

Clemson University

TigerPrints

All Theses

Theses

August 2020

Novel Methods for RGB Aerial Image Analysis

Brennan Teddy

Clemson University, beteddy1@carolina.rr.com

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Recommended Citation

Teddy, Brennan, "Novel Methods for RGB Aerial Image Analysis" (2020). *All Theses*. 3422.

https://tigerprints.clemson.edu/all_theses/3422

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

NOVEL METHODS FOR RGB AERIAL IMAGE ANALYSIS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Plant and Environmental Sciences

by
Brennan Teddy
August 2020

Accepted by:
John D. Mueller, Ph.D., Committee Co-Chair
Kendall R. Kirk, Ph.D., Committee Co-Chair
Michael T. Plumblee, Ph.D.
Michael W. Marshall, Ph.D.

ABSTRACT

Multiple linear regression models were developed to predict sand and clay content along with soil organic matter content from RGB imagery from both commercially available satellite imagery as well as RGB UAV imagery. UAV Imagery was tested at two flight altitudes to determine if lower or higher altitude had an effect on prediction. In cases of sand, clay, and OM content, flight altitudes did not significantly differ in prediction abilities. Satellite imagery was evaluated using data from Planet Labs as well as Google Earth. Regression models were developed to predict sand, clay, and soil organic matter content from these satellite images, which captured fields with bare soil. An alternative to whole field data collection, referred to herein as the point sampling method, was introduced. A survey of currently available neural network and machine learning technologies was performed to establish which of these technologies could benefit the precision agriculture industry. A sample model was trained to detect and classify cotton blooms from low-altitude RGB imagery collected from a DJI Phantom 3 UAV.

TABLE OF CONTENTS

	Page
TITLE PAGE	i
ABSTRACT.....	ii
LIST OF TABLES.....	v
LIST OF FIGURES	viii
CHAPTER	
I. INTRODUCTION	1
II. SOIL TEXTURE AND ORGANIC MATTER CONTENT PREDICTION FROM CONSUMER LEVEL UAV WITH RGB CAMERA:	3
Introduction.....	3
Methods and Materials.....	8
Results and Discussion	18
Conclusion	33
References.....	35
III. SOIL TEXTURE AND ORGANIC MATTER CONTENT PREDICTION FROM INTERNET AVAILABLE SATELLITE IMAGERY	37
Introduction.....	37
Methods and Materials.....	42
Results and Discussion	56
Conclusion	80
References.....	84

IV.	AERIAL IMAGERY POINT SAMPLING METHOD AND SURVEY OF UAV AND IMAGE ANALYSIS USING MACHINE LEARNING IN AGRICULTURE	86
	Introduction.....	86
	UAV Hardware and Software	87
	Machine Learning Technologies.....	97
	Conclusion	103
	References	105
V.	Conclusions.....	107
	APPENDICES	109
A:	Regression Model Bounding Coefficients	110
B:	Dates of Satellite Image Capture	113
C:	Regression Models and Coefficients for Satellite Imagery Models	119

LIST OF TABLES

Table	Page
2.1 C12 Field SSURGO soil type distribution.....	9
2.2 E7 Field SSURGO soil type distribution.....	9
2.3 Pond Field SSURGO soil type distribution	9
2.4 Prediction errors for sand content prediction models	20
2.5 Regression model terms and coefficients for sand content prediction using low altitude training data. $R^2 = 0.76$	22
2.6 Regression model terms and coefficients for sand content prediction using high altitude training data. $R^2 = 0.64$	22
2.7 Regression model terms and coefficients for sand content prediction using combined altitude training data. $R^2 = 0.78$	22
2.8 Prediction errors for clay prediction models.....	23
2.9 Regression model terms and coefficients for clay content prediction using low altitude training data. $R^2 = 0.64$	24
2.10 Regression model terms and coefficients for clay content prediction using high altitude training data. $R^2 = 0.39$	25
2.11 Regression model terms and coefficients for clay content prediction using combined altitude training data. $R^2 = 0.51$	25
2.12 Prediction errors for OM content prediction models	26
2.13 Regression model terms and coefficients for OM content prediction using low altitude training data. $R^2 = 0.79$	26
2.14 Regression model terms and coefficients for OM content prediction using high altitude training data $R^2 = 0.69$	27
2.15 Regression model terms and coefficients for OM content prediction using combined altitude training data $R^2 = 0.74$	27

List of Tables (Continued)

3.1	Big Pivot Field SSURGO soil type distribution	45
3.2	Don Still Field SSURGO soil type distribution.....	45
3.3	Market Front Field SSURGO soil type distribution	45
3.4	Watermelon Rd. Field SSURGO soil type distribution.....	46
3.5	E7 Field SSURGO soil type distribution.....	46
3.6	Rusty Pivot Field SSURGO soil type distribution.....	47
3.7	CP4 & 13 SSURGO soil type distribution.....	47
3.8	Chicken House Field SSURGO soil type distribution.....	47
3.9	B6B Field SSURGO soil type distribution	47
3.10	C12 Field SSURGO soil type distribution.....	48
3.11	An example to illustrate general model construction:	57
3.12	Sand content prediction models utilizing Planet Labs imagery.....	60
3.13	Sand content prediction models utilizing Google Earth imagery.	60
3.14	Sand content prediction models utilizing both Planet Labs and Google Earth imagery.....	61
3.15	Clay content prediction models utilizing Planet Labs imagery	62
3.16	Clay content prediction models utilizing Google Earth imagery.....	63
3.17	Clay content prediction models utilizing both Planet Labs and Google Earth imagery.....	63
3.18	OM content prediction models using Planet Labs imagery.	66
3.19	OM content prediction models using Google Earth imagery.	67
3.20	OM content prediction models using both Planet Labs and Google Earth imagery.	68

List of Tables (Continued)

3.21	Sand content prediction errors across imagery sources.	72
3.22	Clay content prediction errors across imagery sources.....	73
3.23	OM content prediction errors across imagery sources.....	73
3.24	Analysis of “wet” and “dry” points in Field E7. Wet soil significantly alters pixel color characteristics among points with 1% variation in known sand content	74

LIST OF FIGURES

Figure	Page
2.1 Distribution of measured sand (a), silt (b), clay (c) and OM content percentages (d) across all sample sites used in regression modeling.....	11
2.2 Desktop view of FlyLitchi flight plan for Pond Field.....	13
2.3 Bare soil image including areas outside of field boundary, edited to include black box showing exclusion of pixels outside field boundary	14
2.4 Sample distribution of red colorspace bins output by Batch Load Image Processor for an image.....	16
2.5 Actual by predicted data for sand prediction model applied to testing data of model resulting in lowest mean error. 1:1 line included in figure.....	21
2.6 Actual by predicted data for clay prediction model applied to testing data of combined altitude model when applied to low altitude data model resulting in lowest mean error. 1:1 line included in figure.	24
2.7 Actual by predicted data for OM content prediction model applied to testing data of combined altitude model when applied to low altitude data model resulting in lowest mean error. 1:1 line included in figure.	28
2.8 Contoured actual (a) and predicted (b) sand content zones for Field E7. For each map, divisions between Low, Medium, and High were set so that each of the three zones was equal in area, or equal to one third of the field area.	29
2.9 Model making realistic prediction although image is not strictly of bare soil. (a) contains crop residue, (b) contains crop in field.....	31
2.10 Erroneous predictions outside of expected range: anomalous prediction due to lens flare and image overexposure (a) and anomalous prediction due to weed presence in image (b)	32
3.1 Ten captured images of a field using Planet Labs Data to capture average variation caused by time of day, soil moisture, and image sensor variability. The Don Still field is shown as an example	43

List of Figures (Continued)

3.2	Distribution of measured sand (a), silt (b), clay (c) and OM content percentages (d) across all Planet Labs image sample sites used in regression modeling.	49
3.3	Distribution of measured sand (a), silt (b), clay (c) and OM content percentages (d) across all Google Earth image sample sites used in regression modeling.	50
3.4	Georeferencing process within Spatial Image Digitizer. Image on left represents basemap image with field boundary in ArcView Shapefile overlaid. Image to right shows the same field, but image captured from satellite imagery.	51
3.5	Example workflow of reducing SID point dataset (a) to within “sample circles”. Using CPG, “sample circles” of specified diameters were created (b, polygon dataset). Using PPMU, only data within these sample circles is retained for regression model development (figure c, point dataset, clipped to polygons in figure b).....	53
3.6	Actual by predicted data for sand prediction model developed with Planet Labs data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure	58
3.7	Actual by predicted data for sand prediction model developed with Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure	59
3.8	Actual by predicted data for sand prediction model developed with Planet Labs and Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.....	59
3.9	Actual by predicted data for clay prediction model developed with Planet Labs data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure	64
3.10	Actual by predicted data for clay prediction model developed with Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure	64
3.11	Actual by predicted data for clay prediction model developed with Planet Labs and Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.....	65

List of Figures (Continued)

3.12	Actual by predicted data for OM prediction model developed with Planet Labs data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure	68
3.13	Actual by predicted data for OM prediction model developed with Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure	69
3.14	Actual by predicted data for OM prediction model developed with Planet Labs and Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.....	69
3.15	Wet and dry points within Field E7. The two points are within 1% known sand content, but exhibit significantly different R, G, and B values, highlighting the importance physical factors, especially soil moisture, can have on regression modeling	75
3.16	Illustration of zone creation utilizing sand content prediction models. Image (a) shows actual sand content of contoured data, while (b) shows results of applying prediction model	77
3.17	Application of sand content prediction model to field not included in testing dataset located in Sylvania, Georgia. Image (a) illustrates bare soil condition, (b) illustrates results of sand prediction model, and (c) illustrates sand content data as would be utilized for zone delineation.....	78
3.18	Application of sand content prediction model to field not included in testing dataset located in Boiling Springs, North Carolina. Image (a) illustrates bare soil condition, (b) illustrates results of sand prediction model, and (c) illustrates sand content data as would be utilized for zone delineation	80
4.1	Orthophoto of test plots stitched using OpenDroneMap. Orthophoto creates stitched image of an entire sample area. Use of georeferencing points allows for image to be analyzed on a plot by plot basis.....	91
4.2	Illustration of workflow of point data collection using Litchi app. Image locations exported from GIS software (a), and are loaded into FlyLitchi app (b). UAV collects image at each sampling site (c)	94
4.3	Comparison of two UAV images in same field collected using point data method. Point data method allows for quick "spot checking" across large areas.	95

List of Figures (Continued)

4.4	NDVI extracted (a) and contoured (b) from RGB UAV using Point Sampling data collection method	96
4.5	Image illustrating manual image annotation using VGG Image Annotator. Polygon drawn around each cotton bloom to build "dictionary" of training images. Each bloom given a unique number	101
4.6	MaskRCNN cotton bloom identification model example images. MaskRCNN applied grayscale filter to entire image, and applied yellow mask to white cotton blooms.....	102

CHAPTER ONE

INTRODUCTION

Remote sensing technologies allow for data to be collected in ways previously thought to be impossible. This study explores using two remote sensing technologies in order to predict sand, clay, and organic matter content in soil using red, green, blue (RGB) imagery. This study explores the development of regression models which can be used to predict sand, clay, and soil organic matter content from RGB imagery. Two image types are presented in this study; one uses imagery captured using a consumer level DJI UAV capturing RGB imagery, and the other uses Internet- available and free satellite imagery provided by Google Earth and Planet Labs. These studies were not performed to develop a replacement for traditional soil sampling methods, but rather as a supplement to that data, allowing for fast mapping of relative differences across large areas.

When developing models based on UAV imagery, two flight altitudes were chosen to determine if altitude was a significant determinant of soil texture content. The UAV used in this study was a DJI Phantom 3 Advanced, equipped with an RGB camera. While many studies utilize near infrared, infrared, hyperspectral, and multispectral imaging sensors, this study was performed with technology that an end user (grower, crop consultant, etc) can purchase for roughly \$1,500.

For development of satellite imagery-based models, Planet Labs was chosen as an image provider because of their large constellation of satellites, which allow for images of a study area to be captured almost daily. These images, while frequent, offer a lower spatial

resolution than those found on Google Earth. Google Earth, on the other hand, does not provide imagery at regular intervals, but provides higher resolution imagery.

The point sampling method for aerial data collection serves as an alternative to traditional, whole field image stitching and mapping. This method allows for rapid data collection through image analysis. An overview of common technologies associated with UAVs, artificial intelligence, and machine learning is presented to provide the reader with a broad-level view on current technologies in these areas.

CHAPTER TWO

SOIL TEXTURE AND ORGANIC MATTER PREDICTION FROM CONSUMER LEVEL UAV WITH RGB CAMERA

Introduction

Unmanned Aerial Vehicles (UAVs), commonly referred to as drones, are increasing efficiency and data collecting ability in many industries throughout the world; the field of precision agriculture is no exception. In 2016, a market report estimated worldwide UAV usage in agriculture to be a 32.4 billion US dollar industry, with continued growth expected (Pricewaterhouse Coopers, 2016). SZ DJI Technology Co., Ltd. (Shenzhen, China), commonly known as DJI, holds the largest market share of UAV sales, reporting \$2.83 billion in revenue in 2017 (Ying, 2018). The Federal Aviation Administration reported in 2019 that approximately 95% of UAVs that have been registered in the United States are considered consumer-grade, which they define as having a unit cost “below US \$10,000, with an average unit price of around \$2,500” (Federal Aviation Administration, 2019). Many of these devices are equipped with cameras that capture images in the red, green, and blue spectra (RGB cameras). These images are visually similar to those captured by modern digital cameras and smartphones, making them extremely popular in the aerial photography and videography fields. Due to the quality of these images, RGB aerial imagery could provide meaningful data for many applications.

Collection, aggregation, and analysis of soil texture and nutrient data is a necessity in the field of precision agriculture, since soil texture variability is among the most

consistent variables for defining crop yield potential. Soil texture is defined as the relative mass compositions of sand, silt, and clay particles found within a given soil sample, and is classified by the United States Department of Agriculture (Jaja, 2016). Soil texture and properties can influence physical and chemical factors such as drainage, water holding capacity, organic matter content, and cation exchange capacity, and nutrient retention, which can influence crop selection and productivity in an area. Typically, soil texture is determined in one of two ways: by use of the “hand-feel method” (Thien, 1979), or through particle size analysis, using the hydrometer (Bouyoucos, 1962) or pipette method (Miller et al., 1987).

The “hand-feel method”, also known as the “ribbon test”, involves holding a wetted sample of soil in the hand, and drawing conclusions about its texture based upon how well the soil forms a ribbon between the fingers. This process allows for subjectivity in the determination of results; one person may consider a sample to be gritty, while others may consider it to be smooth. Furthermore, the ribboning process is dependent upon moisture levels, and different moisture contents can result in different conclusions. Farmers may also elect to collect soil electrical conductivity (EC) measurements across a field. Soil EC results are derived through the use of an implement such as a Veris Sensor Cart (Veris Technologies, Salina, KS.), which utilizes coulter-electrodes to measure electrical current as the sensor cart is pulled through a field behind a tractor or other vehicle. While soil EC is generally proportional to clay content and inversely proportional to sand content, these relationships can be affected by properties such as soil water content and temperature at the time of mapping (McCutcheon et al., 2006). Therefore, making accurate soil texture

predictions from EC mapping is not practical. A well-defined and repeatable method for quantifying soil texture is referred to as the hydrometer method (Bouyoucos 1962). The hydrometer method is considered accurate because particle settling velocity and particle size are strongly correlated. In this method, dry, screened soil samples are mixed with a surfactant and shaken. Then a hydrometer is used at specific time intervals to measure apparent density of the soil-water suspension. As sand, silt, and clay particles fall out of suspension at different time intervals, the percentage of sand, silt, and clay can be calculated.

Like soil texture classification, soil organic matter (OM) has been considered a key indicator of the productivity and yield potential of a soil. Soil organic matter is defined by the Soil Science Society of America (2020) as “the organic fraction of soil, including plant, animal, and microbial residues, fresh and at all stages of decomposition, and the relatively resistant soil humus”. Knowledge of OM content can greatly influence nutrient recommendations for a specific area. Reeves (1997) compiled and summarized a series of both short and long term studies across multiple cropping systems and soil management practices. He noted a decline of crop yield and plant available nitrogen over time as soil OM was depleted or decreased. This makes quantifying OM content important when considering the addition of soil fertilizer, or reintroducing nutrients into the soil by disking. Soil OM content is strongly correlated to the amount of organic carbon that is contained within a soil. Organic carbon mostly consists of the cells of microorganisms, decomposing plant and animal residues, humus synthesized from residues, and highly carbonized, elemental forms of carbon such as charcoal, graphite, and coal (Nelson and Sommers,

2015). The most widely used process for determining soil organic matter is the “Loss on Ignition” method, during which soil samples are dried in an oven to remove moisture, reweighed, and then ignited in a furnace to incinerate the organic compounds (Ball, 1964). The samples are then weighed again, and the pre- and post-ignition weight difference is used to calculate the amount of organic matter present. Both Ball (1964) and Nelson and Sommers (2015) concede that this method is not precise, nor perfectly quantifiable, as the process destroys both the organic and inorganic carbon, instead of solely the organic carbon used to measure soil organic matter.

In recent years, both Veris (Veris Technologies, Salina, Kans.) and Precision Planting (Precision Planting, Tremont Ill.) have introduced technologies to measure soil properties including organic matter in real time, eliminating the need for laboratory work. Both the Veris iScan and Precision Planting SmartFirmer utilize visible and near infrared (Vis-NIR) sensors to quantify soil reflectance, which can be then correlated to soil organic matter using the process outlined by Sudduth and Hummel (1993). The Vis-NIR module of the Veris iScan can be mounted to many implements including tillage tools and fertilizer bars (Veris.com), and records data during the course of normal field operation. The Precision Planting SmartFirmer mounts to compatible planters behind the seed tube (Precisionplanting.com), and also records data as the planter is being operated. Lund and Maxton (2019), found that once calibrated, the iScan module produced a RMSE of 0.22% OM across all sample sites, and the SmartFirmer produced a RMSE of 0.24% OM across all sample sites studied. Both of these technologies are currently commercially available.

Although soil texture classification has been performed in studies using image analysis, (Zhang et al., 2005; Breul et al., 2006), limited research exists on soil texture classification an aerial imaging platform with an RGB camera instead of a multi-spectral imaging platform. Chung et al., (2010) used RGB imagery from a surface level camera to predict soil texture using linear regression. The most promising model results calculating silt percentages, had a coefficient of determination (R^2) value of 0.20. A conclusion of this study was that better relationships could likely be calculated by using all histograms for R, G, and B values, which would result in more variables to be used in the linear regression. Morais et al., (2019), also attempted to predict soil texture using a ground-based camera. This study analyzed soils at a micro level using a microscope. This study utilized multiple color systems, such as grayscale and HSV (hue, saturation, and value), both of which are derived from RGB, in addition to RGB. When compared to other colorspace models, this study found the lowest R^2 values when using solely RGB data, whereas utilizing a combination of RGB, HSV, and Grayscale resulted in the highest coefficient of determination of 0.933.

Development of a method to measure soil texture and organic matter percentages from a consumer-level UAV could allow soil texture and soil OM analyses to be completed at a much faster pace and on a larger scale for zone delineation used in precision agriculture. If an accurate method with a UAV can be derived, the cost of sampling could be reduced, and this methodology would provide a framework for other UAV based data collection.

The objective of this study is to develop, through the use of linear regression, separate equations to predict sand, clay, and soil OM content from images captured by a consumer-grade UAV equipped with an RGB camera.

Methods and Materials

Field Selection

Fields in this study were selected based on visual observations of the variability of soil color across the field. Field selection was also contingent on ground and crop cover; only fields with bare soil conditions were selected due to the effects standing crop or crop residue would have on the resulting image color values. Three fields were selected for testing, with one site year of data used from each. Soil Survey Geographic Database (SSURGO) soil data (Soil Survey Staff, 2019) were used for general characterization of the soils present, although these data were not used for development of the models presented in this study. Fields C12 (Table 2.1) and E7 (Table 2.2) located at the Clemson University Edisto Research and Education Center in Blackville, South Carolina, were selected for analysis in this study. Field C12 is located at (33.34896°N, 81.32062°W), and Field E7 is located at (33.34465°N, 81.31752°W). These two fields consist exclusively of sand and loamy sand classifications, with exact soil types and percentages shown in the respective table below for each field. Images for these fields were collected on February 26, 2019 using a DJI Phantom 3 Advanced UAV under fair sky conditions as reported by a WeatherUnderground (The Weather Company, San Francisco, Cal.) weather station located near the fields at the time of image collection. Both fields had been disk harrowed

within 10 days prior to image collection to achieve a completely bare soil condition free of any weeds or crop residue.

Table 2.1: C12 Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area, ha(ac)	Percent of Area
DaA	Norfolk loamy sand, 0 to 2 percent slopes	0.32 (0.8)	6.9%
DaB	Barnwell loamy sand, 2 to 6 percent slopes	1.33 (3.3)	28.5%
FuA	Wagram sand, 0 to 2 percent slopes	2.14 (5.3)	45.4%
FuB	Wagram sand, 2 to 6 percent slopes	0.04 (0.1)	1.2%
VaB	Barnwell loamy sand, 2 to 6 percent slopes	0.85 (2.1)	18.0%

Table 2.2: E7 Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area, ha(ac)	Percent of Area
DaB	Barnwell loamy sand, 2 to 6 percent slopes	2.22 (5.5)	29.5%
FuC	Ailey sand, 6 to 10 percent slopes	0.16 (0.4)	2.3%
OrB	Barnwell loamy sand, 2 to 6 percent slopes	5.14 (12.70)	68.2%

The Pond Field (34.6561°N,82.8173°W) is located in Clemson, South Carolina, and is part of the Clemson University Piedmont Research & Education Center. Located in the Piedmont region of South Carolina, the majority of, Pond Field consists of clay loam type soils (Table 2.3). Images for this field were collected on April 27, 2019 under fair sky conditions. This field was disk harrowed within 10 days prior to image collection.

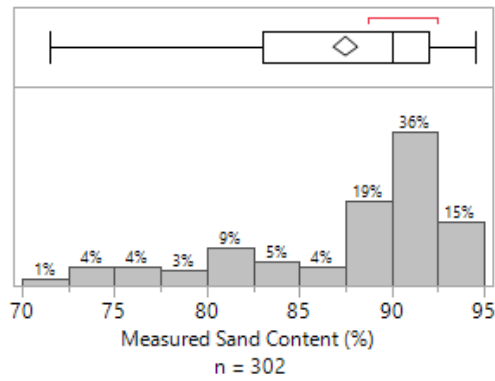
Table 2.3: Pond Field SSURGO Soil Type Distributions

Map Unit Symbol	Map Unit Name	Area, ha(ac)	Percent of Area
CeC3	Cecil clay loam, 6 to 10 percent slopes, severely eroded	3.76 (9.3)	83.8%
CIB2	Cecil sandy loam, 2 to 6 percent slopes, eroded	0.73 (1.80)	16.2%

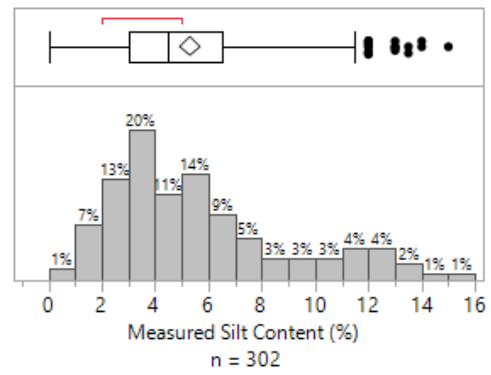
Soil Data Collection and Analysis

Fields selected for this study were assigned 0.101 ha (.25 ac) grid squares using Trimble Ag Desktop Software (v2019.1.0, Trimble Inc., Sunnyvale, Cal.). The GPS coordinates of the grid centers were exported to a comma separated values (CSV) file. The file was then loaded into Soil Sampling Utility (v.1.0.1.10, Clemson University Precision Agriculture, Blackville, SC), and navigation to sample positions was conducted by using a BU-353S4 USB GPS Receiver (GlobalSat WorldCom, New Taipei City, Taiwan) with WAAS, DGPS correction. At each sample site, eight soil cores were collected from a 305 cm (120 in.) radius around the center position of each grid square. Each sample core was collected from the top 15 cm (4 in.) of the soil profile using a soil probe with diameter 2.54 cm (1 in.)

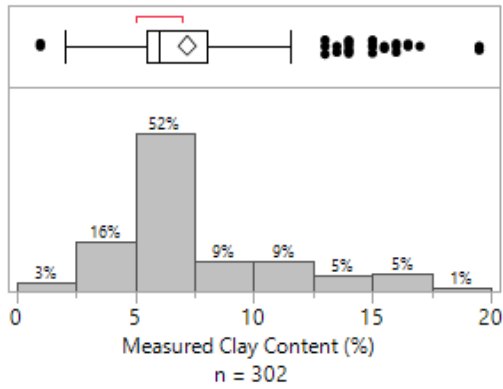
Each sample was passed through a #10 (2 mm) sieve to remove grained sized particles and residue and divided into two equal subsamples, each subsample weighing at least 100 g. One subsample was processed to determine the percentage sand, silt, and clay using the Hydrometer Method as outlined by Huluka and Miller (2010). The other sample was processed to determine OM content using the Loss on Ignition process outlined by Zhang and Wang (2014). Soil texture and OM were reported as mass composition of sand, silt, and clay. Figure 2.1 parts (a), (b), (c), and (d) shows distributions of measured sand, silt, clay and OM content compositions respectively across all sample sites, as these ranges represent the known sand, clay, and OM values which will be represented during regression modeling.



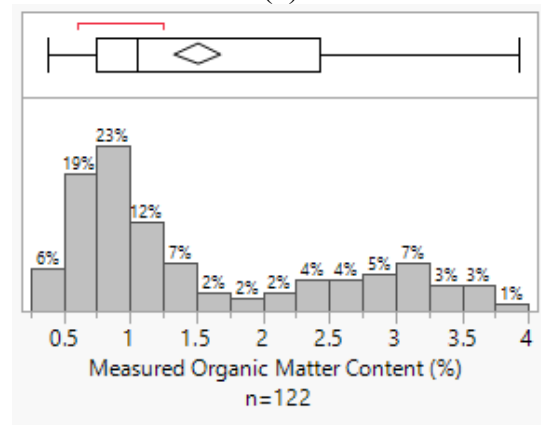
(a)



(b)



(c)



(d)

Figure 2.1: Distribution of measured sand (a), silt (b), clay (c) and OM content percentages (d) across all sample sites used in regression modeling. Sand content ranged from 71.5% - 94.5%, silt content ranged from 0%-15%, clay content ranged from 1% - 19.5%, and OM content ranged from 0.38%-3.92%.

Aerial Imagery Collection

The CSV file containing grid point centers was used in conjunction with FlyLitchi Mission Hub (VC Technology LTD., London, England) to create a flight path for the UAV to follow (Figure 1.3). FlyLitchi was chosen as the application to be used in this study because it allows for UAV flight plans to be programmed before arriving at the sampling site. Additionally, the “Waypoint” flight planning mode allows for easy integration of GPS coordinates to be loaded into the flight plan. For this study, each previously saved grid center was loaded as a waypoint. At each grid center, the UAV was programmed to wait three seconds for the UAV to stabilize, then capture an image. Two separate flight plans were created for each field, one plan at 21 m (70 ft) above ground level (AGL), and one plan at 30 m (100 ft) AGL (Figure 2.2). These altitudes were calculated to ensure that certain areas, as discussed in the section below, could be captured in a single image. Additionally, different altitudes were flown to explore whether varying flight altitude would affect model prediction ability. Using the Pix4D Ground Sampling Distance Calculator (Pix4d S.A., Prilly, Switzerland), along with the camera specifications provided by DJI, it was calculated that the UAV was capable of capturing 0.10 ha (0.25 ac) in a single image at an elevation of 20 m (65 ft) AGL, and 0.20 ha (0.5 ac) at an elevation of 29 m (95 ft) AGL. To account for slight differences in altitude sensing, the flight elevations of this study were selected to be 21 and 30 m (70 and 100 ft) AGL. Each field was flown at a single elevation, then repeated at the second flight elevation.



Figure: 2.2 Desktop view of FlyLitchi flight plan for Pond Field

Aerial imagery was captured using a DJI Phantom 3 Advanced UAV (SZ DJI Technology Co., Ltd., Shenzhen, China). The UAV was equipped with a permanently attached RGB Camera model DJI FC300s, which captures 12.0-megapixel images, each with dimensions of 4000 pixels wide by 3000 pixels tall, an aspect ratio commonly referred to as 4:3. The UAV was controlled using an Apple iPad model A1823 (Apple Inc., Cupertino, Cal.) and FlyLitchi application set in Waypoint mode. All images were captured from a straight-down, or nadir perspective. To fit environmental and available lighting conditions at the time of image capture, all images were captured with white balance set to sunny, ISO set to 100 and an aperture setting of f2.8. The camera automatically determined optimum shutter speed for each image, and shutter speeds ranged from 1/750 sec to 1/100 sec.

Aerial Imagery Processing

All images captured were saved in Portable Network Graphics (PNG) format. Captured images were not modified in resolution or compression. Some images contained areas outside the field boundary, to remove these areas from the images, images were opened in Microsoft Paint (Microsoft Corporation, Redmond, Washington), and edited to include either a black box, RGB (0, 0, 0), or white box, RGB (255, 255, 255), covering areas outside of the field boundary (Figure 2.3). The software used to process the images was programmed to ignore white and black pixels. This was performed to ensure pixels located outside of the field boundaries would not be considered for the development of regression models.



Figure 2.3: Bare soil image including areas outside of field boundary, edited to include black box

Images were loaded into Batch Load Image Processor (BLIP) v.1.1 software developed by Clemson University, to extract and summarize pixel colorspace information. BLIP, a Windows application written in Microsoft Visual Studio Express 2013 (Microsoft Corporation, Redmond, Wash.), is a software application that allows either a single image,

or a directory of images, to be processed to extract and summarize image color data, and save data as a CSV file for further analysis. This software extracts colorspace values for each pixel in an image across three components: red, green, and blue.

For a given pixel, each of these components can range in value from 0 to 255, with the number representing how much light or intensity of the color is added. A color of RGB (255, 0, 0), for example, is a solid red color, and RGB (255, 255, 255) represents solid white. By combining values across all three components, it is possible to identify over 16 million colors. Red, green, blue imagery is the “base color model for most applications” (Ibraheem et al., 2012) because no additional steps are required to display the image. In other words, images captured look identical to those captured from a cell phone, webcam, or digital camera.

Batch Load Image Processor scans each pixel of an image, and extracts the red, green, and blue color values for each pixel. Then the program computes an average value for red, green, and blue values over an entire image. BLIP computes values for other derived color values calculated from the red, blue, and green components such as hue, chroma, and brightness. In addition to these derived values, BLIP separates each of the red, green, blue, hue, and brightness values into further divisions, referred to as “bins” for further analysis. Bins for red, green, and blue colorspace values are divided into 32 equal divisions of the full range of possible values. Due to the values of each component ranging from 0 to 255, each bin represents a range equal to eight colorspace values. For example, bin R0 is defined as red component values between 0 and 7, with R1 counting red component values between 8 and 15. These bins can be used to create a histogram showing

the distribution of pixel counts which fall within each bin, as illustrated in Figure 2.4. All outputs and binned terms are stored by BLIP as a CSV file with each row representing one image. Each column contains values that were factors in this study for developing the regression models to predict sand, clay, and OM content percentages based on soil texture color. The CSV files containing summarized image data were appended with the soil texture and organic matter data corresponding to each image, or sample site. This resulted in a single, tabular file for each of the two flight elevations, the files containing a row for each unique sample site, as well as the extracted image data and soil data for that site.

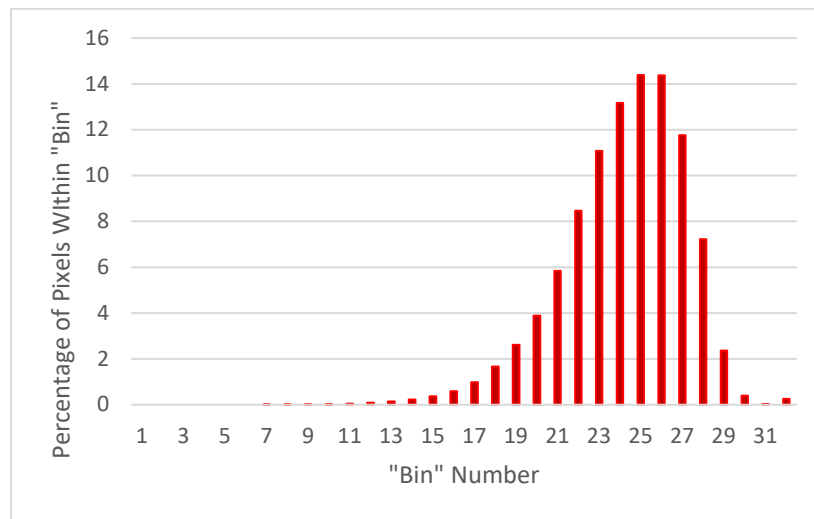


Figure 2.4: Sample distribution of red colorspace bins output by BLIP for an image.

Regression Model Development

All regression modeling and statistical analyses were analyzed in JMP Pro v.14.1.0 (SAS Institute Inc., Cary, North Carolina). Three sets of models were independently created for prediction of sand, clay, and OM content: one set of models for images from each of the two flight altitudes and a third set of models using combined image data from both flight

altitudes. For each set of models, each sample site was randomly assigned to one of two classifications: a training class containing 80% of sample sites and a testing class containing 20% of samples sites. This classification was performed to ensure models were not tested on the same datapoints which were used to develop them. Using the stepwise model fitting personality, multiple linear regression models were created using both forward and backward direction and both minimum AICc and minimum BIC stopping rules. For each model, the term being predicted (e.g. sand, clay, or OM content) was assigned as the response variable, y , while the model effect, x , terms were assigned as being the image color data extracted by BLIP. Transformations of all BLIP outputs were also considered as model effects, including square root, square, cube root, cube, log, and reciprocal. Indices of the red, green, and blue component values were created using combinations of addition, subtraction, multiplication, and division of combinations of the color components. Some examples include: $(R+G+B)$, $(R+B)*(R-B)$, $(R*G*B)$, and $(R+G+B)/3$. Additionally, combinations utilizing addition, subtraction, multiplication, and division of the red, green, and blue variables were added to the regression effect set.

Multiple collinearity was reduced by removing any term with a variance inflation factor (VIF) of greater than 5, as suggested by Kutner, et al. (2005). Regression outliers, or heavily influential points, were removed from consideration using Cook's Distance; any data point with Cook's Distance values greater than 1.0 were excluded as suggested by Hair, et al (1998). Upon exclusion of a regression outlier, the stepwise model iteration was restarted. Terms with low significance ($p\text{-value} > 0.05$) were eliminated until all remaining terms satisfied VIF, Cook's Distance, and $p\text{-value}$ criteria.

Results and Discussion

Models were developed to predict sand, clay, and OM content as a function of RGB, aerial image data. Prediction models were developed for one data subset including only lower altitude imagery, one including only higher altitude imagery, and one including both higher and lower altitude imagery. These models were evaluated to determine whether altitude influenced prediction model accuracy. While both elevations require the same flight time and post processing resources, lower altitude images can capture more details than higher altitude imagery, at the expense of a smaller surface area captured in each image. Higher altitudes and their larger ground surface areas represented may result in a more representative sample of soil conditions in an area being obtained, albeit with less resolution. At an altitude of 21 m (70 ft), captured images were of a ground sample distance (GSD) equal to 0.92 cm px^{-1} (0.36 in. px^{-1}). At an altitude of 30 m (100 ft), GSD for images was 1.32 cm px^{-1} (0.52 in. px^{-1}).

Some models were able to better predict for imagery from elevations for which they were not trained. For example, the model for predicting OM content from imagery at the higher altitude demonstrated numerically less error when applied to imagery collected at the lower altitude than that same model when applied to imagery from the altitude with which it was trained. The model developed using both low and high-altitude datasets, combined, was evaluated for all data, as well as, for data from each altitude, independently. This model resulted in a lower prediction error on the low altitude imagery subset than the model developed solely from the low altitude imagery. Each model developed was tested on both datasets independently in order to compare prediction errors.

Model prediction error was calculated using the testing class for each component modeled and is presented herein in units of percent sand content, percent clay content, and percent OM content. Prediction error was performed only on the data assigned to the testing class, and was calculated using the generalized formula:

$$AbsErr = |Predicted - Actual|, \quad (2.1)$$

where *AbsErr* represents absolute prediction error of the modeled component (e.g. percent sand, clay, or OM content), *Predicted* represents the predicted value of the modeled component for a given image, and *Actual* represents the measured value of the component for the sample collected at that image.

Sand Content Prediction Models

Table 2.4 illustrates results in sand content prediction error from application of the three developed models as applied to the three datasets of varying altitude. The number of images used in each training class of low altitude, high altitude, and both (low and high) altitudes were 119, 123, and 242, respectively. In this table, the columns demonstrating error at 50% and 90% confidence represent the prediction errors for which 50% and 90% of all prediction errors were lower in value. For example, an “error at 90% confidence” value of 6.23 demonstrates that 90% of the absolute values of prediction errors were less than 6.23 % sand content. A means comparison (student’s t-test, $\alpha = 0.05$) suggested no significant differences in prediction ability as related to combinations of flight altitude used for training and flight altitude used for testing. These results suggest that flight altitude was not a critical factor in model development and application in this study. While it does not explain results for all possible altitudes that might be used, it does suggest that the altitude

used for model development may not necessarily have to match altitudes used for model application. Figure 2.5 illustrates actual by predicted data as applied to the testing data of the model combination resulting in the lowest mean error, in this case, the result of applying the model developed using the combined altitude data to the low altitude data. A 1:1 line is also included in the figure.

Table 2.4: Prediction errors for sand content prediction models.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
Low Altitude	Low Altitude	32	2.30	A	1.81	6.23
High Altitude	High Altitude	28	2.51	A	1.61	7.87
Low Altitude	High Altitude	26	2.22	A	1.19	6.22
High Altitude	Low Altitude	32	2.70	A	2.57	5.99
Both Altitudes	Low Altitude	32	2.16	A	1.88	5.84
Both Altitudes	High Altitude	28	2.24	A	1.37	6.03
Both Altitudes	Both Altitudes	60	2.20	A	1.50	5.94

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

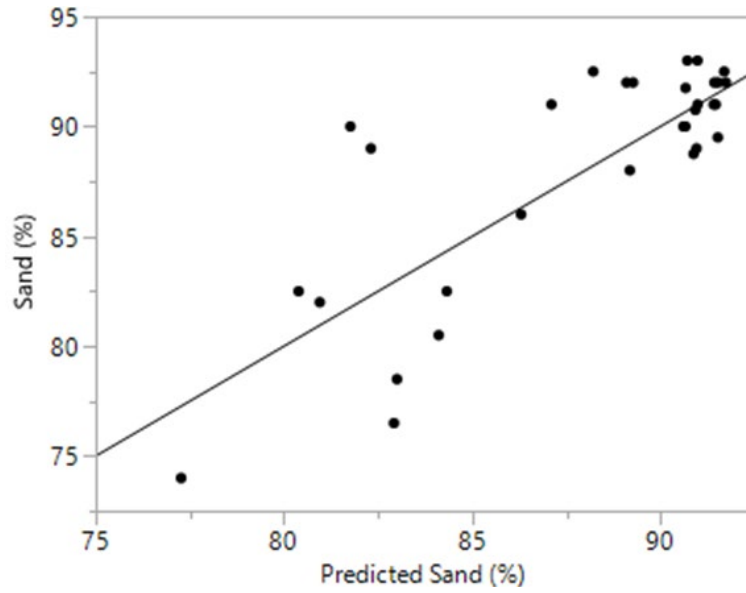


Figure 2.5: Actual by predicted data for sand prediction model applied to testing data of model resulting in lowest mean error. 1:1 line included in figure.

Model terms, coefficient estimates, and standard error values for low, high, and combined altitude models are shown in Tables 2.5, 2.6, and 2.7, respectively. Models may be implemented using the generalized formula provided in Equation 2.2 using values from Tables 2.5, 2.6, and 2.7.

$$y_0 = Intercept + (Term_0 * Estimate_0) + (Term_1 * Estimate_1) \dots + (Term_n * Estimate_n) \quad (2.2)$$

where y_0 represents the variable being predicted (e.g. sand content), *Intercept* represents the Estimate value for the Intercept term, $Term_0$ represents the first term below Intercept in the Term column, and $Estimate_0$ represents the value in the Estimate column corresponding to the row containing $Term_0$. In terms containing parentheses (e.g. G(1)), the value enclosed in parentheses represents the bin number, as previously discussed.

Table 2.5: Regression model terms and coefficients for sand content prediction using low altitude training data. $R^2 = 0.76$

Term	Estimate	Std Error
Intercept	81.7	1.25
BRT(23)	0.285	0.0553
G(1)	-119.0	53.4
G(25)	0.159	0.0589
G(28)	0.951	0.314
G(30)	-231.0	50.9
B(15)	-0.540	0.0536
B(18)	0.326	0.0652

Table 2.6: Regression model terms and coefficients for sand content prediction using high altitude training data. $R^2 = 0.64$

Term	Estimate	Std Error
Intercept	103.0	1.58
CHROMA	-51.9	6.38
R(30)	0.247	0.0714
B(1)	-11.2	4.06

Table 2.7: Regression model terms and coefficients for sand content prediction using combined altitude training data. $R^2 = 0.78$

Term	Estimate	Std Error
Intercept	92.1	0.549
R(31)	-0.124	0.0249
BRT(29) ³	0.00612	0.00176
G(27) ²	0.0105	0.00294
$\sqrt[3]{G(6)}$	-14.4	1.33
$\sqrt[3]{G(29)}$	-1.73	1.06
BRT(22) ³	0.000254	0.000131
$[R-G]^3$	-1.187e-5	2.093e-6

Clay Percentage Prediction Models

Clay content prediction model accuracy was evaluated using the same method used for sand content prediction. Number of images used in training class for low altitude, high

altitude, and both (low and high) altitudes were 119, 123, and 242, respectively. Similar to sand content prediction, no significant differences were noted when a means comparison of model prediction errors across altitudes was performed (student's t-test, $\alpha=0.05$). Error and means comparison results are illustrated in Table 2.8. Model coefficients (for application to Equation 2.2) for low, high, and combined altitude model predictions of clay content are shown in Tables 2.9, 2.10, and 2.11, respectively. In all three models, nearly all terms are “binned” terms. This may suggest that sorting images into color “bins” may help with analysis and prediction. An actual by predicted plot is provided in Figure 2.6, which illustrates the result of applying the combined altitude model to the low altitude data, a combination which, among sand, clay, and OM content prediction regularly resulted in low mean and error at 90% confidence. A 1:1 line is included in the figure.

Table 2.8: Prediction errors for clay content prediction models.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
Low Altitude	Low Altitude	32	2.22	A	1.44	5.86
High Altitude	High Altitude	28	1.96	A	1.30	5.86
Low Altitude	High Altitude	26	1.61	A	0.84	5.92
High Altitude	Low Altitude	32	2.21	A	1.36	5.59
Both Altitudes	Low Altitude	32	1.81	A	1.03	5.07
Both Altitudes	High Altitude	28	1.88	A	1.15	5.34
Both Altitudes	Both Altitudes	60	2.22	A	1.44	5.86

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

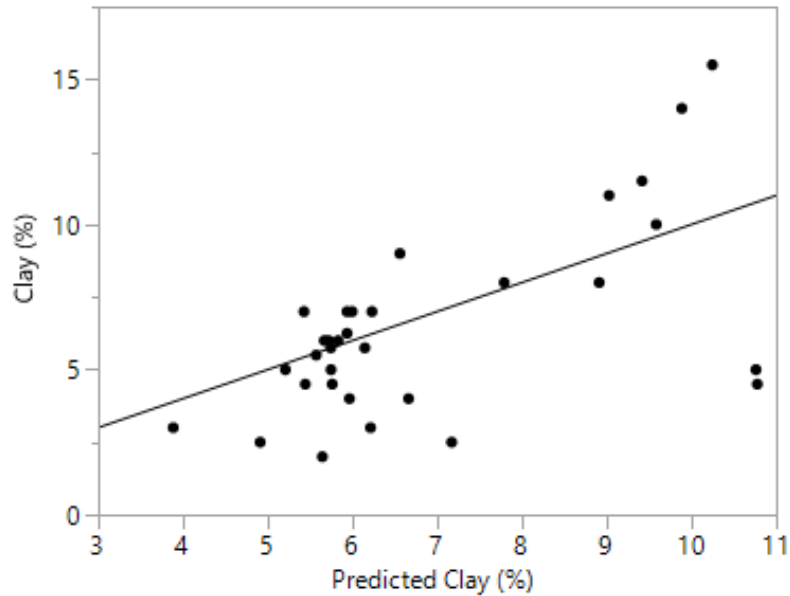


Figure 2.6: Actual by predicted data for clay prediction model applied to testing data of combined altitude model when applied to low altitude data model resulting in lowest mean error. 1:1 line included in figure.

Table 2.9: Regression model terms and coefficients for clay content prediction using low altitude training data. $R^2 = 0.64$

Term	Estimate	Std Error
Intercept	5.49	0.659
$\sqrt{[BRT(29)]}$	-18.1	8.92
$G(8)^3$	0.793	0.206
$G(30)^2$	4800	1030
$B(17)^3$	6.47e-4	9.8e-5
$B(19)^3$	3.55e-5	8.2e-5
$R-B^3$	3.1854e-6	4.6e-7
$BRT(20)^3$	-2.67e-3	1.21e-3
$BRT(21)^3$	1.27e-3	4.07e-4
$G(29)^3$	3.54	1.09

Table 2.10: Regression model terms and coefficients for clay content prediction using high altitude training data. $R^2 = 0.39$

Term	Estimate	Std Error
Intercept	10.5	1.97
B(1)	8.89	4.25
HUE ²	4.91e-3	0.0019
$\sqrt[3]{R(16)}$	-7.11	1.58
$(R-B)/(R+G+B)^3$	451	67.8

Table 2.11: Regression model terms and coefficients for clay content prediction using combined altitude training data. $R^2 = 0.51$

Term	Estimate	Std Error
Intercept	5.19	0.276
R(31)	0.126	0.0129
G(26) ²	5.36e-3	2.37e-3
$(R-G)/(G+B)^3$	157	16.8
G(28) ²	0.0252	7.59e-3
B(17) ³	3.55e-4	8.117e-5
B(28) ³	14500	5080

$R^2: 0.51$

OM Content Prediction Models

OM content prediction was evaluated in using the same methods as sand and clay contents, with results shown in Table 1.8. Number of images used in training class for low altitude, high altitude, and both (low and high) altitudes were 118, 122, and 240, respectively. A means comparison revealed no significant differences between different flight altitudes, with error results illustrated in Table 2.8 Model coefficients for OM content for low, high, and combined altitude models are given in Table 2.9, Table 2.10, and Table 2.11 respectively. An actual by predicted plot is provided in Figure 2.6 which illustrates the result of applying the combined altitude model to the low altitude data. A 1:1 line is included in the figure.

Table 2.12: Prediction errors for OM content prediction models.

Training Basis [a]	Testing Basis [b]	n [c]	Mean Error [d]	T-Test [e]	Error at 50% Confidence [d]	Error at 90% Confidence [d]
Low Altitude	Low Altitude	31	0.43	A	0.31	1.10
High Altitude	High Altitude	27	0.49	A	0.24	1.03
Low Altitude	High Altitude	26	0.35	A	0.20	0.94
High Altitude	Low Altitude	27	0.59	A	0.35	1.33
Both Altitudes	Low Altitude	31	0.38	A	0.20	0.99
Both Altitudes	High Altitude	28	0.34	A	0.22	0.93
Both Altitudes	Both Altitudes	58	0.43	A	0.14	1.22

[a] Imagery dataset whose training class was used for model development

[b] Imagery dataset whose testing class was used for prediction error analysis

[c] Number of images used in testing class

[d] Unit of measure for prediction error presented as % sand content

[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

Table 2.13: Regression model terms and coefficients for OM content prediction using low altitude training data. $R^2 = 0.79$

Term	Estimate	Std Error
Intercept	3.00	0.199
BRT(23)	0.0612	0.00734
B(15)	0.167	0.0108
$\sqrt{[BRT(25)]}$	0.274	0.0467
$BRT(20)^2$	0.00832	0.00199
$B(17)^3$	0.000161	2.237e-5
$G(26)^3$	4.253e-5	1.935e-5

Table 2.14: Regression model terms and coefficients for OM content prediction using high altitude training data. $R^2 = 0.69$

Term	Estimate	Std Error
Intercept	4.12	0.318
$\sqrt{[R(26)]}$	-0.700	0.0683
$BRT(7)^3$	75.9	39.3
$G(28)^3$	0.000441	0.000122
$\sqrt[3]{[G(6)]}$	2.17	0.390
$(G-B)/(R-B)^3$	4.10	0.939

Table 2.15: Regression model terms and coefficients for OM content prediction using combined altitude training data. $R^2 = 0.74$

Term	Estimate	Std Error
Intercept	5.44	0.375
$G(6)$	1.85	0.445
$R(30)^3$	6.389e-5	1.244e-5
$\sqrt{[B(31)]}$	0.463	0.179
$G(21)^2$	0.00422	0.000582
$G(17)^3$	0.00101	0.000175
$R(28)^2$	0.000674	0.000234
$\sqrt[3]{[R(25)]}$	-2.01	0.160
$G(28)^2$	0.00944	0.00162
$B(18)^2$	0.00196	0.00039
$B(16)^3$	6.429e-5	2.477e-5

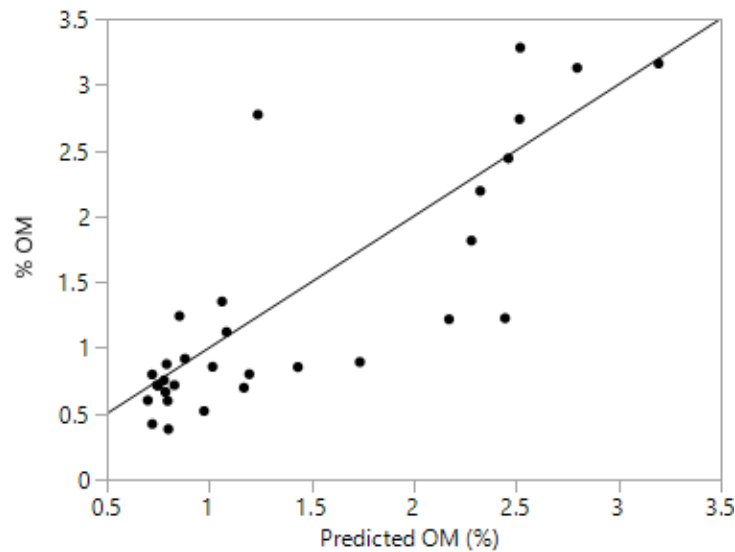


Figure 2.7: Actual by predicted data for OM content prediction model applied to testing data of combined altitude model when applied to low altitude data model resulting in lowest mean error. 1:1 line included in figure.

Application of Sand Content Prediction Model

Models developed have practical application, with one being the development of field management zones. Field management zones in agricultural production are often designed to group homogeneous soils together, seeking to maximize differences [in soil type and/or yield potential] between the zones and minimize differences within zones to aid in developing variable rate or zone base management prescriptions. To demonstrate this application, points with known sand content from Field E7 were compared to sand content predictions developed from the combined altitude model, which produced the lowest error percentage at 90% confidence. Three contoured zones of equal area were created, representing relative sand content percentages: Low, Medium, and High. The model shown in Table 2.7 was applied to bare soil images collected from Field E7 using Equation 2.2. The contoured zone map developed for actual sand content can be compared

to that for predicted sand content in Figure 2.8. Contour maps were produced using Trimble Ag. Desktop Software v.2020.01.



Figure 2.8: Contoured actual (a) and predicted (b) sand content zones for Field E7. For each map, divisions between Low, Medium, and High were set so that each of the three zones was equal in area, or equal to one third of the field area.

Effects of Foreign Imagery and Model Bounding

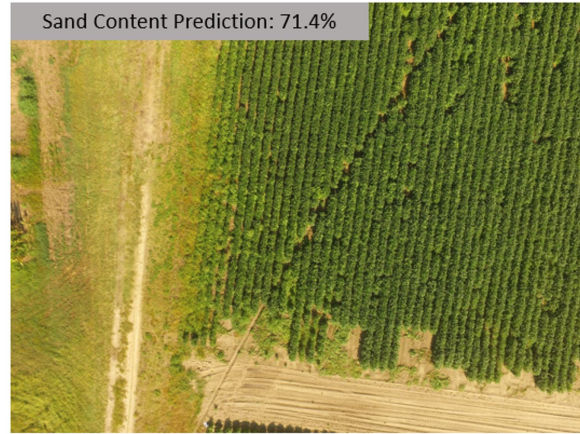
Typical of regression modeling, a model cannot be expected to accurately predict response variables when applied to data unlike that with which it was developed. This allows the potential for two types of erroneous results: predictions that are grossly inaccurate, yet within acceptable range (e.g. wrong prediction), and predictions that are outside of possible ranges (e.g. percentages which are negative or over 100 are impossible values). Examples are provided to illustrate prediction errors; in the examples, the combined altitude sand content prediction model is applied to images unlike those with which it was trained, and which are generally unsuitable for soil texture prediction, such as crop or residue presence. To rectify such errors, model bounds must be applied to each image analyzed. These bounds, defining acceptable ranges of each model term (as

exhibited in the training dataset), allow for determination of whether the image being analyzed is within the range of parameters used for model development. This should generally suggest if a particular image of unknown soil texture would result in a plausible result. If the bounds determine an anomalous image, as compared to the training dataset, the image would be omitted from prediction. For this study, models were bounded using a tolerance factor applied to each regression model term. Model bounding coefficients are provided in Appendix A for the models producing lowest mean error for sand, clay, and organic matter content percentages.

The first type of erroneous result is an inaccurate, yet realistic prediction. In this scenario, demonstrated in Figure 2.9, a model produces results that may not seem to be abnormal or out of range at first glance, but are not accurate. Both images in Figure 2.9 were captured at a similar altitude to that which was used for model development but contain features that were not included in the dataset. These features include crop residue, grass, and a planted field, which do not satisfy the condition of training images to be of “bare soil”. Despite these anomalies, the model attempts to predict sand content, and the results are within a normal, otherwise acceptable range (0 to 100%).



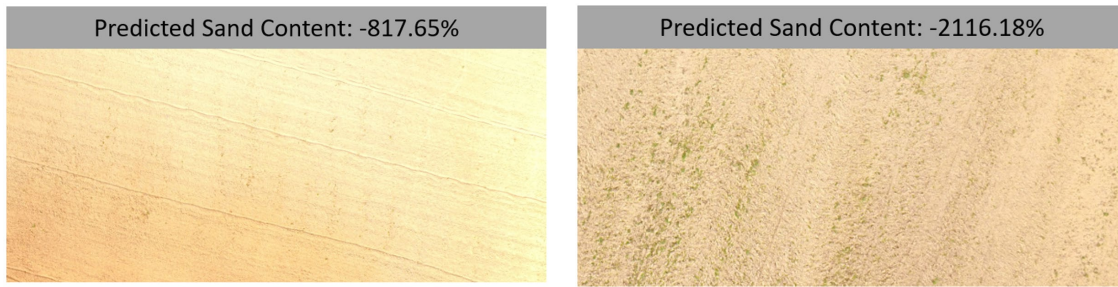
(a)



(b)

Figure 2.9: Model making realistic prediction although image is not strictly of bare soil. (a) contains crop residue, (b) contains crop in field.

Another type of inaccurate prediction occurs when the model produces unrealistic results, or results outside of acceptable values, such as sand, clay, or OM contents outside of the range of 0 to 100%. Some extreme examples of such predictions are illustrated in Figure 2.10, but similar, erroneous predictions can be results of structures, machinery, roads, or water located in an image, or due to inconsistent image characteristics, such as shadows, overexposure, or lens flare. These results are typically outside of the possible range of a texture content, and errors may be extremely large in either the positive or negative direction. Model bounding, as discussed earlier should generally prevent such prediction errors.



(a) (b)

Figure 2.10: Erroneous predictions outside of expected range: anomalous prediction due to lens flare and image overexposure (a) and anomalous prediction due to weed presence in image (b).

With either type of erroneous result, it is necessary to introduce numerical boundaries to model inputs and outputs, which will automatically determine if an image can be accurately analyzed, or whether it should be omitted. Images producing results either above 100% or below 0% can automatically be omitted, as they are out of bounds of a realistic prediction. Alternatively, these predictions within some range could be defaulted to the closer of 0% or 100%. Boundaries for model terms in this study were constructed based on the ranges observed in the training datasets; a binary output was calculated for each term for each processed image to determine whether the term was within the generally observed range of values in the training dataset. As discussed, a tolerance factor was applied to allow for extrapolation. Terms for an image falling within bounds were assigned a BoundCheck value of 1, while terms outside of bounds were assigned a BoundCheck value of 0. The product of the BoundCheck values for each term of a given image, was then used to demonstrate whether all of the terms for an image were within the model's boundary. Products equal to 1 demonstrated that all terms were within model bounds; products equal to 0 demonstrated that the value for at least one term was out of model bounds.

Effects of Environmental and Physical Soil Properties

Physical properties which would alter a soil's perceived color, like soil moisture content, would likely influence texture prediction. These factors were not evaluated in this study but could be included in modeling efforts by capturing images at the same point under varying moisture conditions, such as at specified intervals after a rainfall or irrigation event. Models in this study were developed from soils exhibiting relatively narrow ranges of sand, clay, and OM content. An expansion of this study into different soil types may lead to a universally applicable model or expose the need for different models for different soil textures. Models in this study were constructed using data from fields exhibiting strictly bare soil conditions, as all fields had been recently disk harrowed. These conditions are not always common, this is especially the case in the Southeast due to the recent rise in conservation tillage practices. Development of models using images from fields prepared using either strip till or no till practices could expand the applicability of models. Such models could use pixel classification to ignore pixels not likely to be bare soil pixels.

Conclusion

Soil texture and OM content prediction from a consumer-grade RGB UAV could be accurate, efficient, and effective. In this study multiple linear regression models were developed to predict sand, clay and OM content percentages from RGB imagery from a consumer-grade UAV. While the methods presented in this study are not intended to be replacements for currently accepted practices, models were able to predict within 5.84% sand content, 5.07% clay content, and 0.93% OM content, all at 90% confidence; i.e. 90% of the prediction errors were less than the values listed here. The models developed in this

study could very easily be implemented into an end-user image analysis application, providing a general estimate to those who need soil texture and OM data for zone creation or other applications. With the high speed and turnaround, methods developed utilizing remote sensing could be applied to complement data commonly used for zone development such as SSURGO maps provided by the USDA Web Soil survey (Soil Survey Staff, 2019).

Due to RGB being a device-dependent color model, different image capturing platforms will almost certainly produce different color values, even if they vary only slightly. As a result, models produced may only be accurate within a product family, or at least within a camera sensor specification. Robust models should combine imagery in the training datasets across product families.

Furthermore, grouping color values into “bins” appears to be important for analysis. Almost all terms in all models were “binned” terms, showing there is a strong benefit to splitting individual image color values into groups. In this study, only red, green, blue, and pixel brightness values were binned. Expanding these bins into other color calculations (e.g. hue) has a strong potential to reduce prediction error.

Flight altitudes of 70 and 100ft did not significantly change prediction abilities for sand, clay, or OM content. Additionally, inclusion of more data points across varying soil textures would expand model relevance and possibly increase model accuracy.

References

- Ball, D.F. (1964). Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soils. *J. Soil Soc.*, 15,84-92. doi:10.1111/j.1365-2389.1964.tb00247.x
- Breul, P., Gourves, R. (2006). In field soil characterization: approach based on texture image analysis. *J. Geotech. Geoenviron. Eng.*, 132(1), 102-107. doi: 10.1061/(ASCE)1090-0241(2006)132:1(102)
- Bouyoucos, G. J. (1962). Hydrometer method improved for making particle size analyses of soils. *Agron. J.*, 54(5), 464–465. doi:10.2134/agronj1962.00021962005400050028x
- Chung, S.-O., Cho, K.-H., Kong, J.-W., Sudduth, K. A., & Jung, K.-Y. (2010). Soil texture classification algorithm using RGB characteristics of soil images. *IFAC Proc. Volumes*, 43(26), 34–38. <https://doi.org/10.3182/20101206-3-jp-3009.00005>
- Federal Aviation Administration. (2019). Unmanned Aerial Systems. Retrieved from https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/Unmanned_Aircraft_Systems.pdf
- Hair, J., Anderson, R., Tatham, R. and Black W. (1998). *Multivariate Data Analysis (fifth edition)*. Englewood Cliffs, NJ: Prentice-Hall.
- Huluka, G. and R. Miller. (2014). Particle size determination by hydrometer method. In F. J. Sikora (Ed.), *Southern Cooperative Series Bulletin No. 419: Soil Test Methods From the Southeastern United States*, (pp.180-184). Southern Extension and Research Activity Information Exchange Group - 6
- Ibraheem, N.A., Hasan, M.M., Khan, R.Z., Mishra, P.K. (2012). Understanding color models: a review. *ARPJ. Sci. Technol.*, Vol2, No. 3 265-275.
- Jaja, N. (2016). Understanding the texture of your soil for agricultural productivity. Virginia Cooperative Extension. Publication No. CSES-162P.
- Kutner, M., Nachtsheim, C., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models (fifth edition)*. New York, NY: McGraw-Hill/Irwin
- Lund, E., Maxton, C. (2019). Comparing organic matter estimations using two farm implement mounted proximal sensing technologies. 5th global workshop on proximal soil sensing. PSS 2019 Proc., Columbia, Mo., May 28-31, 2019.
- McCutcheon, M. C., Farahani, H. J., Stednick, J. D., Buchleiter, G. W., & Green, T. R. (2006). Effect of soil water on apparent soil electrical conductivity and texture relationships in a dryland field. *Biosyst. Eng.*, 94(1), 19-32. doi:10.1016/j.biosystemseng.2006.01.002
- Miller, W.P., Miller, D.M. (1987) A micro-pipette method for soil mechanical analysis. *Commun. in Soil Sci and Plant Analysis*. 18(1), 1-15. Doi: 10.1080/00103628709367799
- Morais, P. A. de O., Souza, D. M. de, Carvalho, M. T. de M., Madari, B. E., & de Oliveira, A. E. (2019). Predicting soil texture using image analysis. *Microchemical J.*, 146, 455–463. <https://doi.org/10.1016/j.microc.2019.01.009>
- Nelson, D. and Sommers, L. (2015). Total carbon, organic carbon, and organic matter. in *methods of soil analysis*, A. Page (Ed.). doi:10.2134/agronmonogr9.2.2ed.c29

- Pricewaterhouse Coopers. (2016). Clarity From Above. Retrieved from <https://www.pwc.pl/pl/pdf/clarity-from-above-pwc.pdf>
- Reeves, D.W. (1997). The role of soil organic matter in maintaining soil quality in continuous cropping systems. *Soil Tillage and Research.*, 43, 131-167. [https://doi-org.libproxy.clemson.edu/10.1016/S0167-1987\(97\)00038-X](https://doi-org.libproxy.clemson.edu/10.1016/S0167-1987(97)00038-X)
- Soil Science Society of America. (2020). Glossary of Soil Science Terms. Soil organic matter. Retrieved from: <https://www.soils.org/publications/soils-glossary#>
- Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online. Accessed April 16, 2020.
- Sudduth, K.A. and Hummel, J.W. (1993). Soil organic matter, CEC, and moisture sensing with a portable NIR spectrophotometer. *Trans. ASABE*, 36, 1571–1582.
- Thien, S. J. (1979). A flow diagram for teaching texture-by-feel analysis. *Journal of Agronomic Education*, 8(1), 54–55. <https://doi.org/10.2134/jae.1979.0054>
- Ying, W. (2018). Drone-maker DJI to Develop More Industry Applications. *China Daily*. Retrieved from: <http://www.chinadaily.com.cn/a/201801/27/WS5a6bd252a3106e7dcc1371b0.html>
- Zhang, H. and Wang, J. (2014). Loss on ignition method. In F. J. Sikora (Ed.), *Southern Cooperative Series Bulletin No. 419: Soil Test Methods From the Southeastern United States*, (pp.155-157). Southern Extension and Research Activity Information Exchange Group - 6
- Zhang, X., Younan, N.H., O'Hara, C.G. (2005) Wavelet domain statistical hyperspectral soil texture classification. *IEEE Tran. Geoscience and Remote Sensing*, 43(3), 615-618. doi: 10.1109/TGRS.2004.841476

CHAPTER THREE

SOIL TEXTURE AND ORGANIC MATTER PREDICTION FROM INTERNET AVAILABLE SATELLITE IMAGERY

Introduction

Satellite imaging technologies have improved vastly from their origins in the late 1950's when the United States Government's CORONA program developed the first imaging satellite. For the first time, images were captured on film from space, and this film was then sent back into Earth's atmosphere, where it was collected in mid- air by recovery aircraft equipped with "claws" to snag the film pod (Ruffner, 1995). Today, over 300 earth observation satellites orbit the earth (Mohney, 2018), operated by governments and private entities alike. These satellites are capable of remotely sending data from space, resulting in a near constant feed of imagery and information. Similar to the first satellites, many today are equipped with cameras, providing access to near real time data. Meteorologists track storms, governments conduct surveillance, millions of citizens explore faraway places on platforms such as Google Earth, and researchers track many features, from climate change and the melting of the polar ice caps (NASA 2019a), to sand intrusion along major coastlines (NASA 2019b). Many of these satellites capture imagery in red, green, blue (RGB) format, meaning their resulting images closely resemble those taken by inexpensive, commercially available digital cameras or smartphones.

Alongside the use of unmanned aerial vehicles, utilization of satellite imagery in the agriculture has increased in recent years with internet availability of satellite imagery. These platforms allow for large scale data collection that is cost effective and reduces labor when compared to the traditional "boots on the ground" approach. Houborg and McCabe

(2016) utilized data from satellite images captured via Planet Labs' satellites and Landsat-8 satellites to develop a corrected equation for Normalized Difference Vegetation Index (NDVI). NDVI is commonly associated with plant vigor, or "greenness" and is commonly collected terrestrially using devices such as the Trimble Greenseeker (Trimble Inc., Sunnyvale, Cal.). Yang et al. (2006) correlated satellite imagery to traditional aircraft-captured imagery and found strong relationships between the two when evaluating grain sorghum yield, making yield predictions possible before a crop is harvested. Gholizadeh et al. (2018) utilized Sentinel-2 satellite imagery to, similar to this study, predict soil texture and organic carbon. Results from this study indicated better results when predicting soil organic carbon and clay when compared to results predicting silt and sand. Their study, however, utilized spectral bands outside of the range of the study presented in this document.

Collection, aggregation, and analysis of soil texture and nutrient data is a necessity in the field of precision agriculture, since soil texture variability is among the most consistent variables for defining crop yield potential. Soil texture is defined as the relative mass compositions of sand, silt, and clay particles found within a given soil sample, and is classified by the United States Department of Agriculture (Jaja, 2016). Soil texture and properties can influence physical and chemical factors such as drainage, water holding capacity, organic matter content, and cation exchange capacity, and nutrient retention, which can influence crop selection and productivity in an area. Typically, soil texture is determined in one of two ways: by use of the "hand-feel method" (Thien, 1979), or through

particle size analysis, using the hydrometer (Bouyoucos, 1962) or pipette method (Miller et al., 1987).

The “hand-feel method”, also known as the “ribbon test”, involves holding a wetted sample of soil in the hand, and drawing conclusions about its texture based upon how well the soil forms a ribbon between the fingers. This process allows for subjectivity in the determination of results; one person may consider a sample to be gritty, while others may consider it to be smooth. Furthermore, the ribboning process is dependent upon moisture levels, and different moisture contents can result in different conclusions. Farmers may also elect to collect soil electrical conductivity (EC) measurements across a field. Soil EC results are derived through the use of an implement such as a Veris Sensor Cart (Veris Technologies, Salina, KS.), which utilizes coulter-electrodes to measure electrical current as the sensor cart is pulled through a field behind a tractor or other vehicle. While soil EC is generally proportional to clay content and inversely proportional to sand content, these relationships can be affected by properties such as soil water content and temperature at the time of mapping (McCutcheon et al., 2006). Therefore, making accurate soil texture predictions from EC mapping is not practical. A well-defined and repeatable method for quantifying soil texture is referred to as the hydrometer method (Bouyoucos 1962). The hydrometer method is considered accurate because particle settling velocity and particle size are strongly correlated. In this method, dry, screened soil samples are mixed with a surfactant and shaken. Then a hydrometer is used at specific time intervals to measure apparent density of the soil-water suspension. As sand, silt, and clay particles fall out of

suspension at different time intervals, the percentage of sand, silt, and clay can be calculated.

In recent years, both Veris (Veris Technologies, Salina, Kans.) and Precision Planting (Precision Planting, Tremont Ill.) have introduced technologies to measure soil properties including organic matter in real time, eliminating the need for laboratory work. Both the Veris iScan and Precision Planting SmartFirmer utilize visible and near infrared (Vis-NIR) sensors to quantify soil reflectance, which can be then correlated to soil organic matter using the process outlined by Sudduth and Hummel (1993). The Vis-NIR module of the Veris iScan can be mounted to many implements including tillage tools and implement bars (Veris.com), and records data during the course of normal field operation. The Precision Planting SmartFirmer mounts to compatible planters behind the seed tube (Precisionplanting.com), and also records data as the planter is being operated. Lund and Maxton (2019), found that once calibrated, the iScan module produced a RMSE of 0.22% OM across all sample sites, and the SmartFirmer produced a RMSE of 0.24% OM across all sample sites studied. Both of these technologies are currently commercially available.

Similarly to soil texture classification, soil organic matter (OM) has been considered a key indicator of the productivity and yield potential of a soil. Soil organic matter is defined by the Soil Science Society of America (2020) as “the organic fraction of soil, including plant, animal, and microbial residues, fresh and at all stages of decomposition, and the relatively resistant soil humus”. Knowledge of OM content can greatly influence nutrient recommendations for a specific area. Reeves (1997) compiled and summarized a series of short and long term studies across multiple cropping systems

and soil management practices. He noted a decline of crop yield and plant available nitrogen over time as soil OM was depleted or decreased. This makes quantifying OM content important when considering the addition of soil fertilizer, or reintroducing nutrients into the soil by disking. Soil OM content is strongly correlated to the amount of organic carbon that is contained within a soil. Organic carbon mostly consists of the cells of microorganisms, decomposing plant and animal residues, humus synthesized from residues, and highly carbonized, elemental forms of carbon such as charcoal, graphite, and coal (Nelson and Sommers, 2015). The most widely used process for determining soil organic matter is the “Loss on Ignition” method, during which soil samples are dried in an oven to remove moisture, reweighed, and then ignited in a furnace to incinerate the organic compounds (Ball, 1964). The samples are then weighed again, and the pre- and post-ignition weight difference is used to calculate the amount of organic matter present. Both Ball (1964) and Nelson and Sommers (2015) concede that this method is not precise, nor perfectly quantifiable, as the process destroys both the organic and inorganic carbon, instead of solely the organic carbon used to measure soil organic matter.

Development of a series of regression models that can utilize Internet- available RGB satellite imagery to predict soil texture and organic matter content could aid in nutrient and management zone development (Basnyat et al., 2004). The objectives of this study were to (1) utilize and compare different free, internet available sources for RGB satellite imagery to predict sand, clay and soil organic matter content, and (2) to determine whether these predictions can be made using representative “sample circles” consisting of pixel data surrounding a physical sampling site.

Methods and Materials

Satellite imagery was collected using data from Planet Labs, Inc. (Planet Labs, Inc., San Francisco, Cal.), and Google Earth (Google, LLC, Menlo Park, Cal.). Satellite images from each source were analyzed to extract RGB pixel data from within each image in order to develop regression models. Images were then processed to keep only pixels that occurred within a designated diameter “sample circle” of the selected physical sample site. Two sample circle sizes of 25ft and 50ft were chosen in order to determine if larger or smaller sample circle sizes resulted in reduced model error. Models to predict sand, clay, and OM content were developed independently for each image source, while combined models using data from both sources were also developed.

Planet Labs Satellite Imagery

Images were collected using data obtained from Planet Labs’ “PlanetScope” satellites using the Planet Labs Earth Explorer (Planet.com). This group of approximately 130 satellites allows for daily image capture of a specific area at a resolution of 3 meters/pixel while orbiting at an altitude of approximately 475 km. Each satellite is equipped to capture red, green, blue and near infrared color bands. Planet Labs offers a variety of finished image types, each with varying layers of image processing. Images used in this study were of the “Basic Scene” product, meaning, among other steps, that a color curve was not applied to the imagery. For each field selected, ten separate images were captured using the Planet Labs “Explorer” tool in efforts to account for soil moisture variances due to irrigation or precipitation, as well as differences in capture quality between images. A sample image of variation between capture dates is illustrated in Figure 3.1. A

table of image capture dates for each field is given in Appendix B. Using the Planet Earth Explorer, images with over 15% cloud cover were eliminated from consideration.

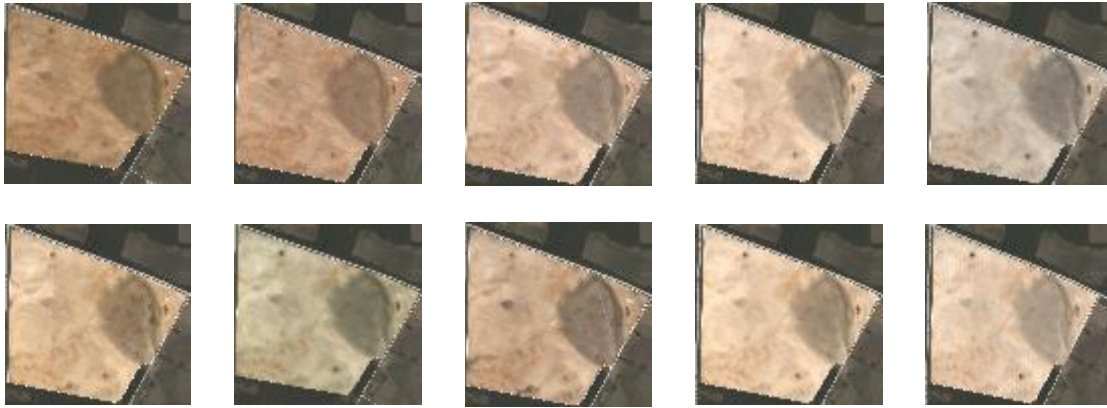


Figure 3.1: Ten captured images of a field using Planet Labs Data to capture average variation caused by time of day, soil moisture, and image sensor variability. The “Don Still” field is shown as an example.

Google Earth Satellite Imagery

Imagery displayed on the Google Earth platform is collected from a variety of sources including high altitude balloons, aircraft, and satellites. Satellite images featured on the platform are collected from sources including Maxar Technologies (Maxar Technologies, Westminster, Colo.), images from the Landsat collection of satellites, and the United States Department of Agriculture Farm Service Agency. Image sources for the current image being viewed can be seen near the bottom of the Google Earth image within the program. Due to this, extracting a single specification, or even list of specifications of the sensors used to capture images is improbable. It is also difficult to identify what, if any, color correction or cloud removal was done when images were provided to Google Earth. The “Historical Imagery” feature of Google Earth v. 7.3.3.7699 was used to locate fields

with previously collected soil texture data that were known to be unplanted and tilled at the time of image collection. From the narrow range of image dates available for viewing as well as the irregularity of image capture date and availability, only one image was captured of each field. Dates of satellite image capture are given in Appendix B.

Field Selection

Fields selected for this study were conventionally tilled and unplanted at the time of satellite imagery collection. Fields with bare soil provide an ideal image, as presence of standing crop or excessive residue of a previous crop would change pixel color values, affecting analysis and model development. Fields were selected based on the ability to visually observe color differences from captured satellite imagery. Soil texture and OM content prediction models were developed using two satellite imagery sources, Planet Labs and Google Earth. The Planet Labs dataset consists of five fields, and the Google Earth dataset consists of six fields. SSURGO soil data (Soil Survey Staff, 2019) were used for general characterization of the soils present, although these data were not used for development of the models presented in this study. Fields used for development of Planet Labs model were: Big Pivot (33.3343°N,81.0946°W) (Table 3.1), Don Still (33.3663°N,81.3399°W) (Table 3.2), “Market Front” (33.3782°N,81.2617°W) (Table 3.3), Watermelon Rd. (33.3169°N,81.0839°W) (Table 3.4), and E- (33.3446°N,81.3176°W) (Table 3.5). Big Pivot and Watermelon Rd. are located near Bamberg, S.C. Don Still, Market Back and E-7 are located near Blackville, S.C.

Table 3.1: Big Pivot Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area ha(ac)	% of Area
BaB	Barnwell loamy sand, 2 to 6 percent slopes	7.7 (19.1)	19.3%
GoA	Goldsboro loamy sand, 0 to 2 percent slopes	5.6 (13.9)	14.1%
McA	McColl loam, 0 to 2 percent slopes	11.0 (27.0)	27.3%
NaB	Nankin loamy sand, 2 to 6 percent slopes	0.9 (2.3)	2.3%
NbB2	Nankin sandy loam, 2 to 6 percent slopes	0.5 (1.3)	1.4%
NrA	Norfolk sand, 0 to 2 percent slopes	14.2 (35.0)	35.5%

Table 3.2: Don Still Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area ha (ac)	% of Area
DaA	Norfolk loamy sand, 0 to 2 percent slopes	3.5 (8.7)	4.6%
DaB	Barnwell loamy sand, 2 to 6 percent slopes	17.9 (44.3)	23.4%
FaC	Nankin loamy sand, 6 to 10 percent slopes	2.8 (7.0)	3.7%
FuA	Wagram sand, 0 to 2 percent slopes	0.9 (2.3)	1.2%
FuB	Wagram sand, 2 to 6 percent slopes	4.7 (11.7)	6.2%
Mc	McColl loam	15.1 (37.4)	19.7%
OrB	Barnwell loamy sand, 2 to 6 percent slopes	7.7 (19.1)	10.1%
VaA	Orangeburg loamy sand, 0 to 2 percent slopes	3.9 (9.6)	5.1%
VaB	Barnwell loamy sand, 2 to 6 percent slopes	15.8 (39.0)	20.6%
VaC	Barnwell loamy sand, 6 to 10 percent slopes	2.1 (5.1)	2.7%
VcD	Neeses soils, 10 to 25 percent slopes	2.1 (5.3)	2.8%

Table 3.3: Market Front Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area, ha (ac)	% of Area
DaB	Barnwell loamy sand, 2 to 6 percent slopes	1.7, (4.2)	5.4%
FuA	Wagram sand, 0 to 2 percent slopes	0.8 (2.1)	2.7%
FuB	Wagram sand, 2 to 6 percent slopes	9.9(24.5)	31.3%
Pu	Plummer loamy sand	0.2 (0.6)	0.8%
VaA	Orangeburg loamy sand, 0 to 2 percent slopes	2.7 (6.6)	8.4%
VaB	Barnwell loamy sand, 2 to 6 percent slopes	16.3 (40.2)	51.4%

Table 3.4: Watermelon Rd. Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area, ha(ac)	Percent of Area
BaB	Barnwell loamy sand, 2 to 6 percent slopes	21.6, (53.4)	69.7%
BoB	Bonneau sand, 2 to 6 percent slopes	0.8, (2.1)	2.8%
CoA	Coxville fine sandy loam, 0 to 2 percent slopes	3.3, (8.2)	10.7%
NoA	Boboco fine sand, 0 to 2 percent slopes	1.1, (2.6)	3.4%
NrA	Norfolk sand, 0 to 2 percent slopes	1.7, (4.2)	5.4%
RaA	Rains fine sandy loam, 0 to 2 percent slopes	2.5, (6.1)	7.9%

Table 3.5: E-7 Field SSURGO Soil Type Distribution

Symb ol	Map Unit Name	Area, ha(ac)	% of Area
DaB	Barnwell loamy sand, 2 to 6 percent slopes	2.22 (5.5)	29.5%
FuC	Ailey sand, 6 to 10 percent slopes	0.16 (0.4)	2.3%
OrB	Barnwell loamy sand, 2 to 6 percent slopes	5.14 (12.70)	68.2%

Fields used in development of Google Earth models included: Rusty Pivot located in Lee County, S.C. (34.1345°N, 80.2533°W) (Table 3.6), CP4 & CP13 (32.9780°N, 81.2811°W) (Table 3.7) located in Allendale, S.C., Chicken House (33.3343°N, 81.3605°W) (Table 3.8), B6B (33.3576°N, 81.3288°W) (Table 3.9), C12 (33.3484°N, 81.3193°W) (Table 3.10), and E7 (33.3448°N, 81.3158°W) (Table 3.5), all of which are located in Barnwell County, S.C. E7 soil data was used for construction of models from both sources, but imagery used in analysis was source- specific.

Table 3.6: Rusty Pivot Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area, ha (ac)	% of Area
CxA	Coxville sandy loam, 0 to 2 percent slopes	17.8 (44.6)	35.2%
GoA	Goldsboro sandy loam, 0 to 2 percent slopes	4.1 (10.4)	8.0%
NnA	Noboco- Goldsboro complex 0 to 2 percent slopes	9.4 (23.4)	18.4%
NoA	Norfolk loamy sand, 0 to 2 percent slopes	10.7 (26.5)	20.9%
RaA	Rains sandy loam, 0 to 2 percent slopes	9.0 (22.3)	17.6%

Table 3.7: CP4 & 13 Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area, ha(ac)	% of Area
BaB	Blanton sand, 0 to 6 percent slopes	6.7 (16.6)	9.6%
BoA	Bonneau fine sand, 0 to 2 percent slopes	.08 (0.2)	0.1%
GoA	Goldsboro sandy loam, 0 to 2 percent slopes	.08 (0.2)	0.1%
NoA	Norfolk loamy sand, 0 to percent slopes	47.3 (117)	67.5%
Pe	Pelham loamy sand, 0 to 2 percent slopes	4.6 (11.4)	6.6%
Ra	Rainy loamy fine sand	11.4 (28.1)	16.2%

Table 3.8: ChickenHouse Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area, ha(ac)	%of Area
BaB	Blanton sand, 0 to 6 percent slopes	0.1 (0.3)	3.0%
DaB	Barnwell loamy sand, 2 to 6 percent slopes	2.1 (5.2)	58.1%
FuB	Wagram sand, 2 to 6 percent slopes	1.4 (3.5)	38.9%

Table 3.9: B6B Field SSURGO Soil Type Distribution

Symbol	Map Unit Name	Area, ha(ac)	% of Area
DaB	Barnwell loamy sand, 2 to 6 percent slopes	0.7 (1.8)	7.0%
FaB	Nankin loamy sand, 2 to 6 percent slopes	0.6 (1.6)	6.3%
FuB	Wagram sand, 2 to 6 percent slopes	2.4 (5.9)	23.2%
VaA	Orangeburg loamy sand, 0 to 2 percent slopes	2.0 (5.0)	19.7%
VaB	Barnwell loamy sand, 2 to 6 percent slopes	4.5 (11.1)	43.8%

Table 3.10: C12 Field SSURGO Soil Type Distribution

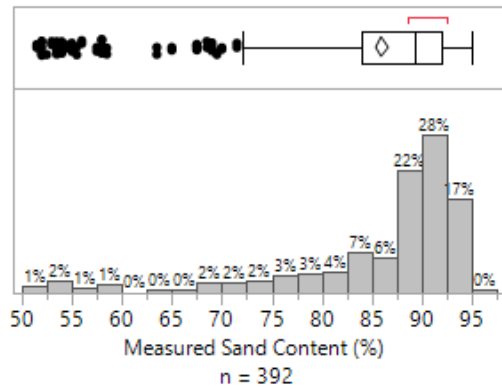
Symbol	Map Unit Name	Area, ha(ac)	% of Area
DaA	Norfolk loamy sand, 0 to 2 percent slopes	0.32 (0.8)	6.9%
DaB	Barnwell loamy sand, 2 to 6 percent slopes	1.33 (3.3)	28.5%
FuA	Wagram sand, 0 to 2 percent slopes	2.14 (5.3)	45.4%
FuB	Wagram sand, 2 to 6 percent slopes	0.04 (0.1)	1.2%
VaB	Barnwell loamy sand, 2 to 6 percent slopes	0.85 (2.1)	18.0%

Soil Texture and OM Content Data Collection

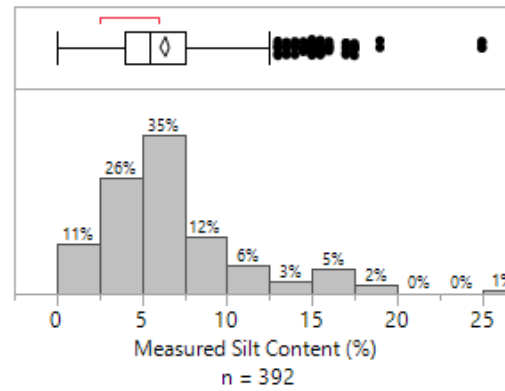
Fields selected for this study were assigned 0.101 ha (.25 ac) grid squares using Trimble Ag Desktop Software (v2019.1.0, Trimble Inc., Sunnyvale, Cal.). The GPS coordinates of the grid centers were exported to a comma separated values (CSV) file. The file was then loaded into Soil Sampling Utility (v.1.0.1.10, Clemson University Precision Agriculture, Blackville, SC), and navigation to sample positions was conducted by using a BU-353S4 USB GPS Receiver (GlobalSat WorldCom, New Taipei City, Taiwan) with WAAS, DGPS correction. At each sample site, eight soil cores were collected from a 305 cm (120 in.) radius around the center position of each grid square. Each sample core was collected from the top 15 cm (4 in.) of the soil profile using a soil probe with diameter 2.54 cm (1 in.)

Each sample was passed through a #10 (2 mm) sieve to remove grained sized particles and residue and divided into two equal subsamples, each subsample weighing at least 100 g. One subsample was processed to determine the percentage sand, silt, and clay using the Hydrometer Method as outlined by Huluka and Miller (2010). The other sample was processed to determine OM content using the Loss on Ignition process outlined by Zhang and Wang (2014). Soil texture and OM were reported as mass composition of sand, silt, and clay. Figure 3.3 parts (a), (b), (c), and (d) shows distributions of measured sand,

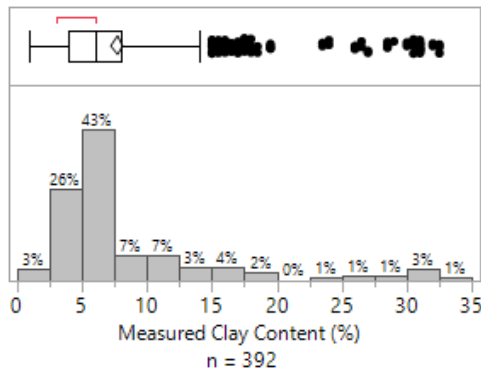
silt, clay and OM content compositions respectively across all sample sites, as these ranges represent the known sand, clay, and OM values which will be represented during regression modeling.



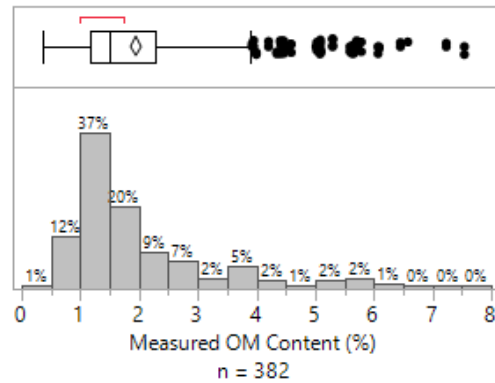
(a)



(b)

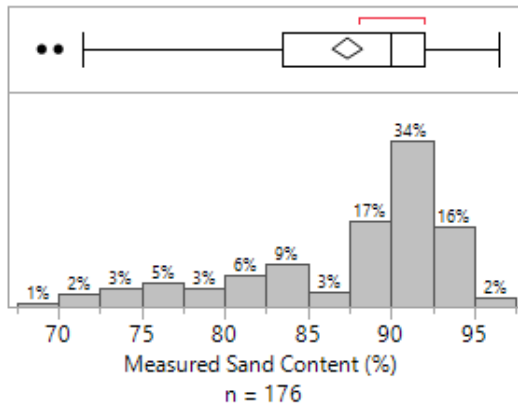


(c)

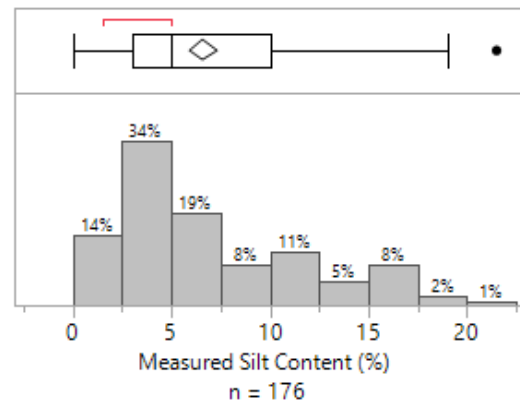


(d)

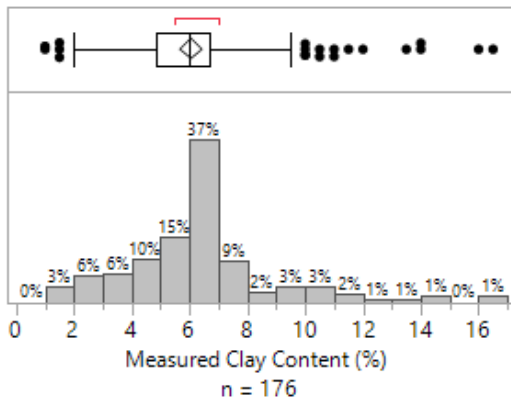
Figure 3.2: Distribution of measured sand (a), silt (b), clay (c) and OM content (d) across all Planet Labs image sample sites used in regression modeling. Sand content ranged from 51.5% - 95.0%, silt content ranged from 0%-25%, clay content ranged from 1% - 32.5%, and OM content ranged from 0.36%-7.53%.



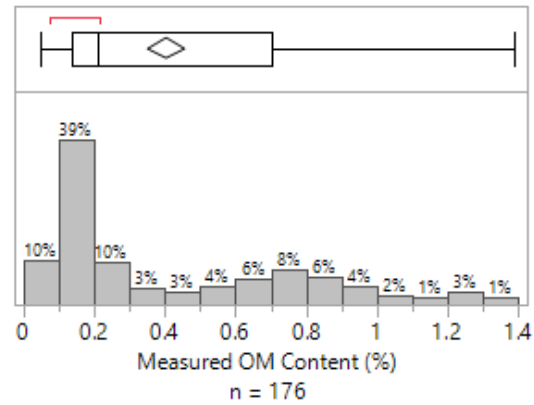
(a)



(b)



(c)



(d)

Figure 3.3: Distribution of measured sand (a), silt (b), clay (c) and OM content (d) across all Google Earth image sample sites used in regression modeling. Sand content ranged from 69.0% - 96.5%, silt content ranged from 0%-21.5%, clay content ranged from 1% - 16.5%, and OM content ranged from 0.05%-1.39%.

Image Processing

Images of each field were loaded into Spatial Image Digitizer v2.0 (SID), software developed by Clemson University, for georeferencing and pixel value extraction. SID works to first assign each pixel within an image a set of coordinates in latitude, longitude

format, a process known as georeferencing. Images can be georeferenced using SID in one of two methods. The first method involves selecting permanent ground control points such as power poles, roadway centerlines, or other land features that are visible both in SID's basemap image, and in the image to be analyzed. The second method involves loading an ArcView shapefile (Environmental Systems Research Institute, Redlands, Cal.) polygon definition containing the field's outer boundary into SID, and also overlaying the same field boundary onto the image to be processed. Using this method, distinct features of a field's boundary such as sharp corners or points can be selected for use as georeferencing points in both images. A sample image of this method is illustrated in Figure 3.4. SID uses these user-specified ground control points to build models for calculation of latitude and longitude as a function of x and y pixel positions.

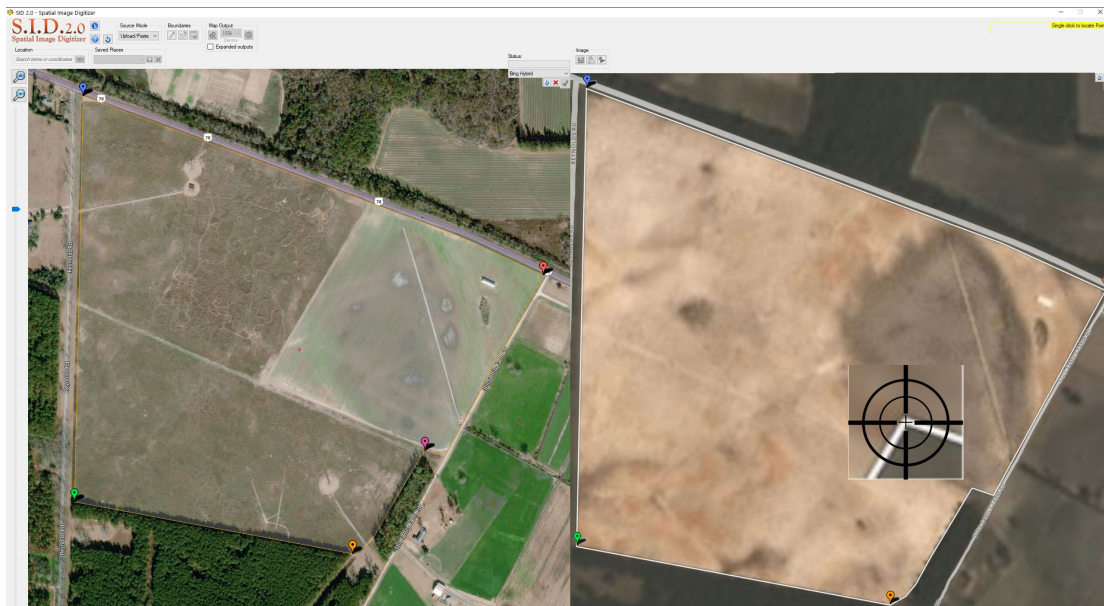


Figure 3.4: Georeferencing process within SID. Image on left represents basemap image with field boundary in ArcView Shapefile overlaid. Image to right shows the same field, but image captured from satellite imagery. User selects one point on field boundary of left image, then selects the same point on right image.

After georeferencing, each image was processed in SID at full resolution to extract red, green, and blue values, as well as other color values calculated from the red, green, and blue components; examples of these derived values include hue, chroma, and brightness. SID outputs are stored in a comma separated values (CSV) file, with each row representing a given pixel within the image, and each column representing color components extracted. The data in these columns forms the foundation for model effects used in regression modeling in later steps.

Pixel data extraction extracts pixel data for the entire shapefile boundary, in this case, the entire field. The first objective of this study was to evaluate different size “sample circles” for prediction. These sample circles consist only of georeferenced pixel data falling within a given diameter of the physical soil sampling site, with each sample circle serving as a representative sample of the pixel data surrounding each physical sample location. Two sample circle diameters were used in this study: 7.6 m (25 ft) represented the “smaller” sample circle, while 15 m (50 ft) represented the “larger” sample circle. These sample circle diameters were chosen to evaluate whether including more or less pixel data resulted in reduced model error. Using Circular Polygon Generator (CPG) and Point Polygon Merge Utility (PPMU), software developed by Clemson University, SID outputs were processed twice independently, once to remove all pixel data outside of a 15 m (50 ft) diameter around the datapoint, and once to remove all data outside of a 7.6 m (25 ft) diameter. CPG is a tool that constructs a shapefile of circular polygons centered on positions included in a point dataset. PPMU is a tool that associates each point in a point dataset with the polygon in which it resides, from a polygon dataset. For this study, the point dataset source was the

SID output and the polygon dataset source was the CPG output. The resulting output datasets from PPMU only contained pixel information for pixels within the diameters specified above; meaning models were developed using only those data. An example of this workflow is illustrated in Figure 3.5.

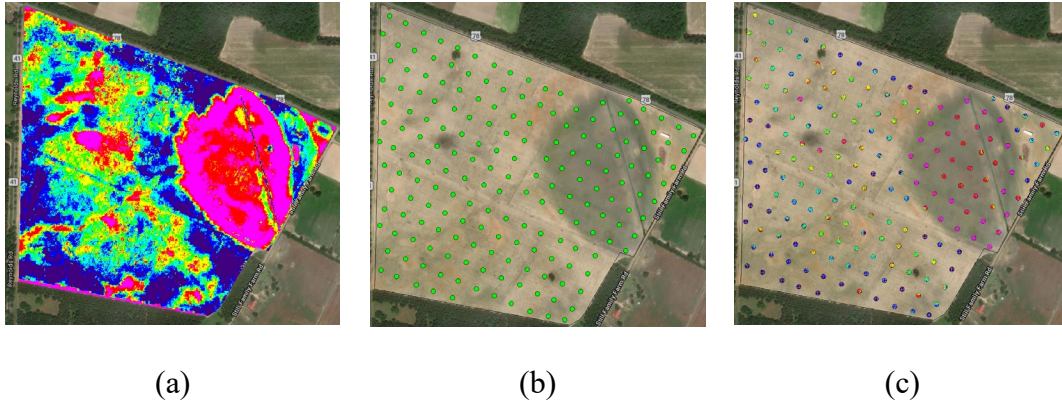


Figure 3.5: Example workflow of reducing SID point dataset (a) to within “sample circles”. Using CPG, “sample circles” of specified diameters were created (b, polygon dataset). Using PPMU, only data within these sample circles is retained for regression model development (figure c, point dataset, clipped to polygons in figure b).

For Planet Labs Imagery, this process was repeated for each of the ten images captured of each field. Each row in the CSV file was associated with a unique SampleID, which defined the point at which the physical soil sample was taken. Pixel values within each sample circle were then averaged, resulting in a single value for each SID output column, within each SampleID. These values were then averaged once again, this time across all ten images captured. The resulting dataset consisted of a single value for each column, for each data point. As an example, if SampleID “1” contained a red colorspace value of 100, the 100 value was derived from averaging all red colorspace values from pixels residing within the sample circle of SampleID “1” across all ten images captured.

Regression Model Development

All regression modeling and statistical analyses were performed using JMP Pro v.14.1.0 (SAS Institute Inc., Cary, North Carolina). Three sets of models were independently created for prediction of sand, clay, and OM contents: one set of models for images from each of the satellite image sources (Planet Labs and Google Earth) and a third set of models using combined image data from both sources. Within each image source, models were also independently created using data from 25- and 50-ft diameter sample circles. For each set of models, each sample site was randomly assigned to one of two classifications: a training class containing 80% of sample sites and a testing class containing 20% of samples sites. This classification was performed to ensure models were not tested on the same datapoints which were used to develop them. Using the stepwise model fitting personality, multiple linear regression models were created using both forward and backward direction and both minimum AICc and minimum BIC stopping rules. For each model, the term being predicted (e.g. sand, clay, or OM content) was assigned as the response variable, y, while the model effect, x, terms were assigned as being the image color data extracted by SID. Transformations of all SID outputs were also considered as model effects, including square root, square, cube root, cube, log+1, and reciprocal. Indices of the red, green, and blue component values were also created using combinations of addition, subtraction, multiplication, and division of combinations of the color components. Some examples include: $(R+G+B)$, $(R+B)*(R-B)$, $(R*G*B)$, and $(R+G+B)/3$. The “3-D Function Finder” feature of ZunZun.com (James Phillips, Birmingham, Ala.) was utilized to create equations consisting of X, Y and Z variables. In

these equations, X and Y were represented by two color component combinations, and Z was represented by the known value of what was being predicted. A sample equation from ZunZun may predict sand from only red and blue. Multiple collinearity was reduced by removing any term with a variance inflation factor (VIF) of greater than 5, as suggested by Kutner, et al. (2005). Regression outliers, or heavily influential points, were removed from consideration using Cook's Distance; any data point with Cook's Distance values greater than 1.0 were excluded as suggested by Hair, et al. (1998). Upon exclusion of a regression outlier, the stepwise model iteration was restarted. Terms with low significance (p-value > 0.05) were eliminated until all remaining terms satisfied VIF, Cook's Distance, and p-value criteria.

Error Reporting

Model prediction error was calculated using the testing class for each component modeled and is presented herein in units of percent sand content, percent clay content, and percent OM content. Prediction error was performed only on the data assigned to the testing class, and was calculated using the generalized formula in Equation 3.1:

$$AbsErr = |Predicted - Actual|, \quad (3.1)$$

where *AbsErr* represents absolute prediction error of the modeled component (e.g. percent sand, clay, or OM content), *Predicted* represents the predicted value of the modeled component for a given image, and *Actual* represents the measured value of the component for the sample collected at that image. In the following tables, the columns demonstrating error at 50% and 90% confidence represent the prediction errors for which 50% and 90%

of all prediction errors were lower in value. For example, an “error at 90% confidence” value of 13.77 as seen in Table 3.12 demonstrates that 90% of the absolute values of prediction errors were less than 13.77% sand content. A means comparison (student’s t-test $\alpha = 0.05$) was performed on each dataset to establish significant differences in prediction ability as a related to combinations of sample circle diameter.

Results and Discussion

Regression models were developed to predict sand, clay, and OM content as a function of data extracted from different RGB satellite imagery sources. Models were developed independently using data from within both 7.6m (25ft) and 15m (50ft) sample circles, the center of which represents the location at which the physical soil samples were collected. These sample circles aim to establish representative sample areas surrounding the point of soil data collection. Two sizes of sample circles were chosen and evaluated to determine whether sample circle size influenced prediction model accuracy, by way of including more pixel data within the larger circle.

While a complete listing of model coefficients and terms is given in Appendix B, an example model is given in Table 3.11 to demonstrate each aspect of the resulting model. Models may be implemented using the generalized formula provided in Equation 3.2 utilizing coefficients and terms from Table 3.11.

$$y_0 = \text{Intercept} + (\text{Term}_0 * \text{Estimate}_0) + (\text{Term}_1 * \text{Estimate}_1) \dots + (\text{Term}_n * \text{Estimate}_n) \quad (3.2)$$

where y_0 represents the variable being predicted (e.g. sand content), *Intercept* represents the Estimate value for the Intercept term, Term_0 represents the first term below Intercept in

the Term column, and $Estimate_0$ represents the value in the Estimate column corresponding to the row containing $Term_0$.

Table 3.11: An example to illustrate general model construction: model terms and coefficients for model predicting sand content using 50-ft diameter sample circle and Planet labs data. Each estimate value is multiplied by the corresponding value of the given term in the dataset.

Term	Estimate	Std Error
Intercept	-5989	654.2
$(R+B)/(G-B)^3$	0.0024	0.0008
$\sqrt[3]{((R+G+B)/(R+B))}$	5346	574.3
$\text{Log}(\text{MeanSatHSL})$	8.822	1.076
$(G-B)/(R-G)^3$	-38.11	3.198

Sand Content Prediction Models

For sand, clay, and organic matter content prediction, models were created in three “banks”, with each soil property representing an independent bank. Within each bank exist models created using Planet Labs, Google Earth, and Combined image source data. For each of these image sources, models were created for both 25ft and 50ft sample circle diameters, as well as a model using data from both sample circle sizes. Models were developed and evaluated independently, but different combinations of sample size diameter were evaluated within each bank.

Tables 3.12, 3.13, and 3.14 illustrate results from application of sand content prediction models developed using Planet Labs, Google Earth, and combined form data from both sources, respectively. None of the three image sources demonstrated significant difference in prediction ability in regard to sample circle diameter size. In some instances, models were able to better predict using data from a sample circle size different than the one with which it was trained. Using an example taken from the dataset, in Table 3.13, the

model trained on the smaller sample circle had both lower mean error and error at 90% confidence when tested on the larger sample circle than when tested on the testing class of the smaller circle. Within each “bank” of models for each source, error numbers remained numerically similar, with Google Earth images Table 3.13 exhibiting the lowest Error at 90% confidence. It should be noted that there is not a direct comparison between errors, since data from only one field was included in modeling for both image sources, and the number of images that were collected is less. Figure 3.6, 3.7, and 3.8 illustrate actual by predicted plots for application of 25-ft sample circle data to the 25-ft sample circle testing class for Planet Labs, Google Earth, and combined image source models. 1:1 lines are also included in these figures.

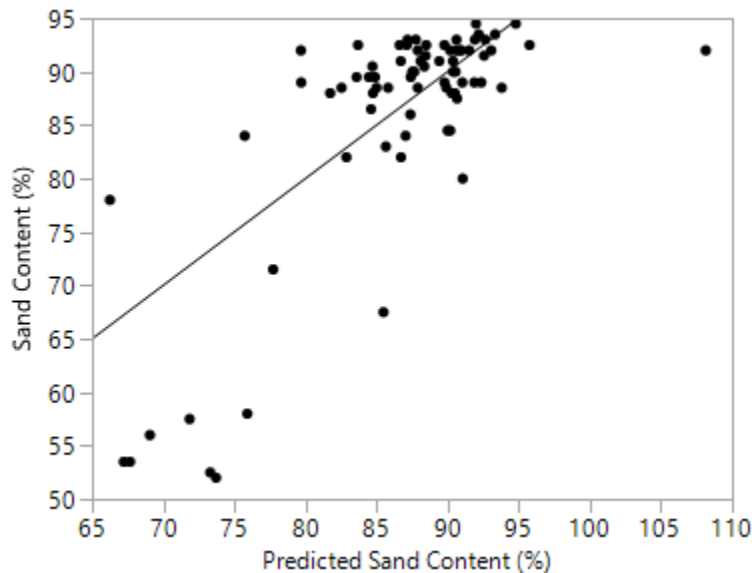


Figure 3.6: Actual by predicted data for sand prediction model developed with Planet Labs data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.

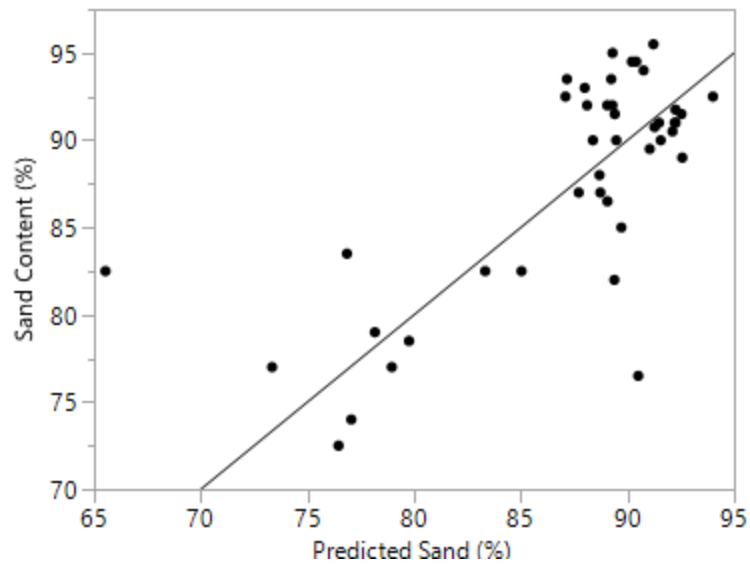


Figure 3.7: Actual by predicted data for sand prediction model developed with Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.

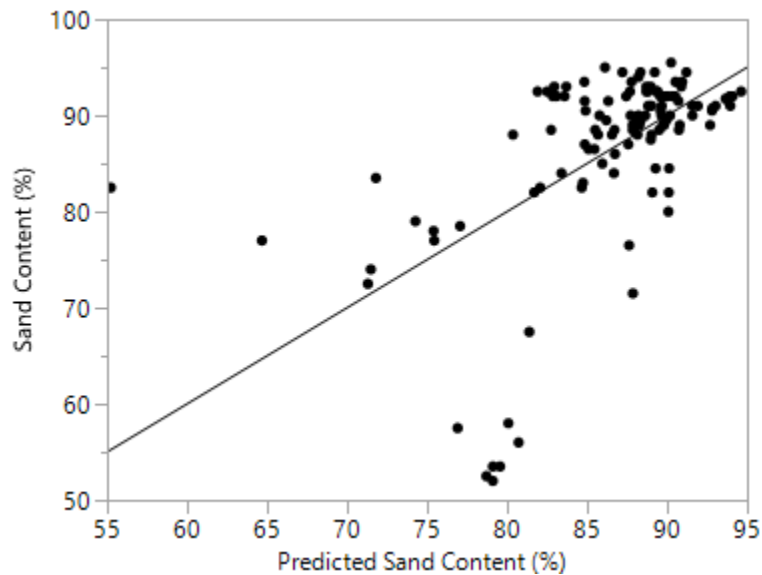


Figure 1.8: Actual by predicted data for sand prediction model developed with Planet Labs and Google Earth (combined) data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.

Table 3.12: Prediction errors of sand content prediction models utilizing Planet Labs imagery, as developed and applied to various subsets of the data.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
25ft Samples	25ft Samples	78	5.04	A	3.11	13.77
25ft Samples	50ft Samples	78	4.98	A	3.04	15.51
25ft Samples	All Samples	156	5.01	A	3.07	13.85
50ft Samples	50ft Samples	78	5.18	A	3.49	14.16
50ft Samples	25ft Samples	77	5.12	A	3.87	11.50
50ft Samples	All Samples	155	5.15	A	3.66	11.69
All Samples	25ft Samples	78	5.02	A	3.14	12.93
All Samples	50ft Samples	78	4.98	A	2.99	14.75
All Samples	All Samples	156	4.00	A	3.03	12.98

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

Table 3.13 Prediction errors of sand content prediction models utilizing Google Earth imagery, as developed and applied to various subsets of the data.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
25ft Samples	25ft Samples	44	3.32	A	2.54	6.49
25ft Samples	50ft Samples	44	3.14	A	2.12	6.45
25ft Samples	All Samples	88	3.23	A	2.32	6.37
50ft Samples	50ft Samples	44	3.39	A	2.68	6.80
50ft Samples	25ft Samples	44	3.54	A	3.06	7.98
50ft Samples	All Samples	88	3.46	A	2.73	6.99
All Samples	25ft Samples	44	3.28	A	2.90	7.36
All Samples	50ft Samples	44	3.16	A	2.36	6.30
All Samples	All Samples	88	3.22	A	2.55	6.85

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

Table 3.14: Prediction errors of sand content prediction models utilizing both Planet Labs and Google Earth imagery, as developed and applied to various subsets of the data.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
25ft Samples	25ft Samples	122	4.99	A	2.31	11.53
25ft Samples	50ft Samples	122	4.95	A	2.44	10.61
25ft Samples	All Samples	244	4.97	A	2.35	10.88
50ft Samples	50ft Samples	122	4.57	A	2.08	10.62
50ft Samples	25ft Samples	122	4.59	A	2.15	10.77
50ft Samples	All Samples	244	4.85	A	2.10	10.55
All Samples	25ft Samples	122	5.06	A	2.86	10.94
All Samples	50ft Samples	122	5.57	A	2.08	10.53
All Samples	All Samples	244	5.04	A	2.83	10.58

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

Clay Content Prediction Models

Clay content prediction model results are illustrated for Planet Labs, Google Earth, and combined source data in Tables 3.15, 3.16, and 3.17 respectively. Significant differences existed only when applying the larger sample circle model to the testing class of the smaller circle, which caused significantly higher error percentages. Similar to sand content prediction, Google Earth models produced the lowest mean and error at 90% confidence interval percentages (Table 3.16). By applying the combined sample circle model to the larger sample circle data, the lowest error in this bank of models was achieved. The clay prediction model bank using both Google and Planet data did not produce any significant differences.

Figures 3.9, 3.10, and 3.11 illustrate actual by predicted plots for application of 25-ft sample circle data to the 25-ft sample circle testing class for Planet Labs, Google Earth, and combined image source models. 1:1 lines are also included in these figures.

Table 3.15: Prediction errors of clay content prediction models utilizing Planet Labs imagery, as developed and applied to various subsets of the data.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
25ft Samples	25ft Samples	78	4.00	A	2.33	10.69
25ft Samples	50ft Samples	78	3.64	A	2.13	9.67
25ft Samples	All Samples	156	3.82	A	2.19	9.87
50ft Samples	50ft Samples	78	3.29	A	2.48	9.55
50ft Samples	25ft Samples	77	8.00	B	6.06	19.55
50ft Samples	All Samples	155	3.22	A	2.43	8.75
All Samples	25ft Samples	77	3.72	A	2.71	10.15
All Samples	50ft Samples	78	3.75	A	2.54	10.20
All Samples	All Samples	155	3.73	A	2.58	10.08

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

Table 3.16: Prediction errors of clay content prediction models utilizing Google Earth imagery, as developed and applied to various subsets of the data.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
25ft Samples	25ft Samples	35	2.58	AB	1.72	5.48
25ft Samples	50ft Samples	34	2.56	AB	1.84	5.46
25ft Samples	All Samples	69	2.57	A	1.78	4.71
50ft Samples	50ft Samples	44	1.98	ABC	1.40	4.45
50ft Samples	25ft Samples	44	2.05	ABC	1.31	4.82
50ft Samples	All Samples	88	2.01	ABC	1.34	4.49
All Samples	25ft Samples	44	1.72	AB	1.47	3.70
All Samples	50ft Samples	44	1.69	BC	1.41	3.34
All Samples	All Samples	88	1.71	C	1.42	3.56

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

Table 3.17: Prediction errors of clay content prediction models utilizing both Planet Labs and Google Earth imagery, as developed and applied to various subsets of the data.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
25ft Samples	25ft Samples	122	3.65	A	2.30	7.50
25ft Samples	50ft Samples	122	3.61	A	2.28	7.65
25ft Samples	All Samples	244	3.63	A	2.29	7.40
50ft Samples	50ft Samples	122	3.59	A	2.08	6.82
50ft Samples	25ft Samples	122	3.62	A	2.03	7.54
50ft Samples	All Samples	244	3.61	A	2.07	7.02
All Samples	25ft Samples	122	3.51	A	2.06	6.49
All Samples	50ft Samples	122	3.49	A	2.07	6.54
All Samples	All Samples	244	3.50	A	2.07	6.29

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

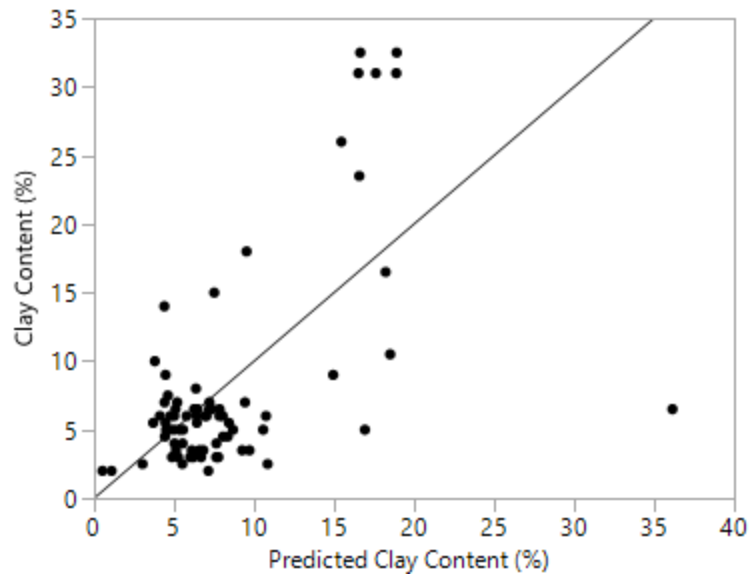


Figure 3.9: Actual by predicted data for clay content prediction model developed with Planet Labs data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.

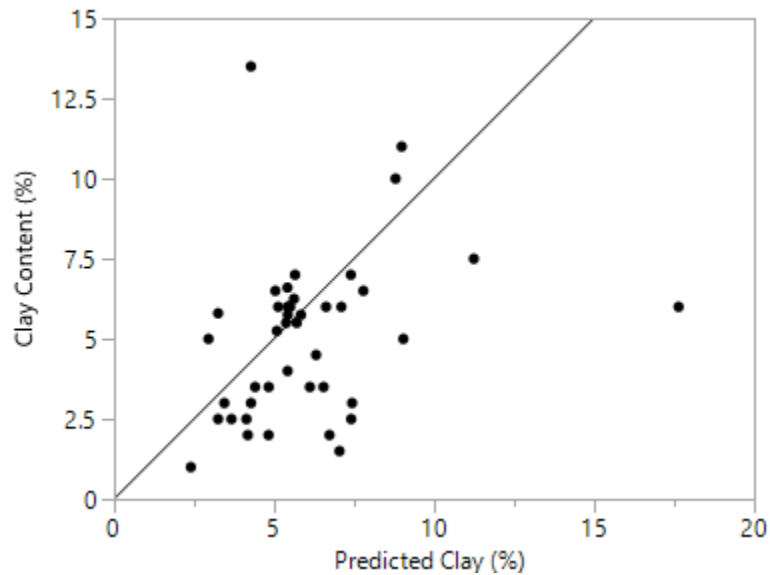


Figure 3.10: Actual by predicted data for clay content prediction model developed with Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.

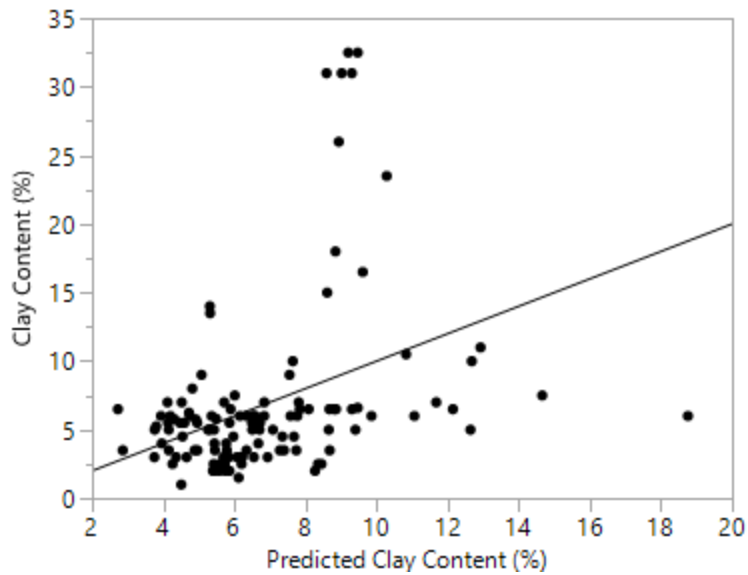


Figure 3.11: Actual by predicted data for clay content prediction model developed with Planet Labs and Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.

OM Content Prediction Models

OM content prediction model results are illustrated for Planet Labs, Google Earth, and combined source data in Tables 3.18, 3.19, and 3.20 respectively. Application of the combined sample circle model to the larger sample circle dataset resulted in significantly lower error values for the Planet Labs bank of models (Table 3.18). OM content prediction models using Google Earth data in Table 3.19 produced significantly different error values, however, the two combinations with the lowest error share the same T-Test value, signifying that although differences exist within the model bank, the two lowest error combinations are not significantly different. The combined source OM content prediction model bank (Table 3.20) did not result in significant differences except when applying the smaller sample circle diameter to any other data that it was trained on, which resulted in extremely high error in both cases.

It is important to note the range of OM content values listed at the beginning of this document, as initially error percentages of sub- 1% may appear as though the model does an excellent job predicting, when in fact the range of OM content used in model training is only ~7%, as illustrated previously in Figure 3.2(d). Figure 3.12, 3.13, and 3.14 illustrate actual by predicted plots for application of 25-ft sample circle data to the 25-ft sample circle testing class for Planet Labs, Google Earth, and combined image source models. 1:1 lines are also included in these figures.

Table 3.18: Prediction errors of OM content prediction models utilizing Planet Labs imagery, as developed and applied to various subsets of the data.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
25ft Samples	25ft Samples	78	0.65	A	0.47	1.60
25ft Samples	50ft Samples	78	0.49	ABC	0.38	0.91
25ft Samples	All Samples	156	0.57	ABC	0.42	1.33
50ft Samples	50ft Samples	78	0.47	BC	0.35	1.01
50ft Samples	25ft Samples	78	0.62	AB	0.44	1.36
50ft Samples	All Samples	156	0.54	ABC	0.41	1.07
All Samples	25ft Samples	78	0.63	AB	0.47	1.47
All Samples	50ft Samples	78	0.45	C	0.33	0.99
All Samples	All Samples	156	0.54	ABC	0.38	1.13

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

Table 3.19: Prediction errors of OM content prediction models utilizing Google Earth imagery, as developed and applied to various subsets of the data.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
25ft Samples	25ft Samples	35	0.24	D	0.12	0.77
25ft Samples	50ft Samples	35	0.80	A	0.52	2.22
25ft Samples	All Samples	70	0.52	BC	0.28	1.44
50ft Samples	50ft Samples	35	0.34	CD	0.32	0.72
50ft Samples	25ft Samples	35	0.83	A	0.54	2.12
50ft Samples	All Samples	70	0.58	B	0.39	1.63
All Samples	25ft Samples	35	0.35	CD	0.28	0.78
All Samples	50ft Samples	35	0.33	CD	0.32	0.59
All Samples	All Samples	70	0.34	D	0.31	0.65

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

Table 3.20: Prediction errors of OM content prediction models utilizing both Planet Labs and Google Earth imagery, as developed and applied to various subsets of the data.

Training Basis ^[a]	Testing Basis ^[b]	n ^[c]	Mean Error ^[d]	T-Test ^[e]	Error at 50% Confidence ^[d]	Error at 90% Confidence ^[d]
25ft Samples	25ft Samples	113	0.71	C	0.44	1.60
25ft Samples	50ft Samples	113	44.39	A	0.56	31.72
25ft Samples	All Samples	226	22.55	B	0.51	4.51
50ft Samples	50ft Samples	113	0.55	C	0.38	1.14
50ft Samples	25ft Samples	113	0.96	C	0.77	2.36
50ft Samples	All Samples	226	0.76	C	0.56	1.99
All Samples	25ft Samples	112	0.81	C	0.54	2.28
All Samples	50ft Samples	109	0.57	C	0.40	1.13
All Samples	All Samples	221	0.69	C	0.46	1.71

^[a] Imagery dataset whose training class was used for model development

^[b] Imagery dataset whose testing class was used for prediction error analysis

^[c] Number of images used in testing class

^[d] Unit of measure for prediction error presented as % sand content

^[e] Means with the same letters are not significantly different ($\alpha = 0.05$)

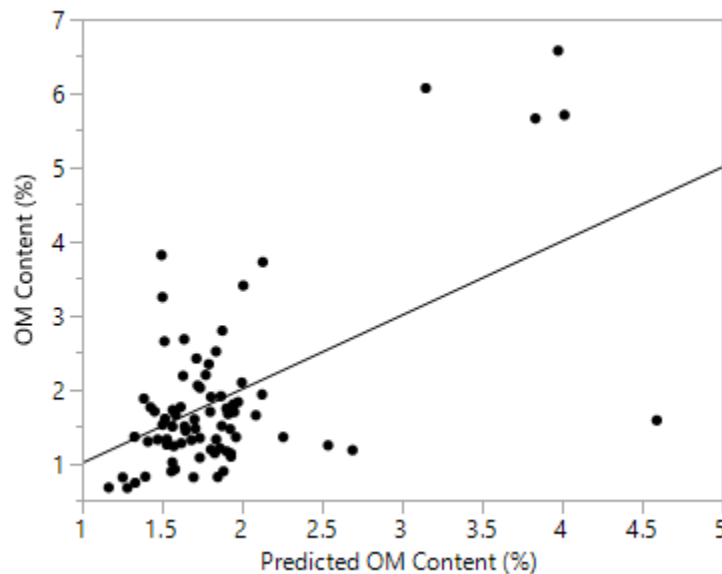


Figure 3.12: Actual by predicted data for OM content prediction model developed with Planet Labs data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.

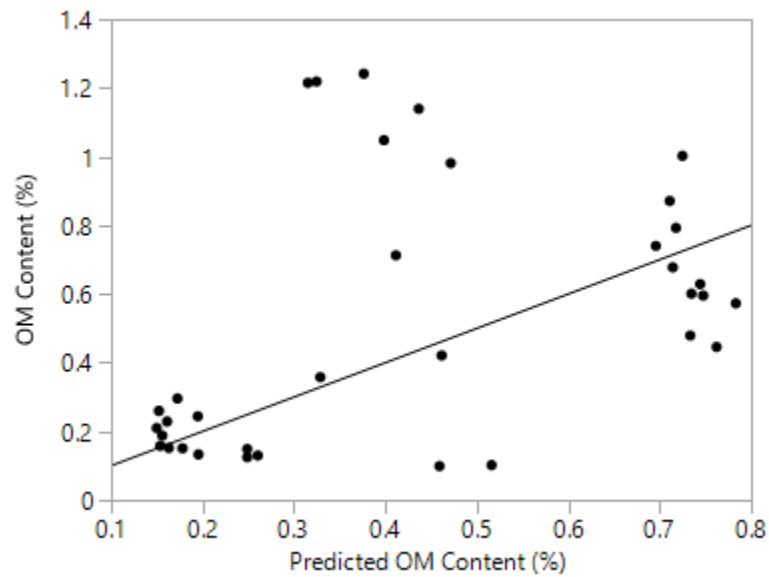


Figure 3.13: Actual by predicted data for clay content prediction model developed with Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.

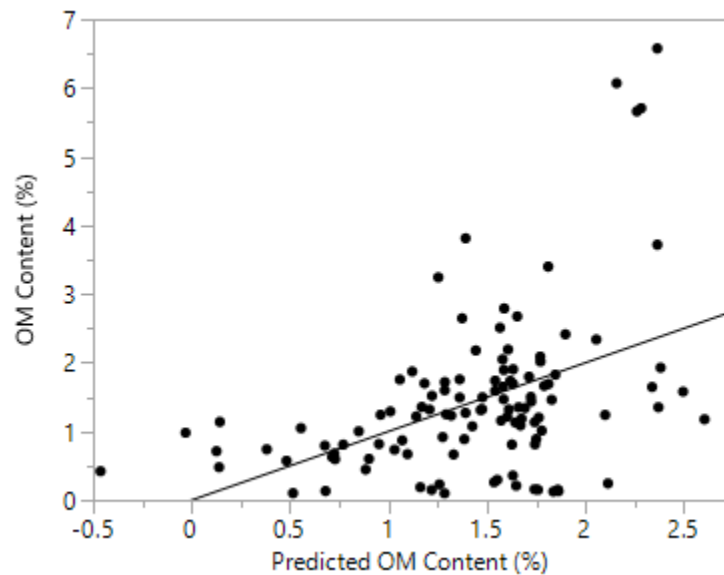


Figure 3.14: Actual by predicted data for clay content prediction model developed with Planet Labs and Google Earth data using 25-ft sample circles applied to testing data of 25-ft sample circle data. 1:1 line provided in figure.

Anomalous Texture Predictions

Regression models are only able to accurately predict values that fall within the range of their training data and cannot be considered reliable when predicting outside of this range or extrapolating. In this study, the result is an anomalous texture prediction. Such predictions often occur in two ways: predictions that are inaccurate, yet within an acceptable range, and predictions that are outside of possible ranges; such as percentages below 0 or greater than 100. To rectify these errors, model input and output boundaries must be applied. These boundaries, defining acceptable ranges of each model term (as exhibited in the training dataset), allow for determination of whether the image being analyzed is within the range of parameters used for model development. This should generally suggest if an image of unknown soil texture would result in a plausible result. If the boundaries determine an anomalous image or group of pixels, as compared to the training dataset, the image would be omitted from prediction. Model bounds are provided for the model resulting in lowest mean error for sand, clay, and organic matter content predictions from each image source in Appendix A. The tolerance factor allowed for some extrapolation outside of the range of values demonstrated in the training dataset.

The first type of erroneous result is an inaccurate, yet realistic prediction. These predictions fall within an acceptable range of data but may vary greatly from points surrounding them. These features are often a result of features captured in the satellite image, such as crop residue, trees, or water falling within the sample circle. Despite these anomalies, the model attempts to make a prediction, and the results are within a normal, otherwise acceptable range (0 to 100%).

Another type of inaccurate prediction occurs when the model produces unrealistic results, or results outside of acceptable values, such as sand, clay, or OM contents outside of the range of 0 to 100%. These predictions may be more likely to result from inconsistent image characteristics, such as shadows, overexposure, or lens flare. These results are typically outside of the possible range of a texture content, and errors may be extremely large in either the positive or negative direction. Model bounding, as discussed earlier, should generally prevent such prediction errors.

With either type of erroneous result, it is necessary to introduce boundaries to the models, which will automatically determine if an image can be accurately analyzed, or whether it should be omitted. Images producing predictions either above 100% or below 0% can automatically be omitted, as they are out of bounds of a realistic prediction. This can be performed by use of an “if...then” statement, calling for images outside of those ranges to be marked, or flagged. Alternatively, these predictions within some range could be defaulted to the closer of 0% or 100%. Boundaries for model terms in this study were constructed based on the ranges observed in the training datasets; a binary output was calculated for each term for each processed image to determine whether the term was within the generally observed range of values in the training dataset. As discussed, a tolerance factor was applied to allow for some extrapolation. Terms for an image falling within bounds were assigned a BoundCheck value of 1, while terms outside of bounds were assigned a BoundCheck value of 0. The product of the BoundCheck values for each term of a given image, was then used to demonstrate whether all of the terms for an image were within the model’s boundaries. Products equal to 1 demonstrated that all terms were within

model bounds; products equal to 0 demonstrated that the value for at least one term was out of model bounds.

Image Source Model Comparison

A means comparison (student's t-test, $\alpha = 0.05$) was performed to establish, across all datapoints, which image source resulted in the lowest error. When predicting both sand (Table 3.21) and clay (Table 3.22), Google Earth resulted in the lowest mean error values. This could be a result of the increased image resolution or could be simply caused by the soil colors of the fields chosen for this study. When predicting OM content (Table 3.23) significant difference existed when using the combined model, this is likely a result of the extremely high error previously mentioned. Although Google Earth produces the lowest error values, this should not be interpreted as a direct comparison between sources due to the differing fields, image capture dates, and number of images captured for each source.

Table 3.21: Sand content prediction errors across imagery sources. Google Earth resulted in lowest mean error, significant differences between Google Earth and both Planet Labs & combined models.

Image Source ^[a]	n ^[b]	Mean Error ^[c]	T-Test ^[d]
Google	528	3.31	B
Planet	924	5.05	A
Combined	1464	4.83	A

^[a] Image source used for model construction

^[b] Number of images used in testing class

^[c] Mean error across all testing classes within an image source

^[d] Means with the same letters are not significantly different ($\alpha = 0.05$)

Table 3.22: Clay content prediction errors across imagery sources. Significant differences exist between all three model sources, with Google Earth resulting in lowest mean error.

Image Source ^[a]	n ^[b]	Mean Error ^[c]	T-Test ^[d]
Google	490	2.06	A
Planet	936	4.39	B
Combined	1464	3.58	C

^[a] Image source used for model construction

^[b] Number of images used in testing class

^[c] Mean error across all testing classes within an image source

^[d] Means with the same letters are not significantly different ($\alpha = 0.05$)

Table 3.23: OM content prediction errors across imagery sources. Significant differences exist only between Combined model and both Google Earth and Planet Labs sources. Google Earth models exhibit lowest mean error, while Combined model mean error is likely a result of previously tabulated errors resulting from combined model combinations.

Image Source ^[a]	n ^[b]	Mean Error ^[c]	T-Test ^[d]
Google	490	0.48	A
Planet	936	0.55	A
Combined	1464	8.06	B

^[a] Image source used for model construction

^[b] Number of images used in testing class

^[c] Mean error across all testing classes within an image source

^[d] Means with the same letters are not significantly different ($\alpha = 0.05$)

Effects of Environmental Conditions and Physical Soil Properties on Prediction

During evaluation of results, it became apparent that soil properties, such as soil moisture, may impact image analysis results relative to texture prediction. Figure 3.6 shows Field E7, in an image that was not used in regression model development. In this Google Earth satellite captured image, a center pivot irrigation system is operating in the field, resulting in one half of the field appearing wet. This image was analyzed using previously

described methods. Then pixel data of a point within the “wet” part of the field was compared to pixel data taken from the “dry” zone in the field from the same satellite image. These two points are within 1% known sand content, but their R, G, and B values vary substantially, resulting in a 10% difference in sand content prediction. Table 3.24 shows the results of analysis of these two points. Utilizing a Google Earth based model, it is apparent that these differing color values result in over a 10% difference in predicted % sand content. For a model to be considered robust, conditions such as these would need to be more completely included within the training datasets. However, an image such as the one shown in Figure 3.15 should not be selected for classifying relative soil differences within a field, such as for zone development.

Table 3.24: Analysis of “wet” and “dry” points in Field E7. Wet soil substantially alters pixel color characteristics among points with 1% variation in known sand content.

SampleID	Moisture Condition	Known Sand %	Predicted Sand %	R	G	B
E7A-7	Wet (Irrigated)	93.0%	78.533%	135.228	89.73	65.78
E7B1-3	Dry (Non-Irrigated)	94.0%	89.98%	198.56	167.67	136.625



Figure 3.15. Wet and dry points within Field E7. The two points are within 1% known sand content, but exhibit substantially different R, G, and B values, highlighting the importance physical factors, especially soil moisture, can have on regression modeling.

Application of Sand Content Prediction Model in Zone Management Scenario

Models developed have practical applications, one of these being in development of field management zones. Field management zones in agricultural production are often designed to group like soils together, seeking to maximize differences in soil type and/or yield potential between the zones and minimize differences within zones. To demonstrate this application, points with known sand content from Chicken House Field were compared to sand content prediction developed from the smaller sample circle area, which produced both the lowest mean error and lowest error percentage at 90% confidence. Three contoured zones of equal area were created, representing relative sand content percentages: Low, Medium, and High. The model shown in Appendix C was applied to bare soil images collected from Chicken House Field using equation 3.2. The contoured zone map developed for actual sand (Figure 3.16a) can be compared to that for predicted sand content in Figure 3.16b. Contour maps were produced using Trimble Ag. Desktop Software v.2020.01. When viewing the actual and predicted data, it is apparent that the predicted values do not align universally with the actual data.

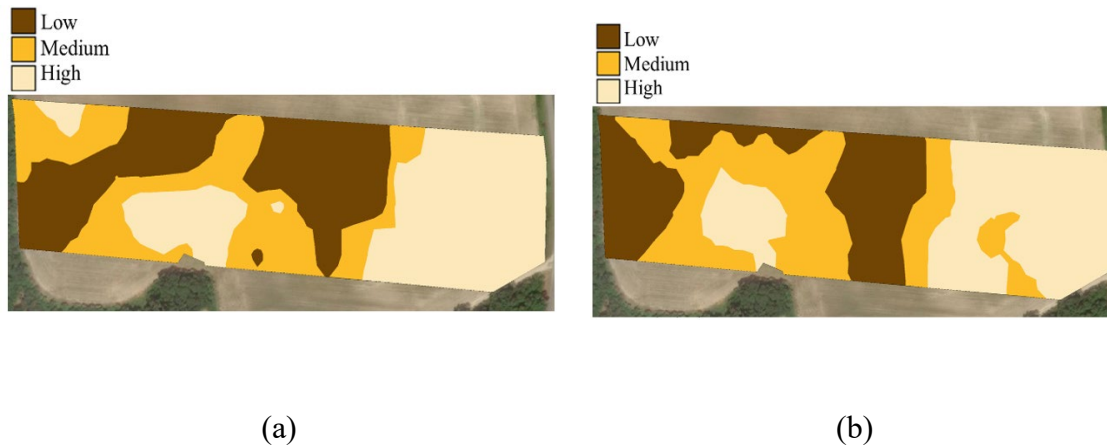


Figure 3.16: Illustration of zone creation utilizing sand content prediction models. Image (a) shows actual sand content of contoured data, while (b) shows results of applying prediction model.

The zone creation process was performed on two additional fields not included in the original study as an example of model robustness when exposed to varying field types. Images were captured using Google Earth and processed using SID, then sand content was predicted using the Google Earth equation developed using all sample circle data, which resulted in lowest mean error when applied to testing data. As these fields were not included in the original study, actual soil texture is not known. Additionally, sampling circles were not utilized for these examples; soil texture was calculated for each and every pixel within field boundaries

The first example field is located in Screven County, G.A. (32.6019°N, 81.5723°W). Google Earth image capture date is 3/28/2013. This field consists of mostly Fuquay loamy sand (USDA Soil Survey Staff), but includes areas of loamy soil which results in a visibly darker soil color (Figure 3.17a). When the prediction model was applied, these differences resulted in lower sand content predictions, and can be visualized in Figure

3.17b. Without applying model bounds, no image points were classified below 0% sand or above 100% sand. A contour map was then created of this data to simulate a management zone application (Figure 3.17c).

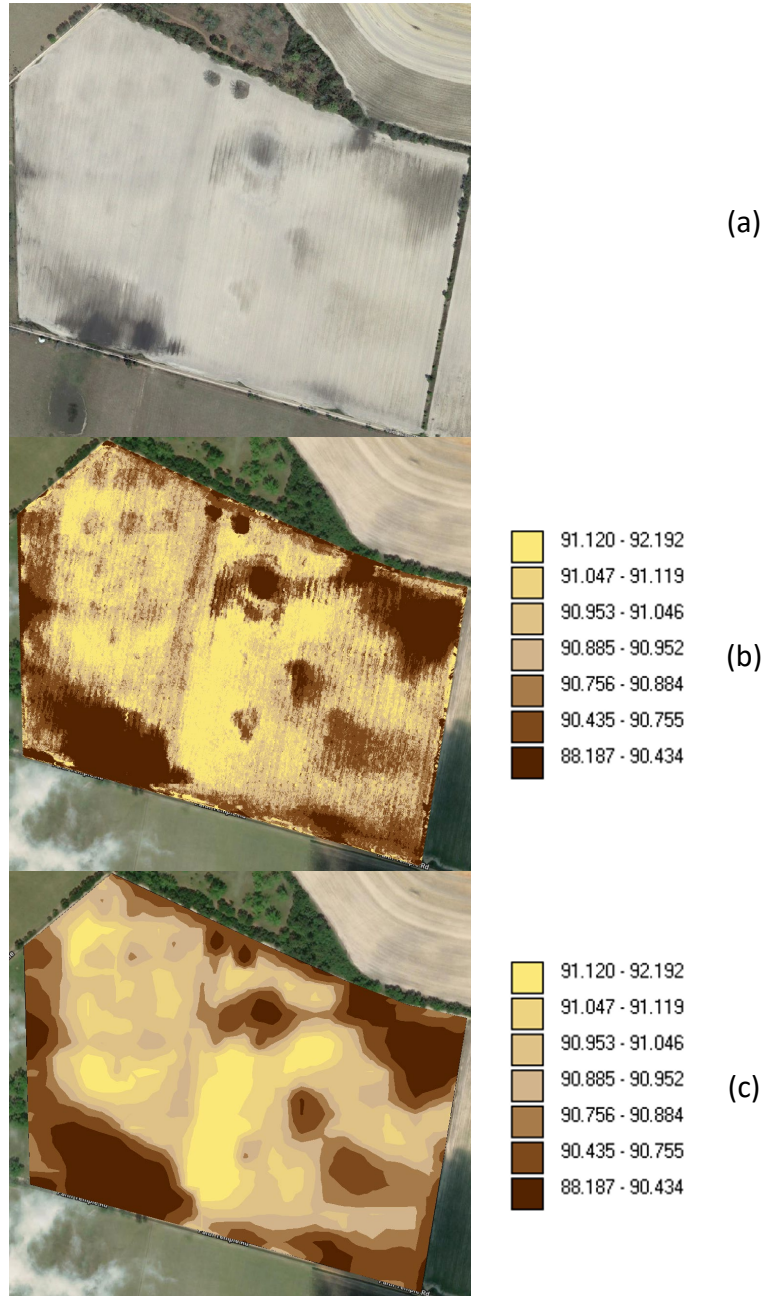


Figure 3.17: Application of sand content prediction model to field not included in testing dataset located in Sylvania, Georgia. Image (a) illustrates bare soil condition, (b) illustrates results of sand

prediction model, and (c) illustrates sand content data as would be utilized for zone delineation, with legend showing predicted percent sand content.

This demonstration was performed again on a field in Boiling Springs, N.C. (35.1874°N, 81.7130°W). Google Earth image capture date was 4/20/2018. This field consists mostly of Cecil sandy clay loam, and visually appears to have more “red clay” in its coloration, and less variation in its bare soil image (Figure 3.18a) than the Sylvania field shown previously. The sand prediction model echoes these visual observations, reporting a lower sand content in these areas. The prediction model also accurately depicts a washout in the field, which would likely have an increased sand content, as illustrated in Figure 3.18b. When contoured (Figure 3.18c), although the model reports a comparatively low range of sand contents, the resulting contour map still depicts relative differences and could be used in a zone delineation application.

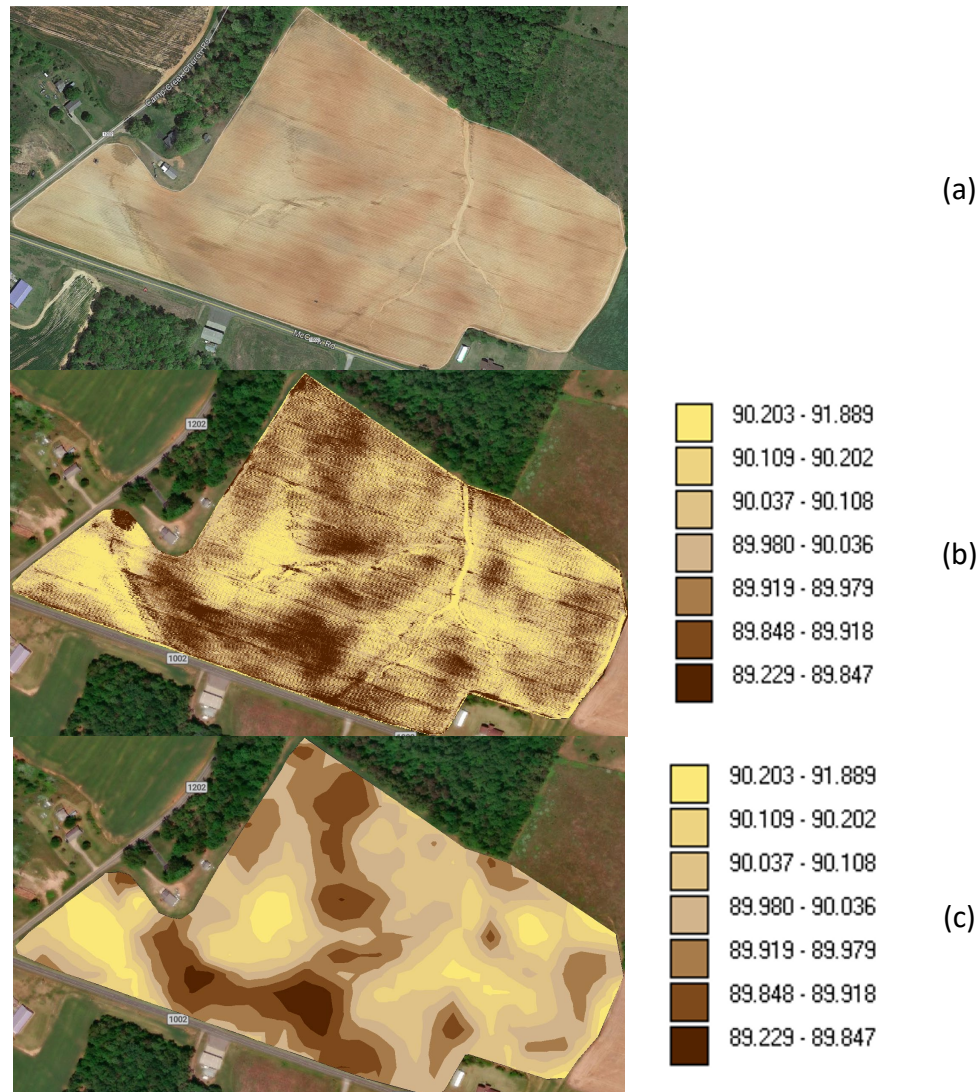


Figure 3.18: Application of sand content prediction model to field not included in testing dataset located in Boiling Springs, North Carolina. Image (a) illustrates bare soil condition, (b) illustrates results of sand prediction model, and (c) illustrates sand content data as would be utilized for zone delineation, with legend showing predicted percent sand content.

Conclusion

The objective of this study was to, using multiple linear regression models, develop equations to predict sand, clay, and organic matter content percentages from internet available, RGB satellite imagery. Two sources of imagery were utilized, Google Earth and

Planet Labs. Separate regression models for predicting sand, clay, and organic matter content percentages were developed independently for each image source.

Because of its mass availability and ease of integration, there is promise in the prospect of utilizing Google Earth satellite imagery for soil texture prediction. Although not directly compared to Planet Labs data, Google Earth image-based models exhibited the lowest prediction error in almost all predictions. The main constraint of using Google Earth satellite imagery is the irregularity and frequency at which updated imagery is provided. As an example, some fields that were considered for this study did not have a bare soil image provided by Google Earth. Best-performing models developed using Google Earth imagery resulted in prediction errors, at a 90% confidence interval, of 6.30% sand, 3.34% clay, and 0.59% organic matter content.

Planet Labs imagery has an edge in that imagery is available almost daily, at the cost of- lower spatial resolution. Image resolution was not evaluated in this study, and as a result, no conclusions can be drawn from it here. Similar to Google Earth, Planet Labs has an application programming interface (API) which allows imagery data to be easily implemented into standalone applications, meaning Planet Labs remains a possible source for future work on this subject. Best-performing models developed using Planet Labs satellite imagery resulted in prediction errors, at a 90% confidence interval, of 11.50% sand, 8.75% clay, and 0.91% organic matter content.

The concept of soil texture prediction from satellite imagery is not at a stage where it will replace conventional soil sampling for soil texture and organic mapping. Factors such as soil moisture and ground cover can have drastic impacts on the resulting image

colors, making an accurate prediction difficult in some cases. The importance of this study lies less in the ability of models to predict exact percentages, and more in the ability of them to recognize relative differences within a field, such as for spatial delineation of management zones.

Future Work

There are many areas in which this study can be improved, from the images used to the range of data collected. This study only utilized data from two image sources; Planet Labs and Google Earth, while there are many other sources of satellite imagery available, such as data from the Landsat series of satellites. Inclusion of more image sources could improve the ability of a single model to accurately predict texture percentages across multiple image sources by taking into account the variations of each satellite type. Collection and logging of data such as soil moisture or other physical properties at the time of image collection may aid in model development. Additionally, inclusion of fields with wider texture content ranges may reduce model error, and would expand applicability of models developed. This study developed prediction models using data within 7.6 m (25 ft) and 15 m (50 ft) diameter circles. Experimentation with other diameter sizes may reveal an “optimal” diameter in which error is reduced. Although ten images were captured of each field utilizing Planet Labs data, the pixel values for all ten images were averaged for each sample site. Utilizing each dataset independently would likely increase variation in pixel values, as the method used in this study may have “averaged out” much of the naturally occurring variation. The concept of “binning” data has shown promise in studies using the same regression model process, but with imagery collected from a UAV. Binning involves

breaking each color component, such as the red, green, and blue, components into groups, with each group representing a fraction of the total range of the component. Binning aids in model development by highlighting “sections” of colors which may be beneficial to model error reduction. During development of Planet Labs data-based models, pixel values were averaged across all ten images captured of a specific field. Treating each image as its own dataset may reduce model error, while at the same time providing more datapoints for model development.

References

- Ball, D.F. (1964). Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soils. *J. Soil Soc.*, 15,84-92. doi:10.1111/j.1365-2389.1964.tb00247.x
- Basnyat, P., McConkey, B., Meinert, B., Gatkze, C., and Noble, G. (2004). Agriculture Field Characterization Using Aerial Photograph and Satellite Imagery. *IEEE Geoscience and Remote Sensing Letters*, 1(1), 7–10. <https://doi.org/10.1109/lgrs.2003.822313>
- Bouyoucos, G. J. (1962). Hydrometer method improved for making particle size analyses of soils. *Agron. J.*, 54(5), 464–465. doi:10.2134/agronj1962.00021962005400050028x
- Gholizadeh, A., Žižala, D., Saberioon, M., & Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment*, 218, 89–103. <https://doi.org/10.1016/j.rse.2018.09.015>
- Hair, J., Anderson, R., Tatham, R. and Black W. (1998). *Multivariate Data Analysis (fifth edition)*. Englewood Cliffs, NJ: Prentice-Hall.
- Houborg, R., and McCabe, M. (2016). High-Resolution NDVI from Planet's Constellation of Earth Observing Nano-Satellites: A New Data Source for Precision Agriculture. *Remote Sensing*, 8(9), 768. <https://doi.org/10.3390/rs8090768>
- Huluka, G. and R. Miller. (2014). Particle size determination by hydrometer method. In F. J. Sikora (Ed.), *Southern Cooperative Series Bulletin No. 419: Soil Test Methods From the Southeastern United States*, (pp.180-184). Southern Extension and Research Activity Information Exchange Group – 6
- Jaja, N. (2016). Understanding the texture of your soil for agricultural productivity. Virginia Cooperative Extension. Publication No. CSES-162P.
- Kutner, M., Nachtsheim, C., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models (fifth edition)*. New York, NY: McGraw-Hill/Irwin
- Lund, E., Maxton, C. (2019). Comparing organic matter estimations using two farm implement mounted proximal sensing technologies. 5th global workshop on proximal soil sensing. PSS 2019 Proc., Columbia, Mo., May 28-31, 2019.
- McCutcheon, M. C., Farahani, H. J., Stednick, J. D., Buchleiter, G. W., & Green, T. R. (2006). Effect of soil water on apparent soil electrical conductivity and texture relationships in a dryland field. *Biosyst. Eng.*, 94(1), 19-32. doi:10.1016/j.biosystemseng.2006.01.002
- Miller, W.P., Miller, D.M. (1987) A micro-pipette method for soil mechanical analysis. *Commun. in Soil Sci and Plant Analysis*. 18(1), 1-15. Doi: 10.1080/00103628709367799
- Mohney, D. (2018, July 28). Retrieved from <https://www.spaceitbridge.com/a-sky-full-of-commercial-eyes-up-to-1300-earth-observation-satellites-in-next-5-years.htm>

- National Aeronautics and Space Administration (NASA) Goddard Space Flight Center. (2019a). *Annual Arctic Sea Ice Minimum 1979-2019 with area graph*. [Video]. Climate.nasa.gov. https://climate.nasa.gov/climate_resources/155/video-annual-arctic-sea-ice-minimum-1979-2019-with-area-graph/
- National Aeronautics and Space Administration (NASA) Earth Observatory. (2019b, July 15). World of Change: Coastline Change. Retrieved from <https://earthobservatory.nasa.gov/world-of-change/CapeCod>
- Nelson, D. and Sommers, L. (2015). Total carbon, organic carbon, and organic matter. in methods of soil analysis, A. Page (Ed.). doi:10.2134/agronmonogr9.2.2ed.c29
- Reeves, D.W. (1997). The role of soil organic matter in maintaining soil quality in continuous cropping systems. *Soil Tillage and Research.*, 43, 131-167. [https://doi-org.libproxy.clemson.edu/10.1016/S0167-1987\(97\)00038-X](https://doi-org.libproxy.clemson.edu/10.1016/S0167-1987(97)00038-X)
- Ruffner, K. C. (1995). *Corona: Americas first satellite program* (United States., Central Intelligence Agency.). Washington, D.C.: History Staff, Center for the Study of Intelligence, Central Intelligence Agency.
- Soil Science Society of America. (2020). Glossary of Soil Science Terms. Soil organic matter. Retrieved from: <https://www.soils.org/publications/soils-glossary#>
- Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online. Accessed April 16, 2020.
- Sudduth, K.A. and Hummel, J.W. (1993). Soil organic matter, CEC, and moisture sensing with a portable NIR spectrophotometer. *Trans. ASABE*, 36, 1571–1582.
- Thien, S. J. (1979). A flow diagram for teaching texture-by-feel analysis. *Journal of Agronomic Education*, 8(1), 54–55. <https://doi.org/10.2134/jae.1979.0054>
- Yang, C., Everitt, J. H., and Bradford, J. M. (2006). Comparison of QuickBird Satellite Imagery and Airborne Imagery for Mapping Grain Sorghum Yield Patterns. *Precision Agriculture*, 7(1), 33–44. <https://doi.org/10.1007/s11119-005-6788-0>
- Zhang, H. and Wang, J. (2014). Loss on ignition method. In F. J. Sikora (Ed.), *Southern Cooperative Series Bulletin No. 419: Soil Test Methods From the Southeastern United States*, (pp.155-157). Southern Extension and Research Activity Information Exchange Group – 6

CHAPTER FOUR

AERIAL IMAGERY POINT SAMPLING METHOD AND SURVEY OF UAV AND IMAGE ANALYSIS USING MACHINE LEARNING IN AGRICULTURE

Introduction

The agriculture industry is actively evaluating methods to use Unmanned Aerial Vehicles (UAV), commonly referred to as “drones” for specific applications to increase efficiency. The Federal Aviation Administration reports that 1,563,263 UAVs have been registered, with 441,709 of those units being registered for commercial use (FAA 2020). The majority of UAVs are equipped with imaging sensors that capture imagery in either red green blue (RGB) format similar to digital cameras or using multispectral image sensors which capture more color bands than can be captured by RGB sensors. Many consumer level UAVs cost less than \$1,500, therefore, for a relatively low investment they provide a way for imagery data to be collected over large areas in a relatively short period of time, for a relatively low investment when compared to traditional plane-mounted aerial imagery.

Additionally, UAVs can be preprogrammed with flight patterns, allowing them to consistently capture the same area in an image when the flight is repeated across different days, growing seasons, or conditions. Since flight plans can be programmed to any legal altitude, UAVs are effective devices for capturing high resolution imagery in a repeatable way. Collected imagery can be used to spot visual differences or can be further processed for image analysis. Data extracted from UAVs imagery can be tied to data collected from

equipment during the growing season to help develop improved management decisions and increase profit.

In the agricultural sector, UAVs are quickly becoming commonplace tools used by researchers, farmers, and crop consultants for scouting and data collection due to their ease of use, low investment, and efficiency. An economic report published by the Association for Unmanned Vehicle Systems International (AUVSI) forecasts that, during the period 2015-2025, UAVs will have a \$75.6 billion impact on the agriculture industry alone, the highest of any impacts in the forecast (AUVSI 2013). While many use their UAV for spotting visual differences within fields, this technology allows for many different applications. Combining data collected from UAVs with data collected from the myriad of sensors available in the precision agriculture market, UAVs are becoming an essential tool in zone creation, management, and crop monitoring.

The objective of this paper is to introduce a “point sampling” method of aerial data collection, and to compare it to the traditional method of whole field, orthophoto-based mapping. Additionally, a review of currently existing links between the fields of machine learning, artificial intelligence (AI) and the use of UAVs as applied to the agricultural sector is included.

UAV Hardware and Software

Among the most frequently studied topics involving the use of UAV in agriculture is yield estimation. Dodge (2019) was able to predict cotton yield using a consumer grade UAV and developed yield estimation equations using 2-dimensional UAV obtained

orthophoto data. Bendig et al., (2014) utilized high resolution imagery captured from a UAV equipped with an RGB camera to estimate barley biomass, and found strong relationships between biomass, plant height, and crop surface models. These crop surface models were developed using a UAV equipped with RGB camera. Reza et al., (2019) utilized a UAV to estimate rice yield using an RGB camera. Several studies utilize UAVs from SZ DJI Technology Co., Ltd. (Shenzhen, China), commonly known as DJI. DJI holds the largest market share of UAV sales at a reported 72% (Lampert 2019), reporting \$2.83 billion in revenue in 2017 (Ying, 2018). Many elect to purchase UAV from DJI because of their low cost, ease of use, and expandability options.

The Phantom platform of DJI UAV is equipped with a GPS module, and can be factory equipped with a real time kinematics (RTK) module allowing sub-inch accuracy which can be used for precise and repeatable navigation to study areas, as well as a multispectral camera sensor, allowing for data collection of parameters beyond what RGB lenses are capable of collecting. Sentera, Inc (Minneapolis, Minn.) has produced a retrofit kit for equipping UAV with a normalized difference vegetation index (NDVI) sensor, allowing for rapid and repeatable measurement of plant vigor and health. However, even the non-specialized Phantoms are very capable devices within the agriculture industry with the help of third-party applications and services. Third party Applications such as Pix4D Capture (Pix4D SA, Prilly, Switzerland), Map Pilot for iOS (Drones Made Easy, San Diego, Cal.), and Litchi (VC Technology Co., Ltd. London, United Kingdom) allow further expansion of the capabilities of the UAV.

Image Capture and Processing

Orthomosaic Mapping

Commonly, UAV data collection is performed on a whole-field basis. Because capturing an entire field in a single image is often impossible due to flight height restrictions, and because most fields would require flying at such an altitude that any level of detail would be lost, this process is commonly used to collect data over a large field or study area. The resulting image, after distortion is removed and individual images are stitched together, is referred to as an “orthophoto”. The orthophoto consists of many single frame images which have been “stitched” together, creating a single image of the entire area of interest. For this type of data collection, flights can be preprogrammed before heading to the imaging site. Users can set their desired altitude, overlap, and flight conditions before the UAV covers the imaging area in a grid-like pattern, capturing images almost continuously. These images then must be joined together using the structure from motion (SFM) process. The SFM process looks for features shared between images and places them together, creating a “stitch” of all images captured (Westoby et. al, 2012).

This data collection method is commonly utilized by fixed-wing UAVs such as the AgEagle series (AgEagle Aerial Systems, Neodesha, Kans.), and is also the method utilized by most crop consulting services. These UAVs are commonly shaped like traditional aircraft, and operate under the same flight principles, meaning they require comparatively little energy to stay in the air when compared to multi-rotor UAVs. Fixed-wing UAVs are capable of covering large areas and can fly for upwards of an hour from a single battery charge, but are not ideally suited for low altitude or tight area flights, due to their inability

to hover in place and their need for a comparatively wide turning radius. Multi-rotor UAV produced by manufacturers such as DJI can also be utilized for orthophoto mapping, but generally can only achieve roughly half of the flight time of a fixed wing UAV. One advantage of multi-rotor UAVs is their ability to hover in place and to change direction at any time. This ability makes multi rotor UAV ideal for plot-scale and smaller field work. An example of a stitched orthophoto image overlaid with field plots is shown in Figure 4.1. The resulting orthophoto is of GeoTiff format, meaning it contains latitude and longitude information, and can be used for analysis on a plot-by plot basis.



(a)



(b)

Figure 4.1: Orthophoto of test plots stitched using OpenDroneMap. Orthophoto creates stitched image of an entire sample area. Use of georeferencing points allows for image to be analyzed on a plot by plot basis.

The number of images required for an orthophoto is directly related to the output quality desired. In order to achieve a higher resolution image, a lower altitude is required, meaning the UAV will not capture as much area in a single shot, requiring more flight time. Additionally, overlap must be increased, meaning the UAV will capture images more frequently, and will travel slower. During flight, spatial events such as a sudden cloud covering the field may render part of the image collected of no benefit. It is not uncommon for the number of images captured to reach into the thousands, taking up gigabytes of

storage space. Additionally, the process of handling and stitching images is computationally intensive and time consuming, making this process unsuitable for most end-users. While many conclusions can be reached by looking at visual differences captured by aerial images, many more relationships can be established by looking below the surface at the data that make up each image. What exactly can be extracted from an image depends mostly on the sensor type used; hyperspectral and multispectral sensors capture color bands not visible to RGB sensors, while lacking the image quality and resolution to spot visual differences. RGB sensors, on the other hand, capture a narrower window of color bands, making them preferred for identification and extraction of spatial features. Most crop consulting firms factor map and recommendation development into their end-user cost, but all major providers of mapping applications including Pix4D, Map Pilot, Agisoft Metashape (Agisoft LLC. St. Petersburg, Russia), DroneMapper (DroneMapper, Cedaredge, Colo.), and DroneDeploy (DroneDeploy, Inc., San Francisco, Cal.) also offer services for stitching and recommendations from user-collected images, either on a credits-based system, or through “software as a service” based subscriptions. OpenDroneMap (OpenDroneMap.org) serves as a free, open source alternative that can be installed on any machine with ample computing resources. OpenDroneMap is constantly being improved and supported by its user base and is beginning support some of these features and analysis tools within its platform.

Point Sampling

Orthophoto imagery is extremely useful for capturing large areas when that a lower resolution image is acceptable. An alternative method for collecting imagery, introduced

here and referred to as the point sampling method, uses an aerial grid sampling technique, capturing single frame images at known points throughout an imaging site as illustrated in Figure 4.2, and creating contoured “heatmap” data using those images. The “FlyLitchi” app while not developed solely for agricultural uses, is available on both Apple and Android platforms, and allows for application of this method when using its “Waypoint” function. The app, similar to orthophoto mapping apps, allows for flight plans to be programmed before arriving at the sampling site. It features integration with comma separated values (CSV) files, allowing for fast importation of sampling points. Point coordinates can be exported from GIS software in CSV format (Figure 4.2 (a)), and then imported into Litchi’s interface (Figure 4.2 (b)). This interface includes options for altitude and flight speed, as well as other commands and settings for image and video capture which can be adjusted according to the desired result.

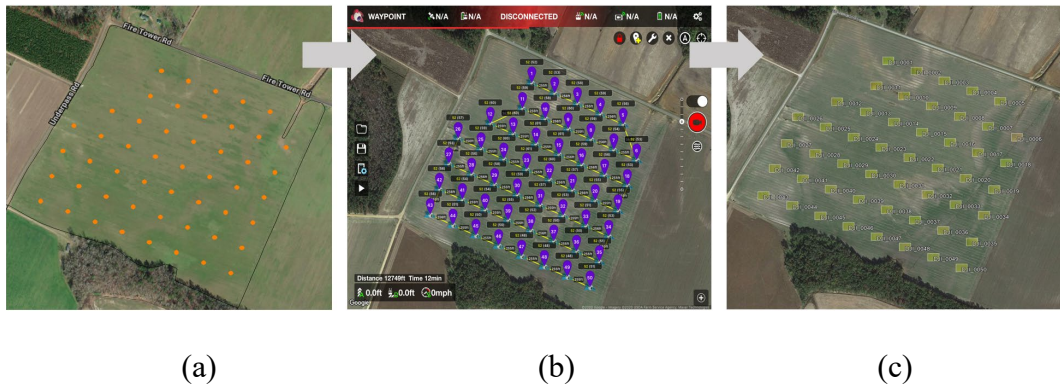


Figure 4.2: Illustration of workflow of point data collection using Litchi app. Image locations exported from GIS software (a), and are loaded into Litchi app(b). UAV collects image at each sampling site (c).

The UAV will then fly to each predetermined point and capture a single image (Figure 4.2 (c)). By creating a sampling grid throughout an entire sample area, a representative sample of the sample area can be represented using these aerial images. This method results in a much lower number of images captured and requires no additional stitching steps. Because images are captured only at desired points, the UAV can cover a larger area during the flight time allowed by battery capacity. Using the point data collection method, a DJI Phantom 3 Advanced UAV can traverse a 40ha (100ac) field, collecting low altitude [30ft above ground level], high resolution images at 50 datapoints in 0.5 ac intervals in approximately 15 minutes under normal flight conditions. Sample images as collected with this process are illustrated in Figure 4.3 (a) and Figure 4.3 (b). To achieve the same image resolution over that same area in orthophoto format could easily extend the flight time into hours.

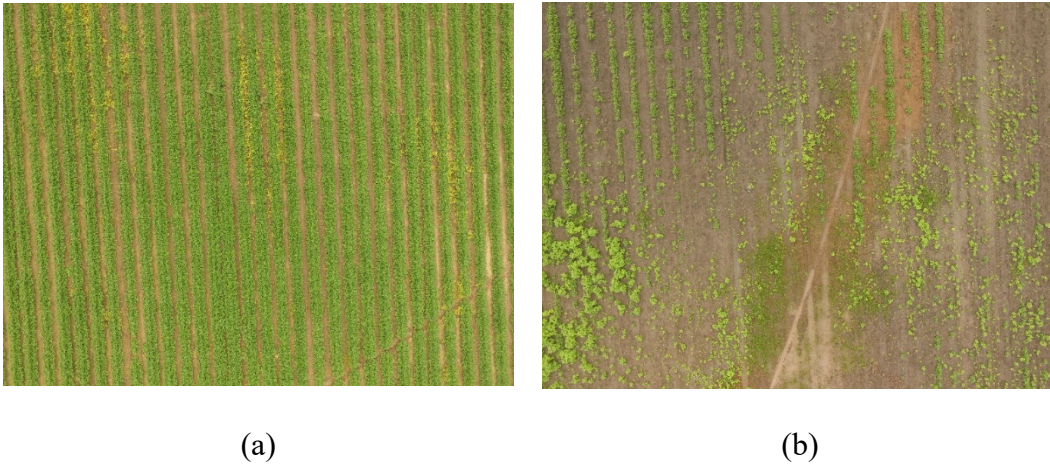


Figure 4.3: Comparison of two UAV images in same field collected on the same date using point data method. Point data method allows for quick "spot checking" across large areas.

This is an ideal method for farmers who are considering entering the UAV market on a low budget, or perhaps have already purchased a UAV and have been utilizing it for visual field inspection. Because each of these images is saved as a single file, they take up no more storage space than images captured with a cell phone or digital camera. Images captured using this method can be analyzed using standalone applications such as Batch Load Image Processor (Clemson University Precision Agriculture, Blackville S.C.) and can provide near real time results, instead of waiting days, or even weeks, for orthophoto processing and results. Using this method, each image sample location will represent a single datapoint, and embedded in each image are the GPS coordinates at which it was taken, making display and analysis of data simple. These points can be loaded into GIS software to be viewed in point form or contoured to create “heatmaps” which can be used to spot relative differences throughout a field. As an example, images from the dataset shown previously in Figure 4.3 were process to extract NDVI from the RGB imagery, and

these results are shown in point data form (Figure 4.4a) and in contoured data form (Figure 4.4b).

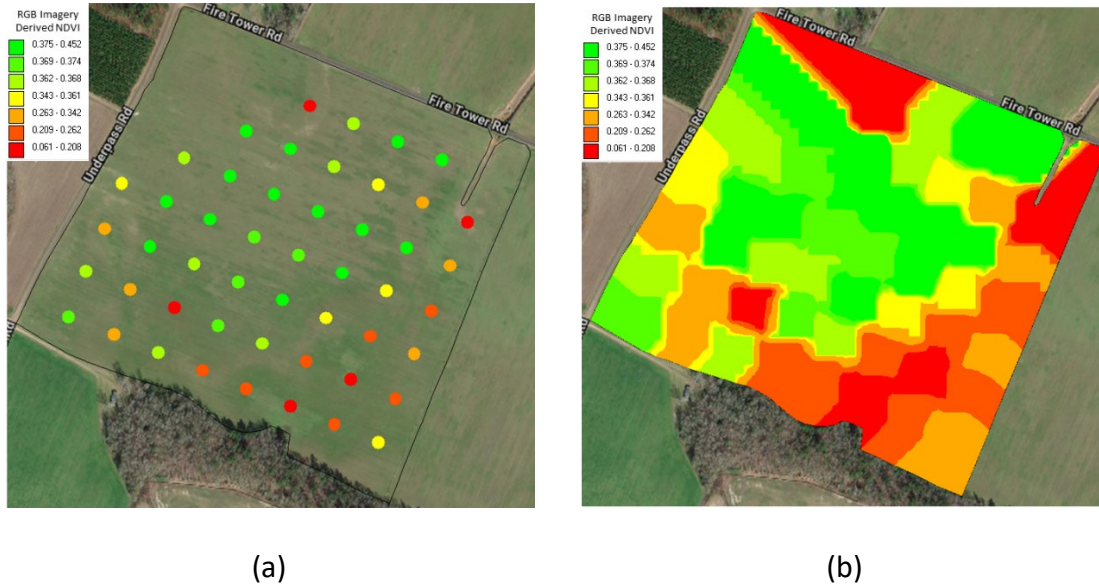


Figure 4.4: NDVI extracted (a) and contoured (b) from RGB UAV using Point Sampling data collection method.

Numerous image and crop features can be extracted using this method, and its uses can be expanded easily. Color features, such as pixel brightness and NDVI can be easily derived, but machine learning and neural networks can also be applied to this data, as seen in the next section, in which data collected using this method was used to develop a cotton boll counting network.

The point sampling method for UAV aerial image data collection provides users the benefits of whole field sampling without the cost, time, and computing power requirements of whole field mapping. By increasing the number of sampling points, more data can be obtained for a negligible increase in flight time. Point sampling is also easily repeatable, making it an ideal method for tracking changes throughout the growing season. It is also extremely adaptable; if a farmer notices a problem with one area of a field, that

area can be designated its own flight plan and can be monitored independently, without consuming time and resources mapping an entire field. This method provides an accurate, “do-it yourself” method for those looking to enter the era of UAV crop scouting without the expense of an external service.

Machine Learning Technologies

The field of artificial intelligence (AI), machine learning, and computer vision are experiencing a period of exponential growth. The job title “Machine Learning Engineer” topped the list of the *Indeed.com* “Best Jobs in the U.S.:2019” with an average base salary of \$146,085, and an impressive 344% growth in job postings from 2015-2018 (Indeed, 2019). Results from studies in these fields are being implemented in numerous aspects of our daily lives, sometimes without our knowing. AI has been implemented at social media giant Facebook in order to rank posts, translate posts between languages, and interpret what is captured in a photograph posted to its website, according to a 2019 paper published by authors working at the company (Wu et al., 2019), and the topic of self-driving cars, which are controlled by AI and machine learning technologies, is of constant importance. The field of agriculture is no exception, and researchers are working globally to solve the world’s food concerns using modern methods, many of which include aspects of machine learning, AI, and computer vision.

Defined by the SAS Institute (2019), machine learning is “a method of data analysis that automates analytical modeling. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention”. Despite its recent explosion in use, machine learning has existed, in

some form, since the dawn of the computer age. Samuel (1959) worked to train a computer to play checkers, and found that within 10 hours of training time, the computer was able to play a better game than the programmer who coded it. Machine learning is commonly broken into two categories of learning; supervised and unsupervised.

Supervised learning, as defined by Sathya and Abraham (2013) is “based on training a data sample from a data source with correct classifications already defined.” This process is perhaps best illustrated by the process of regression, defined by Weisberg (2005) as the study of dependence. In regression and other supervised learning models, a model is developed attempting to match input characteristics that are known to output characteristics that are also known, in attempts to show a relationship. This relationship can then be used to make predictions to new data. A simple practical example tying together the concepts of regression and supervised learning with image analysis can be found in Teddy et al., (2020), in which imagery data collected using a UAV equipped with an RGB camera was used to predict cotton losses caused during the picking process. In this study, all cotton remaining in test plots was collected and weighed, representing the known variable. Images of each test plot were then captured by the UAV, and each image was analyzed to extract, among other characteristics, the red, green, and, blue component values for each image within each plot. These component values were then analyzed statistically to determine what, if any, relationship existed between the weight of cotton collected in a plot, and the color component values of the image of that same plot. In simple terms, was the image color dependent upon the amount of cotton visible in the image? This relationship, in the

form of a model, can be applied to other images of other plots and can issue a prediction using data gained from known points.

In contrast to supervised learning is the method known as unsupervised learning. Using this technique, input data is processed with the goal of discovering hidden patterns, as defined by Liakos et al., (2018). Another way of looking at this method is that, where a supervised learning model will have a list of variables being used to predict an output, an unsupervised learning model will not have this given list of variables but will instead be looking to develop its own patterns and relationships. This method of learning is relied upon by technologies such as neural networks, a technology rapidly being deployed across many industries. In practice, this technology is “trained” using many images which exhibit the trait, pattern, or object that is of interest. After being exposed to many images of the object, the network is then able to identify it when exposed to an image not in the training dataset.

To provide an example of machine learning and neural networks in agriculture, a sample model was developed to identify white cotton blooms from low altitude aerial imagery collected via a DJI Phantom 3 UAV. This model was not created to report optimal model-specific settings, but rather to illustrate an example of machine learning, neural networks, and UAV imagery.

MaskRCNN Neural Network Sample Model

The UAV was programmed with the waypoint feature of the Litchi app, using the point data collection method previously described, and was programmed to capture images at an altitude of approximately 11m (35ft) above ground level (AGL). The resulting images

were captured using the UAV's integrated RGB camera, model DJI FC300s. Images captured were of resolution 4,000 x 3000 pixels and were stored in .JPEG format. This altitude and camera sensor combination resulted in images with a ground sample distance of 0.46 cm px⁻¹ (0.18 in. px⁻¹). A total of 10 images were captured on 10 August 2018 over sample sites of blooming cotton in Barnwell County, S.C. To aid in training speed, each image was tiled into a grid of four rows and eight columns using the latest version of IrfanView (Irfan Škiljan, Austria, Europe). Each of the resulting images was of size 500 pixels x 750 pixels. 80% of the resulting tiled images were randomly assigned to a training dataset, while the remaining 20% were assigned to the validation dataset. Each image within the training dataset was then annotated using the VGG Image Annotator (Dutta and Zisserman 2019). The annotation process involves loading each image, then manually selecting features to be identified, and sorting them into a category. In this experiment, the only category was “white cotton bloom”; each image was loaded, and a polygon was drawn around each bloom visible in the image as illustrated in Figure 4.5. Annotating images serves to create a “dictionary” of known images within each category, that will be used to train the model and eventually identify the same category in validation images. This process was repeated for all images within the training dataset, and the output was saved as a .JSON file.

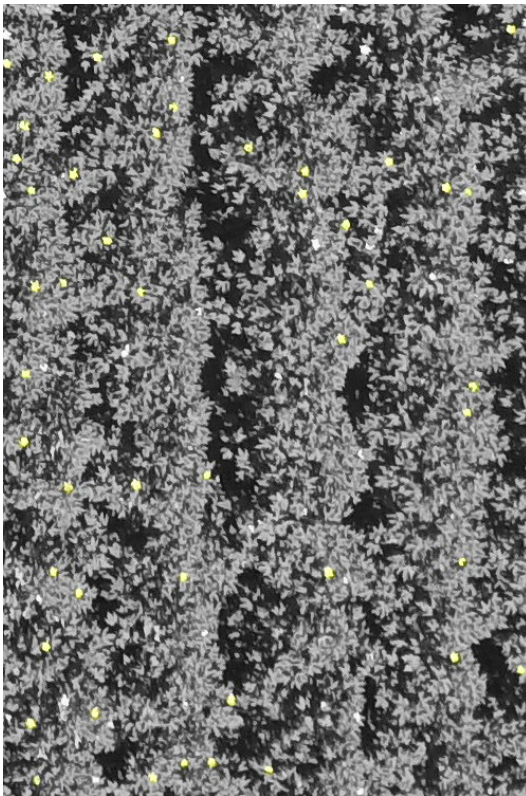


Figure 4.5: Image illustrating manual image annotation using VGG Image Annotator. Polygon drawn around each cotton bloom to build "dictionary" of training images. Each bloom given a unique number.

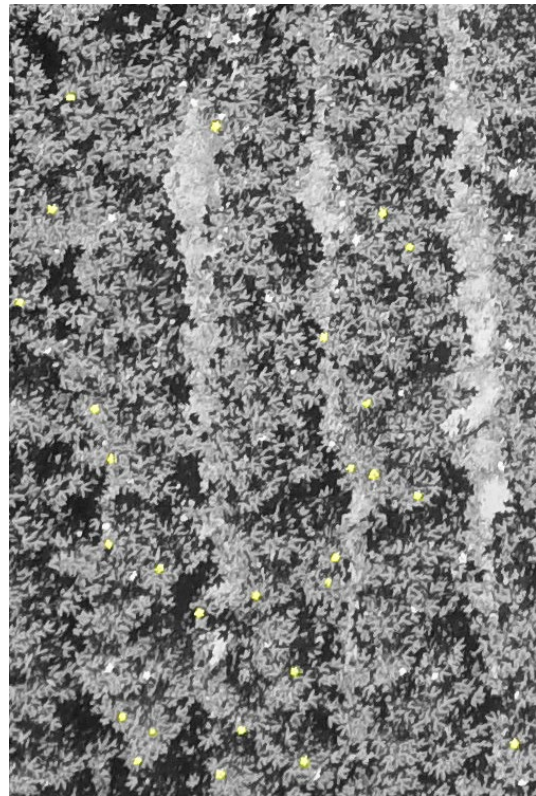
The neural network was then developed using MaskRCNN, an instance segmentation framework developed by Facebook AI Research and published by He et al., (2017). Instance segmentation, as defined by Parades and Torr (2016) is the problem of detecting and delineating each distinct object of interest appearing in an image. Whereas object detection works to simply identify if a particular object exists in an image, image segmentation aims to identify all instances of that object within the image. MaskRCNN is able to analyze each pixel in an image and assign it to one of the classification categories. It then “masks” each instance of the object being classified. MaskRCNN has been implemented to extract and count buildings (Zhao et al., 2018), identify cell nuclei (Johnson 2018), and detect individual strawberries in development of an automated harvester (Yu et al., 2019). The MaskRCNN framework provides results which can be

implemented with pixel and object counts, as well as color value categorization. These results can then be implemented into end-user applications.

The demonstration model was heavily based upon the MaskRCNN “balloon” example (Matterport 2018), and was trained using mostly default settings using the latest versions of Python 3, Keras, and Tensorflow. Model training was accomplished by use of the Clemson University Palmetto Cluster, and by utilizing available nVidia CUDA-capable graphics processing units (GPUs). Upon conclusion of model training, some example images from the “validation” dataset were processed, with the results illustrated in Figure 4.6 (a) and (b). MaskRCNN applied a grayscale filter to the entire image, except for the segmented objects, white blooms in this case.



(a)



(b)

Figure 4.6. MaskRCNN cotton bloom identification model example images. MaskRCNN applied grayscale filter to entire image, and applied yellow mask to white cotton blooms.

Model Implementation

Although model evaluation is beyond the scope of this discussion, it is visibly apparent that the majority of blooms were identified and masked. Model robustness would be improved with a wider training dataset, as well as more training epochs. Additionally, the final model would likely provide a benefit to the end user, such as a single number representing the number of blooms identified. This could then be parsed with yield data, irrigation data, or soil data to aid in zone management decisions. Because model training requires the brunt of computing power, execution and application of the model on non-training imagery can be performed on nearly all consumer computers and smartphones. Models such as the above can be implemented using platforms such as Django (Django Software Foundation, Lawrence, Kans.) or Flask (Armin Ronacher, Austria).

Conclusion

The objective of this document has been to provide, with specific examples, links between agriculture, UAV / remote sensing, and machine learning technologies, and to introduce a point sampling method for aerial data collection. All three are vast and rapidly expanding areas of interest for researchers, professionals such as crop consultants, and growers.

The point sampling method discussed here allows images to be used as representative samples of a field or other area of interest, reducing flight time, processing time, and data storage required for aerial image collection. Data extracted from images

collected using this point sampling method can be subsequently processed using traditional, regression models, or machine learning algorithms. These data outputs can then be mapped to show a heatmap or contour map of the resulting data.

While currently it would be an overestimation to expect to replace physical data collection with remotely sensed data, new technologies are being developed constantly that are working to narrow the divide, and remote sensing and artificial intelligence have cemented their place in the future of agriculture.

References

- Ali, I., Cawkwell, F., Dwyer, E., & Green, S. (2017). Modeling managed grassland biomass estimation by using multitemporal remote sensing data—a machine learning approach. *IEEE J. Selected Topics in Applied Earth Observations and Remote Sensing*, 10(7), 3254–3264. doi: 10.1109/jstars.2016.2561618
- Association for Unmanned Aerial Systems International (AUVSI). (March 2013). The Economic Impact of Unmanned Aircraft Integration in the United States. Retrieved from: <https://www.auvsi.org/our-impact/economic-report>
- 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), 10395–10412. <https://doi.org/10.3390/rs61110395>
- Dodge, W.G. (2019). Image based yield estimation in cotton using UAS. Master's Thesis. Texas Tech. University. Lubbock, Texas. Retrieved from: <https://ttuir.tdl.org/handle/2346/85003>
- Dutta, A., & Zisserman, A. (2019). The VIA annotation software for images, audio and video. *Proc. of the 27th ACM International Conference on Multimedia. MM '19: The 27th ACM International Conference on Multimedia*. <https://doi.org/10.1145/3343031.3350535>
- He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). Mask R-CNN. *Intl. Conf. Computer Vision (ICCV)*. IEEE (2017), pp.2980-2988
- Indeed Editorial Team. (2019 March 14). The Best Jobs in the U.S.:2019. Indeed Blog. <http://blog.indeed.com/2019/03/14/best-jobs-2019/>
- Johnson, J. (2018). Adapting Mask-RCNN for automatic nucleus segmentation. *Proc. 2019 Computer Vision Conf.* doi: 10.1007/978-3-030-17798-0
- Lampert, A. (2019 November 13). Amid privacy backlack, China's DJI unveils drone-to-phone tracking. Reuters. <https://www.reuters.com/article/us-aviation-drones/amid-privacy-backlash-chinas-dji-unveils-drone-to-phone-tracking-idUSKBN1XN2JR>
- Liakos, K., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors*, 18(8), 129. doi: 10.3390/s18082674
- Machine learning: What it is and why it matters. (2019 September 28). SAS Institute. Retrieved from: https://www.sas.com/en_us/insights/analytics/machine-learning.html
- Matterport. MaskRCNN “Balloon Example”. (2018). GitHub Repository, https://github.com/matterport/Mask_RCNN
- Reza, M. N., Na, I. S., Baek, S. W., & Lee, K.-H. (2019). Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images. *Biosystems Engineering*, 177, 109–121. <https://doi.org/10.1016/j.biosystemseng.2018.09.014>
- Romera-Paredes, B., & Torr, P. H. S. (2016). Recurrent instance segmentation in computer vision. *ECCV 2016* (pp. 312–329). Springer International Publishing. https://doi.org/10.1007/978-3-319-46466-4_19

- Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210–229.
<https://doi.org/10.1147/rd.33.0210>
- Sathya, R., & Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *Intl. J. of Advanced Research in Artificial Intelligence*, 2(2), 34–38. doi: 10.14569/ijarai.2013.020206
- UAS by the Numbers. (2020, March 10). Retrieved from
https://www.faa.gov/uas/resources/by_the_numbers/
- Weisberg, S. (2005). *Applied linear regression*. Hoboken, N.J: Wiley-Interscience
- Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., & Reynolds, J. M. (2012). “Structure-from-Motion” photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179, 300–314.
<https://doi.org/10.1016/j.geomorph.2012.08.021>
- Wu, C.-J., Brooks, D., Chen, K., Chen, D., Choudhury, S., Dukhan, M., ... Zhang, P. (2019). Machine Learning at Facebook: Understanding Inference at the Edge. *2019 IEEE Intl. Symp. on High Performance Computer Architecture (HPCA)*, 331–334. doi: 10.1109/hpca.2019.00048
- Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, 163, 104846.
<https://doi.org/10.1016/j.compag.2019.06.001>
- Zhao, K., Kang, J., Jung, J. and Sohn, G., (2018). Building extraction from satellite images using Mask R-CNN with building boundary regularization. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, 2018, pp. 242-2424, doi: 10.1109/CVPRW.2018.00045.

CHAPTER FIVE

CONCLUSIONS

The objectives of these studies were to develop linear regression models to predict sand, clay, and organic matter content using RGB images collected from two types of remotely sensed aerial imagery; (1) images collected from a consumer-level UAV and (2) images collected using, Internet- available satellite imagery, and (3) to introduce the point sampling method of aerial data collection and provide an outline of current methods in which machine learning and artificial intelligence can be implemented in precision agriculture using UAV imagery.

As their initial cost decreases, adaptation rates of UAV are rising, as are their uses. Using imagery data collected from a consumer level UAV equipped with an RGB camera, regression models were consistently able to predict within 5.84% sand content, 5.07% clay content, and 0.93% OM content, all at 90% confidence; i.e. 90% of the prediction errors were less than the values listed here. When tested at two flight altitudes (21m and 30m), there were no significant differences in prediction ability of sand, clay, or OM content, however flying at the lower altitude provides a greater ground sample distance and may be preferred if other data is to be extracted from the image especially because the flight time does not change between altitudes.

Similarly, Internet- available satellite imagery allow for rapid data collection without having to enter a field. In this study, satellite imagery data from Planet Labs and Google Earth was used to predict sand, clay, and OM content over unplanted, bare soil fields. Due to imagery availability, the two imagery sources utilized different fields. With

that in mind, models created using Google Earth images had significantly lower error when predicting sand, clay, and OM content when compared to Planet Labs imagery. This is likely due to the lower spatial resolution (3m/pix) of Planet Labs data, but also potentially due to analysis methods. The best performing Google Earth regression models were consistently able to predict within 6.37% sand content, 3.34% clay content, and 0.59% OM content, all at 90% confidence; i.e. 90% of the prediction errors were less than the values listed here. Of the two methods utilized, models developed from UAV captured imagery, generally result in lower error than those developed utilizing satellite captured imagery, at the added cost of the initial investment of software and hardware purchases.

Although these results are promising, they do not suggest that these methods are replacements for traditional soil sampling procedures, as they are unable to provide added benefits such as nutrient analysis. Additionally, these models are affected by physical soil properties, such as soil moisture, which can lead to increased error. These methods do have benefit in management zone development, as they allow for mapping of relative textural differences throughout an area.

The point sampling method for aerial data collection provides a faster, less resource intensive method for those looking to collect data over fields when compared to whole field sampling and orthophoto generation. This method can be performed without extra fees or services once the initial application and hardware are purchased. As future prediction models for soil and crop qualities are developed, this method can provide a framework for them to be utilized on, and provide near real time field data collection.

APPENDICES

Appendix A

Examples of Regression Model Bounding Coefficients

UAV Sand Content Prediction model bounding using “Combined” altitude model (Table 2.7). Model terms and bounding results continue in second image.

Tolerance		1	
BRT(23)			
100.00% maximum	34.572	G(1)	
99.50%	33.14442	100.00% maximum	0.063
97.50%	27.08095	99.50%	0.04549
90.00%	24.0656	97.50%	0.014425
75.00% quartile	21.65	90.00%	0.0027
50.00% median	18.8535	75.00% quartile	0
25.00% quartile	13.8225	50.00% median	0
10.00%	1.5436	25.00% quartile	0
2.50%	0.7423	10.00%	0
0.50%	0.40407	2.50%	0
0.00% minimum	0.333	0.50%	0
		0.00% minimum	0
Range	34.239	Range	0.063
StDev	8.55975	StDev	0.01575
Upper	68.811	Upper	0.126
Lower	-33.906	Lower	-0.063
G(25)			
100.00% maximum	29.693	G(28)	
99.50%	25.95977	100.00% maximum	21.511
97.50%	18.81968	99.50%	18.38289
90.00%	15.3393	97.50%	9.285475
75.00% quartile	12.1145	90.00%	1.5207
50.00% median	8.021	75.00% quartile	0.133
25.00% quartile	5.12225	50.00% median	0.007
10.00%	0.0256	25.00% quartile	0.001
2.50%	0.003	10.00%	0
0.50%	0.001	2.50%	0
0.00% minimum	0.001	0.50%	0
		0.00% minimum	0
Range	29.692	Range	21.511
StDev	7.423	StDev	5.37775
Upper	59.384	Upper	43.022
Lower	-29.691	Lower	-21.511
G(30)			
100.00% maximum	9.59	B(15)	
99.50%	8.94316	100.00% maximum	23.103
97.50%	0.8991	99.50%	21.85104
90.00%	0.0067	97.50%	18.15443
75.00% quartile	0.002	90.00%	14.0525
50.00% median	0	75.00% quartile	9.382
25.00% quartile	0	50.00% median	4.2195
10.00%	0	25.00% quartile	2.62825
2.50%	0	10.00%	1.7686
0.50%	0	2.50%	1.144025
0.00% minimum	0	0.50%	0.58797
		0.00% minimum	0.383
Range	9.59	Range	22.72
StDev	2.3975	StDev	5.68
Upper	19.18	Upper	45.823
Lower	-9.59	Lower	-22.337
B(18)			
100.00% maximum	22.992		
99.50%	22.7963		
97.50%	20.66075		
90.00%	16.1612		
75.00% quartile	13.641		
50.00% median	10.774		
25.00% quartile	7.7975		
10.00%	1.3607		
2.50%	0.278425		
0.50%	0.04024		
0.00% minimum	0.032		
Range	22.96		
StDev	5.74		
Upper	45.952		
Lower	-22.928		

UAV Clay Content Prediction model bounding using “Combined” altitude model (Table 2.11). Model terms and bounds continue in second image.

Tolerance 0.99

R(31) ²		G(26) ²		((R-G)/(G+B)) ³	
100.00% maximum	62.606	100.00% maximum	878.2332	100.00% maximum	0.042748
99.50%	61.473	99.50%	659.7471	99.50%	0.04247
97.50%	49.81885	97.50%	224.8158	97.50%	0.034593
90.00%	37.311	90.00%	103.7949	90.00%	0.02591
75.00% quartile	18.98825	75.00% quartile	25.38904	75.00% quartile	0.011568
50.00% median	0.1745	50.00% median	4.050187	50.00% median	0.00599
25.00% quartile	0.02	25.00% quartile	0.541888	25.00% quartile	0.003442
10.00%	0.003	10.00%	0.000009	10.00%	0.001918
2.50%	0	2.50%	0	2.50%	0.000796
0.50%	0	0.50%	0	0.50%	0.000207
0.00% minimum	0	0.00% minimum	0	0.00% minimum	0.000164
Range	62.606		878.2332		0.042584
StdDev	15.6515		219.5583		0.010646
Upper	124.5859		1747.684		0.084906
Lower	0.62606		8.782332		0.00059
G(28) ²		B(17) ³		B(28) ³	
100.00% maximum	462.7231	100.00% maximum	11846.12	100.00% maximum	0.000389017
99.50%	347.1457	99.50%	11394.49	99.50%	0.000330105
97.50%	86.22533	97.50%	7295.925	97.50%	2.29877E-05
90.00%	2.324726	90.00%	3982.033	90.00%	2.5799E-06
75.00% quartile	0.017737	75.00% quartile	1605.424	75.00% quartile	0.000000064
50.00% median	0.000049	50.00% median	535.8827	50.00% median	0
25.00% quartile	0.000001	25.00% quartile	187.3255	25.00% quartile	0
10.00%	0	10.00%	15.05342	10.00%	0
2.50%	0	2.50%	0.464514	2.50%	0
0.50%	0	0.50%	0.022651	0.50%	0
0.00% minimum	0	0.00% minimum	0.015069	0.00% minimum	0
	462.7231		11846.1		0.000389017
	115.6808		2961.525		9.72543E-05
	920.819		23573.75		0.000774144
	4.627231		118.4761		3.89017E-06

UAV OM Content Prediction model bounding using “Combined” altitude model (Table 2.11). Model terms and bounds continue on second line.

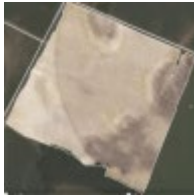
Tolerance		0									
G(6)		R(30)^3		SQRT((B(31))		G(21)^2		G(17)^3			
100.00% maximum	1.055	100.00% maximum	40551.58	100.00% maximum	0.921412	100.00% maximum	390.8134	100.00% maximum	1726.704		
99.50%	0.866405	99.50%	31702.57	99.50%	0.856958	99.50%	387.3647	99.50%	1714.156		
97.50%	0.5184	97.50%	19087.84	97.50%	0.659943	97.50%	316.1986	97.50%	1420.765		
90.00%	0.2426	90.00%	10200.14	90.00%	0.528393	90.00%	242.5837	90.00%	763.08		
75.00% quartile	0.002	75.00% quartile	5484.627	75.00% quartile	0.29612	75.00% quartile	192.3078	75.00% quartile	20.62154		
50.00% median	0	50.00% median	4.398125	50.00% median	0	50.00% median	132.3549	50.00% median	6.128685		
25.00% quartile	0	25.00% quartile	0.050142	25.00% quartile	0	25.00% quartile	73.55301	25.00% quartile	2.480936		
10.00%	0	10.00%	9.56E-05	10.00%	0	10.00%	13.21905	10.00%	0.710623		
2.50%	0	2.50%	7.29E-07	2.50%	0	2.50%	4.613378	2.50%	0.081281		
0.50%	0	0.50%	1.34E-08	0.50%	0	0.50%	1.231725	0.50%	0.00073		
0.00% minimum	0	0.00% minimum	0	0.00% minimum	0	0.00% minimum	0.665856	0.00% minimum	0.000343		
Range	1.055		40551.58		0.921412		390.1475		1726.704		
StdDev	0.26375		10137.9		0.230353		97.53688		431.676		
Upper	1.055		40551.58		0.921412		390.8134		1726.704		
Lower	0		0		0		0.665856		0.000343		
R(28)^2		CubeRoot[R(25)]		G(28)^2		B(18)^2		B(16)^3			
100.00% maximum	1126.542	100.00% maximum	2.863311	100.00% maximum	462.7231	100.00% maximum	528.6321	100.00% maximum	14934.19		
99.50%	1086.324	99.50%	2.836085	99.50%	351.6342	99.50%	520.054	99.50%	14369.26		
97.50%	671.9876	97.50%	2.67974	97.50%	86.49864	97.50%	427.1571	97.50%	6080.735		
90.00%	417.3296	90.00%	2.537545	90.00%	2.660632	90.00%	262.4456	90.00%	3853.228		
75.00% quartile	312.5528	75.00% quartile	2.412435	75.00% quartile	0.022658	75.00% quartile	190.3711	75.00% quartile	1552.328		
50.00% median	182.5471	50.00% median	2.200892	50.00% median	0.000049	50.00% median	116.5538	50.00% median	188.285		
25.00% quartile	81.32987	25.00% quartile	1.66936	25.00% quartile	7.5E-07	25.00% quartile	61.47738	25.00% quartile	51.55576		
10.00%	7.850143	10.00%	1.252372	10.00%	0	10.00%	1.690785	10.00%	20.65888		
2.50%	1.101997	2.50%	1.005807	2.50%	0	2.50%	0.075743	2.50%	3.421798		
0.50%	0.286844	0.50%	0.902796	0.50%	0	0.50%	0.001658	0.50%	0.425263		
0.00% minimum	0.2116	0.00% minimum	0.897938	0.00% minimum	0	0.00% minimum	0.001024	0.00% minimum	0.233745		
	1126.33		1.965373		462.7231		528.631		14933.95		
	281.5826		0.491343		115.6808		132.1578		3733.489		
	1126.542		2.863311		462.7231		528.6321		14934.19		
	0.2116		0.897938		0		0.001024		0.233745		

Appendix B

Dates of Satellite Image Capture

Planet Labs Data

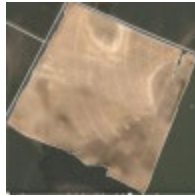
Big Pivot West Field



5-15-2016



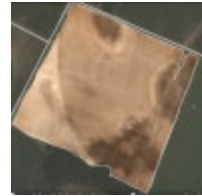
5-6-2017



5-7-2017



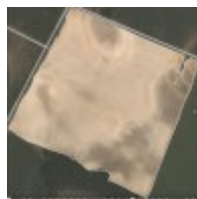
5-11-2017



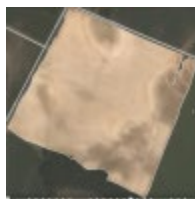
5-17-2017



5-27-2017



4-29-2018



5-1-2018



5-4-2018



5-13-2018

Don Still Field



3-22-2018



3-31-2018



4-5-2018



4-17-2018



4-19-2018



4-20-2018



4-29-2018



5-1-2018



5-20-2018



5-10-2018

Market Front Field



3-27-2017



4-23-2017



5-7-2017



5-11-2017



5-14-2017



6-4-2017



4-18-2018



4-21-2018



5-4-2018

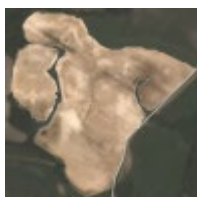


5-7-2018

Watermelon Road Field



4-28-2017



5-7-2017



5-17-2017



5-27-2017



4-11-2018



4-28-2018



4-29-2018



5-4-2018



5-9-2018



5-13-2018

E-7 Field



6-17-2016



5-19-2017



6-8-2017



3-31-2018



4-2-2018



5-4-2018



5-12-2018



6-3-2018



6-7-2018



6-19-2018

Google Earth Data

Rusty Pivot Field



Rusty Pivot Field Image Dated: 3-11-2004

CP4 & 13 Field



CP4 & 13 Field Image Dated: 6-11-2015

Chickenhouse Field



Chickenhouse Field Image Dated 6-9-2011

B6B Field



B6B Field Image Dated 6-9-2011

C12 Field



C12 Field Image Dated 6-9-2011

E7 Field



E7 Field Image Dated 6-11-2011. Only upper third of right field and field left of pivot used for model development.

Appendix C

Regression Models and Coefficients for Satellite Imagery Models

Sand Content Prediction

Planet Labs sand content prediction model using 25-foot sample circles.

Term	Estimate	Std Error
Intercept	-21.34	15.66
$(R-G)/R^2$	-517.3	54.56
$\text{Log}(R)$	27.77	2.82
$(G-B)/(R-G)^3$	-25.52	2.12

Planet Labs sand content prediction model using 50-foot sample circles

Term	Estimate	Std Error
Intercept	-5989	654.22
$((R+B)/(G-B))^3$	0.0024	0.0009
$\sqrt[3]{[(R+G+B)/(R+B)]}$	5346	574.3
$\text{Log}(\text{SatHSL})$	8.822	1.076
$(G-B)/(R-G)^3$	-38.11	3.198

Planet Labs Sand content prediction model using all sample circles

Term	Estimate	Std Error
Intercept	100.0	1.832
$R+G^3$	3.543e-7	2.563e-8
$(G-B)/(R-G)^3$	-23.33	1.197
SatHSL^3	-35.65	5.131
$(R-B)/(G+B)^3$	-540.9	46.986

Google Earth sand content prediction model using 25-foot sample circles.

Term	Estimate	Std Error
Intercept	-53.45	12.49
$(R-G)/(R-B)$	5.856	1.345
$\text{Log}(R+B)$	24.16	2.101
$(G+B)/(R-B)^3$	3.196e-6	1.236e-6

Google Earth sand content prediction model using 50-foot sample circles.

Term	Estimate	Std Error
Intercept	92.04	0.59
$G-B^3$	-9.611e-5	4.643e-5
$(R-G)/B^3$	-370.6	33.36
$(R-G)/(G-B^3)$	-0.1047	0.0415

Google Earth sand content prediction model using all sample circles.

Term	Estimate	Std Error
Intercept	89.55	0.6319
B^3	2.289e-7	9.769e-7
$(R-G)/B^3$	-332.9	29.51

Both sources Sand content prediction model using 25ft sample circles.

Term	Estimate	Std Error
Intercept	52.17	20.02
$SatHSL^2$	-11.51	3.502
$(R+G+B)/(R+B)^2$	-64.61	11.57
$\text{Log}(BRT)+1$	32.11	2.465

Both sources sand content prediction model using 50ft sample circles.

Term	Estimate	Std Error
Intercept	71.23	1.901
R^2	0.000369	3.488e-5
$(R+B)/(G-B)^2$	0.00277	0.000862
$\sqrt[3]{(R-G)}$	0.7598	0.3285
$(R-G)/B^3$	-62.18	19.04

Both sources sand content prediction model using all sample circles.

Term	Estimate	Std Error
Intercept	23.14	6.539
R^3	2.2712e-6	1.464e-7
$(R+B)/G^3$	5.122	0.5782
$\sqrt[3]{SatHSI}$	22.18	5.731

R-B ³	-3.244e-5	3.696e-6
------------------	-----------	----------

Clay Content Prediction

Planet Labs clay content prediction model using 25-foot sample circles

Term	Estimate	Std Error
Intercept	46.73	18.36
$\sqrt[3]{((G-B)/(R-G))}$	80.78	14.28
Log(R+B)	-21.58	1.981
$(G+B)/(R-G)^3$	0.0019	.0007
$(R-G)/B^3$	176.1	19.47

Planet Labs clay content prediction model using 50-foot sample circles

Term	Estimate	Std Error
Intercept	15.88	15.88
$(G+B)/(R-G)^3$	0.0076	.0005
Log(R)	3.276	3.275
G-B ³	0.0006	4.234e-5

Planet Labs Clay content prediction model using all sample circles

Term	Estimate	Std Error
Intercept	-19.37	1.307
$(G+B)/(R-G)^3$	0.0035	0.0005
SatHSL	-7.425	0.5216
SatHSI ³	1118	67.76
$(G-B)/(R-G)^3$	7.296	1.087

Google Earth clay content prediction model using 25-foot sample circles.

Term	Estimate	Std Error
Intercept	127.1	22.95
$\sqrt{[(R-G)/(R-B)]}$	-7.730	2.198
$(G-B)/R^2$	-270.3	78.65
Log[R+1]	-22.45	4.159
G-B ³	0.0003	7.513e-5

Google Earth clay content prediction model using 50-foot sample circles.

Term	Estimate	Std Error
Intercept	79.62	10.81
R-B	-0.2592	0.0384
Log[R+1]	-13.29	1.967
G-B ³	0.0003	4.45e-5
R-G ³	0.0001	4.755e-5

Google Earth clay content prediction model using all sample circles.

Term	Estimate	Std Error
Intercept	5.519	0.7757
SatHSL	-15.58	3.707
$\sqrt{[(R+G)/(R-B)]}$	0.3710	0.0962
$(G-B)/(R+B)^2$	392.1	57.33

Both sources Clay content prediction model using 25ft sample circles.

Term	Estimate	Std Error
Intercept	75.56	7.669
$(G-B)/(R-B)$	4.100	0.9218
Log[(BRT_HSP)+1]	-14.15	1.582
G-B ³	7.634e-5	1.839e-5

Both sources clay content prediction model using 50ft sample circles.

Term	Estimate	Std Error
Intercept	14.96	1.629
$(G-B)/R^3$	1193	154.4
$\sqrt[3]{\text{SatHSL}}$	-17.35	2.704

Both sources clay content prediction model using all sample circles.

Term	Estimate	Std Error
Intercept	17.41	1.313
$\sqrt[3]{\text{SatHSL}}$	-15.54	2.275
G ³	-3.078e-7	6.17e-8
R-G ³	5.7259e-5	8.58e-6

OM Content Prediction

Planet Labs OM content prediction model using 25-foot sample circles

Term	Estimate	Std Error
Intercept	0.7803	0.2483
(RedGreen Taylor Series M) ³	0.0178	0.0016
Log(SatHSL)	-0.6583	0.2036
(R-G) ³	5.341e-6	2.262e-6

Planet Labs OM content prediction model using 50-foot sample circles

Term	Estimate	Std Error
Intercept	2.571	0.3625
SatHSL	-0.9773	0.3393
RedBlue Taylor Series P ²	0.1557	0.0072
Log(SatHSV)	0.8431	0.2729

Planet Labs OM content prediction model using all sample circles

Term	Estimate	Std Error
Intercept	9.236	1.435
SatHSL	-1.192	0.2771
RedGreen Taylor Series M ³	0.0139	0.0016
Log[(G+B)/G]	-12.11	2.377
GreenBlue Taylor Series N ³	0.0135	0.0021

Google Earth OM content prediction model using 25-foot sample circles.

Term	Estimate	Std Error
Intercept	0.5663	.0.265
$\sqrt[3]{(R-G)}$	-0.1356	0.0116
(G+B)/(R-B) ³	-1.091e-7	5.727e-7

Google Earth OM content prediction model using 50-foot sample circles.

Term	Estimate	Std Error
Intercept	0.9883	0.0951
R-G ³	6.122	9.305e-6

Google Earth OM content prediction model using all sample circles.

Term	Estimate	Std Error
Intercept	-0.1498	0.0917
RedGreen Taylor Series P	-0.0111	0.0029
$\sqrt[3]{\text{GreenBlue}}$ Taylor Series P	0.2072	0.0151
$R-G^3$	1.102e-5	5.059e-6

Both sources OM content prediction model using 25ft sample circles.

Term	Estimate	Std Error
Intercept	-6.614	0.7843
$\sqrt[3]{[(G-B)/(R-G)]}$	-0.180	0.0757
$\sqrt[3]{[(G-B)/R]}$	11.63	1.253
$\sqrt[3]{\text{RedBlue}}$ Taylor Series M	1.854	0.2285
RedBlue Taylor Series P ³	-0.0006	0.0002

Both sources OM content prediction model using 50ft sample circles.

Term	Estimate	Std Error
Intercept	-0.2370	0.3727
$\sqrt[3]{\text{RedGreen}}$ Taylor Series O	2.415	0.2565
RedBlue Taylor Series M ²	-0.000101	1.081e-5
G^3	-1.674e-7	1.925e-8
$(R-G)/(G-B)^3$	-0.1025	0.01136

Both sources OM content prediction model using all sample circles.

Term	Estimate	Std Error
Intercept	-0.2602	0.141
$\sqrt{\text{RedGreen}}$ Taylor Series M	1.585	0.101
$(R+B)/(G-B)^2$	-0.0012	8.578e-5
RedGreen Taylor Series P ³	-0.0006	8.67e-5