Assessing the Performance of Multispectral Sensors Mounted on Unmanned Aerial Vehicles for the Prediction of Soil Organic Carbon Levels at Field-Scale

by

Michael March

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Master of Science
in
Geography

Guelph, Ontario, Canada
© Michael March, January, 2020
ABSTRACT

ASSESSING THE PERFORMANCE OF MULTISPECTRAL SENSORS MOUNTED ON UNMANNED AERIAL VEHICLES FOR THE PREDICTION OF SOIL ORGANIC CARBON LEVELS AT FIELD-SCALE

Michael March
University of Guelph, 2019

Advisor: Dr. Aaron Berg
Committee Member: Dr. Ahmed Laamrani

The quantification of soil organic carbon (SOC) is critical for sustainable agricultural production. Conventional field measurements for assessing SOC content are time-consuming, costly and require large soil sampling efforts. The remote monitoring of SOC using unmanned aerial vehicles (UAVs) possesses the capability to be faster and more economically advantageous when compared to conventional soil sampling methods. This research sought to examine the potential of UAV-mounted multispectral (400-800nm) sensors for SOC prediction at the sub-field scale. To do so, UAV-based imagery was acquired over agricultural fields under bare soil conditions; and a total of 806 georeferenced soil samples were collected at 20m intervals for each study site. We used multivariate regression analysis to assess the relationship between SOC and reflectance. The R² and RMSE were calculated between estimated and observed SOC. Laboratory and UAV reflectance were combined to explore the potential of transferrable models that could estimate SOC across various platforms.
ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere and utmost gratitude to Dr. Aaron Berg for his support, advice and direction in the preparation of this thesis. Additionally, I would like to thank my committee member, Dr. Ahmed Laamrani, for all his assistance, feedback and suggestions that he provided throughout my entire thesis. I would also like to acknowledge Alexander (Sandy) McLaren for his outstanding efforts and countless hours of drone data collection.

To my parents, Marie and Glenn March, my brother Nicholas March and my sister in law Katie March, I would like to recognize your endless support and encouragement throughout this entire process. To my mother and father in law, Allison Rippin-Armstrong and John Armstrong, sisters in law, Emily and Clara Armstrong and brother in law Tyler Grant, thank you for your advice, inspiration and encouragement. Finally, I would like to thank my partner, Amanda Armstrong for all her unwavering guidance, optimism, advice and love throughout the entirety of my thesis. This thesis would not have been possible without each of you.
# TABLE OF CONTENTS

Abstract ........................................................................................................................................... ii

Acknowledgements ....................................................................................................................... iii

Table of Contents ......................................................................................................................... iv

List of Tables ............................................................................................................................... vi

List of Figures ............................................................................................................................. vii

1 Introduction ............................................................................................................................ 1

2 Literature Review ..................................................................................................................... 4

  2.1 Introduction .......................................................................................................................... 4

  2.2 Significance of SOC for Agricultural Systems ................................................................. 4

  2.3 Effects of Agricultural Management Practices on SOC Levels ................................... 7

    2.3.1 Cropping Systems ....................................................................................................... 7

    2.3.2 Quantification of Spatial Variability of SOC ............................................................. 10

  2.4 Current and Emerging Analytical Methods Used for Measuring SOC .................. 12

    2.4.1 Dry Combustion ......................................................................................................... 12

    2.4.2 Wet Digestion ............................................................................................................. 13

    2.4.3 Laboratory Spectroscopy .......................................................................................... 14

    2.4.4 Field Spectroscopy ................................................................................................... 16

  2.5 UAV Applications in SOC Mapping .................................................................................. 18

  2.6 SOC Modelling ..................................................................................................................... 21

    2.6.1 Current Models for SOC Prediction ........................................................................... 21

  2.7 Conclusion ............................................................................................................................ 23

3 Quantifying soil organic carbon in a sub-field agricultural setting in Southern Ontario using multispectral UAV imagery ......................................................................................................................... 25
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Abstract</td>
<td>25</td>
</tr>
<tr>
<td>3.2</td>
<td>Introduction</td>
<td>26</td>
</tr>
<tr>
<td>3.3</td>
<td>Methods</td>
<td>29</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Study Area Description</td>
<td>29</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Soil Sample Collection, Processing and Measurements</td>
<td>31</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Methodology and Materials</td>
<td>33</td>
</tr>
<tr>
<td>3.3.4</td>
<td>UAV Image Collection</td>
<td>33</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Laboratory Spectral Measurements</td>
<td>41</td>
</tr>
<tr>
<td>3.3.6</td>
<td>Model Development</td>
<td>46</td>
</tr>
<tr>
<td>3.3.7</td>
<td>Statistical Analysis</td>
<td>51</td>
</tr>
<tr>
<td>3.4</td>
<td>Results</td>
<td>53</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Overview</td>
<td>53</td>
</tr>
<tr>
<td>3.4.2</td>
<td>UAV Models</td>
<td>54</td>
</tr>
<tr>
<td>3.4.3</td>
<td>ASD FieldSpec3 Models</td>
<td>62</td>
</tr>
<tr>
<td>3.4.4</td>
<td>USDA RaCA Models</td>
<td>68</td>
</tr>
<tr>
<td>3.4.5</td>
<td>Integration of Laboratory Models with UAV Datasets</td>
<td>70</td>
</tr>
<tr>
<td>3.4.6</td>
<td>Discussion</td>
<td>79</td>
</tr>
<tr>
<td>3.5</td>
<td>Conclusion</td>
<td>82</td>
</tr>
<tr>
<td>4</td>
<td>Summary and Conclusion</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>87</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 3.1 Number of samples associated with each study site.................................................. 32
Table 3.2: General SOC percentage statistics from each study site. ........................................ 32
Table 3.3: Tetracam and MicaSense sensor specifications.......................................................... 34
Table 3.4: The theoretical limits of SOC estimation using field measured reflectance values from
the UAV imagery for the MicaSense and Tetracam sensors...................................................... 56
Table 3.5: The theoretical limits of SOC estimation using field measured reflectance values from
the ASD FieldSpec3 within the MicaSense and Tetracam sensor wavelength ranges............ 63
Table 3.6: SOC estimation using field measured reflectance values from the USDA RaCA
dataset within the MicaSense and Tetracam sensor wavelength ranges................................. 69
Table 3.7: R² and RMSE values for the models created using ASD FieldSpec3 band ratio models
and spectral indices derived from UAV reflectance data for each sensor. The equations that
were used are included to demonstrate the most important band ratios for this analysis..... 72
Table 3.8: R² and RMSE values for the models created using the USDA band ratio models and
spectral indices derived from UAV reflectance data for each sensor. The resulting model
equations are also included to demonstrate band ratio selection............................................. 74
LIST OF FIGURES

Figure 3.1: An overview of the Elora study sites (A) and the Woodrill study site (B) within the Lake Erie Basin (C). .......................................................... 30

Figure 3.2: Soil sampling and drone image extraction points for each study area overlayed on a DEM. The study areas are as follows: MicaSense Smeltzer extraction (A), Tetracam Smeltzer extraction (B), Elora long term trial plots (C), Elora site 1 (D), and Elora site 2 (E). ................................................................. 40

Figure 3.3: A flowchart outlining the UAV methodology used in this study. .............................. 41

Figure 3.4: Digital elevation models and soil sampling points for all study areas. The study areas are as follows: Smeltzer (A), Elora long term trials (B), Elora site 1 (C), Elora site 2 (D). 43

Figure 3.5: A flowchart outlining the ASD FieldSpec 3 and USDA RaCA methodology used in this study. ................................................................. 46

Figure 3.6 (A-H): Graphs illustrating the predicted and observed SOC for each UAV model type, grouped by study site. The abbreviations are as follows: MicaSense (MS), Tetracam (TTC), reflectance model (Reflectance), reflectance and topography model (Reflectance Topo), band ratio model (BR), and band ratio and topography model (BR Topo). .......................... 62

Figure 3.7 (A-F): Graphs illustrating the predicted and observed SOC for each ASD FieldSpec 3 model type, grouped by study site. The abbreviations are as follows: MicaSense (MS), Tetracam (TTC), reflectance model (Reflectance), reflectance and topography model (Reflectance Topo), band ratio model (BR), and band ratio and topography model (BR Topo). ................................................................. 68

Figure 3.8: Graph displaying the USDA RaCA and Tetracam SOC predicted and SOC observed values produced from the models. The abbreviations for the models are as follows: Tetracam (TTC), reflectance models (Reflectance) and band ratio models (BR). ........................... 70

Figure 3.9 (A-C): Graphs displaying the ASD FieldSpec 3 SOC predicted and SOC observed with MicaSense and Tetracam models using UAV data. The abbreviations for the models are as follows: MicaSense (MS), Tetracam (TTC), band ratio models (BR) and band ratio and topography models (BR Topo). ................................................................. 74

Figure 3.10 (A-F): Graphs displaying the USDA RaCA MicaSense and Tetracam model derived SOC predicted and SOC observed using UAV data. The abbreviations for the models are as follows: MicaSense (MS), Tetracam (TTC) and band ratio models (BR). ........................... 78
1 Introduction

The level of soil organic carbon (SOC) in agricultural soils is the key factor regulating soil health as it directly benefits soil physical, chemical and biological properties (Lal, 2004; Brady and Weil, 2008; Keesstra et al., 2016; Srivastava et al., 2019). Additionally, it determines the ability of the soil to function in maintaining plant productivity and the cycling of water and nutrients (Havlin et al., 1990; Rawls et al., 2003). Management of agricultural lands poses many challenges, as the soils are highly variable. The spatial variability of SOC can be attributed to many factors, such as topography, regional climatic variables and anthropogenic land use practices (Batjes, 1996; Jandl et al., 2014; Lal, 2011; Weissert et al., 2016). These factors each have differing effects on the soil to varying degrees. Furthermore, topographic variables arguably have the most prominent effect on the spatial distribution of SOC because the carbon-rich topsoil will tend to accumulate at the trough of a hill (Dialynas et al., 2016; Sun et al., 2015). Climatic variables, such as precipitation and temperature also have an effect on soil carbon dynamics. They can influence the promotion or depletion of SOC depending on seasonal cycles (Luo et al., 2017). Lastly, anthropogenic land use management have a major effect on SOC. Cropping systems, which include crop rotations, crop residue management, and cultivation, have a direct impact on the field-scale variability of SOC (Gebhart et al., 1994; Kong et al., 2005; Varvel and Wilhelm, 2010; Xiong et al., 2016) While it may take years to exhibit a significant change in SOC percentages, the complex relationships that exist between SOC and its controlling variables hold vital evidence to help understand spatial variability trends.

Currently, SOC content is being obtained by collecting in situ soil samples or soil cores from the field and bringing them back to the lab for further analysis. These conventional
methodologies for assessing SOC content are time-consuming and costly (Tabatabai and Bremner, 1970; Allison 1960; Wotherspoon et al., 2015). Therefore, there is a need for more accurate, rapid and cost-effective methods. The recent use of multispectral sensors on drone-based platforms (also known as Unmanned Arial Vehicles, UAV, or Unmanned Aerial Systems, UAS) has introduced new opportunities for providing detailed spatially explicit spectral information of several soil properties, including SOC concentration (Odeh et al., 1995; López-Granados et al., 2005; Kumar, 2015; Laamrani et al., 2019). In addition, prediction models fitted with Geographic Information System (GIS) and medium resolution remote sensing-based data have provided key benefits to understanding the relationships that occur between SOC concentrations and the numerous factors that must be incorporated to produce accurate results (Zhi et al., 2014; Kumar, 2013; Kumar, 2015).

Stepwise multivariate linear regression (SMLR) techniques are commonly used for SOC prediction using remotely sensed data (Selige et al., 2006; Wu et al., 2009; Bhunia et. al., 2018). Although this technique has been mostly applied to varied remote sensing data (i.e., satellite, airborne and laboratory spectroscopy), Vasques et al., 2008 and Bhunia et al., 2018 found that reliable prediction of SOC content can be achieved using UAV-derived reflectance. The SMLR is advantageous to use because it selects the best linear regression model according to input parameter significance (p<0.05) and returns a formula that can be used for estimation (Hummel et al., 2001; Vasques et al., 2008; Vohland et al., 2011; Lu et al., 2013). In this study, the SMLR model will be implemented to evaluate its effectiveness to estimate SOC from UAV-mounted multispectral sensors.
The purpose of this research is to evaluate the potential of two commercially available UAV-mounted multispectral sensors for predicting SOC concentrations in an agricultural setting. To this end, multispectral UAV images taken at agricultural fields near Guelph, Ontario, Canada were collected and used with field measured SOC content to develop a prediction model. Specific objectives of this study were (1) to conduct a literature review regarding the academic research relevant to SOC and its measurement; (2) to estimate SOC concentration using SMLR analysis with multispectral data collected from UAV; and (3) to assess the accuracy of laboratory models inputted with UAV reflectance data for SOC estimation.
2 Literature Review

2.1 Introduction

This chapter contains a review of current literature regarding the importance of SOC and agriculture, the various established methods that are used to quantify and model SOC. Firstly, section 2.2 will discuss the significance of SOC and its association to landscape. This will be followed by a review of how agricultural practices interact with SOC accumulation or release. Finally, this chapter will highlight the current SOC observation, quantification and modelling methods.

2.2 Significance of SOC for Agricultural Systems

SOC is a constituent of soil organic matter (SOM) and is comprised of the carbon components found within SOM (Brady and Weil, 2008; Lehmann et al., 2008; Lehmann and Kleber, 2015). SOC is essential to stabilize the long-term volatility of agricultural soil and soil health (Congreves et al., 2014). Organic carbon can be found within the entire soil profile, but the highest concentrations are found within the O and A horizons. The O horizon is generally present in forested regions because the organic matter is made up of undecomposed plant and animal byproducts. The A horizon can be found across all soils and is the topmost mineral horizon. The A horizon contains completely or partially decomposed, humified products, giving the soil a darker color than other horizons. The microbial biodiversity that is present in the A horizon can also improve soil productivity, aggregate stability and overall soil health (McLaughlin and Mineau, 1995; Brady and Weil, 2008; Martin, 2008; Keesstra et al., 2016). This microbial biodiversity has a direct impact on decomposition because microbial community concentration within the soil affect decay rates (Drijber et al., 2000; Beyaert and Voroney, 2011).
SOC can also reach the soil through plant roots, as carbon flows into the rhizosphere through root exudates. These exudates tend to be simple compounds that are created into carbon by soil microbes (Kallenbach et al., 2016; Poirier et al., 2017).

Within a soil, higher levels of SOC can improve soil health through nutrient and water retention, as well as enhancing biotic activity (Lal, 2004). Increased levels in SOC causes a higher presence of organic binding agents, improving aggregate stability and creating a more desirable soil structure (Six et al., 2000; Ashman et al., 2003; Srivastava et al., 2019). This promotion in soil structure typically increases the soil water retention potential and can be explained because SOC concentration is indicative of surface soil structure and texture (Bauer and Black, 1992; Hudson, 1994; Rawls et al., 2003; Manns and Berg, 2014). Manns and Berg (2014) suggest that a linear relationship exists between SOC and soil water content and show that particle size has a direct link to SOC and soil water retention potential. They conclude that SOC is an essential predictor of soil water content and could be used in future studies to accurately explain more variance than with normal model predictor variables (Manns and Berg, 2014). However, studies demonstrate that all soils experience some positive change and fine-textured soils (sandy and silty soils) exhibit the largest increase in soil water retention when SOC concentration increases (Rawls et al., 2003; Minasny and McBratney, 2017; Kallenbach et al., 2019). This can have a major impact on agricultural regions, since agricultural practices are highly dependent on soil water capacity and availability, especially in drought-prone areas.

Certain management practices, such as irrigation, could be significantly affected by an increase in soil water capacity (Chartzoulakis and Bertaki, 2015).
SOC in southern Ontario is typically controlled by the dominant Ontario field crops (corn and soybean) and some combination of the various land management practices that are present (i.e., tillage practices and fertilizer application) (Yang and Kay, 2001b; VandenBygaart and Kay, 2004; Congreves et al., 2014). It has been estimated that 44% of Ontario’s agricultural land are prone to soil erosion and soil degradation (Ontario, 2009). This requires land management practices in Ontario, especially southern Ontario where most agriculture occurs, to be carefully considered to prevent both soil erosion and degradation as SOC is vital to sustain stable crop yields (Congreves et al., 2014). Decreases of 20 to 25% SOC have been documented in short-term studies (2-5 years), but 6 to 10-year study periods were required to detect SOC changes using a 90% confidence interval (Smith, 2004). Long-term field experiments in Ontario have been conducted to understand how land management practices affect SOC concentrations (VandenBygaart et al., 2003; Deen and Katak, 2003; Meyer-Aurich et al., 2006; Van Eerd et al., 2014). These studies have demonstrated that land management practices, such as no-till vs. conventional tillage, crop rotations vs. continuous cropping, or nitrogen fertilizer systems vs. no nitrogen fertilizer application, can all have drastic effects on SOC at plough depth and depths below the plough layer (Gregorich et al., 1996; Wanniarachchi et al., 1999; Yang and Kay, 2001a; Yang and Kay, 2001b; VandenBygaart et al., 2004; Congreves et al., 2014). These effects include increases and decreases of SOC depending on the land management practices that are being analyzed. For example, complex crop rotations are beneficial to increase SOC when compared to continuous cropping because they allow residue to accumulate and decompose, which creates an increase in SOC (Studdert and Echeverr, 2000).
2.3 Effects of Agricultural Management Practices on SOC Levels

Soils have the potential to sequester 0.4-1.2 Gt C/year of atmospheric CO$_2$, but this number is significantly reduced when cropping systems and recommended agricultural land use management practices are ignored (Lal, 2004; Angelopoulou et al., 2019). It is well known that land management practices (i.e., crop rotations, tillage practices, or fertilizer applications) can impact the SOC concentrations. For instance, manure applications, crop residues and biomass can add essential nutrients to the soil, which promote higher levels of SOC (Johns et al., 2015; Bader et al., 2016). Conversely, poor tillage practices and a low diversity of crops used in cropping rotations can lead to a decrease in SOC (Congreves et al., 2014).

2.3.1 Cropping Systems

SOC concentration levels can indicate the effect that certain agricultural cropping systems have on the soil. The three main parameters that depict the largest effect on soil organic carbon concentrations are: crop rotations, residue cover, and tillage practices. SOC can be managed the most accurately when these three parameters work in unison with one another.

There are many different crop rotations when determining which one would provide the largest improvement to SOC. Through cropping rotations, it is possible to manipulate the opportunity that exists to reduce the SOC loss that will inherently happen with the uptake of nutrients by plants (Campbell, 1978). The main advantage that different cropping rotations have on SOC concentration is the varying opportunity that each different plant possesses to sequester atmospheric C and return it to the soil (Witt et al., 2000; West and Post, 2002). There are many studies that show that perennial pastures or crop rotations that include cover crops are advantageous for increasing SOC when compared to annual crop rotations (Dalal et al., 1995;
Franzluebbers et al., 2000; Gentile et al., 2005). Studdert et al. (1997) showed that under regular cropping rotations, SOC decreased by 4.4 g kg$^{-1}$ in 6 to 7 years and under pasture was able to replenish in 3 to 4 years. Complex cropping rotations also influence SOC dynamics. For example, the introduction of various summer crops can have a major impact on SOC levels when compared to just using one summer crop (Studdert and Echeverr, 2000). Research that was conducted in southern Ontario displayed similar results when comparing continuous and complex crop rotations. Alfalfa-corn crop rotations had higher SOC concentrations when compared to continuous corn rotations, including at depths beyond the plough layer (Gregorich et al., 2001). Meyer-Aurich et al. (2006) further supported this finding with SOC storage being largest using alfalfa-corn and soybean-corn rotations.

Crop residue management is an essential factor on SOC concentration. A data mining study of literature (n=660) illustrated that the removal of crop residues in temperate areas resulted in an average SOC loss of 12% (Raffa et al., 2015). Large biomass crops, such as corn or sorghum, have proven to significantly increase the SOC present at the surficial layers of the soil because these crops produce a large amount of residue cover that decomposes into organic carbon that the soil eventually absorbs (Barzagli and Mani, 1979; Havlin et al., 1990). Furthermore, corn residue has established a larger increase in SOC when compared to corn-legume mixed residue (Chen et al., 2018). Research conducted in Elora, Ontario suggests that this is likely attributed to the above and below-ground biomass that accumulates with such a high residue crop such as corn, as well as the lower C:N ratio of legumes compared to corn rotations (Meyer-Aurich et al., 2006). Although surface residue cover has a large effect on the SOC content, root structures are also capable of exuding a large amount of organic carbon back into the soil. Corn roots have been shown to release 37% of their organic carbon back into the soil,
contributing to the maintenance of SOC in the soil, more considerably under no-till conditions (Barber, 1979). Research also suggests that this root biomass can influence SOC concentrations at depths below the plough layer (>45 cm) in Ontario (Johnson et al., 2006).

Cultivation can directly impact the change that SOC concentration experiences within an agricultural soil (Mann, 1986). Cultivation, or tillage, inherently disrupts the soil, ultimately leading to a rapid mineralization of soil organic carbon and releasing it into the atmosphere in the form of CO₂ (Bader et al., 2016; Schlesinger and Andrews, 2000; Alvarez et al., 1995). A 3-year study that examined no-till and various types of conventional tillage (i.e., strip-till, deep rip, chisel plow, and moldboard plow), also found a positive correlation between no-till methods and SOC at the 0-10 cm portion of the soil profile. The total SOC was found to be 32% greater at the 0-5 cm level and 41% greater at the 5-10 cm level when compared to the other tillage techniques (Al-Kaisi and Yin, 2005). This drastic increase is not present at all depths of the soil profile, but the SOC contents are higher when using no-till methods and a continuous high biomass crop (Havlin et al., 1990). For example, the accumulation of crop debris within the first 5 cm of a no-till system has been proven to increase SOC levels by 42-50% of their previous state (Alvarez et al., 1995). Long-term no-till practices in southern Ontario illustrated higher SOC levels in the surface layer (0-10 cm) when compared to conventional tillage practices, with 41 g C m⁻² yr⁻¹ more SOC at shallow soil depths (Angers et al., 1997; VandenBygaart et al., 2002; Halpern et al., 2010; VandenBygaart et al., 2011). A review of Ontario literature concluded that no-till systems distribute SOC within the plough layer (<20 cm) but did not improve SOC levels beyond the plough layer (Deen and Katakí, 2003; Angers and Eriksen-Hamel, 2008; Yang et al., 2008; Shi et al., 2011). At depths beyond the plough layer (>20 cm), Ontario studies suggest that the climate dictate the rates of decomposition after the soil has been subject to cultivation. Results indicate
that at depths greater than 20 cm, conventional tillage is superior to no-till systems to distribute SOC (Angers et al., 1997; Angers and Eriksen-Hamel, 2008).

Microbial activity can be considerably affected by cropping systems and they have the ability to modify the soil environment. This can lead to a change in the soil microbial biomass, which can have a large affect the SOC nutrient availability of plants (Ghidey and Alberts, 1993; Kaiser and Heinemeyer, 1993; Kong et al., 2011). Crop rotations, tillage practices, and fertilization application regulate microbial biomass populations, which catalyzes certain processes such as nutrient cycling and decomposition rates, increasing the organic carbon turnover rate (Beiderbeck et al., 1984). Since the microbial biomass can increase SOC dynamics, it is important to instill beneficial agricultural practices into cropping systems. No-till practices have been found to increase the microbial biomass by 57% when compared to soil that does not have any residues present (Collins et al., 1992). Franzluebbers et al. (1995) used the chloroform fumigation-incubation method, first proposed by Jenkinson and Powlson in 1976, to accurately extract microbial biomass C from various plots that included continuous and rotated sorghum, as well as conventional and no-till methods at depths ranging from 0-200 mm. The results showed that under no-till conditions, the microbial biomass C was 65%, 22%, and 9% greater at 0-50mm, 50-125 mm, and 125-200 mm depths, respectively. Also, crop rotations stimulated the microbial biomass C by an average of 31.3% across the 0-200mm depth range (Franzluebbers et al., 1995).

2.3.2 Quantification of Spatial Variability of SOC

Determining the spatial variability of SOC has shown to be costly and time consuming because it largely relies on gathering high resolution in situ data that needs to be collected from
field-scale sampling (Conant and Paustian, 2002). The lateral distribution of SOC across the surface and the vertical distribution of SOC throughout the soil profile are inherently impacted by the land management practices that are implemented on any given soil (VandenBygaart, 2006; Pennock et al., 2007). Accurate predictions are often difficult to attain because of the various terrain attributes and agricultural practices that contribute to the soil landscape from region to region (Huang et al., 2007a; Mishra et al., 2009). Topographic variables must be considered because they can play a role in the spatial distribution of SOC. For instance, hilly areas with large slopes have the potential for more soil erosion to occur, which causes the SOC to oxidize and deplete (Smith et al., 2001; Dialynas et al., 2016). These topographic variables can be incorporated into prediction models to improve accuracy and uncertainties within the model (Zhi et al., 2014; Kumar, 2015). While elevation, slope, profile curvature and plan curvature are usually incorporated into topographic models, elevation is consistently the most statistically significant (p = 0.01) variable to use (Oueslati et al., 2013). Alternatively, agricultural practices (i.e., no-till and conventional till) have a profound effect on both the lateral and vertical distribution of SOC (Shi et al., 2012; VandenBygaart et al., 2007). VandenBygaart et al. (2007) determined that the vertical distribution of SOC in soil cores taken from Saskatchewan, Quebec and Ontario could be attributed to tillage and agricultural land management practices. Additionally, research was conducted in southern Ontario to examine the relationship between no-till, ridge till, and moldboard plough tillage and SOC. Results illustrated that no-till systems had much greater lateral distribution of SOC at surface and subsurface soil layers (0-5cm; 5-20cm). However, ridge tillage accounted for the spatial variation of SOC at depth (25-60cm) because crop residues were incorporated into the soil profile and decomposition could occur more readily, promoting SOC increase (Shi et al., 2012; Yao et al., 2019). Geostatistical
methods, multiple linear regressions (MLR), and regression kriging (RK) approaches have been explored to determine their effectiveness to predict spatial variation in spatial distribution of SOC (Kumar, 2015; Wang et al., 2018). Geostatistical kriging methods produced the best results from a variety of different techniques that were used on various datasets. Specifically, ordinary kriging (OK) and geographically weighted regression kriging (GWRK) reported the lowest RMSE and the highest $R^2$ values, especially when predicting SOC variability at depth (Kumar, 2015; Liu et al., 2015; Bhunia et al., 2018).

2.4 Current and Emerging Analytical Methods Used for Measuring SOC

Traditional established methods of SOC quantification tend to be in a controlled, laboratory setting. These methods include wet digestion, dry combustion analysis and laboratory spectral analysis of *in situ* soil samples. However, proximal sensors that attach to agricultural machinery to gather real-time SOC measurements are currently being developed. These sensors are still in the preliminary stages of development and require extensive research to be considered a reliable source of SOC quantification in the future (Nayak et al., 2019). The following section will outline three main types of commonly methods currently being used for SOC measurement.

2.4.1 Dry Combustion

The LECO carbon analyzer is a satisfactory dry combustion, or loss-on-ignition (LOI), method of measuring the total carbon that is found in soil samples retrieved from the field. This analysis is performed in a laboratory where a soil sample is burned at a high-temperature (i.e., 1,300°C) in an induction furnace (Aula et al., 2019; Graham et al., 2019). The carbon dioxide that is released from the sample provides an automatic carbon quantification of the sample by calculating the percentage of carbon using before and after weights. For the LECO method to be
properly performed, the soil must be subject to dry combustion, rather than wet combustion (Tabatabai and Bremner, 1970). Allison (1960) explored the potential of wet combustion when trying to extract organic and inorganic C from the soils. However, this study has been contrasted to dry combustion studies, proving that dry combustion is the better method because the range and standard deviation of estimated percent C is much smaller (Tabatabai and Bremner, 1970).

The disadvantage of using the LECO Carbon Analyzer is the extensive pre-processing of soil samples. When using this technique all inorganic carbon must be removed from the soil, leaving the desired organic carbon (Wotherspoon et al., 2015). This method exposes the soil to a pre-determined amount of sulfurous acid, which evaporates easily, thus causing minimal oxidation of organic carbon (Nelson and Sommers, 1996; Bremner, 1949). The results of exposure to sulfurous acid digestion indicate that the SOC contents that are extracted have a higher precision compared to samples that still contain soil inorganic carbon (Wotherspoon et al., 2015). Also, the soils must be either air or oven dried and passed through a mesh sieve to extract any materials that may have been present during soil sampling (Wang and Anderson, 1998). The processing that all samples must undergo makes this method time consuming, which is not desirable.

2.4.2 Wet Digestion

Wet digestion originated from the Walkley-Black technique, where SOC is oxidized using potassium dichromate, and ferrous sulfate is added as a titrant to react with the remaining dichromate. The organic carbon content is then estimated from the volume of dichromate that is consumed during the reaction (Walkley and Black, 1934; Chen et al., 2015). This is a widely accepted SOC extraction process because it is a rapid, simple way to extract SOC with minimal
equipment (Grewal et al., 1991). However, this method has a major disadvantage because only active SOC is oxidized, thus a correction factor of 1.33 is commonly used to account for the unoxidized SOC (Walkley and Black, 1934; Nelson and Sommers, 1996). A newer form of wet digestion consists of a mixture of dichromate and sulfurous acid being added to a sample, which oxidizes the carbon in the solution without adding any heat (Schwartz, 1995). This method can also be used in conjunction with heat to accelerate the reaction (Rosell et al., 2001).

Although wet digestion methods can achieve reasonably accurate results in a short amount of time, there are many drawbacks to using these techniques. The type and properties of soil can vary results widely because of the oxidation process that is involved (Neal and Younglove, 1993). Possibly the most contributing disadvantage is the correction factor that must be used in the Walkley-Black method. This factor can be affected by soil type, texture, depth, horizon and SOC concentration, which will in turn affect the accuracy of the resulting SOC estimation (Kimble et al., 2001; Rossel et al., 2006a; De Vos et al., 2007).

2.4.3 Laboratory Spectroscopy

In recent years, soil spectroscopy (visible, near- and/or mid-infrared) has proven to be a useful analytical tool for rapid and precise estimation of SOC (Katuwal et al., 2018). Spectroscopy concentrates on the electromagnetic radiation that reflects from the surface of the soil, thus creating a spectrum from the reflectance or absorbance, using wavelength (nm), as a spectral indicator (Hunt, 1977; Brown et al., 2006). These spectra, or spectral signatures, provide qualitative and quantitative data on the soil properties that are being analyzed (Nocita et al., 2015). Spectroscopy can be conducted in the laboratory or using multispectral and/or hyperspectral sensors mounted on aircraft, UAVs, or satellite platforms to measure surface SOC.
Laboratory spectroscopy has generally been the most accepted method, due to the vast quantity of research that has been performed on it in the past (Dalal et al., 1986; Lu et al., 2013; Doetterl et al., 2013; Laamrani et al., 2019). Since all parameters can be optimized in a laboratory setting, the quantification of SOC using multispectral and hyperspectral spectroscopy has exhibited positive results (Dalal and Henry, 1986; Nocita et al., 2013; Conforti et al., 2015). The use of laboratory near infrared (NIR) spectroscopy has seen a large increase in research in the last 10-15 years (Malley et al., 2004). This technology is experiencing an increase in the agricultural sector because of its potential to detect soil properties (Malley et al., 2004). The use of visible near infrared (VIS-NIR) spectroscopy can reduce error and computation time when determining SOC, making this research highly advantageous in the agricultural sector (Morra et al., 1991, Angelopoulou et al., 2019). VIS-NIR spectroscopy has mainly been employed to identify certain soil properties, but in the last 20 years, there has been an increased desire to study the estimation of SOC specifically (Sudduth and Hummel, 1991; Chang et al., 2001; Nocita et al., 2013; Stevens et al., 2013). Standard practice in the laboratory is to first chemically determine the SOC using the Walkley-Black method to produce observed SOC values from soil samples (Walkley and Black, 1934; Chen et al., 2015). Then, under a controlled laboratory environment, the desired spectrometer is used to determine the SOC content. The VIS-NIR spectrometer is then mounted on a tripod at a specific angle and distance away from the soil sample to ensure that the field of view can be accurately calculated (Lucà et al., 2015). Illuminated halogen lamps (1,000W) in the nadir position are used as the reflectance light source (Nocita et al., 2011). Parameters in the field, such as soil moisture, stoniness, and shadows, can be controlled in a laboratory setting. Therefore, the signal to noise ratio (SNR) can be significantly reduced by performing laboratory spectroscopy and uncontrollable field soil conditions can be regulated for
the best results (Stevens et al., 2008). The NIR (400-2,500 nm) and mid infrared (MIR) (2,500-
25,000 nm) range of the electromagnetic spectrum can be highly correlated with the
measurement of soil carbon (McCarty et al., 2002). Although MIR has the potential to measure
SOC at a higher degree of accuracy, it is costly to acquire a camera that senses in this portion of
the electromagnetic spectrum (Rossel et al., 2006b). However, the NIR absorption coefficients
are much less than the MIR, which is superior because it enables light to have better penetration
into the soil parameters that are being examined (Brown et al., 2006; Bellon-Maurel and
McBratney, 2011). Previous studies that have focused on SOC spectroscopy have shown certain
regions within the VIS-NIR spectrum to be more sensitive to the presence of SOC based on
reflectance and absorbance factors within the spectra (Aldana-Jague et al., 2016; Cambou et al.,
2016; Crucil et al., 2019). This sensitivity, or change in spectral signature, can be attributed to
the vibrations that the molecule experiences at the atomic level. (Mohamed et al.,
2018; Angelopoulou et al., 2019). These changes in the spectral signature can be specifically
found within the 0.5-1.0 nm region, with peaks and troughs distinctly portraying soil
characteristics (Shepherd and Walsh, 2002; Islam et al., 2003). A recent study performed by
Laamrani et al (2019) supports this finding and shows that ten specific spectral bands throughout
the NIR, ranging from 433-998 nm, were the most important for this spectral region.
Furthermore, laboratory spectroscopy is considered a well-established method to estimate SOC,
with the error values being similar to those seen when using conventional laboratory analyses
(Stevens et al., 2006; Doetterl et al., 2013; Nocita et al., 2015; Aldana-Jague et al., 2016)

2.4.4 Field Spectroscopy

Portable spectroscopy allows the user to record data very quickly and efficiently, while
preserving the soil that is present. If the sensor is calibrated properly and the images are taken
under the proper lighting and soil conditions, the proximal sensor may have a similar effect to laboratory spectroscopy. There are different types of proximal sensors that may be used for in-situ measurements of SOC. Some examples include: tractor mounted sensors, off nadir soil sensors, and soil profile attachments; the last two are referred to as spectroradiometers. An example of a spectroradiometer include the ASD FieldSpec, which is a high resolution, hand-held, VIS-NIR device that uses fiberoptic cables, calibrations, and reflectance values to create a spectral signature graph that can be used to identify certain criteria, including soil properties. Each sensor measures in a certain range of the electromagnetic spectrum, which causes the accuracy to increase or decrease depending on the size of the range (Croft et al., 2012).

Although these sensors can be placed in optimized areas on the soil, it is impossible to fully neutralize all of the variables that negatively influence the sensor result. Surface roughness causes microshading to occur, which skews the measured reflectance. However, this accuracy reduction is seemingly non-linear and can be removed to preserve the shape of the spectrum (Stevens et al., 2008). Fiber optic VIS and NIR spectrometers are available for purchase and can be an effective on-the-go proximal soil sensor that measures the mean SOC content at a field scale (Rossel et al., 2006). An example of a proximal sensor would be the SmartFirmer released by Precision Planting. It offers a sensor that adjusts planting depths based on real-time organic carbon measurements. The data is stored within the system and can be inputted into Soil Information Systems (SIS), where carbon maps are created to show the spatial variability of SOC at the field-scale. The error associated with the SOC estimations using these types of sensors are directly dependent on the conditions that the sensor is subject to. The least amount of error was found when a sensor was integrated into a chisel plough, allowing the fiber optic spectrometer to become homogenous with the soil, rather than measuring the soil surface (Mouazen et al.,
If NIR sensors were to be taken into the field, they are a reasonable replacement for conventional soil sampling techniques because of their trade-off between low cost and accuracy. However, these methodologies are still in their infancy and further testing must be performed for this to be a practical method. Firstly, bias must be reduced by making the techniques reproducible, which would lead to the accuracy being increased. Secondly, estimation equations of the volumetric C found within the soil must be developed to keep costs low, since NIR sensors only detect C on the surface soil. Lastly, improvements must be made to estimate the root C that is found within the soil, allowing the SOC approximations from root decomposition over time to be incorporated into the estimation model (Bellon-Maurel and McBratney, 2011).

2.5 UAV Applications in SOC Mapping

UAVs are an affordable, rapid, non-destructive method that have the potential to assist in the quantification of the spatial and temporal variability of SOC (Zhang and Kovacs, 2012). UAVs have the ability to collect detailed, field-scale imagery, along with analysis in a matter of hours, depending on the sophistication of the system (Angelopoulou et al., 2019). VIS-NIR sensors mounted on UAVs have become increasingly popular in recent years because of the high image resolution that these sensors produce (Aldana-Jague et al., 2016). These sensors use image reflectance values to estimate numerous soil properties (Bartholomeus et al., 2011). The reflectance properties of soils are usually dependent on investigated soil parameters. For example, an increase in SOC results in a decrease in the reflectance in the VIS-NIR spectra (Baumgardner et al., 1985). Numerous studies have been conducted on the optical remote detection of SOC in the VIS-NIR and MID-IR regions of the electromagnetic spectrum.
However, research tends to primarily focus on the VIS-NIR multispectral portion due to sensor accessibility (Rossel et al., 2006b).

There are many challenges and limitations associated with the optical remote sensing of SOC from UAV platforms (Nayak et al., 2019). Technological challenges arise from the UAV system that is being used, as well as the software that is used to process the images. The image acquisition is limited to the duration of the battery that is being used on the UAV. This can be attributed to flying height or the payload that is present on the UAV. When adjusting these parameters, the amount of spatial coverage that can be attained per battery will either increase or decrease. Also, the images that are acquired take a significant amount of time to process because they must be inputted into an orthomosaicking software to create usable reflectance maps (Aldana-Jague et al., 2016).

Some limitations need to be dealt with prior to using UAV imagery for SOC estimation and mapping. For instance, there are a variety of field-based and atmospheric parameters that need to be corrected when trying to produce usable SOC results from images taken using the optical remote sensing. Past laboratory research shows that soils with a soil organic matter (SOM) content less than or equal to 2% result in poor reflectance values, which makes analysis very challenging, especially when trying to quantify it from UAV imagery (Croft et al., 2012). Optimal flight parameters, (i.e., sun illumination at nadir, no cloud cover, etc.) must also be present for remote sensing to produce usable spectral information. These optimal parameters also include bare soil being present, since vegetation cover will dramatically affect the usability of the imagery. Additionally, image acquisition must be performed during a period of low soil moisture, since soil moisture can greatly affect the reflectance values (Aldana-Jague et al., 2016;
Crucil et al., 2019). Although there are a variety of factors present in all agricultural soils that inherently reduce the accuracy of the statistical prediction methods that are being applied, correction techniques can be used to lessen the effect that these factors have on the results.

Since the target of multispectral UAV imaging is strictly limited to the surface soil, one of the most important parameters that determines SNR is soil moisture. Soil moisture is an issue as it directly impacts the albedo of the soil. A film is produced around wetter soil aggregates, which causes an ancillary reflection, ultimately leading to a larger amount of light to penetrate deeper into the soils. This is closely linked to the atmospheric attenuation and unstable illumination conditions, specifically illumination angles, which can be present during flight (Nocita et al., 2013). Surface roughness is similar to soil moisture in that it causes light scattering, leading to a decrease in reflectance values (Rodionov et al., 2014). Soil moisture and surface roughness are intricate problems as they differ spatially and temporally depending on the soil that is being studied (Rodionov et al., 2014).

Macrostructure and stoniness of the soils, which are non-existent or controlled in a laboratory environment, coupled with vegetation and residue cover are necessary parameters to include in the interpretation of the data (Cambou et al., 2016). The limitation induced by the presence of soil cover and residues is a large issue when trying to identify SOC with UAV imagery. Without a viable means to correct the overestimation of SOC that is inherently introduced through surface cover, an accurate SOC estimation can only be measured using bare soil. However, it is possible to use supervised classification methods to differentiate between soil cover and bare soil in UAV imagery. This would allow VIS-NIR imaging of agricultural fields
without to the presence of bare soil to be usable for SOC estimation and a more practical, universal SOC prediction method could be created (Rodionov et al., 2014).

2.6 SOC Modelling

2.6.1 Current Models for SOC Prediction

Once the spectral measurements are taken from the SOC samples, a variety of calibration and validation techniques can be applied to create prediction models. These models are appropriate to use for data from laboratory, proximal, and remote sensing collection methods. However, most previous studies focus mainly on laboratory and proximal spectroscopic models (Mouazen et al., 2007; Gomez et al., 2008; Aldana-Jague et al., 2016; Crucil et al., 2019; and Laamrani et al., 2019). This can be explained by the uncertainty surrounding the estimation of SOC using remote sensing techniques. The four most pertinent models currently found in literature used to estimate SOC are: random forest (RF), support vector machine (SVM), partial least square regression (PLSR), and cubist (CB) (Stevens et al., 2006; Stevens et al., 2010; Stevens et al., 2013; Aldana-Jague et al., 2016; and Laamrani et al., 2019).

Although all four methods are recognized in literature, the PLSR technique is by far the most common algorithm that is used for soil spectroscopy applications (Stevens et al., 2012). The PLSR model is relevant to soil spectroscopy because of its ability to construct predictive models with large amounts of factors that are highly collinear. The factors that most commonly include high multicollinearity are all spectral bands that are usually included with this prediction model type. This allows the model to correct for over-fitting and the problem of multicollinearity (Wold et al., 2001 and Tobias, 2003). Multicollinearity within variables is a major problem, especially in the field of spectroscopy, because most bands within the same regions will have
very similar reflectance values. If the algorithm can automatically detect these bands and adjust the equations accordingly, then the statistical power of the model could potentially be improved. The PLSR is also able to account for variations within soil variables (i.e., soil type, soil texture) and spectral variables (i.e., individual bands from hyperspectral imagery), depending on the calibration of the model. These variables tend to change with the in-field conditions (illumination and surface) when sampling, along with differences within the soil (type, texture, etc.) (Chang et al., 2001; Stevens et al., 2008; Kelcey and Lucieer, 2012; and Nocita et al., 2013).

The other three algorithms for SOC estimation have been used much less in spectroscopic literature. The CB and RF models fit predictor variables to numerous tree structures, usually 500, which are based off rules that are created by the model. A regression model is then fit to each of the data subsets, which are defined by the model trees. When the CB models were used to predict SOC, Minasny and McBratney (2008) and Stevens et al. (2013) recorded that they surpassed the other competing models in terms of root mean square error (RMSE) and $R^2$ values in numerous studies. According to the paper by Stevens et al. (2013), the SVM regression, which fits data to linear regressions that separate the data into classes, was then ranked as the second-best model, while the RF and PLSR models ranked third and fourth, respectively. Laamrani et al., 2019 found similar results, reporting that the SVM model achieved the best results followed by the RF model, which performed much better than the PLSR model. This was likely due to the robustness of the SVM and RF algorithms when dealing with large volumes of data. Ultimately, the CB and RF models were found to attain the best results (Minasny and McBratney, 2008 and Stevens et al., 2013). Both models account for outliers, multicollinearity, and adjust for variable importance, rerunning the algorithm in all possible combinations (Gromping, 2009; Stevens et al., 2013; Gregorutti et al., 2017; and Sorenson et al., 2017). The CB model also returns the
formula for each multivariate equation based on a set of rules that are used to determine the best possible equation (Minasny and McBratney, 2008 and Sorenson et al., 2017). Although the PLSR is the most commonly used model, the recommendation is to fit specific data to all four approaches and choose the best result.

2.7 Conclusion

SOC levels are essential to quantify because they have a massive impact on soil health, as well as nutrient and water retention. Agricultural land management practices impact SOC concentration, thus beneficial practices must be implemented to retain stable SOC levels. SOC is one of the main factors that influences crop growth and digital SOC estimation maps could improve agricultural management (Lal, 2004). This would occur because these SOC estimation maps could be incorporated into precision agriculture techniques to determine areas of high and low SOC. The areas of low SOC could then be identified and attention could be directed to improving and increasing SOC levels within these areas. However, due to the variable spatial distribution of SOC, these estimation maps are only possible to be produced at a reasonable rate if the development of resilient, efficient and accurate methods are generated (Eswaran et al., 1993; Lal, 2004; Nayak et al., 2019). The challenges that arise when dealing with quantification of SOC are many, as numerous variables must be accounted for to acquire an estimation with an acceptable accuracy. Also, the different methods that are used to quantify SOC, especially UAV sensors, must have their parameters adjusted according to soil and topographical changes that are present within differing spatial regions (Nocita et al., 2011; Jiang et al., 2016; Aldana-Jague et al., 2016; Croft et al., 2012; Chang et al., 2001). These models could include more general linear models for ease of use, such as SMLR, which could account for more variability. UAV-based
remote sensing platforms and their associated models have the potential for accurate SOC estimation. The challenges that surround the remote detection of SOC allow for an opportunity to better understand how these systems work and how they can effectively be related to the prediction of SOC. Increasing the knowledge in this area of research could catalyze the development of more rapid, affordable, and accurate SOC prediction techniques.
Quantifying soil organic carbon in a sub-field agricultural setting in Southern Ontario using multispectral UAV imagery

3.1 Abstract

Spectral remote sensing using the visible near infrared (VIS-NIR) for SOC mapping has been limited due to the requirement of high-resolution multispectral (400-860 nm) imagery. This study evaluates the potential of two commercially available unmanned aerial vehicle (UAV) mounted multispectral sensors for the measurement of field-scale SOC in a region of Southern Ontario. Furthermore, laboratory models using an ASD FieldSpec 3 spectroradiometer (350-2500 nm) and the USDA Rapid Carbon Assessment (RaCA) database were incorporated into the study to determine the theoretical limits that are present under optimal sensing conditions. Multivariate linear regression analysis was used for model creation and variable extraction. The $R^2$ values for the UAV models ranged from 0.02 to 0.93, while the ASD FieldSpec 3 laboratory models exhibited an $R^2$ range of 0.10 to 0.95. Finally, a generalized, transferrable model that can be used across various remote sensing platforms was tested to create an optimized SOC-specific model.
3.2 Introduction

Soil organic carbon (SOC) is a critical component of the soil, affecting key elements within agricultural and climatic processes. SOC constitutes the largest terrestrial carbon pool, both affecting and being affected by agricultural practices (Tiessen et al., 1994 and Schlesinger, 1997). Current knowledge of sequestration potential of our soils could reduce the effects of global warming, but our confidence in this potential is based on large scale modelling, at the biosphere and biome scales (Ballock and Nelson, 2000; Johnston et al., 2009; O’Rourke et al., 2015). Since SOC has complex, variable interactions between soil properties, topographic indices, and decomposition kinetics, these types of models are difficult to achieve at the farm level (O’Rourke et al., 2015). This inaccuracy can also be attributed to the high spatial variability of SOC, along with ever changing land management practices that occur at the global scale (Eswaran, 1993; Angelopulou et al., 2019). To capture an accurate, detailed representation of SOC over scales larger than an agricultural field can be tremendously expensive and labor intensive. Current global SOC prediction maps are limited, while those that do exist are inherently inaccurate (Kochy et al., 2015; Stockmann et al., 2015).

Laboratory estimates of SOC using VIS-NIR spectroscopy has been well documented (Dalal and Henry, 1953; Stoner and Baumgardner, 1981; Ben-Dor and Banin, 1995; Stevens et al., 2008; Vasques et al., 2009; Nocita et al., 2013; Shi et al., 2014; Nocita et al., 2015). Recent studies have illustrated the advances that laboratory spectroscopy has made in accuracy of SOC predictions (Nocita et al., 2015; Allory et al., 2019). The United States Department of Agriculture (USDA) conducted the Rapid Carbon Assessment (RaCA) study, in which 144,833 soil samples were collected from across the United States and reflectance values (350-2500 nm) at 1 nm increments were recorded for each sample (Soil Survey Staff and Loecke, 2016). This
database is of interest for this research because laboratory reflectance values of soils with similar attributes are already produced. Using this data, a generalized laboratory model could be created and used across various platforms (i.e., UAVs). A similar concept has been proven to work with high resolution satellite data, with varying results (Thaler et al., 2019).

Optical remote sensing from UAVs is a viable alternative to in situ soil measurements, using multispectral sensors within the VIS-NIR spectrum (350-1,000nm) to estimate detailed SOC data. These systems have the potential to cover large areas that would be unattainable with conventional soil sampling methods (Nocita et al., 2015; Aldana-Jague et al., 2016). The new generation of multispectral sensors that are available for soil mapping have the capability to integrate with nearly any UAV platform that can support the weight of the sensor (Ben Dor, 2012; Crucil et al., 2019). These sensory systems are relatively low cost, have more flexibility, and can capture data at a high spatial and temporal resolutions when compared to traditional airborne or satellite imagery (Matese et al., 2015; Aldana-Jague et al., 2016; Soriano-Disla et al., 2017). However, the offset to using these sensors is the low radiometric and spectral resolution that is fundamentally present (Sona et al., 2016). This problem can be remedied by lower flying heights and using a downwelling light sensor (DLS) on the drone to account for accurate calibration of surface reflectance during image processing. The compatibility of UAV multispectral imagery models is limited due to the numerous factors that drive the spatial variability of SOC (Crucil et al., 2019). Currently, there are very few studies that examine the application of VIS-NIR spectral information derived from UAV imagery (Zhang and Kovacs, 2012; Nebiker et al., 2016; Soriano-Disla et al., 2017; Crucil et al., 2019). Topographic indices are also essential to include in SOC estimation models, as they attribute to accurate carbon mapping and prediction modelling (Huang et al., 2007b). These topographic variables are also
important to understand the spatial distribution of SOC throughout the soil (Schillaci et al., 2017). Research that comprises spatial and topographic factors must be carried out in order to create robust models to detect SOC from UAV multispectral imagery.

SMLR analysis is widely used to estimate SOC using spectroscopic information (Krishnan et al., 1980; Sudduth and Hummel, 1991; Hummel et al., 2001; Mueller and Pierce, 2003; Uno et al., 2005; Vasques et al., 2009). These models allow the user to input a variety of different predictor variables and receive an output multiple linear regression equation with coefficients. The cut-off threshold for including/excluding predictor variables is typically $p < 0.05$, which is a standard alpha significance threshold value (Vasques et al., 2009). The significant predictor variables, along with their associated coefficients and the intercept value are then combined to create a prediction equation. Values from the predictor variables can then be inputted into this equation to create an estimation of the response variable. This process was essential to this study because SOC estimation equations were produced using reflectance values, topographic variables, and spectral indices. From these equations, SOC predictions were produced and compared with values obtained from soil samples to determine the degree of accuracy between *in situ* SOC observations and SOC estimations derived from laboratory and UAV data.

The purpose of this research is to assess the integration of UAV-mounted multispectral sensors for SOC estimations in four agricultural fields in Southern Ontario. Multispectral UAV images were acquired near Guelph, Ontario, Canada to develop several prediction models, which were validated with *in situ* soil samples. Specific objectives of this study were (1) to estimate SOC concentration using two commercially available UAV-mounted multispectral sensors; and (2) to evaluate the effectiveness of laboratory spectral models that are assimilated with UAV reflectance data for SOC prediction.
3.3 Methods

3.3.1 Study Area Description

The four study sites used in this research are situated within the Canadian Lake Erie basin in Southwestern Ontario. These study sites were chosen to determine SOC concentrations in areas with varying topographic attributes and agricultural land management practices. All the sites are located within the Guelph-Eramosa township in Wellington County, with the Elora Long Term Trial (LTT) plots, site 1 and site 2 being located at the Elora Research Station (43°38'25.87"N, 80°24'36.67"W), and the remaining site belonging to the Woodrill Ltd. Farming Organization (43°42'30.72"N, 80°15'52.89"W) (Figure 3.1). These sites are located within close proximity to one another, being just 20 km apart, meaning that the climatic variables do not differ significantly between the four sites. Climate data were collected from the Environment Canada Canadian Climate Normals 1981-2010, at the Waterloo Wellington A station, located at the Waterloo Regional Airport. These study sites receive approximately 918 mm of precipitation per year, with 777 mm coming in the form of rain during the summer months (May – October), while 160 cm of snow generally falls in the winter months (November – April). The mean annual temperature in Wellington County is 7°C, with July being the warmest month with an average 20°C and January being the coldest month with an average of −6.5°C. The three warmest summer months are June, July and August with an average temperature of 18.8°C, while the three coldest winter months are December, January and February with an average temperature of −5.1°C. The rotation trials at Elora LTT originated in 1980 and include four repetitions of 15 crop rotations with plots that measure 15 m x 5 m. Each of these rotations have a conventional tillage and no-till plot (for a total of 30 plots x 4 rotations = 120 plots overall).
The topography at the LTT site has a very gentle slope (usually <1°). On these specific plots, the soil type is predominantly loamy in texture (London Loam). The field sites, site 1 and site 2 have are both conventionally tilled with the use of a moldboard plough and more prominent topographic differences. Site 1 has the largest elevation changes (357 m – 378 m), the highest slopes (6.4°), and has catchment features that are defined by flat areas between hilly terrain, which allows for water and nutrient pooling to occur, these drainage and pooling features could increase the SOC concentration in these areas. The final study site in Elora is site 2, which does not have high topographic differences, but rather exhibits little elevation and slope change.
The cropping history at each of the sites is similar and features predominantly a corn (*Zea mays*), soybean (*Glycine max*), and winter wheat (*Triticum aestivum*) rotation, which is a typical Southern Ontario cropping rotation.

The selected site from the Woodrill enterprise was conventionally tilled with a moldboard plough, however in recent years a no-till practices are incorporated. Similar to the three Elora sites, the main cropping rotation used is corn (*Zea mays*), soybean (*Glycine max*), and winter wheat (*Triticum aestivum*), with either cattle manure or municipal organic waste being applied to the soil each spring. The topography at this study site varies, with numerous hills and troughs located throughout the field. The soil texture at the Woodrill study site is sandy loam, with the majority of the field belonging to the Hillsburg fine sandy loam series, while a small portion of the field belongs to the Caledon fine sandy loam and Fox sandy loam series.

### 3.3.2 Soil Sample Collection, Processing and Measurements

Topsoil samples (0-15cm) were collected at predetermined locations for the four study sites in fall/spring of 2018 (Table 3.1). Once the collected samples were taken in field, they were transferred to a labelled bag and homogenized for SOC laboratory analysis. The soil samples were then air-dried, ground, passed through a 2mm sieve, and analyzed using an LECO CR-12 Carbon analyzer. This machine weighs the soil samples before and after combustion to quantify the amount of carbon contained in the soil. The inorganic carbon present in the soil samples was not extracted during the processing steps, so only the total SOC was determined. The Elora LTT soil samples used in this study were taken in the center of each plot that had bare soil, with no residue from the previous cropping year (Figure 3.2 B). The remaining Elora study sites (site 1 and site 2) were sampled at 20m intervals along predetermined transects that were found using
clustering analysis in ArcGIS (Figure 3.2 C, D). This clustering analysis recognized areas of high and low topographic variance and importance, ultimately creating a map that showed optimal

**Table 3.1 Number of samples associated with each study site.**

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Number of Samples</th>
<th>Soil Type</th>
<th>Topography</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smeltzer</td>
<td>726</td>
<td>Hillsburgh and Caledon Fine Sandy Loam; Fox Sandy Loam</td>
<td>Hilly Terrain</td>
<td>43°42’30”</td>
<td>80°15’50”</td>
</tr>
<tr>
<td>Elora LTT</td>
<td>32</td>
<td>London Loam</td>
<td>Gentle Slope</td>
<td>43°38’26”</td>
<td>80°24’31”</td>
</tr>
<tr>
<td>Site 1</td>
<td>20</td>
<td>Guelph and Burford Loam</td>
<td>Large Slope</td>
<td>43°39’03”</td>
<td>80°24’46”</td>
</tr>
<tr>
<td>Site 2</td>
<td>28</td>
<td>Guelph Loam</td>
<td>Gentle Slope</td>
<td>43°37’40”</td>
<td>80°23’52”</td>
</tr>
<tr>
<td>Total</td>
<td>806</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
</tbody>
</table>

sampling transects. For the Woodrill site, a 20m x 20m grid was formed over the entire field and soil samples were taken at each 20m interval (Figure 3.2 A). This allowed a high resolution SOC map to be created after the samples were processed in the laboratory. The GPS coordinates for the soil samples were recorded for each site. These locations were stored in the form of a point shapefile and linked with their associated SOC concentrations for each point. Table 3.2 outlines the mean, minimum and maximum SOC percentages produced from soil samples from each study site.

**Table 3.2: General SOC percentage statistics from each study site.**

<table>
<thead>
<tr>
<th>Study Site</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTT</td>
<td>2.6424</td>
<td>1.7435</td>
<td>3.6325</td>
</tr>
<tr>
<td>Site 1</td>
<td>3.4448</td>
<td>2.7188</td>
<td>4.3502</td>
</tr>
<tr>
<td>Site 2</td>
<td>2.8284</td>
<td>2.2843</td>
<td>6.2808</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>3.6408</td>
<td>1.6200</td>
<td>7.6480</td>
</tr>
</tbody>
</table>
3.3.3 Methodology and Materials

3.3.4 UAV Image Collection

*UAV Image Acquisition:*

MicaSense Red Edge and Tetracam RGB + 3 cameras were mounted on a DJI Matrice 100 drone to collect the images required from each of the study areas. These cameras will be referred to hereafter as MicaSense and Tetracam, respectively. They were used because they are commonly used and capture slightly different spectral ranges (Table 3.2). Therefore, a comparison can be made between how these two commercial multispectral sensors performed in relation to one another. A global positioning system logger, inertial measurement unit (IMU), and downwelling light sensor (DLS) on the drone were also integrated into the system for proper sensor calibration and orthorectification of the final image mosaic. Since the incorporation of the DLS was only usable for the MicaSense camera, a white calibration tile image was acquired prior to each flight to capture the sun illumination for calibration of the Tetracam images.

The MicaSense sensor has 5 channels with central wavelengths of 475nm, 560nm, 668nm, 717nm, and 840nm, while the Tetracam has 4 channels, with a regular RGB camera that can be separated into 3 individual bands and 3 additional cameras that capture the red edge and NIR. The central wavelengths of the Tetracam are 435nm, 515nm, 635nm, 680nm, 700nm, and 800nm. Both cameras capture imagery in the VIS-NIR, but differ in channel number, central wavelength, and bandwidth (Table 3.2). Flight lines for each study site were predetermined using Google Earth images integrated into the Pix4D Capture or DroneDeploy applications on an iPad. The Pix4D software suite is a professional photogrammetry and drone mapping computer program. It allows the user to capture images using the tablet/smartphone-based application and
process the resulting images with the desktop computer software to create maps.

These applications communicate with the UAV using GPS coordinates to create a flight path that will be followed. There are additional variables needed to initiate a mission, including flying height and image overlap. Due to flight duration constraints, the chosen flying height was 50m, with an 85% frontal and side overlap of each image for image mosaicking purposes. Before imaging in the field, ground control points (GCPs) were placed in the corners of the fields for proper orthorectification during processing. The time of image acquisition was between 10:00 AM and 2:00 PM, on clear-sky.

**Table 3.3: Tetracam and MicaSense sensor specifications.**

<table>
<thead>
<tr>
<th>Band</th>
<th>Center Wavelength (nm)</th>
<th>Bandwidth (nm)</th>
<th>Band</th>
<th>Center Wavelength (nm)</th>
<th>Bandwidth (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Red)</td>
<td>635</td>
<td>65</td>
<td>1 (Blue)</td>
<td>475</td>
<td>20</td>
</tr>
<tr>
<td>2 (Green)</td>
<td>515</td>
<td>35</td>
<td>2 (Green)</td>
<td>560</td>
<td>20</td>
</tr>
<tr>
<td>3 (Blue)</td>
<td>435</td>
<td>35</td>
<td>3 (Red)</td>
<td>668</td>
<td>10</td>
</tr>
<tr>
<td>4 (NIR1)</td>
<td>675</td>
<td>20</td>
<td>4 (NIR)</td>
<td>840</td>
<td>40</td>
</tr>
<tr>
<td>5 (NIR2)</td>
<td>700</td>
<td>20</td>
<td>5 (RedEdge)</td>
<td>717</td>
<td>10</td>
</tr>
<tr>
<td>6 (NIR3)</td>
<td>800</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

days to capture optimal sun illumination conditions. To reduce the affects of soil moisture on the resulting images, 3-4 days was waited after a precipitation event to allow the soil to air dry in the field. The acquisition timing on the MicaSense and Tetracam were set at 1 second intervals. The MicaSense captured images using a polygon method, which allowed the camera to start and stop capturing images depending on the orientation along the flight line. However, the Tetracam did not have this capability and it continuously obtained images for the entirety of the pre-set flights.
**UAV Image Processing:**

The initial image processing between the MicaSense and the Tetracam were different, because the raw data taken from each camera were stored in different formats. The MicaSense images were recorded in .TIFF format on removable SD cards, each having the capacity for 32 GB of storage. The GPS and IMU orientation data are embedded into the metadata of each image that is acquired by the MicaSense camera. The images used in the stitching process must then be manually sorted to remove any unwanted or irrelevant images. However, there were very few images that needed to be discarded because the polygon method used with the MicaSense was able to control the camera to capture images only from the predetermined flight line.

The Tetracam images were recorded in RAW format on 4 camera integrated SD cards, each capable of storing 16GB of data. These RAW images were processed through PixelWrench2 software, which was developed by Tetracam. Through this process, the images were converted from RAW to TIFF format. Furthermore, each image was processed individually and split to create 6 single page rasters, which corresponded with each channel on the camera. A GPS log file created for each flight was also converted from decimal degrees to degrees-minutes-seconds using an Excel formula and a KML file was produced from the resulting coordinates. Since the Tetracam captures constant images at 1 second intervals, nearly half of the data needed to be manually sorted and discarded as redundant. This typically occurred when the UAV needed to return to the home point for a new battery.

The processing of the MicaSense images was carried out in Pix4Dmapper to create a final orthomosaic. This software allows the user to either create a single or multiband raster that can be further processed for feature extraction using alternative software, such as ArcGIS. Once
the images were inputted into Pix4D, a display showed the travelled flight path of the UAV and the software also indicates where each image was captured. The final step allows for radiometric processing and calibration for the resulting orthomosaic. The MicaSense has a white calibration tile that is supported by the DLS that automatically incorporates itself into the images for each channel. Once this was complete, the initial processing took place, the ground control points (GCP) were identified, and the orthorectified mosaic was produced. This process was followed for each of the study area that was imaged by the MicaSense camera. These mosaics were then input into ArcGIS and resampled from 3cm to 50cm to remove some of the noise that is inherently present in high resolution imagery. The resampling method used was nearest neighbor (NN), in which the nearest cell center of the input raster will be identified, and the output raster will be assigned that value. The NN method was used on this data because it is a computationally rapid technique and it preserves accurate reflectance data from the area closest to the soil sample that was taken (vega et al., 2015; Bagheri, 2016). This is the optimal assignment technique because it does not change the value of the input cell.

A different process was used for the Tetracam images because they were not supported by the Pix4Dmapper software. The GPS log files that were created during the UAV flights were loaded in ArcGIS and created into point shapefiles. The point shapefiles taken during soil sampling were also inputted into ArcGIS, and the points that overlapped with one another were identified and further processed. Since the Tetracam only used a white calibration tile and not a DLS, the single page tiff images were still in 256-bit format. This meant the images contained digital numbers (DNs) rather than absolute reflectance. To counteract this problem, an image from the Tetracam taken prior to and after flights of a calibration tile was used. The reflectance value from this calibration panel, which was usually 255, was then divided by each individual
pixel value in the image and multiplied by 100 to convert the DNs to percent absolute reflectance. This process was carried out for every image that was used for extraction from each of the study sites. These images were then resampled from 1cm to 50cm, once again using the NN method in ArcGIS.

**UAV Imagery Spectral Extraction:**

The point shapefiles created during soil sampling were overlayed on the MicaSense images in preparation for reflectance extraction. A 0.5m buffer was then created around each point and the reflectance values were extracted for each channel on the multispectral camera using the ‘Zonal Statistics to Table’ tool in ArcGIS. This tool averaged the reflectance values and automatically recorded the values with their associated SOC values at each point. This averaging technique was used to reduce the spectral noise that would be present if only a single reflectance value pixel was used. This process was repeated for site 1, site 2 and the bare soil plots in the Elora LTT. However, only half of the Smeltzer field was imaged, due to a camera malfunction. There was also crop residue present on the field, which produced different reflectance values from bare soil. To counteract this issue, a supervised classification was carried out to identify crop residue, bare soil, and green vegetation. Spectral mixing graphs were created to compare how closely the channel values were to one another and ensure that minimal spectral overlap was present between channels and their associated values for crop residue, bare soil, and green vegetation. The resulting supervised classification was then overlayed with the Smeltzer soil sampling grid points, and the points that intersected the bare soil areas in the supervised classification were exported as a new shapefile. Since there was a significant amount of residue in this field, the reflectance values were only extracted from the single 50 x 50 cm resampled
pixel that was deemed bare soil after the supervised classification was performed (Figure 3.2 A, C, D, E).

Once the usable Tetracam images were identified in the previous steps, the average reflectance was extracted from a 0.5m buffer located in the center of each chosen image, depending on proximity to crop residue. This extraction was completed using the ‘Zonal Statistics to Table’ tool in ArcGIS, which links the averaged reflectance values with the specific SOC value at each point (Figure 3.2 B, C, D, E). This process was duplicated for all study areas, and the resulting tables were exported for further statistical analysis. The flowchart in Figure 3.3 displays a summary of the UAV imagery processing methodology in the top half of the flow chart, while the bottom half of the flow chart explains the SMLR approach that is explained further in this thesis.
Figure 3.2: Soil sampling and drone image extraction points for each study area overlayed on a DEM. The study areas are as follows: MicaSense Smeltzer extraction (A), Tetracam Smeltzer extraction (B), Elora long term trial plots (C), Elora site 1 (D), and Elora site 2 (E).
Figure 3.3: A flowchart outlining the UAV methodology used in this study.

3.3.5 Laboratory Spectral Measurements

Laboratory spectral measurements were conducted using an ASD FieldSpec3 spectroradiometer on each soil sample that was collected from Smeltzer, site 1 and site 2. The bandwidth of the UAV multispectral sensors (Table 3.2) were then extracted from the laboratory spectral data to create models that illustrate the theoretical limits of the Tetracam RGB + 3 and MicaSense sensors. Additionally, reflectance data from the United States Department of
Agriculture (USDA) Rapid Carbon Assessment (RaCA) database, within the bandwidth of the UAV multispectral sensors, was extracted to fit this data to SMLR and generate equations from the USDA RaCA database spectral information. The models created from the laboratory data are referred to as theoretical limits, in terms of $R^2$ and RMSE, because the spectral data is collected under optimal measuring conditions. The models derived from the laboratory spectral information were then applied to UAV data to assess the transferability of the equations.

**ASD FieldSpec 3 Data:**

The soil samples that were collected for this study (Figure 3.4 A-D) were analyzed using an ASD FieldSpec3 spectroradiometer (Analytical Spectral Devices Inc.) and were sorted according to the wavelength ranges that are present on the channels found on the Tetracam and MicaSense multispectral cameras (Table 3.2). The initial measurements taken ranged from 350-2500 nm, which was well above the highest wavelength of the multispectral camera at 860 nm. A 50W halogen lamp was positioned 50 cm above the samples at nadir to capture
Figure 3.4: Digital elevation models and soil sampling points for all study areas. The study areas are as follows: Smeltzer (A), Elora long term trials (B), Elora site 1 (C), Elora site 2 (D).
optimal illumination and a white calibration tile were used to calibrate the sensor before any measurements were taken. The fiber optic cable that receives the light was positioned 7cm from the sample at a 35-degree angle, which created a 4.5cm in diameter field of view (FOV) on each soil sample. This FOV was small enough to capture only the soil reflectance, while avoiding the plastic edges of the container that the sample was in. The reflectance was measured 3 times for each sample, with a white panel calibration occurring every 5 samples to maintain the consistency of soil measurements. This procedure was performed on site 1 and site 2. Due to time constraints, laboratory reflectance values for the 726 Smeltzer samples was not included. However, a section of the Smeltzer field soil samples (n=37) were selected for laboratory spectral measurement. The Elora LTT samples were also excluded from this portion of the study because of time constraints.

USDA Rapid Carbon Assessment Data:

An overarching goal of the research was to provide techniques for rapid soil organic carbon extraction from multi-spectral data. One of the largest SOC data bases available was composed by the USDA as part of the (RaCA) data (Soil Survey Staff, 2013). The models derived using this dataset were of high interest because they could be tested against the other datasets present in this study. This would help explain if the models created using this well-established, public database were transferrable to other datasets. If the models were robust enough to encompass different datasets, a generalized model could be created to estimate SOC across various study areas.

The SOC data extracted from the RaCA database needed to correspond with the glacial geology of the Southern Ontario study areas. The Elora and Smeltzer fields ranged from glacial till to glaciofluvial deposits, with an overarching loamy soil texture according to the Ontario
Geological Survey for surficial geology. The near surface geology (bedrock) is of the middle and lower Silurian period consisting of dolostone, shale, sandstone and siltstone. Regions of similar geological history were matched with regions in United States of America covered by data from the RaCA, with a specific focus on Michigan and Ohio because of their glacial similarities to the study sites. Sites selected within these states have very similar parent materials and were part of the Carboniferous period, which enriched the soil with carbonates. These glacial and geological similarities were essential in identifying the RaCA extraction areas because they needed to closely resemble the soils that were being examined through our study. The final extraction step was to sort the data samples, so they only included soils with loamy texture.

The entire RaCA database consists of 144,833 soil samples that were collected in the upper 1m of 32,084 soil profiles at approximately 6,000 randomly selected locations for measurement of organic and inorganic carbon by VIS-NIR spectroscopy using a LabSpec 2500 Visible Near Infrared Spectrometer. A subset from this database was retrieved using the ‘soilDB’ R package ([https://cran.r-project.org/web/packages/soilDB/soilDB.pdf](https://cran.r-project.org/web/packages/soilDB/soilDB.pdf)). More specifically, the bbox method was used as a spatial bounding box to retrieve spatial data, where the user inputs a maximum 5-degree x 5-degree box of samples to extract. This method is advantageous because it uses proper geographic locations to get certain samples that surround the relevant study areas. The approximately 600 samples that are used in this study were taken from areas neighboring states to Southern Ontario in an attempt and capture the soil types that have experienced similar climates (including glaciation). These samples were then sorted to extract agricultural fields with SOC less than or equal to 10%, which are similar to the values present in the study areas. This data was then further sorted to only include the wavelength ranges that are
present on the MicaSense and the Tetracam multispectral cameras that were mounted on the UAV. The flowchart found in Figure 3.5 provides an overview of the laboratory methodology.

Figure 3.5: A flowchart outlining the ASD FieldSpec 3 and USDA RaCA methodology used in this study.

3.3.6 Model Development

*USDA and ASD FieldSpec 3 Pre-Modelling Data Processing:*

The USDA, ASD FieldSpec 3 and UAV imagery datasets were incorporated into individual SMLR analysis to find a reproducible workflow to estimate SOC using reflectance
data. After the reflectance was extracted from the UAV imagery for each study site, the data was initially run using a simple general linear model that was coded using the R statistical software. However, the USDA and ASD FieldSpec 3 data needed additional processing to capture the full bandwidth of the sensors that were active on the UAV at the time of image acquisition. Each channel on the MicaSense and Tetracam sensors are made up of a different number of wavelengths, which needed to be combined to create a complete dataset. This was achieved by using the full width at half maximum (FWHM) technique according to the sensor specifications provided by the manufacturer (Laliberte et al., 2011; Logie and Coburn, 2018). These datasets were retained for further processing.

The USDA hyperspectral data was further processed to include soils of similar glaciation, texture, and SOC concentration when compared to the study areas. A model was created with the downloaded USDA data and the resulting formula was used to estimate the SOC within the USDA dataset. These estimations, which were generated from the USDA hyperspectral reflectance equations, were then compared to the observed data that was included in the USDA downloaded dataset. This data violated the normality and homoskedasticity assumptions of a linear model and as a result was used purely for testing purposes. The reason the USDA equations were retained for testing purposes was to analyze the degree of accuracy from equations that were created from this hyperspectral soil reflectance library, inputted with UAV reflectance data.

*Spectral Indices:*

In addition to band reflectances, several spectral indices were also evaluated in this study. The spectral indices used are based on previous analysis (Mulder et al., 2011; Nocita et al., 2013; Bhunia et al., 2018). The four spectral indices that were chosen are: optimized soil adjusted
vegetation index (OSAVI), modified secondary soil adjusted vegetation index (MSAVI2), bare soil index (BSI), and normalized difference vegetation index (NDVI). Specifically, these spectral indices were chosen because they focus on combinations of the red and NIR multispectral ranges, which are most sensitive to SOC estimation (Qi et al., 1994; Rondeaux et al., 1996; Bhunia et al., 2018; Xue and Su, 2017). The calculations were implemented into each dataset to test the effectiveness of spectral indices when estimating SOC. It is important to note that the NIR Tetracam reflectance values came from the final channel (NIR3) in the sensor which is located around 800nm. This was because the wavelength ranges of NIR1 and NIR2 are overlapped by the Red channel on this instrument.

The OSAVI and MSAVI2 methods are related to the soil line concept in remote sensing because they use red and NIR reflectances in their equations (Lawrence and Ripple, 1998; Fox and Sabbagh, 2002; Xue and Su, 2017; Baghi and Oldeland, 2019). Although these indices are generally used to differentiate between soil and vegetation for the purpose of monitoring vegetation, this research aims to use them as a means of identifying the sensitivity of reflectance normalizing indices for SOC estimation. OSAVI is an optimized version of the soil adjusted vegetation index (SAVI), using a fixed correction factor of 0.16, rather than the traditional value range of 0 – 1 that is used with SAVI (Rondeaux et al., 1996). This fixed value has been calculated and tested using the soil line method. OSAVI is defined where NIR and Red are the near-infrared and red reflectance values respectively, and L is the soil brightness correction factor.

\[
\text{OSAVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red} + \text{L}} \times (1 + \text{L})
\]

MSAVI2 uses a more complex formula produces values between -1.0 and +1.0. The same variables are inputted into the formula, with NIR and Red being the near-infrared and red
reflectance values, respectively. This formula was derived from the original modified soil adjusted vegetation index (MSAVI) and is beneficial to the user because the soil line and soil brightness correction factors are no longer needed to produce an accurate result (Qui et al., 1994). The formula is as follows:

\[
\text{MSAVI}_2 = \frac{(2\times\text{NIR}+1)\sqrt{(2\times\text{NIR}+1)^2-8*(\text{NIR}-\text{Red})}}{2}
\]

The BSI uses a combination of the red, green, blue and NIR channels to create a spectral indicator that identifies soil variations. BSI values attempt to normalize brighter and darker areas that are unevenly lit by sunlight, as well as reduce the differences between wet and dry soil areas (Kumar, 2013 and Shabou et al., 2015) with modifications from Jamalabad and Abkar (2004). The BSI is calculated from the following equation:

\[
\text{BSI} = \frac{[(\text{Red}+\text{Green})-(\text{Red}+\text{Blue})]/[(\text{NIR}+\text{Green})+(\text{Red}+\text{Blue})]+100*100
\]

NDVI is perhaps the most common of the proposed spectral indices, with abundant research and literature pertaining to this topic. Values of NDVI range from -1.0 to +1.0, with positive values accounting for highly reflective surfaces such as vegetation, and negative values indicative of areas with low vegetation (Jackson and Huete, 1991; Qi et al., 1993; Gbolo et al., 2015). NDVI can be calculated using the following equation:

\[
\text{NDVI} = \frac{(\text{NIR}-\text{Red})/\text{(NIR+Red)}}
\]

When this index is applied, bare soil is expected to have a low value due to the red and NIR reflectance values being very similar (Bhunia et al., 2018).

*Topographic Variables:*

The topographic variables used in this study were topographic wetness index (TWI), slope and flow accumulation (Sun et al., 2015; Mondal et al., 2017; Schillaci et al., 2017). These three variables were identified as essential to help explain the effect that topography exhibits on
the spatial variability of SOC (Pei et al., 2010; Oueslati et al., 2013; Grinand et al., 2017; Mondal et al., 2017). The raw topography data was downloaded using the Land Information Ontario (LIO) portal. This data was collected by Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) using an airborne Lidar system from 2016-2018. For the purpose of this project, certain point cloud tiles that encompassed the study areas were downloaded from the Lake Erie dataset. Once the raw data was downloaded from LIO, the point clouds were filtered and the resulting filtered point clouds were interpolated. This was an initial de-noising technique to create tiles that only included ground return points. The interpolated tiles were then mosaicked to create an image of the entire study area. To further reduce the noise of the point cloud, objects classified as “off terrain” were removed. Finally, any cells that represented no data or missing data were filled using the inverse distance weighting (IDW) scheme. These values were derived from neighboring cells. The slope, TWI and flow accumulation were then created using the final, corrected DEM to ensure the topographic values were accurate.

Model Building:

A stepwise multiple linear regression analysis was used to create equations that display the most influential combinations of reflectance bands, topographic variables and band ratios. This type of analysis aims to create a model that explains the relationship between two or more independent predictor variables and the response variable. This is achieved by fitting a linear equation to the observed data that is inputted into the algorithm. The multiple linear regression formula is as follows:

\[ Y = \beta_0 + \beta_1 \chi_1 + \beta_2 \chi_2 + \ldots + \beta_n \chi_n \]

where \( Y \) is the value of the response variable for the \( i \)th case, \( \beta_0 \) is the intercept, \( \beta_1, \beta_2, \ldots, \beta_n \) are the regression coefficients for the predictor variables for the \( i \)th case, and \( \chi_1, \chi_2, \ldots, \chi_n \) are the
predictor variables for the $i^{th}$ case. This research always used SOC as the response variable and used a combination of spectral bands, topographic variables and spectral indices as the predictor variables. These equations were created using the step AIC function in R statistical software, which chooses the most accurate permutation of input variables. This particular function allows for the stepwise regression to be built “forward”, “backward” or “both” directions and the permutation with the lowest AIC value is chosen as the model. For the purpose of this study, “both” directions were chosen because the model starts processing with no predictors and progressively adds the most influential predictor variables. The statistical software then removes the variables that are confirmed to no longer add improvement to the fit of the model. As a result, this technique also accounts for overfitting of the model, which is essential when trying to formulate a generalizable model outside of the original dataset (Dotto et al., 2018; Venables and Ripley, 2002).

3.3.7 Statistical Analysis

The Cook’s Distance Threshold was applied to each individual reflectance dataset derived from the ASD FieldSpec 3 and the UAV sensors to remove outliers. The main objective of this method was to identify influential outliers present in the predictor variables within each dataset. The threshold of $4/n$ was used, which is directly affected by the removal of data points within the dataset (Cook, 1977; Hair et al., 1998). If there are data points that are deemed influential, the point with the highest influence is to be removed from the dataset and the statistics must be run again. This process should be repeated until there are no influential points remaining according to the threshold that is calculated. The stepAIC function in R was used to determine the best combination of independent variables to use to create the SOC estimation formula.
Since the USDA RaCA data had over 550 data points, the Cook’s Distance was an impractical method to use for outlier removal. The interquartile rule for outliers was used instead for effective outlier removal. This technique uses the interquartile range (IQR), multiplied by 1.5, and added to the third quartile to create the outlier threshold value. This method is synonymous with the widely known box and whisker plot, which uses the IQR, as well as the first and third quartile to create the well-known graphic (Hubert and Veeken, 2008; Dawson, 2011; Spitzer et al., 2014; Hofmann et al., 2017)

Once the influential points were identified and removed using the Cook’s Distance method or interquartile range and a model was created, the five assumptions of a linear regression were tested on 68 individual model. The Mardia skewness and Mardia kurtosis tests were used to test for multivariate normality within the dataset. The p-value was required to be greater than 0.05 for the assumption to be accepted. To test the issue of multicollinearity, a variance inflation factor (VIF) test was performed. This test states a VIF lower than 4 and no higher than 10 should be present for the assumption of little to no multicollinearity to be present. A Shapiro-Wilk test was run to detect the violation of the normality assumption, with the alpha value set at 0.05. The null hypothesis was rejected if the p-value was less than 0.05, meaning the data was not normally distributed. The Breusch Pagan test was used to determine homoskedasticity or heteroskedasticity, with the alpha set at 0.05. In this case, the null hypothesis was also rejected if the p-value was less than 0.05, which meant the data was heteroskedastic. The occurrence of this would result in the data being transformed in an attempt to create a homoskedastic dataset. The RMSE and R-squared values were recorded for each model to show the error that is present and proportion of variance that the independent variables can predict on the dependent variable. The prediction sum of squares (PRESS),
and Akaike Information Criteria (AIC) values were the final values recorded. The PRESS states how well the model will predict a new dataset, while choosing models with lower AIC values provides a further indicator of the model quality based on a penalty for over parameterization. Once the assumptions were tested and satisfactory, the linear model was performed on the fitted dataset. The formula and independent value coefficients were derived using R statistical software on each individual model to create a series of prediction formulas.

3.4 Results

3.4.1 Overview

In this study, we created four SMLR model groups (Table 3.3) using ASD FieldSpec 3 and UAV datasets: reflectance, reflectance and topography, band ratios, and band ratios and topography. The variables used in these multivariate linear regression models were consistent with what has been used in literature when trying to estimate SOC (Pei et al., 2010; Bhunia et al., 2018; Mondal et al., 2017). The two focal assumptions of a linear model, normality and homoskedasticity were recorded as the Shapiro-Wilk test and Cooks-Weisberg test, respectively. These assumptions were met in every model (p > 0.05), except for the USDA models, which were previously stated to be integrated into this study for testing purposes only. There was some multicollinearity present between variables, but the SMLR was able to select variables to account for this to a certain degree. A calibration-validation (60%-30%) analysis was not used in this study because we used in situ data as means of model validation. The observed and predicted SOC for each dataset were graphed to visualize the linearity that was present. The graphs that produced negative values were not included in this study because a negative SOC value is not possible.
3.4.2 UAV Models

The UAV models produced promising results when comparing the performance of various models and their associated data (Table 3.3). The observed and predicted SOC was graphed to visualize the linearity that was present for the UAV data (Figure 3.6 A-H). The MicaSense models that used only reflectance always included a combination of red, red edge, and NIR bands, which is in agreement with literature as sensitive bands when estimating SOC from UAV imagery (Fox and Sabbagh, 2002; Bartholomeus et al., 2008; Vasques et al., 2009; Peng et al., 2014). However, the Tetracam models seemed to focus around the NIR1, NIR2 and NIR3 spectral regions, which was likely affected by band overlap from the NIR channels into the red spectral region. The model performance when using only reflectance values is variable and seems to be dataset dependent. This trend continues throughout the models that incorporate topography along with reflectance and little increase or decrease in $R^2$ or RMSE occurs. However, the addition of topographic variables is seemingly important, as many models include all three topographic features. The TWI and Slope seen to have the greatest inclusion in all the models and are generated in six of the eight possible equations. There was some expectation that model fits using the band ratios would work better as they were selected to minimize lighting affects by the sun, as well as areas that are lighter and darker from differing soil moisture content. Of the band ratios, the OSAVI and BSI are the most prevalent among all models, which is possibly due to their incorporation of red and NIR spectral bands into their equations. The most accurate UAV models in terms of $R^2$ were produced using band ratios and topographic data. The $R^2$ values for the UAV models ranged from 0.02 to 0.93, while the RMSE values ranged from 0.08 to 0.44. The site 2 study site from the Tetracam sensor using band ratios and topography produced the highest $R^2$ and the lowest RMSE, which indicated that this was the best
model. The combination of band ratios and topographic variables clearly made a difference as it outperformed nearly every other model in terms of $R^2$ and RMSE. This is an especially interesting finding when compared to the reflectance and topography models, with the Tetracam and MicaSense sensors producing better $R^2$ and RMSE values in three of the four models. All spectral indices and topographic variables were included in the equations quite regularly, with OSAVI and NDVI occurring the most frequently for spectral indices and slope occurring for five of the eight models, while TWI and flow accumulation only occurred three times.
Table 3.4: The theoretical limits of SOC estimation using field measured reflectance values from the UAV imagery for the MicaSense and Tetracam sensors.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Reflectance</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R2</td>
</tr>
<tr>
<td><strong>MicaSense</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTT</td>
<td>soc = Blue(1.437)+Red(-0.917)+NIR(0.264)+0.332</td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>soc = Green(0.426)+Red(-0.206)+NIR(-0.064)+3.633</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>soc = Blue(0.54)+Green(0.221)+Red(-0.107)+RedEdge(-0.296)+3.776</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>soc = RedEdge(0.264)+NIR(-0.332)+2.371</td>
<td></td>
</tr>
<tr>
<td><strong>Tetracam</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTT</td>
<td>soc = Blue(-0.025)+Green(0.049)+NIR1(0.096)+NIR2(-0.147)+NIR3(-0.055)+4.014</td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>soc = NIR1(0.033)+2.771</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>soc = Green(0.058)+NIR1(0.078)+NIR2(-0.130)+2.376</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>soc = Blue(0.102)+Red(-0.046)+NIR1(-0.017)+4.058</td>
<td></td>
</tr>
<tr>
<td><strong>Reflectance and Topography</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MicaSense</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTT</td>
<td>soc = Blue(1.618)+Red(-0.713)+TWI(-3.092)+Slope(-2.232)+FlowAcc(0.243)+34.843</td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>soc = Green(0.772)+Red(-0.398)+NIR(-0.115)+TWI(-0.650)+Slope(-0.203)+10.680</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>soc = Blue(0.343)+RedEdge(-0.195)+TWI(0.186)+Slope(0.137)+FlowAcc(-0.076)+1.911</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>soc = Green(0.284)+RedEdge(0.437)+NIR(-0.713)+FlowAcc(0.106)+1.518</td>
<td></td>
</tr>
<tr>
<td><strong>Tetracam</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTT</td>
<td>soc = Green(0.036)+NIR1(0.035)+NIR2(-0.101)+NIR3(-0.063)+TWI(-2.636)+Slope(-2.321)+33.387</td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>soc = NIR1(0.033)+2.771</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>soc = Green(0.284)+Red(-0.172)+NIR1(0.042)+NIR3(-0.076)+TWI(0.463)+Slope(0.165)+(-1.869)</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>soc = Green(0.081)+Red(-0.058)+NIR1(-0.022)+NIR2(0.021)+NIR3(-0.012)+Slope(-0.045)+3.744</td>
<td></td>
</tr>
</tbody>
</table>

**Band Ratios**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MicaSense</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R2</td>
</tr>
<tr>
<td>LTT</td>
<td>soc = Blue(1.437)+Red(-0.917)+NIR(0.264)+0.332</td>
</tr>
<tr>
<td>Site 1</td>
<td>soc = Green(0.426)+Red(-0.206)+NIR(-0.064)+3.633</td>
</tr>
<tr>
<td>Site 2</td>
<td>soc = Blue(0.54)+Green(0.221)+Red(-0.107)+RedEdge(-0.296)+3.776</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>soc = RedEdge(0.264)+NIR(-0.332)+2.371</td>
</tr>
</tbody>
</table>

56
<table>
<thead>
<tr>
<th></th>
<th>LTT</th>
<th>0.40</th>
<th>0.24</th>
<th>3.37</th>
<th>0.08</th>
<th>1.42</th>
<th>0.12</th>
<th>soc = OSAVI(-0.606)+BSI(0.004) + 6.375</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>0.13</td>
<td>0.32</td>
<td>11.98</td>
<td>0.49</td>
<td>1.68</td>
<td>0.64</td>
<td>soc = MSAVI2(29.718)+NDVI(-32.622) + (-0.304)</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>0.62</td>
<td>0.14</td>
<td>-15.92</td>
<td>0.48</td>
<td>0.39</td>
<td>0.44</td>
<td>soc = OSAVI(-0.294)+BSI(0.003)+NDVI(20.321) + (-3.158)</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.14</td>
<td>0.43</td>
<td>44.01</td>
<td>0.50</td>
<td>6.83</td>
<td>0.79</td>
<td>soc = OSAVI(-0.064) + 4.149</td>
<td></td>
</tr>
</tbody>
</table>

**Tetracam**

<table>
<thead>
<tr>
<th></th>
<th>LTT</th>
<th>0.66</th>
<th>0.19</th>
<th>-6.95</th>
<th>0.79</th>
<th>0.80</th>
<th>0.76</th>
<th>soc = OSAVI(-0.129) + 5.654</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>0.38</td>
<td>0.34</td>
<td>17.38</td>
<td>0.91</td>
<td>2.33</td>
<td>0.25</td>
<td>soc = OSAVI(0.025)+NDVI(-1.027) + 3.021</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>0.24</td>
<td>0.20</td>
<td>-2.89</td>
<td>0.30</td>
<td>0.89</td>
<td>0.41</td>
<td>soc = OSAVI(-0.123)+BSI(0.001) + 2.149</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.19</td>
<td>0.38</td>
<td>97.69</td>
<td>0.78</td>
<td>15.07</td>
<td>0.90</td>
<td>soc = OSAVI(-0.103)+BSI(0.001)+NDVI(9.224) + 3.318</td>
<td></td>
</tr>
</tbody>
</table>

**Band Ratios and Topography**

**MicaSense**

<table>
<thead>
<tr>
<th></th>
<th>LTT</th>
<th>0.80</th>
<th>0.15</th>
<th>-11.78</th>
<th>0.20</th>
<th>0.59</th>
<th>0.50</th>
<th>soc = OSAVI(-1.088)+MSAVI2(160.518)+BSI(0.013)+NDVI(-147.746)+TWI(-1.847) +Slope(-1.128)+FlowAcc(0.222) + (-7.679)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>0.21</td>
<td>0.24</td>
<td>3.80</td>
<td>0.81</td>
<td>1.03</td>
<td>0.54</td>
<td>soc = MSAVI2(23.972)+NDVI(-25.921) + 0.218</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>0.85</td>
<td>0.08</td>
<td>-29.12</td>
<td>0.13</td>
<td>0.15</td>
<td>0.67</td>
<td>soc = OSAVI(0.105)+TWI(0.206)+FlowAcc(-0.082) + (-0.738)</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.16</td>
<td>0.42</td>
<td>41.29</td>
<td>0.72</td>
<td>6.20</td>
<td>0.28</td>
<td>soc = NDVI(6.190) + 2.416</td>
<td></td>
</tr>
</tbody>
</table>

**Tetracam**

<table>
<thead>
<tr>
<th></th>
<th>LTT</th>
<th>0.82</th>
<th>0.15</th>
<th>-9.52</th>
<th>0.22</th>
<th>0.42</th>
<th>0.99</th>
<th>soc = OSAVI(-0.191)+MSAVI2(-6.392)+BSI(-0.001)+NDVI(4.597)+Slope(-0.430) + 10.349</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>0.85</td>
<td>0.15</td>
<td>-7.46</td>
<td>0.90</td>
<td>0.46</td>
<td>0.75</td>
<td>soc = OSAVI(0.024)+Slope(-0.098)+FlowAcc(-0.025) + 3.487</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>0.93</td>
<td>0.08</td>
<td>-39.00</td>
<td>0.12</td>
<td>0.15</td>
<td>0.44</td>
<td>soc = OSAVI(-0.412)+MSAVI2(2.836)+BSI(0.002)+NDVI(-6.896)+TWI(0.438)+Slope(0.146) + (-2.320)</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.02</td>
<td>0.44</td>
<td>142.31</td>
<td>0.11</td>
<td>22.58</td>
<td>0.85</td>
<td>soc = Slope(-0.038) + 3.267</td>
<td></td>
</tr>
</tbody>
</table>

R² values are the model fits achieved within each dataset, RMSE is the measure of absolute fit, AIC is the Akaike information criterion and PRESS is the predicted sum of squares. The equations are included with the variables used and their corresponding coefficients.
A) Elora LTT UAV MicaSense SOC Predicted vs SOC Observed

\[ R^2 = 0.8686 \]
\[ R^2 = 0.7788 \]
\[ R^2 = 0.4343 \]
\[ R^2 = 0.8686 \]

B) Elora LTT UAV Tetracam SOC Predicted vs SOC Observed

\[ R^2 = 0.5825 \]
\[ R^2 = 0.8123 \]
\[ R^2 = 0.6766 \]
\[ R^2 = 0.7775 \]
C) Site 1 UAV MicaSense SOC Predicted vs SOC Observed

R² = 0.3377
R² = 0.2653
R² = 0.4687
R² = 0.2776

D) Site 1 UAV Tetracam SOC Predicted vs SOC Observed

R² = 0.8863
R² = 0.2428
R² = 0.26
E) Site 2 UAV MicaSense SOC Predicted vs SOC Observed

F) Site 2 UAV Tetracam SOC Predicted vs SOC Observed
G) Smeltzer UAV MicaSense SOC Predicted vs SOC Observed

H) Smeltzer UAV Tetracam SOC Predicted vs SOC Observed
Figure 3.6 (A-H): Graphs illustrating the predicted and observed SOC for each UAV model type, grouped by study site. The abbreviations are as follows: MicaSense (MS), Tetracam (TTC), reflectance model (Reflectance), reflectance and topography model (Reflectance Topo), band ratio model (BR), and band ratio and topography model (BR Topo).

3.4.3 ASD FieldSpec3 Models

The ASD FieldSpec3 and USDA data models were incorporated into this study to establish the theoretical limits of SOC using laboratory spectrometers. Table 3.4 and Figure 3.7 (A-F) illustrate the accuracy of the various models that were created for the ASD FieldSpec 3 models. This table also includes all ASD FieldSpec 3 model equations, with their variables and associated coefficients that were created for SOC estimation. When all ASD FieldSpec3 models were compared, they produced a minimum $R^2$ value of 0.10 and a maximum $R^2$ value of 0.95, with a maximum RMSE of 0.28 and a minimum RMSE of 0.07. These highest $R^2$ value and lowest RMSE value were found with the site 2 dataset, using the reflectance dataset of the Tetracam. Topographic variables were only included in the MicaSense and Tetracam models for the Smeltzer dataset. The $R^2$ value increased in both cases, 0.47 for the MicaSense sensor and 0.60 for the Tetracam sensor. There was a similar occurrence in the Smeltzer dataset when topography was included with reflectance to create a model. The Smeltzer soil samples were extracted from areas of large topographic variation, which could help explain why the topographic variables were so essential to create a model with a high $R^2$ value.
Table 3.5: The theoretical limits of SOC estimation using field measured reflectance values from the ASD FieldSpec3 within the MicaSense and Tetracam sensor wavelength ranges.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>R2</th>
<th>RMSE</th>
<th>AIC</th>
<th>Shapiro-Wilk</th>
<th>PRESS</th>
<th>Cooks-Weisberg</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflectance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MicaSense</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>soc = Green(-1.827)+Red(0.63)+NIR(0.262) + 8.102</td>
</tr>
<tr>
<td>Site 1</td>
<td>0.75</td>
<td>0.26</td>
<td>7.58</td>
<td>0.34</td>
<td>1.29</td>
<td>0.34</td>
<td>soc = Green(-1.827)+Red(0.63)+NIR(0.262) + 8.102</td>
</tr>
<tr>
<td>Site 2</td>
<td>0.86</td>
<td>0.09</td>
<td>-19.36</td>
<td>0.31</td>
<td>0.15</td>
<td>0.84</td>
<td>soc = Blue(0.726)+RedEdge(-1.525)+NIR(0.988) + 2.43</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.22</td>
<td>0.20</td>
<td>-4.48</td>
<td>0.69</td>
<td>1.22</td>
<td>0.83</td>
<td>soc = RedEdge(-0.449)+NIR(0.380) + 1.694</td>
</tr>
<tr>
<td>Tetracam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>soc = Blue(1.215)+Green(-3.152)+Red(-1.665)+NIR1(2.431) + 8.881</td>
</tr>
<tr>
<td>Site 1</td>
<td>0.80</td>
<td>0.25</td>
<td>6.89</td>
<td>0.96</td>
<td>1.12</td>
<td>0.36</td>
<td>soc = Blue(1.215)+Green(-3.152)+Red(-1.665)+NIR1(2.431) + 8.881</td>
</tr>
<tr>
<td>Site 2</td>
<td>0.95</td>
<td>0.07</td>
<td>-30.24</td>
<td>0.90</td>
<td>0.08</td>
<td>0.76</td>
<td>soc = Blue(-1.010)+Green(3.044)+Red(-3.949)+NIR1(2.077) + 3.782</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.66</td>
<td>0.13</td>
<td>-16.71</td>
<td>0.59</td>
<td>0.47</td>
<td>0.59</td>
<td>soc = Green(0.602)+Red(-4.021)+NIR1(16.286)+NIR2(-15.134)+NIR3(2.233) + 2.141</td>
</tr>
<tr>
<td>Reflectance and Topography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MicaSense</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>soc = Green(-1.827)+Red(0.630)+NIR(0.262) + 8.102</td>
</tr>
<tr>
<td>Site 1</td>
<td>0.70</td>
<td>0.26</td>
<td>7.58</td>
<td>0.34</td>
<td>1.29</td>
<td>0.34</td>
<td>soc = Green(-1.827)+Red(0.630)+NIR(0.262) + 8.102</td>
</tr>
<tr>
<td>Site 2</td>
<td>0.87</td>
<td>0.09</td>
<td>-29.30</td>
<td>0.13</td>
<td>0.23</td>
<td>0.07</td>
<td>soc = Blue(1.828)+Green(-0.914)+RedEdge(-1.601) +NIR1(1.095) +TWI(0.872) +Slope(1.064) +FlowAcc(0.034) + (-6.289)</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.83</td>
<td>0.18</td>
<td>-8.51</td>
<td>0.95</td>
<td>0.99</td>
<td>0.31</td>
<td>soc = Blue(-1.226)+Green(1.282)+Red(-0.496)+TWI(1.058)+Slope(0.493) + (-6.288)</td>
</tr>
<tr>
<td>Tetracam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>soc = Blue(2.018)+Green(-5.106)+NIR1(1.737)+TWI(0.160) + 7.850</td>
</tr>
<tr>
<td>Site 1</td>
<td>0.75</td>
<td>0.24</td>
<td>5.45</td>
<td>0.62</td>
<td>1.04</td>
<td>0.15</td>
<td>soc = Blue(2.018)+Green(-5.106)+NIR1(1.737)+TWI(0.160) + 7.850</td>
</tr>
<tr>
<td>Site 2</td>
<td>0.64</td>
<td>0.16</td>
<td>-12.05</td>
<td>0.56</td>
<td>0.74</td>
<td>0.28</td>
<td>soc = Blue(-2.052)+Green(4.145)+Red(-5.578)+NIR1(7.731)+NIR2(-5.359 ) +NIR3(0.990) +TWI(0.727)+Slope(0.863) + (-5.388)</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.64</td>
<td>0.22</td>
<td>1.69</td>
<td>0.05</td>
<td>1.64</td>
<td>0.62</td>
<td>soc = NIR2(-0.596)+NIR3(0.444)+TWI(1.010)+Slope(0.509) + (-7.786)</td>
</tr>
<tr>
<td>Band Ratios</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MicaSense</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>soc = OSAVI(1.036)+MSAVI2(-24.065)+BSI(-0.012) + 17.807</td>
</tr>
<tr>
<td>Site 1</td>
<td>0.69</td>
<td>0.26</td>
<td>8.04</td>
<td>0.38</td>
<td>1.27</td>
<td>0.35</td>
<td>soc = OSAVI(1.036)+MSAVI2(-24.065)+BSI(-0.012) + 17.807</td>
</tr>
<tr>
<td>Site 2</td>
<td>0.41</td>
<td>0.19</td>
<td>-5.87</td>
<td>0.63</td>
<td>0.97</td>
<td>0.68</td>
<td>soc = OSAVI(-0.334)+MSAVI2(-658.460)+BSI(0.004)+NDVI(985.922) + 21.049</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.39</td>
<td>0.28</td>
<td>13.48</td>
<td>0.92</td>
<td>2.26</td>
<td>0.66</td>
<td>soc = MSAVI2(-1250.116)+BSI(0.002)+NDVI(1807.252) + 47.188</td>
</tr>
<tr>
<td>Tetracam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>soc = MSAVI2(-1250.116)+BSI(0.002)+NDVI(1807.252) + 47.188</td>
</tr>
<tr>
<td>Site</td>
<td>R2</td>
<td>B2</td>
<td>R3</td>
<td>B4</td>
<td>B5</td>
<td>soc = OSAVI(B2)+MSAVI(B3)+BSI(B4)+NDVI(B5) + C6</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>0.67</td>
<td>0.28</td>
<td>8.08</td>
<td>0.19</td>
<td>1.38</td>
<td>0.17 soc = OSAVI(0.381)+BSI(-0.006) + 8.414</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>0.61</td>
<td>0.16</td>
<td>-12.40</td>
<td>0.84</td>
<td>0.65</td>
<td>0.78 soc = OSAVI(-0.398)+MSAVI(33.955)+BSI(0.004) + (-7.489)</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.10</td>
<td>0.22</td>
<td>-2.74</td>
<td>0.32</td>
<td>1.37</td>
<td>0.66 soc = MSAVI2(4.922) + 0.502</td>
<td></td>
</tr>
</tbody>
</table>

### Band Ratios and Topography

**MicaSense**

<table>
<thead>
<tr>
<th>Site</th>
<th>R2</th>
<th>B2</th>
<th>R3</th>
<th>B4</th>
<th>B5</th>
<th>soc = OSAVI(B2)+MSAVI(B3)+BSI(B4)+NDVI(B5) + C6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>0.69</td>
<td>0.26</td>
<td>8.04</td>
<td>0.38</td>
<td>1.28</td>
<td>0.35 soc = OSAVI(1.036)+MSAVI(-24.065)+BSI(-0.012) + 17.807</td>
</tr>
<tr>
<td>Site 2</td>
<td>0.47</td>
<td>0.17</td>
<td>-11.68</td>
<td>0.67</td>
<td>0.74</td>
<td>0.27 soc = OSAVI(-0.317)+MSAVI(-678.919)+BSI(0.004)+NDVI(1013.795) + 22.016</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.86</td>
<td>0.16</td>
<td>-14.21</td>
<td>0.56</td>
<td>0.76</td>
<td>0.18 soc = MSAVI2(3.993)+TWI(1.130)+Slope(0.548)+FlowAcc(0.126) + (-11.309)</td>
</tr>
</tbody>
</table>

**Tetracam**

<table>
<thead>
<tr>
<th>Site</th>
<th>R2</th>
<th>B2</th>
<th>R3</th>
<th>B4</th>
<th>B5</th>
<th>soc = OSAVI(B2)+BSI(-0.006) + 8.414</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>0.67</td>
<td>0.27</td>
<td>8.07</td>
<td>0.19</td>
<td>1.38</td>
<td>0.17 soc = OSAVI(0.381)+BSI(-0.006) + 8.414</td>
</tr>
<tr>
<td>Site 2</td>
<td>0.64</td>
<td>0.14</td>
<td>-17.01</td>
<td>0.75</td>
<td>0.48</td>
<td>0.59 soc = OSAVI(-0.374)+MSAVI(31.664)+BSI(0.003) + (-6.833)</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.70</td>
<td>0.23</td>
<td>1.80</td>
<td>0.16</td>
<td>1.73</td>
<td>0.38 soc = MSAVI2(6.887)+TWI(0.925)+Slope(0.496) + (-9.928)</td>
</tr>
</tbody>
</table>

R² values are the model fits achieved within each dataset,
-RMSE is the measure of absolute fit
-AIC is the Akaike information criterion and PRESS is the predicted sum of squares.
The equations are included with the variables used and their corresponding coefficients.
**A) Site 1 ASD FieldSpec 3 MicaSense SOC Predicted vs SOC Observed**

R² = 0.7412  
R² = 0.7412  
R² = 0.7536  
R² = 0.7536

**B) Site 1 ASD FieldSpec 3 Tetracam SOC Predicted vs SOC Observed**

R² = 0.7144  
R² = 0.7144  
R² = 0.8185  
R² = 0.8014
C) Site 2 ASD FieldSpec 3 MicaSense SOC Predicted vs SOC Observed

D) Site 2 ASD FieldSpec 3 Tetracam SOC Predicted vs SOC Observed
E) Smeltzer ASD FieldSpec 3 MicaSense SOC Predicted vs SOC Observed

\[ R^2 = 0.8806 \]
\[ R^2 = 3 \times 10^{-5} \]
\[ R^2 = 0.866 \]
\[ R^2 = 0.2179 \]

F) Smeltzer ASD FieldSpec 3 Tetracam SOC Predicted vs SOC Observed

\[ R^2 = 0.6438 \]
\[ R^2 = 0.1341 \]
\[ R^2 = 0.5753 \]
\[ R^2 = 0.6783 \]
3.4.4 USDA RaCA Models

The models using USDA RaCA reflectance data were only able to be created using reflectance and band ratios because there was no topographic information associated with the data (Table 3.5; Figure 3.8). These models reflected similarities to the ASD FieldSpec3 models because the red, red edge, and NIR portions of the spectrum (all of which pertain to SOC reflectance sensitivity) were included in the reflectance models. The model outcomes were practically identical when comparing the MicaSense and Tetracam reflectance model $R^2$ values measuring 0.52 and 0.53, respectively, while an RMSE value of 0.71 was produced for both reflectance models. However, different predictor variables were chosen, which was most likely a result of the differences in spectral bandwidth from each channel of the multispectral sensors. Both band ratio models used all possible predictor variables. Similar to the reflectance models, the band ratio models also produced nearly identical results. This can be attributed to the same spectral channels used for each of the spectral indices. The Tetracam band ratio model had a slight advantage with an $R^2$ of 0.62 and an RMSE of 0.63. Although the results these models were promising, two key assumptions of a linear regression (normality and homoskedasticity) were violated. Unfortunately, this meant that the application of the equations produced from these models is not recommended. The AIC and PRESS values were also extremely high, indicating a poor model fit. An explanation of the poor model performance may come from the different outlier test that was used for these datasets. Outliers were removed using the IQR method, which is not as strict as the Cook’s Distance technique. IQR method was used in this case because of
the large number of data points that would need to be individually examined. This difference in outlier tests could be one of the reasons as to why the data fails to meet two of the main assumptions of a multivariate linear regression, leading to poor model performance.

Table 3.6: SOC estimation using field measured reflectance values from the USDA RaCA dataset within the MicaSense and Tetracam sensor wavelength ranges.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>R2</th>
<th>RMSE</th>
<th>AIC</th>
<th>Shapiro-Wilk</th>
<th>PRESS</th>
<th>Cooks-Weisberg</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflectance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MicaSense</td>
<td>0.52</td>
<td>0.71</td>
<td>1199.89</td>
<td>0.00</td>
<td>282.65</td>
<td>0.00</td>
<td>soc = Blue(0.063) + Green(0.127) + Red(-1.009) + RedEdge(0.765) + 2.915</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetracam</td>
<td>0.53</td>
<td>0.71</td>
<td>1195.88</td>
<td>0.00</td>
<td>280.57</td>
<td>0.00</td>
<td>soc = Green(0.113) + Red(0.316) + NIR1(-2.204) + NIR2(1.719) + 2.959</td>
</tr>
<tr>
<td>Band Ratios</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MicaSense</td>
<td>0.60</td>
<td>0.65</td>
<td>1096.31</td>
<td>0.00</td>
<td>235.67</td>
<td>0.00</td>
<td>soc = OSAVI(-0.073) + MSAVI2(-65.523) + NDVI(127.285) + 1.149</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetracam</td>
<td>0.62</td>
<td>0.63</td>
<td>1069.01</td>
<td>0.00</td>
<td>224.38</td>
<td>0.00</td>
<td>soc = OSAVI(-0.122) + MSAVI2(-78.613) + BSI(0.001) + NDVI(148.75) + 0.653</td>
</tr>
</tbody>
</table>

R² values are the model fits achieved within each dataset, RMSE is the measure of absolute fit, AIC is the Akaike information criterion and PRESS is the predicted sum of squares. The equations are included with the variables used and their corresponding coefficients.
3.4.5 Integration of Laboratory Models with UAV Datasets

One of the goals of this research was to determine the degree of accuracy SOC could be predicted when using UAV data and equations developed a priori based on spectral reflectance libraries. This was achieved by comparing observed SOC derived from soil sample data with the predicted SOC estimated using UAV reflectance data inputted into the equations developed from the laboratory ASD FieldSpec 3 and the USDA RaCA reflectance data. The $R^2$ and RMSE were then calculated to show how well the models performed and graphs were created to illustrate the linearity of the observed and predicted SOC (Table 3.6; Figure 3.9 A-C; and Table 3.7; Figure 3.10 A-F). Although the UAV imagery was processed to reduce image noise and other geometric

---

**Figure 3.8:** Graph displaying the USDA RaCA and Tetracam SOC predicted and SOC observed values produced from the models. The abbreviations for the models are as follows: Tetracam (TTC), reflectance models (Reflectance) and band ratio models (BR).
distortions small topographic variation and surface roughness will cause issues related to bidirectional reflectance. Bidirectional reflectance issues were not directly assessed. However, use of band ratioing approaches has been shown to reduce topographic effects (Lelong et al., 2008). Therefore, only the laboratory models that included band ratioing were used for analysis. The laboratory and UAV raw reflectance values will exhibit large differences because all parameters were optimized in the ASD FieldSpec 3 laboratory data. However, the same optimization process could not occur when acquiring UAV imagery, which would result in bidirectional reflectance error being introduced into the dataset. When bare soil is present, these errors can be mainly attributed to the distribution of scattered and emitted radiation (Labarre et al., 2019). Surface roughness, microtopography, multiangular reflectance scattering and soil moisture concentration also influence bidirectional reflectance (Roosjen et al., 2015; Labarre et al., 2019). The band ratios that were included in this study were selected to help normalize the data, regarding soil moisture and brightness factors, which should allow the laboratory band ratio models to be transferrable to UAV imagery (Huete et al., 1985; Mattikalli, 1997; Bhunia et al., 2019). The ASD FieldSpec 3 and USDA RaCA laboratory band ratio models were analyzed and tested for transferability. The \( R^2 \) values for the MicaSense wavelength range were found to either be higher or comparable to the Tetracam sensor range across all datasets and models. The RMSE values that the Tetracam RMSE values (min=2.92; max=46.01) far exceeded the MicaSense values (min=1.48; max=7.89), which indicates that there is a high amount of variability in the predicted SOC when compared to the observed SOC. The \( R^2 \) values were relatively low for all the models, with the highest value being 0.32 for the MicaSense ASD FieldSpec3 band ratio.
model. There were several other models with $R^2$ close to 0.30 within the ASD FieldSpec3 models. The RMSE values for these models were quite large, which indicates

Table 3.7: $R^2$ and RMSE values for the models created using ASD FieldSpec3 band ratio models and spectral indices derived from UAV reflectance data for each sensor. The equations that were used are included to demonstrate the most important band ratios for this analysis.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Band Ratios</th>
<th>R2</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MicaSense</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>0.00</td>
<td>4.53</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>0.32</td>
<td>4.04</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.01</td>
<td>6.76</td>
<td></td>
</tr>
<tr>
<td><strong>Tetracam</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>0.08</td>
<td>4.78</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>0.10</td>
<td>46.01</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.00</td>
<td>2.92</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Band Ratios and Topography</th>
<th>R2</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MicaSense</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>0.01</td>
<td>4.67</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>0.26</td>
<td>3.60</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.02</td>
<td>1.48</td>
<td></td>
</tr>
<tr>
<td><strong>Tetracam</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>0.04</td>
<td>4.89</td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>0.28</td>
<td>44.63</td>
<td></td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.00</td>
<td>4.10</td>
<td></td>
</tr>
</tbody>
</table>
A) ASD FieldSpec 3 Site 1 Models with UAV MicaSense Spectral Data SOC Predicted vs SOC Observed

B) ASD FieldSpec 3 Site 2 Models with UAV MicaSense Spectral Data SOC Predicted vs SOC Observed
Figure 3.9 (A-C): Graphs displaying the ASD FieldSpec 3 SOC predicted and SOC observed with MicaSense and Tetracam models using UAV data. The abbreviations for the models are as follows: MicaSense (MS), Tetracam (TTC), band ratio models (BR) and band ratio and topography models (BR Topo).

Table 3.8: $R^2$ and RMSE values for the models created using the USDA band ratio models and spectral indices derived from UAV reflectance data for each sensor. The resulting model equations are also included to demonstrate band ratio selection.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>R2</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Band Ratios</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MicaSense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>0.00</td>
<td>7.89</td>
</tr>
<tr>
<td>Site 2</td>
<td>0.22</td>
<td>1.76</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.16</td>
<td>1.87</td>
</tr>
<tr>
<td>Tetracam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>0.03</td>
<td>4.84</td>
</tr>
<tr>
<td>Site 2</td>
<td>0.18</td>
<td>35.60</td>
</tr>
<tr>
<td>Smeltzer</td>
<td>0.01</td>
<td>4.02</td>
</tr>
</tbody>
</table>
A) USDA RaCA Site 1 MicaSense Model and UAV MicaSense Spectral Data SOC Predicted vs SOC Observed

SOC Predicted vs SOC Observed

USDA RaCA MS BR Estimates

Linear (USDA RaCA MS BR Estimates)

R² = 0.0006

B) USDA RaCA Site 1 Tetracam Model and UAV Tetracam Spectral Data SOC Predicted vs SOC Observed

SOC Predicted vs SOC Observed

USDA RaCA TTC BR Estimates

Linear (USDA RaCA TTC BR Estimates)

R² = 0.0836
C) USDA RaCA Site 2 MicaSense Model and UAV MicaSense Spectral Data SOC Predicted vs SOC Observed

D) USDA RaCA Site 2 Tetracam Model and UAV Tetracam Spectral Data SOC Predicted vs SOC Observed
E) USDA RaCA Smeltzer MicaSense Model and UAV MicaSense Spectral Data SOC Predicted vs SOC Observed

\[ R^2 = 0.1568 \]

F) USDA RaCA Smeltzer Tetracam Model and UAV Tetracam Spectral Data SOC Predicted vs SOC Observed

\[ R^2 = 0.0064 \]
Figure 3.10 (A-F): Graphs displaying the USDA RaCA MicaSense and Tetracam model derived SOC predicted and SOC observed using UAV data. The abbreviations for the models are as follows: MicaSense (MS), Tetracam (TTC) and band ratio models (BR).

that the accuracy to which a prediction of SOC can be made is relatively low. High RMSE values and low $R^2$ values were also a problem within the USDA models. However, this could be linked to the models not meeting all linear regression assumptions, which could have negatively impacted their performance. Currently, it is very challenging to have an accurate prediction of SOC when implementing UAV reflectance data with equations generated from laboratory analysis. The equations were excluded in Table 3.6 and Table 3.7 because the equations are not stable, producing very high RMSE values. Using laboratory generated equations with UAV reflectance data is not advised at this time.

The spectral indices that were included in the ASD FieldSpec3 models were mostly OSAVI, MSAVI2 and BSI. The band ratios included in each model seemed to have a negligible effect on the total performance of the model. This also seemed to occur when topographic variables were included in the model. Topographic variables were only present in two of the six models, with TWI and Slope appearing in both, while flow accumulation only appeared in one. These two models had little to no change in $R^2$, but the RMSE values lowered in the MicaSense model that included topography and were higher in the Tetracam model that included topography. The USDA band ratio models for both sensors included all band ratios. The MicaSense models clearly outperformed the Tetracam models, with higher $R^2$ values ($R^2 = 0.22, 0.16$) and much lower RMSE values associated with these $R^2$ values (RMSE = 1.76, 1.87).
3.4.6 Discussion

This study has demonstrated that a multivariate linear regression analysis using various datasets (reflectance, reflectance and topography, band ratios, band ratios and topography) is a reasonable way to measure SOC with both laboratory and UAV VIS-NIR data. However, there are obviously still problems with these techniques, as the results tend to vary, especially for the UAV derived data. As well, when working with multispectral data (400-860 nm) for prediction analysis, there is inherently multicollinearity that is present within the bands. This is because spectral information tends to be very similar within this portion of the electromagnetic spectrum. Although the SMLR models accounted for some multicollinearity, it was still present within the datasets and could have been a cause for error.

The findings in this study align with the literature surrounding laboratory and UAV-based SOC estimation modelling (Chang et al., 2001; Nocita et al., 2011; Nocita et al., 2013; Rodionov et al., 2014; Bhunia et al., 2018). This study has demonstrated that the laboratory methods (i.e. reflectance, band ratio) were able to outperform the UAV models, which is consistent with previous work (Stevens et al., 2006; Crucil et al., 2019). This was most likely a result of optimal sensing conditions, including perfect illumination at nadir, soils that were dried and sieved to eliminate moisture and roughness effects, and the lack of any soil cover whatsoever because it was removed during the soil processing stage (Chang et al., 2001; Nayak et al., 2019). Although the UAV methods did not outcompete the laboratory models, they achieved an average $R^2$ of 0.48 for both the MicaSense and Tetracam sensors, which is quite high considering the minimal correction factors included in the image processing. These models could be used to derive a
field-based map of surface SOC however a significant amount of in situ observation were required.

$R^2$ and RMSE that were derived from the two UAV sensors produced quite similar results (i.e., $R^2$ ranged from 0.02-0.93 for Tetracam and $R^2$ ranged from 0.14-0.85 for MicaSense; RMSE for Tetracam ranged from 0.08-0.78 and RMSE for MicaSense ranged from 0.08-0.43). The MicaSense sensor was more consistent throughout all models when compared to the Tetracam sensor, while the Tetracam excelled when band ratios were applied. This was likely due to the sun angle and illumination being embedded into the MicaSense image metadata from the DLS, while the Tetracam had to rely on spectral indices for correction. Both sensors did not experience notable increases in $R^2$ and RMSE when topographic variables were added to reflectance. This is because the reflectance models experienced poor performance and the topographic variables did not provide enough additional information to increase the $R^2$ or lower the RMSE a considerable amount. When added to band ratios however, the models improved greatly. This was most likely attributed to the influx of information and the lighting and moisture corrections that are inherent with band ratios (Bhunia et al., 2019). Perhaps the most noticeable irregularity throughout the entire UAV model results is the low $R^2$ and high RMSE values that are present for the Smeltzer study site datasets across both sensors. A logical explanation for these low values could be the required amount of preprocessing (i.e., supervised classification) on the UAV images. The supervised classification may have introduced unwanted error and variability within the dataset that could have had a negative effect on model results.

The ASD FieldSpec 3 models achieved similar $R^2$ and RMSE values across all models with very few inconsistencies. This study identified that using the Tetracam wavelength range
creates better models when compared to the MicaSense wavelength range when using ASD
FieldSpec 3 reflectance values for SOC estimation. This is likely caused by the Tetracam
capturing more data because it has an additional band relative to the MicaSense sensor, as well
as three narrow-band NIR channels that can acquire more specific spectral information. The
ASD FieldSpec 3 models that incorporated reflectance data outperformed the band ratio datasets.
Since there were no illumination differences, moisture difference or bidirectional reflectance
concerns, the reflectance for these models did not need any additional processing to perform
well. Based on the assessment of this study, the Tetracam wavelength range can be
recommended for use.

Currently, the feasibility of using laboratory equations with UAV data is still relatively
unknown. The ASD FieldSpec 3 results suggest that the laboratory models that are calibrated
using soil samples from the same study sites as the UAV models perform better than the USDA
RaCA models. This is likely due to the identical nature of the soil used in the ASD FieldSpec 3
and UAV models, while the USDA RaCA models use similar, but not identical soils. Another
solution could exist in creating the model from a larger dataset to account for some of the
variation that is inherently present in the USDA RaCA data. Also, the models developed from
the USDA data did not follow a normal distribution and was not transformed, which could have
ultimately been the reason that this dataset does not show promising results for transferrable
relationships.

Although SOC estimation has been performed using spectroscopic derived data, there is a
need to further this research to improve the accuracy, efficiency, and assimilation of these
models. This would include models that could be used on varying types of soils (i.e., texture or
class) and completing research that focuses on UAV models. (Chang et al., 2001; Stevens et al., 2008; Aldana-Jague et al., 2016). Although the past few decades has seen a notable increase in research studies surrounding multispectral and hyperspectral measurement of SOC, a more robust model that is not location-specific must be created to determine the most efficient input parameters to improve predictive accuracy (Malley et al., 2004; Angelopoulou et al., 2019). While the extensive studies are present regarding the importance of SOC, there is a lack of field-scale studies that focus on the quantification of SOC.

Unlike conventional laboratory methods, UAV sensing in particular offers a reasonably quick and affordable method to detect SOC. This technique is versatile and can essentially be applied anywhere that agricultural practices exist (Chang et al., 2001; Stevens et al., 2008; Stevens et al., 2010; Croft et al., 2012; Aldana-Jague et al., 2016). Although several studies have been performed to quantify SOC using UAV mounted sensors, the incorporation of spectroscopic derived data for SOC prediction is still not well understood. Since the spatial variability of soils is such a complex mechanism, a challenge arises when trying to create a model that would be able to encompass all aspects of the soil, as well as a model that could be applied to fields under differing conditions. More likely, multiple models should be created in order to capture specific locational characteristics and be applied at large spatial scales. These models would have the potential to produce results that would benefit the understanding of the relationship that exists between UAV acquired VIS-NIR data and SOC prediction.

3.5 Conclusion

In this study, we illustrated that multivariate linear regression analysis using VIS-NIR (475-800 nm) laboratory spectrometry and UAV multispectral images can be used to estimate
field-scale SOC. Multivariate regression analysis is a promising technique to create equations for SOC concentration assessment. The equations selected by the AIC approach used in this study proves that a similar workflow can be applied to various datasets to develop SOC prediction models for laboratory and UAV derived spectral data. Using UAV VIS-NIR data, coupled with laboratory derived models must be further researched to understand relationships that exist. When integrating laboratory models with UAV data, the ASD FieldSpec 3 models that were calibrated with local soil sample spectral data performed better than the USDA RaCA models because the soil data (i.e., texture, types) was exactly the same for the ASD FieldSpec 3. Further research should be focused in this area in order to create a generalized, transferrable model. Realistically, a series of models would most likely need to be created to account for all soil variations, even within smaller regions, such as Southern Ontario. Since the models would rely on the VIS-NIR spectrum and publicly available topographic data, they provide a relatively low-cost alternative to traditional SOC measurements. If these types of models are generated, it would allow for UAV images to be quickly transformed into a digital SOC map and provide potential benefits to agricultural land management practices. These maps could be integrated into precision agriculture systems to identify areas of high and low SOC, which could affect the tillage and crop rotation used in those certain areas.

Future recommendations to further this research include studies that include a large number of sampling points that can be used for model calibration and validation. This study had quite a small sample size from each location and more accurate, robust results could have been achieved if a larger sample size was used. Despite this limitation, the results of this study demonstrate that UAV-derived VIS-NIR imagery can be used for SOC content estimation and
future research should be conducted to fully capture the potential of this technique for measuring SOC. The optimal flight conditions (i.e., flying height, flying speed, illumination conditions) of UAV and soil conditions should also be considered in future studies, as they have a direct impact on reflectance values that are present in the acquired images. If these conditions could be accounted for, the accuracy of SOC estimation using UAV could increase and the uncertainties present with these methods would be better understood.

4 Summary and Conclusion

This thesis focused on two main research objectives and assessed the feasibility of using two commercially available UAV-mounted multispectral cameras for SOC prediction. The first objective consisted of estimating SOC concentrations using SMLR analysis from multispectral imagery acquired from UAV. Georeferenced in situ soil samples were collected for the analysis from four study sites in southern Ontario near Guelph, Ontario in the spring of 2018. Images were collected from UAV-mounted multispectral cameras from the same study sites in southern Ontario and the georeferenced soil sampling points were overlayed on the resulting orthomosaics for spectral extraction. Topographic variables were created using DEM data from the same fields and spectral indices were formed using the reflectance values. Stepwise multivariate linear regression analysis was performed on the reflectance, reflectance and topography, band ratio, and band ratio and topography datasets for every individual study area. The results illustrate that the best UAV model to use in terms of $R^2$ and RMSE was band ratio with topographic variables. Graphical interpretations further support this finding and indicate that a positive linear relationship exists between the in situ observed SOC and the predicted SOC derived from UAV reflectance data.
The second objective was focused on evaluating how effective laboratory models inputted with UAV reflectance data could estimate SOC. This objective was split into two separate laboratory datasets; ASD FieldSpec 3 spectroradiometer and USDA RaCA database. To begin this objective, theoretical limits were established for both laboratory datasets, with the ASD FieldSpec 3 models being derived from in situ soil sample data and the USDA RaCA models being created from the data provided within this database. Initial results suggested that these models performed well, with the highest ASD FieldSpec 3 R\textsuperscript{2} being 0.95 and the highest R\textsuperscript{2} from the USDA RaCA database being 0.62. However, when UAV reflectance data was used in these models to estimate SOC, the results were not as promising, with the highest R\textsuperscript{2} achieved being 0.32. Overall, the findings in this study suggest that SOC estimation is feasible using UAV models, but the assimilation of UAV reflectance data with laboratory models is currently not plausible to an acceptable accuracy.

UAV prediction methods are a time and cost-efficient way for SOC monitoring, although ground validation is still recommended. Although these methodologies are still in their infancy, it is expected that this research has applications for the measurement of SOC using commercially available UAV-mounted multispectral sensors at a regional scale. Scaling this research could prove difficult, as variations in soil, agricultural land management practices and climate have a direct impact on spectral reflectance properties. However, it is well documented in academic literature that using spectral indices as a means of bidirectional reflectance normalization could provide a potential solution to these issues. Future research should replicate this experiment to determine if these results are location specific, or if they are reproducible at scale. The
integration of UAV reflectance data with laboratory derived models should also be further explored to compute the potential of these techniques.
REFERENCES


