PREDICTING NUTRIENT CONTENT, PLANT HEALTH, AND SITE SUITABILITY: A CASE STUDY OF *ERAGROSTIS TEF*

By

K. Colton Flynn
Bachelor of Science in Earth Science
University of Arkansas
Fayetteville, AR
2010

Master of Arts in Geography
University of Arkansas
Fayetteville, AR
2013

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2019

PREDICTING NUTRIENT CONTENT, PLANT HEALTH, AND SITE SUITABILITY: A CASE STUDY OF *ERAGROSTIS TEF*

Dissertation Approved:

Dr. Amy E. Frazier

Dissertation Adviser

Dr. Carlos Cordova

Dr. Adam Mathews

Dr. Andrew Doust

ACKNOWLEDGEMENTS

This work was heavily supported by various grants and fellowships and thus would not have been a success without the support of the J. William Fulbright U.S. Student Research Award, Walter Kolb STEM Graduate Studies Scholarship, Robberson Summer Research Fellowship, and other funding from grants provided by Dr. Amy Frazier. Outside of funding I also had strong support from the Geography department at Oklahoma State University and will be forever grateful for the education they provided me throughout my PhD studies.

I have also had the great pleasure of working closely with individuals that have pushed me to be a better researcher, person, and educator. This list is long, and thus, I could never name everyone, but I would like to thank Emily Ellis, Liz Brian, Andy Han, and Nathalia Graf Grachet for constantly supporting me without question academically, personally, and in field/lab work; your friendship and support has truly led to this success. As for mentors, I would like to thank Dr. Jon Comer, Dr. Jennie Popp, and Sintayehu Admas. These three don't know each other but they have pushed and supported me without question (even when they probably should have questioned) throughout my educational career. I want to say thank you for believing in me, providing the learning opportunities, and answering the knock at the door/constant emails even when you knew it was that same guy on the other end…again. Additionally, I would like to thank the committee members that are probably the only ones that will ever read a majority of this document as they provided generous and constructive feedback throughout the many versions. Thus, thank you for your patience and time Dr. Carlos Cordova, Dr. Adam Mathews, and Dr. Andrew Doust.

Academically and professionally, no one has ever had a greater impact on me than my advisor Dr. Amy Frazier. The first time I met her in person I was deathly afraid, shell shocked, and fan-girling a bit as I knew she was amazing in her field/profession and I wanted to emulate that one day. What I didn't know at the time is how great of an advisor I would end up with. The time and patience she has offered me over the past four years are things I could never thank her enough for as she endlessly helped with filling out grants/fellowships (there were many), told me I had to go to Ethiopia when I thought I might not, advised me from half way around the world, put up with my stubbornness (because sometimes I think I'm right, and often am not), cared enough to see me through to graduation despite a relocation, and perhaps the most important for me was that she led by example not only words. All of these things and so much more that could fill up an entire page, but I will spare you. Dr. Frazier, I am a better researcher, academic, and person because of you, so thank you.

Finally, nobody has had a greater effect on me than my family. I was raised by two loving parents, Kyle and Linda Flynn, that always supported my brother and I in anything we wanted to pursue, with one mostly unspoken caveat: You must pursue goals

with all that you have leaving nothing on the field (we are competitive and into sports lol). I am thankful for my parents instilling this drive in me from early on as it has made this PhD experience much smoother than it would have been otherwise. Outside of this, thank you mom and dad for constantly providing love and guidance throughout this process and whatever else I decide to take on in the future, as I never have to question that you will be there whenever I need you.

Name: K. COLTON FLYNN

Date of Degree: MAY, 2019

Title of Study: PREDICTING NUTRIENT CONTENT, PLANT HEALTH, AND SITE SUITABILITY: A CASE STUDY OF *ERAGROSTIS TEF*

Major Field: GEOGRAPHY

Abstract: Advancements in agricultural and geographic principals have led to worldwide food and agricultural globalization. Because agricultural production continues to further in global interconnectedness, confirmed precision agriculture (PA) methods are required to monitor crops in-field. PA utilizes a remote sensing method referred to as imaging spectroscopy (IS). IS is often performed using a field spectroradiometer that identifies reflectance values. The reflectance values obtained have been utilized in agricultural studies to correlate spectral reflectance to biochemical and biophysical properties. However, while there is a large body of research focusing on IS predicting these agricultural characteristics, many studies have only employed the research in a single region/location resulting in findings that may lacking reproducibility and replication (R&R) for more than a single environment. The lack of regionally comparative IS methods for nutrient and plant health analysis is important as varying geographies may prove to have an effect on IS findings. Therefore, the proposed research utilizes IS methods to predict nutrient and plant health values utilizing tef (Eragrostis tef) as a case study as its cultivated in Ethiopia and the United States. Currently, in the United States, the cultivation of tef is limited thus the United States could benefit from an exploration of site suitability analysis to aid expansion of tef cultivation in the U.S. It is through this interdisciplinary study that potential improvement to geography and remote sensing theory/methods can be obtained to achieve goals within food/agriculture geography.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

1.1 THE GLOBALIZATION OF AGRICULTURE

Advancements in the studies of crops, genetics, languages, and phytogeography have allowed the identification of geographical origins of food crops. The identification of the geographical origins of food crops have led to a greater awareness that worldwide food and agricultural production is interconnected globally (Khoury et al. 2016). The globalization of food crops is most often motivated by an aim to overcome various pests and pathogens, provide season specific cultivation options, and meet dietary requirement of a region's inhabitances (Jenning and Cock 1977). These food crop globalization trends are further amplified by an increase in purchasing power, a movement towards supermarkets, increased consumption outside the home, urbanization, subsidized agricultural practices, refrigerated transport, and industrialized agriculture (Khoury et al. 2014; Kearney 2010). As a part of the industrialization of agriculture, precision agriculture has resulted in improvements in food production practices.

1.2 PRECISION AGRICULTURE

Precision agriculture is an innovative field integrating technology with agricultural practices that aims to increase efficiency of resource utilization (i.e. water, fertilizer, etc.)

and decrease the ambiguity of decisions required during crop production, such as fertilizer placement and irrigation practices, that are often highly variable (Schellberg 2008). Precision agriculture is often considered a subfield of geography because many of the technologies utilized in geography are also implemented in precision agriculture such as global positioning systems (GPS), geographic information systems (GIS), and remote sensing methodologies (Zhang et al. 2002). Precision agriculture began in earnest in the late 1970s when soil survey/sampling started to become a regular practice, and aerial photography was introduced as a way to scout crops (Robert 2002). Since then, studies focused on advancements in precision agriculture have increased exponentially (Schellberg 2008) as precision agricultural practices can be beneficial to the environment and to profitability (Zhang et al. 2002; Schellberg 2008). Some of these practices include the incorporation of spectral scanning technologies and algorithms to predict site suitability of crops.

Precision agriculture in the U.S. has been driven by four motives: (1) strict environmental legislation; (2) public fear of overuse of chemicals; (3) profitability from a decrease in inputs; and (4) an increased need for large-scale farm management (Zhang et al. 2002). Furthermore, precision agriculture has also contributed to global advances in production (Oliver et al. 2013) through improvements in crop mapping, phenological analysis, crop health, pest/weed management, and monitoring nutrient levels of cultivated crops (Calvao and Pessoa 2015). However, based on my experience the presence of precision agriculture in Ethiopia is lacking. Historically, nutrient and health analysis of crops, foods, and agricultural grasses have relied on procedures that are time-consuming, expensive, destructive, and must be completed in a lab; where precision agriculture methods such as remote sensing and imaging spectroscopy can be done *in situ*, are often non-invasive, and

results can be determined almost instantaneously in some situations (Martinez-Valdivieso et al. 2014). Accordingly, precision agriculture techniques, specifically those using imaging spectroscopy, are becoming a preferred technique for nutrient and health analysis. Additionally, site suitability analysis can be incorporated for decision making of where crops have the potential to be grown to aid farm management decisions made across differing environments.

1.3 REMOTE SENSING AND IMAGING SPECTROSCOPY

Remote sensing is the science of measuring electromagnetic energy that has been reflected or emitted from an object or phenomenon through a device that is not in contact with the object/phenomenon. Throughout history, remote sensing has utilized many different types of platforms including hot air balloons, kites, pigeons, airplanes, satellites, unmanned aerial systems (UASs), and field spectroradiometers (Jensen 2016). Each of these platforms carry sensors that are designed to capture electromagnetic energy through varying spatial, spectral, and radiometric resolutions. Spectral resolution is the number and dimension of the specific wavebands of electromagnetic energy that are discernable by a sensor (Jensen 2016). A low spectral resolution, or broad-band spectral sensor, covers a spectrum of wavelengths using a few, broad intervals; while a high spectral resolution, or a narrow-band sensor, covers a spectrum of wavelengths using many, narrow intervals. Broad-band sensors are common among satellites as they are often utilized to differentiate simple targets such as cultural features and vegetation characteristics, additionally, hyperspectral sensors are costly. Satellite sensors are also useful for imaging large areas of interest because they often have low spatial resolutions that allow them to cover a large area.

The earliest application of remote sensing for precision agriculture relied on the multispectral scanner (MSS) broad-band sensor flown onboard the Earth Resources Technology Satellite 1 (ERTS-1; eventually renamed Landsat-1) satellite. This sensor was used to classify fields according to crop type as either soybean or corn agriculture with an 83% accuracy rate (Bauer and Cipra 1973). Since then, studies have focused on the implementation of broad-band sensors to create indices for vegetation health (Overgaard et al. 2013a) including the widely used normalized difference vegetation index (NDVI: Rouse et al. (1973), which utilizes the red and NIR (near-infrared) portions of the electromagnetic spectrum to determine vegetation vigor. Other broad-band indices utilizing the red and NIR portions of the electromagnetic spectrum include the simple ratio index (Jordan 1969; Rouse et al. 1973) and the modified simple ratio (Chen 1996). However, NDVI and other broad-band vegetation indices have been found to lack accuracy due to oversaturation of the analysis (Schlerf et al. 2005), its sensitivity to varying climates (Wessels et al. 2007), and cases of underestimation (Krupenikov et al. 2011).

Nevertheless, many of the broad-band sensors mounted on satellite platforms lack the spectral precision required for some fields of research (Moran et al. 1997), including many precision agriculture applications that are mainly only employing NDVI. Narrow-band spectral imaging is useful in research that requires a greater spectral sensitivity (Thenkabail 2000), because it can analyze smaller portions of the electromagnetic spectrum allowing for more specified reflectance/absorption readings. As sensors have improved, a method known as imaging spectroscopy (Thenkabail 2000) has become widespread in precision agriculture. Imaging spectroscopy is the simultaneous acquisition of a large number of usually contiguous, narrow spectral bands and is sometimes referred to as hyperspectral remote

sensing (Thenkabail 2000; Mohan and Porwal 2015; Overgaard et al. 2013a; Overgaard et al. 2013b). Imaging spectroscopy is often performed using a field spectroradiometer instead of a sensor mounted on a plane or satellite platform. When using a field spectroradiometer, reflectance is measured within a cone of light reception by a sensor, and the reflectance values for many narrow wavelengths can be charted to produce what is known as a spectral curve (Jensen 2016). The spectral curve (Figure 1.1) produced is dependent upon the physical properties of the object being analyzed, within the footprint of light reception, and results in a unique spectral signature, like a fingerprint, for all phenomena capable of reflectance, transmittance, and absorption (Rabideau et al. 1946; Seigal and Howell 2002; Shaw and Burke 2003; Thenkabail 2000).

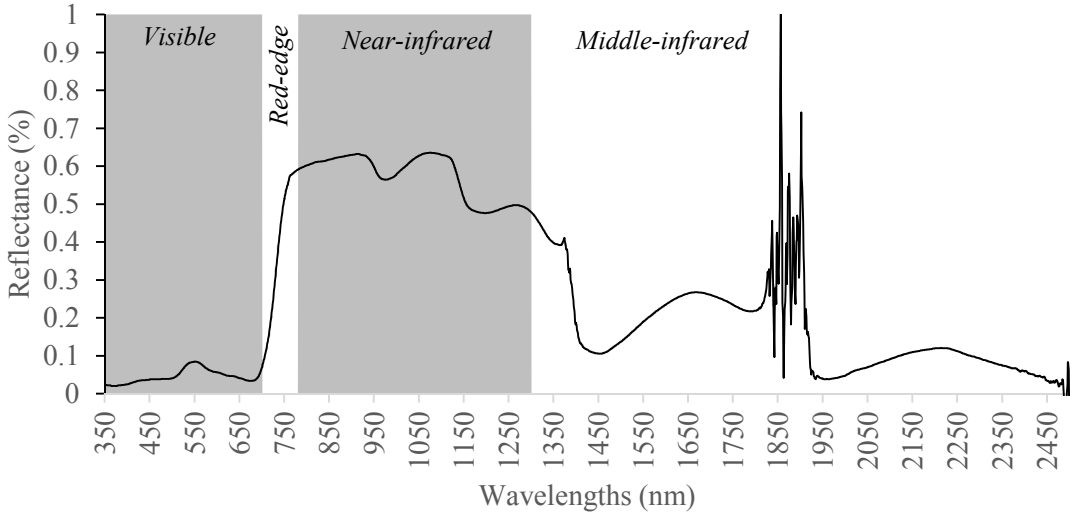

Figure Error! No text of specified style in document..**1.** A sample spectral curve of Eragrostis tef with labels for the visible, red-edge, near-infrared, and middle-infrared portions of the spectrum. Data collected by author (2017).

For plants, the reflectance, absorption, and transmittance of light can be complex due to the structure of leaves having many different overlapping layers (Curran 1983; Calvao and

Pessoa 2015). The spectral reflectance of plants is further complicated by the structure of the leaves and the spatial organization of canopies, causing reflectance issues such as background reflectance and scattering (Homolova et al. 2013). Nevertheless, the leaves are typically the main focus for spectral imaging of plants (Daughtry and Walthall 1998), because the leaves are, in most cases, the broadest portion of the plant and are typically good indicators of plant health and nutrient content.

Major advancements in imaging spectroscopy methods began in the late 1980s and 1990s (Vane and Goetz 1988; Curran and Dungan 1989; Wessman et al. 1989; Curran et al. 1990; Dawson et al. 1999; Kokaly and Clark 1999), but it was earlier on that Hoffer (1978) identified the light in the visible spectrum (blue and red) that was absorbed by plants due to high levels of photosynthetic pigments such as chlorophylls. As field spectroradiometers improved, it was found that the NIR region of the spectrum correlated considerably with plant health (Blackburn 1998; Carter 1998; Elvidge and Chen 1995; Shibayama and Akiyama 1991; Curran et al. 1990).

The general spectral curve for healthy green plants is usually characterized by low reflectance in the visible spectrum, high reflectance in the NIR, and variable reflectance in the middle-infrared (Hoffer 1978; Curran 1983; Hardisky et al. 1983; Schneider 1984; Goward et al. 1985; Milton and Mouat 1989). These reflectance values will differ depending on the biophysical characteristics of the plant. For instance, the photosynthetic pigments present in the leaf are important in controlling reflectance of visible light, which results in absorption of blue (446-520nm) and red (630-690 nm) light, and reflectance of green (520-600 nm) light in a healthy plant. The NIR reflectance of a plant is dependent on the scattering of solar radiation in the air-cell interfaces, which controls the absorption levels within the

700 nm to 1,300 nm portion of the spectral curve. The middle-infrared reflectance is mostly controlled by the water content of the plant (Sinclair et al. 1971; Hoffer 1978; Barrett and Curtis 1982; Boyer et al. 1988), where high water content results in high absorption of the electromagnetic energy represented by the spectral curve. Most of the variation within plant reflectance is found in the NIR and middle-infrared regions, but for healthy plants, there is a large difference in reflectance between the red and NIR regions because the red portion of the spectrum is absorbed by chlorophyll, while the NIR portion is scattered and reflected based upon the physical traits (leaf abundance) of the plant being sensed. This rapid increase in reflectance between the red and NIR for healthy plant reflectance is often referred to as the 'red-edge' (Hoffer 1978; Boyer et al. 1988). This 'red-edge' falls around 0.7 and 0.75 μm (Figure 1.1), and its precise location is often correlated with canopy biophysical characteristics and photosynthetic pigments (Broge and Leblanc 2000). Further, studies have found that the red-edge travels towards shorter wavelengths as stress is induced on the plant (Boyer et al. 1988; Pinter et al. 2003).

Due to these known spectral reflectance characteristics of plants, it is common to utilize spectroradiometers and cameras in agriculture as a means of non-destructively assessing plant health (Alvaro et al. 2007; Beeri et al. 2007; Belanger et al. 2007; Feng et al. 2008; Morindo et al. 2007; Overgaard et al. 2013a; Rao et al. 2008). Agricultural field conditions often vary, therefore, imaging spectroscopy has been employed to better understand and prepare for larger scale analysis (i.e. satellite, aircraft) to analyze within-field variations (Thenkabail 2000; Overgaard et al. 2013a). Improved systems for yield and nutrient mapping of agricultural fields are highly desired as such methods have potential to

locate problem areas (i.e. low yielding/nutrient zones) of the field prior to the following growing season (Overgaard et al. 2013a).

1.3.1 Agricultural Imaging Spectroscopy for Food Crops

Specifically, in agriculture, imaging spectroscopy has generally been utilized to derive plant health, trace minerals, protein concentration, and the quality/quantity of crops, grasses, forage, and grains/fruit for many cultivation practices (Thenkabail et al. 2000; Cozzolino and Moron 2004; Apan et al. 2006; Overgaard et al. 2013a). Furthermore, these remote sensing methods employed are utilized to identify correlations to photosynthetic pigments, water content, nutrient content, and internal structure of the plant/crop (Raikes and Burpee 1998; Nellis et al. 2009). Studies concerned with these biophysical and biochemical characteristics often focus on the wavelengths best associated with biophysical properties (Thenkabail 2000), in-field chlorophyll values (Zarco-Tejada et al. 2005), trace mineral prediction (Cozzolino and Moron 2004), protein analysis (Apan et al. 2006), and quality/quantity analysis (Overgaard et al. 2013a).

Thenkabail et al. (2000) aimed to identify the spectral bands needed on satellites that were best suited for characterizing biophysical properties (health, height, and yield) of cotton, potatoes, soybeans, corn, and sunflowers in Syria. Using NDVI, OMNBR (optimum multiple narrow band reflectance), and soil-adjusted indices (indices accounting for the reflectance of the underlying soils and other plant characteristics) dependent on 490 different wavelengths within the red and NIR portions of the electromagnetic spectrum, the authors found that wavelengths within the range of 650 nm and 750 nm were most important in identifying biophysical properties of these crops (Thenkabail et al. 2000).

Studies also began to identify the effects of background soil reflectance on the spectral readings collected by spectroradiometers, prompting the development of soil-adjusted vegetation indices (Huete 1988; Qi et al. 1994; Rondeaux et al. 1996; Thenkabail 2000). Specifically, Zarco-Tejada et al. (2005) used aerial systems to correlate spectral reflectance with chlorophyll content in vines in the Algoma Region of Canada. Viticulture remote sensing presents challenges because the soil below the vine is exposed to the sensor, causing issues when utilizing indices to predict the chlorophyll content. These soil background effects are important as soils also reflect light, which is often mixed with the plant reflectance within the footprint of light reception. Thus, a challenge in imaging spectroscopy and remote sensing studies is the removal of these background affects through mathematical computations that remove the effects of the reflectance of soil and other materials underlying the plant (Zarco-Tejada et al. 2005; le Maire et al. 2004). Zarco-Tejada et al. (2005) found that broad-band indices did not have the appropriate spectral sensitivities to account for background reflectance and other variation in the environment, but the narrow band indices they tested were found to be highly correlated with chlorophyll (e.g., R^2=0.90).

In addition to plant chlorophyll content, researchers have identified a strong relationship between electromagnetic reflectance values and chemical constitutes, such as nutrient values, for some plants/crops (Curran 1989; Ebbers et al. 2002). Nutrient values can be identified through imaging spectroscopy because the molecular vibrations caused by chemical bonds absorb electromagnetic energy, particularly in the NIR region of the electromagnetic spectrum (Kokaly and Clark 1999). Thus, researchers have utilized imaging spectroscopy to estimate mineral content of crops in the field and lab (Cozzolino and Moron

2004; Apan et al. 2006) as well as nitrogen plant content assessment for fertilizer application purposes (Haboudane et al. 2002; Goel et al. 2003; Overgaard et al. 2013a).

While mineral content analysis using remote sensing methods has been successful in some cases, it has been challenging in others. For instance, Cozzolino and Moron (2004) developed a method to predict trace minerals (i.e., sodium (Na), sulphur (S), copper (Cu), iron (Fe), manganese (Mn), zinc (Zn) and boron (B)) in animal feed legumes using an in-lab spectroradiometer (400-2500 nm). The two plants tested were Lucerne (Medicago sativa) and white clover (Trifolium repens). Past research (Cornforth 1984; Mills and Jones 1996; Pinkerton et al. 1997) found differing ranges of wavelengths correspond to levels of C–H, N–H, and O–H bonds, which have been found to be the primary constituents of organic molecules within forage (Osborne et al. 1993; Coleman and Murray 1993). Cozzolino and Moron's (2004) results demonstrated the potential for using near-infrared spectroscopy (NIRS) to predict trace minerals such as B, Cu, Mn, and Zn as well as two macro elements that included Na and S, but the correlating relationships were too weak to obtain the exact amounts.

Since it can be difficult to remotely sense micronutrients in the field, some studies have focused on macronutrients such as protein. Employing imaging spectroscopy in Queensland, Australia, Apan et al. (2006) attempted to identify the spectral bands (NIR region; 935 nm and 1122 nm) correlating to leaf protein content in different parts of a wheat field. The study tested specific wavelengths since using all of the wavelengths sensed by a spectroradiometer (341 to 2500 nm) could lead to statistical overfitting. The researchers reduced wavelengths by excluding overlapping regions of the near-infrared, short wavelengths (341-399 nm), wavelengths associated with water vapor absorption (1356-1480

nm; 1791-2021 nm); and all bands beyond 2396 nm where noise is common. The spectral wavelengths in which protein was most easily identified fell in the 935 nm band (R^2=0.76) and the 1122 nm band (R^2=0.76).

1.3.2 Forage Grasses Imaging Spectroscopy

Grasses used for forage are particularly important to global food and nutrition as they support the livestock consumed by humans (Tilman et al. 2002). Within agricultural studies, a subset (Beeri et al. 2007; Rabbotnikof et al. 1995; Ruan-Ramos et al. 1999; Overgaard et al. 2013a) have focused on utilizing imaging spectroscopy to derive forage quality of C_4 (plants that perform well in warmer temperatures due to the type of photosynthesis performed) and C_3 (plants that perform well in cooler temperatures due to the type of photosynthesis performed) forage grasses. C_4 grasses are more efficient than C_3 grasses at converting solar energy into biomass, have improved water use efficiency (WUE), and greater nitrogen use efficiency (NUE). The differences in efficiency result from the internal leaf structure, which further complicates remote sensing as internal structure alters the reflectance and absorption characteristics of the incoming solar radiation. These complications result in challenges when employing imaging spectroscopy on grasses to derive quality and quantity, coupled with other complexities of grass imaging spectroscopy that include plant community distribution (i.e. degree of heterogeneity and biomass volume among vegetation being remotely sensed; Boelman et al. 2005), soil color (Gao et al. 2000), hydrology (Todd and Hoffer 1998), and topography (Kawamura et al. 2005). Nevertheless, research in monitoring forage quality using imaging spectroscopy is a key focus in remote sensing as it can aid in precision agriculture practices (Haboudane et al. 2002; Goel et al. 2003). However, while most of the

nutrient imaging spectroscopy has been focused on the leaf level, the grain level has not been one of the key foci of past research.

For instance, Rabotnikof et al. (1995) applied imaging spectroscopy methods to analyze the quality of warm-season grasses (C_4) in La Pampa, Argentina by testing the sensitivities six NIRS bands (1680, 1940, 2100, 2180, 2230, and 2310 nm). However, the authors were able to accurately identify digestibility/solubility quality for animals, otherwise known as in vitro digestibility (R^2=0.827) and crude protein content (R^2=0.918). Through utilization of a sensors with greater spectral sensitivities, beyond only six bands, the accuracy of the relationship between reflectance and plant properties could be improved. Similarly, Ruan-Ramos et al. (1999) analyzed forage for livestock to better predict P, K, Ca, and Mg using non-invasive laboratory methods. The specific wavelengths correlating to the nutrients (Table 1.1) closely relate to specific wavelengths (2250, 2325, 2350, 1350, 2210, 2410, 2325, 2350, and 2250 nm) correlating to properties such as phospholipids, protein-phosphorus bonds, amino acids, and phosphates, found in past studies (Murray and Williams 1987; De Boever et al. 1994; Osborne and Fearn 1986; Shenk et al. 1979; Valdes et al. 1985; Clark et al. 1987; Convertini et al. 1991; VaÂzquez de Aldana et al. 1995). However, this study included additional wavelengths within the 1730 to 1760 nm range as they were significant contributors to the correlations. Nevertheless, many studies have utilized wavelength ranges outside those identified in this study to delineate grass nutrients through imaging spectroscopy (Valdes et al. 1985; Clark et al. 1987; Redshaw et al. 1986; Saiga et al. 1989; Convertini et al. 1991; Linn and Martin 1991; Vâzquez de Aldana et al.1995).

Table Error! No text of specified style in document.**.1.** Spectral bands utilized to delineate nutritional content in forage, grasses, legumes, and forbs based on multiple linear regression. Data derived from: Ruan-Ramos et al (1999).

Mineral	**Mathematical Treatment**	**Calibration**		
		Wavelengths (nm)	**R^2**	**Std. Err.**
Phosphorus (P)	*log*	1128, 1172, 2188, 2292, 2308, 2336, 2352	0.72	0.26
	1st derv. log	1179, 1187, 1759, 2151, 2203, 2331, 2351	0.84	0.21
	2nd derv. log	1346, 1382, 1558, 1782, 1870, 2170, 2310	0.74	0.27
Potassium (K)	*log*	1480, 1684, 1776, 1964, 2332, 2432	0.91	1.43
	1st derv. log	1547, 1563, 1595, 1747, 2175, 2371, 2423	0.90	1.50
	2nd derv. log	1218, 1318, 1394, 1526, 1618, 2138, 2242	0.89	1.58
Calcium (Ca)	*log*	1108, 1120, 1156, 1172, 1284, 1892, 2004	0.89	0.25
	1st derv. log	1531, 1675, 1751, 1951, 2043, 2175, 2307	0.91	0.22
	2nd derv. log	1242, 1610, 1742, 1786, 1890, 1938, 1970	0.91	0.21
Magnesium (Mg)	*log*	1676, 1800, 1940, 2000, 2188, 2328, 2452	0.92	0.08
	1st derv. log	1127, 1671, 1947, 2203, 2351, 2371, 2423	0.94	0.07
	2nd derv. log	1254, 1346, 1682, 1874, 1922, 2342, 2418	0.91	0.09

1.4 OBJECTIVE

While there is a large body of research focusing on predicting plant biophysical and biochemical characteristics from imaging spectroscopy in agriculture, and specifically grasses used for forage, it is important to note that many of the studies were employed in a single region and/or location (often in a single crop field), thus limiting the applicability of findings across multiple environments and geographical contexts. Furthermore, many of the methods developed and tested lack the ability to account for environmental variances such as soil (Gao et al. 2000), hydrological (Todd and Hoffer 1998), and topographical (Kawamura et al. 2005) differences, which may cause variations in background reflectance dependent on location. Thus, a comparative study of biochemical and biophysical analysis using imaging spectroscopy is also important for generalizing results across regions. More specifically, rarely have studies investigated reproducibility and replication (R&R) for more than a single environment. Past studies have also not attempted to derive nutritional value of the grain coming from different regions.

While past studies have focused on utilizing imaging spectroscopy to delineate crude protein (Apan et al. 2006; Overgaard et al. 2013a; Rabotnikof et al. 1995; Beeri et al. 2007), micronutrients (Cozzolino and Moron 2004; Ruan-Ramos et al. 1999), and plant health (Thenkabail 2000; Zarco-Tejada et al. 2005) of forage grasses and/or grains, there are gaps in past research that include analyzing the imaging spectroscopy methods between two differing environments and the effects the differing environments have on reflectance and absorption values used to correlate to plant health and nutrients. Further, the globalization of forage crops high in nutrients have the potential to serve as sequential crops potentially alleviating

the demands of a growing population, requiring an improved understanding of where such a crop could be grown within the U.S.

This research will utilize tef (*Eragrostis tef*) as a case study as it is cultivated in many agroecologies of Ethiopia, the geographic origin of the crop. In Ethiopia, tef monitoring using imaging spectroscopy has the potential to alleviate food insecurities. Thus, this study aims to determine imaging spectroscopy methods to delineate nutrient content of the plant and the grain as well as plant health through chlorophyll detection. Furthermore, to contribute to the globalization of agriculture, a site suitability analysis for tef cultivation in the U.S. will be executed.

1.5 TEF

Tef is a C_4 grass that is grown for both human consumption and forage, making it an important and versatile crop (Miller 2014). Native to Ethiopia, tef is known for its rich nutrient content compared to other cereals (Table 1.2) and its widespread distribution as a cereal crop in Ethiopia today (Taffesse et al. 2011).Tef is widely used to produce a food staple known as *injera*, a fermented bread central to the traditional Ethiopian diet; although other uses for tef such as porridge and beers are also being explored (Gerbremariam et al. 2014; Zewdie and Muchie 2014). Four other major cereal crops are cultivated in Ethiopia including wheat, maize, sorghum, and barley, but from 2004/2005 to 2007/2008, tef accounted for 20.9% (2,337,850 ha) of the total number of hectares cultivated in Ethiopia (Taffesse et al. 2011). In the U.S., tef has recently started to be incorporated as a forage, with few examples of tef grown for grain (Miller 2014). Although the crop has been introduced to the US, its cultivation has been limited (Figure 1.2). However, it has potential to serve as a

sequential crop, meaning it is grown within the rotation of major food crops such as wheat, possibly alleviating some of the pressures of the rising demand for meat among the US population by serving as an annual forage that can be harvested during the warm months (Delgado 2005; Thornton 2010).

Table Error! No text of specified style in document.**.2.** Micro- and macro-nutrient values within teff compared to other cereals. Data derived from: Gerbremariam et al. (2014).

Component	Gluten-free cereals					Glutenous cereals		
	Teff	***Maize***	***Brown rice***	***Sorghum***	***Pearl millet***	***Barley***	***Wheat***	***Rye***
Starch (%)	73.0	72	64.3	62.9	67.0	60.6	71.0	69.0
Crude Protein (%)	11.0	8-11	7.3	8.3	11.5	11.1	11.7	7.98
Crude fat (%)	2.5	4.9	2.2	3.9	4.8	3.2	2.0	1.98
Moisture (%)	10.5	14.0	14.0	14.0	9.5	10.6	12.6	-
Ash (%)	2.8	1.4	1.4	1.6	1.7	2.4	1.6	1.72
Crude fiber (g/100 g)	3.0	-	0.6-1.0	0.6	0.5	3.7	2.0	1.56
Food energy (kJ/100g)	1406					-	1105	
Calcium (mg/100g)	165.2	48.3	6.85	50.0	46.0	34	39.45	31.5
Copper (mg/100g)	2.6	1.3	0.16	0.41	1.06	0.52	0.23	
Iron (mg/100g)	15.7	4.8	0.57	6.0		2.43	3.5	2.7
Magnesium (mg/100g)	181.0	107.9	16.88	180.0	137.0	94.3	103.5	92.0
Manganese (mg/100g)	3.8	1.0	0.36			8.97	0.95	
Phosphorus (mg/100g)	425.4	299.6	61.7	263.3	379.0	563.0	-	359.0
Potassium (mg/100g)	380.0	324.8	181.71	225.23		507.0	-	412.0
Sodium (mg/100g)	15.9	59.2	0.54	6.18		25.4	-	

Zinc (mg/100g)	4.8	4.6	2.0	2.0	3.1	2.2	1.94	3.0

CHAPTER II

REPRODUCIBILITY AND REPLICATION OF IMAGING SPECTROSCOPY METHODS BETWEEN DIFFERING AGRICULTURAL ENVIRONMENTS

Abstract

Achieving reproducibility and replication (R&R) within any scientific discipline is of utmost importance for future development of a given field. Yet, the topic of R&R has not received much attention in the field of imaging spectroscopy (IS). In particular, R&R in IS could benefit precision agriculture potentially resulting in increased efficiency of resource utilization. Thus, this study aims to investigate the reproducibility of research findings across study sites, environmental contexts, and international boundaries to determine whether the same process of IS data collection, processing, and analytical methods can be used to predict the nutrient content (Ca, Mg, protein) of *Eragrostis tef* plant and grain samples from the United State and Ethiopia. The methods incorporate the use of waveband creation, spectral preprocessing (e.g., Savisky-Golay, first derivative, and second derivative), waveband selections, and partial least square regression (PLS) with bootstrapping procedures. The results suggest high correlations for both the plant and grain in single environments, with problems of overfitting when combining environments. Additionally, results suggest that spectral preprocessing methods and wavebands selected for PLS models will differ amongst differing environments. Thus,

this research suggests for the purpose of reproducibility and accuracy, IS models aiming to predict nutrient values of agricultural products should be developed for single geographies.

KEYWORDS: R&R, hyperspectral, waveband selection, partial least squares, Ethiopia, *Eragrostis tef*, tef

LIST OF ABBREVIATIONS

ASD	Analytical Spectral Devices
Ca	Calcium
ET	Ethiopia
ET1	Debre Zeit, Ethiopia
ET2	Akaki, Ethiopia
FDR	First derivative
GPS	Global Positioning System
IS	Imaging spectroscopy
Mg	Magnesium
MIR	Mid-infrared
MLR	Multiple linear regression
NIR	Near-infrared
NLV	Number of latent variables
PCA	Principle component analysis
PCR	Principle component regression
PLS	Partial least squares regression
R&R	Reproducibility and replication
R^2 std	R^2 standard deviation
R^2_{CV}	Cross-validated coefficient of determination
$RMSE_{CV}$	Root mean squared error from cross validation
$RMSE_P$	Root mean square error of prediction
$RMSE_P$ std	Root mean square error of prediction standard deviation
SDR	Second derivative
SG	Savitsky-Golay
tef	*Eragrostis tef*
US	United States
US1 and US2	Hydro, Oklahoma
USET	Combined United States and Ethiopia
βw	Weighted regression coefficient

2.1 INTRODUCTION

Achieving reproducibility and replication (R&R) of scientific findings is critical for advancing scientific discoveries, particularly in the field of remote sensing. The topic of R&R has recently moved to the forefront of many fields of study such as economics, psychology, and medicine (Asendorph et al. 2013; Begley and Ioannidis 2015; Baker 2016; Camerer et al. 2016; Ioannidis et al. 2017) where it has widely been discovered that many studies cannot be reproduced or replicated (Ioannidis 2005; Baker 2015). Yet, the topic of R&R has not received much attention in geography, remote sensing, and the spatial sciences, where investigations tend to be observational instead of experimental or theoretical (Kedron et al., under revision). The field of remote sensing is uniquely positioned to contribute to R&R in the spatial sciences for several reasons. First, remote sensing scientists work with large datasets and often perform complex spectral and spatial manipulations of the data (Lindberg et al. 1983; Naes and Martens 1984; Lorber et al. 1987; Kawamura 2008), which makes reproducibility—where the same data and methods are used to produce the same results—difficult to achieve if processing steps are not adequately reported. Second, there is a rich archive of publicly available remote sensing data online, which permits independent investigations using the same datasets.

One sub-field of remote sensing that would benefit from R&R standards is imaging spectroscopy (IS), particularly as it is used in precision agriculture. Precision agriculture integrates technology with agricultural practices and aims to increase efficiency of resource utilization (e.g., water, fertilizer, etc.) and decrease the ambiguity of decisions required on agricultural lands that are often highly variable (Schellberg 2008). If findings from one field or study area are to be transferred into practice in

another region, the results must necessarily be replicable. However, most IS studies capture data in a single region or location (often in a single crop field) under uniform conditions (Flynn et al., under review), thus limiting the ability to replicate findings across multiple environments and geographical contexts. In addition, many studies lack explanation for environmental variances such as soil (Gao et al. 2000), hydrological (Todd and Hoffer 1998), and topographical (Kawamura et al. 2005) differences that can cause variations in reflectance dependent on location.

The objective of this study is to investigate the reproducibility of research findings across study sites, environmental contexts, and international boundaries to determine whether the same process of IS data collection, processing, and analytical methods can be used to predict the nutrient content of *Eragrostis tef* (tef), a grain that is primarily grown in Ethiopia but cultivation has recently expanded to other areas of the world. Predicting the nutrient status of plants in the field has proven difficult (Curran et al. 2001; Mutanga et al. 2003), mainly due to plant water content masking absorption values in the NIR that had been found to correlate well with biochemicals (Clevers 1999; Kokaly and Clark 1999). Background effects of soil and atmospheric absorption resulted in further challenges for in-field nutrient analysis (Curran et al. 2001). Additionally, there have not been many studies that have attempted to correlate IS data to non-milled grain (Caporaso et al. 2018). I captured complementary IS data and plant/grain samples from crops in two diverse locations (United States and Ethiopia) and tested the reproducibility of nutrient prediction across the two environments using partial least square regression (PLS). I aim to identify specific wavebands that can predict nutrient content of both plant

and grain material. I compare the results from each location separately and then combine the datasets to produce a comprehensive model.

2.2 DATA COLLECTION AND PROCESSING

2.2.1 Tef

Tef is a grass of the family Poaceae that has received very little attention from the remote sensing community despite its versatile cultivation characteristics—it is drought and heat resistant—and its high nutrient content (Flynn et al., under review). This lack of attention may be due to the fact that while tef can be cultivated across many environments, it is predominantly grown in Ethiopia, where it is the most commonly harvested crop because it produces a highly nutritious and gluten-free grain (Boe et al. 1986; Twidwell et al. 2002; Bultosa and Taylor 2004; Dekking et al. 2005; Gerbremariam et al. 2014; Hopman et al. 2008). In the United States, tef is becoming popular as a sequential forage crop for cattle and horses (Flynn, under revision), but it is only grown in a handful of locations. Its cultivation is expected to increase though, given the rise in popularity of gluten-free diets (Stallknecht 1993; Boe et al. 1986; Gerbremariam et al. 2014; Miller 2014).

2.2.2 Study Sites

The two sites for this study are located in the United States (US) and Ethiopia (ET). Within the US, I sampled two fields (US1 and US2), both located in Hydro, Oklahoma (Figure 2.1). Hydro, Oklahoma is located in the Central Great Plains ecoregion and experiences drastic temperature changes throughout all seasons but generally has cold winters (average minimums from 4 - -12°C) and hot summers (reaching greater than 38°C) with low and highly variable precipitation and humidity

rates. The two field sites are located within two miles of one another, thus the environmental characteristics were similar. Both fields had similar soils (vertisols) and were located at the same elevation (474 m). In-field spectroscopy and plant/grain samples were collected immediately prior to harvest in mid-summer 2016.

The two sites in Ethiopia (ET1 and ET2) are located in Debre Zeit and Akaki (Figure 2.1). The International Food Policy Institute separates Ethiopia into 18 agro-ecological zones based on environmental conditions (e.g., elevation, precipitation, etc.). The two Ethiopian sites are located in different agro-ecological zones (ET1: Warm Sub-Moist Lowlands, and ET2: Warm Humid Lowlands). Soil composition in both sites is similar (vertisols), but elevations are different (1919 m and 2201 m, respectively for ET1 and ET2). Both Ethiopian sites were sampled immediately prior to harvest in October 2017.

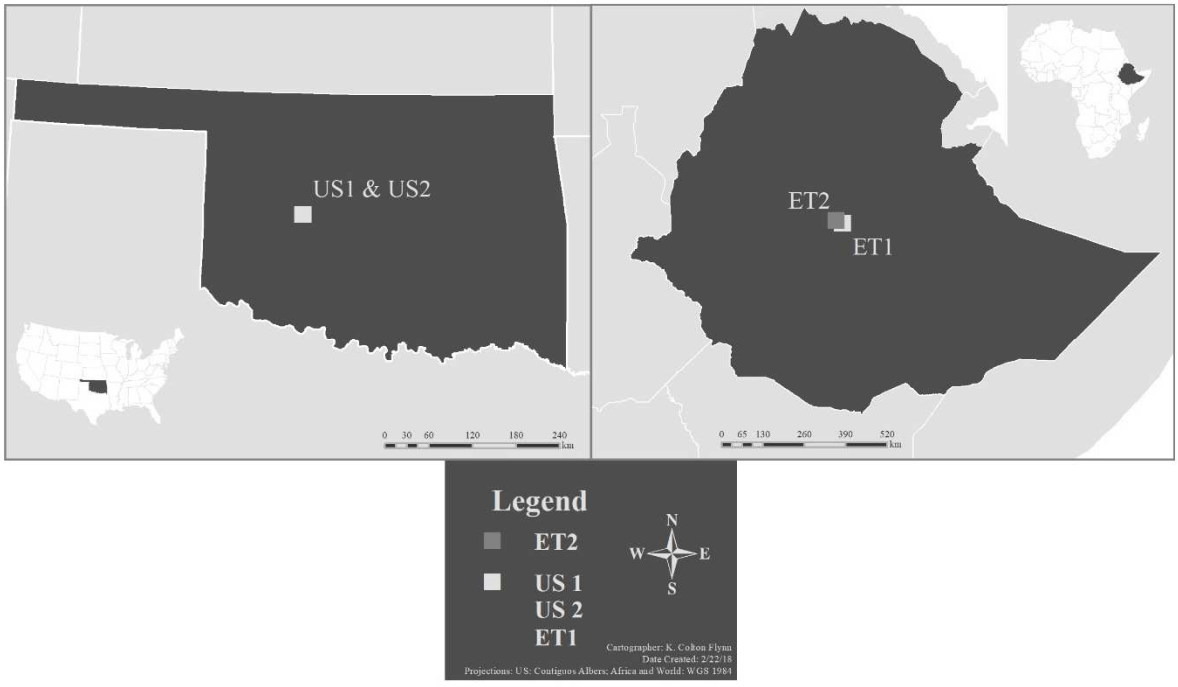

Figure 2.1. Study Site locations for both the United States (US) and Ethiopia (ET).

2.2.3 Reflectance Measurements

The process for collecting spectral data and sampling both the plant material and grain for nutrient testing is illustrated in Figure 2.2. The plant material was imaged *in situ*, while the grain material was imaged *ex sit*u after it had been separated from the parent plant in the lab.

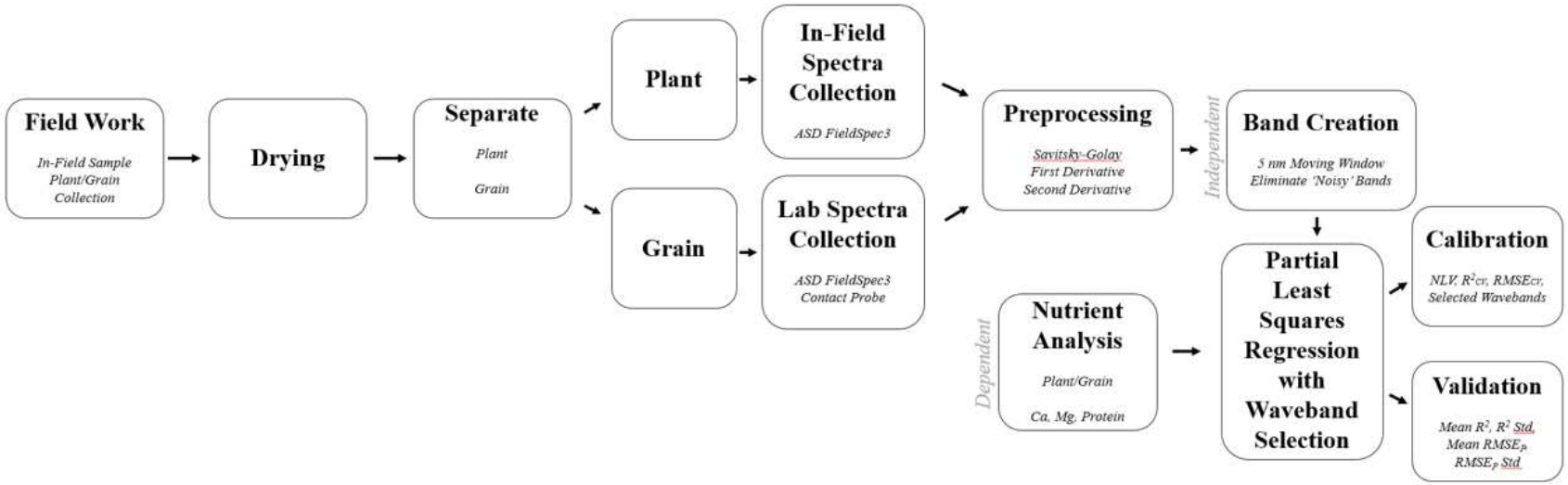

Figure 2.2. Field research methods flow chart.

Table 2.1. Number (*n*) of plant and grain samples collected in the United States (US) and Ethiopia (ET).

Plant/Grain	Nutrient	Number of Samples (*n*)	
		United States	Ethiopia
Plant	Ca	67	78
	Mg	67	79
	Protein	67	79
Grain	Ca	66	78
	Mg	66	79
	Protein	65	79

Canopy in-field spectral data were collected using a spectroradiometer (FieldSpec Pro FR: Analytical Spectral Devices [ASD], Boulder, CO), measuring reflectance from 350-2500 nm with a spectral sampling of 1.4 nm from 350-1000 nm and 2.0 nm from 1000-2500 nm. Forty random points within each field were imaged, after which samples of the plant material and grains were collected. The spectroradiometer fiber was held 1.4 m above the ground which equated to a 50.7 cm diameter circle of cover on the ground. Based on the size of sample needed (10 grams of grain; SSSA 1990; Forage Analysis Procedures 1993) for the nutrition testing, this diameter allowed for the inclusion of enough area to correspond to the amount of plant/grain matter required. Clippers were used to remove only the above ground biomass. Samples were dried to remove excess moisture. Following drying, the plant and grain were separated using the traditional methods of hand threshing and the use of a basket weaved surface. The grains were spectrally imaged in a dark room using a contact probe (Contact Probe: Analytical Spectral Devices [ASD], Boulder, CO) with a light source (Halogen bulb) emitting

spectral wavelengths (350-2500 nm) capable of being identified by the same spectroradiometer used in the field. For both canopy and in-lab methods, five spectral readings were collected for each sample, and the five readings were averaged to provide a single, representative curve for preprocessing.

2.2.4 Spectral Curve Preprocessing

The raw spectral curves from both the plant and grain IS data, were processed following Kawamura et al. (2008) and Kawamura et al. (2018) (Figure 2.2). Each spectral curve was smoothed using the Savitsky-Golay (1964) method (SG; Figure 2.3) method. First (FDR) and second (SDR) derivatives were computed from the smoothed spectra (Figure 2.3). Computing derivatives is a common practice within IS as the derivatives exploit minor difference highlighting key regions such as inflection points (FDR) and shoulder inflections (SDR). These minor changes are often difficult to acknowledge computationally and visually when data are in raw form (Kokaly et al. 2009). To further reduce noise within each spectral curve, the hyperspectral data were averaged using a moving window into 5 nm centered bands (e.g. a band centered at 600 nm would be the average value of wavelengths 598-602 nm). This step did not alter the spectral resolution of the data. Additionally, bands associated with atmospheric noise (1290-1495; 1705-2045; and 2355-2500 nm) and splicing points within the spectroradiometers (350-395 and 1005-1015 nm) were removed. This pre-processing resulted in 277 spectral wavebands between 400-2350 nm, which will ultimately serve as the independent variables for the PLS statistical analyses discussed below.

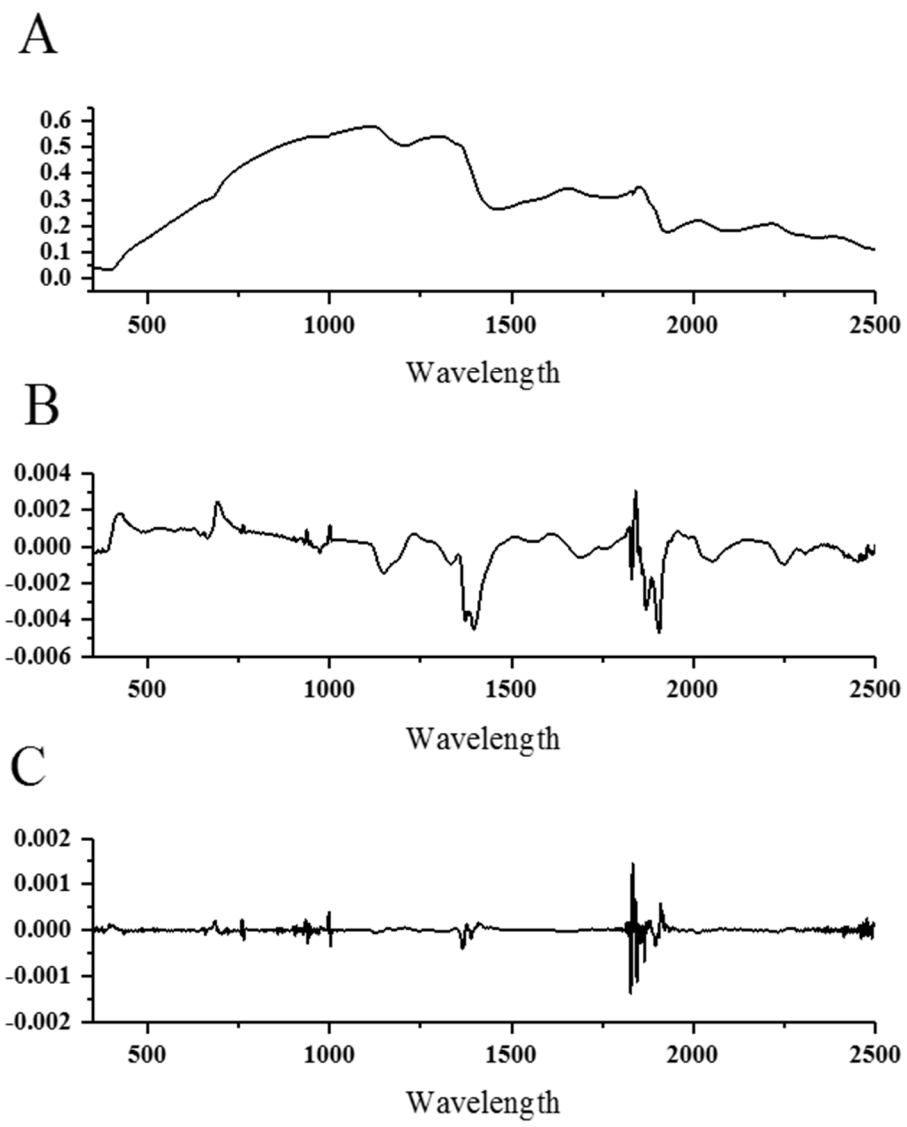

Figure 2.3. Depiction of typical canopy spectral curves of tef subject to (A) Savitsky-Golay, (B) 1st derivative, (C) 2nd derivative.

2.2.5 Nutrient Analysis

Nutrient analysis for samples was performed in the United States at the Oklahoma State University Soil, Water, and Forage Analytical Laboratory and in Ethiopia by the Ethiopian Public Health Institute of Addis Ababa, Ethiopia. The same procedures were used in both place to analyze nutrients, so the difference in processing locations is not

expected to have a significant impact on the results. Samples were analyzed for calcium (Ca), magnesium (Mg), and protein content. These measures are common in agronomic research. For details on Ca and Mg laboratory procedures, please refer to SSSA (1990). Additionally, for details on protein laboratory analysis, please refer to Forage Analysis Procedures (1993). Ca and Mg values are expressed in ppm mg/kg, while protein values are expressed in percent (%) of total sample weight. These nutrient data will serve as the dependent variable in the PLS analyses (discussed below).

2.3 ANALYTICAL METHODS

Partial least squares regression with waveband selection (PLS) is employed to assess the relationship between imaging spectroscopy data (independent variable) and nutrient content (dependent variable) of the plant and grain (Figure 2.2). PLS was selected over other types of regression because it accounts for overfitting errors that are common with other methods (i.e., multiple linear regression) when analyzing IS data (Kawamura et al. 2008; Kawamura et al. 2018). PLS standardizes the construction of models created from the preprocessed IS data, which are ultimately used to predict the nutrient levels of the plant and grain. Additionally, the construction of successful models in PLS can be tested through calibration and validation steps.

2.3.1 Partial Least Squares Regression with Waveband Selection

Multiple linear regression (MLR) is appropriate in situations where there is more than one independent variable, and those variables are not collinear. However, in situations where the independent variables are collinear, using MLR will often overfit the model. With hyperspectral data, the multiple wavebands that serve as the independent variables are often highly collinear. In this study, there are 277 individual wavebands

serving as independent variables. Thus, MLR is not appropriate. Instead, PLS can be used to construct predictive models in situations where the independent variables are collinear (Geladi and Kowalski 1986; de Jong 1993). PLS has been compared to principal component regression (PCR) because both can aid in overcoming multicollinearity and reducing dimesionality, but the two differ in the construction of the factors. PCR aims to reduce collinearity amongst the independent variables by regressing the principal components of the explanatory variables against the dependent variables instead of using the independent variables themselves. In PLS, instead of finding hyperplanes of maximum variance between the response and independent variables, the technique fits a linear regression model by projecting both the independent and dependent variables into a new space. In other words, a PLS model aims to find the multidimensional direction in *X* space that explains the maximum multidimensional variance direction in *Y* space. The PLS factors, which are often referred to as latent factors, aim to capture the variability of the dependent variable(s), often resulting in a smaller number of variables than PCR. More specifically, PLS establishes models by extracting what are called *X*-scores from the latent variables to construct a model to predict the *Y*-scores. In PLS the *X*- and *Y*-scores are subject to redundancy analysis that seeks directionality in the factor space until the most accurate prediction is found (Geladi and Kowalski 1986; Wold et al. 2001; Wu et al. 2016).

When implementing PLS with spectroscopic data, it is important to consider the number of latent variables (*NLV*) and the number of independent variables being used, as overfitting can occur in situations where the number of latent variables far exceeds the number of independent variables (Kawamura et al. 2008). Thus, based on results from

past work (Kawamura et al. 2008), I limited acceptable models to those using seven latent variables or fewer. If a model exceeded seven latent variables, I did not consider the results.

2.3.1.1 Selection of relevant wavebands

A modified form of PLS has been developed specifically for imaging spectroscopy studies known as the waveband selection method (Kawamura et al. 2008). This modification implements a waveband selection process to reduce the number of wavebands down to only those most relevant for plant material prediction before implementing the regression. This waveband selection method can be compared to stepwise linear regression in that it drops the least important wavebands (independent variables) from the creation of the latent factors. In PLS, each independent variable is assigned a weighted regression coefficient (βw) (Ramadan et al. 2001; Schmidtlein and Sassin 2004). Thus, independent variables (wavebands) with high βw have greater contribution to the models than independent variables with low βw. Using these coefficient values, the waveband selection method begins with all 277 wavebands, and following PLS methodologies described above, the waveband contributing the least to the model (lowest βw) is removed, and PLS is run again with only 276 variables. This process continues until there is one dependent variable left. After each iteration, the root mean squared error ($RMSE_{CV}$) for the prediction ability of each set of wavebands is computed as:

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i^c - y_i^c)^2}{n}} \qquad (Eq.\ 1)$$

where $\hat{y}_i^c$ represents nutrient prediction values according to the set of wavelengths (e.g., 277, 276, …, 1), y_i^c represents the actual nutrient values from the laboratory sampling,

and *n* represents the number of iterations performed (in this case, 276). The PLS model with the lowest $RMSE_{CV}$ is used to determine the relevant wavebands that will be used below. The waveband selection step does not identify a best fit PLS model.

In addition to the $RMSE_{CV}$, cross-validated coefficient of determination (R^2_{CV}), and *NLV* are also recorded to aid in the interpretation of the results (R^2_{CV}) and to collect measures for to avoid overfitting (*NLV*).

2.3.1.2 Calibration and validation

Following the identification of the optimal number of wavebands and their locations on the electromagnetic spectrum, the datasets of dependent variables (i.e., plant and grain samples; Table 2.1) are each divided into a set of calibration data and a set of validation data. The literature suggests using 65-75% for calibration and 25-35% for validation (Efron 1979). So, for the U.S. plant samples in which there were 67 total samples (Table 2.1), the set would be broken into approximately 47 for calibration and 20 for validation. A bootstrapping procedure (n=1000) with replacement was used to readily test the calibrated models (Mutanga et al. 2004; Kawamura et al. 2008; Kawamura et al. 2018). For each iteration, the set of 67 samples from the example above would be randomly resampled into a set of 47 values for calibration and a set of 20 for validation. This random resampling was completed 1000 times with a PLS model calibrated and validated each time. The performance of the validation models was assessed using root mean square error of prediction ($RMSE_P$). $RMSE_P$ is computed as:

$$RMSE_p = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i^v - y_i^v)^2}{n}} \qquad (Eq.\ 2)$$

where $\hat{y}_i^v$ represents the predicted nutrient values, y_i^v represents the measured nutrient values, and *n* represents the number of samples assigned to the validation subset. The

mean coefficient of determination (R^2), R^2 standard deviation (R^2 std), and $RMSE_P$ standard deviation ($RMSE_P$ std) were also calculated for the validation sample. The standard deviation for each measure of accuracy were calculated using the values of the n=1000 iterations recorded for both the R^2 and $RMSE_P$, respectively. R^2 serves as a measure of predictability of the models. $RMSE_P$ provides a sense of error and whether the error is within the range of the nutrient values. R^2 std and the $RMSE_P$ std both serve as measures of consistency across the many iterations within the bootstrapping procedures. PLS analyses were performed in MatLab v2016a (MathWorks, Sherborn, MA, USA) using a waveband selection package (Kawamura et al. 2018).

2.3.2 Reproducibility Across Multiple Environments

To assess the reproducibility of nutrient prediction from IS, I compared the results from the US and ET sites and also combined the datasets together (USET) to investigate the influence of a larger, more variable sample population on results. In these comparisons, I closely analyzed the spectral preprocessing and subsequent wavebands important to each nutrient prediction of each location and plant or grain. Analyses were derived using figures depicting wavebands used in each spectral preprocessing to see if the wavebands used were similar among plant/grain materials. Furthermore, I analyzed how and if the wavebands for nutrients across the different sites differed or were the same using the same figures. I also compared the performance (R^2) of the validation sets across environments and nutrient types to better understand differences and similarities of correlation between different environments. Finally, I focused on which dataset (single or combined environments) provided the least over fitting issues for the IS methodologies to better direct future studies looking to employ similar IS methods.

2.4 RESULTS

2.4.1 Descriptive Statistics

Overall, plant nutrient values for Ca, Mg, and protein were considerably higher in the US than in ET (Table 2.2). Moreover, mean Ca was five to six times higher; mean Mg was almost four times higher; and mean protein was nearly 2.5 times higher. Overall results were similar for grain samples (Table 2.2) with values in the US typically two to four times higher than those in Ethiopia. Additionally, the standard deviation and ranges of both the plant and grain in the US were higher than in ET for Ca and protein. However, the standard deviation and ranges for the Mg were relatively similar across both locations and plant/grain (Table 2.2).

Table 2.2. Descriptive statistics for each location, sample type, and nutrient measured in the laboratory.

Location	Plant/Grain	Nutrient	Descriptive Statistics					
United States			*n*	Min.	Mean	Max.	Range	SD
	Plant	Ca (ppm mg/kg)	67	3760	6651	9360	5600	1371
		Mg (ppm mg/kg)	67	1860	2753	3620	1760	327
		Protein (%)	67	5.74	15.68	23.52	17.78	5.42
	Grain	Ca (ppm mg/kg)	66	1620	2267	3240	1620	465
		Mg (ppm mg/kg)	66	1750	2015	2510	760	186
		Protein (%)	65	12.13	17.59	34.42	22.29	4.71
Ethiopia								
	Plant	Ca (ppm mg/kg)	78	437	1223	1772	1335	195
		Mg (ppm mg/kg)	79	115	740	1181	1066	297
		Protein (%)	79	3.02	5.77	9.93	6.91	1.69
	Grain	Ca (ppm mg/kg)	78	716	1283	2128	1411	460
		Mg (ppm mg/kg)	79	270	553	841	571	155
		Protein (%)	79	8.16	10.84	14.57	6.41	1.43

2.4.2 Relevant Wavebands

The wavebands selected as relevant for predicting plant and grain nutrient content showed crucial differences between the two regions, and these differences were not resolved when the US and ET datasets were combined (Figure 2.4). For example, at the plant level, the wavebands identified as relevant for prediction of Ca in the US and ET samples did not share any commonalities. For the US, wavebands in the red-edge portion of the electromagnetic spectrum (710-725, 745, 750, 760, and 765 nm) were selected while for ET, wavebands in the blue and ultra-blue portions of the electromagnetic spectrum (425 and 430 nm) were selected. Results for protein were similar, with no overlapping wavebands selected for prediction in US and ET. In ET, only three wavebands in the blue, red, and NIR regions were identified as relevant: 435, 655, and 965 nm. In the US, several blocks corresponding to the green (515-540 nm), red-edge (720-760 nm), and SWIR (1140-1165 nm) were identified. The results for Mg showed more similarities in terms of the wavebands selected as relevant. There were 12 wavebands (510-530, 720, 730-735, 930, 960, 1000, and 2250 nm; Figure 2.4) that were similar.

As for the grain level, the wavebands identified as relevant for prediction of Ca in the US and ET samples did share commonalities. There results suggested 11 wavebands (665, 705-710, 810, 825, 860, 1165-1170, and 1220-1230 nm; Figure 2.4) that overlapped. In the US, wavebands for Ca prediction spanned portions of the red (660-670 nm), red-edge (705-710 nm), TIR (1135-1190 and 1220-1230 nm), and SWIR (1670-1700 nm) of the electromagnetic spectrum. In ET, selected wavebands for Ca were within the ultra-blue (415, 450, and 456 nm), red-edge (700-740), NIR (800-815, 825, 835-840,

and 855-860 nm), and TIR (1220-1230 nm) portions of the spectrum. For Mg, there were no commonalities amongst the wavebands selected. In the US, only seven bands were selected (555, 945, 965, 970, 1000, 1655, and 1670 nm) spanning the green, NIR, and SWIR portions of the spectrum. In ET, wavebands relevant to Mg prediction included portion of the ultra-blue (410-415 nm), green (560-575, and 590 nm), red (680 nm), NIR (835-845, 890, 920, 960, and 975-985 nm). Results for protein had three wavebands (400 and 710-715 nm; Figure 2.4) that were similar for protein across US and ET waveband selection. In the US, the ultra-blue (400, 415, and 425 nm), green (555 and 570-575 nm), red-edge (710-720, and 730 nm), NIR (960 and 995-1020 nm), TIR (1060-1070, 1115, 1135, 1230, and 1235 nm), and SWIR (1665-1680 and 2125-2160 nm) portions of the electromagnetic spectrum were relevant to protein. In ET, selected wavebands for protein were within the ultra-blue (400, 440 nm), blue (440, 495, 500, and 510 nm), green (560 and 580 nm), red (660 nm), red-edge (700-715 nm), TIR (1160 nm), SWIR (1615-1650 and 2190-2195 nm) portions of the electromagnetic spectrum.

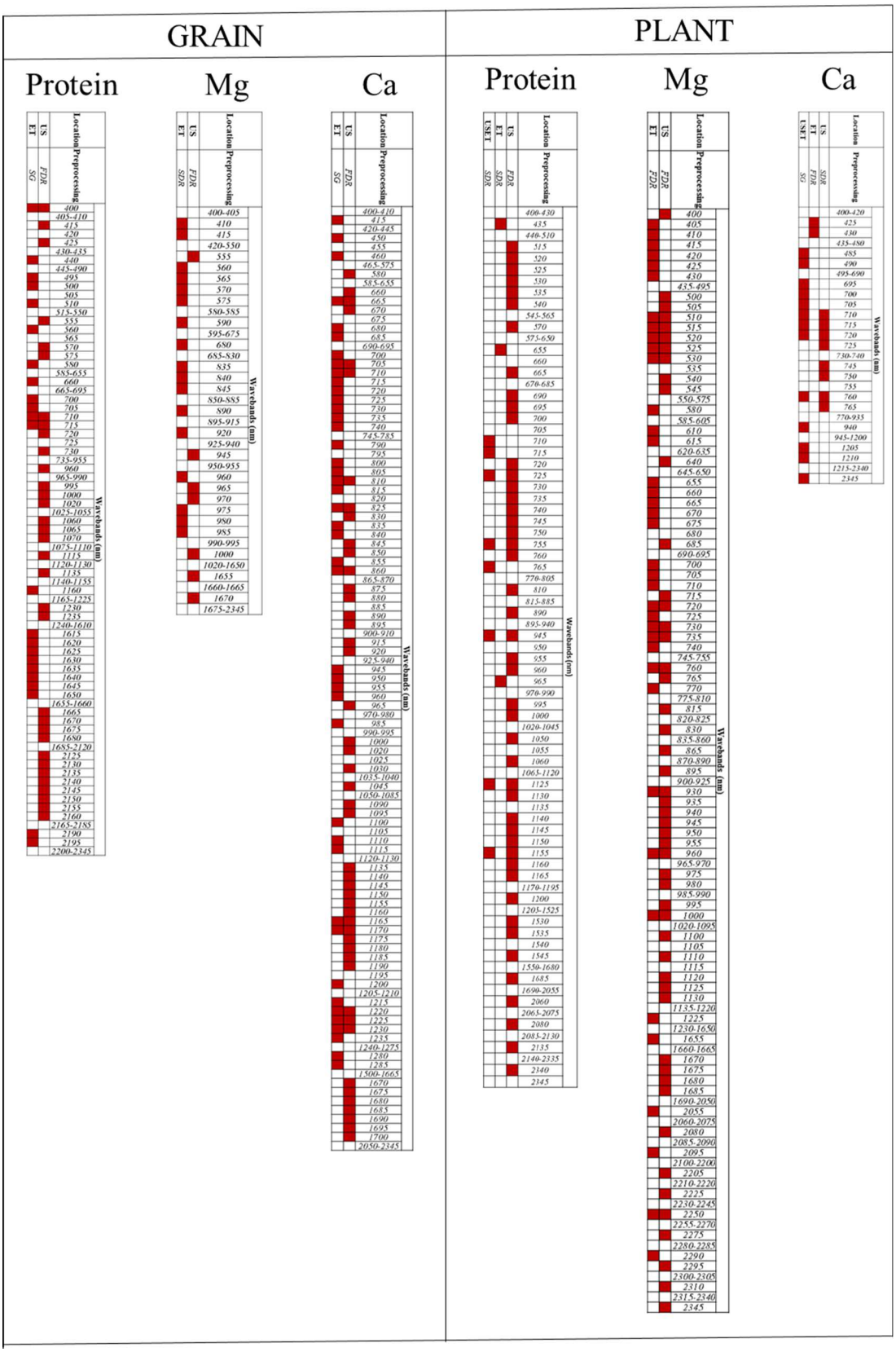
GRAIN
PLANT
Protein
Mg
Ca
Protein
Mg
Ca
Location
Preprocessing
Wavebands (nm)
SG
FDR
SDR
US
ET
USET

Figure 2.4. Selected wavelengths (shaded in red) for partial least square regression analysis for each plant/grain, nutrient, location, and spectral preprocessing. Abbreviations: SG: Savitsky-Golay; FDR: first derivative; SDR: Second derivative; NLV: Number of latent variables; US: United States; ET: Ethiopia.

2.4.3 PLS Model Results

2.4.3.1 United States

Generally, the models performed well overall with model fits ranging from 0.28-0.92 for both the plant and the grain PLS model results for plant material using the US data showed the highest mean correlation (R^2=0.90; Table 2.3) for protein content. Although the mean correlation was high (R^2≥0.88; Table 2.3) for all three spectral preprocessing methods, FDR resulted in the highest mean correlation for protein content validation using five NLV and 46 wavebands (Table 2.3).

R^2 values for Ca prediction using PLS were moderate (Table 2.3). The highest mean correlation (R^2=0.56; Table 2.3) used the SDR spectral preprocessing along with one NLV and only eight wavebands (Table 2.3). R^2 values for Mg prediction were low (R^2≤0.28; Table 2.3). Similar to protein, the spectral preprocessing FDR correlated best with Mg using four NLV and 50 wavebands. Additionally, upon review of the distribution of the R^2 values in the bootstrapping procedures both Ca and protein resulted in normal distributions (Figure 2.5) supporting the validity of the methods. However, when observing the distribution of the R^2 values in the bootstrapping procedures for Mg, the distributions tended to skewed (Figure 2.5), though the spectral preprocessing model that resulted in the highest coefficients of determination (FDR) did near normal distribution (Figure 2.5).

Table 2.3. United States (US) partial least square regression (PLS) calibration and validation results for the plant and grain samples. Bolded values represent the best spectral preprocessing, model, and correlation for the individual nutrient. Abbreviations: SG: Savitsky-Golay; FDR: first derivative; SDR: Second derivative; NLV: Number of latent variables; std: Standard deviation.

Plant/Grain	Nutrient	Spectral Processing	Calibration				Validation			
			NLV	R^2_{CV}	$RMSE_{CV}$	# Waves	Mean R^2	R^2 std	Mean $RMSE_P$	$RMSE_P$ std
Plant	**Ca**	*SG*	3	0.56	899.24	9	0.55	0.10	954.24	113.58
		FDR	1	0.53	929.65	9	0.55	0.10	937.88	125.67
		SDR	**1**	**0.53**	**928.94**	**8**	**0.56**	**0.09**	**927.72**	**113.39**
	Mg	*SG*	2	0.08	312.41	121	0.10	0.11	323.94	56.66
		FDR	**4**	**0.30**	**277.97**	**50**	**0.28**	**0.14**	**296.41**	**48.51**
		SDR	5	0.41	256.43	61	0.25	0.14	316.29	59.74
	Protein	*SG*	6	0.88	1.84	46	0.88	0.04	1.93	0.29
		FDR	**5**	**0.92**	**1.53**	**46**	**0.90**	**0.04**	**1.78**	**0.28**
		SDR	12	0.91	1.63	36	0.87	0.04	2.05	0.28
Grain	**Ca**	*SG*	8	0.90	149.80	14	0.88	0.03	169.35	25.42
		FDR	**2**	**0.91**	**142.23**	**47**	**0.89**	**0.04**	**156.88**	**25.36**
		SDR	3	0.89	152.54	220	0.88	0.04	164.27	26.66
	Mg	*SG*	4	0.73	96.74	77	0.73	0.10	102.28	22.78
		FDR	**4**	**0.78**	**86.09**	**7**	**0.78**	**0.06**	**90.85**	**11.74**

	SDR	1	0.74	93.87	174	0.75	0.08	96.72	17.86
Protein	*SG*	12	0.95	1.00	74	0.94	0.03	1.17	0.29
	FDR	**4**	**0.93**	**1.24**	**33**	**0.92**	**0.03**	**1.43**	**0.33**
	SDR	3	0.94	1.19	73	0.93	0.03	1.28	0.25

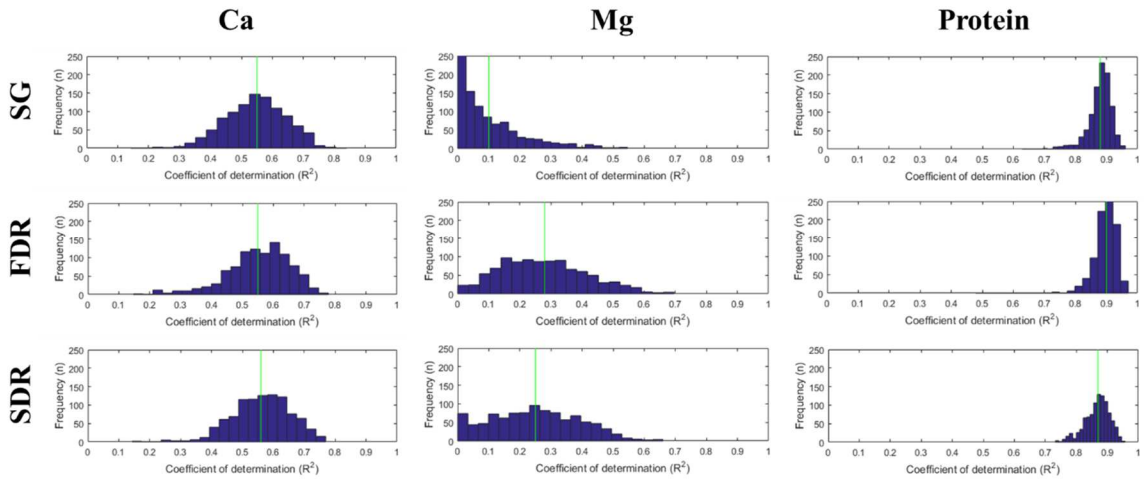

Figure 2.5. United States (US) plant bootstrapping (n=1000) validation histograms. The vertical green line represent the position of the mean R^2. Abbreviations: SG: Savitsky-Golay; FDR: first derivative; SDR: Second derivative.

At the grain level, PLS model results for all nutrients showed high correlations (R^2≥0.78; Table 2.3). For all three nutrients, the FDR resulted in the highest mean correlations without exceeding the stipulation for overfitting (NLV≤7; number of wavebands≤50). PLS model results for protein, again, resulted in the highest mean correlation (R^2=0.92; Table 2.3); closely followed by the high mean correlation of Ca (R^2=0.89; Table 2.3). The PLS model results suggest that Mg correlated the lowest to the FDR spectral preprocessing (R^2=0.56; Table 2.3), but the model did use the least number of wavebands of the three nutrients (7 wavebands; Table 2.3). Furthermore, upon review of the distribution of the R^2 values in the bootstrapping procedures all three nutrients resulted in normal distributions (Figure 2.6) supporting the validity of the methods.

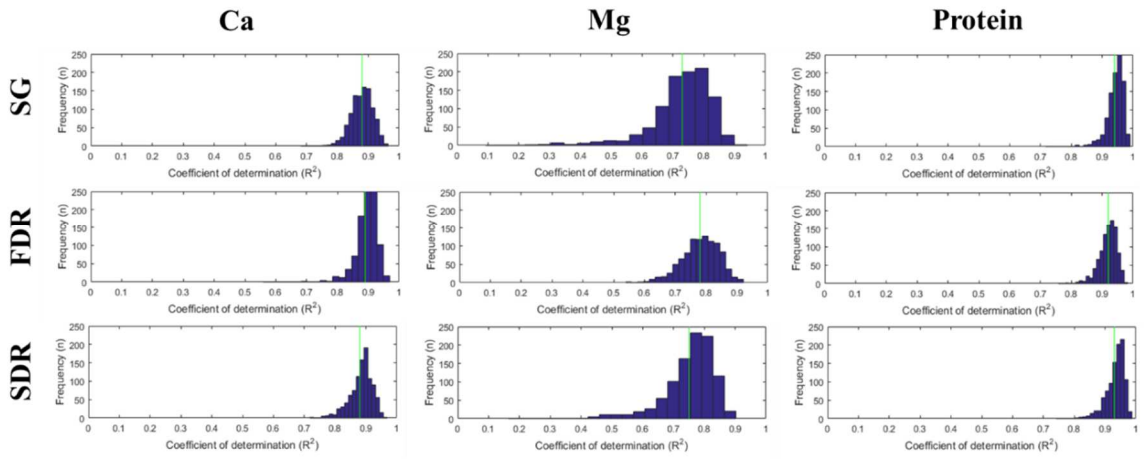

Figure 2.6. United States (US) grain bootstrapping (n=1000) validation histograms. The vertical green line represent the position of the mean R^2. Abbreviations: SG: Savitsky-Golay; FDR: first derivative; SDR: Second derivative.

2.4.3.2 Ethiopia

In ET, at the plant level, the model fit well as indicated by an R2 value of 0.92 for Mg (Table 2.4). The model used FDR preprocessing along with five NLV and 38 wavebands. The other PLS model results for Ca and protein did not correlate as well (R^2=0.25 and R^2=0.57, respectively; Table 2.4) when using the FDR (Ca) and SDR (protein) spectral preprocessing methods. Both of the models required only one NLV along with two wavebands for Ca and three wavebands for protein (Table 2.4). Additionally, upon review of the distribution of the R^2 values in the bootstrapping procedures both Mg and protein resulted in normal distributions (Figure 2.7) supporting the validity of the methods. However, when observing the distribution of the R^2 values in the bootstrapping procedures for Ca, a bimodal distribution was observed (Figure 2.7). Upon further investigation of the original ET Ca data the distribution deviated from a normal distribution and instead was both bimodal and skewed to the right, explaining the bimodal and skewed distribution of the R^2 bootstrapping values (Figure 2.7).

Table 2.4. Ethiopia (ET) plant and grain results for the partial least square regression (PLS) calibration and validation for nutrient analysis. Bolded values represent the best spectral preprocessing, model, and correlation for the individual nutrient. Abbreviations: SG: Savitsky-Golay; FDR: first derivative; SDR: Second derivative; NLV: Number of latent variables; std: Standard deviation.

Plant/Grain	Nutrient	Spectral Processing	Calibration				Validation			
			NLV	R^2_{CV}	$RMSE_{CV}$	# Waves	Mean R^2	R^2 std	Mean $RMSE_P$	$RMSE_P$ std
Plant	**Ca**	*SG*	2	0.12	226.93	49	0.23	0.17	221.46	74.21
		FDR	**1**	**0.11**	**228.37**	**2**	**0.25**	**0.19**	**219.72**	**75.70**
		SDR	2	0.09	231.79	7	0.17	0.16	227.00	71.00
	Mg	*SG*	10	0.89	100.04	15	0.88	0.03	109.74	14.80
		FDR	**5**	**0.92**	**84.14**	**38**	**0.92**	**0.02**	**88.49**	**12.98**
		SDR	3	0.91	88.93	69	0.90	0.03	98.52	14.72
	Protein	*SG*	3	0.45	1.25	31	0.46	0.11	1.32	0.12
		FDR	1	0.42	1.28	166	0.47	0.12	1.31	0.16
		SDR	**1**	**0.51**	**1.18**	**3**	**0.57**	**0.09**	**1.18**	**0.11**
Grain	**Ca**	***SG***	**7**	**0.81**	**200.20**	**43**	**0.79**	**0.06**	**223.68**	**28.61**
		FDR	12	0.93	122.84	44	0.92	0.02	142.35	20.65
		SDR	2	0.87	166.65	57	0.87	0.04	176.02	26.79
	Mg	*SG*	12	0.73	80.54	15	0.73	0.07	87.75	11.56
		FDR	8	0.87	56.29	33	0.85	0.04	63.38	9.02

	SDR	**6**	**0.81**	**67.04**	**17**	**0.81**	**0.05**	**72.36**	**8.96**
Protein	***SG***	**6**	**0.52**	**0.99**	**23**	**0.53**	**0.14**	**1.05**	**0.17**
	FDR	7	0.67	0.82	159	0.65	0.11	0.91	0.14
	SDR	8	0.81	0.62	94	0.75	0.09	0.72	0.13

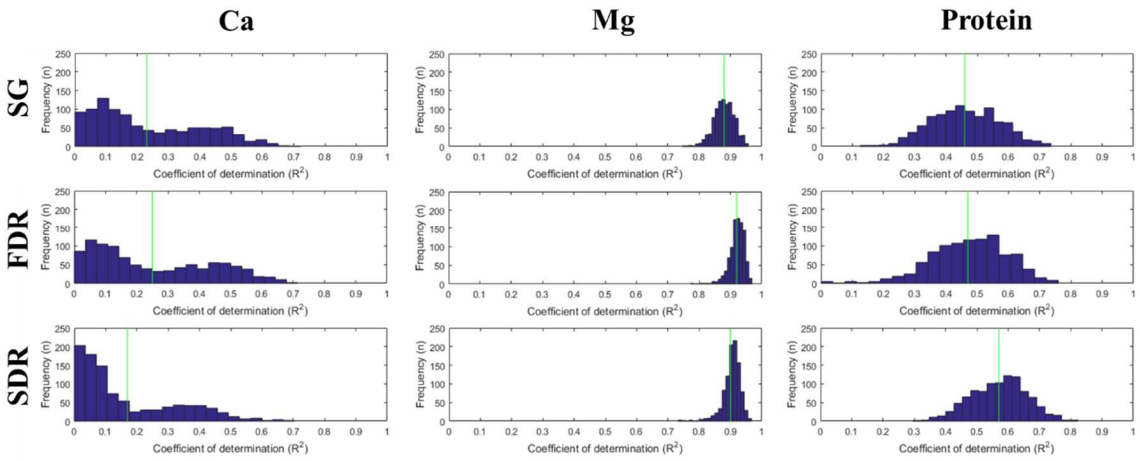

Figure 2.7. Ethiopia (ET) plant bootstrapping (n=1000) validation histograms. The vertical green line represent the position of the mean R^2. Abbreviations: SG: Savitsky-Golay; FDR: first derivative; SDR: Second derivative.

Similar to the plant level, for the ET grain the PLS model fit well as indicated by an R^2 value of 0.81 for Mg (Table 2.4). The main difference between grain and plant PLS models for Mg is the use of SDR spectral preprocessing. This PLS model required seven NLV and 43 wavebands (Table 2.4). PLS model fit for Ca was relatively close in mean correlation to that of PLS model fit results for Mg at the grain level (R^2=0.79; Table 2.4). The PLS model fit for Ca utilized the SG spectral preprocessing along with seven NLV and 43 wavebands. The PLS models fit comparatively poor for grain protein (R^2=0.53; Table 2.4). Nevertheless, the PLS model utilized six NLV and 23 wavebands from the SG spectral preprocessing. Furthermore, upon review of the distribution of the R^2 values in the bootstrapping procedures all three nutrients resulted in normal distributions (Figure 2.8) supporting the validity of the methods.

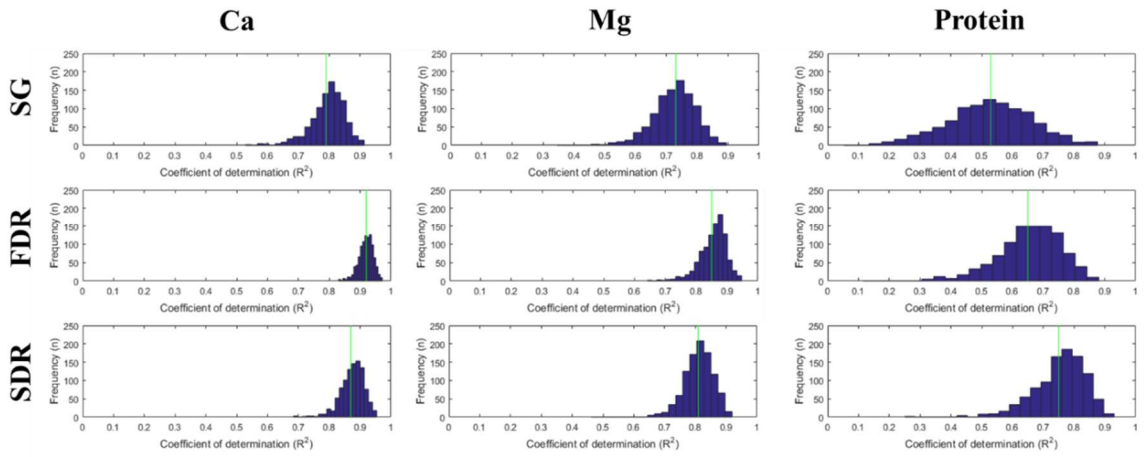

Figure 2.8. Ethiopia (ET) grain bootstrapping (*n*=1000) validation histograms. The vertical green line represent the position of the mean R^2. Abbreviations: SG: Savitsky-Golay; FDR: first derivative; SDR: Second derivative.

2.4.3.3 Combined Environments: United States and Ethiopia

The PLS models performed well for combined environments for all nutrients (Tables 2.7 and 2.8). However, only two PLS models (Ca and protein) at the plant level did not exceed the criteria for overfitting (NLV≤7; number of wavebands≤50). In this case, for Ca the model fit well as indicated by an R^2 value of 0.95 (Table 2.5). This model utilized SG spectral preprocessing along with five NLV and 13 wavebands. It is important, however, to note that the $RMSE_P$ (664.46; Table 2.5) of this PLS model is greater than some of the sample Ca levels (minimum=437 ppm mg/kg; Table 2.1) in ET, suggesting the model may not be a good fit for both locations. For protein, the model fit well indicated by an R^2 value 0.92 (Table 2.5). This model used the SDR spectral preprocessing along with six NLV and eight wavebands. This PLS model is more suitable than that of the Ca model as the $RMSE_P$ is much smaller than any of the protein results for either region. Additionally, upon review of the distribution of the R^2 values in the

bootstrapping procedures both Ca and protein resulted in normal distributions (Figure 2.9) supporting the validity of the methods.

Table 2.5. United States and Ethiopia (USET) plant and grain results for the partial least square regression (PLS) calibration and validation for nutrient analysis. Bolded values represent the best spectral preprocessing, model, and correlation for the individual nutrient. Abbreviations: SG: Savitsky-Golay; FDR: first derivative; SDR: Second derivative; NLV: Number of latent variables; std: Standard deviation.

Plant/Grain	Nutrient	Spectral Processing	Calibration				Validation			
			NLV	R^2_{CV}	$RMSE_{CV}$	# Waves	Mean R^2	R^2 std	Mean $RMSE_P$	$RMSE_P$ std
Plant	**Ca**	***SG***	**5**	**0.95**	**658.69**	**13**	**0.95**	**0.01**	**664.46**	**72.39**
		FDR	4	0.95	664.55	38	0.94	0.01	690.87	79.27
		SDR	3	0.94	698.37	146	0.94	0.01	713.82	76.80
	Mg	*SG*	9	0.94	265.07	44	0.93	0.01	276.48	31.00
		FDR	5	0.94	261.27	112	0.94	0.01	271.65	33.62
		SDR	**4**	**0.93**	**286.80**	**156**	**0.92**	**0.02**	**296.87**	**35.27**
	Protein	*SG*	5	0.92	1.78	13	0.92	0.02	1.81	0.18
		FDR	8	0.93	1.68	52	0.92	0.02	1.79	0.16
		SDR	6	0.92	1.80	8	0.92	0.02	1.85	0.18
Grain	**Ca**	*SG*	11	0.91	206.43	59	0.90	0.02	221.41	21.83
		FDR	10	0.93	173.75	59	0.93	0.02	186.35	17.83
		SDR	5	0.93	182.89	133	0.92	0.02	194.10	18.06

Mg	*SG*	12	0.98	106.71	39	0.98	0.00	114.98	11.90
	FDR	7	0.98	101.05	51	0.98	0.00	103.95	9.54
	SDR	11	0.98	97.48	22	0.98	0.00	103.03	10.01
Protein	*SG*	10	0.94	1.12	19	0.94	0.02	1.16	0.19
	FDR	9	0.95	1.04	31	0.95	0.01	1.10	0.17
	SDR	6	0.94	1.15	66	0.94	0.02	1.21	0.16

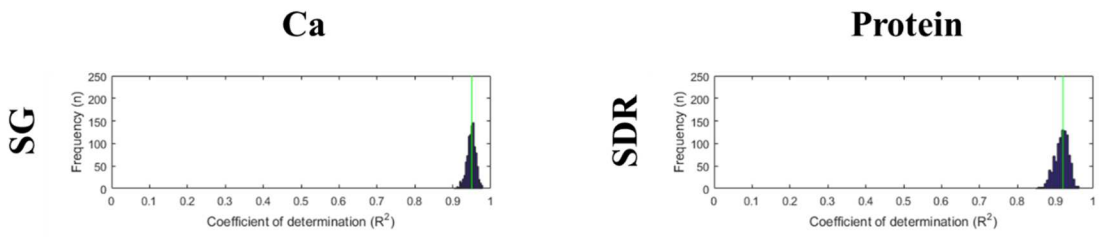

Figure 2.9. United States and Ethiopia (USET) plant bootstrapping (*n*=1000) validation histograms. The vertical green line represent the position of the mean R^2. Only two results are depicted as the rest of the results were found to have issues of overfitting. Abbreviations: SG: Savitsky-Golay; FDR: first derivative; SDR: Second derivative.

All PLS regressions for the combined environments at the grain level resulted in overfitting (NLV>7; number of wavebands>50; Table 2.5). Thus, no model was successful for predicting grain nutrients.

2.5 DISCUSSION

Biochemical properties, such as nutrient content, can be derived from plant canopies and grain samples using IS data as differing molecular interaction cause various scattering and absorption features within the electromagnetic spectrum. Thus, this study aimed to utilize IS data to predict plant nutrients and grain nutrients for tef and to test the replicability of PLS models for predicting nutrients from IS data across multiple environments in the US and Ethiopia. The implications of the findings on precision agriculture and replication of scientific results are discussed below.

2.5.1 Prediction of Plant Nutrients

As our understanding of the relationship between plant content and spectral reflectance increases, our ability to conduct quantitative modifications and applications in IS analysis for nutrient prediction is also improving. Previous studies have focused heavily on biochemical constituents such as lignin, chlorophyll, nitrogen, cellulose, and

water, but there has not been systematic exploration of these methods to predict nutrient content of plant canopies across environmental gradients (Kokaly 2009). Therefore, a major contribution of this study is the identification of important wavebands for the prediction of Ca, Mg, and protein of plant canopies in differing environments (Figure 2.4). Spectral preprocessing is an important step to reduce noisy portions of the spectral signatures and highlight changes across the spectral curve to identify these wavebands. The findings suggest that implementing FDR preprocessing can aid in identifying the most important wavebands for nutrients prediction as evidenced by four of the six single-environment analyses in which the PLS model performed best with a FDR transformation. FDR preprocessing is commonly used in remote sensing analyses to highlight inflection points across spectral curves since inflection points at specific locations along the electromagnetic spectrum are widely known to correlate with certain biochemical properties like chlorophyll (Cho and Skidmore 2006; Clevers et al. 2002). Our findings suggest that certain nutrients may also have molecular interactions that are correlated to specific inflection points (Figure 2.4). While I did not test this relationship specifically, it is a promising avenue for future research. SDR transformations also performed well for two of the nutrient in two different environments (Figure 2.4). SDR preprocessing is commonly used in remote sensing studies to pinpoint 'shoulder points' in sigmoidal shaped regions of the spectral curve. These points can often indicate wavebands where reflectance is transitioning from troughs to edges or edges to peaks. These findings suggest that in certain environments, relationships between nutrients and spectral reflectance may emerge more readily when the curves are transformed using

derivatives, and varying environmental factors should guide explorations of preprocessing decisions for future studies.

A key finding from this research is that the optimal wavebands for predicting nutrients (Ca, Mg, protein) via plant canopy measurements differed between the two environments. These differences could be due to varying levels of water content because of irrigation or non-irrigation practices (Kokaly 2009), varying field fertilizer applications (Mulla 2013), and/or latitudes and sun angles causing different rates of scattering and absorption (Jensen 2016). For instance, the ET fields were rain fed while the US fields were irrigated through to harvest. The varying plant water contents likely played a role in waveband selection as water is known to cause access noise in spectral signatures. Moreover, the number of wavebands selected in the PLS models for ET samples were often less than the number of wavebands selected for the US samples (Tables 2.3 and 2.5). This may be representative of the PLS models combating the noise caused by increased levels of water content in the US plant canopies. In summary, the finding that the optimal wavebands and regions of the electromagnetic spectrum most suited for nutrient prediction vary widely between sites has critical implications for the replicability of methods across environments. Furthermore, these findings suggest that environmental impacts on reflectance may be much greater than previously assumed in IS studies, prompting the need for further research into methods to remove these effects.

2.5.2 Prediction of Grain Nutrients

The relevant wavebands for predicting grain nutrient content varied between the two environments (Figure 2.4). Due to the controlled environment that the IS data for the grain samples were collected in, we can assume that these differences likely were not the

result of external factors (i.e. sun angles, latitudes, etc.). However, when comparing the nutrient content of the grains between US and ET, the US grains had two to four times higher amounts of nutrients than the ET grains (Table 2.2). These large differences in nutrient content likely result in varying chemical property relationships within the grain, which in turn result in differential absorption and scattering of electromagnetic energy sensed by the IS methods. The large variance amongst biochemicals within the grain may result in noise for some nutrients as nutrient reflectance properties are often associated with near or similar portions of the electromagnetic spectrum (Kokaly 2009).

Another key contribution of this study is the use of IS for predicting nutrient content, as there have been very few studies using IS data to predict nutrient content of non-milled grains (Caporaso et al. 2018). I was able to identify wavebands important to the prediction of Ca, Mg, and protein (Figure 2.4) in the grain. However, it should be noted that the IS data for the grain samples was collected in a controlled environment (dark room using a contact probe). Thus, coefficients of determination tended to be higher compared to measurements collected of the plant canopy *in situ* (Tables 2.3-2.6). It is likely that the method of data collection for the grain samples reduced some of the background noise (e.g., from soil) that can affect *in situ* imaging spectroscopy data, and modeling techniques may perform better when data are collected in controlled conditions.

Spectral preprocessing and PLS model results for the prediction of grain nutrients again suggest that FDR is an appropriate transformation technique for relating IS data to nutrient content. In this case, three of the six models performed best when the data were transformed with a first derivative. The three models that performed best with FDR data came from the same environment (US; Figure 2.4). In Ethiopia there was a mix of

preprocessing methods (two SG and one SDR) that resulted in the best fit PLS models (Figure 2.4). These differences in best performing spectral pretreatments were somewhat expected given results in prior studies, although the reasons for these differences are not well understood (Caporasao et al. 2018; Kokaly et al. 2009; Agelet et al. 2012). For instance, when comparing independent results of studies such as Caporaso et al. (2018) and Agelet et al. (2012), which performed similar analyses with grains from one location (namely United Kingdom and Iowa, US, respectively), the spectral preprocessing transformations identified as correlating best to the nutrient (protein) levels were also different. Thus, analysis of these inconsistencies in spectral preprocessing performance is warranted for future studies exploring IS data implementation for grain nutrient content analysis.

2.5.3 Replicability of Models Across Differing Environments

Studies exploring the use of IS data to predict biochemical properties are numerous (Phan-Thien et al. 2011; Martinez-Valdivieso et al. 2014; Caporaso et al. 2018; Agelet et al. 2012), but most of these studies lack any analysis of the replicability of prediction models for different crops or in differing environments. These results suggest that the replicability of PLS models across differing environments may not be the best practice as the models created to correlate IS data to nutrient content had varying spectral preprocessing, selected wavebands, and resulted in varying coefficients of determination (Mean R^2). Additionally, when synthesizing samples and their corresponding data for two differing environments, the PLS models created generally result in overfitting. Thus, considerations should be taken in future research when combining samples from differing

regions that may be experiencing varying environmental factors, agricultural practices, and large disparities in nutrient content.

2.6 CONCLUSION

This study aimed to investigate the R&R of an IS method to predict nutrients (e.g. Ca, Mg, protein) between two environments (e.g. US and ET) for *Eragrostis tef* at the plant and grain levels. The IS methods employed included waveband creation, the use of spectral preprocessing (e.g. SG, FDR, SDR), and the PLS waveband selection method (Kawamura et al. 2008; Kawamura et al. 2018). Results suggest that at both the plant and grain level the reproducibility of models created are best when developed in a single location as models created often incorporate differing spectral preprocessing methods, waveband selections, and results in differing coefficients of determination (Mean R^2). Additionally, combining environments generally results in overfitting of models using the same methods. Thus, this research suggests for the purpose of reproducibility and accuracy, IS models aiming to predict nutrient values of agricultural products should be developed for single geographies.

CHAPTER III

TRANSLATING HYPERSPECTRAL INDICES TO MULTI-SPECTRAL SENSORS FOR CHLOROPHYLL PREDICTION AT SAMPLED LOCALITIES IN ETHIOPIA AND OKLAHOMA, USA

Abstract

As remotely sensed data becomes more readily available around the world, satellites such as Landsat-8 and Sentinel-2 have great potential to support precision agriculture. In particular, sensors with high spectral and spatial resolutions are optimal for smallholder farmers to improve land management. The objective of this study is to translate chlorophyll prediction indices that are typically computed using ground-based hyperspectral data into a form that can be captured with multispectral imagery. I then test the performance of those indices across two differing environments. The methodological approach is tested for tef (*Eragrostis tef*), an endemic grass crop native to Ethiopia that forms a major component of Ethiopian diets and is grown by smallholder farmers. It is also grown commercially by farmers within the United States. Hyperspectral reflectance data captured in situ at the canopy level were convolved into bands matching the Landsat-8 and Sentinel-2 instruments, and all three sets of data were used to compute a set of commonly-used chlorophyll prediction indices. Results show that simple pigment

indices using visible bands perform best for predicting chlorophyll when translated to multispectral imagery. Specifically, the Blue/Red Index showed the highest correlations for total chlorophyll (*a*+*b*) across the three datasets. The red-edge index also performed well. These findings suggest that publicly available, multispectral imagery can potentially substitute for hyperspectral data, thereby improving the accessibility of precision agriculture methods for smallholder farmers.

3.1 INTRODUCTION

Precision agriculture has been advancing global food security and production through improvements in crop mapping, phenological analysis, pest/weed management, nutrient analysis, and crop health analysis (Calvao and Pessoa 2015; Oliver et al. 2013), while also helping to decrease malnutrition (Gupta et al. 2014). Methods that measure chlorophyll are particularly useful, since crop health is often associated with plant chlorophyll content, and natural and anthropogenic stressors can cause fluctuations in chlorophyll (Carter 1994; Lichtenthaler 1998). Additionally, chlorophyll correlates directly with nitrogen content, and thus, photosynthesis (Evans 1989; Field and Mooney 1986; Niinemets and Tenhunen 1997; Yoder and Pettigrew-Crosby 1995). However, analyzing in situ chlorophyll content can be time-consuming, expensive, destructive, and often requires laboratory resources. Remote sensing methods provide a non-invasive alternative for chlorophyll prediction and can produce instantaneous results in some situations (Martinez-Valdivieso et al. 2014), but many methods have not been universally tested, thus limiting their effectiveness.

In particular, imaging spectroscopy, in which a spectrum of reflected energy from the plant target is collected by a spectroradiometer, has been utilized to derive plant health traits since the 1980s (Thenkabail et al. 2000; Cozzolino and Moron 2004; Apan et al. 2006; Overgaard et al. 2013a; Raikes and Burpee 1998; Nellis et al. 2009; Vane and Goetz 1988; Curran and Dungan 1989; Wessman et al. 1989; Curran et al. 1990; Dawson et al. 1999; Kokaly and Clark 1999). Curran (1989) identified specific portions of the electromagnetic spectrum that correlated strongly with foliar chemistries. Building on these findings, researchers began developing indices to predict leaf- and canopy-level chlorophyll content from remotely sensed data (Huete 1988; Qi et al. 1994; Rondeaux et al. 1996; Lyon et al. 1998; Thenkabail 2000; see le Maire et al. 2004 for a review). These indices typically incorporate portions of the electromagnetic spectrum (e.g., visible, near- and mid-infrared) that are sensitive to chlorophyll content (Gitelson and Merzlyak 1994; Gitelson and Merzlyak 1996; Lichtenthaler et al. 1996) and correlate with key plant components (Xue and Su 2017). Additionally, many of these indices are able to account for confounding factors such as background soil reflectance (Rondeaux et al. 1996; Qi et al. 1994), differing plant types (Haboudane et al. 2004; Haboudane et al. 2002; Daughtry et al. 2000; Broge and Leblanc 2000; Rougean and Breon 1995), and varying chlorophyll types (i.e. chlorophyll a/b; Zarco-Tejada 2005).

However, one major drawback to using imaging spectroscopy is the cost of hyperspectral sensing equipment, which is often too expensive for farmers, especially smallholders. Thus, exploring the potential for incorporating publicly available datasets (e.g., Sentinel-2, Landsat-8) may lead to more applicable methods for determining crop health globally. In particular, Sentinel-2 has relatively high spatial (10, 20, and 60m),

temporal (5 days), and spectral (13 bands, including red-edge) resolutions (Bari et al. 2014; ESA 2018) that motivate exploration of whether it can provide a substitute for previously-developed hyperspectral indices. If properly employed, satellite-based proxies of chlorophyll prediction indices can provide affordable agricultural monitoring (Oliver et al. 2013; Gupta et al. 2014).

Due to the high costs of equipment and the time required for *in situ* data collection, hyperspectral indices are often developed based on data from a single study area (Cho and Skidmore 2006; Zarco et al. 2005; Thenkabail et al. 2000), and thus may not be appropriate for differing environments or management practices (i.e., irrigated and non-irrigated lands). Despite wide application of chlorophyll prediction indices in the developed world (e.g., Australia, Denmark, Spain, France, United States, etc.), methods have not been widely tested or adopted in the developing world (Mulla 2013). In developing areas, farming practices often differ in terms of fertilizer use, irrigation, and harvesting practices, to name a few, which may lead to different relationships between plant biochemical characteristics and spectral reflectance.

The objective of this study is to determine whether chlorophyll prediction indices previously developed using hyperspectral data can be mimicked using multi-spectral imagery to successfully predict a wide range of chlorophyll content in plants. This study analyzes hyperspectral data captured in four fields of planted tef, two in Ethiopia and two in Oklahoma, United States. I synthesize in situ hyperspectral data collected at each site to mimic Landsat-8 and Sentinel-2 sensors and test the performance of various chlorophyll prediction indices across all three datasets.

3.1.1 Study Areas and Cultivation Practices

Eragrostis tef (tef) is the most commonly cultivated grain among Ethiopian smallholder farmers (Taffesse et al. 2011) and is used to produce injera, a fermented bread central to the traditional Ethiopian diet (Gerbremariam et al. 2014; Zewdie and Muchie 2014). Tef has received little attention in the literature, but it is growing in popularity worldwide because the grain is highly nutritious and gluten free (Gerbremariam et al. 2014), making it an optimal case study for analyzing the predictive power of chlorophyll indices.

In the United States, the two sampled sites (US1 and US2) are located in Hydro, Oklahoma (Figure 3.1). Both sites have similar soil composition (vertisols), are located at the same elevation (Table 3.1), and were sampled during peak crop maturity in summer 2016. In Ethiopia, the two sampled sites are located in Debre Zeit (ET1) and Akaki (ET2) (Figure 3.1). The International Food Policy Research Institute (IFPRI 2006) recognizes 18 agro-ecological zones within Ethiopia. Each zone has environmental conditions (i.e., elevation, precipitation, etc.) that result in differing agricultural land use practices (IFPRI 2006). The E1 site is located in the Warm Sub-Moist Lowlands, and the E2 site is located in the Warm Humid Lowlands (E2). Both farms have similar soil composition (vertisols) but different elevations (Table 3.1). Both E1 and E2 were sampled at peak maturity in October 2017.

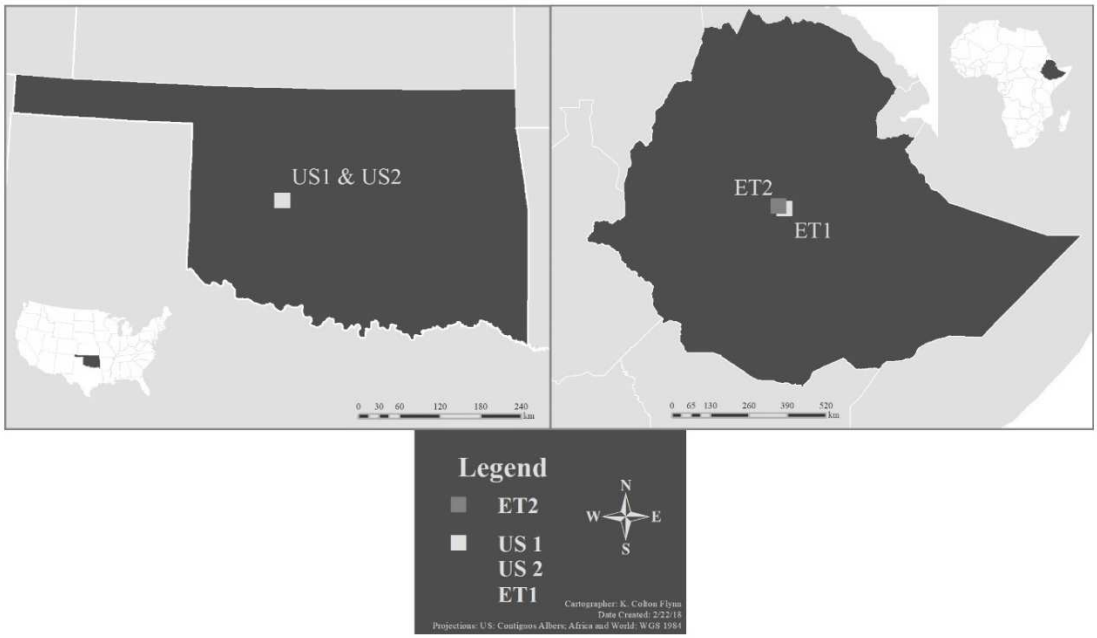

Figure 3.2. Locations of the study sites in the United States (US) and Ethiopia (ET).

Table 3.3. Study site characteristics.

Site	Sowing Date	Harvest Date	Data Collection Date	# of Samples	Ecoregion/ Agroecology	Elevation (m)	Size (ha)
US1	4/18/16	6/23/16	6/21/16	40	Central Great Plains	474	21.45
US2	4/28/16	7/5/16	7/5/16	27	Central Great Plains	474	24.19
ET1	7/26/17	12/23/17	10/25/17	40	Warm Sub-Moist Lowlands	1919	0.77
ET2	8/1/17	12/30/17	10/27/17	40	Warm Humid Lowlands	2201	1.23

3.2 METHODS

3.2.1 Data Collection and Chlorophyll Extraction

Canopy level spectra were collected at each site using a spectroradiometer (FieldSpec Pro FR: Analytical Spectral Devices [ASD], Boulder, CO), which measures spectral reflectance from 350-2500 nm with a sampling width of 1.4nm from 350-1000 and 2.0 nm from 1000-2500 nm. For all sites, 40 random points were selected for spectra and sample collection. Spectra were collected from 1.2 m above ground with a 25-degree cone of acceptance, leading to a 0.53 m diameter footprint for data collection. Following spectra collection, a representative sample of leaves were collected from within the sensor field of view (Figure 3.2). Samples were stored in plastic bags and placed on ice in a cooler for transportation back to the lab.

Figure 3.3. Depiction of *Eragrostis tef* (tef) in the field.

Laboratory methods for chlorophyll extraction followed Cole-Parmer (2014) whereby a 0.1 g sample was used for chlorophyll analysis. Tef blades are too narrow to utilize the common hole punch methodology (Figure 3.2), therefore, weighed samples were used to meet the mass requirement. After extracting the 0.1 g sample, the remainder of the sample was weighed, dried for 24 hours at 60℃, and reweighed to obtain a dry weight. The 0.1 g sample was pulverized using a mortar and pestle with acetone to extract chlorophyll. The sample and acetone were put in a sealed test tube that was placed in a centrifuge. After the centrifuge, the sample was filtered using a Whatman 11 nm filter that allowed the chlorophyll-saturated acetone to pass through while simultaneously filtering grass particles. Once filtered, the sample was diluted prior to being placed in a spectrophotometer (Jenway, Cole-Parmer, Beacon Road, Stone, Staffordshire, ST15 OSA, UK; measuring absorption level readings of liquids from 320-1100 nm with a sampling width of 0.1 nm). Cole-Parmer (2014) suggests utilizing absorption values obtained from a spectrophotometer at both the 662.6 nm and 645.6 nm wavelengths. Absorption values (A) at each of these wavelengths were utilized to calculate chlorophyll *a* and *b* concentrations:

$$C_a = 11.75\, A_{662.6} - 2.35\, A_{645.6} \qquad (Eq.\ 3.1)$$

$$C_b = 18.61\, A_{645.6} - 3.96\, A_{662.6} \qquad (Eq.\ 3.2)$$

where A662.6 is the absorption reading at 662.6 nm, and A645.6 is the absorption reading at 645.6 nm.

Combined (a+b) chlorophyll content was computed as:

$$C_{a+b} = C_a + C_b \qquad (Eq.\ 3.3)$$

Chlorophyll measurements were computed in mg/L and converted to g Chl/g (hereafter noted as g/g) mass using the mass measurements described above.

3.2.2 Computational Analyses

To better evaluate the effectiveness of chlorophyll indices reflecting the variation of chlorophyll content in a large range, it is necessary to include samples with a wide range of chlorophyll content. As such, the sites in Ethiopia and the United States are managed differently in terms of irrigation and fertilization. Therefore, a *t*-test (α=0.05) was conducted to determine whether there were significant differences between the Ethiopia and US sites in terms of chlorophyll *a* and *b* and total chlorophyll (*a*+*b*). Significant results will indicate that the samples are drawn from different populations and therefore include a wide range of chlorophyll values. Next, a set of chlorophyll prediction indices identified in the literature (Table 3.2) was computed using three separate data sources: (1) the hyperspectral reflectance data (Hy); (2) Landsat-8 Operational Land Imager (L8) bands synthesized from the hyperspectral data; and (3) Sentinel-2 Multispectral Instrument (S2) bands synthesized from the hyperspectral data. For the L8 and S2 synthesized bands, hyperspectral wavelengths were convolved to match each sensor's band ranges (Table 3.3). Actual L8 and S2 imagery could not be used due to the lack of suitable imagery at the time of peak crop maturity. The indices are grouped into six types based on their naming conventions in the literature. Certain indices (e.g., chlorophyll absorption) can only be computed using hyperspectral data and therefore do not have multispectral proxies (Table 3.2). However, these indices were included in the analysis of the hyperspectral data for completeness. Similarly, I included several indices that were developed directly from multispectral imagery (e.g., NDVI) for comparison.

To test the prediction power of each index, the index value (dependent variable) was regressed against chlorophyll content (*a*, *b*, *a*+*b*), which served as the independent variable, using ordinary least squares (OLS) and fifth-order polynomial following Cho and Skidmore (2006) and Pu et al. (2003). Fifth-order polynomial regression was included because the relationship between chlorophyll content and chlorophyll indices is often discontinuous (Cho and Skidmore 2006; le Maire et al. 2004), and this type of regression model can account for those discontinuities. Relationship strength between the dependent and independent variables was assessed using the proportion of variability (adjusted R^2) and root mean square error (RMSE).

Table 3.4. Commonly used chlorophyll prediction indices computed using hyperspectral (Hy) data and synthesized Landsat-8 OLI (L8) and Sentinel-2 MSI (S2) data (derived from Zarco-Tejada et al. [2005] and le Maire et al. [2004]). See Table 3.3 for band convolutions for synthesized Landsat-8 OLI and Sentinel-2 MSI.

Type	Vegetation Index	Hyperspectral (Hy) Indices	Synthesized Landsat-8 OLI (L8) Indices	Synthesized Sentinel-2 MSI (S2) Indices	Reference
Band Ratios	Normalized Difference Vegetation Index (NDVI)	--	$L8\_NDVI = (B5 - B4)/(B5 + B4)$	$S2\_NDVI = (B7 - B4)/(B7 + B4)$	Rouse et al. (1974)
	Simple Ratio (SR)	--	$L8\_SR = B5/B4$	$S2\_SR = B7/B4$	Jordan (1969); Rouse et al. (1974)
	Modified Simple Ratio (MSR)	--	$L8\_MSR = \frac{\frac{B5}{B4} - 1}{(\frac{B5}{B4})^{0.5} + 1}$	$S2\_MSR = \frac{\frac{B7}{B4} - 1}{(\frac{B7}{B4})^{0.5} + 1}$	Chen (1996)
Triangulated	Triangular Vegetation Index (TVI)	$Hy\_TVI = 0.5 * [120 * (R_{750} - R_{550}) - 200 * (R_{670} - R_{550})]$	$L8\_TVI = 0.5 * [120 * (B5 - B3) - 200 * (B4 - B3)]$	$S2\_TVI = 0.5 * [120 * (B7 - B3) - 200 * (B4 - B3)]$	Broge and Leblanc (2000)

Soil Adjusted	Improved SAVI with self-adjustment factor L (MSAVI)	$Hy\_MSAVI = \frac{1}{2}[2 * R_{800} + 1 - \sqrt{(2 * R_{800} + 1)^2 - 8 * (R_{800} -$	$L8\_MSAVI = \frac{1}{2}[2 * B5 + 1 - \sqrt{(2 * B5 + 1)^2 - 8 * (B5 -$	$S2\_MSAVI = \frac{1}{2}[2 * B7 + 1 - \sqrt{(2 * B7 + 1)^2 - 8 * (B7 -$	Qi et al. (1994)
	Optimized Soil-Adjusted Vegetation Index (OSAVI)	$Hy\_OSAVI = (1 + 0.16) * (R_{800} - R_{670})/(R_{800} + R_{670} + 0.16)$	$L8\_OSAVI = (1 + 0.16) * (B5 - B4)/(B5 + B4 + 0.16)$	$S2\_OSAVI = (1 + 0.16) * (B7 - B4)/(B7 + B4 + 0.16)$	Rondeaux et al. (1996)
Simple Pigment	Blue/Green and Blue/Red Pigment indices (RGI, BGI, BRI)	$Hy\_RGI = R_{690}/R_{550}$ $Hy\_BGI_1 = R_{400}/R_{550}$ $Hy\_BGI_2 = R_{450}/R_{550}$ $Hy\_BRI_1 = R_{400}/R_{690}$ $Hy\_BRI_2 = R_{450}/R_{690}$	$L8\_RGI = B4/B3$ $L8\_Ultra\ BGI_{(BGI1)} = B1/B3$ $L8\_BGI_2 = B2/B3$ $L8\_Ultra\ BRI_{(BRI1)} = B1/B4$ $L8\_BRI_2 = B2/B4$	$S2\_RGI = B4/B3$ $S2\_Ultra\ BGI_{(BGI1)} = B1/B3$ $ST\_BGI_2 = B2/B3$ $S2\_Ultra\ BRI_{(BRI1)} = B1/B4$ $ST\_BRI_2 = B2/B4$	Zarco-Tejada et al. (2005)
	Simple Ratio Pigment Index (SRPI)	$Hy\_SRPI = R_{430}/R_{680}$	$L8\_Ultra\ SRPI = B1/B4$	$S2\_Ultra\ SRPI = B1/B4$	Peñuelas et al. (1995)

Chlorophyll Absorption	Modified C_{ab} Absorption in Reflectance Index (MCARI)	$Hy\_MCARI = [(R_{700} - R_{670}) - 0.2 * (R_{700} - R_{550})] * (\frac{R_{700}}{R_{670}})$	--	--	Daughtry et al. (2000)
	Transformed CARI (TCARI)	$Hy\_TCARI = 3 * [(R_{700} - R_{670}) - 0.2 * (R_{700} - R_{550}) * (\frac{R_{700}}{R_{670}})]$	--	--	Haboudane et al. (2002)
Red-Edge	Red-Edge Linear Extrapolation	*Inflection point:* $R_{re} = (R_{670} + R_{780})/2$ $Hy\_REP = 700 + 40\left(\frac{R_{re} - R_{700}}{R_{740} - R_{700}}\right)$	--	*Inflection point:* $R_{re} = (B4 + B7)/2$ $S2\_REP = 700 + 40\left(\frac{R_{re} - B5}{B6 - B5}\right)$	Cho and Skidmore (2006)

Table 3.5. Convolved hyperspectral wavelengths (nm) to match Landsat-8 OLI (L8) and Sentinel-2 MSI (S2) bands (*B*).

Satellite	*B1*	*B2*	*B3*	*B4*	*B5*	*B6*	*B7*	*B8*
Landsat-8	Ultra Blue	Blue	Green	Red	NIR	--	--	--
	435-451	452-512	533-590	636-673	851-879	--	--	--
Sentinel-2	Ultra Blue	Blue	Green	Red	Red-Edge 1	Red-Edge 2	Red-Edge 3	Narrow NIR
	430-457	447-546	538-583	646-684	695-713	731-749	770-797	848-882

3.3 RESULTS

Leaf-level chlorophyll values differed between the four sites, with average values higher for the US samples compared to Ethiopia (Table 3.4). Chlorophyll values (*a*, *b*, *a*+*b*) were significantly ($p<0.05$) different for the samples collected from Ethiopia and the United States, ensuring the chlorophyll indices were tested over a wide range of chlorophyll values. These significant differences are likely due to the irrigation practices that are employed in the United States, in which the crop is irrigated through to harvest, that are not practiced in Ethiopia.

Table 3.6. Descriptive statistics of chlorophyll (*a, b, a*+*b*) (g/g) for each site.

Site	Mean (*a*)	Mean (*b*)	Mean (*a*+*b*)	Minimum (*a*+*b*)	Maximum (*a*+*b*)	Standard Deviation (*a*+*b*)
US1 (*n*=40)	0.009	0.004	0.013	0.007	0.022	0.003
US2 (*n*=27)	0.006	0.003	0.010	0.005	0.013	0.002
ET1 (*n*=40)	0.002	0.001	0.002	0.001	0.007	0.001
ET2 (*n*=40)	0.003	0.001	0.004	0.001	0.007	0.002

3.3.1 Chlorophyll Prediction Using Hyperspectral Indices

Twelve indices were computed using the hyperspectral data (Table 3.2), and the OLS regression results show that chlorophyll (*a*, *b*, *a*+*b*) correlates best with simple pigment and red-edge indices (Table 3.5). The strongest relationships were found between total chlorophyll (*a*+*b*) and several of the simple pigment indices including Hy_BRI2 (R^2=0.710), Hy_BRI1 (R^2=0.703), and Hy_SRPI (R^2=0.649). Beyond these, the red-edge index Hy_Red-Edge (R^2=0.621) was the only index to perform at a comparable level. Relationship strength was similar for chlorophyll *a* and chlorophyll *b* with the simple

pigment and red-edge indices (Table 3.5). However, it should be noted that the adjusted R^2 values for chlorophyll *b* were consistently lower than those for chlorophyll *a* and total chlorophyll (*a*+*b*). RMSE values were similar for all relationships (Table 3.5).

Table 3.7. OLS regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*). See Table 3.2 for index descriptions.

Hyperspectral Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		Total C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj. R^2*	*RMSE*	*Adj. R^2*	*RMSE*
Hy_TVI	0.306	0.003	0.179	0.001	0.266	0.004
Hy_MSAVI	0.415	0.003	0.257	0.001	0.366	0.004
Hy_OSAVI	0.442	0.003	0.277	0.001	0.396	0.004
Hy_RGI	0.658	0.002	0.548	0.001	0.631	0.003
Hy_BGI_1	0.642	0.002	0.552	0.001	0.622	0.003
Hy_BGI_2	0.348	0.003	0.329	0.001	0.347	0.004
Hy_BRI_1	***0.742***	***0.002***	***0.593***	***0.001***	***0.703***	***0.003***
Hy_BRI_2	**0.747**	**0.002**	**0.603**	**0.001**	**0.710**	**0.003**
Hy_SRPI	*0.693*	*0.002*	*0.532*	*0.001*	*0.649*	*0.003*
Hy_MCARI	0.402	0.003	0.434	0.001	0.418	0.004
Hy_TCARI	0.324	0.003	0.324	0.001	0.329	0.004
Hy_Red-Edge (nm)	0.657	0.002	0.521	0.001	0.621	0.003

*Highest R^2 in **bold**, second highest in ***bold italic***, third highest in *italic*

A discontinuous relationship was found between chlorophyll and several of the indices tested (see Figure 3.3 for an example). Therefore, fifth-order polynomial regression was performed for the entire set of indices, which improved R^2 values (Table 3.6) but did not alter the general findings. Again, the highest performing indices were of the simple pigment type including Hy_BRI2 (R^2=0.731), Hy_BRI1 (R^2=0.727), Hy_BGI1 (R^2=0.675), and Hy_SRPI (R^2=0.646). Hy_Red-edge (R^2=0.638) also performed comparably well. The rankings according to R^2 values were similar among the different

variants of chlorophyll; however, R^2 values were lower for chlorophyll *b* compared to chlorophyll (*a* and *a*+*b*) (Table 3.5). RMSE values were similar to the OLS results.

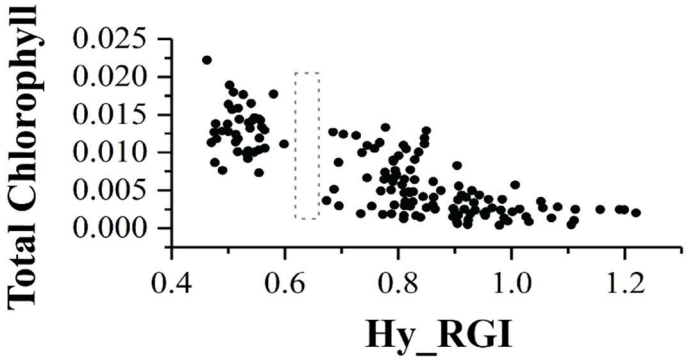

Figure 3.4. Example of the discontinuous relationship between index values (*x*-axis) and total chlorophyll (*y*-axis) outcomes. The index depicted includes the hyperspectral Red/Green Index (Hy_RGI). The discontinuity is highlighted with a dotted-border box.

Table 3.8. Fifth-order polynomial regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*) with indices computed using hyperspectral data. See Table 3.2 for index descriptions.

Hyperspectral Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj.* R^2	*RMSE*	*Adj.* R^2	*RMSE*
Hy_TVI	0.555	0.002	0.434	0.001	0.521	0.003
Hy_MSAVI	0.620	0.002	0.465	0.001	0.575	0.003
Hy_OSAVI	0.613	0.002	0.452	0.001	0.566	0.003
Hy_RGI	0.689	0.002	0.556	0.001	0.653	0.003
Hy_BGI_1	*0.696*	*0.002*	*0.610*	*0.001*	*0.675*	*0.003*
Hy_BGI_2	0.380	0.003	0.372	0.001	0.383	0.004
Hy_BRI_1	***0.758***	***0.002***	***0.643***	***0.001***	***0.727***	***0.003***
Hy_BRI_2	**0.760**	**0.002**	**0.647**	**0.001**	**0.731**	**0.003**
Hy_SRPI	0.688	0.002	0.537	0.001	0.646	0.003
Hy_MCARI	0.402	0.003	0.434	0.001	0.418	0.004
Hy_TCARI	0.390	0.003	0.401	0.001	0.399	0.004

Hy_Red-Edge (nm)	0.682	0.002	0.524	0.001	0.638	0.003

*Highest R^2 in **bold**, second highest in **bold italic**, third highest in italic

3.3.2 Chlorophyll Prediction Using Synthesized Landsat and Sentinel Data

Indices with high (>0.60) R^2 values for total chlorophyll (a+b) relationships are shown along with several of the most commonly used indices (e.g., NDVI, SR, MSR; complete results are reported in Appendix A). OLS regression results for the synthesized L8 data show that the simple pigment indices outperformed all other indices (Table 3.7; Appendix A). L8_BRI2 (R^2=0.732), which incorporates the blue region (452 nm-512 nm), resulted in the strongest relationship with total chlorophyll (*a*+*b*). L8_BRI1 (R^2=0.693) and L8_BGI1 (R^2=0.620), which incorporates the ultra-blue region (435 nm-451 nm), resulted in comparable relationships. The red-edge index could not be calculated due to the spectral band placement of L8. RMSE values of all computed indices were similar to the Hy indices.

Table 3.9. OLS regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*) with indices computed using synthesized Landsat-8 (L8) data. See Table 3.2 for index descriptions.

Synthesized L8 Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		Total C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj. R^2*	*RMSE*	*Adj. R^2*	*RMSE*
L8_NDVI	0.487	0.002	0.348	0.001	0.447	0.004
L8_SR	0.575	0.002	0.403	0.001	0.525	0.003
L8_MSR	0.568	0.002	0.398	0.001	0.518	0.003
L8_RGI	0.628	0.002	0.506	0.001	0.596	0.003
L8_Ultra BGI (BGI_1)	*0.640*	*0.002*	*0.550*	*0.001*	*0.620*	*0.003*
L8_Ultra BRI (BRI_1)	***0.733***	***0.002***	***0.581***	***0.001***	***0.693***	***0.003***
L8_BRI_2	**0.767**	**0.002**	**0.628**	**0.001**	**0.732**	**0.003**

*Highest R^2 in **bold**, second highest in ***bold italic***, third highest in *italic*
**L8_SRPI is redundant to L8_Ultra BRI.

The indices computed using synthesized L8 data also showed discontinuous relationships, so again a fifth-order polynomial regressions were performed. The R^2 values (Table 3.8) were slightly higher than the OLS regression (Table 3.7), although the rankings according to R^2 remained constant (Appendix A). The highest R^2 for total chlorophyll (*a*+*b*) was with L8_BRI2 (R^2=0.761), while L8_BRI1 (R^2=0.703) and L8_BGI1 (R^2=0.653) also produced high R^2 values. RMSE values for all indices were similar to previous results.

Table 3.10. Polynomial regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*) and synthesized Landsat-8 (L8) indices. See Table 3.2 for index descriptions.

Synthesized L8 Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		Total C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj. R^2*	*RMSE*	*Adj. R^2*	*RMSE*
L8_NDVI	0.584	0.002	0.424	0.001	0.537	0.003
L8_SR	0.611	0.002	0.445	0.001	0.563	0.003
L8_MSR	0.579	0.002	0.409	0.001	0.529	0.003
L8_RGI	0.650	0.002	0.505	0.001	0.610	0.003
L8_Ultra BGI (BGI_1)	*0.674*	*0.002*	*0.584*	*0.001*	*0.653*	*0.003*
L8_Ultra BRI (BRI_1)	***0.738***	***0.002***	***0.609***	***0.001***	***0.703***	***0.003***
L8_BRI_2	**0.785**	**0.002**	**0.689**	**0.001**	**0.761**	**0.002**

*Highest R^2 in **bold**, second highest in ***bold italic***, third highest in *italic*
**L8_SRPI is redundant to L8_Ultra BRI.
***Red Edge (nm) could not be included due to lack of spectral resolution.

OLS regression results for the synthesized S2 data show that the simple pigment index S2_BRI2 (R^2=0.682) had the strongest relationship with chlorophyll content (*a*, *b*, *a*+*b*), while S2_BRI1 (R^2=0.678) resulted in similar findings. With S2, it is possible to

compute the red-edge index (R^{2c}=0.613), which was also one of the highest performing indices for total chlorophyll (*a*+*b*) correlations for the synthesized S2 data. RMSE values resulted in only slight differences (<0.001) between the three chlorophyll (*a*, *b*, *a*+*b*) concentrations. Because the Sentinel Multispectral Imagery program includes two satellites (2A and 2B), it is possible to compute multiple versions of each index that correspond to the approximate time that samples were collected in the field (Appendix A). However, the differences between the two satellites in terms of R^2 values were negligible. In all cases, the higher performing of the two was included within the results.

Table 3.11. OLS regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*) and indices computed using synthesized Sentinel-2 (S2) data. See Table 3.2 for index descriptions.

Synthesized S2 Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		Total C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj. R^2*	*RMSE*	*Adj. R^2*	*RMSE*
S2_NDVI	0.477	0.003	0.337	0.001	0.437	0.004
S2_SR	0.572	0.002	0.399	0.001	0.522	0.003
S2_MSR	0.564	0.002	0393	0.001	0.514	0.003
S2_RGI	0.589	0.002	0.466	0.001	0.557	0.003
S2_Ultra BGI (BGI_1)	0.626	0.002	0.542	0.001	0.607	0.003
S2_Ultra BRI (BRI_1)	***0.719***	***0.002***	***0.564***	***0.001***	***0.678***	***0.003***
S2_BRI_2	**0.721**	**0.002**	**0.572**	**0.001**	**0.682**	**0.003**
S2_Red-Edge (nm)	*0.648*	*0.002*	*0.515*	*0.001*	*0.613*	*0.003*

*Highest R^2 in **bold**, second highest in ***bold italic***, third highest in *italic*
**S2_SRPI is redundant to S2_Ultra BRI.

The indices computed using synthesized S2 data revealed a discontinuous relationship, so fifth-order polynomial regressions were computed. R^2 results (Table 3.10) indicate relationships were again slightly higher than with OLS (Table 3.9), although the

rankings remained similar (Appendix A). The simple pigment indices performed best with S2_BRI2 (R^2=0.695) resulting in the strongest relationship with total chlorophyll (*a*+*b*), while S2_BRI1 (R^2=0.683) and S2_BGI1 (R^2=0.645) produced similar results. While still a good fit, the total chlorophyll (*a*+*b*) and red-edge index (R^2=0.637) results decrease in ranking from the OLS (3rd highest; Table 3.9) to the polynomial (4th highest; Table 3.10) regression. However, the relationship between chlorophyll *a* and the red-edge index (R^2=0.648) did result in a higher correlation for chlorophyll a than BGI1 (R^2=0.626) within the polynomial regression. RMSE values were similar to previous regression results and within each of the chlorophyll (*a*, *b*, *a*+*b*) concentrations for the polynomial regression.

Table 3.12. Polynomial regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*) with indices computed using synthesized Sentinel-2 (S2) data. See Table 3.2 for index descriptions.

Broad-Band Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		Total C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj. R^2*	*RMSE*	*Adj. R^2*	*RMSE*
S2_NDVI	0.585	0.002	0.425	0.001	0.538	0.003
S2_SR	0.615	0.002	0.448	0.001	0.566	0.003
S2_MSR	0.584	0.002	0.414	0.001	0.534	0.003
S2_RGI	0.622	0.002	0.468	0.002	0.578	0.003
S2_Ultra BGI (BGI_1)	*0.664*	*0.002*	*0.582*	*0.001*	*0.645*	*0.003*
S2_Ultra BRI (BRI_1)	***0.720***	***0.002***	***0.583***	***0.001***	***0.683***	***0.003***
S2_BRI_2	**0.730**	**0.002**	**0.603**	**0.001**	**0.695**	**0.003**
S2_Red-Edge (nm)	0.678	0.002	0.530	0.001	0.637	0.003

*Highest R^2 in **bold**, second highest in ***bold italic***, third highest in *italic*
**S2_SRPI is redundant to S2_Ultra BRI.

3.4 DISCUSSION

The simple pigment indices showed the strongest relationships with chlorophyll amongst all three data types: Hy, synthesized L8, and synthesized S2. The simple pigment indices are ratios of visible light using red, blue, and green portions of the electromagnetic spectrum. From this family of indices, the ratios incorporating the blue and red wavelengths exhibited the strongest relationships with chlorophyll for all data types. Additionally, indices that included the ultra-blue wavelengths, which are captured by both the L8 and S2 sensors (Table 3.3), outperformed all other indices. It is possible that the reason for the strong performance of the indices computed with the ultra-blue bands is due to the fact that chlorophyll a absorbs light from two portions of the electromagnetic spectrum: the blue region (centered on 430 nm) and the red region (centered on 660 nm) (Jensen 2016). Many of the indices developed previously in the literature did not incorporate light in the region of 430 nm because there was no corresponding band on previous versions of Landsat. However, these results indicate that this region of the spectrum may have strong potential for chlorophyll monitoring, and it may be worth revisiting some well-established indices to determine if incorporation of the ultra-blue band can improve prediction results.

In addition to the ultra-blue bands, the Sentinel-2 platform also carries a sensor positioned in the red-edge region, which shows similar promise for chlorophyll monitoring. Red-edge computed using the synthesized S2 data showed comparable results to the simple pigment indices. The red-edge index could not be calculated for the synthesized L8 data.

Many of the soil adjusted indices that performed well for other crops (Qi et al. 1994; Rondeaux et al. 1996) did not perform well in this study. It is possible that the lodging nature of tef may have played a role in the poor performance of the soil adjusted indices as well. Lodging occurs when the crop, in this case a grass, grows to a height that it cannot sustain its own weight, and it falls over (Figure 3.2). Lodging causes two issues for remote sensing studies. First, for crops that are planted densely, such as tef, the lodging often obscures reflectance from any background materials (e.g., soil). The increased canopy coverage from lodging may explain why the soil adjusted indices did not perform particularly well. Second, lodging may impact reflectance from the plant itself, since the leaves fall in different directions, thus exposing different sides of the leaf to the sensor. While I did not explicitly measure these angular impacts in this study, the effects of lodging on remotely sensed imagery are an important area for future research.

Hyperspectral data are also subject to 'noise' caused by atmospheric interactions, water, and other background targets (i.e. soils), which can limit their effectiveness for applications in agriculture (Gao et al. 2000; Todd and Hoffer 1998; Kawamura et al. 2005). In some cases, particularly when using the synthesized L8 data, I observed stronger relationships for the convolved data than the original hyperspectral data. It is possible that the indices computed using the hyperspectral data included noise from surrounding environmental factors that was muted when the hyperspectral bands were aggregated into the synthesized data. Future studies incorporating hyperspectral data may benefit from analyzing differing combinations of band averaging to minimize the effects of noise in the calculation of chlorophyll indices.

Researchers and agriculturalists seeking to put the indices identified here into practice will need to choose which of the two satellites to use. While the results presented here were similar, the synthesized L8_BRI2 index outperformed the synthesized S2_BRI2 index. However, the Sentinel-2 MSI sensor may be more beneficial to smallholder farmers than the Landsat-8 OLI sensor since it has both higher spatial and temporal resolutions (Bari et al. 2014; ESA 2018). In particular, the temporal resolution is important as it could allow for more precise phenological studies. Additionally, the Sentinel-2 platform includes a sensor in the red-edge region, which provides the opportunity to compute the red-edge index.

3.5 CONCLUSION

The aim of this study was to test whether chlorophyll prediction indices originally developed using hyperspectral data can be computed using synthesized Landsat-8 OLI and Sentinel-2 MSI data to successfully predict leaf chlorophyll content across differing environments. The simple pigment chlorophyll indices performed best for tef in all studied sites in both Ethiopia and the United States. Specifically, the Blue/Red simple pigment (BRI2) index resulted in the strongest relationships with chlorophyll (*a*, *b*, *a*+*b*) concentrations for all three datasets. Additionally, the Red-Edge index performed well within the synthesized S2 data. These findings have implications for farmers and those interested in using Landsat-8 OLI and Sentinel-2 MSI for monitoring crop health. The increased spectral sensitivities (i.e. ultra blue and red-edge bands) of the sensors onboard Landsat-8 and Sentinel-2 provide future research opportunities to focus on the incorporation of new bands to improve past or create new chlorophyll indices..

CHAPTER IV

SITE SUITABILITY ANALYSIS FOR TEF (*ERAGROSTIS TEF*) WITHIN THE CONTIGUOUS UNITED STATES

Abstract

Livestock production has increased globally over the past 50 years to meet the demand of population growth and, along with it, increased demand for meat. Agricultural managers are pressed to increase production of forage for grazing to meet these demands, along with crop production for both humans and livestock. Therefore, nutritious forage for livestock that can serve as a sequential crop grown within the rotation of major food crops (e.g., wheat) is desired. *Eragrostis tef* (tef) is native to Ethiopia and is the most commonly cultivated crop in that region. However, because it is a C_4 grass, it has the potential to be grown in the United States during the summer months as it is drought resistant, not requiring irrigation. Despite the potential benefits of incorporating tef in the United States, there is limited research about the suitability of this crop in the country. Using a weighted overlay approach and the analytical hierarchy process (AHP), this study identifies suitable sites for tef production within the 48 contiguous United States. Using AHP in a Geographic Information System (GIS), this study assesses elevation, slope, insolation, soil type, average precipitation (without irrigation), average temperature, minimum temperature, and land cover to identify sites suitable for the

cultivation of this nutritious forage. Results show that within the early-summer (May, June, July; MJJ) 32.38% of the contiguous United States is highly suitable for tef cultivation. In the late-summer (July, August, September, JAS), 32.68% is highly suitable. The findings at the state level suggest that Texas, Kansas, South Dakota, North Dakota, Nebraska, Iowa, Delaware, Kentucky, and Ohio can take on the cultivation of tef to support livestock, sequentially grow tef with wheat, or begin cultivation of a specialty crop. Additionally, using the known United States locations in which tef is currently cultivated, an accuracy assessment of the GIS and site suitability method found the technique to be highly accurate.

4.1 INTRODUCTION

Livestock production has increased globally over the past 50 years to meet the demand of population growth and, along with it, increased demand for meat (Delgado 2005; Thornton 2010). Due to these increases, a rising demand (300% increase) for crop production for feed has manifested (FAO 2016a). Today, livestock production has been able to keep pace with the growing population but has significant implications on food security as croplands have been devoted to feed production rather than crops for human consumption (Tilman et al. 2002). Agricultural managers are under pressure to increase production of forage for grazing along with crop production to meet the demands of both humans and the intensified livestock industry (Steinfield et al. 2006; Delgado et al. 1999). A nutritious forage for livestock that can serve as a sequential crop grown within the rotation of major food crops such as *Triticum aestivum* (wheat) has been highly desired among agricultural managers for some time (Boe et al. 1986). Such a practice could mirror the *Vigna unguiculata* (cowpea) and wheat rotation methods that are accomplished

by the indigenous of Ethiopia, resulting in maximum productivity of land (Berrada et al. 2006).

Eragrostis tef (tef), a C_4 grass native to Ethiopia, has potential to satisfy the need for a sequential crop in the United States. Of all the 350 species of *Eragrostis*, *E. tef* is the only tropical cereal variety that is cultivated (Demissie 2000). The parent plant can be grown across many environments as it is drought resistant and tends to resist insect pests (Gerbremariam et al. 2014). Many farmers who grow tef in Ethiopia are smallholders (<25.2 ha). To raise the crop, farmers are heavily dependent on rainfall as most of them are not utilizing irrigation systems. Its strong dependence on rainfall in its native climate suggests that the crop is drought resistant and does not require irrigation, an advantage for its potential use as a summer crop in the United States (Taffesse et al. 2012).

Tef produces both forage and grain, serving a dual agricultural purpose (Miller 2014). The nutrient properties of both the forage and the grain are highly regarded for their high contents of protein, iron, and calcium supporting the nutritional needs of livestock as well as human dietary requirements (Boe et al. 1986; Twidwell et al. 2002; Bultosa and Taylor 2004; Dekking et al. 2005; Gerbremariam et al. 2014; Hopman et al. 2008). The forage is similar in nutritional quality to many other grasses utilized for hay (Boe et al. 1986, Twidwell et al. 2002), and the digestibility and protein content (as high as 19.5%) of test plots grown in the U.S. meet the levels desired by livestock managers (Stallknecht 1993; Boe et al. 1986).

Due to its many desirable qualities and farming requirements, trials and commercial cultivation of the crop have been explored throughout the United States. Tef

is a C_4 grass and has strong potential to serve as a sequential crop for C_3 grasses due to its efficient pathways for carbon concentration and the number of days required for hay cutting. C_4 grasses are more efficient than C_3 grasses at converting solar energy into biomass, have improved water use efficiency (WUE), and have greater nitrogen use efficiency (NUE) because of the many pathways used to process carbon. Therefore, C_4 grasses generally outperform C_3 grasses during the warm months within the contiguous United States. Additionally, tef only needs 45 to 55 days to have enough biomass to cut for use of hay (Miller 2014). Tef thrives best in temperatures between 30°C to 40°C (Bjorkman et al. 1970; Bjorkman and Pearcy 1971; Long et al. 1975; Loomis 1983; Ludlow and Wilson 1971; Monson et al. 1983; Pearcy and Harrison 1974; Tieszen and Delting 1983) but will die if temperatures fall below freezing (Miller 2014). Further, tef has specific temperatures in which productivity is most significant (Figure 4.1; Kebede 1986). Additionally, tef does not require irrigation in regions with greater than 43.18 cm (17 in) of precipitation, though it is tolerant of some drought and waterlogging (FAO 2016b; Westphal 1975; Miller 2014). It is also suggested that tef thrives in soils with some clay for structure purposes between plant and soil as well as a variety of levels of nutrient development, some good examples are Alfisols, Vertisols, and some Inceptisols when it comes to soil types (Westphal 1975). Furthermore, research has found that species of Eragrostis cultivate well in large amounts of insolation (Roseberg 2005). For C_4 grasses, higher amounts of insolation generally have positive linear relationships to biomass (Roseberg 2005; Gosse et al. 1986). C_4 grasses also thrive at differing elevations (Figure 4.2; Long 1983; Rundel 1980; Miller 2011).

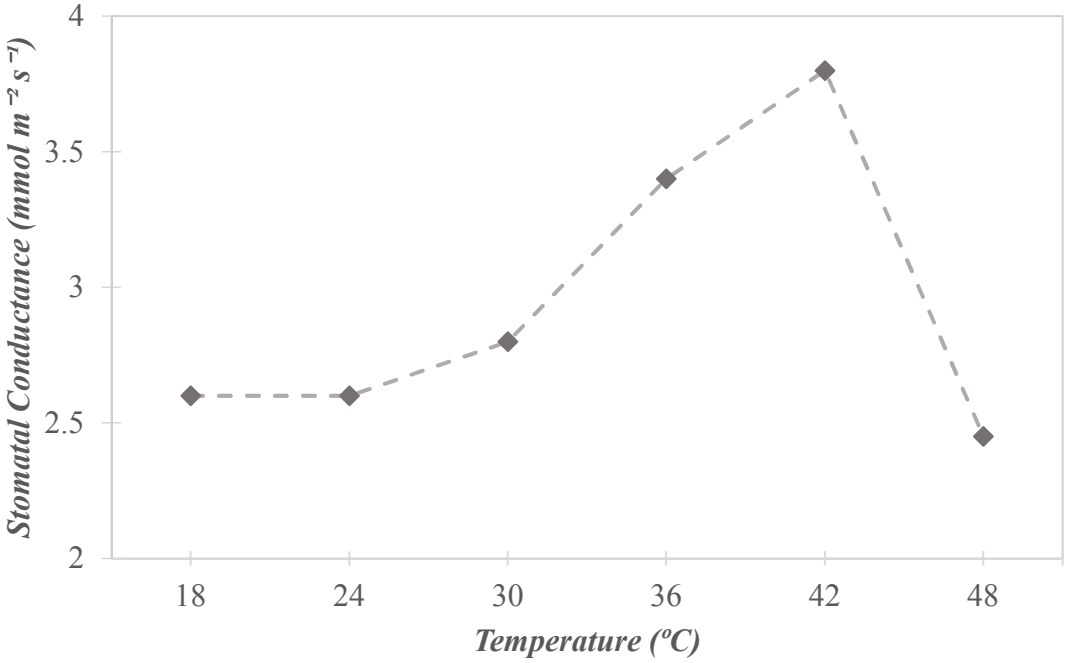

Figure 4.5. Stomatal conductance of *Eragrostis tef* (Zucc.) Trotter under varying temperatures. Adapted from: Kebede (1986).

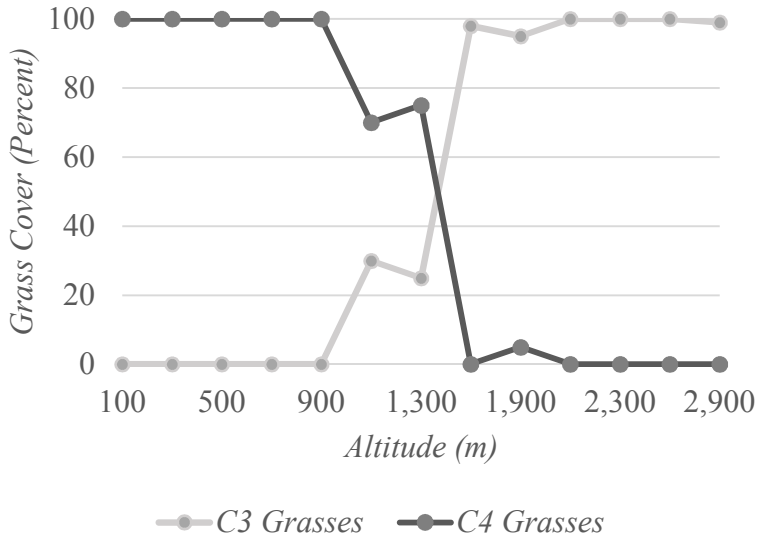

Figure 4.6. A model depicting the average elevations and temperatures in which C_4 and C_3 grasses perform. Adapted from: Rundel (1980).

Using these known requirements for the cultivation of tef, the objective of this research is to identify a method of site suitability analysis for tef in the United States that will locate where this crop is highly suitable. Thus, the suitability of tef within the United States requires a detailed regional analysis of environmental conditions suitable for the introduction of the crop. Crop suitability studies frequently employ weighted overlay techniques to locate suitable areas for crop production from known environmental and physical conditions (i.e., site suitability analysis) (Akinci et al. 2013; Pramanik 2016; Bandyopadhyay et al. 2009; Feizizadeh and Blaschke 2012). Weighted overlay is a form of multi-criteria suitability evaluation (MCE) (Pramanik 2016), which utilizes differing criteria such as geologic, biophysical, and climatic elements in a decision-making process that leads to the solving of a problem using multiple inputs (Wang et al. 1990; Jankowski 1995; Yu et al. 2011; Aydi 2018). Geographic Information Systems (GIS) are frequently employed in MCE because they allow overlay analyses of multiple geospatial data layers (Mokarram and Aminzadeh 2010; Mendas and Delali 2012; Aydi et al. 2016).

The determination of appropriate weights for each of the input layers is one of the most critical steps in a weighted overlay method since the multi-level hierarchical arrangement of the variables is not necessarily known a priori (Triantaphyllou and Mann 1995). There are a few techniques for determining the weighting system (Voogd 1983; Yager 1988). While many of these techniques for determining weights are relatively easy and straightforward, their weight assignments are subjective. These techniques also limit the ability to check for biases, as often comparing variables as a whole rather than a comparison of the distinct relationship among each variable. Thus, many studies rely on the analytical hierarchy process (AHP) to determine the weight of importance of different

variables in the analysis of site suitability (Pramanik 2016). AHP is based upon a pairwise comparison of variables that utilizes relative significance. The relative significance of the variables establishes weighting parameters for complex data relationships (Saaty 1980; Miller et al. 1998; Tiwari et al. 1999). The process sub-divides the often intangible task of ranking many variables into miniature sub-problems, providing an opportunity for pairwise comparison instead of comparing the variables as a whole (Saaty 1987). Thus, utilizing AHP and MCE this study aims to identify suitable sites and states within the contiguous United States best suited for the cultivation of tef using known environmental conditions such as elevation, slope, insolation, soil type, average precipitation (without irrigation), average temperature, minimum temperate, and land cover. Additionally, the study aims to confirm the accuracy of the AHP and MCE methods by comparing site suitability rankings to known sites of current tef cultivation within the contiguous United States. Confirming the current method will allow researchers to incorporate the geospatial analytics within future agricultural site suitability studies.

4.2 MATERIALS AND METHODS

4.2.1 Study Area

The study area will be comprised of the 48 contiguous United States as many within this region have tested tef's cultivation potential, suggesting it beneficial to obtain an extent at which further trials and cultivation are worth pursuing. Furthermore, the terrain and environmental conditions are greatly heterogeneous across the study area, providing an optimal region for site suitability analysis and the AHP.

4.2.2 Criteria Utilized to Construct Site Suitability and Data Manipulation

Eight criteria that are known to influence tef cultivation were included in this analysis: elevation, slope, insolation, soil type, average precipitation, average temperature, minimum temperate, and land cover. These criteria and the manipulation of the data are further explained in the following sections. The rankings and choice of data were based on a review of past literature (Rundel 1980; Kebede 1986; Delden et al. 2012; Pramanik 2016). Additionally, for all spatial data it was found that the heterogeneity of the study area was preserved within a 90m resolution, thus, if a set of spatial data had lower spatial resolution a gridding tool (*Create Fishnet*, ArcMap 10.3.1) was used to grid the data at 90m, though the spatial data were not altered.

4.2.2.1 Elevation

In Ethiopia, tef is best cultivated 2,000 m (7,726 ft) mean sea level (MSL) because the elevation works like a temperature control for the crop due to Ethiopia being near the equator. However, in the United States, C_4 grasses generally cannot thrive at such high elevations at higher latitudes (Figure 4.2; Rundel 1980). This tradeoff increases the amount of time required to cultivate certain plants as they need extended periods to mature due to the differing climatic conditions (Pramanik 2016). Thus, elevation data were obtained from the United States Geologic Survey (USGS) National Elevation Dataset (NED) at 90 m resolution. The standardization of the elevation values for the weighted overlay and AHP (Table 4.1) were implemented.

4.2.2.2 Slope

Steep slopes generally produce poorer soil development compared to shallow slopes due to erosion and runoff (Barnes 2010). Thus, the mass of materials being eroded away depend heavily on the degree of slope (Koulouri and Giourga 2007). Slope also has secondary effects that include an adverse effect on irrigation control, and if the degree of slope is too steep, an inability to utilize machinery (Pramanik 2016). Slope data were derived from a digital elevation model (DEM) obtained from the USGS NED at a nominal resolution of 90 m. The variable was then standardized (Table 4.1) for the weighted overlay and AHP.

4.2.2.3 Insolation

Tef is sensitive to insolation and like other C_4 grasses performs best with more insolation (Roseberg et al. 2005; Gosse et al. 1986). Roseberg et al. (2005) have found that flowering of the crop reduces with decreased amounts of sunlight availability. The modeling of insolation, in this context, is for varieties that are insolation insensitive, such as tef (Delden et al. 2012). Thus, this variable was based upon average monthly total insolation from a 7-year data set (1985-1991) (National Renewable Energy Laboratory, 2017). The data for May through September were averaged, then the 40 km spatial resolution data were gridded to a 90m resolution. The variable was then standardized for the weighted overlay and AHP (Table 4.1).

4.2.2.4 Soil type

In this study, soil type is soil taxonomy according to the World Soil Survey (WSS), a subsidiary of the United States Department of Agriculture (USDA). Tef is known to grow well in Alfisols, Vertisols, and Inceptisols when it comes to soil

categorization provided by the WSS as these soils often provide enough structure and nutrition for cultivation (Miller 2014; Westphal 1975). The data obtained from the WSS came in a vector format. Thus, the data in the vector were altered to a number format where 10 was suitable (Alfisols, Vertisols, and Inceptisols) and 1 was not suitable (all other soil taxonomies). The newly categorized vector was transformed into a 90 m resolution raster for each raster cell where cells with soil boundaries assigned the soil with maximum area within the cell. The newly formed numerical values were then implemented into the weighted overlay and AHP analysis (Table 4.1).

4.2.2.5 Average precipitation

In Ethiopia, tef is cultivated during the wet seasons known as *Belg* (February-April) and *Meher* (June-September) (FAO 2016c). It is during these times that the crop can obtain enough water without being irrigated. Miller (2014) suggests that for tef to be grown without irrigation, average annual precipitation must be between 43.18 and 60.96 cm (17 and 24 in). The crop is also known for its ability to grow under dry or overly wet conditions (Miller 2014). Thus, there can be some variability in the amount of rain within regions for tef cultivation. Precipitation data for the U.S. were obtained from the Prism Climate Group (2016) at 800 m resolution and were gridded to 90 m to match the resolution of other datasets. The average precipitation values were then standardized for the weighted overlay and AHP (Table 4.1).

4.2.2.6 Average temperature

C_4 grasses such as tef thrive best in temperatures between 30°C to 40°C (Loomis 1983; Monson et al. 1983; Tieszen and Delting 1983). This productivity level is often

related to the stomatal conductance, or the ability to pass carbon dioxide, which generally decreases as temperatures decrease for C_4 grasses. For tef, the stomatal conductance performs best around average temperatures of 42°C (Figure 4.1; Kebede 1986). Because this study is particularly interested in identifying regions in which tef can be cultivated as a sequential crop in rotation with C_3 crops, the average temperature data were split into three month averages for two potential rotation periods: early-summer (May, June, July; MJJ) and late-summer (July, August, September; JAS). Each period is approximately 90 days, which could potentially allow for two harvesting cycles, as tef can mature in 45 to 55 days for the first cutting of hay and approximately 20 days later be mature enough for a second cutting under proper conditions (Miller 2014). Temperature data were obtained from the Prism Climate Group (2016) dataset at 800 m spatial resolution, and 3-month averages were computed for the early- (MJJ) and late-summer (JAS) periods. Additionally, similar to average precipitation, the data were gridded to 90 m and standardized for the weighted overlay and AHP (Table 4.1).

4.2.2.7 Minimum temperature

Tef is relatively hardy for a C_4 grass but is susceptible to frost throughout growth. If temperatures drop below 0°C, tef will die, and the crop will be lost (Miller 2014). The number of months in which the 30-yr average minimum temperature did not drop below 0°C for each 800 m pixel were summed for the five months to account for frost potential. The study was concerned with the summer months (May, June, July, August, and September), thus, resulting in the most optimum value being five (equating to five months without an occurrence of freezing). Following these procedures, the data were at 800 m resolution; therefore, the data were gridded to 90 m to match the resolution of

other datasets. The data were then standardized in preparation for the weighted overlay and the AHP (Table 4.1).

4.2.2.8 Land cover

Land cover plays a vital role in where tef can be feasibly cultivated. Thus, based on the USGS's National Land Cover Database (NLCD) (Homer et al. 2012), this analysis utilized the categories of grassland, pasture/hay, and cultivated crops (71, 81, 82) as the highest suitability, while the remaining categories were assigned the lowest suitability. The NLCD was available at 30 m resolution and was resampled to 90 m using majority rules classification. It was then standardized using numerical values instead of categorical values for the weighted overlay and the AHP (Table 4.1).

Table 4.13. Standardized rankings within criteria.

Main Criteria	Sub-criteria	Ranking
Average Annual Precipitation (cm)	≥43.18	10
	43.18-25.4	5
	25.4-15.24	2
	15.24-0	1
Slope (°)	0-3	10
	3-10	8
	10-20	6
	20-30	4
	30-89	2
Elevation (m)	0-1000	10
	1000-1350	7
	1350-1750	3
	1750-2000	2
	>2000	1
Average Temperature (°C) *Both MJJ and JAS (Separately)*	42-36	10
	36-30	7
	30-24	5
	24-18	3

	<18 or >42	1
Minimum Temperature (°C)	5 Months Above Freezing	10
	4 Months Above Freezing	8
	3 Months Above Freezing	6
	2 or less Months Above Freezing	1
Soils	Alfisols, Vertisols, Inceptisols	10
	Other Soils	1
Insolation (kW/m^2/day)	≥7	10
	7-5	8
	5-4	6
	4-3	4
	<3	2
Land Cover	71, 81, 82	10
	Other Categories	1

4.2.3 Calculation of Weights for Individual Criteria

The AHP, originally introduced by Saaty (1980), permits the weight decision-making process using a pairwise comparison matrix. The matrix is composed of relative significance of criteria in pairs (Miller et al. 1998). The significance values ranging from 1 (equal significance) to 9 (greatest significance) compose the matrix. The significance values establishing the importance of the environmental variables for the individual pairwise comparisons are based upon literature review and local expert knowledge. Saaty (1980) provides a precise definition for the significance values (Table 4.2). The pairwise comparison matrix is mainly supported by the criteria of reciprocity, mathematically expressed as:

$$Reciprocity = \frac{n(n-1)}{2} \qquad \textit{(Eq. 1)}$$

where *n* represents the number of components within the pairwise comparison matrix. The pairwise comparison matrix is then subject to calculation of weights based upon the

eigenvalue that corresponds to the greatest eigenvector of the matrix, followed by a normalization of the sum of the factors (Saaty 1980; Pramanik 2016). This step is completed by using the matrix of relative significance that is made of columns and rows of the variables. The relative significance determined for this study were based on past studies that have similar methods focused on agricultural production (Mokarram and Aminzadeh 2010; Mendas and Delali 2012; Akinci 2013; Pramanik 2016). Following the determination of relative significance, the value of each cell is divided by the corresponding column's sum. After that is completed, each row's average is calculated for the newly created values, from the first step. The averages for each row represent the relative weights. Following completion of the two newly calculated matrices, the final steps of the AHP include the use of the consistency ratio (*CR*), random index (*RI*), and the consistency index (*CI*) in order to calculate and/or identify whether there are any inconsistencies in the weighting of the criteria (Saaty 1980; Pramanik 2016). The efficiency of the criteria is assessed by the *CR*, mathematically expressed as:

$$CR = \frac{CI}{RI} \qquad \textit{(Eq. 2)}$$

where *RI* represents the random index, (Saaty 1980) for studies using 1 to 10 different variables. The *CI* represents the consistency index calculated using the following equation:

$$CI = \frac{\lambda_{max} - n}{n-1} \qquad \textit{(Eq.3)}$$

where λ_{max} is the principle eigenvector of the matrix and *n* represents the order of the matrix. Once the *CR* is computed, the value must be less than 0.10 or the matrix is

suggested to have inconsistencies, resulting in conclusions that could be meaningless (Saaty 1980). If there are no inconsistencies, the weights can be used as percentages to employ the weighted overlay method.

Table 4.14. Fundamental scale proposed by Saaty (1980) for the pairwise comparison matrix.

Importance Rank	Definition	Explanation
1	Equal importance	Two criteria contribute equally to the objective
3	Low importance of one over another	Experience and judgment slightly favor one criterion over another
5	Strong or essential importance	Experience and judgment strongly favor one criterion over another
7	Established importance	A criterion is strongly favored, and its dominance is established in practice
9	Absolute importance	The evidence favoring one criterion over another is of the highest probable order of affirmation
2,4,6,8	Intermediate values between the two adjacent judgments	When compromise is needed
Reciprocals	If activity *i* has one of the above numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*	

4.2.4 Weighted Overlay Method for Site Suitability

After the weights were calculated, the eight rasters were combined using equation 4 to identify areas in the U.S. most suitable for incorporating tef as a sequential crop:

$$SS = \sum_{i=1}^{n} W_i X_i \qquad (Eq.\ 4)$$

where *SS* denotes the site suitability score (the higher the score, the more suitable the region), W_i is the weight for individual criteria being measured, X_i indicates the ranking based on sub-criteria for the same criteria, and *n* is the total number of criteria (*n*=8; Pramanik 2016). The combining of the raster layers was conducted in ArcGIS (10.5.1) utilizing the raster calculator data management tool. This software was incorporated because of the need to organize and calculate large geospatial datasets.

4.3 RESULTS

4.3.1 Individual Criteria

For elevation, based upon criteria rankings associated with past research, 68.13% of the land in the contiguous U.S. is optimal (0-1000m MSL) for C_4 grasses (Table 4.3). For slope, over half (64.53%) of the study area is between 0 to 3 degrees (highly suitable) and another 22.23% of the contiguous U.S. is between 3 to 10 degrees (moderately suitable) (Table 4.1; Pramanik 2016). The soil type results suggest that 29.87% of the study area is optimal for tef cultivation when considering this criterion solely (Table 4.3). Nevertheless, precipitation results alone suggest most (90%) of the study area is suitable for tef production receiving at least 43.18 cm (17 in) of rain annually (Table 4.3). The data for the average temperatures of MJJ and JAS meeting the temperature needs (moderately suitable) of a C_4 include 0.47% and 1.49% of the study area, respectively (Table 4.3). Further, the temperature data suggests that most (96.39%) of the study area had no freezing events across the months observed (Table 4.3). The calculated insolation raster suggests that 46.13% of the land within the contiguous U.S. has optimal amounts of sunlight for tef growth (Table 4.3). Finally, the resulting raster output suggests 37.98%

of the land cover in the contiguous U.S. falls within the three categories (71, 81, 82) of interest. Along with all of the criteria is a geospatial raster depiction for each criterion (Figure 4.3).

Table 4.15. Area (percentage of total area) for each criterion.

Main Criteria	**Sub-criteria**	**Area (ha)**	**Area (%)**
Average Annual Precipitation (in)	≥43.18	703,991,032	90.14
	43.18-25.4	34,790,313	4.45
	25.4-15.24	18,640,282	2.39
	15.24-0	23,584,356	3.02
Slope (°)	0-3	504,007,329	64.53
	3-10	173,608,323	22.23
	10-20	68,435,101	8.76
	20-30	26,440,676	3.39
	30-89	8,510,751	1.09
Elevation (m)	0-1000	532,133,313	68.13
	1000-1350	69,592,752	8.91
	1350-1750	77,814,361	9.96
	1750-2000	37,477,195	4.80
	>2000	63,984,560	8.19
Average Temperature (°C) *MJJ*	42-36	0	0
	36-30	4,064,004	0.47
	30-24	179,956,330	20.72
	24-18	328,099,630	37.77
	<18 or >42	356,511,307	41.04
Average Temperature (°C) *JAS*	42-36	0	0
	36-30	11,899,801	1.49
	30-24	236,534,770	29.53
	24-18	380,339,340	47.49
	<18 or >42	172,120,220	21.49
Minimum Temperature (°C)	5 Months Above Freezing	752,785,332	96.39
	4 Months Above Freezing	24,260,291	3.11
	3 Months Above Freezing	2,705,290	0.35
	2 or fewer Months Above Freezing	1,255,069	0.16

Soils	Alfisols, Vertisols, Inceptisols Other Soils	233,292,294 547,714,697	29.87 70.13
Insolation	≥7 6-5 5-4 4-3	360,241,320 6,566,195 410,415,090 3,669,055	46.13 0.84 52.56 0.47
Land Cover	71, 81, 82 Others Categories	296,611,759 484,395,100	37.98 62.02

**MJJ: March, June July; JAS: June, August, September.*

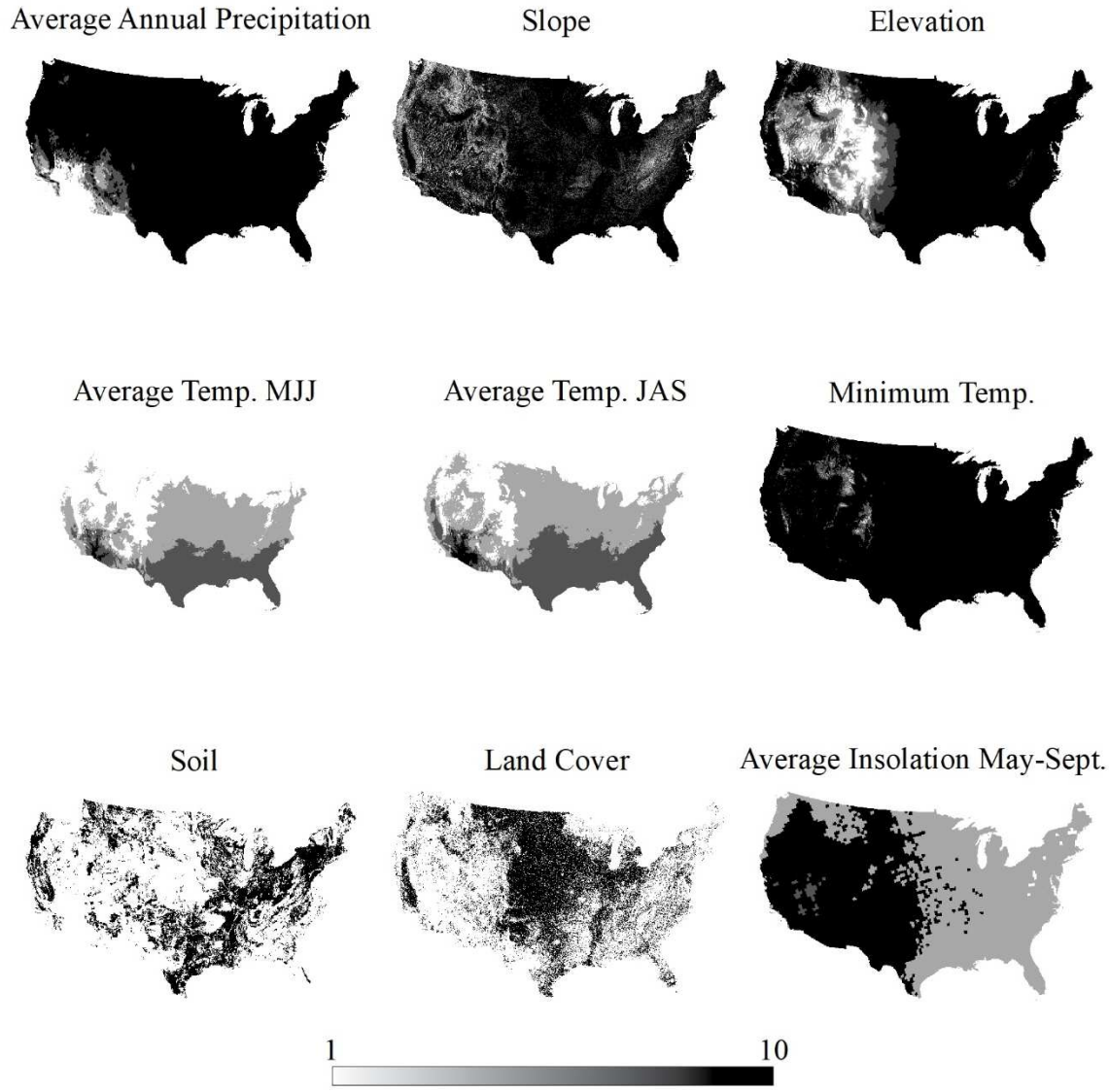

Figure 4.7. Suitability maps for each criterion.

4.3.2 AHP

Using Saaty's (1980) method, a pairwise comparison matrix was created to calculate a quantitative relationship based on past research (Akinci et al. 2013; Pramanik et al. 2016) for each of the unique pairs of criteria (Table 4.4). Furthermore, the insolation criterion has not been incorporated in past studies; thus, the relationship of insolation to other criteria was based on the findings of Muchow et al. (1989). The synthesized matrix for relative weights was then calculated based on the methods of Saaty (1980) (Table 4.5). The weights for each criterion were defined by calculating the average of each row of the synthesized matrix for relative weights (Table 4.5). Each of these weights was incorporated in the calculations of site suitability (*Eq. 4*). Through utilization of the random inconsistency (*RI*) indices provided by Saaty (Table 4.6), the calculation of the consistency index (*CI*; *Eq. 3*), and the number of criteria incorporated in the study (n=8), the calculated consistency ratio (*CR*; *Eq. 2*) resulted in a value of 0.048. The resulting *CR* was less than 0.10; therefore there were no inconsistencies in the calculation of the weights Saaty (1980).

Table 4.16. Pairwise comparison matrix.

Criteria	Slope	Elevation	Land Cover	Precipitation	Average Temp.	Minimum Temp.	Soil	Aspect
Slope	1	2	2	3	3	3	6	9
Elevation	1/2	1	2	3	3	3	6	8
Land Cover	1/2	1/2	1	3	3	3	6	8
Precipitation	1/3	1/3	1/3	1	1	1	5	7
Average Temp.	1/3	1/3	1/3	1	1	1	5	7
Minimum Temp.	1/3	1/3	1/3	1	1	1	5	7
Insolation	1/6	1/6	1/6	1/5	1/5	1/5	1	4
Soil	1/9	1/8	1/8	1/7	1/7	1/7	1/4	1

Table 4.17. Synthesized matrix for relative weights.

Criteria	Slope	Elevation	Land Cover	Precipitation	Average Temp.	Minimum Temp.	Soil	Aspect	Weights
Slope	0.305	0.417	0.318	0.243	0.243	0.243	0.175	0.176	0.265
Elevation	0.153	0.209	0.318	0.243	0.243	0.243	0.175	0.157	0.218
Land Cover	0.153	0.104	0.159	0.243	0.243	0.243	0.175	0.157	0.185
Precipitation	0.102	0.070	0.053	0.081	0.081	0.081	0.146	0.137	0.094
Average Temp.	0.102	0.070	0.053	0.081	0.081	0.081	0.146	0.137	0.094
Minimum Temp.	0.102	0.070	0.053	0.081	0.081	0.081	0.146	0.137	0.094
Insolation	0.051	0.035	0.026	0.016	0.016	0.016	0.029	0.078	0.034
Soil	0.034	0.026	0.020	0.012	0.012	0.012	0.007	0.020	0.018

Max. eigenvalue (λ_{max}) = 8.474.
N = 8.
Consistency index (CI) = (λ_{max} – n)/(n – 1) = 0.068.
Random index (RI) = 1.41.
Consistency ratio (CR) = (CI/RI) = 0.048.

Table 4.18. Random inconsistency (*RI*) indices as provided by Saaty (1980).

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.46	1.49

Random index (RI)

4.3.3 Weighted Overlay

Using the criteria rasters (Figure 4.3), the weighted overlay method resulted in two products: one computed using MJJ average temperature (Figure 4.4) and another computed using JAS average temperatures (Figure 4.5). Site suitability classifications for agricultural systems within the FAO are generally divided into five classes that include [values in brackets correlate to a range of corresponding rankings]: 1) Highly suitable [10-8], 2) moderately suitable [8-6], 3) marginally suitable [6-4], 4) currently not suitable [4-2], and 5) permanently not suitable [<2]. For the early-summer (MJJ) analysis, the weighted overlay resulted in 32.38% highly suitable, 46.46% moderately suitable, 19.18% marginally suitable, 1.98% currently not suitable, and <0.00% permanently not suitable (Table 4.7). Similarly, for the late-summer (JAS) analysis, 32.68% is highly suitable, 47.05% is moderately suitable, 18.35% is marginally suitable, 1.92% is currently not suitable, and <0.00% is permanently not suitable (Table 4.7). While the two sets of results are very similar, there is slightly more highly suitable land for tef cultivation during the early summer period, likely due to higher average temperatures and increased insolation.

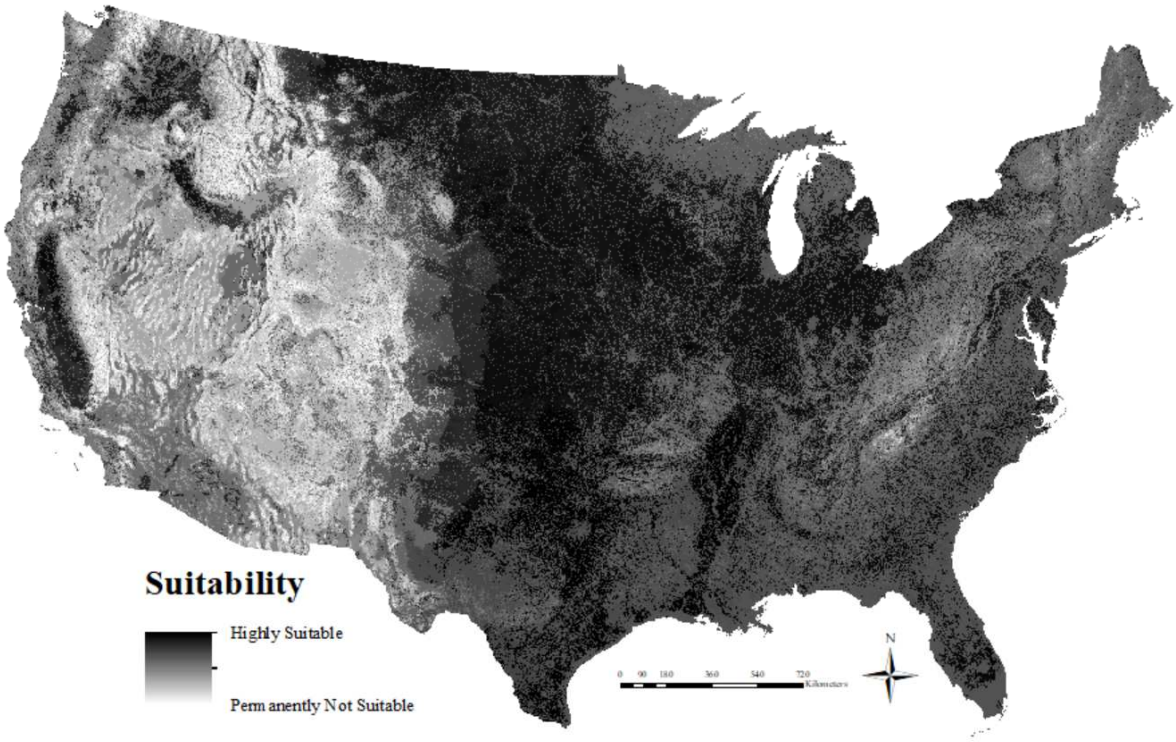

Figure 4.8. Tef site suitability for the early-summer period (May, June, July; MJJ).

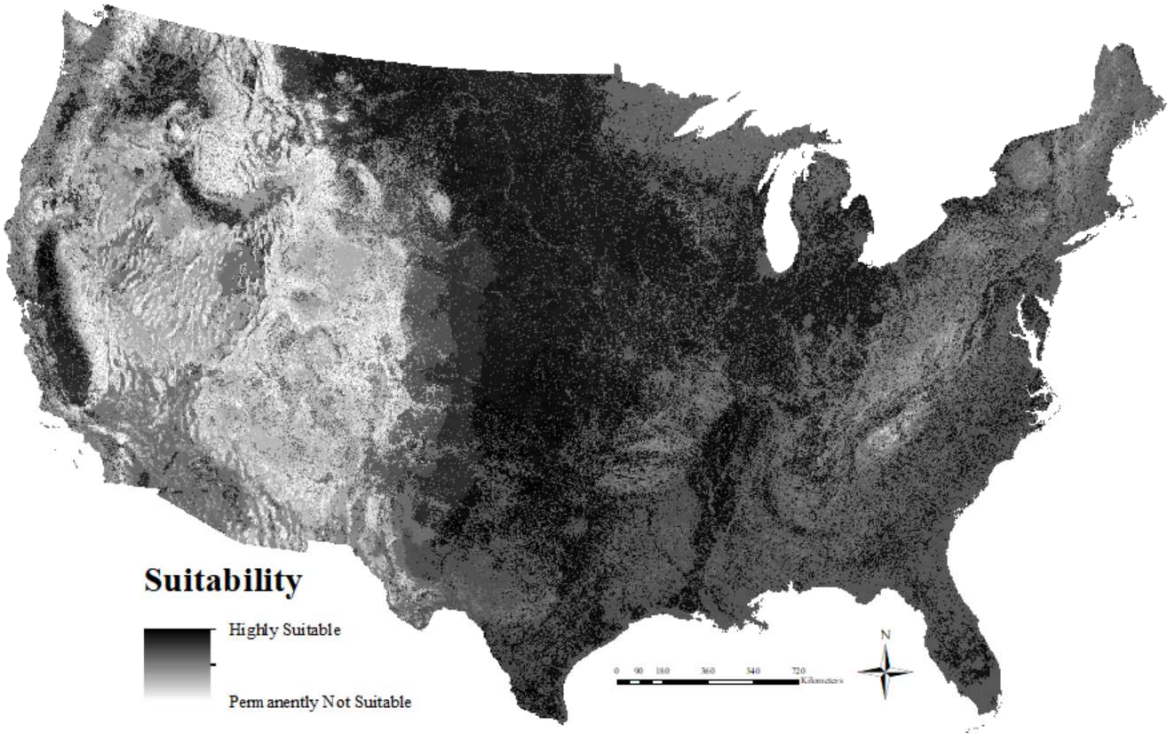

Figure 4.9. Tef site suitability for the late-summer period (July, August, September; JAS).

Table 4.19. Areas and percentages of tef site suitability results using FAO categorization.

FAO Suitability Categories	MJJ Tef Suitable Land Within Study Area		JAS Tef Suitable Land Within Study Area	
	ha	***%***	***ha***	***%***
Highly Suitable	252,837,773	32.38	255,207,542	32.68
Moderately Suitable	362,808,748	46.46	367,367,921	47.05
Marginally Suitable	149,747,360	19.18	143,260,145	18.35
Currently Not Suitable	15,448,493	1.98	15,006,766	1.92
Permanently Not Suitable	2	<0.00	2	<0.00
Total	780,842,378	100	780,842,378	100

4.4 DISCUSSION

The results suggest it is practical to incorporate tef as a sequential crop during the summer months in the United States. The later months of the summer are slightly more suitable for the higher latitudes and/or elevations as temperature can be a controlling factor in the successful cultivation of the C_4 crop (Bjorkman et al. 1970; Bjorkman and Pearcy 1971; Long et al. 1975; Loomis 1983; Ludlow and Wilson 1971; Monson et al. 1983; Pearcy and Harrison 1974; Tieszen and Delting 1983). Nevertheless, tef is highly suitable in many regions of the United States. Several states that have not yet participated in tef trials (Miller 2014) can implement tef into their agricultural practices. Texas is the state with the highest amount of land (25,900,471 ha; Figure 4.6) highly suitable for tef production, in the late summer, which could be a benefit considering the potential to further support the livestock industry within the state with a nutritious forage. Kansas,

South Dakota, North Dakota, Nebraska, and Iowa have 80% or more land highly suitable for cultivation of tef in the late summer. For Kansas and North Dakota a C_4 grass, such as tef, could be included as a sequential crop for wheat, their highest yielding crop in the state, as it is a C_3 grass (USDA, 2016). Other states such as Delaware, Kentucky, or Ohio could also take on tef as a specialty crop as their percentage of land with highly suitable rankings were high in comparison to the number of total hectares in the state. The farmers of these states can benefit from including a gluten-free grain for human consumption as gluten-free diets are on the rise (Lis et al. 2015).

Of the locations in which tef is known to be currently cultivated for grain use (Table 4.8), the site suitability calculations were all moderately suitable or highly suitable within a kilometer's (0.62 mi) radius of their location. The lowest suitability rankings were for the Idaho location with all of the land within the 1 km (0.62 mi) radius containing moderately suitable rankings. Nevertheless, promisingly, the Nevada location was comprised of 66% high suitability ranking. This suggests the proposed methodology is highly accurate, considering Nevada is one of the lowest for highly suitable acreage and percent of total state land (Figure 4.6). Montana and Oklahoma locations had similar percentages for highly suitable land. However, Oklahoma's tef farm was found to have the highest site suitability as nearly the entire (96%) 1 km (0.62 mi) radius was composed of highly suitable rankings. This highly suitable location in Oklahoma is a prime example of how tef can be utilized as a sequential crop for C_3 crops, as SS Farms cultivate hot peppers (C_3) at the beginning of growing season, and directly following harvest, the farm cultivates tef for forage and grain.

There are some potential limitations in the use of tef as a sequential crop as the demands of the soil by the preceding crop or the crop to be grown following may be too high to incorporate the cultivation of tef. Nevertheless, following such an example as the farm in Oklahoma, or the sequential cowpea growers in Ethiopia (Berrada et al. 2006), the possibilities of areas of highly suitable land proposed could lead to an expansion of the implementation of tef into agricultural practice. The inclusion of this practice can alleviate some of the stresses that the growing livestock industry has presented (Tilman et al. 2002; Steinfield et al. 2006; Delgado et al. 1999; Boe et al. 1986).

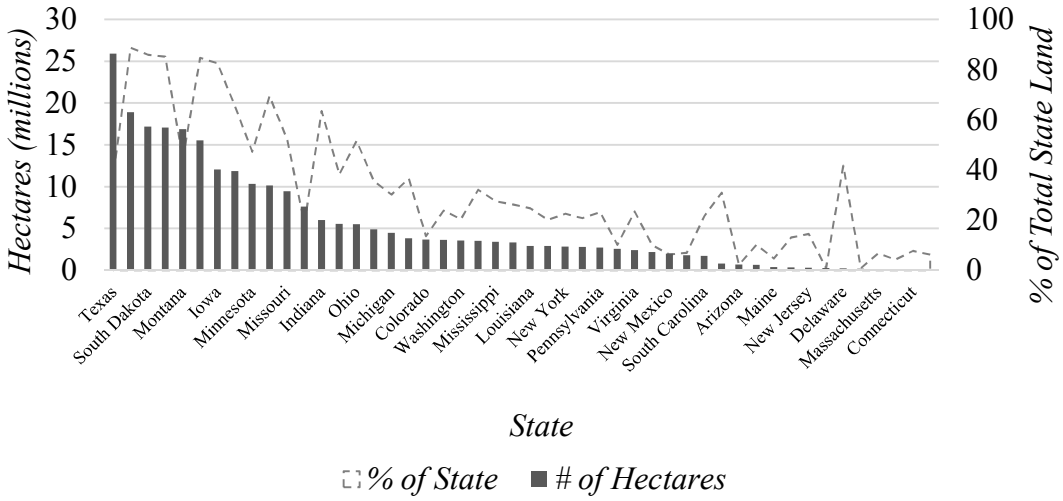

Figure 4.10. Highly suitable hectares and percentage of land within each state for late summer (July, August, September; JAS).

Table 4.20. Suitability analysis (JAS) for farms currently growing tef as a grain within the contiguous US (1 km radius).

Farms	**Location**	**Late Summer (JAS) Tef Suitability Within 1 km (%)**	
		Highly Suitable	***Moderately Suitable***
SS Farms	Hydro, OK	94	6
Prairie Heritage Farms	Great Falls, MT	92	8
Desert Oasis Tef	Fallon, NV	66	34
The Tef Company	Nampa, ID	0	100

4.5 CONCLUSION

This study aimed to combine the methods of weighted overlay and AHP to produce an assuring result for the regions within the U.S. that are suitable for tef to be employed as a sequential crop. The study was accomplished using eight criteria that reflected various environmental components affecting the cultivation of tef. The weights for eight criteria were found by employing AHP. The results of the weighted overlay suggest that within May, June, and July (MJJ) 32.38% of the U.S. is considered highly suitable for the cultivation of tef. Additionally, within July, August, and September (JAS) 32.68% of the U.S. is considered highly suitable for the cultivation of tef.

It was also found that many states could incorporate tef as a way to support livestock, sequentially grow with wheat, and begin cultivation of a specialty crop as the geographic regions and their current agricultural practices would allow for such an addition. The accuracy of the results was confirmed by the incorporation of the four current farm

locations cultivating tef grain for market. Within a 1km (0.62 mi) radius of each of these farms, the rankings of the land were moderately and highly suitable, the two highest ranks. These findings suggest the use of the weighted overlay and AHP are accurate in the identification of locations in which tef can be cultivated. Overall, the geospatial information provided could aid agricultural managers in their decision to implement tef as a sequential crop to support the needs of the livestock industry to meet the demands of population growth and consumers.

CHAPTER V

CONCLUSION

5.1 INTRODUCTION

As food crops such as tef (*Eragrostis tef*) continue to experience globalization with respect to cultivation, it is imperative that precision agriculture methods are fitted to the new environments in which tef is grown. Fitting the precision agriculture methods is pertinent to farmers that rely on the methods to inform their farm management practices. Thus, the research within this dissertation aimed to: (1) investigate the ability to use imaging spectroscopy (IS) to predict nutrient contents of tef plant and grain, and determine if these methods are replicable across environmental and international contexts; (2) determine whether the chlorophyll content of tef plants can be predicted using chlorophyll indices originally developed for use with hyperspectral data but mimicked using multi-spectral imagery bands, and similarly determine if these methods are replicable across different environments; and (3) identify suitable locations for tef cultivation in the United States employing the analytical hierarchy process (AHP) and weighted overlay using a multi-criteria suitability evaluation. The first two research aims were designed to investigate remote sensing methods that benefit tef farmers and researchers, while the third research aim was designed to inform and identify regions of the United States that could both cultivate and benefit from the identified precision

agriculture methods in the first two research aims. All three research aims share the common theme of contributing to geospatial analyses of an under-researched crop with the potential to contribute to food security issues.

5.2 REVISITING THE METHODS

Geospatial methods and technologies are responsible for the advancement of GPS, GIS, and remote sensing technologies (Zhange et al. 2002). The field of precision agriculture integrates many of these geospatial technologies to inform efficient resource use such as fertilizer and irrigation practices. Of the geospatial methods and technologies incorporated into precision agriculture, the use of IS has become a method of interest as many studies have found that IS can provide instantaneous insights into biochemical makeup of scanned plants (Apan et al. 2006; Overgaard et al. 2013a; Rabbotnikof et al. 1995; Beeri et al. 2007). When investigating potential for IS methods and data to predict nutrient contents (e.g. calcium, magnesium, and protein) of tef plant and grain and the replicability of those methods across differing environmental contexts, the methods employed the use of waveband creation from hyperspectral data, spectral preprocessing (e.g. Savisky-Golay smoothing, first derivative, and second derivative), and a partial least squares waveband selection method (Kawamura et al. 2008; Kawamura et al. 2018). In addition to the scientific methods for these studies we also met with farmers to ensure they knew the potential benefits of the research. In particular, in Ethiopia I was interested in identifying whether the farmers would be willing to take on such a technology. While many of the regression fit results identified between IS data and nutrients were high (up to R^2=0.92), amongst differing environments the research identified contradictory portion of the electromagnetic spectrum, differing spectral preprocessing methods, and varying

coefficients of determination for a single nutrient between the environmental and international contexts. Thus, for the first research aim the results suggest, for the purpose of reproducibility and accuracy, when predicting nutrient content of agricultural products, IS models are best developed for single geographies.

Agricultural cultivation is global, and to explore plant health globally satellite platforms must be incorporated. However, many chlorophyll indices were originally developed for hyperspectral data rather than satellite platform spectral resolutions. An exploration of publicly available satellite data (e.g., Sentinel-2, Landsat-8) across multiple environments is required to lead to more applicable methods of determining plant health globally. Thus, to determine whether chlorophyll content of tef crops can be predicted using chlorophyll indices originally developed for hyperspectral data mimicked using multi-spectral imagery bands across differing environments this study used hyperspectral data to synthesize wavelengths to the spectral bands of Landsat-8 OLI and Sentinel-2 MSI. Of the 15 chlorophyll indices calculated for hyperspectral and satellite level spectral sensitivities, it was found that simple pigment chlorophyll indices resulted in the highest correlations for the study sites within both Ethiopia and the United States. Moreover, a simple pigment index incorporating blue and red pigments resulted in the greatest correlations (up to R^2=0.79) for all chlorophyll concentrations (*a*, *b*, *a*+*b*). This finding greatly benefits farmers and researchers aiming to obtain an aerial view of tef crop health within their fields as Landsat-8 OLI and Sentinel-2 MSI are open source data platforms with high spatial and temporal resolutions.

Another geospatial technique frequently used in agriculture is site suitability analysis. Site suitability in agriculture can aid farm managers in determining whether a

crop has potential to be grown on their lands. Tef is a C_4 grass that could serve as a sequential forage crop for C_3 crops (such as wheat) during the summer months, potentially providing support for the growing United States cattle industry. Thus, the third study in this dissertation implemented a weighted overlay site suitability analysis using geospatially referenced elevation, slope, insolation, soil type, average precipitation, average temperature, minimum temperate, and land cover data to determine regions within the United States that are suitable for tef cultivation. Using the eight criteria reflecting various environmental components important to the cultivation of tef, weights of importance were identified using the AHP. The weights identified were used to conduct a weighted overlay analysis that resulted in the identification 32.38% of the United States being highly suitable in early summer months (e.g. May, June, July) and 32.68% in late summer months (e.g. July, August, September). In addition to the identification of regions highly suitable for tef cultivation, the accuracy of the method was also tested by incorporating the use of four known tef cultivation locations across the United States (i.e. Oklahoma, Montana, Nevada, and Idaho). It was found that this method of site suitability was highly accurate as for within a 1 km radius of each of the known locations the land was highly or moderately suitable for tef cultivation. The results of the highly accurate method provides agricultural managers geospatial information that will support decision to potentially incorporate tef cultivation.

5.3 SYNTHESIS

Agricultural specialty crops such as tef are currently subject to rapid growing globalization. Food geographers and remote sensors must work together to identify the new environments that could, or currently are, cultivating these newly introduced crops

around the globe. Within precision agriculture, it is of utmost importance to ensure that sensors and models used in the remote sensing methods of precision agriculture are calibrated to different environments around the globe. Furthermore, based on my experiences in Ethiopia, Ethiopian farmers are enthusiastic about the potential to incorporate precision agriculture technologies into their cultivation practices.

Thus, the first study aimed to identify whether IS methods and data to predict nutrient contents (e.g. calcium, magnesium, and protein) of tef plant and grain could be replicated across differing environmental contexts. The results identified relevant wavebands that future remote sensing studies can explore for each nutrient at the plant and grain level across two differing environments. The study also concluded that with current methods of employing IS data to predict nutrients, the models created should be formed for single environments as the spectral preprocessing, relevant wavebands, and coefficients of determination differed for each respective environment and nutrient. Additionally, future studies in precision agriculture are cautioned when synthesizing samples and their corresponding nutrient data for two differing environments as the PLS models generally suffered from overfitting. Nevertheless, the models created to predict nutrients for single geographies performed well for Ca, Mg, and protein at both the plant and grain levels with R^2 values ranging from 0.28-0.92.

The second study aimed to translate chlorophyll prediction indices that are typically computed using ground-based hyperspectral data into a form that can be captured with multispectral imagery for two differing environments. The results suggested that simple pigment indices performed the best when predicting chlorophyll across two different environments. Furthermore, these types of indices correlated best to

chlorophyll content when hyperspectral wavelengths were convolved to bands matching the Landsat-8 and Sentinel-2 sensing instruments. The findings are beneficial for the globalization of food crops such as tef, as farmers and remote sensing scientist can use similar methods to convolve hyperspectral data to satellite spectral sensitivities allowing a more global collection of remote sensing data and thus globalization of precision agriculture practices.

As precision agriculture practices are fitted globally, the methods can contribute to the successful globalization of crops such as tef. Thus, a preconceived understanding of where tef can be grown is beneficial to farm managers searching for new crops already fitted to precision agriculture methods. Thus, the third study aimed to utilize another geospatial method known as site suitability to identify locations in the United States that tef could be cultivated as a sequential crop during summer months. The results suggested that the contiguous United States is highly suitable for tef cultivation for 32.38-32.68% of available lands. In particular, the findings suggest that states such as Texas, Kansas, South Dakota, North Dakota, Nebraska, Iowa, Delaware, Kentucky, and Ohio could readily cultivate tef as a sequential crop for wheat to support livestock forage demands. Additionally, for future food geographers interested in site suitability, this study assessed the accuracy of the methods employed by comparing locations known within the United States to those known to currently cultivate tef and found the technique to be highly accurate.

Each of these research aims employed methods rooted in geographic theory and application. The first two aims utilized remote sensing techniques that many would consider a sub-field of geography. Additionally, both of these studies incorporated the

basic geographic principle of place as the studies were concerned with how the remote sensing methods worked across differing international boundaries and environments. The third research aim was a bit more obvious in its geographic procedures as maps were produced throughout the methods and results. Nevertheless, using spatial analysis (i.e. AHP and weighted overlay) this study identified geographic regions that farms could incorporate tef cultivation, and thus, potentially benefit from the precision agriculture methods investigated in the previous two studies.

5.4 LIMITATIONS

The limitation of the first two research aims are similar in that the field, collections, and methods were similar. The main limitations would first be the locations and number of locations in which the data and samples were collected. IS and plant sample collection takes a lot of time, energy, and most importantly money. While the research could have benefitted from the inclusion of additional environments, the time allotted for crops near harvest along with clear sunny days pose challenges with time. Additionally, the cost of visiting locations for sample collection was a limiting factor as well.

As for the site suitability analysis in the third study, the main limitation is with the accuracy assessment. For this study there were only four sites used for the accuracy assessment as these were the only locations the research could identify that were currently cultivating tef. With more locations for the accuracy assessment more in depth statistics could be utilized. Thus, future studies may want to incorporate crops that are more readily grown in a given location to further analyze the accuracy of the methods constructed.

5.5 AVENUES FOR FUTURE RESEARCH

Future research can further exam the performance and accuracy of the methods explored. Additionally, further exploration of nutrient prediction using IS data across environments should be analyzed. In particular, the identification of a method of model creation that would result in high correlation coefficients without overfitting would be a great contribution to the field of remote sensing. Within chlorophyll analysis, the exploration of the red-edge index could lead to greater accuracy across environments. Many studies have begun to analyze alternative methods for red-edge calculation, thus a further exploration of these methods across environments would be beneficial to the field of remote sensing. Finally, future research for site suitability methods could explore the use of new variables and new rankings of those variables to better understand the importance of various variables and ranking structures. Overall, future research within the fields of food geography and precision agriculture can utilize the methods proposed within these studies to serve as a foundation for applications in other crops currently experiencing globalization that agricultural managers may be interested in incorporating into their cultivation practices.

REFERENCES

Akinci, H., A.Y. Ozalp, and B. Turgut. 2013. Agricultural land use suitability analysis using GIS and AHP technique. *Computers and Electronics in Agriculture*, 97: 71-82.

Al-Abbas, A.H., R. Barr, J.D. Hall, F.L. Crane, and M.F. Baumgardner. 1974. Spectra of normal and nutrient deficient maize leaves. *Agronomy Journal*, 66: 16-20.

Apan, A., R. Kelly, S. Phinn, W. Strong, D. Lester, D. Butler, and A. Robson. 2006. Predicting grain protein content in wheat using hyperspectral sensing of in-season crop canopies and partial least squares regression. *International Journal of Geoinformatics*, 2(1): 93-108.

Asendorpf, J.B., M. Conner, F. De Fruyt, J. De Houwer, J.J. Denissen, K. Fiedler, S. Fiedler, D. C. Funder, R. Kliegl, B.A. Nosek, and M. Perugini, 2013. Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27(2): 108-119.

Aydi A. 2018. Evaluation of groundwater vulnerability to pollution using a GIS-based multi-criteria decision analysis. *Groundwater for Sustainable Development,* 7: 204–211.

Aydi A., T. Abichou, I. Hamdi Nasr, M. Louati, M. Zairi. 2016. Assessment of land suitability for Olive mill wastewater Disposal Site Selection by integrating fuzzy logic, AHP and WLC in a GIS. *Environmental Monitoring and Assessment,* 188:59.

Baker, M. 2016. 1500 scientists lift the lid on reproducibility. *Nature*, 533(7604): 452-454.

Bandyopadhyay, S., R.K. Jaiswal, V.S. Hedge, V. Jayarman. 2009. Assessment of land suitability potentials for agriculture using remote sensing and GIS based approach. *International Journal of Remote Sensing*, 30(4): 879-895.

Baret, F. and G. Guyot. 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of the Environment*, 35: 161-173.

Baret, F., S. Jacquemoud, G. Guyot, and C. Leprieur. 1992. Modelled analysis of the biophysical nature of spectral shifts and comparison with information content of broad bands. *Remote Sensing of the Environment*, 41: 133-142.

Bari, J.A., K. Lee, G. Kvaran, B.L. Markham, and J.A. Pedelty. 2014. The spectral response of the Landsat-8 operational lab imager. *Remote Sensing*, 6: 10232-10251.

.

Barnes, G. 2010. Soil mechanics: principles and practice. Basingstoke, Palgrave Macmillan

Beeri, O., R. Phillips, J. Hendrickson, A.B. Frank and S. Kronberg. 2007. Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie. *Remote Sensing of Environment*, 110: 216-225.

Begley, C.G. and J.P. Ioannidis 2015. Reproducibility in science: improving the standard for basic and preclinical research. *Circulation research*, 116(1): 116-126.

Bjorkman, O. and R.W. Pearcy. 1971. Effect of growth temperature on the temperature dependence of photosynthesis in vivo and CO3 fixation by carboxydismutase in vitro in C3 and C4 species. *Caregie Inst Yearbook*, 70: 511-520.

Bjorkman, O., R.W. Pearcy, and M.W. Nobs. 1970. Photosynthetic characteristics. *Carnegie Inst Yearbook*, 70: 511-520.

Blackburn, G. A. 1998. Quantifying chlorophylls and carotenoids at leaf and canopy scales; An evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, 66(3): 273 -285.

Boe, A., J. Somerfieldt, R. Wynia, and N. Thiex. 1986. A preliminary evaluation of the forage potential of teff. *Proceedings of the South Dakota Academy of Science*, 65, 75-82.

Boochs, F., G. Kupfer, K. Dockter, and W. Kuhbauch. 1990. Shape of the red edge as vitality indicator for plants. *International Journal of Remote Sensing*, 11(10): 1741−1753.

Broge, N. H. and E. Leblanc. 2000. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76: 156 – 172.

Bultosa, G. and J.N.R. Taylor. 2004. Tef. In: Wringley C., H. Corke, and C. Walker (eds) Encyclopedia of grain science. Academic, Oxford, 281-289.

Buschmann, C. and E. Nagel. 1993. In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing*, 14(4): 711−722.

Calvoa, T. and M.F. Pessoa. 2015. Remote sensing in food production – a review. Emirates *Journal of Food and Agriculture*, 27(2): 138-151.

Camerer, C.F., A. Dreber, E. Forsell, T. H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, and E. Heikensten 2016. Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280): 1433-1436.

Caporaso, N., M.B. Whitworth, and I.D. Fisk. 2018. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chemistry*, 240: 32-42.

Carter, G.A. 1994. Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing*, 15(3): 697-703.

Chang, S.H. and W. Collins. 1983. Confirmation of the airborne biogeophysical mineral exploration technique using laboratory methods. *Economic Geology and the Bulletin of the Society of Economic Geologists*, 78: 723-736.

Chen, J. 1996. Evaluation of vegetation indices and modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, 22: 229–242.

Cho, M.A. and A.K. Skidimore. 2006. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment*, 101: 181-193.

Clevers, J.G.P.W. 1999. The use of imaging spectrometry for agricultural applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54: 299-304.

Clevers, J.G.P.W. et al. 2002. Derivation of the red edge index using MERIS standard band setting. *International Journal of Remote Sensing*, 223(16): 3169-3184.

Climate-data.org. 2017. Climate. Climate-data.org. Retrieved from: https://en.climate-data.org/. Last visited: July 18, 2017.

Collins, W., G.L. Raines, and F.C. Canney. 1977. Airborne spectroradiometer discrimination of vegetation anomalies over sulphide mineralisation—a remote sensing technique. Abstract with programs (pp. 932−933). Seattle, Washington: Geological Society of America.

Cozzolino, D. and A. Moron. 2004. Exploring the use of near infrared reflectance spectroscopy (NIRS) to predict trace minerals in legumes. *Animal Feed Science and Technology*, 111: 161-173.

Curran, P.J. 1989. Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30: 271-278.

Curran, P.J. and J.L. Dungan. 1989. Estimation of signal-to-noise: a new procedure applied to AVIRIS data. *IEEE Transactions on Geoscience and Remote Sensing*, 27: 620-628.

Curran, P.J., J.L. Dungan, and H.L. Gholz. 1990. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree Physiology,* 7: 33-48.

Curran, P.J., J.L. Dungan, D.L. Peterson. 2001. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: Testing the Kokaly and Clark methodologies. *Remote Sensing of Environment*, 76: 349-359.

Curran, P.J., W.R. Windham, and H.L. Gholz. 1995. Exploring the relationship between reflectance red edge and chlorophyll concentration in slash pine leaves. *Tree Physiology*, 15: 203-206.

Daughtry, C. S. T., C.L. Walthall, M.S. Kim, E. Brown de Colstoun, and J.E. McMurtrey III. 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of the Environment*, 74: 229-239

Dawson, T. P. and P.J. Curran. 1998. A new technique for interpolating red edge position. *International Journal of Remote Sensing*, 19(11): 2133−2139.

Dawson, T.P., P.J. Curran, P.R.J. North, S.E. and Plummer. 1999. The propagation of foliar biochemical absorption features in forest canopy reflectance: a theoretical analysis. *Remote Sensing of Environment,* 67: 147-159.

de Jong, S. 1993. SIMPLS:An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18: 251-263.

Dekking, L.S., Y.K. Winkelaar, and F. Koning. 2005. The Ethiopian cereal tef in celiac disease. *The New England Journal of Medicine*, 353: 1748-1749.

Delden, S., J. Vos, T. Stomph, G. Brouwer, and P. Stuik. 2012. Photoperiodism in *Eragrostis tef*: Analysis of ontogeny and morphology in response photoperiod. *European Journal Journal of Agronomy*, 37, 105-114.

Delgado, C.L. 2005. Rising demand for meat and milk in developing countries: implications for grasslands-based livestock production. In: DA, M.G. (ed.), Grassland: A global resource. Wageningen Academic Publishers, Wageningen(The Netherlands), 29-39.

Delgado, C.L., M.W. Rosegrant, H. Steinfeld, S.K. Ehui, and C. Courbois. 1999. Livestock to 2020: The next food revolution. IFPRI, Washington DC.

Demetriades-Shah, T. H., M.D. Steven, and J.A. Clark. 1990. High resolution derivative spectra in remote sensing. *Remote Sensing of the Environment*, 33: 55-64.

Demissie, A. 2000. Tef genetic resources in Ethiopia. In: Tefera, H., G. Belay, and M. Sorrells (eds) Narrowing the rift: tef research and development. Debrezeit, Ethiopia, 27-31.

ESA. 2018. Spectral and spatial resolutions: Sentinel-2. European Space Agency. https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath. (Last Accessed: 11 April 2018)

Evans, J.R. 1989. Photosynthesis and nitrogen relationships in leaves of C3 plants. *Oecologia*, 78: 9-19.

Feizizadeh, B., and T. Blaschke. 2012. Land suitability analysis for Tabriz County, Iran: a multi criteria evaluation approach using GIS. *Journal of Environmental Planning and Management*, 1: 1-23.

Field, C and H.A. Mooney. 1986. The photosynthesis-nitrogen relationship in wild plants. In T.J. Givnish (ed.). *On economy of plant from and function* (pp. 22-55). Cambridge, UK: Cambridge University Press.

Flynn, K.C. Under revision. Site suitability analysis for tef (Eragrostis tef) within the contiguous United States.

Flynn, K.C., A.E. Frazier, and S.A. Admas. Under review. Translating hyperspectral indices to multi-spectral sensors for chlorophyll prediction across multiple environment.

Food and Agriculture Organization of the United Nations (FAO). 2016a. FAOSTAT database. (Last Accessed: 12 October 2016).

Food and Agriculture Organization of the United Nations (FAO). 2016b. Grassland species profiles: Eragrostis tef (Zucc.) Trotter. http://www.fao.org/ag/agp/AGPC/doc/Gbase/data/pf000247.htm. (Last Accessed: 12 October 2016).

Food and Agriculture Organization of the United Nations (FAO). 2016c. GIEWS – Global Information and Early Warning Signs. http://www.fao.org/giews/countrybrief/country.jsp?code=ETH. (Last Accessed: 12 October 2016).

Forage Analyses Procedures. 1993. National Forage Testing Association. Pp. 79-81.

Foster, A.J., V.G. Kakani, and J. Mosali. 2017. Estimation of bioenergy crop yield and N status by hyperspectral canopy reflectance and partial least square regression. *Precision Agriculture*, 18: 192-209.

Gao, X., A. Huete, W. Ni, and T. Miura. 2000. Optical biophysical relationships of vegetation spectra without background contamination. *Remote Sensing of Environment*, 74: 609−620.

Gates, D. M., H. J. Keegan, J.C. Schleter, and V.R. Weidner. 1965. Spectral properties of plants. *Applied Optics*, 4(1): 11−20.

Geladi, P. and B. Kowalski. 1986. Partial least squares regression: A tutorial. Analytica Chimica Acta, 185:1-17.

Gerbremariam, M.M., M. Zarnkow, and T. Becker. 2014. Tef (*Eragrostis tef*) as a raw material for malting, brewing and manufacturing of gluten-free foods and beverages: a review. *Journal of Food Science and Technology*, 51(11): 2881-2895.

Gitelson, A.A. and M.N, Merzlyak. 1994. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, 143: 286-292.

Gitelson, A.A. and M.N. Merzlyak. 1996. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology*, 148: 494-500.

Gosse, G., C. Varlet-Grancher; R. Bonhomme; M. Chartier; J.M. Allirand and G. Lemaire. 1986. Maximum dry matter production and solar radiation intercepted by a canopy. *Agronomie*, 6: 47-56.

Gupta, P., K. Singh, V. Seth, S. Agarwal, and P. Mathur. 2014. Association of food insecurity and malnutrition among young children (6-36 Months). *The Indian Journal of Nutrition and Dietetics*, 51(3).

Guyot, G., F. Baret, and S. Jacquemoud. 1989. Imaging spectroscopy for vegetation studies. In: F. Toselli and J. Bodechtel (Eds.), Imaging spectroscopy: fundamentals and prospective applications (pp. 145 - 165). Kluwer Academic Publishers, Dordrecht.

Haboudane, D., J.R. Miller, E. Pattey, P.J. Zarco-Tejada, and I. Strachan. 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3): 337-352.

Haboudane, D., J.R. Miller, N. Tremblay, P.J. Zarco-Tejada, and L. Dextraze. 2002. Integration of hyperspectral vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, 81(2 – 3): 416-426.

Hall, D. O. and K.K. Rao. 1987. *Photosynthesis* (4th ed.). Great Britain: Edward Arnold.

Hare, E.W., J.R. Miller, and G.R. Edward. 1984. Studies of vegetation red reflectance edge in geobotanical remote sensing in eastern Canada. *Proceedings of the 9th Canadian Symposium on Remote Sensing*, held at St, John's, Newfoundland, 13–17 August 1984 (pp. 433−440). Ottawa: Canadian Aeronautics and Space Institute.

Homer, C.H., J.A. Fry, and C.A. Barnes. 2012. The National Land Cover Database. U.S. Geological Survey Fact Sheet, 2012-3020: 4.

Hopman, G.D., E.H.A. Dekking, M.L.J. Blokland, M.C. Wuisman, W.M.K.F. Zuijderduin, and J.J. Schweizer. 2008. Tef in the diet of celiac patients in the Netherlands. Scandinavian Journal of Gastroenterology, 43: 277-282.

Hopman, G.D., E.H.A. Dekking, M.L.J. Blokland, M.C. Wuisman, W.M.K.F. Zuijderduin, and J.J. Schweizer. 2008. Tef in the diet of celiac patients in the Netherlands. Scandinavian *Journal of Gastroenterology*, 43: 277-282.

Horler, D. N. H., J. Barber, and A.R. Barringer. 1980. Effects of heavy metals on the absorbance and reflectance spectra of plants. *International Journal of Remote Sensing*, 1: 121.

Horler, D.N.H., M. Dockray, and J. Barber. 1983. The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2): 273-288.

Huete, A.R. 1988. A soil adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 36: 54-53.

Ioannidis, J.P., T.D. Stanley, and H. Doucouliagos. 2017. The power of bias in economics research. *The Economic Journal*, 127(605): F236-F265.

Jankowski, P. 1995. Integrating geographical information system and multiple criteria decision making methods. *International Journal of Geographic Information Systems*, 9(3): 251-273.

Javadian, M., H. Shamskooshki, and M. Momeni. 2011. Application of sustainable urban development in environmental suitability analysis of educational land use by using AHP and GIS in Tehran. *Procedia Engineering*, 21: 72-80.

Jensen, J.R. 2016. *Introductory Digital Image Processing: A Remote Sensing Perspective.* Upper Saddle River, N.J.: Prentice Hall.

Jordan, C.F. 1969. Derivation of leaf area index from quality of light on the forest floor. *Ecology*, 50: 663–666.

Kawamura, K., H. Ikeura, S. Phongchanmaixay, and P. Khanthavong. 2018. Canopy hyperspectral sensing of paddy fields at the booting stage and PLSL regression can assess grain yield. *Remote Sensing*, 10: 1249.

Kawamura, K., N. Watanabe, S. Sakanoue, and Y. Inoue. 2008. Estimating forage biomass and quality in a mixed sown pasture based on partial least squares regression with waveband selection. *Japanese Society of Grassland Science*, 54: 131-145.

Kawamura, K., T. Akiyama, H. Yokota, M. Tsutsumi, T. Yasuda, O. Watanabe, et al. 2005. Quantifying grazing intensities using geographic information system and satellite remote sensing in the Xilingol steppe region, Inner Mongolia, China. Agriculture, *Ecosystems and Environment*, 107: 83−93.

Kebede, H. 1986. Photosynthetic response to temperature in *Eragrostis tef* (Zucc,) Trotter (Masters Thesis). Retrieved from https://shareok.org/bitstream/handle/11244/15855/Thesis-1986-K25p.pdf?sequence=1&isAllowed=y.

Kedron, P.J., A.E. Frazier, A.B. Trgovac, T. Nelson and A.S. Fotheringham. Under revision. Reproducibility and replicability in geographical analysis.

Kim, M.S., C.S.T. Daughtry, E.W. Chappelle, J.E. McMurtrey III, and C.L. Walthall. 1994. The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (Apar). *Proceedings of the 6th Symposium on Physical Measurements and Signatures in Remote Sensing*, January 17 – 21, 1994, Val D'Isere, France (pp. 299-306).

Kokaly, R.F. and R.N. Clark. 1999. Spectroscopic determination of leaf biochemistry using bad-depth analysis of absorption features and step-wise multiple linear regression. *Remote Sensing of Environment*, 75: 153-161.

Kokaly, R.F. and R.N. Clark. 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment,* 67: 267-287.

Kokaly, R.F., G.P. Asner, S.V. Ollinger, M.E. Martin, and C.A. Wessman. 2009. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment*, 113: 578-591.

Koulouri, M. and C. Giorga. 2007. Land abandonment and slope gradient as key factors of soil erosion in Mediterranean terraced lands. *Catena*, 69(3): 274-281.

Kumar, L., K.S. Schmidt, S. Dury, and A.K. Skidmore. 2001. Imaging spectrometry and vegetation science. In F. van der Meer, and S.M. de Jong Eds), Imaging spectrometry (111-155). Dordrecht, The Netherlands: Kluwer Academic Publishing.

Lamb, D.W., M. Steyn-Ross, P. Schaare, M.M. Hanna, W. Silvester, and A. Steyn-Ross. 2002. Estimating leaf nitrogen concentration in ryegrass (*Lolium* spp.) pasture

using the chlorophyll red-edge: Theoretical modelling and experimental observations. *International Journal of Remote Sensing*, 23(18): 3619-3648.

le Maire, G., C. Francois, and E. Dufrene. 2004. Toward universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment*, 89: 1-28.

Lichtenthaler, H. K., A.A. Gitelson, and M. Lang. 1996. Nondestructive determination of chlorophyll content of leaves of a green and an aurea mutant of tobacco by reflectance measurements. *Journal of Plant Physiology*, 148: 483 - 493.

Lichtenthaler, H.K. 1998. The stress concept in plants: An introduction. *Annals of the New York Academy of Science*, 851: 187-198.

Lindberg, W., J.A. Persson, and S. Wold. 1983 Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and linginsulfonate. *Analytical Chemistry*, 55: 643-648.

Lis, D.M., T. Stellingweriff, C.M. Shing, K.D.K. Ahuja, and J.W. Fell. 2015. Exploring the popularity, experiences, and beliefs surrounding gluten-free diets in nonceliac athletes. *International Journal of Sports Nutrition and Exercise Metabolism*, 25: 37-45.

Liu, L., J. Wang, X. Song, C. Li, W. Huang, C. Zhao. 2005. The canopy spectral features and remote sensing of wheat lodging. *Journal of Remote Sensing,* 9(3): 323-327.

Long, S.P. 1983. C4 photosynthesis at low temperatures. *Plant, Cell and Environment*, 6: 345-363.

Long, S.P., L.D. Incoll, and H.W. Woolhouse. 1975. C_4 photosynthesis in plants from cool temperate regions, with particular reference to Spartina townsendii. *Nature*, 257: 622-624.

Loomis, R.S. 1983. Productivity of agricultural ecosystems. In: Lange, O.L., P.S. Nobel, C.B.Osmond, and H. Ziegler (eds) Physiological Plant Ecology IV. Ecosystem Processes: Mineral Cycling, Productivity and Man's Influence. Encyclopedia of Plant Physiology New Series, Vol 12D, 151-203.

Lorber, A., L.E. Wangen, and B.R. Kowalski. 1987. A theoretical foundation for the PLS algorithm. *Journal of Chemometirics*, 1: 19-31.

Ludlow, M.M. and G.L. Wilson. 1971. Photosynthesis of Tropical Pasture Plants. I. Illuminance, carbon dioxide concentration, leaf temperature and leaf air-vap pressure difference. *Australian Journal of Biological Sciences*, 24: 449-470.

Lyon, J.G., D. Yuan, R.S. Lunetta, and C.D. Elvidge. 1998. A change detection experiment using vegetation indices. *Photogrammetric Engineering and Remote Sensing*, 62: 143-150.

Mariotti, M., L. Ercoli, and A. Masoni. 1996. Spectral properties of iron deficient corn and sunflower leaves. *Remote Sensing of Environment*, 58(3): 282 - 288.

Martinez-Valdivieso, D., R. Font, P. Gomez, T. Blanco-Diaz, and M.D. Rio-Celestino. 2014. Determining the mineral composition in Cucurbita pepo fruit using near

infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture*, 94: 3171-3180.

Mauser, W. and H. Bach. 1995. Imaging spectroscopy in hydrology and agriculture: determination of model parameters. In: J. Hill and J. Megier (Eds.), Imaging spectrometry: a tool for environmental observations (pp. 261 - 283). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Mendas A. and A. Delali. 2012. Integration of multi-criteria decision analysis in GIS to develop land suitability for agriculture: application to durum wheat cultivation in the region og Mleta in Algeria. *Computers and Electronics in Agriculture*, 83: 117-126.

Miller, D. 2014. Tef grass: crop overview and forage production guide. http://www.kingsagriseeds.com/blog/wp-content/uploads/2014/12/Tef-Grass-Management-Guide.pdf. (Last Accessed: 12 October 2016).

Miller, D. 2014. Teff grass: crop overview and forage production guide. http://www.kingsagriseeds.com/blog/wp-content/uploads/2014/12/Teff-Grass-Management-Guide.pdf. (Last Accessed: 12 October 2016).

Miller, W., W. Collins, F.R. Steiner, and E. Cook. 1998. An approach for greenway suitability and analysis landscape and urban planning. *International Journal of Geographic Information Science*, 42(2-4): 91-105.

Mokarram, M., F. Aminzadeh. 2010. GIS-based multi-criteria land suitability evaluation using ordered weight averaging wiwht fuzzy quantifier: a case study in Shavur Plain, Iran. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(2): 508-512.

Monson, R.K., R.O. Littlejohn, and G.J. Williams. 1983. Photosynthetic adaptation to temperature in four species from the Colorado shortgrass steppe – a physiological model for coexistence. *Oecologia*, 58: 43-51.

Muchow, R.C., T.R. Sinclair, and J.M. Bennett. 1989. Temperature and Solar Radiation Effects on Potential Maize Yield across Locations. *Agronomy Journal*, 82(2): 338-343.

Mueller, N.D., J.S. Gerber, M. Johnston, D.K. Ray, N. Ramankutty, and J.A. Foley. 2012. Closing yield gaps through nutrient and water management. *Nature*, 490: 254-257.

Mulla, D.J. 2013. Twenty-five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4): 358-371.

Mutanga, O., A.K. Skidmore, and H.H.T. Prins. 2004. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment*, 89: 393-408.

Mutanga, O., A.K. Skidmore, and S. Van Wieren. 2003. Discriminating tropical grass canopies (C. ciliaris) grown under different nitrogen treatments using

spectroradiometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57: 263-272.

Naes, T. and H. Martens. 1984. Multivariate calibration II Chemometric methods. Trends in *Analytical Chemistry*, 3: 266-271.

National Renewable Energy Lab. 2017. NREL GIS data: Continental United States photovoltaic low resolution. Department of Energy. https://catalog.data.gov/dataset/nrel-gis-data-continental-united-states-photovoltaic-low-resolution. (Las Accessed: 11 November 2017).

Nellis, M.D., K.P. Price, and D. Rundquist. 2009. Remote sensing of cropland agriculture. Papers in Natural Resources, Paper 217. University of Nebraska-Lincoln. http://digitalcommons.unl.edu/natrespapaers/217.

Niinemets, U. and J.D. Tenhunen. 1997. A model separating leaf structural and physiological effects on carbon gain along light gradients for the shade-tolerant species *Ace saccharum. Plant, Cell and Environment*, 20: 845-866.

Oliver, M.A., T.F.A. Bishop, and B.P. Marchant. 2013. Precision agriculture for sustainability and environmental protection. New York, NY: Routledge.

Overgaard, S.I., T. Isaksson, and A. Korsaeth. 2013a. Prediction of wheat yield and protein using remote sensors on plots-Part I: Assessing near infrared model robustness for year and site variation. *Journal of Near Infrared Spectroscopy*, 21: 117-131.

Pearcy, R.W. and A.T. Harrison. 1974. Comparative photosynthetic and reparatory gas exchange characteristics of Atriplex lentiformis (Torr) Wats. In coastal and desert habitats. *Ecology*, 55: 1104-1111.

Phan-Thien, K.-Y., M. Golic, G.C. Wright, and N.A. Lee. 2011. Feasibility of estimating peanut essential mineral by near infrared reflectance spectroscopy. *Sensing and Instrumentation of Food Quality and Safety*, 5: 43-49.

Pramanik, M.K. 2016. Site Suitability analysis for agricultural land use of Darjeeling district using AHP and GIS techniques. *Modeling Earth Systems and Environment*, 2:56.

PRISM Climate Group, Oregon State University. 2016. Northwest Alliance for Computational Science and Engineering. http://prism.oregonstate.edu. (Last Accessed: 1 November 2016).

Pu, R., P. Gong, G.S. Biging, and M.R. Larrieu. 2003. Extraction of red edge optical parameters from Hyperion data for estimation of forest leaf area index. *IEEE Transaction on Geoscience and Remote Sensing*, 41(4): 916-921.

Qi, J., A. Chehbouni, A.R. Huete, Y.H. Kerr, and S. Sorooshian. 1994. A modified soil adjusted vegetation index (MSAVI). *Remote Sensing of Environment*, 48: 119-126.

Rabotnikof, C. M., G.M. Planas, J. Silva Colomer, N.P. Stritzler. 1995. Near infrared reflectance spectroscopy (NIRS) for predicting forage quality of perennial warm-season grasses in La Pampa, Argentina. Annales de Zootechnie, 44 (1): 97-

100.Royal Society of London. 2008. Sustainable Biofuels: Prospects and Challenges. Royal Society, London.

Raikes, C. and L.L. Burpee. 1998. Use of multispectral radiometry for assessment of Rhizoctonia blight in creeping bentgrass. *Photopathology Journal*, 88: 446-449.

Ramadan, Z., X-H. Song, P.K. Hopke, M.J. Johnson, K.M. Scow. 2001. Variable selection in classification of environmental soil samples for partial least square and neural network models. *Analytica Chimica Acta*, 466: 231-242.

Robert, P.C. 2002. Precision agriculture: a challenge for crop nutrition management. *Plant and Soil*, 247:143-149.

Rondeaux, G., M. Steven, and F. Baret. 1996. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55: 95-107.

Roseberg, R.J., S. Norberg, J. Smith, B. Charlton, K. Rykobst, C. Shock. 2005. Yield and quality of tef forage as a function of varying rates of applied irrigation and nitrogen. Annual Report of Klamath Experiment Station. http://oregonstate.edu/dept/kbrec/sites/default/files/yield_and_quality_of_tef_forage_as_a_function_of_varying_rates_of_applied_irrigation_and_nitrogen_2005.pdf . (Last Accessed: 11 November 2017).

Rougean, J. L. and F.M. Breon. 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, 51: 375-384.

Rouse, J.W., R.H. Haas, J.A. Schell, D.W. Deering, J.C. Harlan. 1974. Monitoring the vernal advancements and retrogradation of natural vegetation. NASA/GSFC final report, Greenbelt, MD, USA: 371.

Ruan-Ramos, A., A. Garcia-Ciudad, and B. Garcia-Criado. 1999. Near infrared spectroscopy prediction of mineral content in botanical fractions from semi-arid grasslands. *Animal Feed Science and Technology*, 77: 331-343.

Rundel, P.W. 1980. The ecological distribution of C4 and C3 grasses in the Hawaiin Islands. *Oecologia*, 45: 354-359.

Saaty, R.W. 1987. The analytic hierarchy process—what it is and how it is used. *Math Model* 9:161–176.

Saaty, T.L. 1980. The analytic hierarchy process: planning, priority setting, resource allocation. McGraw Hill International, New York.

Savitzky, A. and M.J.E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36: 1627-1639.

Schellberg, J. et al. 2008. Precision agriculture on grassland: Applications, perspectives and constraints. *European Journal of Agronomy*, 29: 59-71.

Schmidtlein, S. and J. Sassin. 2004. Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. *Remote Sensing of Environment*, 92: 126-138.

Schwarz, A.G. and R.E. Redmann. 1988. C4 grasses from the boreal forest region of northwestern Canada. *Canadian Journal of Botany*, 66: 2424-2430.

Smith, K.L., M.D. Steven, and J.J. Colls. 2004. Use of hyperspectral derivative ratios in the red edge region to identify plant stress responses to gas leak. *Remote Sensing of Environment*, 92: 207-217.

Soil Science Society of America (SSSA). Soil Testing and Plant Analysis. 3rd Ed. 1990: 404-411. Western States Laboratory Proficiency Testing Program. Soil and Plant Analytic Methods. Ver 4.00. 1997: 117-119.

Stallknecht, G.F., K.M. Gilbertson, and J.L. Eckhoff. 1993. Teff: Food crop for humans and animals. In: J. Janick and J.E. Simon (eds.), New crops. Wiley, New York, 231-234.

Steinfield, H. et al. 2006. Livestock's Long Shadow: Environmental Issues and Options. FAO, Rome.

Taffesse, A.S., P. Dorosh, and S. Asrat. 2011. Crop production in Ethiopia: Regional patterns and trends. International Food Policy Research Institute, Ethiopia Strategy Support Program, Addis Ababa. ESSP-2 Discussion Paper No. 16.

Thenkabail, P.S., R.B. Smith, and E. De Pauw. 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing Environment*, 71: 158-182.

Thomas, J.R. and G.F. Oerther. 1972. Generic combustion method for determination of crude protein in feeds: Collaborative study. *Journal Association of Official Analytical Chemists*, 72: 770-774.

Thornton, P.K. 2010. Livestock production: recent trends, future prospects. *Philosphical Transactions of the Royal Society B*, 365(1554), 2853-2867.

Tieszen, L.L. and J.K. Detling. 1983. Productivity of grassland and tundra. In: Lange, O.L., P.S. Nobel, C.B.Osmond, and H. Ziegler (eds) Physiological Plant Ecology IV. Ecosystem Processes: Mineral Cycling, Productivity and Man's Influence. Encyclopedia of Plant Physiology New Series, Vol 12D, 173-203.

Tilman, D., K.G. Cassman, P.A. Matson, R. Naylor, and S. Polasky. 2002. Agricultural sustainability and intensive production practices, *Nature*, 418, 671-677.

Tiwari, D.N., R. Loof, and G.N. Paudyal. 1999. Environmental-economic decision-making in lowland irrigated agriculture using multi-criteria analysis techniques. *Agricultural Systems*, 60(2): 99-112.

Todd, S.W. and R.M. Hoffer. 1998. Responses of spectral indices to variations in vegetation cover and soil background. *Photogrammetric Engineering & Remote Sensing*, 64: 915−921.

Triantaphyllou, E. and S.H. Mann. 1995. Using analytic hierarchy process for decision making in engineering applications: some challenges. *International Journal of Industrial Engineering: Theory, Applications, and Practice*, 2(1): 35-44.

Twidwell, E.K., A. Boe, and D.P. Casper. 2002. Tef: a new annual forage grass for South Dakota? South Dakota State University Extra Extension, 8071.

U.S.D.A. 2017. United States Department of Agriculture – National Agricultural Statistics Service. U.S. Department of Agriculture. https://www.nass.usda.gov/Data_and_Statistics/. Accessed: May 23, 2017.

Vane, G. and A.F.H. Goetz. 1988. Terrestrial imaging spectroscopy. *Remote Sensing of Environment*, 24: 1-29.

Voogd, H. 1983. Multicriteria evaluation for urban and regional planning. London: Pion.

Wang, F., G.B. Hall, and Subaryono. 1990. Fuzzy information representation and processing in conventional GIS software: data base design and applications. *International Journal of Geographic Information Systems*, 4(3): 261-283.

Wessman, C.A., J.D. Aber, D.L. and Peterson. 1989. An evaluation of imaging spectrometry for estimating forest canopy chemistry. *International Journal of Remote Sensing*, 10: 1293-1316.

Westphal, E. 1975. Agricultural systems in Ethiopia. Wageningen, the Netherlands, Centre for Agricultural Publishing and Documentation. Agriculture Resources Rept No. 826.

Xue, J. and B. Su. 2017. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, 2017: 1-17.

Yager, R.R. 1988. On Ordered Weighted Averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 8(1):183-190.

Yang, H., C. Erxue, L. Zengyuan, et al. 2015. Wheat lodging monitoring using polarimetric index from RADARSAT-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 34(1): 157-166.

Yang, M.D., K.S. Huan, Y.H. Kuo, H.P. Tsai, and L.M. Lin. 2016. Spatial and spectral hybrid image classification for rice lodging assessment through UAV imagery. *Remote Sensing,* 9: 583-602.

Yoder, B.J. and R.E. Pettigrew-Crosby. 1995. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400-2500 nm) at leaf and canopy scales. *Remote Sensing of Environment*, 53(3): 199-211.

Yu, J., Y. Chen, J. Wu, and S. Khan. 2011. Cellular automata-based spatial multi-criteria and land suitability simulation for irrigated agriculture. *International Journal of Geographic information Science*, 25(1): 131-148.

Zarco-Tejada, P.J. et al. 2005. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sensing of Environment*, 99:271-287.

Zewdie, H.T. and M. Muchie. 2014. The significance of whole grain tef fo improving nutrition: From injera to ready to eat porridge using extrusion cooking technology. *International Journal of African Development (IJAD)*, 2(1): 40-56.

Zhang J., X. Gu, J. Wang et al. 2012. Evaluating maize grain quality by continuous wavelet analysis under normal and lodging circumstances. *Sensor Letters*, 2012, 10(1/2): 580－585.

Zhang, N., Wang, M., and Wang, N. 2002. Precision agriculture: A worldwide overview. *Computers and Electronics in Agriculture*, 36: 113-132.

APPENDICES

APPENDIX A: All Synthesized Landsat-8 and Sentinel-2 Regression Results

LANDSAT-8 OLI

OLS regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*) and synthesized Landsat-8 (L8) indices. See Table 3.2 for index descriptions.

Broad-Band Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		Total C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj.* R^2	*RMSE*	*Adj.* R^2	*RMSE*
L8_NDVI	0.487	0.002	0.348	0.001	0.447	0.004
L8_SR	0.575	0.002	0.403	0.001	0.525	0.003
L8_MSR	0.568	0.002	0.398	0.001	0.518	0.003
L8_SAVI	0.464	0.003	0.310	0.001	0.418	0.004
L8_MSAVI	0.493	0.002	0.337	0.001	0.447	0.004
L8_TVI	0.330	0.003	0.196	0.001	0.288	0.004
L8_RGI	0.628	0.002	0.506	0.001	0.596	0.003
L8_Ultra BGI (BGI_1)	0.640	0.002	0.550	0.001	0.620	0.003
L8_BGI_2	0.398	0.003	0.415	0.001	0.409	0.004
L8_Ultra BRI (BRI_1)	0.733	0.002	0.581	0.001	0.693	0.003
L8_BRI_2	0.767	0.002	0.628	0.001	0.732	0.003

*L8_SRPI not included as it is the same index as L8_Ultra BRI at Landsat-8 OLI spectral resolution.

**Red Edge (nm) could not be included due to lack of spectral resolution.

Polynomial regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*) and synthesized Landsat-8 (L8) indices. See Table 3.2 for index descriptions.

Broad-Band Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		Total C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj.* R^2	*RMSE*	*Adj.* R^2	*RMSE*
LS_NDVI	0.584	0.002	0.424	0.001	0.537	0.003
LS_SR	0.611	0.002	0.445	0.001	0.563	0.003
LS_MSR	0.579	0.002	0.409	0.001	0.529	0.003
LS_SAVI	0.612	0.002	0.450	0.001	0.565	0.003
LS_MSAVI	0.599	0.002	0.437	0.001	0.552	0.003
LS_TVI	0.562	0.002	0.435	0.001	0.526	0.003
LS_RGI	0.650	0.002	0.505	0.001	0.610	0.003
LS_Ultra BGI (BGI_1)	0.674	0.002	0.584	0.001	0.653	0.003
LS_BGI_2	0.420	0.003	0.450	0.001	0.435	0004
LS_Ultra BRI (BRI_1)	0.738	0.002	0.609	0.001	0.703	0.003
LS_BRI_2	0.785	0.002	0.689	0.001	0.761	0.002

*L8_SRPI is redundant to L8_Ultra BRI.
**Red Edge (nm) could not be included due to lack of spectral resolution.

SENTINEL-2 MSI

OLS regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*) and synthesized Sentinel-2 (S2) indices. See Table 3.2 for index descriptions.

Broad-Band Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		Total C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj.* R^2	*RMSE*	*Adj.* R^2	*RMSE*
S2_$NDVI_1$	0.477	0.003	0.337	0.001	0.437	0.004
S2_$NDVI_2$	0.472	0.003	0.335	0.001	0.433	0.004
S2_SR_1	0.572	0.002	0.399	0.001	0.522	0.003
S2_SR_2	0.568	0.002	0.396	0.001	0.518	0.003
S2_MSR_1	0.564	0.002	0393	0.001	0.514	0.003
S2_MSR_2	0.559	0.002	0.390	0.001	0.510	0.004
S2_$SAVI_1$	0.461	0.003	0.305	0.001	0.414	0.004
S2_$SAVI_2$	0.450	0.003	0.298	0.001	0.404	0.004
S2_$MSAVI_1$	0.491	0.002	0.330	0.001	0.443	0.004
S2_$MSAVI_2$	0.480	0.002	0.324	0.001	0.434	0.004

S2_TVI_1	0.355	0.003	0.214	0.001	0.311	0.004
S2_TVI_2	0.315	0.003	0.184	0.001	0.274	0.004
S2_RGI	0.589	0.002	0.466	0.001	0.557	0.003
S2_Ultra BGI (BGI_1)	0.626	0.002	0.542	0.001	0.607	0.003
S2_BGI_2	0.580	0.002	0.568	0.001	0.584	0.003
S2_Ultra BRI (BRI_1)	0.719	0.002	0.564	0.001	0.678	0.003
S2_BRI_2	0.721	0.002	0.572	0.001	0.682	0.003
S2_Red Edge (nm)	0.648	0.002	0.515	0.001	0.613	0.003

*S2_SRPI is redundant to S2_Ultra BRI.

Polynomial regression results (R^2 and the *RMSE*) for chlorophyll *a* (C*a*), chlorophyll *b* (C*b*), and total chlorophyll (C*a*+*b*) and synthesized Sentinel-2 (S2) indices. See Table 3.2 for index descriptions.

Broad-Band Indices	Chlorophyll (C) Measures					
	C*a* (g/g)		C*b* (g/g)		Total C*a*+*b* (g/g)	
	Adj. R^2	*RMSE*	*Adj.* R^2	*RMSE*	*Adj.* R^2	*RMSE*
S2_$NDVI_1$	0.585	0.002	0.425	0.001	0.538	0.003
S2_$NDVI_2$	0.578	0.002	0.419	0.001	0.532	0.003
S2_SR_1	0.615	0.002	0.448	0.001	0.566	0.003
S2_SR_2	0.608	0.002	0.442	0.001	0.560	0.003
S2_MSR_1	0.584	0.002	0.414	0.001	0.534	0.003
S2_MSR_2	0.575	0.002	0.405	0.001	0.525	0.003
S2_$SAVI_1$	0.613	0.002	0.451	0.001	0.567	0.003
S2_$SAVI_2$	0.611	0.002	0.449	0.001	0.564	0.003
S2_$MSAVI_1$	0.603	0.002	0.439	0.001	0.555	0.003
S2_$MSAVI_2$	0.596	0.002	0.435	0.001	0.549	0.003
S2_TVI_1	0.601	0.002	0.470	0.001	0.564	0.003
S2_TVI_2	0.555	0.002	0.432	0.001	0.520	0.003
S2_RGI	0.622	0.002	0.468	0.002	0.578	0.003
S2_Ultra BGI (BGI_1)	0.664	0.002	0.582	0.001	0.645	0.003
S2_BGI_2	0.624	0.002	0.619	0.001	0.631	0.003
S2_Ultra BRI (BRI_1)	0.720	0.002	0.583	0.001	0.683	0.003
S2_BRI_2	0.730	0.002	0.603	0.001	0.695	0.003
S2_Red Edge (nm)	0.678	0.002	0.530	0.001	0.637	0.003

*S2_SRPI is redundant to S2_Ultra BRI.

VITA

K. Colton Flynn

Candidate for the Degree of

Doctor of Philosophy

Dissertation: PREDICTING NUTRIENT CONTENT, PLANT HEALTH, AND SITE SUITABILITY: A CASE STUDY OF *ERAGROSTIS TEF*

Major Field: Geography

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in geography at Oklahoma State University, Stillwater, Oklahoma in May, 2019.

Completed the requirements for the Master of Arts in geography at the University of Arkansas, Fayetteville, AR in 2013.

Completed the requirements for the Bachelor of Science in earth science at the University of Arkansas, Fayetteville, AR in 2010.

Experience:

8/2015 – 5/2019	Graduate Teaching Assistant, Dept. of Geography, OSU
1/2017 – 12/2018	Biological Science Aid (GS-0404-02), USDA, Forage and Livestock Production Unit of El Reno, Oklahoma
8/2017 – 5/2018	Visiting Research Fellow, Fulbright Program, Ethiopian Biodiversity Institute of Addis Ababa, Ethiopia

Professional Memberships:

2018 – Present	American Geophysical Union (AGU)
2017 – Present	International Exchange Alumni: U.S. State Department