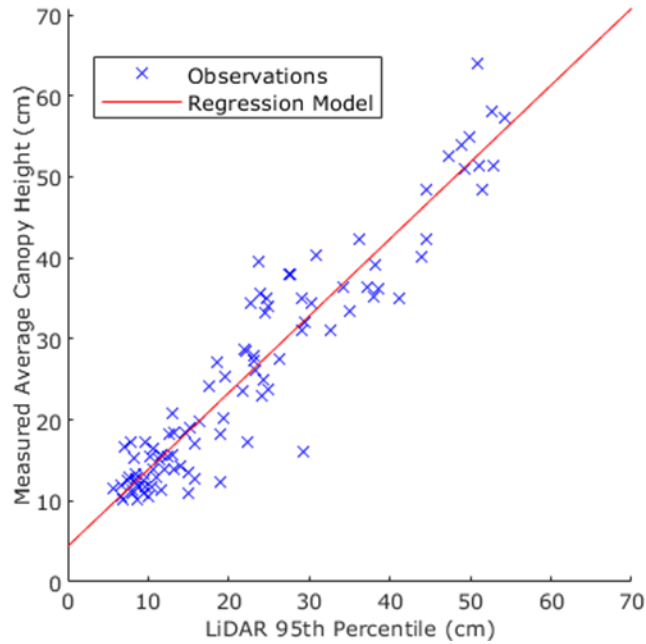


Figure 2.7 This plot shows the relationship between the 95th percentile LiDAR heights and observed average canopy heights.

The above regression model is quantitatively described by:

$$h_p = 0.949l_{95} + 4.38 \tag{2.1}$$

where

h_p = predicted average canopy height (cm)

l_{95} = 95th percentile of the height distribution from the LiDAR scan (cm)

The model can also be represented/visualized by comparing the model estimates from the LiDAR data with the actual observations (Figure 2.8). The model performs well, and most predictions are close to the actual measured heights. Errors appear evenly distributed at different heights, so the linear model is an appropriate fit.

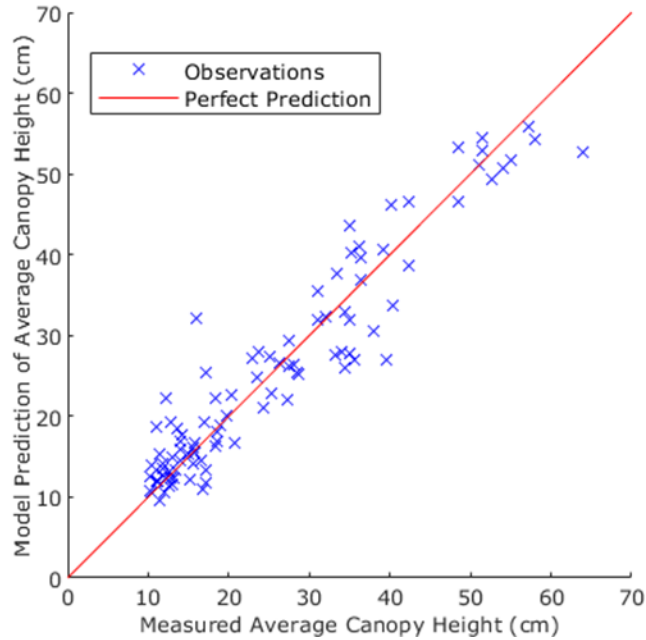


Figure 2.8 This is a goodness of fit plot between the observed average canopy height and the predicted average canopy height from the model described by Equation 2.1.

Although the 95th percentile was found to be the optimal LiDAR measurement to use, models with other LiDAR derived measurements were tested for comparison. All other models that utilized a single LiDAR measurement (e.g. 99th percentile alone) or two LiDAR measurements (e.g. 50th percentile and maximum height together) had R^2 values less than 0.90 and higher RMSE. Most models that used more than two LiDAR measurements or that included alfalfa health observations also had R^2 values less than 0.90. There were a couple of models that had slightly higher R^2 values. A model using the 25th, 50th and 95th percentiles had slight improvements in RMSE and R^2 (Table 2.3). The best model that was found included three LiDAR measurements (25th percentile, 50th percentile, maximum height) and two alfalfa health descriptors (insect and disease pressure). However, these models are much more complex than the basic model using the

95th percentile and only provided slight improvements in predictive accuracy as shown by the small differences in RMSE and R^2 .

Table 2.3 Best linear models of average canopy height and their properties

Variables	RMSE (cm)	R^2
95 th percentile	4.5	0.90
25 th , 50 th , 95 th percentiles	4.1	0.91
25 th & 50 th percentile, max height, insect, disease	3.9	0.92

2.3.2 Modeling Yield

The best predictive model of yield was a fine gaussian support vector machine that utilized the 95th percentile of LiDAR derived canopy height (Figure 2.9). The model was able to achieve a R^2 of 0.75. This indicates that 75% of the variation that is present in the yield data can be explained by the single predictor model. Consequently, the estimated yield values from the model were close to the observed yield values with the exception of a couple of predictions when the observed yield was at its highest (Figure 2.10). Other models were tested with the variables determined in the above section (Table 2.4). Although most models performed moderately well ($R^2 \geq 0.65$), there was not a combination of variables (LiDAR and/or alfalfa health) or a particular modeling technique that could outperform the support vector machine model using only the 95th percentile.

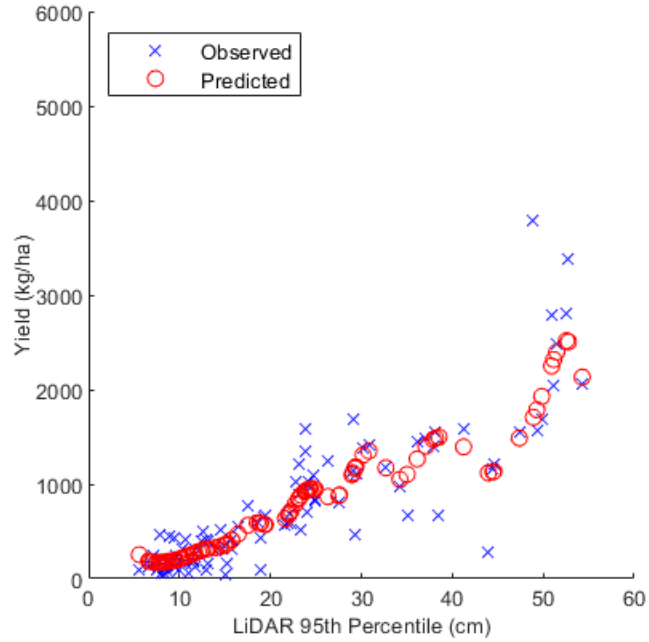


Figure 2.9 This plot shows the relationship between the 95th percentile of LiDAR derived canopy height and manually measured yield. This plot also shows the predictions from the optimal yield model.

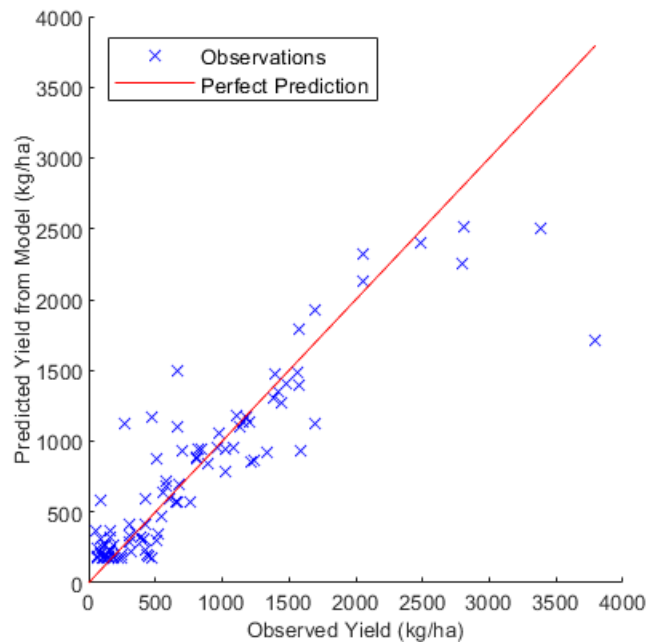


Figure 2.10 This is a goodness of fit plot showing the relationship between the observed yield and the yield predictions from the optimal yield model

Table 2.4 Best models of yield and their properties

Variables	Types	RMSE (kg/ha)	R ²
95 th percentile	Fine Gaussian SVM ^a	376	0.75
25 th ,50 th ,95 th percentiles	Linear SVM	401	0.71
25 th & 50 th percentile, max height, insect, disease	Linear SVM	402	0.71

^a support vector machine

2.4 Discussion

2.4.1 Predictive Models of Canopy Height

All the models presented in this study produced accurate predictions of average canopy height. Even with the various complexities of each model, the model accuracies were all close to each other. The largest difference between R² values was .02, while the largest difference in RMSE was 0.6 cm. With all the models performing to a similar degree of accuracy, the amount of labor that went into data collection and variable calculation is a deciding factor on which model is the best. The basic model performed well and only required one variable, the 95th percentile of the scanned LiDAR plant heights. Producing this model required a minimal amount of labor. The single LiDAR measurement model only required data that could be rapidly collected from the LiDAR sensor. Once the LiDAR data were processed, statistics, like the 95th percentile, were easily and quickly attainable.

The model that utilized a combination of LiDAR measurements to predict average canopy height performed slightly better than the single variable model, but it was also more complex. The data collection process for the combination model was the same as the single variable model. This allowed for easy and quick data acquisition without much

manual labor. Software could easily handle calculating the additional statistical descriptors of the scanned LiDAR plant heights, so computational time was not a major issue with the additional complexity. However, given the minimal improvement, the simpler model appears more appropriate. If this system is deployed more broadly, it would be wise to monitor for occasions when the model performs poorly and see if these alternative models turn out to be more robust.

Adding field observations to the models had only minor benefits to the overall accuracy of the models. Although the model using LiDAR measurements and field observations exhibited the best R^2 and RMSE values, this model required significantly more labor to produce. In conjunction with LiDAR measurements, the status of the crop within the quadrats had to be manually observed and measured (Table 2.1- 2.2). Requiring manual field observations would make deployment of a LiDAR system for field scale monitoring much more difficult. The model's accuracy is only marginally better than the other models that require less inputs, so it is hard to justify this model for practical applications. It may still be worth monitoring this system in broader deployments to make sure that corrections for weed pressure, pest pressure and disease pressure are not necessary in particular extreme cases.

2.4.2 Predictive Models of Yield

The variables that were determined to be the best to use in simple linear regression models of canopy height were also used to a moderate degree of success in yield modeling. All models of yield presented in this study had moderately high R^2 values ($R^2 \geq 0.70$). The model using the 95th percentile of LiDAR derived canopy height performed better than the rest of the tested models. This finding is not only important because of the

relatively high R^2 value ($R^2 = 0.75$), but that such a high value can be achieved with the use of one LiDAR derived height variable. The other models use various combinations of LiDAR variables and manual observations, but they still under-performed when compared to the single variable model. Because of its high performance and simplicity, the support vector machine model with the 95th percentile can be recommended for use in yield monitoring. The other yield models presented here should still be considered in future cases of yield modeling, especially if conditions of the alfalfa fields are significantly different from the conditions in this study.

2.5 Conclusion

This study shows that the rapid and accurate acquisition of alfalfa canopy height and prediction of yield are possible by means of a simple single beam LiDAR sensor. The canopy height and yield data that can be modeled using these methods can inform end-users on the status of the alfalfa and harvest timing. From the LiDAR data presented in this study, three predictive models of alfalfa canopy height were developed. The simplest model was comprised of a single LiDAR height measurement, the 95th percentile. The other two models incorporated a combination of LiDAR height percentiles and LiDAR height data with field observations, respectfully. The prediction accuracies of all three models were high ($R^2 \geq 0.90$ and $RMSE < 4.5$ cm) and close to each other in value. The model incorporating the 95th percentile was able to achieve the same level of accuracy as the other models by using fewer explanatory variables. Similarly, three predictive models of alfalfa yield were developed from the results of the alfalfa canopy height modeling process. All three models performed moderately well ($R^2 \geq 0.70$), but the best model of yield was determined to be a fine gaussian support vector machine using the 95th

percentile of LiDAR derived canopy height. Under the conditions of this experiment, the 95th percentile value of LiDAR derived canopy height can be recommended as a good indicator of actual alfalfa canopy height and as an explanatory variable for yield modeling.

CHAPTER 3. DETERMINING STABLE METHODS OF GENERATING AND APPLYING UAV DERIVED CANOPY HEIGHT MODELS FOR ALFALFA MONITORING

3.1 Introduction

Researchers are using canopy height models (CHMs) for a variety of crop production purposes. Song and Wang (2019) used CHMs to simply measure the height of a winter wheat canopy. Yanbo (2016) used CHMs to estimate the yield of a cotton crop. Canopy height models have also been used to aid in plant breeding efforts of tomatoes and sorghum (Enciso et al., 2019; Watanabe et al., 2017). Above ground biomass of barley has also been accurately predicted using data derived from CHMs (Bendig et al., 2014). Numerous crops have been monitored using CHMs, but few researchers have explored using CHMs to monitor alfalfa.

Alfalfa is often labeled the “Queen of the Forages” because it is considered one of the most widely grown legumes in the world (Cumò, 2013). One reason for alfalfa’s popularity is how nutritious it is for livestock, such as horses and cattle. The dairy industry greatly depends on the efficient production of forage crops, like alfalfa, as a feed source for dairy cows (Martin et al., 2017). Another reason for alfalfa’s popularity is its ability to form a symbiotic relationship with rhizobium bacteria that allows for atmospheric nitrogen to be fixed into a plant available form (Oke & Long, 1999). Due to this, alfalfa often improves the performance and yield of subsequent crops grown in rotation with the alfalfa (Yost, Coulter, Russelle, Sheaffer, & Kaiser, 2012).

Profitable production of alfalfa greatly relies on harvest timing. Certain tradeoffs exist between alfalfa yield and alfalfa quality that must be considered when planning to harvest alfalfa (Undersander, 2011). Traditional methods of assessing alfalfa have

required manually carrying poles through a field and taking different types of canopy measurements like the length of the longest stem, PEAQ method, or the height at which the canopy reaches sufficient density to provide visual obstruction, Robel pole method. The structure of an alfalfa canopy is complex, and these traditional measurements of different types of canopy height have been proven to be stable and valuable in estimating various properties of growing alfalfa (Smith, 2008; Vittetoe & Lang, 2019). On the other hand, these traditional methods come with major downfalls. They can require a significant amount of labor and time to complete, and they use a small number of sampling points that ultimately give low-resolution data of the alfalfa field. Having a timely way of assessing yield and quality at the field scale would greatly benefit alfalfa producers. Additionally, having an innovative way of measuring these quantities will be useful in judging the merit of new cultivars, like reduced lignin alfalfa (Cherney, Smith, Sheaffer, & Cherney, 2020). Canopy height models derived from UAV imagery could be a solution (Figure 3.1).

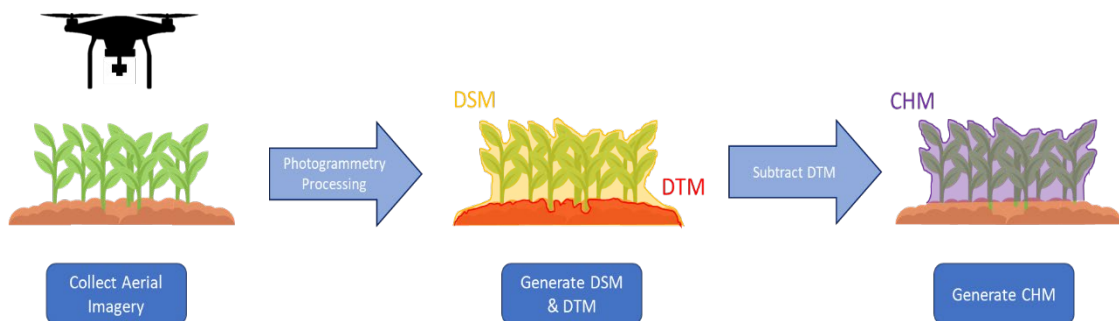


Figure 3.1 This flowchart shows the process of creating CHMs from UAV imagery.

Researchers have proven that remotely sensed data can be used in alfalfa applications at the sub-field scale where data collection was localized to the specific sampling area within a field as opposed to the entire field. Noland et al. (2018) were able

to effectively estimate the yield and nutritive value of growing alfalfa with a handheld sensor that measured canopy reflectance and a LiDAR sensor that measured canopy height. Destructive sampling of various alfalfa plots at different growth stages were taken, and the yield/nutritive data rendered from the samples were used as response variables in predictive models. The models that utilized the remote sensing data as explanatory variables achieved high R^2 values between 0.81 and 0.90. Similarly, Dvorak et al. (2021) were able to use remotely sensed data from a UAV to predict the quality and yield of a growing alfalfa crop at a sub-field scale. Instead of spectral reflectance or LiDAR returns, they focused on photogrammetrically processing overlapping RGB images of alfalfa plots into CHMs. They found that alfalfa yield and nutritive value can be predicted using the mean canopy height and standard deviation derived from the CHMs at the sub-field scale. By using the mean canopy height, standard deviation, and measures of field health as explanatory variables, the researchers were able to predict yield and nutritive values with R^2 values around 0.80. These studies serve as proofs of concept for using CHMs to monitor growing alfalfa, but further research should be conducted to assess the factors that impact CHM generation at the field scale. Before implementing the techniques mentioned in the above articles for field-scale monitoring of alfalfa production, it is necessary to determine which statistical analyses, flight parameters, and field conditions provide a stable CHM that could be considered for further processing to identify more production relevant points of information, such as nutritive value and yield.

When using a UAV to collect data for a CHM, the current standard flight controls maintain a constant height above ground level as determined at the launch site. In large

fields in regions with even moderate slopes or hills, there can easily be a difference of tens of meters from one region of the field to another. If conservation practices like terracing are used (common in the Great Plains), these elevation changes can occur quickly and may even be captured in the same image, which could prevent even new UAVs from adjusting the height to match ground contours. Until UAVs offer the ability to perfectly maintain a desired height above the ground while rapidly flying over a field, it is critical to assess whether the data that are used to build CHMs are sensitive to these differences. This is an important step in moving from experimental research to a widespread application by producers in many different regions.

CHMs can be constructed in different ways from the raw photogrammetry or LiDAR point clouds. The general approach is to collect all the points in a given area and perform a statistical analysis to assign a single value for that area. Statistics like the mean, median, 75th percentile, 99.5th percentile, and max canopy height have all been used to build predictive models (Bendig et al., 2014; Chang, Jung, Maeda, & Landivar, 2017; Madec et al., 2017; Watanabe et al., 2017; Wijesingha, Moeckel, Hensgen, & Wachendorf, 2019). Although many of these statistics have been used to varying degrees of success in canopy height modeling, little research has been performed on which statistic would be the most stable to use in monitoring alfalfa. While researchers have focused on establishing CHMs that provide a single measurement for a given area of a field, the true canopy structure in a crop like alfalfa is much more complex. A healthy stand of alfalfa quickly forms a closed canopy with many stems extending from each crown where they attach to the roots. These stems of various lengths are covered with leaves and reach different heights above the ground. Given the complexity of the alfalfa

canopy structure and the variety of methods in which it has been traditionally measured, there is a need to determine appropriate methods for generating CHMs from UAV images.

Even before statistics can be derived from a CHM, certain factors must be considered during the image collection process. One of these factors is the field condition. Point clouds can look different depending on the conditions of the field being imaged. High-resolution CHMs are extremely sensitive to even very small changes in the ecosystem structure (Cunliffe, Brazier, & Anderson, 2016). In particular, H. Zhang et al. (2018) found that stand density can have a major effect on the CHM generation of grasslands. They concluded that an accurate point cloud could not be generated from UAV imagery if the canopy was too dense or too thin. Similarly, Zahawi et al. (2015) found that substantial errors occurred within CHMs of vegetation exhibiting a relatively low canopy height and lack of height variation within the canopy. These differences in field conditions can affect the values for the various statistical descriptors of the point clouds even over the same locations at the same time. Seeing how field conditions, like stand density, can affect the creation of alfalfa CHMs would help producers and researchers determine if their operations would benefit from CHMs.

Other factors that can have a drastic effect on the outcome of UAV derived CHMs are the flight parameters. Research has shown that factors like flight altitude and camera gimbal angle can significantly impact the 3D reconstruction of a scene (Jaud et al., 2019; Mesas-Carrascosa, García, De Larriva, & García-Ferrer, 2016). With flight altitude, there exists an inverse relationship between altitude and image resolution. Acquiring high resolution imagery typically requires low flight altitude. Certain disadvantages of low

altitude flights are longer flight durations and increased data storage requirements (Mesas-Carrascosa et al., 2016). For certain crop monitoring applications, extremely high spatial resolution is not necessary, and, for those specific use cases, higher altitudes can be used for increased efficiency (Mesas-Carrascosa et al., 2015). Determining an optimal flight altitude or a range of altitudes for which alfalfa CHMs can be efficiently and reliably generated would give guidance to UAV operators imaging alfalfa.

A lot of UAV missions are flown in a nadir imaging configuration, but there have been various studies that have shown the utility of flying with camera angles off from nadir (Figure 3.2). James and Robson (2014) found that using oblique imagery can reduce errors, like occlusion, in surface models. Another advantage of using oblique imagery is that it can give a better view of the vertical relief of the vegetation structure (Lin, Wang, Ma, & Lin, 2018). Other researchers have found that combining images, taken in nadir and off nadir configurations of the same scene, can reduce errors in the surface modeling of complex surfaces (Tu et al., 2021). Determining the optimal gimbal angle or combination of gimbal angles that would create stable CHMs would be valuable information for future UAV users in monitoring their alfalfa crop.

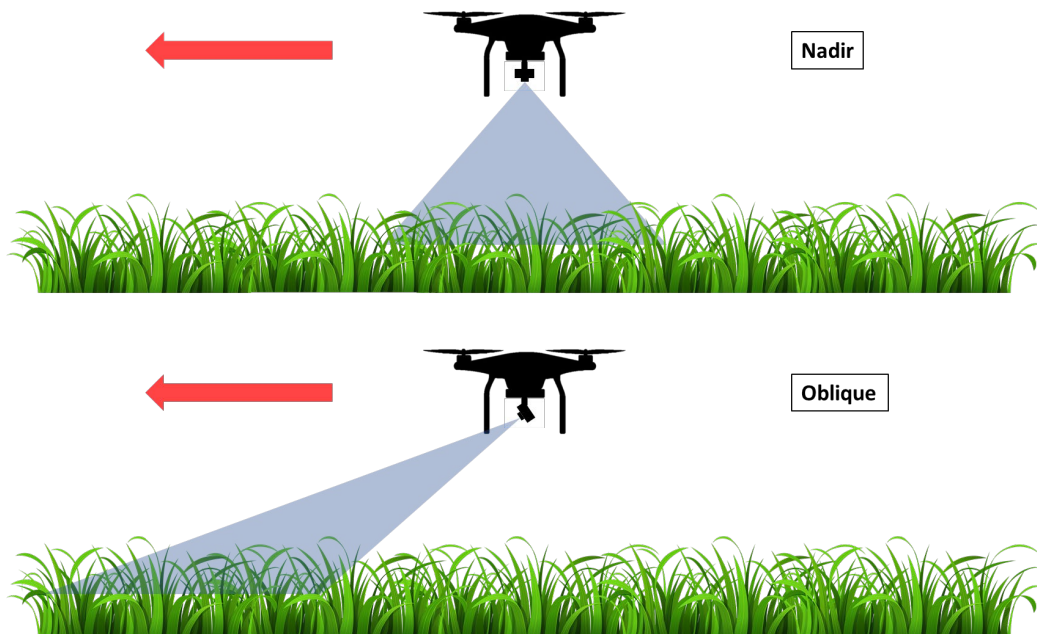


Figure 3.2 This figure depicts the fields of view from two UAVs as they cross over a field from right to left. The UAV at the top of the figure is in a nadir imaging configuration, and the UAV at the bottom is in an oblique configuration.

For practical application, it is important to select stable methods to create the CHM. The CHM needs to be robust and not sensitive to variations in flight or field conditions. A robust method may require not using certain statistical descriptors derived from the CHM or precluding the use of the method with certain field conditions or flight parameters. The goal of this project was to determine the field conditions, flight parameters, and statistical processing methods that would provide a stable, robust CHM and that should be considered when creating predictive models of more valuable, production-focused data at the field scale. To accomplish this goal, data taken from a thick stand and a thin stand of alfalfa was compared. Second, correlations between canopy height data taken at different flight parameters were assessed. Third, correlations made between common statistical descriptors of modeled canopy height were evaluated.

Lastly, predictive models of canopy height and yield were developed from the data deemed stable enough for use in modeling.

3.2 Methods

3.2.1 Field Conditions

Two alfalfa fields were used during this study. Both fields were located at the University of Kentucky's North Farm with field 1 (Figure 3.3) approximately located at (38.128688, -84.509497) and field 2 (Figure 3.4) at (38.118587, -84.509612). Field 1 had an area of 7.02 ha, while field 2 had an area of 3.09 ha. Each field exhibited different soil types and was planted with a different variety of alfalfa. The soil in field 1 consisted of Armour silt loam (Fine-silty, mixed, active, thermic Ultic Hapludalfs), Egam silt loam (Fine, mixed, active, thermic Cumulic Hapludolls), and Huntington silt loam (Fine-silty, mixed, active, mesic Fluventic Hapludolls). Ameristand 400 HVXRR (Forage Genetics International, LLC; Nampa, ID), a reduced lignin and glyphosate resistant variety of alfalfa, was planted in field 1. Field 2 contained Armour silt loam, Huntington silt loam, and Bluegrass-Maury silt loam (Fine-silty, mixed, active, mesic Typic Paleudalfs). Field 2 was planted with Allied 428RR (Allied Seed, LLC; Nampa, ID), a glyphosate resistant variety of alfalfa. The first cut of the alfalfa took place on May 7, 2019 with a John Deere 630 Discbine (Moline, Illinois). The second cutting was performed after data collection had stopped. Herbicide was applied once, on May 21, 2019.



Figure 3.3 The image depicts field 1. The area highlighted in blue is the alfalfa field, and the green lines represent the path of the UAV.

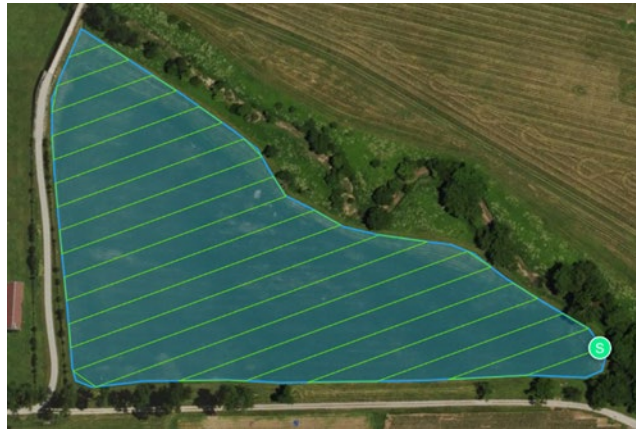


Figure 3.4 This is an aerial image of field 2. The region highlighted in blue shows the extent of field 2, and the green lines depict the flight path of the UAV.

Throughout the sampling period, manual observations of weed, insect, and disease pressure were collected using the scales in tables 2.1-2.2. Early in the data sampling period, most of fields 1 and 2 exhibited low (<5%) weed presence and damage due to disease and insects. Weed pressure stayed relatively constant throughout the

sampling period, and only a few locations within both fields exhibited weed pressure greater than 5% by the end of data collection. The insect and disease pressure increased throughout the data collection period to be between 5% and 20% for most locations in both fields.

Manual measurements of stand density and yield were also taken as measures of the field conditions on each collection date. Ten randomly selected plots within both fields were used as subsamples for the entire fields. This amounted to 1.42 samples per hectare for field 1 and 3.24 samples per hectare for field 2. For each 1 m² plot, all of the alfalfa crowns were counted and used to calculate the stand density of that plot.

Destructive sampling of the plots was performed to determine yield. The alfalfa harvested from each plot was collected, dried, and weighed. Throughout the sampling period, field 1 had a higher stand density than field 2 (Figure 3.5). The yields followed a similar pattern with field 1 exhibiting higher yields than field 2 (Figure 3.6). Field 1 represents a standard stand of alfalfa, while field 2 represents a weaker stand of alfalfa.

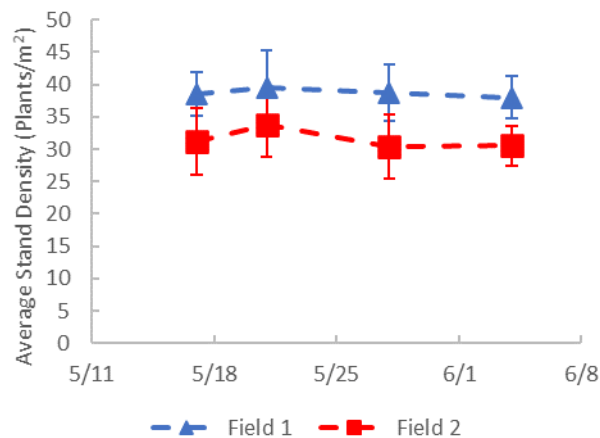


Figure 3.5 This time series plot shows the average stand density and standard deviation for field 1 and 2 during the summer of 2019.

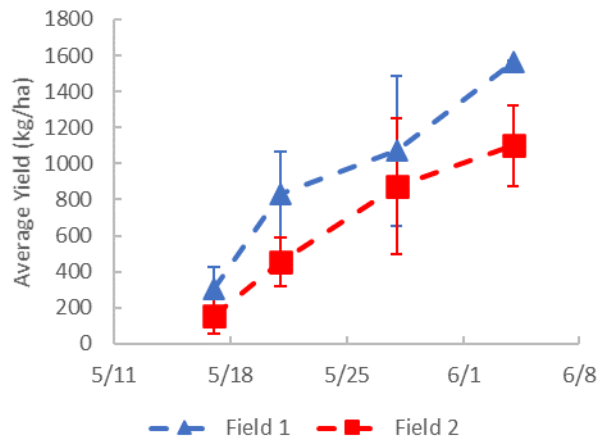


Figure 3.6 This time series plot shows the average yield and standard deviation for field 1 and 2 during the summer of 2019.

3.2.2 Photogrammetry Data Collection

Data collection began on May 17, 2019 and ended June 4, 2019. On each collection date, 10 quadrats were randomly placed in each field (Figure 3.7). These quadrats helped to designate the sampling areas and provided a rigid, easily recognizable structure to record ground control points (GCPs). In total, there were 10 GCPs collected on each field with a Trimble 5800 RTK GNSS receiver (Sunnyvale, California). Each GCP corresponded to one of the corners on a quadrat. Field notes were kept to ensure that the GCP information collected in the field could be properly associated with the correct UAV images within the photogrammetry software.

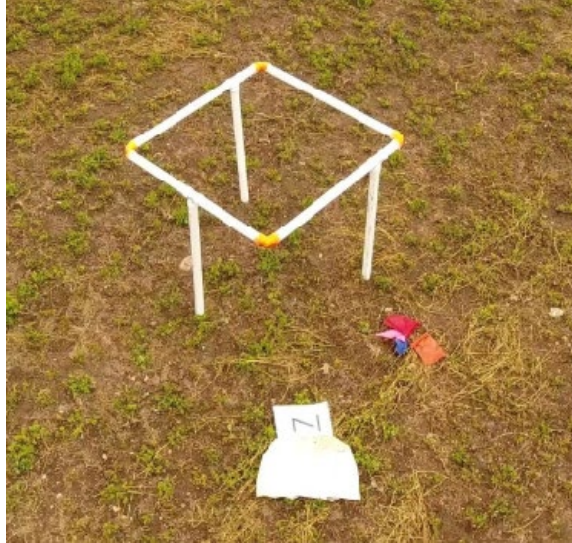


Figure 3.7 This image shows one of the quadrats that was used in the data collection process. It is a simple PVC structure with a 1 m² area elevated 1 m from the ground surface.

For each collection date, fields 1 and 2 were flown a total of three times using a DJI Phantom 4 Pro (Shenzhen, China) equipped with an RGB sensor. Using flight planning software, the UAV was autonomously flown at 30 m above ground level with a 90° gimbal angle, at 50 m with a 90° gimbal angle, and at 50 m with a 75° gimbal angle. At the 30 m elevation, the UAV was flown at a speed of 3.4 m/s and was able to capture images at a 0.8 cm/pixel resolution. At the 50 m elevation, the UAV reached speeds of 5.7 m/s and achieved an image resolution of 1.4 cm/pixel. The UAV flew in a back and forth pattern for all three sets of flight parameters (Figure 3.3-3.4). A standard front overlap of 85% and side overlap of 75% were also used for all three sets of flight parameters. The extent of each flight mission covered the entire area of each field and some of the area adjacent to the fields. The UAV was also used to collect low-elevation, high-resolution imagery of each quadrat sampling area in each field on each date. The UAV was flown at heights near ground level (approximately 10 m) and manually piloted

to circle around the quadrat. The RGB sensor acquired oblique imagery of the quadrat sampling area from many angles around the quadrat. Overall, there were four different flight parameters, and each one was tested on each sample collection day (four days, each separated by one week) in each field (field 1 and field 2), which had different stand densities (Table 3.1).

Table 3.1 Flight parameters tested in each of the two fields.

	Flight Parameters 1 (50-90°)	Flight Parameters 2 (50-75°)	Flight Parameters 3 (30-90°)	Flight Parameters 4 (Quadrat)
Elevation (m)	50.0	50.0	30.0	~10
Speed (m s ⁻¹)	5.70	5.70	3.40	varied
Gimbal Angle ^a	-90.0°	-75.0°	-90.0°	varied: -10° to -45°
Resolution (cm px ⁻¹)	1.40	1.40	0.80	varied

^a Gimbal Angle was measured from horizontal.

3.2.3 Data Processing

3.2.3.1 Photogrammetry Software Processing

Pix4Dmapper (Pix4D S.A., Prilly, Switzerland) was utilized to process all of the UAV imagery into CHMs. The process began by taking the images from each field scan and importing them into a Pix4D project. There were separate field scans for each of the first three flight parameters (50-90°, 50-75°, and 30-90°) in each field and for each of the four data collection days. This provided 24 different field scans. Once the images were loaded, the flight path and location of each image were visible in the program. These were checked for relative accuracy from field notes that were taken. If certain images were missing or if the generated flight path did not align with the actual flight path, the process of importing the images was restarted. Once the field scan's images were imported properly, the GCP information, collected in the field, was imported into the

project. The coordinates recorded from the RTK GPS unit were associated to their corresponding locations within the images of the project.

After importing all the relevant images and GCP data, the actual image processing was able to begin. Pix4D breaks up this process into three stages: initial processing, point cloud densification, and digital surface model (DSM) generation. After the initial processing stage was complete, the ground sampling distance, RMSE, and camera calibration of the initial point cloud were assessed. If these values fell outside of an acceptable range, the initial processing was restarted. Once the initial point cloud was deemed acceptable, the final two stages of processing were performed. This processing resulted in the creation of 24 point clouds describing the top of the canopy of fields 1 and 2 on each date with each flight parameter.

An additional eight more point clouds were created by combining all the field scan images (from the 50-90°, 50-75°, and 30-90° flight parameters) collected on a particular date in a particular field into one Pix4D project. Additionally, 80 point clouds, describing only the quadrat sampling area, were generated using individual Pix4D projects from the UAV flights flown near the quadrat. In total, 112 unique point clouds were created.

3.2.3.2 Point Cloud Processing

The point clouds were processed using Python (version 3.8). Python libraries, matplotlib, numpy, transforms3D, and open3D, were utilized to complete the processing. The point cloud output from Pix4D was imported into the Python environment and all the x, y, and z coordinates describing the field point cloud were extracted. The field point

cloud was then segmented into regions around each quadrat. These regions were a circle centered on each quadrat with a 2 m radius, but the inner 1 m directly around the quadrat was excluded (Figure 3.8). Thus, the analysis region was an annulus with an outer radius of 2 m and an inner radius of 1 m around each quadrat. All points within this annulus were selected for further analysis. An annulus was used because close inspection of the point clouds revealed that most exhibited significant stitching errors within and immediately around the quadrat structure. There were no obvious stitching errors found within these annulus regions.

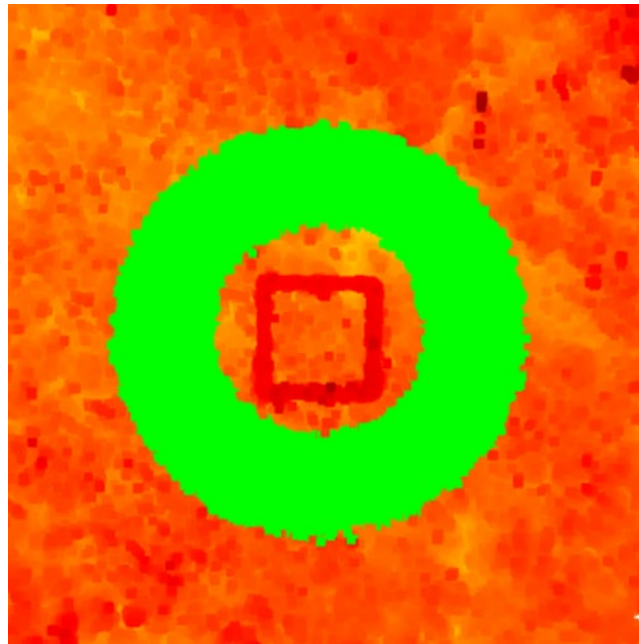


Figure 3.8 This is a top view of a point cloud depicting a sampling area imaged with the UAV. The points in green represent the annulus of points that were segmented out from each dataset. The red points were disregarded and not used in any further analysis. The dark red square in the middle of the annulus is the top of the quadrat.

After segmentation, there were still minor artifacts in the point clouds that frequently created extreme outliers and needed to be accounted for. To do so, a simple

filtering procedure, using the mean, standard deviation, 5th percentile, and 95th percentile values of the z coordinates, was utilized. Points with a z coordinate within two standard deviations of the mean z value and between the 5th and 95th percentile of z values were kept for further analysis. Once filtering of extreme outliers was complete, proper CHM values were extracted from the point clouds using the fixed geometric relationships between the canopy, quadrat, and ground surface. Statistical descriptors of the CHM's height distribution, the mean, 25th, 50th, 75th, and 95th percentile values, were calculated for further statistical analysis.

3.2.4 Statistical Analysis

Simple linear regressions were performed with the data generated from the point cloud processing. This was to determine the relationship between the data collected with different flight parameters. The CHM distribution statistics from each flight condition were compared to the corresponding statistics of the other flight parameters. These regression models were split between field 1 and 2. Data from field 1 was only compared to other data from field 1, and the same is true for field 2 data. The outcome of this analysis was 50 R^2 values (five flight parameters, two fields, five CHM descriptors) describing the correlations between each possible combination of flight parameters and CHM descriptors for each field (Table 3.2). For a given quadrat (sampling location), it would be desirable to have a CHM descriptor that was consistent across different flight parameters. This consistency would be revealed by a high R^2 for relationship between the descriptor created using different flight parameters. It is difficult to interpret 50 different correlations, so to identify situations that do not produce consistent results, the correlations were grouped by field, CHM descriptors, and flight parameters. If a certain

grouping had much lower correlations, it would be revealed as inconsistent and unsuitable for use in creating CHMs.

Table 3.2 Experimental variable for creating canopy height models

Flight Parameters	Fields	CHM descriptors
50-90°	Field 1: Thick Stand	mean
50-75°	Field 2: Thin Stand	25 th percentile
30-90°		50 th percentile
Quadrat		75 th percentile
Combination		95 th percentile

3.2.5 Predictive Modeling

Once stable methods of collecting UAV derived canopy height data were determined, the data from those methods were employed to produce predictive models of measured average canopy height and yield. The modeling process was performed with Python (version 3.8). A machine learning library called sci-kit learn was used to build, train, and test all of the predictive models used in this analysis. Python code was developed to iteratively use statistical descriptors, found to be reliable measures of canopy height, as explanatory variables in various regression models of manually measured canopy height and destructively sampled yield. All regression modeling techniques available within the sci-kit learn library were tested: simple linear, multiple linear, support vector machine, gaussian process, decision trees, k nearest neighbor (KNN), and neural networks. All the models were validated using repeated k fold cross validation. This resulted in each model being trained and tested 50 times (5 fold cross validation repeated 10 times) to account for any bias and/or variance that could be present in the modeling results. The reported statistics, R^2 and RMSE, were the averages from the 50 rounds of testing.

3.3 Results

3.3.1 Canopy Height Model Stability

Data collected from a standard alfalfa stand at the five flight parameters were highly correlated to each other (Table 3.3). Table 3.3 displays the correlations between each CHM descriptor created with one flight parameter with the same CHM descriptor created with another flight parameter (e.g. the correlation between the mean height from 30-90° and the mean height from 50-90°). Simple linear regression between distribution statistics resulted in moderate to high R^2 values, ranging from 0.75 to 0.97, with many greater than 0.90. Data collected within a weak stand of alfalfa were found to be highly unstable (Table 3.4). The data had low correlations among the five flight parameters. Regression analysis revealed that the majority of the R^2 values were below 0.75 with a range between 0.06 and 0.91. On average, the correlations found between data taken from a thick alfalfa field are higher than that of an alfalfa field with a weak stand. Consequently, the distributions of canopy height taken from field 1 at different flight parameters share similar values, while the distributions from field 2 have dissimilar values (Figure 3.9).

Table 3.3 R² values between statistical descriptors collected in field 1

	Distribution Statistics				
	Mean	25th Percentile	50 Percentile	75th Percentile	95th Percentile
Quadrat & 30-90°	0.94	0.93	0.94	0.93	0.90
Quadrat & 50-90°	0.80	0.75	0.81	0.83	0.84
Quadrat & 50-75°	0.94	0.91	0.94	0.94	0.93
Quadrat & Combo	0.93	0.92	0.93	0.93	0.91
30-90° & 50-90°	0.79	0.76	0.79	0.81	0.82
30-90° & 50-75°	0.93	0.92	0.94	0.94	0.94
30-90° & Combo	0.95	0.95	0.96	0.96	0.96
50-90° & 50-75°	0.89	0.87	0.89	0.90	0.92
50-90° & Combo	0.90	0.88	0.90	0.91	0.92
50-75° & Combo	0.97	0.95	0.97	0.97	0.97
Average:	0.90	0.88	0.91	0.91	0.91
STD:	0.06	0.07	0.06	0.05	0.05
Range:	0.18	0.20	0.17	0.16	0.15

Table 3.4 R² values between statistical descriptors collected in field 2

	Distribution Statistics				
	Mean	25th Percentile	50 Percentile	75th Percentile	95th Percentile
Quadrat & 30-90°	0.07	0.06	0.07	0.09	0.17
Quadrat & 50-90°	0.65	0.51	0.67	0.71	0.76
Quadrat & 50-75°	0.49	0.37	0.49	0.58	0.67
Quadrat & Combo	0.51	0.43	0.49	0.54	0.61
30-90° & 50-90°	0.22	0.18	0.22	0.28	0.38
30-90° & 50-75°	0.17	0.09	0.17	0.26	0.38
30-90° & Combo	0.49	0.45	0.48	0.56	0.64
50-90° & 50-75°	0.86	0.82	0.85	0.86	0.91
50-90° & Combo	0.84	0.75	0.83	0.86	0.88
50-75° & Combo	0.78	0.71	0.79	0.80	0.83
Average:	0.51	0.43	0.51	0.55	0.62
STD:	0.27	0.26	0.27	0.26	0.23
Range:	0.78	0.76	0.78	0.77	0.74

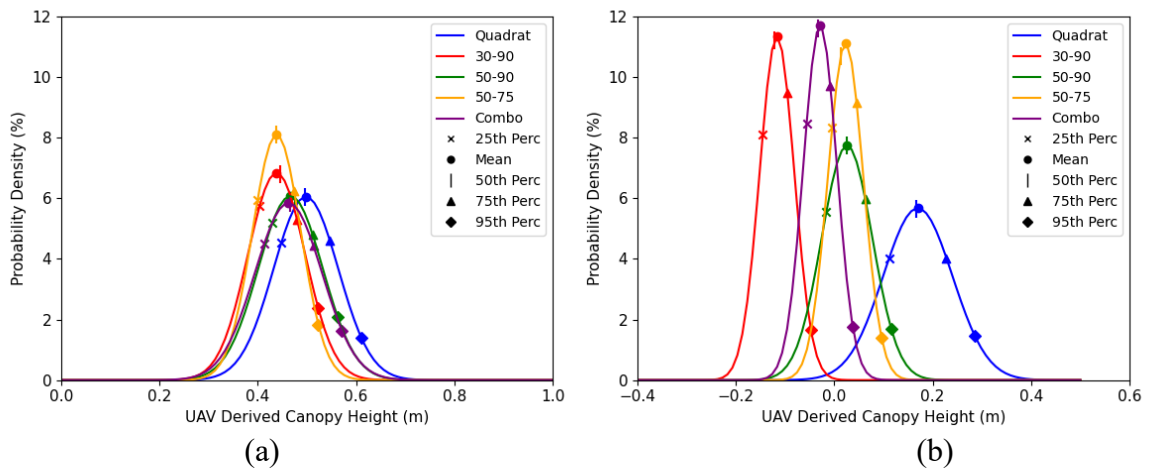


Figure 3.9 The Gaussian curve plots show the probability density of UAV-derived canopy height of each flight condition for quadrat 7 in field 1 (a) and quadrat 20 in field 2 (b) on 06/04/2019.

While Table 3.3 demonstrated that the various CHM descriptors were stable between flight parameters in thick alfalfa stands, it is also necessary to consider if any particular flight parameter is much worse or less consistent when compared to the others. In this analysis, the correlations for all the CHM descriptors between each pair of flight parameters were averaged together. Looking at the correlations between each flight parameter in a thick stand of alfalfa, data derived from the 50-90° flight condition seemed to be the least stable (Table 3.5). The 50-90° data was the least correlated to all other data collected at different flight parameters. The average R^2 values ranged from 0.80 to 0.90. All other relationships assessed between each of the other flight parameters exhibited higher correlations ($R^2 > 0.90$).

Table 3.5 Average R^2 matrix for flight parameters

	Quadrat	30-90°	50-90°	50-75°	Combination
Quadrat	**	0.93	0.81	0.93	0.92
30-90°	0.93	**	0.80	0.93	0.96
50-90°	0.81	0.80	**	0.89	0.90
50-75°	0.93	0.93	0.89	**	0.97
Combination	0.92	0.96	0.90	0.97	**
Average ¹	0.90	0.91	0.85	0.93	0.94

** R^2 will be 1 for each condition with itself.

¹Includes the R^2 with all flight parameters other than with itself.

For a thick alfalfa stand, the correlations between the descriptive statistics from different flight parameters showed that all of the statistical descriptors are stable enough to use in predictive modeling (Table 3.3). The mean, 25th, 50th, 75th and 95th percentiles had high average R^2 ($R^2 \geq 0.88$). The standard deviation and range of the R^2 values are also similar between the five variables. On average, the 25th percentile exhibited the lowest R^2 values within both types of alfalfa stands. For the thin stand, the 95th percentile had the highest average R^2 value, 0.62 (Table 3.4). The 95th percentile also had the lowest standard deviation and range of R^2 values, 0.23 and 0.74 respectively.

3.3.2 Predictive Models of Canopy Height and Yield

The best model of measured average canopy height in a thick stand of alfalfa is a simple linear regression model described by:

$$H_p = 0.93h_{50} + 0.165 \quad (3.1)$$

where

H_p = predicted average canopy height (m)

h_{50} = 50th percentile of the height distribution from the UAV flown at 30-90° (m)

Utilizing the 50th percentile of UAV derived canopy height flown at an elevation of 30 m in a nadir imaging configuration, the model fits the observed data nicely with evenly distributed residual errors (Figure 3.10a) and was able to produce predictions that were similar to the observed average canopy height (Figure 3.10b). Even though the model is relatively simple, it was capable of explaining 89% of the variation in the average canopy height and achieve a RMSE of 0.043 m.

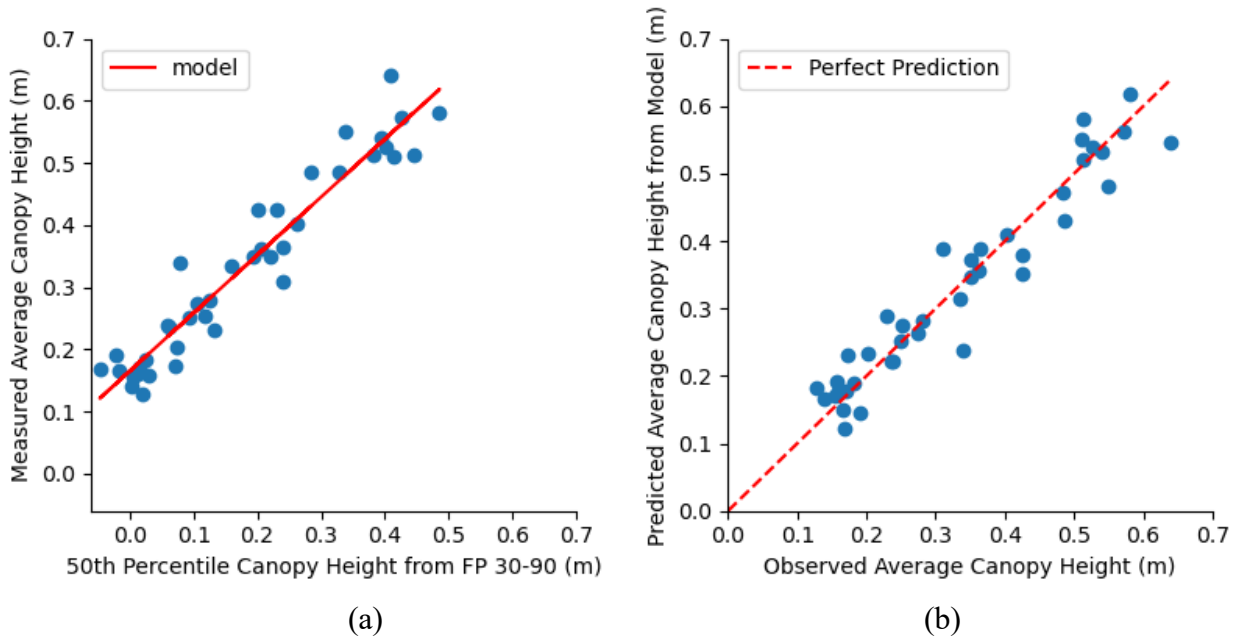


Figure 3.10 This figure shows two scatterplots depicting the simple linear regression model described in Equation 3.1 (a) and the goodness of fit for the model (b).

This specific model was chosen among the various models that were tested because using the 50th percentile as a single predictor outperformed all other statistical descriptors (Table 3.6). All models that used a statistical descriptor of canopy height between the mean/50th percentile and the 95th percentile performed relatively well ($R^2 \geq 0.85$). Although all the models had acceptable performances, the model incorporating the 50th percentile exhibited slightly better model evaluation metrics which makes it the optimal model to use.

Table 3.6 Linear regression models of measured average canopy height with different statistical predictors from the 30-90° flight parameter

Predictor	Intercept (m)	Coefficient	R ²	RMSE (m)
50th percentile	0.165	0.93	0.89	0.043
mean	0.165	0.96	0.88	0.044
75th percentile	0.136	0.90	0.87	0.044
95th percentile	0.104	0.87	0.85	0.047

The model (Equation 3.1) was also decided upon because other models utilizing the 50th percentile at different flight parameters did not perform as highly (Table 3.7). Most of the other models performed well and had R² values higher than 0.80. The exception to this is the model using canopy height data collected from a UAV flown at 50 m with a 90° gimbal angle. This finding coincides with the previous finding from the stability analysis which found that the data from the 50-90° flight parameter was the least stable. Out of the models that performed well, the model using data from a 30-90° flight slightly outperformed the rest with a R² of 0.89 and RMSE of 0.043 m.

Table 3.7 Linear regression models of measured average canopy height using the 50th percentile at different flight parameters

Flight Parameters	Intercept (m)	Coefficient	R ²	RMSE (m)
30-90°	0.165	0.93	0.89	0.043
Quadrat	0.102	0.91	0.86	0.049
Combination	0.118	0.86	0.86	0.048
50-75°	0.114	0.90	0.83	0.050
50-90°	0.193	0.65	0.62	0.078

Thus far, the results have reported the findings from simple linear regression models of average canopy height. To ensure the robustness of the reported model (Equation 3.1), other modeling techniques were evaluated and compared to it (Table 3.8). The various techniques rendered models that performed at a similar level or worse than the model in Equation 3.1. The KNN model and the multiple linear regression model performed comparably to the simple linear model. Technically, the KNN model had a better RMSE value, but this model is more complex to define and requires the user to set a k value. For the purposes of this study, the simple linear model was deemed the best because of its high performance and simplicity.

Table 3.8 Different model types using the 50th percentile from the 30-90° flight parameter as a predictor for average canopy height

Predictor(s)	Regression Model Type	R ²	RMSE (m)
30-90° 50th percentile	KNN ^a	0.89	0.041
30-90° 50th percentile	Simple Linear	0.89	0.043
30-90° mean & 50th percentile	Multiple Linear	0.88	0.043
30-90° 50th percentile	Gaussian Process	0.85	0.050
30-90° 50th percentile	Decision Tree	0.81	0.054
30-90° 50th percentile	Support Vector Machine	0.79	0.058

^a k = 5

The optimal model (Equation 3.1) was tested on data from field 2 to evaluate its performance in a thin stand of alfalfa. The model was not able to accurately predict the average canopy height of field 2. The R² value between the actual average canopy height

and the model's predictions of average canopy height was a negative number. Not only does this show that Equation 3.1 is not able to account for any of the variation seen in the canopy height data of field 2, but it also reveals that Equation 3.1 performed poorer than a constant model that always outputs the mean value of the canopy height dataset no matter the input data (which would have a R^2 of 0.0). Additionally, the RMSE value was 0.192 m which was worse than the RMSE achieved by Equation 3.1 when tested on the data from the thick stand (RMSE = 0.043 m).

The best model of yield in a thick stand of alfalfa is a KNN regression model that utilizes $k = 5$ nearest neighbors and the 95th percentile of UAV derived canopy height from the 50-75° flight parameter ($R^2 = 0.62$, RMSE = 476 kg/ha) (Figure 3.11). Other modeling techniques and explanatory variables were used in the modeling process, but most of the resulting models performed poorly ($R^2 < 0.60$). Only 4 other models were able to achieve comparable performances to the KNN model (Table 3.9). These models performed to a similar degree as the KNN, but they used canopy height data from the 30-90° flight parameter. These models were also simple linear regression models that did not require a user defined k value.

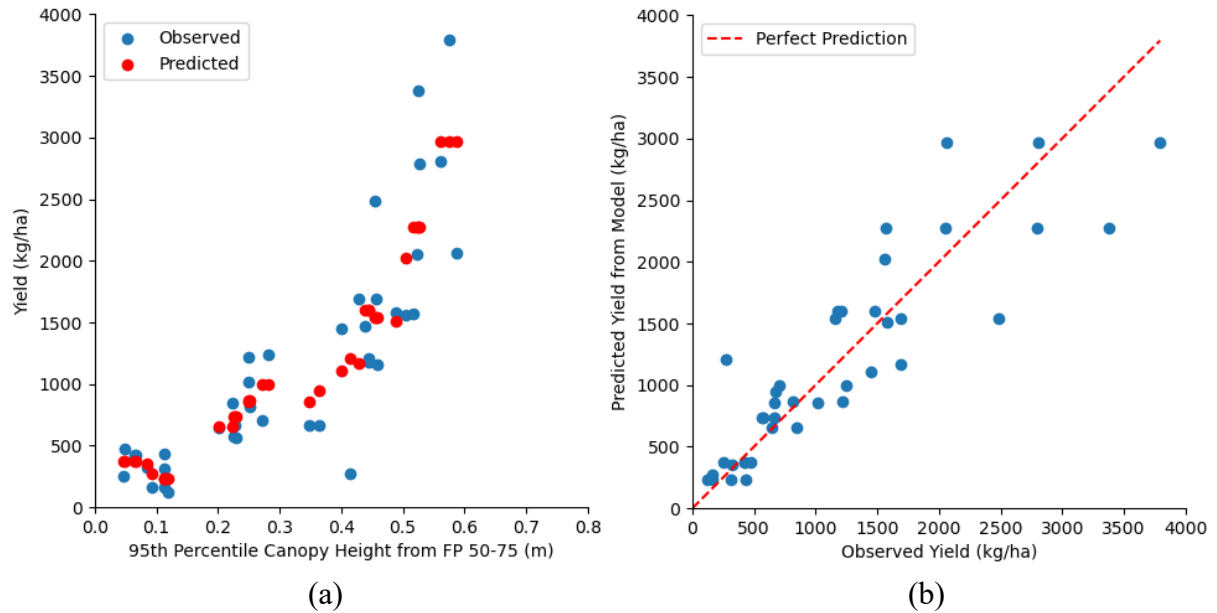


Figure 3.11 This figure shows two scatterplots depicting the KNN yield model using the 95th percentile of canopy height acquired by a UAV at 50-75° (a) and the goodness of fit of that model (b).

Table 3.9 Highest performing models of yield

Predictor	Regression Model Type	R ²	RMSE (kg/ha)
50-75° 95th percentile	KNN ^a	0.62	476
30-90° 75th percentile	Simple Linear	0.62	486
30-90° 50th percentile	Simple Linear	0.62	488
30-90° mean	Simple Linear	0.62	489
30-90° 95th percentile	Simple Linear	0.61	494

^a k = 5

The optimal model for yield in the thick stand was also tested on data from field 2 to evaluate its performance in a thin stand. The R² and RMSE values between the model's predictions of yield and the observed yield were 0.06 and 438 kg/ha, respectively. Even though the model achieved a lower RMSE value in the thin stand than in the thick stand, the model would still not be useful for thin stand applications because it exhibited a low R² value. Such a small R² value is indicative of a model that is hardly able to capture any of the variation within the response variable, in this case yield of a

thin stand. This shows that the optimal model for thick stands cannot be used for accurate predictions within a thin stand of alfalfa.

3.4 Discussion

3.4.1 Canopy Height Model Stability

From the results, it can be seen that the accuracy of UAV derived CHMs is sensitive to field conditions. Specifically, stand density played a major role in the correlations between statistical descriptors at various flight parameters. Canopy height models derived from UAV imagery of a thick alfalfa stand (field 1) showed high correlations between the statistical descriptors. Canopy height models from a thin alfalfa stand (field 2) had low correlations between the statistical descriptors. These low correlations can be attributed to stitching errors that most of the thin stand CHMs exhibited. The point clouds derived from field 2 data were improperly stitched below the ground surface (Figure 3.12). Having a thin stand allows for more of the ground surface to be exposed around the individual plants. The bare soil surface can interfere with UAV data collection and photogrammetry processing. Thompson et al. (2019) saw similar results with soil surface interference when creating CHMs. The magnitude of these errors fluctuated at different flight parameters. Consequently, there was little linear association between the datasets from each flight condition. This rendered the UAV data collected on the thin stand near useless in reliably creating robust CHMs of the alfalfa.

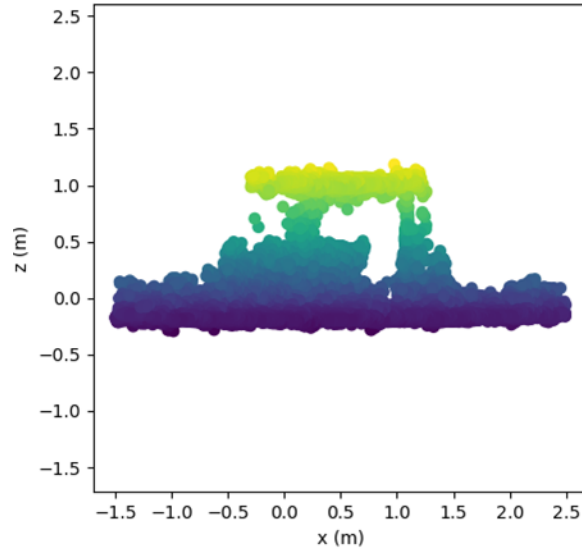


Figure 3.12 This scatterplot shows a profile view of the point cloud describing a sampling area in field 2 on 06/04/2019. The quadrat has been left in the point cloud to help visualize the stitching error due to a thin alfalfa stand. Points are color-coded based on height on the z axis.

All flight parameters utilized in this study were significantly correlated to each other on a thick stand of alfalfa. The 50-90° flight condition had the smallest correlations among all other flight parameters with an average R^2 of 0.85. Changing the camera gimbal angle from 90° to 75° increased the stability of the data. Flying at 50-75° showed improved R^2 values over that of 50-90°. This conveys how UAV canopy height collected at altitudes ranging from near canopy to 50 m are highly related. Higher altitude flights are advantageous because they typically have decreased flight times, but they usually produce lower-resolution data. The stability of CHMs generated from higher altitude flights might be high, but high stability does not directly correlate to high accuracy when compared to manual measurements of canopy height. Also, combining all of the data from the flight missions rendered CHMs that were similar to CHMs rendered from single flight missions. Data taken at 30-90° and 50-75° achieved comparable stability with

significantly less flight time and processing time than the combination flight condition. This suggests that multiple flights are unnecessary to capture reliable canopy height data and that the most efficient approach would be to use data from a single flight mission.

All UAV plant height variables were stable in thick alfalfa stands with the 25th percentile being the least stable. This suggests that general distribution statistics are stable enough to be incorporated into predictive models of canopy height of thick stands of alfalfa. As alfalfa stands become thinner, the distribution statistics become less stable. The correlations between statistical descriptors of a thin alfalfa stand were drastically different than that of a thick stand. Most descriptors had limited correlations with their counterparts at different flight parameters. The most stable descriptor in the thin stand was the 95th percentile. Because the 95th percentile performed the best in both thick and thin stands of alfalfa, it would be the best variable to use in predictive models of canopy height in fields with a normal to poor stand.

3.4.2 Predicting Canopy Height and Yield

Using the insights gained from the CHM stability analysis, models of alfalfa canopy height for a thick stand were developed. All of the models reported in this study were accurate in predicting manually measured average canopy height. The best model (Equation 3.1) used a single predictor, the 50th percentile of UAV derived canopy height at 30-90°. Although Table 3.6 shows the 50th percentile to be the best statistic to use, separate models built with other descriptive statistics, such as the mean, had comparable performances. Due to this, models using a statistic other than the 50th percentile should be considered in future analysis.

Although the data from all the tested flight parameters were highly stable, the data from the 30-90° flight parameter resulted in the best model of alfalfa canopy height (Table 3.7). The model built from the 30-90° flight parameter data even outperformed the model built from the quadrat flight parameter data. The quadrat flight parameter data was collected from an UAV at an elevation near 10 m while circling a specific quadrat. The 30-90° data was collected while performing a full field scan at an elevation 20 m above that of the quadrat flight parameter. This would suggest that not only can accurate data be acquired from field scale UAV scans, but the resulting data from the field scans could produce more accurate models than data taken at a higher resolution on a sub-field scale.

In contrast to the very high accuracies that were achieved in canopy height modeling, yield modeling of the thick stand with the UAV derived data resulted in R^2 values less than 0.63 (Table 3.9). The best model of yield used the 95th percentile of canopy height data from the 50-75° flight parameter with a KNN regression technique. All of the other top performing yield models were simple linear models that used a statistical descriptor from the 30-90° flight parameter. Notably, the 50th percentile of UAV derived canopy height from the 30-90° flight parameter resulted in one of the top yield models as well. This particular data index should be considered in future analysis because of its high performance in both canopy height and yield modeling.

3.5 Conclusion

The findings from this study outline the methods that should be used to ensure stable and robust UAV CHMs of alfalfa. One major factor that can affect CHM generation is the field that is being imaged. Data were collected on two different fields, one with a

thick stand and one with a thin stand of alfalfa. The resulting CHMs from each field were noticeably different. Stable canopy height data was not able to be collected on the thin stand with the UAV. This indicated how important stand density can be to the outcome of CHMs. Along with the field conditions, the sensitivity of CHMs to UAV flight parameters was also tested. Flights ranged from near ground level to 50 m and had camera gimbal angles in oblique and nadir configurations. For a thick alfalfa stand, all of the flight parameters rendered moderately to highly stable data. Lastly, all of the descriptive statistics used in this study were determined to be stable enough to be incorporated into predictive models of alfalfa production variables. The mean, 50th, 75th, and 95th percentiles of CHMs at different flight parameters were highly correlated to each other ($R^2 \geq 0.90$). Ultimately, the top of an alfalfa canopy can be reliably captured and defined by general statistical descriptors derived from UAV flights at heights equal to or below 50 m with nadir and oblique imagery for a thick stand of alfalfa.

From these determined methods, models of alfalfa canopy height and yield were developed. The best model of alfalfa canopy height was a simple linear regression model using the 50th percentile of UAV derived canopy height from the 30-90° flight parameter. The best model of yield was a KNN regression model that used the 95th percentile of UAV derived canopy height from the 50-75° flight parameter. By using the methods and models reported in this study, moderately to highly accurate field-wide estimations of alfalfa canopy height and yield can be achieved on fields with similar conditions as this study.

CHAPTER 4. CONCLUSION

The findings from these two studies validate and support the use of LiDAR and UAV technologies in monitoring growing alfalfa. The first study showed that the rapid and accurate acquisition of alfalfa canopy height and yield predictions were possible with a simple LiDAR sensor. Multiple predictive models of canopy height and yield were developed and tested. The best model of alfalfa canopy height was a simple linear regression model that utilized the 95th percentile of LiDAR derived canopy height as the single explanatory variable. The best model of alfalfa yield also used the 95th percentile as a single predictor, but the regression technique used was a fine gaussian support vector machine. This research lays the foundation for implementing LiDAR sensors into more autonomous, field scale monitoring systems, such as UAVs equipped with LiDAR sensors.

The second study proposed a methodology for reliable generation of UAV derived canopy height data and applied the methods to generate predictive models of alfalfa canopy height and yield at the field scale. Reliable canopy height data from a thick stand of alfalfa was able to be captured by general distribution statistics from a UAV flown at 50 m or below in nadir and oblique imaging configurations. Using this criteria, various field scale models of alfalfa canopy height and yield were created. The best model of canopy height was a simple linear regression model using the 50th percentile of detected canopy height from an UAV flown at 30 m in a nadir image configuration. The best model of yield was a KNN regression model using the 95th percentile of detected canopy height from an UAV flown at 50 m in an oblique image configuration. The findings from

this study serve as a means of reliably generating and practically applying UAV canopy height modeling for alfalfa monitoring at the field scale.

APPENDICES

[APPENDIX 1. POINT CLOUD PROCESSING CODE]

```
from laspy.file import File
import numpy as np
import open3d as o3d
import transforms3d as t3d
import matplotlib.pyplot as plt
from ralign import ralign
import pandas as pd

# Import 2019 Field Scans quadrat corner coordinates
df = pd.read_excel(
    'C:/Users/Tuck/OneDrive - University of Kentucky/Grad Work/'
    'Comparing Field Scan and Quadrat Models/'
    'Field Scan-Quadrat Corners-TuckCopy.xlsx')
df = df[['Date','Field','FP','Quadrat','X1','Y1','Z1','X2','Y2','Z2','X3','Y3','Z3']]
df = df.dropna()
df = df.reset_index(drop=True)

# Iterates through the 2019 quadrat corner rows to process each point cloud
for i in range(270,290):
    row = df.loc[i,:]
    if row['Field'] == 1 and row['FP'] == 'FP 30-90':
        inFile = File('E:/Pix4D Alfalfa Data/Field Scans/'+row['Date']+'KY/'
            +row['Date']+ '_entofieldscan_30m-90-part/2_densification/point_cloud/'
            +row['Date']+ '_entofieldscan_30m-90-
part_group1_densified_point_cloud.las', mode = "r")
    elif row['Field'] == 1 and row['FP'] == 'FP 50-75':
        inFile = File('E:/Pix4D Alfalfa Data/Field Scans/'+row['Date']+'KY/'
            +row['Date']+ '_entofieldscan_50m-75-part/2_densification/point_cloud/'
            +row['Date']+ '_entofieldscan_50m-75-
part_group1_densified_point_cloud.las', mode = "r")
    elif row['Field'] == 1 and row['FP'] == 'FP 50-90':
        inFile = File('E:/Pix4D Alfalfa Data/Field Scans/'+row['Date']+'KY/'
            +row['Date']+ '_entofieldscan_50m-90-part/2_densification/point_cloud/'
            +row['Date']+ '_entofieldscan_50m-90-
part_group1_densified_point_cloud.las', mode = "r")
    # elif row['Field'] == 1 and row['FP'] == 'FP 30/50-75/90':
    #     inFile = File('E:/Pix4D Alfalfa Data/Field Scans/'+row['Date']+'KY/'
```

```

#         +row['Date']+ '_entofieldscan_50m-90-part/2_densification/point_cloud/'
# got to change the FP
#         +row['Date']+ '_entofieldscan_50m-90-
part_group1_densified_point_cloud.las', mode = "r") # got to change the FP
elif row['Field'] == 2 and row['FP'] == 'FP 30-90':
    inFile = File('E:/Pix4D Alfalfa Data/Field Scans/'+row['Date']+ 'KY/'
                 +row['Date']+ '_trianglefieldscan_30m-90-part/2_densification/point_cloud/'
#change field
                 +row['Date']+ '_trianglefieldscan_30m-90-
part_group1_densified_point_cloud.las', mode = "r") #change field
elif row['Field'] == 2 and row['FP'] == 'FP 50-75':
    inFile = File('E:/Pix4D Alfalfa Data/Field Scans/'+row['Date']+ 'KY/'
                 +row['Date']+ '_trianglefieldscan_50m-75-full/2_densification/point_cloud/'
# change field
                 +row['Date']+ '_trianglefieldscan_50m-75-
full_group1_densified_point_cloud.las', mode = "r") # change field
elif row['Field'] == 2 and row['FP'] == 'FP 50-90':
    inFile = File('E:/Pix4D Alfalfa Data/Field Scans/'+row['Date']+ 'KY/'
                 +row['Date']+ '_trianglefieldscan_50m-90-full/2_densification/point_cloud/'
# change field
                 +row['Date']+ '_trianglefieldscan_50m-90-
full_group1_densified_point_cloud.las', mode = "r") # change field
# elif row['Field'] == 2 and row['FP'] == 'FP 30/50-75/90':
#     inFile = File('E:/Pix4D Alfalfa Data/Field Scans/'+row['Date']+ 'KY/'
#                 +row['Date']+ '_trianglefieldscan_50m-90-
part/2_densification/point_cloud/' # got to change the FP and field
#                 +row['Date']+ '_trianglefieldscan_50m-90-
part_group1_densified_point_cloud.las', mode = "r") # got to change the FP and field

# Get the raw xyz points from LAS
x = inFile.x
y = inFile.y
z = inFile.z

# Create an array of raw xyz coordinates
xyz = np.array([x,y,z]).transpose()

# Convert and save raw xyz coordinates as a .ply file because Open3D requires this
format
pcd = o3d.geometry.PointCloud()
pcd.points = o3d.utility.Vector3dVector(xyz)

# Define source and target coordinates

```

```

source_coords = np.array([[row['X1'],row['Y1'],row['Z1']],
                          [row['X2'],row['Y2'],row['Z2']],
                          [row['X3'],row['Y3'],row['Z3']]]).transpose()
target_coords = np.array([[0, 0, 1], [0, 1, 1], [1, 1, 1]]).transpose()

# Computing rotation, scaling and translation with ralign code from GitHub
R, c, t = ralign(source_coords,target_coords)

# Create 4x4 transformation matrix with transforms3D library
s = np.array([c,c,c])
T = t3d.affines.compose(t,R,s)

# Apply transformation
pcd_trans = pcd.transform(T)

# Creating annulus
trans_xyz = np.asarray(pcd_trans.points)
x = trans_xyz[:,0]
y = trans_xyz[:,1]
ann = trans_xyz[((x-0.5)**2+(y-0.5)**2)>1**2]&(((x-0.5)**2+(y-0.5)**2)<2**2)]

# save unfiltered pcd
pcd_unfilt = o3d.geometry.PointCloud()
pcd_unfilt.points = o3d.utility.Vector3dVector(ann)
o3d.io.write_point_cloud("E:/Pix4D Alfalfa Data/2019_processed_pcds"
                        "/annulus/field_scans/pcds/unfiltered/F"

+str(row['Field'])+"T"+str(row['Quadrat'])+"_"+row['Date']+"_"+row['FP']+".ply",
pcd_unfilt)

# Get array of filtered PCD
z = ann[:,2]
mean = z.mean()
std = z.std()
perc5 = np.percentile(z,5)
perc95 = np.percentile(z,95)
filt_xyz = ann[(z > mean - 2*std) & (z < mean + 2*std) & (z > perc5) & (z < perc95)]

# Histogram of filtered z values
plt.hist(filt_xyz[:,2],bins = 100,range=(-0.5,1.5),color = "cornflowerblue",edgecolor =
'black')
plt.title("Histogram of Plant Heights")
plt.xlabel('Heights (m)')

```

```

plt.ylabel('Count')
plt.savefig("E:/Pix4D Alfalfa Data/2019_processed_pcds"
           "/annulus/field_scans/histo/F"
+str(row['Field'])+"T"+str(row['Quadrat'])+"_"+row['Date']+"_"+row['FP']+".jpg")
plt.close()

# Save filtered pcd
pcd_filt = o3d.geometry.PointCloud()
pcd_filt.points = o3d.utility.Vector3dVector(filt_xyz)
o3d.io.write_point_cloud("E:/Pix4D Alfalfa Data/2019_processed_pcds"
                        "/annulus/field_scans/pcds/F"
+str(row['Field'])+"T"+str(row['Quadrat'])+"_"+row['Date']+"_"+row['FP']+".ply",
pcd_filt)

# Save filtered pcd in Matplotlib with axes
fig = plt.figure()
plt3d = fig.add_subplot(111, projection='3d')
plt3d.scatter(filt_xyz[:,0], filt_xyz[:,1], filt_xyz[:,2],c = filt_xyz[:,2])
plt3d.set_xlabel('X (m)')
plt3d.set_ylabel('Y (m)')
plt3d.set_zlabel('Z (m)')
plt3d.set_xlim(-2,3)
plt3d.set_ylim(-2,3)
plt3d.set_zlim(-0.5,1.5)
plt3d.set_title('Point Cloud')
plt.savefig("E:/Pix4D Alfalfa Data/2019_processed_pcds"
           "/annulus/field_scans/3D_scatter_plots/F"
+str(row['Field'])+"T"+str(row['Quadrat'])+"_"+row['Date']+"_"+row['FP']+".jpg")

plt.close()

# Add descriptive statistics of filtered pcd to excel file
desc_stats = pd.read_excel('E:/Pix4D Alfalfa
Data/2019_processed_pcds/annulus/field_scans/stats.xlsx', 'Sheet1', index_col=None,
na_values=['NA'])
res = pd.DataFrame(
    np.array([[row['Date'],
              row['Field'],
              row['FP'],
              row['Quadrat'],

```



```

    filt_xyz[:,2].mean(),
    filt_xyz[:,2].std(),
    np.percentile(filt_xyz[:,2],25),
    np.percentile(filt_xyz[:,2],50),
    np.percentile(filt_xyz[:,2],75),
    np.percentile(filt_xyz[:,2],95)]]),
    columns=['date','field','FP','quadrat','mean','std','25 perc','50 perc','75 perc','95 perc'])
desc_stats = desc_stats.append(res,True)
desc_stats.to_excel('E:/Pix4D Alfalfa
Data/2019_processed_pcds/annulus/field_scans/stats.xlsx', sheet_name='Sheet1', index =
False)

```

[APPENDIX 2. CANOPY HEIGHT PREDICTIVE MODELING CODE]

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression as lr
from sklearn.model_selection import cross_validate, RepeatedKFold
# import numpy as np

# import manually measured data
df_man = pd.read_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad
Work/"
                    "Comparing Field Scan and Quadrat Models/"
                    "Alfalfa Field Measurements Data 2019 - Tuck Copy.xlsx",0)
df_man['max'] = df_man['Canopy Height (Max)]*(1/100) # converting cm to m
df_man['mean'] = df_man['Canopy Height (Avg)]*(1/100) # converting cm to m
df_man = df_man[['Date','Plot','max','mean']]
df_man = df_man[~(df_man.Plot.str.startswith('S')).fillna(False)]
df_man = df_man[df_man.Date != '2019-05-14 00:00:00'] # remove 05-14-19
df_man = df_man[df_man['Plot']<11] # getting only field 1
df_man4q = df_man.drop([53,98]) # removing gaps that are present in quad data
df_man4combo = df_man.drop(27)

# import quadrat data
df_quad = pd.read_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad
Work/"
                    "Comparing Field Scan and Quadrat Models/statistical methods/"
                    "annulus/statistical_analysis_obj1.xlsx",0)
df_quad = df_quad.drop([87,91]) # remove outliers
df_quad = df_quad.drop(list(range(19))) # remove 05-14-19
df_quad = df_quad.reset_index(drop=True)
df_quad =
df_quad.drop(list(range(10,20))+list(range(29,39))+list(range(48,57))+list(range(67,73)))
df_quad = df_quad.drop(columns=['std','25 perc','90 perc'])

# import field scan data
df_fs = pd.read_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad Work/"
                    "Comparing Field Scan and Quadrat Models/statistical methods/"
                    "annulus/statistical_analysis_obj1.xlsx",1)
df_fs = df_fs[df_fs['field']==1]
df_fs = df_fs.drop(columns=['std','25 perc'])
df_3090 = df_fs[df_fs['FP']=='FP 30-90']
```

```

df_3090 = df_3090.drop(columns=['date','field','quadrat'])
df_5090 = df_fs[df_fs['FP']=='FP 50-90']
df_5090 = df_5090.drop(columns=['date','field','quadrat'])
df_5075 = df_fs[df_fs['FP']=='FP 50-75']
df_5075 = df_5075.drop(columns=['date','field','quadrat'])
df_combo = df_fs[df_fs['FP']=='FP 30/50-75/90']
df_combo = df_combo.drop(57)
df_combo = df_combo.drop(columns=['date','field','quadrat'])

dflst = [df_3090,df_5090,df_5075,df_combo,df_quad]

# automated linear regression with cross validation
for n in range(5):
    data = dflst[n]
    for i in range(1,5):

        if data.iloc[0,0] == 'FP 30-90':
            FP = '3090 '
            y = df_man[['mean']]
        elif data.iloc[0,0] == 'FP 50-90':
            FP = '5090 '
            y = df_man[['mean']]
        elif data.iloc[0,0] == 'FP 50-75':
            FP = '5075 '
            y = df_man[['mean']]
        elif data.iloc[0,0] == 'FP 30/50-75/90':
            FP = 'combo '
            y = df_man4combo[['mean']]
        elif data.iloc[0,0] == 'F28T1_05-17-2019':
            FP = 'quad '
            y = df_man4q[['mean']]

        label = FP+data.columns[i]
        x = data[[data.columns[i]]]
        model = lr().fit(x,y)
        intp = model.intercept_
        coeff = model.coef_
        ypred = model.predict(x)
        k = 5 # k number of folds
        cv = RepeatedKfold(n_splits=k,n_repeats=10,random_state=0)
        scores = cross_validate(lr(),x,y,cv=cv,scoring=(r2,'neg_root_mean_squared_error'))
        r2=scores['test_r2'].mean()
        rmse=-scores['test_neg_root_mean_squared_error'].mean()

```

```

# scatter plot and goodness of fit plot
fig,ax = plt.subplots(1,2,figsize=(8,4))

ax[0].scatter(x,y)
ax[0].plot(x,ypred,c='r',label = "model")
ax[0].plot([], [], ' ', label='y = '+str(round(float(coeff),2))+ 'x +
'+str(round(float(intp),2)))
ax[0].plot([], [], ' ', label= 'R\u00b2: '+str(round(r2,2)))
ax[0].spines['right'].set_visible(False)
ax[0].spines['top'].set_visible(False)
ax[0].set_xlabel(f'{label} Canopy Height (m)')
ax[0].set_ylabel('Manually Measured Mean Canopy Height (m)')
ax[0].set_aspect('equal','box')
ax[0].legend()

ax[1].scatter(y,ypred)
ax[1].plot([0,y.max()], [0,y.max()],'r--',label='Perfect Prediction')
ax[1].spines['right'].set_visible(False)
ax[1].spines['top'].set_visible(False)
ax[1].set_ylabel(f'Predicted Mean Canopy Height from {label} (m)')
ax[1].set_xlabel('Observed Mean Canopy Height (m)')
ax[1].set_aspect('equal', 'box')
ax[1].legend()
fig.tight_layout()
fig.savefig("C:/Users/Tuck/OneDrive - University of Kentucky/Grad Work/FP
paper/figures/sim_linear/"+label+'.png')
plt.close()

# Add models and model performance to excel file
stats = pd.read_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad
Work/FP paper/models.xlsx",
                    'Sheet1', index_col=None, na_values=['NA'])
res = pd.DataFrame(
    [['Average Manual Height',label,float(intp),float(coeff),r2,rmse,'simple linear
regression']],
    columns=['response','predictor','intp','coeff','R2','RMSE','type'])
stats = stats.append(res,True)
stats.to_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad Work/FP
paper/models.xlsx",
              sheet_name='Sheet1', index = False)

```

[APPENDIX 3. YIELD PREDICTIVE MODELING CODE]

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression as lr
from sklearn.model_selection import cross_validate, RepeatedKFold
# import numpy as np

# import yield data
df_yield = pd.read_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad
Work/"
                        "FP paper/Alfalfa Field Measurements Data 2019 - Yield.xlsx",0)
df_yield = df_yield[['Date','Plot','Yield (kg/ha)']]
df_yield = df_yield[df_yield.Date != '2019-05-14 00:00:00']
df_yield = df_yield[df_yield['Plot']<11].reset_index(drop=True)
df_yield_C = df_yield.drop(7)
df_yield_Q = df_yield.drop([13,28])

# import quadrat data
df_quad = pd.read_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad
Work/"
                        "Comparing Field Scan and Quadrat Models/statistical methods/"
                        "annulus/statistical_analysis_obj1.xlsx",0)
df_quad = df_quad.drop([87,91]) # remove outliers
df_quad = df_quad.drop(list(range(19))) # remove 05-14-19
df_quad = df_quad.reset_index(drop=True)
df_quad =
df_quad.drop(list(range(10,20))+list(range(29,39))+list(range(48,57))+list(range(67,73)))
df_quad = df_quad.drop(columns=['std','25 perc','90 perc'])

# import field scan data
df_fs = pd.read_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad Work/"
                        "Comparing Field Scan and Quadrat Models/statistical methods/"
                        "annulus/statistical_analysis_obj1.xlsx",1)
df_fs = df_fs[df_fs['field']==1]
df_fs = df_fs.drop(columns=['std','25 perc'])
df_3090 = df_fs[df_fs['FP']=='FP 30-90']
df_3090 = df_3090.drop(columns=['date','field','quadrat'])
df_5090 = df_fs[df_fs['FP']=='FP 50-90']
df_5090 = df_5090.drop(columns=['date','field','quadrat'])
df_5075 = df_fs[df_fs['FP']=='FP 50-75']
df_5075 = df_5075.drop(columns=['date','field','quadrat'])
```

```

df_combo = df_fs[df_fs['FP']=='FP 30/50-75/90']
df_combo = df_combo.drop(57)
df_combo = df_combo.drop(columns=['date','field','quadrat'])

dflst = [df_3090,df_5090,df_5075,df_combo,df_quad]

# automated linear regression with cross validation
for n in range(5):
    data = dflst[n]
    for i in range(1,5):

        if data.iloc[0,0] == 'FP 30-90':
            FP = '3090 '
            y = df_yield[['Yield (kg/ha)']]
        elif data.iloc[0,0] == 'FP 50-90':
            FP = '5090 '
            y = df_yield[['Yield (kg/ha)']]
        elif data.iloc[0,0] == 'FP 50-75':
            FP = '5075 '
            y = df_yield[['Yield (kg/ha)']]
        elif data.iloc[0,0] == 'FP 30/50-75/90':
            FP = 'combo '
            y = df_yield_C[['Yield (kg/ha)']]
        elif data.iloc[0,0] == 'F28T1_05-17-2019':
            FP = 'quad '
            y = df_yield_Q[['Yield (kg/ha)']]

        label = FP+data.columns[i]
        x = data[[data.columns[i]]]
        model = lr().fit(x,y)
        intp = model.intercept_
        coeff = model.coef_
        ypred = model.predict(x)
        k = 5 # k number of folds
        cv = RepeatedKFold(n_splits=k,n_repeats=20,random_state=0)
        scores = cross_validate(lr(),x,y,cv=cv,scoring=('r2','neg_root_mean_squared_error'))
        r2=scores['test_r2'].mean()
        rmse=-scores['test_neg_root_mean_squared_error'].mean()

        # scatter plot and goodness of fit plot
        fig,ax = plt.subplots(1,2,figsize=(8,4))

        ax[0].scatter(x,y)

```

```

ax[0].plot(x,ypred,c='r',label = "model")
ax[0].plot([], [], ' ', label='y = '+str(round(float(coeff),2))+ 'x +
'+str(round(float(intp),2)))
ax[0].plot([], [], ' ', label= 'R\u00b2: '+str(round(r2,2)))
ax[0].spines['right'].set_visible(False)
ax[0].spines['top'].set_visible(False)
ax[0].set_xlabel(f'{label} Canopy Height (m)')
ax[0].set_ylabel('Manually Measured Mean Canopy Height (m)')
ax[0].set_aspect('equal','box')
ax[0].legend()

ax[1].scatter(y,ypred)
ax[1].plot([0,y.max()], [0,y.max()], 'r--', label='Perfect Prediction')
ax[1].spines['right'].set_visible(False)
ax[1].spines['top'].set_visible(False)
ax[1].set_ylabel(f'Predicted Mean Canopy Height from {label} (m)')
ax[1].set_xlabel('Observed Mean Canopy Height (m)')
ax[1].set_aspect('equal', 'box')
ax[1].legend()
fig.tight_layout()
fig.savefig("C:/Users/Tuck/OneDrive - University of Kentucky/Grad Work/FP
paper/figures/sim_linear/"+label+'.png')
plt.close()

# Add models and model performance to excel file
stats = pd.read_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad
Work/FP paper/yield_models.xlsx",
                    'Sheet1', index_col=None, na_values=['NA'])
res = pd.DataFrame(
    [['yield',label,float(intp),float(coeff),r2,rmse,'simple linear regression']],
    columns=['response','predictor','intp','coeff','R2','RMSE','type'])
stats = stats.append(res,True)
stats.to_excel("C:/Users/Tuck/OneDrive - University of Kentucky/Grad Work/FP
paper/yield_models.xlsx",
              sheet_name='Sheet1', index = False)

```

REFERENCES

- Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., & Bareth, G. (2014). Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. *Remote Sensing*, 6(11). doi:10.3390/rs61110395
- Chang, A., Jung, J., Maeda, M. M., & Landivar, J. (2017). Crop height monitoring with digital imagery from Unmanned Aerial System (UAS). *Computers and Electronics in Agriculture*, 141, 232-237. doi:<https://doi.org/10.1016/j.compag.2017.07.008>
- Cherney, J. H., Smith, S. R., Sheaffer, C. C., & Cherney, D. J. R. (2020). Nutritive value and yield of reduced-lignin alfalfa cultivars in monoculture and in binary mixtures with perennial grass. *Agronomy Journal*, 112(1), 352-367. doi:<https://doi.org/10.1002/agj2.20045>
- Crommelinck, S., & Höfle, B. (2016). Simulating an Autonomously Operating Low-Cost Static Terrestrial LiDAR for Multitemporal Maize Crop Height Measurements. *Remote Sensing*, 8(3), 205.
- Cumo, C. (2013). *Encyclopedia of Cultivated Plants: From Acacia to Zinnia [3 Volumes] : From Acacia to Zinnia*. Santa Barbara, California: ABC-CLIO.
- Cunliffe, A. M., Brazier, R. E., & Anderson, K. (2016). Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. *Remote Sensing of Environment*, 183, 129-143. doi:<https://doi.org/10.1016/j.rse.2016.05.019>
- Dvorak, J. S., Pampolini, L. F., Jackson, J. J., Seyyedhasani, H., Sama, M. P., & Goff, B. (2021). Predicting Quality and Yield of Growing Alfalfa from a UAV. *Transactions of the ASABE*, 64(1), 63-72. doi:<https://doi.org/10.13031/trans.13769>
- Eitel, J. U. H., Magney, T. S., Vierling, L. A., Brown, T. T., & Huggins, D. R. (2014). LiDAR Based Biomass and Crop Nitrogen Estimates for Rapid, Non-destructive Assessment of Wheat Nitrogen Status. *Field Crops Research*, 159, 21-32.
- Enciso, J., Avila, C. A., Jung, J., Elsayed-Farag, S., Chang, A., Yeom, J., . . . Chavez, J. C. (2019). Validation of agronomic UAV and field measurements for tomato varieties. *Computers and Electronics in Agriculture*, 158, 278-283. doi:10.1016/j.compag.2019.02.011
- Feng, A., Zhang, M., Sudduth, K. A., Vories, E. D., & Zhou, J. (2019). Cotton Yield Estimation from UAV-Based Plant Height. *Transactions of the ASABE*, 62(2), 393-404. doi:10.13031/trans.13067
- Fricke, T., Richter, F., & Wachendorf, M. (2011). Assessment of forage mass from grassland swards by height measurement using an ultrasonic sensor. *Computers*

and *Electronics in Agriculture*, 79(2), 142-152.
doi:<https://doi.org/10.1016/j.compag.2011.09.005>

- Grev, A. M., Wells, M. S., Sheaffer, C. C., & Martinson, K. L. (2017). A comparison of reduced lignin and conventional alfalfa varieties and their potential for use as equine forage sources. *Journal of equine veterinary science*, 52, 100-100. doi:10.1016/j.jevs.2017.03.150
- James, M. R., & Robson, S. (2014). Mitigating systematic error in topographic models derived from UAV and ground-based image networks. *Earth Surface Processes and Landforms*, 39(10), 1413-1420. doi:<https://doi.org/10.1002/esp.3609>
- Jaud, M., Letortu, P., Théry, C., Grandjean, P., Costa, S., Maquaire, O., . . . Le Dantec, N. (2019). UAV survey of a coastal cliff face – Selection of the best imaging angle. *Measurement*, 139, 10-20. doi:<https://doi.org/10.1016/j.measurement.2019.02.024>
- Jimenez-Berni, J., Deery, D. M., Rozas-Larraondo, P., Condon, A. G., Rebetzke, G. J., James, R. A., . . . Sirault, X. R. R. (2018). High Throughput Determination of Plant Height, Ground Cover, and Above-Ground Biomass in Wheat with LiDAR. *Frontiers in Plant Science*, 9(237).
- Lacefield, G. D. (1988). Alfalfa Hay Quality Makes the Difference. Retrieved from https://uknowledge.uky.edu/anr_reports
- Lacefield, G. D., Henning, J. C., Collins, M., & Swetnam, L. (1996). Quality Hay Production. Retrieved from <https://forages.ca.uky.edu/foragepublications>
- Lacefield, G. D., Henning, J. C., Rasnake, M., & Collins, M. (1997). Alfalfa The Queen of Forage Crops. Retrieved from https://www.google.com/url?client=internal-element-cse&cx=003398099011942252124:gmpps7ysdwm&q=http://www.ca.uky.edu/agc/pubs/agr/agr76/agr76.pdf&sa=U&ved=2ahUKewiD_oa6xdvxAhXTKM0KHa2wBvAQFjAAegQIAhAB&usg=AOvVaw2T2iB0YnIxkiQNdm7faWyu
- Lin, J., Wang, M., Ma, M., & Lin, Y. (2018). Aboveground Tree Biomass Estimation of Sparse Subalpine Coniferous Forest with UAV Oblique Photography. *Remote Sensing*, 10(11). doi:10.3390/rs10111849
- Lyons, T., Undersander, D., Welch, R., & Donnelly, D. (2016). Estimating Alfalfa Yield from Plant Height. *Crop, Forage & Turfgrass Management*, 2(1), cftm2015.0203. doi:<https://doi.org/10.2134/cftm2015.0203>
- Madec, S., Baret, F., Solan, B. d., Thomas, S., Dutartre, D., Jezequel, S., . . . Comar, A. (2017). High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. *Frontiers in Plant Science*, 8(2002).
- Martin, N. P., Russelle, M. P., Powell, J. M., Sniffen, C. J., Smith, S. I., Tricarico, J. M., & Grant, R. J. (2017). Invited review: Sustainable forage and grain crop production

- for the US dairy industry. *Journal of Dairy Science*, 100(12), 9479-9494. doi:<https://doi.org/10.3168/jds.2017-13080>
- Mathanker, S. K., Maughan, J. D., Hansen, A. C., Grift, T. E., & Ting., K. C. (2014). Sensing Miscanthus Swath Volume for Maximizing Baler Throughput Rate. *Transactions of the ASABE*, 57(2), 355-362.
- Mesas-Carrascosa, F.-J., García, M. D. N., De Larriva, J. E. M., & García-Ferrer, A. (2016). An analysis of the influence of flight parameters in the generation of unmanned aerial vehicle (UAV) orthomosaicks to survey archaeological areas. *Sensors (Basel, Switzerland)*, 16(11), 1838. doi:10.3390/s16111838
- Mesas-Carrascosa, F.-J., Torres-Sánchez, J., Clavero-Rumbao, I., García-Ferrer, A., Peña, J.-M., Borra-Serrano, I., & López-Granados, F. (2015). Assessing optimal flight parameters for generating accurate multispectral orthomosaicks by uav to support site-specific crop management. *Remote sensing (Basel, Switzerland)*, 7(10), 12793-12814. doi:10.3390/rs71012793
- Noland, R. L., Wells, M. S., Coulter, J. A., Tiede, T., Baker, J. M., Martinson, K. L., & Sheaffer, C. C. (2018). Estimating alfalfa yield and nutritive value using remote sensing and air temperature. *Field Crops Research*, 222, 189-196. doi:<https://doi.org/10.1016/j.fcr.2018.01.017>
- O. Payero, J., M. U. Neale, C., & L. Wright, J. (2004). Comparison of Eleven Vegetation Indices for Estimating Plant Height of Alfalfa and Grass. *Applied Engineering in Agriculture*, 20(3), 385-393. doi:<https://doi.org/10.13031/2013.16057>
- Oke, V., & Long, S. R. (1999). Bacterial genes induced within the nodule during the Rhizobium-legume symbiosis. *Molecular Microbiology*, 32(4), 837-849. doi:<https://doi.org/10.1046/j.1365-2958.1999.01402.x>
- Pittman, J. J., Arnall, D. B., Interrante, S. M., Moffet, C. A., & Butler, T. J. (2015). Estimation of Biomass and Canopy Height in Bermudagrass, Alfalfa, and Wheat Using Ultrasonic, Laser, and Spectral Sensors. *Sensors*, 15(2). doi:10.3390/s150202920
- Rosell, J. R., & Sanz, R. (2012). A Review of Methods and Applications of the Geometric Characterization of Tree Crops in Agricultural Activities. *Computers and Electronics in Agriculture*, 81, 124-141.
- Schaefer, M. T., & Lamb, D. W. (2016). A Combination of Plant NDVI and LiDAR Measurements Improve the Estimation of Pasture Biomass in Tall Fescue (*Festuca arundinacea* var. Fletcher). *Remote Sensing*, 8(2). doi:10.3390/rs8020109
- Smith, M. A. (2008). Robel Pole Technique and Data Interpretation [Extension Article]. Retrieved from https://www.wyoextension.org/publications/Search_Details.php?pubid=207&pub=MP-111.10

- Song, Y., & Wang, J. (2019). Winter Wheat Canopy Height Extraction from UAV-Based Point Cloud Data with a Moving Cuboid Filter. *Remote sensing (Basel, Switzerland)*, 11(10), 1239. doi:10.3390/rs11101239
- Thompson, A. L., Thorp, K. R., Conley, M. M., Elshikha, D. M., French, A. N., Andrade-Sanchez, P., & Pauli, D. (2019). Comparing Nadir and Multi-Angle View Sensor Technologies for Measuring in-Field Plant Height of Upland Cotton. *Remote Sensing*, 11(6). doi:10.3390/rs11060700
- Tu, Y. H., Johansen, K., Aragon, B., Stutsel, B. M., Ángel, Y., Camargo, O. A. L., . . . McCabe, M. F. (2021). Combining Nadir, Oblique, and Façade Imagery Enhances Reconstruction of Rock Formations Using Unmanned Aerial Vehicles. *IEEE Transactions on Geoscience and Remote Sensing*, 1-13. doi:10.1109/TGRS.2020.3047435
- Undersander, D. J. (2011). *Alfalfa management guide*. Madison, WI: Madison, WI : American Society of Agronomy : Crop Science Society of America : Soil Science Society of America.
- USDA-NASS. (2021). *Crop Values Annual Summary*. Retrieved from <https://usda.library.cornell.edu/concern/publications/k35694332>
- Vittetoe, R. K., & Lang, B. J. (Producer). (2019). Estimate Alfalfa First Crop Harvest with PEAQ. *Integrated Crop Management News*.
- Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., . . . Iwat, H. (2017). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Frontiers in Plant Science*, 8, 421-421. doi:10.3389/fpls.2017.00421
- Wiering, N. P., Ehlke, N. J., & Sheaffer, C. C. (2019). Lidar and RGB Image Analysis to Predict Hairy Vetch Biomass in Breeding Nurseries. *Plant phenome journal*, 2(1), 1-8. doi:10.2135/tppj2019.02.0003
- Wijesingha, J., Moeckel, T., Hensgen, F., & Wachendorf, M. (2019). Evaluation of 3D point cloud-based models for the prediction of grassland biomass. *International Journal of Applied Earth Observation and Geoinformation*, 78, 352-359. doi:<https://doi.org/10.1016/j.jag.2018.10.006>
- Yanbo, H. (2016). Cotton Yield Estimation Using Very High-Resolution Digital Images Acquired with a Low-Cost Small Unmanned Aerial Vehicle. *Transactions of the ASABE*, 59(6), 1563-1574. doi:10.13031/trans.59.11831
- Yost, M. A., Coulter, J. A., Russelle, M. P., Sheaffer, C. C., & Kaiser, D. E. (2012). Alfalfa Nitrogen Credit to First-Year Corn: Potassium, Regrowth, and Tillage Timing Effects. *Agronomy Journal*, 104(4), 953-962. doi:<https://doi.org/10.2134/agronj2011.0384>

- Zahawi, R. A., Dandois, J. P., Holl, K. D., Nadwodny, D., Reid, J. L., & Ellis, E. C. (2015). Using lightweight unmanned aerial vehicles to monitor tropical forest recovery. *Biological Conservation*, *186*, 287-295. doi:<https://doi.org/10.1016/j.biocon.2015.03.031>
- Zhang, H., Sun, Y., Chang, L., Qin, Y., Chen, J., Qin, Y., . . . Wang, Y. (2018). Estimation of Grassland Canopy Height and Aboveground Biomass at the Quadrat Scale Using Unmanned Aerial Vehicle. *Remote Sensing*, *10*(6). doi:10.3390/rs10060851
- Zhang, L., & Grift, T. E. (2012). A LIDAR-based crop height measurement system for *Miscanthus giganteus*. *Computers and Electronics in Agriculture*, *85*, 70-76. doi:10.1016/j.compag.2012.04.001

VITA

Stuart Tucker Sheffield

EDUCATION

Mississippi State University, Starkville, Mississippi

BS in Agricultural Engineering Technology and Business August 2015 - May 2019

- Concentration in Geographical Information Systems and Remote Sensing
- 4.0/4.0 GPA (Summa Cum Laude)

WORK EXPERIENCE

University of Kentucky, Lexington, Kentucky

Biosystems and Agricultural Engineering Research Assistant August 2019 - Present

USDA-Agricultural Research Services, Starkville, Mississippi

Student Technician for Sustainable Agriculture Division May 2018 - May 2019

Brent Engineering Service, Starkville, Mississippi

Field Crew and Draftsman May 2017 - May 2018

SCHOLASTIC AND PROFESSIONAL HONORS

Alpha Epsilon Engineering Honor Society, Univ. of Kentucky Lexington, Kentucky
August 2020 - Present

Stephen D. Lee Scholar, Mississippi State Univ. Starkville, Mississippi
May 2019

Gamma Sigma Delta Ag. Honor Society, Mississippi State Univ. Starkville, Mississippi
August 2018 - May 2019

Phi Kappa Phi Honor Society, Mississippi State Univ. Starkville, Mississippi
January 2019 - May 2019

PROFESSIONAL PUBLICATIONS

Sheffield, S. T., Dvorak, J., Smith, B., Arnold, C. & Minch, C. (in press). Using LiDAR to Measure Alfalfa Canopy Height. *Transactions of the ASABE*.

Minch, C., Dvorak, J., Jackson, J., & **Sheffield, S. T.** (2021). Creating a Field-Wide Forage Canopy Model Using UAVs and Photogrammetry Processing. *Remote Sensing (Basel, Switzerland)*, 13(13), 2487–. <https://doi.org/10.3390/rs13132487>

Minch, C., Dvorak, J., Jackson, J., & **Sheffield, S. T.** (2020). UAV How-To: Create a Forage Canopy Model with Photogrammetry. *University of Kentucky Cooperative Extension*. Available at: http://dept.ca.uky.edu/agc/pub_prefix.asp?series=AEN