Photocopy and Use Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at Idaho State University, I agree that the Library shall make it freely available for inspection. I further state that permission for extensive copying of my thesis for scholarly purposes may be granted by the Dean of the Graduate School, Dean of my academic division, or by the University Librarian. It is understood that any copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Signature: _____

Date: _____

Precision Agriculture for Improving Potato Crop Management in Lebanon

by

Hanan Abou Ali

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Geographic Information Science

Idaho State University

Fall 2018

Committee Approval

To the Graduate Faculty:

The members of the committee appointed to examine the thesis of HANAN ABOU ALI find it satisfactory and recommend that it be accepted.

_____

Dr. Donna Delparte,
Associate Professor, Department of Geosciences,
College of Science and Engineering
Idaho State University
Major Advisor

_____

Dr. Carrie Bottenberg,
Assistant Lecturer, Department of Geosciences,
College of Science and Engineering
Idaho State University
Committee Member

_____

Dr. Yolonda Youngs,
Associate Professor, Department of Global Studies and Languages
College of Arts and Letters
Idaho State University
Graduate Faculty Representative

Dedication

إلى من علّماني المثابرة والسعي خلف أهدافي وكانا بجانبي في كل خطوة يشجعانني على تحقيق طموحاتي،

إلى من كان مصدر قوّتي عند الصّعاب وفي أوقات الشّدّة،

إلى من كان وما زال رمز العطاء والتضحية والصبر بلا حدود،

إلى من تعب وسهر وناضل في سبيل تعليمي وتربيتي وتأمين حياة كريمة لي ولإخوتي ،

إلى أبي الغالي وأمي الحبيبة،

تقف الحروف والكلمات عاجزة عن تعبير مدى حبي وامتناني لكما ولكل ما بذلتماه في سبيلي منذ ان أبصرت النور. لن أستطيع أن أُكافئكما يوما ولكن اسمحا لي بإهدائكما ثمرة تعبكما التي تتلخص في هذه الرسالة، راجية من الله عزّ وجلّ أن يطيل أعماركما ويديم عليكما الصحة والعافية والسعادة فأنتما أكبر نعم ربي.

أدعو الله أن تريا اخوتي يحققون احلامهم وطموحاتهم جمعاء.

دمتم لي السند والقوة والمثال الذي يحتذى به.

Dedication in English

To those who taught me perseverance and pursuing my goals, and were by my side encouraging me every step of the way, to achieve my ambitions;

To the source of my strength at times of difficulties and distress;

To the ones who were and still are icons in giving, sacrificing and patience without limits;

To those who worked hard, stayed up long hours, and fought for my education and upbringing, and to provide a good life for my siblings and I;

To my precious dad and my dear mom;

Letters and words fail to express the love and gratitude I have for you both and for everything you did for me since I opened my eyes in this world. There won't come a day where I can reward you but please allow me to present you with the outcome of your hard work that is summarized in this thesis. I pray to God to lengthen your life and bless you with health, wellness and happiness as you are my biggest blessing from God.

I ask God that you witness my siblings achieving all of their dreams and ambitions.

May you always be my support, strength and role models.

Acknowledgements

I would like to take this opportunity to thank everyone who helped and supported me during this work and made it possible to finish despite all challenges throughout. To Dr. Donna Delparte for being a great mentor and adviser and all her patience along this journey, to Dr. H. Carrie Bottenberg for teaching me everything I know about remote sensing, to Dr. Yolonda Youngs for her comments and input on this work and for representing the graduate school, and to Dr. Shannon Kobs Nawotniak for nurturing my love for coding and helping me automate processing to save my sanity! To L. Mike Griffel for his valuable input and advice throughout the project. To all the support and help from the loving Melissa Neiers, Kate Zajanc and Breane Fitzgerald in the Geosciences front office. Special thanks to Diana Boyack for her help in the Digital Mapping Lab.

To my amazing support system: Jerad Anderson, Leslie Montoya, Hanan Alkhateeb, Chris Borg, Angela Garcia, Ashley Shields, Brittany Guzzo, Kate Wehrs and everyone in the Geoscience department for becoming my family in Idaho and being there every step of the way. Thank you for all your support, love and endless shared coffee! To my American family, the Leitermans, for always being just a phone call away no matter when or where. To my parents and siblings for being supportive and loving despite being 10,936 km apart. To my Lebanese family and friends for encouraging my ambitions. To everyone I met along this journey, it's been an honor.

Most importantly, thanks to the US Department of State, the Fulbright program, US Embassy in Beirut, and AMIDEAST as they were the reason I was able to come to ISU and follow my dreams. Thank you for believing in me, my potential and my research interest to give back to Lebanon. Great thanks to Planet for granting me access to their imagery through the Education and Research program that made this work possible. Thanks to AmericaView for helping fund part of my travels to present at the Pecora20 conference.

Table of Contents

List of Figures x
List of Tables xii
List of Abbreviations xiii
Abstract xiv
Chapter 1 Introduction 1
1.1 Agricultural Production 1
1.2 Research Questions 3
1.3 Research Significance 8
Chapter 2 Literature Review 9
2.1 Potato Production in Lebanon 9
2.2 Satellite Imagery 13
2.3 Precision Agriculture 15
2.4 Vegetation Indices 16
2.5 Water Indices 19
2.6 Crop Monitoring 19
2.7 Yield Forecasting 20
Chapter 3 Satellite Imagery for Crop Monitoring and Yield Forecasting in Idaho 24
3.1 Introduction 24
3.2 Methods 29
3.2.1 Study Area 29
3.2.2 Image Processing 31
3.2.3 Yield Mapping 32
3.2.4 Removing Data Outliers 33
3.2.5 Crop Variability 33
3.2.6 Vegetation and Water indices 34
3.2.7 Yield Forecasting 35
3.3 Results 36
3.3.1 Variation of Indices over Growing Season 36
3.3.2 Planet Data Yield Regression Model 40
3.3.2.1 Russet Fields (Early Season) Regression Model 41
3.3.2.2 Norkotah Fields (Early Season) Regression Model 42

3.3.2.3 Russet Fields (Late Season) Regression Model 42
3.4 Discussion 43
Chapter 4 Utilizing Satellite Imagery for Precision Agriculture in Lebanon 46
4.1 Introduction 46
4.2 Methods 50
4.2.1 Study Area 50
4.2.2 Data Resources 53
4.2.3 Data Pre–Processing 54
4.2.4 Data Processing 54
4.2.5 Data Post–Processing 55
4.2.6 Information from Grower 56
4.3 Results 57
4.3.1 Variation of Indices over Growing Season Sentinel-2 vs. Planet 57
4.3.2 Yield Forecasting 62
4.4 Discussion 63
Chapter 5 Conclusion 67
References 71
Appendix A – Planet Python Code 77
Appendix B – Regression Models 87

List of Figures

Figure 1 Bekaa Valley, Lebanon and SouthEast Idaho locations on the world map. 4

Figure 2 Bekaa Valley boundaries and location within Lebanon. 5

Figure 3 Snake River Plain and study area location within Idaho. 6

Figure 4 Distribution of agricultural lands in Lebanon (DAR-IAURIF 2005)..... 9

Figure 5 Locations of major potato production in the Bekaa Valley and Akkar north Lebanon (Choueiri et al. 2017) 10

Figure 6 Potato Production in Lebanon between 1961 and 2014. (Knoema 2015)..... 11

Figure 7 The varying reflectance response of the leaf to the different wavelengths of the electromagnetic spectrum (Crum n.d.)..... 18

Figure 8 Interaction of leaf structure with Visible and Infrared Radiation (Arnold 2010)..... 18

Figure 9 Study area showing ten field locations near Parker, Idaho. (WGS 84 - UTM 12 N)..... 30

Figure 10 Satellite imagery and yield data processing workflow..... 32

Figure 11 Speed sensor allowing cop yield monitoring. Source:http://www.geektechforag.com/Products/BYM/Beet_Yield_Hardware/Beet_Yield_Hardware.aspx 32

Figure 12 Idaho: Russet Variety - G1 Fields 2, 3, 4 & 5 37

Figure 13 Idaho: Norkotah Variety - G2 Fields 1, 7, 8a, 8b, 9 & 10 38

Figure 14 Idaho: Russet Variety - G3 Fields 11a & 11b 39

Figure 15 Weekly average precipitation (mm) and temperature (ºC) over the growing season of 2017..... 40

Figure 16 Summary statistics of Russet fields regression model 41

Figure 17 Summary statistics of Norkotah fields regression model 42

Figure 18 Summary statistics of Russet fields 11 regression model 43

Figure 19 Lebanon Elevation Map. Elevation Dataset: Global Multi-resolution Terrain Elevation Data 2010 (GMTED 2010) – USGS EarthExplorer Website 47

Figure 20 Location of Tal Znoub within the Bekaa Valley, Lebanon. 50

Figure 21 Study Area showing the location of 16 fields within Lebanon in the WGS 84 - UTM 36N coordinate system at a scale of 1:16,000 52

Figure 22 Variation of indices over the 2017 growing season 58

Figure 23 Variation of NDVI over the 2017 growing season Sentinel (A1: April 11, B1: May 11, C1: June10) vs. Planet (A2: April 17, B2; May 9, C2: June9) 59

Figure 24 Variation of SAVI over the 2017 growing season Sentinel (A1: April 11, B1: May 11, C1: June10) vs. Planet (A2: April 17, B2; May 9, C2: June9) 60

Figure 25 Variation of NDWI over the 2017 growing season Sentinel (A1: April 11, B1: May 11, C1: June10) vs. Planet (A2: April 17, B2; May 9, C2: June9) 61

Figure 26 Predicted yield values (A) and the corresponding slope values (B) 62

Figure 27 Average daily temperature 2017 growing season (°C) 64

List of Tables

Table 1 Sentinel-2 bands (ESA - https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric) 14

Table 2 PlanetScope 4 Band Imagery 15

Table 3 Vegetation Indices 17

Table 4 Water Indices 19

Table 5 Field Areas 30

Table 6 Crop Variety and Growing Information 31

Table 7 Spectral Indices 35

Table 8 Russet fields regression model variables 41

Table 9 Norkotah fields regression model variables 42

Table 10 Russet field 11 regression model variables 43

Table 11 Field Areas in m^2 51

Table 12 Satellite Imagery Sources 53

Table 13 Vegetation and Water Indices 55

Table 14 2017 Growing season information 56

List of Abbreviations

AVHRR	Advanced Very High Resolution Radiometer
BOA	Bottom of Atmosphere
DEM	Digital Elevation Model
ESA	European Space Agency
FAO	Food and Agriculture Organization
GDP	Gross Domestic Product
GNDVI	Green Normalized Difference Vegetation Index
GWR	Geographically Weighted Regression
LAI	Leaf Area Index
MLR	Multi Linear Regression
MODIS	Moderate Resolution Imaging Spectroradiometer
MSAVI2	Modified Soil Adjusted Vegetation Index 2
NAAS	National Agricultural Statistics Service
NAIP	National Agriculture Imagery Program
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near Infrared
OLS	Ordinary Least Square
PCN	Potato Cyst Nematodes
PVY	Potato Virus Y
RMSE	Root Mean Square Error
SANDVI	Saturation Adjusted Normalized Difference Vegetation Index
SAVI	Soil Adjusted Vegetation Index
SPOT	Satellite Pour l'Observation de la Terre
TOA	Top Of Atmosphere
UAS	Unmanned Aircraft Systems
USGS	United States Geological Survey
WDRVI	Wide Dynamic Range Vegetation Index

Abstract

High spatial and temporal resolution satellite imagery provides a reliable resource for crop monitoring throughout the growing season. Spectral indices including NDVI, SAVI and NDWI deliver crucial information about crop health to aid growers with decision making for precision agriculture practices. These vegetation and water indices show crops' response to changing weather conditions and indicate critical times where extra irrigation or nutrients are needed. NDVI and slope are critical indicators for yield prediction. Using the Idaho potato yield data and taking into consideration different potato varieties, we are able to predict yield values for crops in Lebanon.

Precision Agriculture for Improving Potato Crop Management in Lebanon

Thesis Abstract – Idaho State University (2018)

Key Words: crop monitoring, indices, Planet, remote sensing, Sentinel-2, yield prediction, satellite imagery, yield forecasting, precision agriculture

Chapter 1 Introduction

1.1 Agricultural Production

Agriculture is an important sector in the global economy and is a crucial component for fighting hunger and food insecurity. According to the United Nations Food and Agriculture Organization (2017), the expectation is that the world population will reach 10 billion by 2050 and there is a need to produce around 50% more food than in 2012. Lebanon relies heavily on its agricultural sector as it contributes 637% to Gross Domestic Product (GDP) and provides work to 15% of the Lebanese population (FITA 2008). The Bekaa Valley, located in the center of Lebanon, is one of the largest agricultural regions in the country. Potato crops account for 56% of vegetable production in the country, mainly in the Bekaa Valley and North Lebanon states (Hatoum 2005a). Hence, it is important to ensure the health of potato crops and to improve production in order to revive and improve the local and regional market to where it was prior to the Lebanese civil war (1975 – 1990).

Lebanese growers are having a hard time selling their produce which is having an impact on the Lebanese economy. With the pressures of potato viruses, pests, diseases and other threats attacking crops along with the increased cost of pesticides, herbicides, fertilizers and workers needed to inspect plants from the fields, Lebanese farmers have been suffering economically and so has potato production (An-Nahar Newspaper 2015). Since potatoes are an important irrigated crop that can be vulnerable to water availability, pests, disease and other crop threats, precision agriculture has the potential to help minimize such issues. Precision agriculture uses information technology to better manage crop production by taking into consideration the variations within the field to increase profitability and sustainability (Q. Zhang 2016). Precision agriculture can improve crop yield by empowering farmers with timely scientific knowledge on crop condition. By utilizing

remote sensing data from satellite and unmanned aircraft systems (UAS), farmers can leverage cost-effective technologies to mitigate crop threats with targeted approaches for grower decision-making such as variable rate fertilizer application, timely irrigation, early disease detection, and pest control. Introducing the concept of precision agriculture plays a major role in empowering local farmers and stakeholders by educating them about new technologies that have the potential to improve their crop productivity and enhance their economic sustainability.

Though there are many factors affecting the agricultural and economy sectors in Lebanon, one of the major ongoing issues is the ongoing Syrian crisis and its corresponding refugee influx's impact on Lebanon. The war that started in 2011 has resulted in over 1.5 million Syrians taking Lebanon as their shelter away from the war. Despite the refugees living all over Lebanon, the main concentration areas were the Bekaa Valley and Northern Lebanon as they are the common boundaries between Lebanon and Syria. The majority of the Syrian refugees set up tents near bodies of water within the valley and mainly along rivers to have access to water. However, most of those refugee camps lack access to sanitary living conditions and the residents use the water for all their daily needs including cooking, showering and toilet water. As a result, there has been an increase in pollution levels within the rivers which irrigate neighboring agricultural lands. The random and unorganized distribution of refugee tents lead to a decrease of natural resources including water bodies and agricultural lands (FAO 2014).

In Lebanon, the traditional method of crop inspection is for growers to individually inspect plants in the field (An-Nahar Newspaper 2015). This is time-consuming and not very efficient. In addition, the cost of applying fertilizers to crops is higher since farmers apply treatments uniformly instead of using a variable rate treatment. This places a greater financial burden on the farmers due to the increased amounts of fertilizers needed to cover the entire field. Precision agriculture offers

a solution for such problems; however, in a country such as Lebanon, precision agriculture is still in its early stages of adoption. Lebanon recently teamed up with the Food and Agriculture Organization of the United Nations (2017) to launch the Country Programming Framework from 2016 to 2019 in order to develop more sustainable practices to improve the agricultural sector. Currently, Syria is a politically unstable region bordering Lebanon, and it is challenging to utilize UAS for crop inspection in certain areas since they could mistakenly be perceived as a threat. This poses a major challenge until the region achieves increased stability. In the meantime, satellite imagery has the potential to act as a substitute for UAS precision agriculture data collection and cover larger areas without the risk of having people interpret it as a threat.

1.2 Research Questions

This comparative thesis study assesses potato crop stressors that affect yield in Lebanon's Bekaa Valley by drawing from a model generated over the 2017 growing season using satellite imagery and potato yield data collected in Idaho, USA. Since Idaho and the Bekaa Valley share similar topographic, geographic and environmental features (Figure 1, Figure 2 & Figure 3) including related crop threats, we hypothesized that a model trained with yield data and satellite imagery from Idaho can characterize a yield prediction model for the Bekaa Valley.

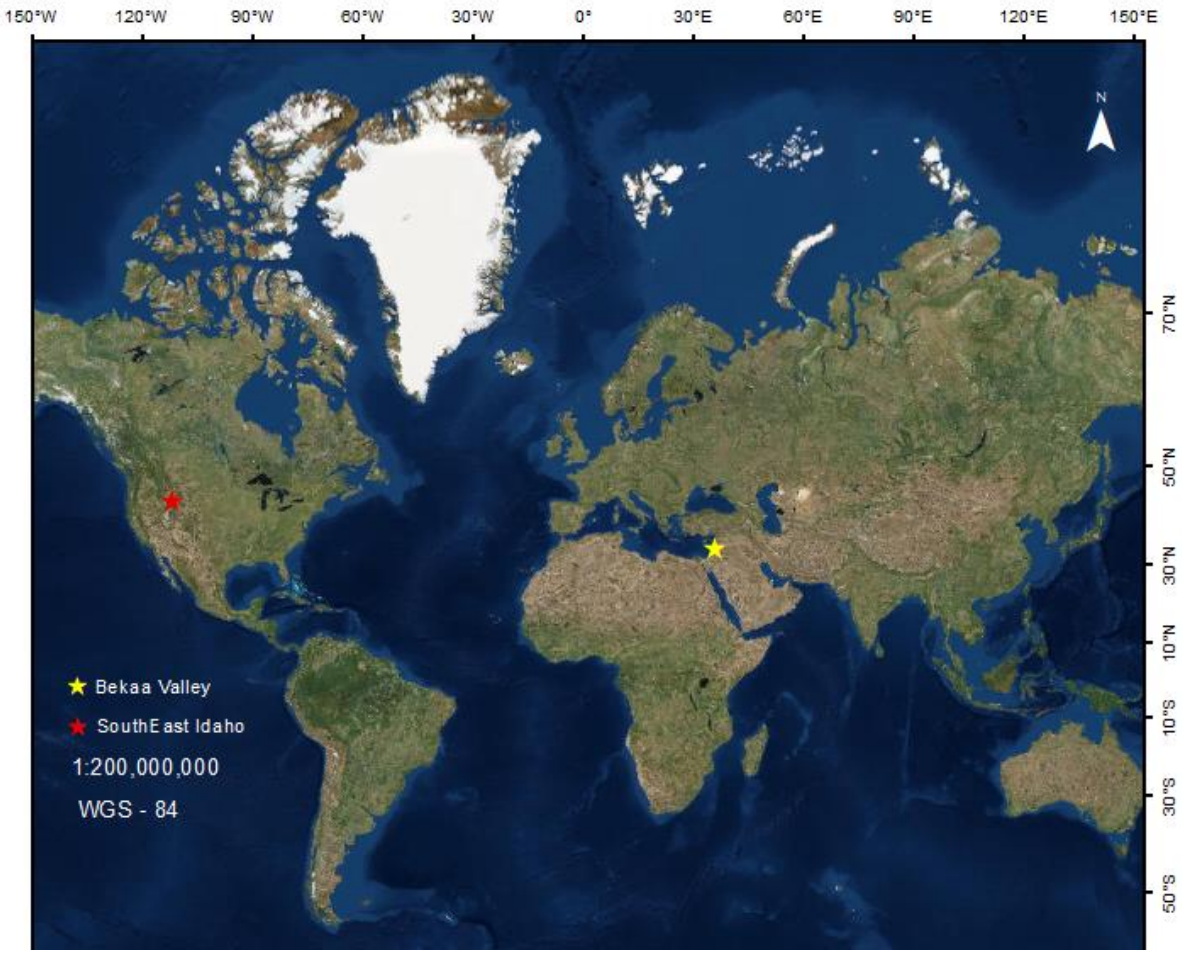

Figure 1 Bekaa Valley, Lebanon and SouthEast Idaho locations on the world map. Basemap source: Esri, DigitalGlobe, GeoEye, Earthsatr Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

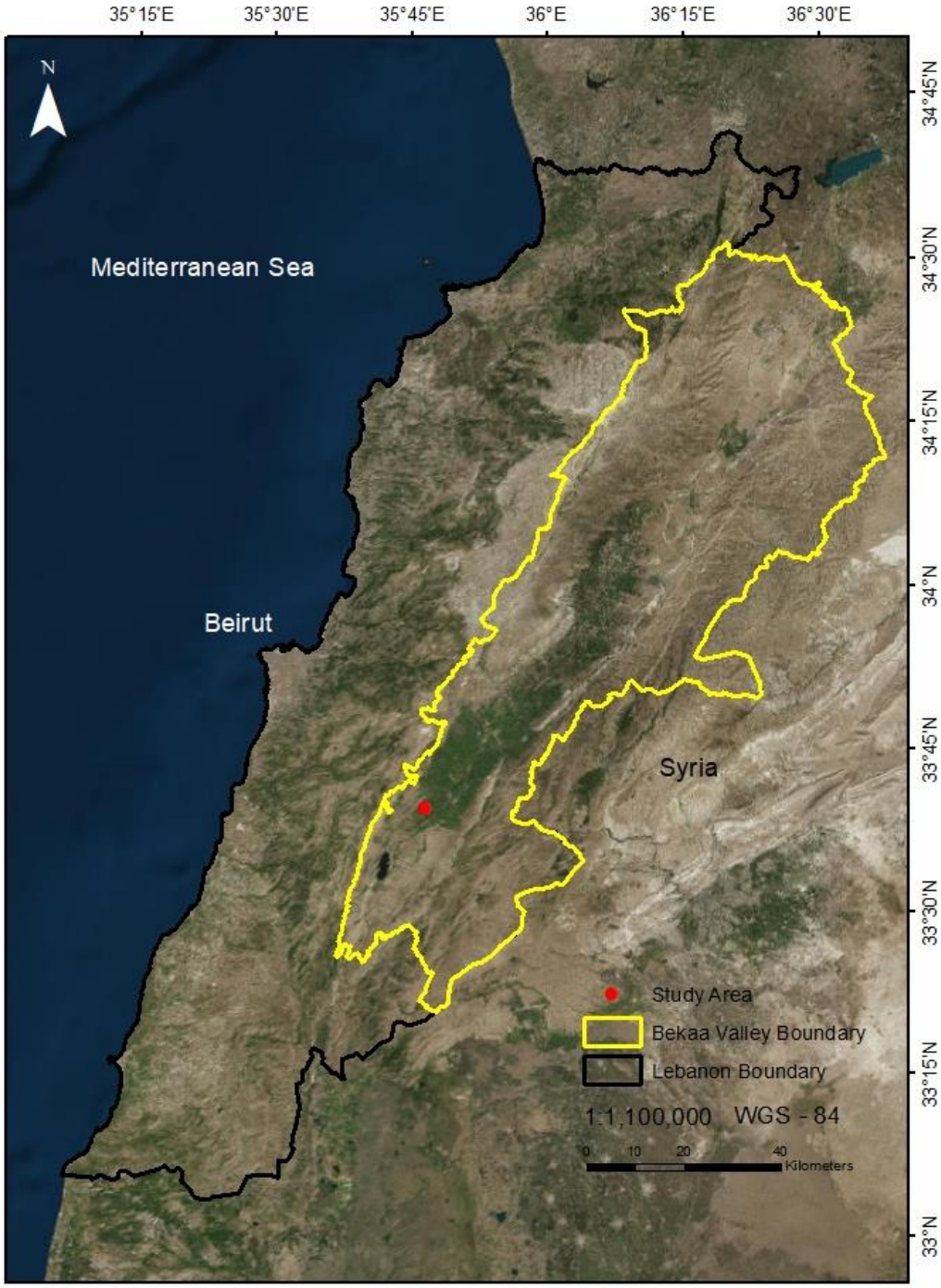

Figure 2 Bekaa Valley boundaries and location within Lebanon.
Boundaries layers source:
https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/LBN_Boundaries_2016/FeatureServer

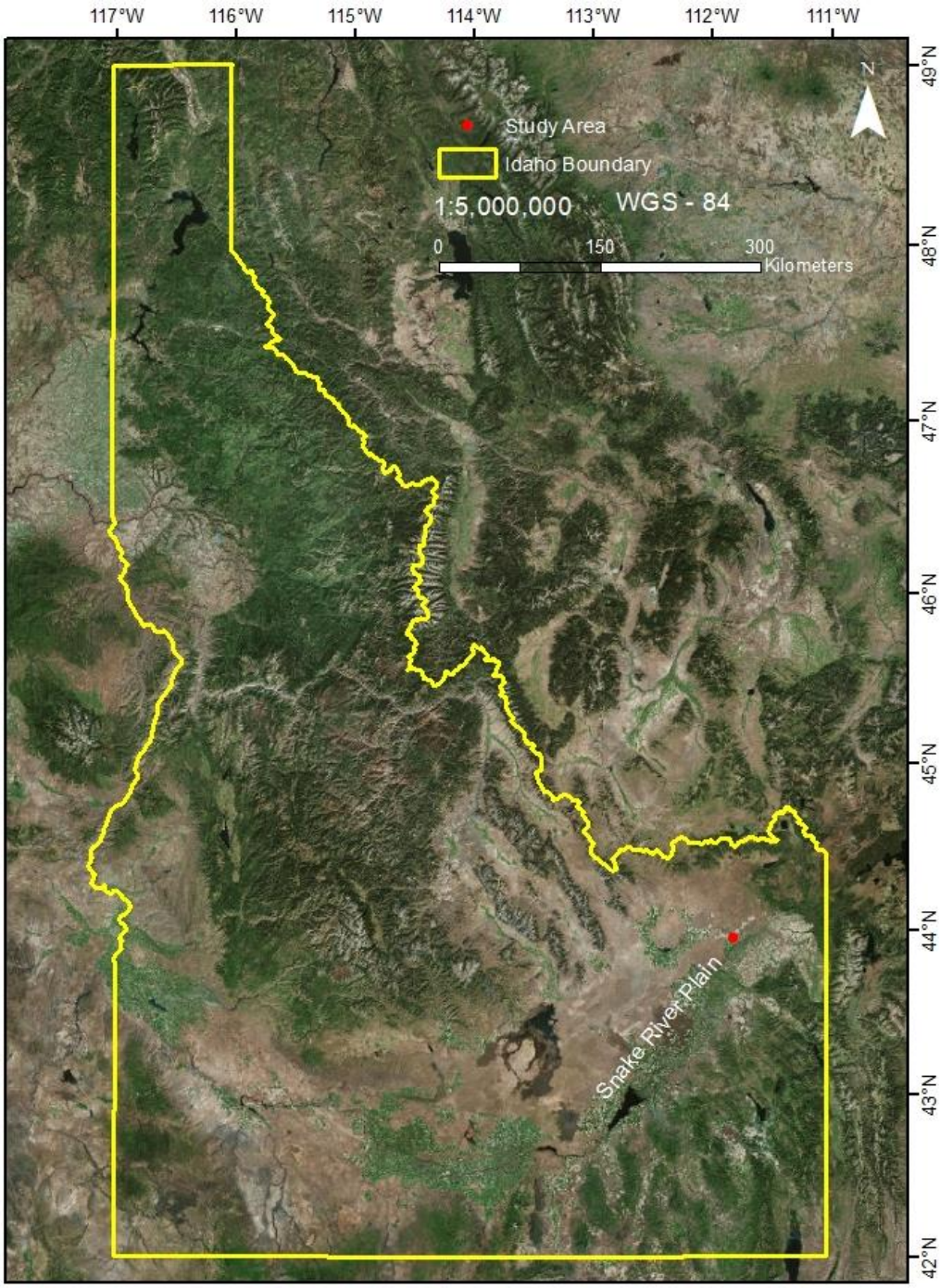

Figure 3 Snake River Plain and study area location within Idaho.
Boundary layer source:
https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/USA_Boundaries_2016/FeatureServer

Both study areas have cold wet winters and hot dry summers and are within river plains and part of the North Temperate Zone. The Idaho study area is located at a latitude of 43.9º and altitude of 1502 m above mean sea level with a wet season from October through June with a high average temperature during the 2017 season of 24ºC. The Lebanon study area is at 10º lower than the Idaho study area, located at 33.9º latitude and elevation of 872 m above sea level and a high average temperature of 34ºC during the 2017 growing season.

This research also evaluated general irrigation efficiency and crop health over the growing season. Multispectral satellite imagery from PlanetScope sensor over Idaho and Lebanon, including derived vegetation and water indices, provided the predictor variables for the regression model to forecast yield.

Specifically, this thesis research addresses the following questions:

- What are the best indices for yield prediction?
- How do vegetation indices vary over the growing season?
- What is the relationship between vegetation indices and water indices?
- What is the relationship between vegetation and water indices and potato crop yield?
- What is the potential of multispectral satellite imagery for agricultural applications?
- Can an Idaho yield model predict yield in Lebanon?

1.3 Research Significance

With the use of satellite imagery, there is the potential for a low-cost decision support system to increase crop yield. As a result, this can help strengthen the Lebanese economy by reviving the domestic market and potentially increasing potato exports. This study builds upon existing work and aims to expand the use of satellite imagery for studying crop threats through analysis of spectral response throughout the growing season. In addition, the use of newly available Sentinel-2 along with PlanetScope satellite imagery for irrigation management and general crop health will act as a consistent and reliable resource to evaluate crops over the growing season.

This research will empower farmers and stakeholders with tools that are cost and time effective in mitigating crop threats. Moreover, in a world suffering from climate change impacts of rising temperatures and irrigation shortfalls, it is very important to improve water usage and promote water management as part of a broader strategy of encouraging farmers to adopt more sustainable farming practices. Upon successful completion of this work, I will present the results of this work to the Lebanese Ministry of Agriculture to demonstrate the efficiency of precision agriculture and its potential application on Lebanese crops upon successful completion.

Chapter 2 Literature Review

2.1 Potato Production in Lebanon

With an area of 10,452 km^2, Lebanon has approximately 2,730 km^2 of its land used for agriculture (Figure 4 & Figure 5). The agricultural sector is the third most important sector in the country contributing between 6 to 7% of the GDP (El Gazzar 2015).

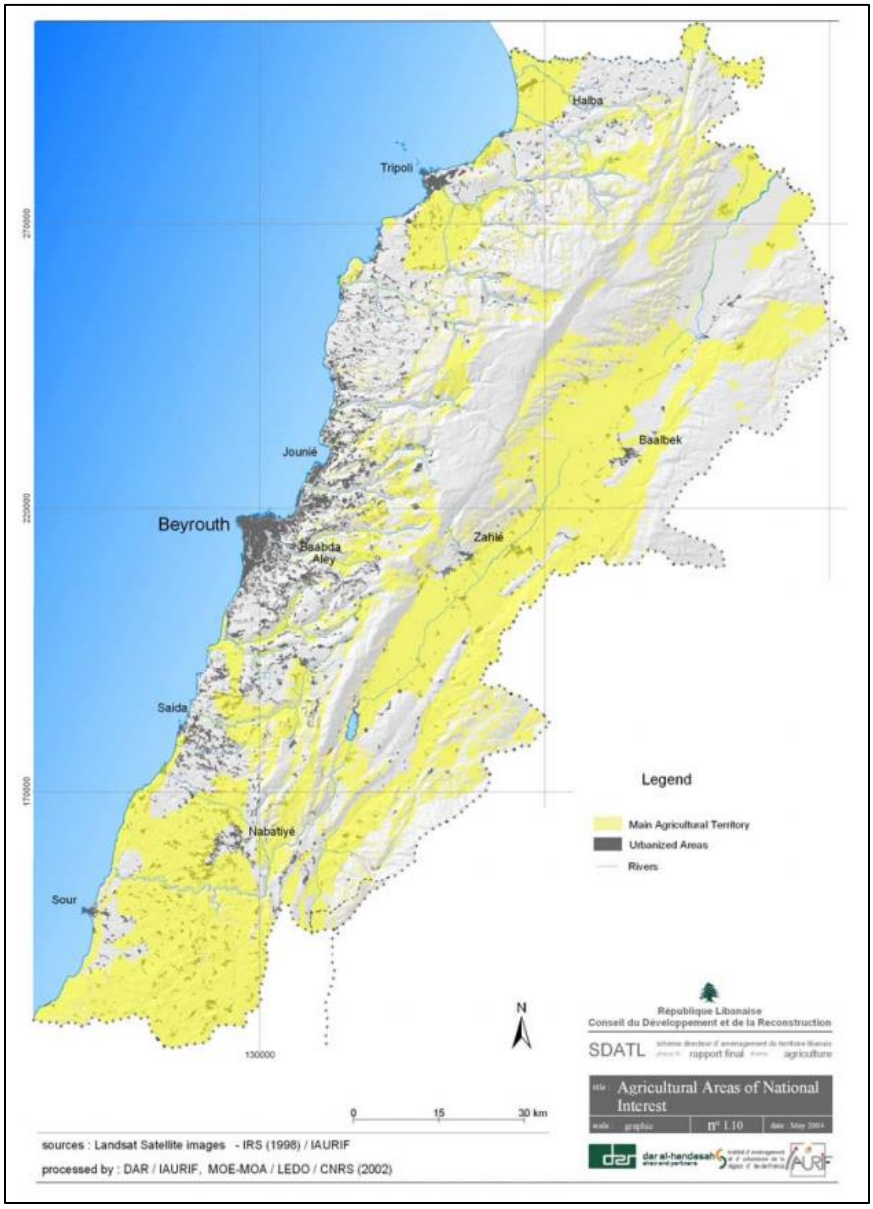

Figure 4 Distribution of agricultural lands in Lebanon
(DAR-IAURIF 2005)

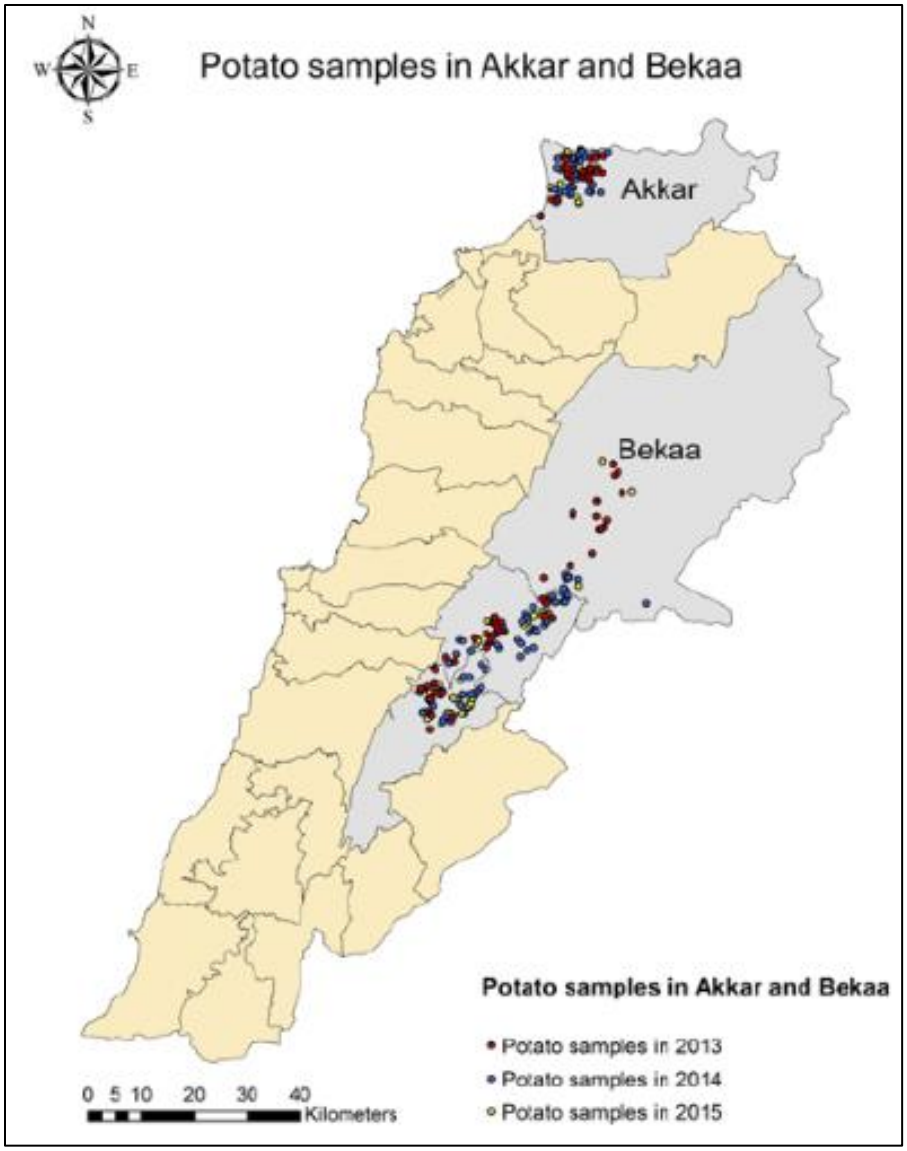

Figure 5 Locations of major potato production in the Bekaa Valley and Akkar north Lebanon (Choueiri et al. 2017)

Lebanon has traditionally been a major potato exporter to neighboring countries and the region where 60% of its production was exported to Arab countries, the United Kingdom and Brazil (Ktheien 2008). During the years of the country's civil war from 1975 to 1990, potato production was severely impacted similar to other sectors where production and exports dropped drastically by about 50% (Gale 2007). As shown in (Figure 6), for instance, potato production

during the beginning of the civil war in 1975 reached a value as low as 20 kt (kilotons) from 120 kt in the early 1970's but started to increase with relatively steady rates after that. Though previously all potato seeds were imported, Lebanon has recently set up a seed production system with certification (Quere 2009). Although the highest potato production for Lebanon on record is in 2007 with 514 kt, it has been decreasing to as low as 275kt in 2011 and the last recorded value is from 2014 with 451.8 kt (Actualitix 2016). When it comes to potato varieties in the country, Alpha and Arran Banner are the main two while others such as Spunta, Jaerla and Cloustar are making their way into the Lebanese market (Ktheien 2008).

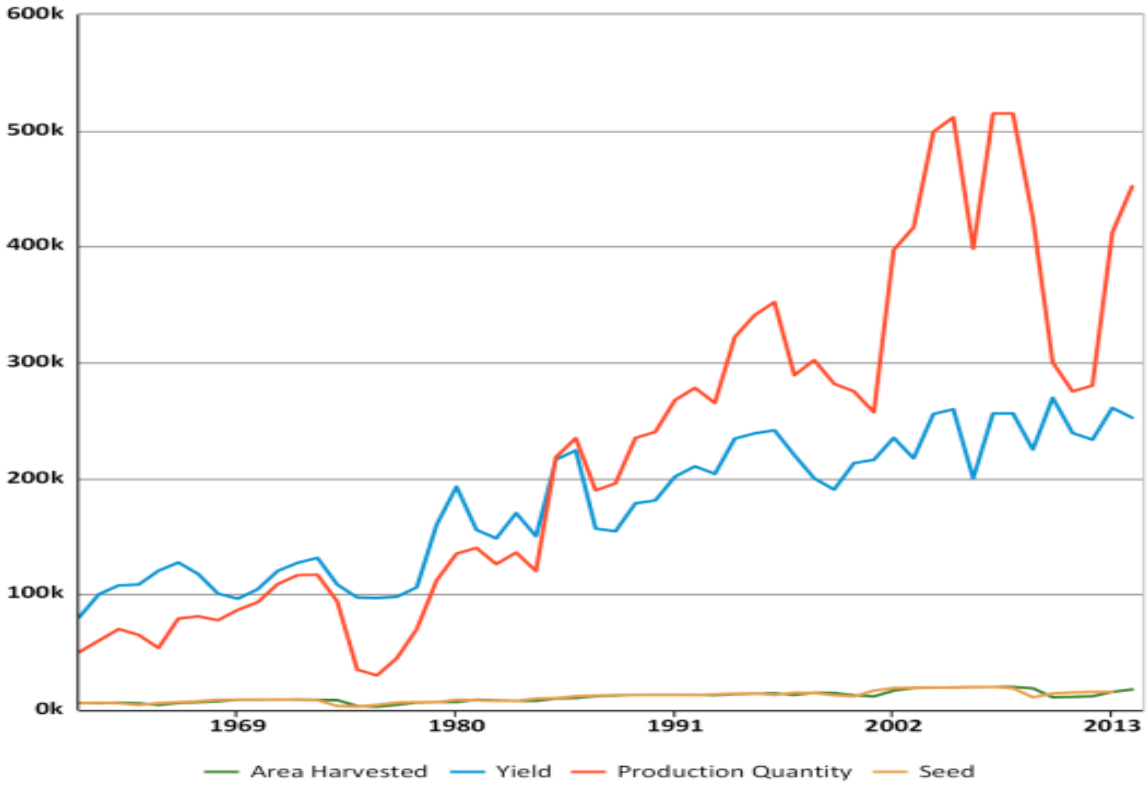

Figure 6 Potato Production in Lebanon between 1961 and 2014. (Knoema 2015)

Although there is very limited data on potato viruses and threats in Lebanon, there are several studies to provide a statistical overview for a baseline to build future work (Knoema 2015). Since water is a crucial element in potato crop growing, efficient irrigation methods for potato crops are important. My research shows that the main irrigation systems in Lebanon are, in addition to rainfall, basin and furrow irrigation, followed by sprinklers and micro irrigation techniques (Karaa and Karam 2000). Among different irrigation methods, Darwish et al. (2004) came to a conclusion that drip systems were the most effective for crop growth results and should be implemented more widely. For potato viruses in Lebanon, Abou-Jawdeh et al. (2001) sampled growth seasons from major agricultural regions in the country and various potato viruses were detected; potato virus Y *(Potyvirus)* was the most dominant one among all samples. PVY represents a significant threat to potato crops worldwide as it is considered the most harmful to potato fields (Steinger, Gilliand, and Hebeisen 2014), In Turkey, for example, they are recording the highest percentages of infections among other potato viruses (Yardımcı, Kılıç, and Demir 2015). Another potato threat in Lebanon is the Potato Cyst Nematodes (PCN) *(Globodera)* which is the major potato pest in the country (Ibrahim, Abi Saad, and Moussa 2004). PCN are the most severe potato pest where losses could reach up to 80% within a growing area (Hassan et al. 2013). The biggest challenge in detecting PCN is that it needs skilled and experienced workers to visually identify infected plants and soil in the field or by using, real-time Polymerase Chain Reaction in the labs, which saves time but is very costly (Hassan et al. 2013). As important as it is to study potato viruses and threats in Lebanon, it is equally crucial to have an overview of potato threats worldwide to gain a better understanding of the challenges faced worldwide. Studying a major potato producing region such as Idaho that employs precision agriculture techniques to manage crop threats can inform practices in Lebanon.

2.2 Satellite Imagery

Over time, satellite imagery technology improved as reflected by an increased sensor resolution and expanding the usages of satellite images to cover various applications such as precision agriculture starting in the early 1970's (Mulla 2013). Not only has satellite imagery improved in quality, but these images have also become more accessible to the public (Turner et al. 2015) through constantly updated datasets. Though satellite images are generally of large data volume (Skakun et al. 2016), various algorithms and software packages have made it easier and less time consuming to process and analyze large datasets. Image classification is done on a pixel by pixel level and thus the higher the spatial resolution ,the more detailed the result (Yang et al. 2013).

The Sentinel-2 is a mission by the European Space Agency (ESA). Sentinel-2A, launched in June 2015, and more recently, Sentinel-2B launched in March 2017, are two multispectral imagers covering 13 spectral bands (443 nm – 2190 nm) at resolutions of 10-20 and 60 m (Table 1). The Sentinel-2 program is filling a void for open source freely available, medium resolution imaging with five day revisit times depending upon latitude, cloud cover and other factors, to assess plant health and vigor during growing seasons (Dash and Ogutu 2016).

Table 1 Sentinel-2 bands (ESA - https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric)

Band	**Central Wavelength (µm)**	**Resolution (m)**
Band 1 – Coastal Aerosol	0.443	60
Band 2 – Blue	0.490	10
Band 3 – Green	0.560	10
Band 4 – Red	0.665	10
Band 5 – Vegetation Red Edge	0.704	20
Band 6 – Vegetation Red Edge	0.740	20
Band 7 – Vegetation Red Edge	0.783	20
Band 8 – NIR	0.842	10
Band 8A – Vegetation Red Edge	0.865	20
Band 9 – Water Vapor	0.945	60
Band 10 – SWIR , Cirrus	1.375	60
Band 11 – SWIR	1.610	20
Band 12 - SWIR	2.190	20

In addition, there is the PlanetScope satellite constellation operated by Planet that is freely available for university researchers. It has four spectral bands (445 nm – 860 nm) at a resolution of 3 m (Table 2) with a daily revisit. Both Sentinel-2 mission and PlanetScope data are useful for precision agriculture applications. While some previous studies on modeling yield prediction used Landsat (Song et al. 2016) and Sentinel-2 (Al-Gaadi et al. 2016), this work utilizes Sentinel-2A and PlanetScope analytical scene for the analysis due to their higher temporal and spatial resolution.

Table 2 PlanetScope 4 Band Imagery

Band	**Wavelength (nm)**
Blue	455 – 515
Green	500 – 590
Red	590 – 670
NIR	780 – 860

2.3 Precision Agriculture

Satellite imagery has a vital role in identifying crop stress to aid crop management strategies. However, Guo et al. (2012) argued that satellite imagery technology is tightly limited and unavailable during peak times of need and is yet unable to balance the high-resolution low-cost equation. Hunt et al. (2014) agreed by explaining that satellite images only provide large pixel size and are faced with infrequent flying times and limited in presence of clouds. However, as new data emerges and becomes available to the public, satellite imagery is revolutionizing the field of precision agriculture application aiding decision making starting with Sentinel -2 imagery since 2015 and Planet imagery with a daily revisit more recently. The main aim for precision agriculture is to take into consideration the variability in the field rather than assuming that the field is uniform. Hence, all field variables influencing spatial variability in growth and yield such as soil and topography are taken into account with precision agriculture approaches (Sivarajan 2011). Using satellite imagery for precision agriculture was introduced in the early 1970's (Mulla 2013) and has been expanding ever since especially with the increasing availability of satellite imagery and options for analysis. The increase in both spatial and temporal resolution of satellite imagery provides a valuable resource for crop monitoring and specifically yield modeling and prediction. Not only has satellite imagery improved in quality, but they have also become more accessible to the public (Turner et al. 2015) through constantly updated datasets.

2.4 Vegetation Indices

Indices best suited for analyzing crop health status are outlined in Table 3 and include: Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Soil Adjusted Vegetation Index (SAVI) and Modified Soil Adjusted Vegetation Index 2 (MSAVI2) (Candiago et al. 2015). NDVI is a normalized ratio of near-infrared and red bands that range between -1 and 1 where higher values indicate more vegetated cover while lower values correspond to non-vegetated areas and features. Due to the ratio calculation of NDVI, it reduces noise and provides an effective approach for comparing changes in vegetation over periods of time (A. Huete et al. 2002). Similarly, GNDVI is used to further focus on the greenness of plants as it is more sensitive to chlorophyll content than NDVI because it is related to the Leaf Area Index (LAI) and biomass (Candiago et al. 2015) and has a range between 0 and 1. The SAVI index, on the other hand, aims at minimizing the effect of soil in vegetation unlike NDVI (A. R. Huete 1988). SAVI has a range between -1 and 1 where values between -1 and 0.1 are likely not vegetated. This index has the variable L in its equation which is related to the density of vegetation, so as the density increases, the value of L decreases (Candiago et al. 2015). MSAVI2, however, is a more accurate estimation of wide range vegetation cover (Liu et al. 2007) and is a tool to help compensate for images where there is a large amount of bare soil exposed in the field. Previous yield studies demonstrate that the most efficient indices for yield prediction are NDVI (Prasad et al. 2006; Song et al. 2016; Yang et al. 2013), SAVI (Al-Gaadi et al. 2016; Satir and Berberoglu 2016; Sivarajan 2011) and Normalized Difference Water Index (NDWI) (Satir and Berberoglu 2016; Wojtowicz, Wojtowics, and Piekarczyk 2016; You et al. 2017). Al-Gaadi et al. (2016), highlighted that yield was highly correlated to SAVI as it minimized the soil reflection.

Table 3 Vegetation Indices

Vegetation Index	**Formula**	**Reference**
Normalized Difference Vegetation Index	$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$	(Rouse et al. 1973)
Green Normalized Difference Vegetation Index	$\frac{\rho_{NIR} - \rho_{Green}}{\rho_{NIR} + \rho_{Green}}$	(Gitelson, Kaufman, and Merzlyak 1996)
Soil Adjusted Vegetation Index	$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red} + L} * (1+L)$	(A. R. Huete 1988)
Modified Soil Adjusted Vegetation Index 2	$\frac{2\rho_{NIR} + 1 - \sqrt{(2\rho_{NIR} + 1)^2 - 8(\rho_{NIR} - \rho_{Red})}}{2}$	(Qi et al. 1994)

Chlorophyll, the green pigment in vegetation, is primarily responsible for the green color of the plant and controls leaf spectral response within the visible portion of the electromagnetic spectrum (Campbell and Wynne 2011) (Figure 7). However, chlorophyll absorbs light with unequal proportions as it mainly absorbs blue and red light while reflecting green light (Figure 8) and thus the green color visible to the human eye reflects a healthy plant. When it comes to the near infrared section, the structure of mesophyll tissue is in control. A very small portion of the infrared light gets absorbed while most of it is scattered (Figure 8).

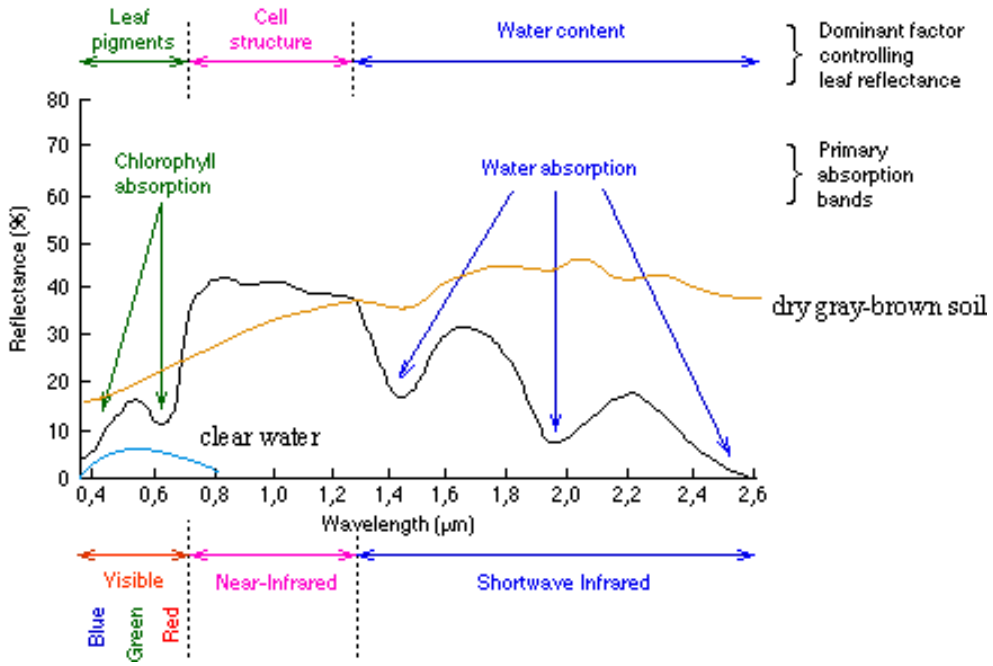

Figure 7 The varying reflectance response of the leaf to the different wavelengths of the electromagnetic spectrum (Crum n.d.)

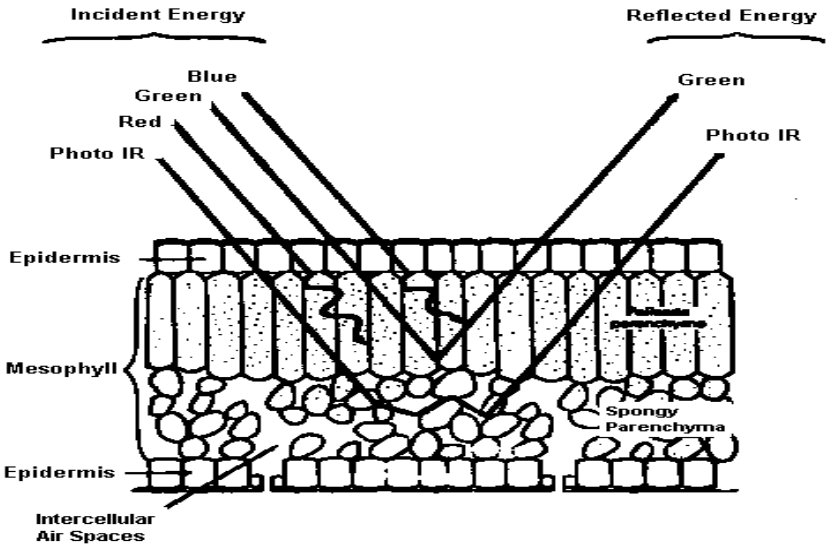

Figure 8 Interaction of leaf structure with Visible and Infrared Radiation (Arnold 2010).
Chlorophyll absorbs blue and red light while green is partially reflected and Near-Infrared (NIR) is scattered and reflected by the cell walls in the mesophyll.

2.5 Water Indices

As irrigation plays a major role in the health of potato crops, it is important to analyze water stress and content in crops to manage water resources more efficiently and to determine what irrigation techniques are the most efficient. Using data collected over 3 years, Rud et al. (2014) showed that UAS could be used for water management by testing several irrigation methods over the fields of interest. In addition, according to research performed in Egypt, (Nahry, Ali, and Baroudy 2011), satellite imagery proved to be very effective in guiding land management practices to apply fertilizers and irrigation only where needed with a result of increased economic and environmental profitability. Further, using a Normalized Difference Water Index (NDWI) (Table 4) calculated from Sentinel multispectral satellite imagery, water stress in plants was successfully determined (Gao 1996).

Table 4 Water Indices

Water Index	Formula	Reference
Normalized Difference Water Index	$\frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$	(Gao 1996)
Normalized Difference Water Index	$\frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}}$	(S.K. 1996)

2.6 Crop Monitoring

One of the basic methods for crop yield prediction is to form an empirical relationship between values of vegetation indices and ground measures (Lobell 2013). By using samples collected from the field, the correlation between vegetation indices is determined through plot scatter graphs (Al-Gaadi et al. 2016) and is used to determine the accuracy of the predictions by

validating them on the ground. According to Satir and Berberoglu (2016), in order to identify the crops within the field, the analysis process starts with a supervised classification. Their work showed that using an object-based classification with the Supervised Maximum Likelihood Classifier is more efficient than running a pixel by pixel based classification to identify crops as it increases accuracy and decreases classification time by assuming that each classified class has a normal distribution and thus calculates the probability of each pixel and adds it to the corresponding class. Vegetation indices are then used for the prediction, each based on the significance of the index (Satir and Berberoglu 2016). Models are then subsequently generated to predict crop yield by taking into consideration these variables. Such models include linear and multilinear regression models where they show the relationship between variables (Satir and Berberoglu 2016). While both models as well as regression tree models proved to be effective (Song et al. 2016), another study showed that while some bands are being neglected in those models, they could be of great use through a Gaussian model that also accounts for the relations on the spatial and temporal levels using random variables (You et al. 2017).

2.7 Yield Forecasting

Crop yield forecasting is very crucial for farmers as it helps with management decisions regarding harvesting, storage, pricing and marketing. The availability of satellite imagery with higher spatial and temporal resolution provides detailed information about crop health over the duration of the growing season. Such information along with other variables such as precipitation and temperature data make valuable tools for crop monitoring. Yield data is the dependent variable and the remaining variables are explanatory for building a regression model. While there are numerous methods for regression models, regression analysis uses different variables (known as explanatory variables) to fit a relation with the dependent variable to predict it numerically (Aditya

Shastry 2017; Sellam and Poovammal 2016). In this study the dependent variable is yield collected at 0.8m x 10m resolution by a device attached to the potato harvester. Although different studies use a variety of regression model approaches for yield prediction, the majority of them use NDVI as the only index considered as the major explanatory variable due to its correlation with plant health (Prasad et al. 2006; Rembold et al. 2013; Song et al. 2016; Tadesse et al. 2015; L. Zhang, Lei, and Yan 2010). Few other works use more than one index or incorporate other indices such as SAVI and/or NDWI to compare which better explains and predicts yield (Aboelghar, Ali, and Arafat 2014; Al-Gaadi et al. 2016; Alganci et al. 2014; Dempewolf et al. 2014; Satir and Berberoglu 2016).

Despite the availability of a wide range of regression models, one of the most commonly used in yield prediction is linear regression. Linear regression uses one variable at a time to explain and predict the yield using separate equations for each (Aditya Shastry 2017) assuming that there is a linear relation between the yield and the explanatory variable, model residuals are almost normally distributed such that there is no clustering in the data (Sellam and Poovammal 2016). In their paper, Rembold et al. (Rembold et al. 2013), used linear regression with NDVI from SPOT (Satellite Pour l'Observation de la Terre) imagery to explain yield values and obtained an R^2 of 0.930 and 0.799 for the Morocco and Egypt study areas, respectively. Al-Gaadi et al. (2016) also used linear regression but used Landsat-8 and Sentinel-2 imagery for computing both NDVI and SAVI. They generated a linear regression equation per index per sensor per field and concluded that higher resolution imagery yielded better regression models where the Landsat-8 imagery resulted in an R^2 range between 0.39 and 0.65 compared to 0.47 and 0.65 for the Sentinel-2 dataset.

Similarly, different studies used linear regression on more than one index for yield prediction and obtained higher R^2 values when considering other climate and soil variables

utilizing different sensors such as SPOT, Landsat 7, IKONOS as well as aerial imagery from UAS (Aboelghar, Ali, and Arafat 2014; Alganci et al. 2014; Geipel, Link, and Claupein 2014). A study compared the linear regression, stepwise multiple-linear regression and regression tree methods using the NDVI from Landsat-8 to predict winter wheat yield and concluded that the regression tree gave the best fit model with an R^2 of 0.87 (Song et al. 2016). Song et al. (2016) explained that linear regression enabled them to understand if there was a relationship between NDVI and yield, while stepwise-multiple linear regression simplified the equation and avoided multicollinearity among variables and the regression tree showed the importance of every growth stage as to how it related to yield. The regression tree method (Dempewolf et al. 2014) used NDVI, SANDVI (Saturation-Adjusted Normalized Difference Vegetation Index) along with additional indices based on Landsat 5 and 7 and MODIS (Moderate Resolution Imaging Spectroradiometer) data to predict wheat yield in Pakistan. The results suggested that Landsat imagery provides better regression models due to the higher resolution, specifically with the Wide Dynamic Range Vegetation Index (WDRVI). A study comparing multi-linear regression (MLR) to step-wise linear regression showed with MLR the error is largely due to uncertainty in the variables. The reason is that MLR attempts to fit a linear model to more than one explanatory variable. However, the step-wise linear model reduced the error by selecting the most relevant variables through binary relations between the dependent variable and the yield (Satir and Berberoglu 2016) using various indices including NDVI and NDWI. Contrary to the work of Satir and Berberoglu (2016), multiple studies showed the potential of multilinear regression with NDVI, in particular for tea crop yield prediction (Sitienei, Juma, and Opere 2017) and soybean and corn yield prediction in Iowa (Prasad et al. 2006) using AVHRR (Advanced Very High Resolution Radiometer) and Landsat data. In both works, one equation explained the yield by assigning a coefficient for each explanatory

variable to predict crop yield. Another recent approach in crop yield prediction is deep learning. This approach uses historical data as a training tool for the machine learning algorithm to forecast yield and is showing potential using various training models (Kuwata and Shibasaki 2015; You et al. 2017). Ordinary Least Squares (OLS) is another linear regression approach for yield prediction using indices derived from satellite imagery over the growing season or over the period of several seasons. However, the main issue with OLS is that it doesn't account for spatial autocorrelation (L. Zhang, Lei, and Yan 2010). This is why research that applied spatial auto regression and Geographically Weighted Regression (GWR) did a better job in yield prediction when compared to OLS using NDVI from MODIS data (Tadesse et al. 2015; L. Zhang, Lei, and Yan 2010).

Chapter 3 Satellite Imagery for Crop Monitoring and Yield Forecasting in Idaho

3.1 Introduction

As the world continues to battle food insecurity, potatoes are considered important, fast growing, cheap and nutritious food source only behind rice, wheat and corn (FAO 2017; Rad et al. 2015; Sivarajan 2011). The US Pacific Northwest is a leading potato producer and processor with a contribution of 57% towards the nation's potato production in 2012 (Lewin et al. 2011). According to the National Agricultural Statistics Service (NAAS) most recent press release document (published online September 14, 2018) in 2017 Idaho produced 33.7% of the total US potato production, followed by Washington with 24.8% and Oregon 5.3% (https://www.nass.usda.gov/Statistics_by_State/Idaho/Publications/Crops_Press_Releases/2018/PT09_1.pdf). Idaho alone is responsible for one third of the country's potato shipments making potatoes an essential part of Idaho's economy as it is responsible for employing 46% of the Idahoan agricultural processing workforce (Lewin et al. 2011).

The implementation of precision agriculture informed by satellite imagery has the potential to aid agricultural practices and improve decision- making (Al-Gaadi et al. 2016; Khot et al. 2016; Kussul et al. 2017; Prasad et al. 2006; Satir and Berberoglu 2016; Wojtowicz, Wojtowics, and Piekarczyk 2016). The main aim for precision agriculture is to take into consideration the variability in the field rather than assuming that the field is uniform. Hence, all the field factors influencing growth and yield such as soil and topography are taken into account with precision agriculture approaches (Sivarajan 2011).

Satellite imagery for precision agriculture was introduced in the early 1970's (Mulla 2013) and has been expanding ever since especially with its increasing availability and options for analysis. Both multispectral and hyperspectral satellite imagery provide a resource for crop management, yield prediction, and vegetation health to support decision making (Atzberger 2013; Govender, Chetty, and Bulcock 2007; Johannsen 2010; Wojtowicz, Wojtowics, and Piekarczyk 2016). A challenge to satellite imagery analysis in the early 1970's and 1980's was the low spatial resolution of the available sensors and long revisit times (Atzberger 2013; Wojtowicz, Wojtowics, and Piekarczyk 2016). The increase in both spatial and temporal resolutions of satellite imagery in recent years provides a valuable resource for crop monitoring and specifically yield modeling and prediction. Not only has satellite imagery improved in quality, but it has also become more accessible to the public (Turner et al. 2015) through constantly updated datasets. For example, programs such as the European Space Agency's Sentinel missions, Planet's imagery products and Digital Globe's WorldView 2 and 3 provide imagery with resolutions reaching up to 50 cm and high temporal resolution allowing daily availability of imagery for a given area of interest. With access to such rich datasets, researchers are able to explain how fields change over the course of the growing season and utilize this information to inform managers' decision making for crop inputs and to forecast yield.

Sentinel-2A, launched in June 2015, and more recently, Sentinel-2B launched in March 2017, are two new multispectral imagers operated by the European Space Agency (ESA) covering 13 spectral bands (443 nm – 2190 nm) at resolutions of 10 – 20 m and 60 m. The Sentinel-2 program is filling a void for low-cost, medium resolution imaging with a five day revisit times depending upon latitude, cloud cover and other factors, to assess plant health and vigor during growing seasons (Dash and Ogutu 2016; Kussul et al. 2017). Gaadi et al. (2016) showed the

efficiency of using Sentinel-2 imagery along with various vegetation indices in predicting potato crop yield with a prediction model Root Mean Square Error (RMSE) as low as 4.96% in some fields. In addition, there is the PlanetScope satellite constellation operated by Planet that is freely available for university researchers. It has four spectral bands (455 nm – 860 nm) at a resolution of 3 m with a daily revisit. Both the Sentinel-2 mission and PlanetScope data are useful for precision agriculture applications.

Despite Landsat imagery providing an important tool for yield monitoring over time, the emergence of higher resolution datasets with a shorter satellite return period demonstrates an increased potential for precision agriculture data collection. For example, a study that assessed the performance of different sensors with varying resolutions (5 m for RapidEye, 3 m for PlanetScope, 10 m for Sentinel-2 and 30 m for Landsat) for yield monitoring concluded that higher resolution imagery gave more accurate results (Burke and Lobell 2017). While some previous studies on modeling yield prediction used Landsat (Song et al. 2016) and Sentinel-2 (Al-Gaadi et al. 2016), this study utilizes Sentinel-2A and PlanetScope analytical scenes for the analysis processing.

The most common vegetation indices used for monitoring crop health status are Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Soil Adjusted Vegetation Index (SAVI) and Modified Soil Adjusted Vegetation Index 2 (MSAVI2). The main difference is that NDVI reduces noise and provides an approach for comparing change over periods of time (A. Huete et al. 2002) while GNDVI is aimed towards detecting the greenness of the plant due to its sensitivity to chlorophyll content (Candiago et al. 2015). Similarly, SAVI and MSAVI2 are vegetation indices less sensitive to bare soil in individual pixels. While SAVI does a great job with minimizing the effect of soil in vegetation, MSAVI2 gives a more accurate estimation of wide range vegetation cover (Liu et al. 2007).

Satellite imagery has the potential to predict and monitor crop yield by correlating yield to different spectral bands, vegetation and water indices. These indices highlight crop health status and water stress. Previous yield studies demonstrated that the most efficient indices for yield prediction were NDVI (Prasad et al. 2006; Song et al. 2016; Yang et al. 2013), NDWI (Satir and Berberoglu 2016; Wojtowicz, Wojtowics, and Piekarczyk 2016; You et al. 2017) and SAVI (Al-Gaadi et al. 2016; Satir and Berberoglu 2016; Sivarajan 2011). Based on the work done by Al-Gaadi et al. (2016), yield was highly correlated to SAVI as it minimized the soil reflection.

One of the basic methods for crop yield prediction is to form an empirical relation between values of vegetation indices and ground measures (Lobell 2013). By using samples collected from the field, the correlation between vegetation indices and yield is determined through plot scatter graphs (Al-Gaadi et al. 2016) and is used to determine the accuracy of the predictions by validating them on the ground. A crop yield-prediction model takes into consideration various vegetation indices to correlate yield values. Satir and Berberoglu (2016), used seven indices for MLR yield model to explain the relation between yield and the indices variables

Regression analysis is a common approach to filter out the most significant spectral bands to explain crop yield and uses these band combinations and indices to create a prediction model. Sensors with varying spatial resolutions lead to different yield-prediction model accuracy. Prasad et al. (2006) used Advanced Very High Resolution Radiometer (AVHRR) with a 1.1 km resolution to calculate NDVI for predicting corn and soybean yield and obtained R^2 values of 0.78 and 0.86, respectively. Moderate Resolution Imaging Spectroradiometer (MODIS) data ranging between 250 m up to 1,000 m resolution, for soybean yield prediction using the water index NDWI gave a model accuracy of around 85% (Wojtowicz, Wojtowics, and Piekarczyk 2016; You et al. 2017) while the Hyperspectral Imager (Hyperion) sensor with a 30 m ground resolution NDWI values

gave an R^2 of 0.75 for soybean prediction (Wojtowicz, Wojtowics, and Piekarczyk 2016). As for the Landsat 8, with a spatial resolution varying from 15 m to 100 m, predicting potato yield via SAVI resulted in an R^2 of 0.81 (Sivarajan 2011) compared to a range of 0.39 – 0.65, according to the work of Al-Gaadi et al. (2016). Using NDVI based on Landsat 8 for wheat yield prediction resulted in an R^2 of 0.87 (Song et al. 2016) and 0.67 (Satir and Berberoglu 2016). Sentinel-2 imagery bands' resolution of 10 m, 20 m and 60 m, for potato yield prediction gave an R^2 range of 0.47-0.65 based on Al-Gaadi et al.'s work (2016).

This work aims to expand the use of multispectral satellite imagery for potato yield monitoring and prediction. The research objectives are to support the following hypotheses:

- Spectral indices provide an explanatory tool to understanding crop yield, in particular, SAVI and NDWI indices' values and yield values have a direct correlation.
- Spectral indices vary over the growing season based on the different growth stages of potato plants thereby allowing the indication of problems across the fields during critical times of the growing season.
- Vegetation and water indices are inversely proportional as the higher the vegetation index value is, the lower the water stress index value is.
- Higher temporal resolution imagery improves yield prediction models.

A limitation identified early in this study was the temporal resolution contrast between Sentinel-2A with a 10-day revisit period and Planet imagery with a daily revisit. In order to detect changes in the field as the season progresses, higher temporal resolution imagery is essential. Not only does Planet imagery allow the observance of change on a more detailed level, it also compensates for cloudy days due to the frequent revisit time. However, with Sentinel-2A, a

temporal resolution of 10 days means that one cloudy day results in the loss of a crucial data point when compared to Planet data. Although the launching of Sentinel-2B on March 7th 2017increased the temporal resolution by cutting the revisit time to every five days, Sentinel-2B data was unavailable over the 2017 growing season. Thus, this study relied solely on Sentinel-2A imagery. Despite the limitation, this approach demonstrated in this work has the potential to apply globally to any potato producing region/country. Similarly, it is modifiable to other crops such as wheat, corn, cotton, etc. based on the specific main crop of any given area.

3.2 Methods

3.2.1 Study Area

The study area is located near Parker in southeast Idaho within Fremont County, USA. There are ten fields within the area of interest as shown in Figure 9 with a total area of 402.98 ha (Table 5) Parker is located within Idaho's agriculturally productive Snake River Plain at an elevation of 1,502 m above mean sea level and characterized by a humid continental climate where the winters are cold and summers are hot. Generally, the wettest month of the year is May while the driest is August with a total average annual precipitation of 357.6 mm (AgriMet 2017). Crops rely on irrigation with water drawn from the Henry's Fork of the Snake River, which flows through the study area. Potatoes are the dominant crop in the area and the fields are on a potato and cereal grain rotation cycle where one year of potato planting is followed by two years of cereal crops before going back to planting potatoes (Grow 1993; Lewin et al. 2011).

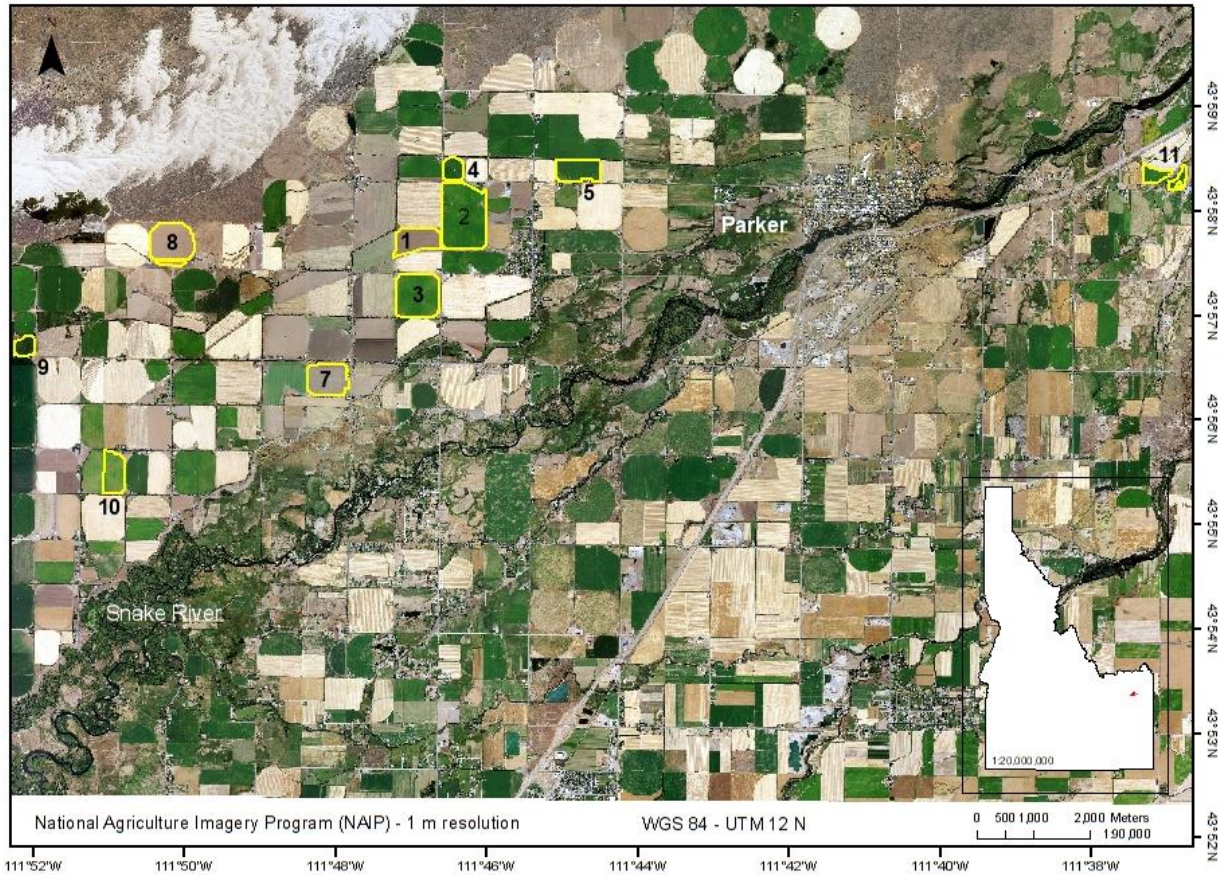

Figure 9 Study area showing ten field locations near Parker, Idaho. (WGS 84 - UTM 12 N)

Table 5 Field Areas

Field	1	2	3	4	5	7	8	9	10	11	Total
Area *10^2 (m^2)	3,145	8,666	5,655	1,390	2,848	3,772	3,689	1,096	2,656	7,381	40,298

The Idaho farmers supported the data analysis by providing information for each field that included planting dates and the potato variety (Table 6). Based on planting date and potato variety, the fields were sub grouped into three categories so that fields within the same group have similar growing timeline which influences the variables for the yield forecasting models depending on date and variety. Different planting dates result in offset between models while each potato variety has its own growth cycle (different maturity periods).

- G1: fields 2, 3, 4 and 5 planted with Norkotah potatoes on April 15th and 16th
- G2: fields 1, 7, 8 (a & b), 9 and 10 planted with Russet potatoes on April 13th and 14th
- G3: fields 11 a & b planted with Norkotah potatoes on May 3rd

Table 6 Crop Variety and Growing Information

Field	Potato Variety	Planting Date	Group	Number of Planet Scenes
1	Norkotah	April 13	G1	24
2	Russet	April 15	G2	24
3	Russet	April 15	G2	24
4	Russet	April 16	G2	24
5	Russet	April 15	G2	24
7	Norkotah	April 13	G1	24
8 (a & b)	Norkotah	April 13	G1	24
9	Norkotah	April 14	G1	24
10	Norkotah	April 14	G1	24
11 (a & b)	Russet	May 3	G3	28

3.2.2 Image Processing

The workflow is comprised of two main processes: satellite imagery processing and yield data processing (Figure 10) to which the results from both processes input into the yield prediction model. Sentinel-2 imagery was obtained from the U.S. Geologic Survey's EarthExplorer website (https://earthexplorer.usgs.gov/). In order to be able to perform the analyses needed on the Sentinel-2 dataset, ESA's SNAP software converted the top of atmosphere (TOA) values to surface reflectance using the Sen2Cor plugin for atmospheric correction. Twenty four 4 – band imagery scenes were downloaded from Planet website (https://www.planet.com/products/planet-imagery/) via their education and research application interface. The PlanetScope imagery is the surface reflectance ready product provided by Planet so no further atmospheric correction was needed. A python code (Appendix A – Planet Python Code) extracted the individual bands needed for processing.

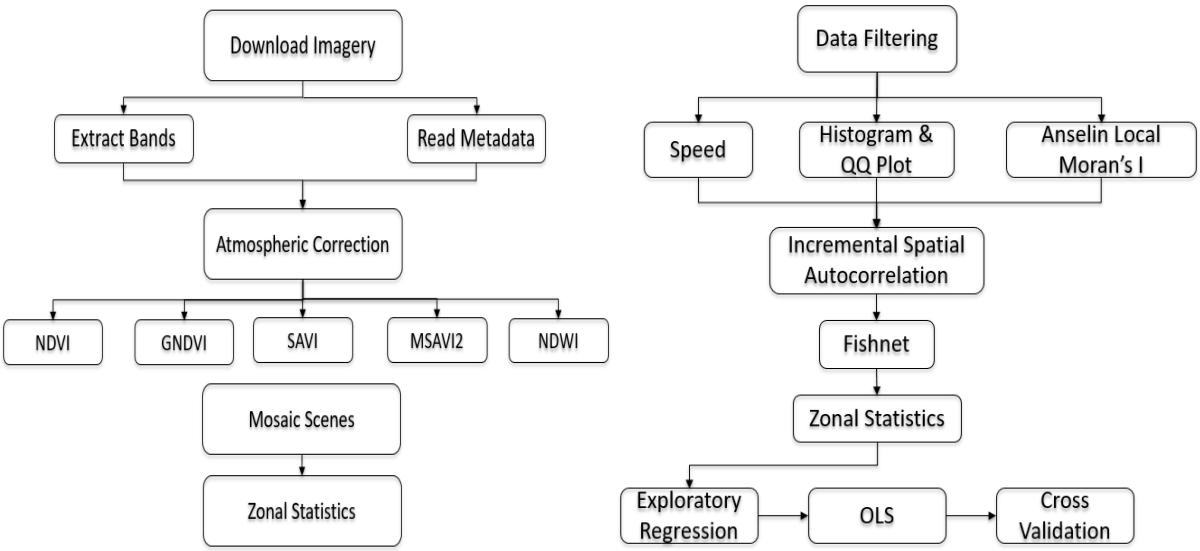

Figure 10 Satellite imagery and yield data processing workflow

3.2.3 Yield Mapping

The farmer provided the yield data from the 2017 growing season, which was collected using the GK Technology for Agriculture sensor (Figure 11). The sensor attaches to the harvester and records the speed, yield, and pounds per second along with the geographic coordinates every 0.8 meters by 10 meters. The raw yield data format is compatible with Microsoft Access.

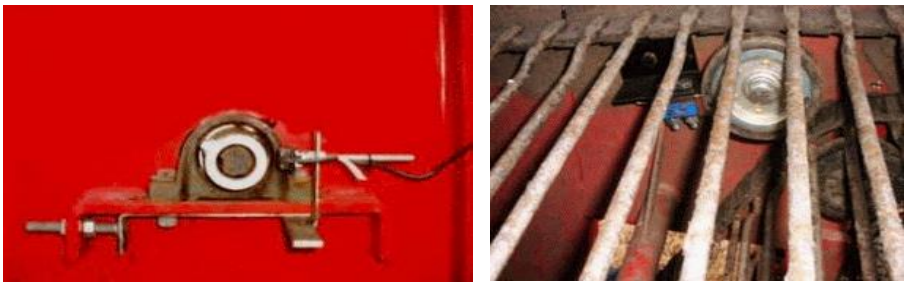

Figure 11 Speed sensor allowing cop yield monitoring.
Source:http://www.geektechforag.com/Products/BYM/Beet_Yield_Hardware/Beet_Yield_Hardware.aspx

3.2.4 Removing Data Outliers

Yield data filtering consisted of removing outlier yield points from the dataset through three steps: (1) accounting for machine and human errors, (2) speed, yield values and (3) Anselin Local Moran's I. Human and machine errors consisted where there was harvester taking partial paths, turn-around, in-field roads where there is no yield data and field edges with outlier yield values which resulted in a lack of sufficient data within such spots and were thus omitted. Since speed is an important factor that influences the data collection, it was essential to reduce the noise in the data through maintaining a margin of $\pm 2\sigma$ (speed standard deviation) by removing points outside this range. In addition, yield values were constrained to a range of $\pm 2.5\sigma$ (yield standard deviation). Examining the histogram and QQ plot of each field comprised the second step of the filtering process. The statistics of the histogram, the skewness and kurtosis, as well as the tail of the histogram provided a base for filtering outliers. The skewness value determines the asymmetry of the data and its histogram as well as determining the relationship between the data and the mean. As for the kurtosis, it determines the peakedness of the histogram. These two statistical values along with the visualization of the QQ plot allowed the filtering of the data. The last step was executing Anselin Local Moran's I (Tiefelsdorf and Boots 1997) via the Geostatistical Analysis toolkit in Esri's ArcGIS Pro 2.2.2 to determine any remaining outliers and/or clustering in the data. After removing the outliers, the data exhibited a normal distribution and was ready for processing.

3.2.5 Crop Variability

One of the main advantages of precision agriculture is taking into consideration the variability within the field rather than assuming it is uniform. Hence, dividing the field into subplots ensures accurate representation of indices' variation and yield output within each zone. In order to determine an optimal subplot size, Esri's Incremental Spatial Autocorrelation tool

measured the spatial autocorrelation for the data points in each field and reported the z-scores at which distance clustering is prominent. A cell size of 80 by 80 meters was selected to ensure consistency in analysis and represented the range of values returned from this step. After establishing cell size, the Fishnet tool creates the grids over each field. The Zonal Statistics tool reads in each raster layer from satellite imagery variable along with yield values and reports the mean value per grid cell. Due to the large number of available variables for both Sentinel and Planet datasets (three bands and five indices per date for the duration of the growing season), a Python code facilitated automated using the Anaconda Spyder compiler. This code generated a geodatabase table and an Excel spreadsheet that had the variable name followed by the date as a header and the grid cells of each field consisted of the rows of the table. Similarly, the Zonal Statistics tool summarized the indices' values throughout the field for every given date as shown in Figure 10.

3.2.6 Vegetation and Water indices

Due to the leaf cell structure and the varying reflectance response to different wavelengths of the electromagnetic spectrum, vegetation and water provide information about crop health (Table 7). Chlorophyll, the green pigment in vegetation, absorbs blue and red wavelengths while the green wavelength reflects partially. As for the near infrared section of the spectrum, it scatters and reflects by the cell walls in the mesophyll. In order to visualize the calculated indices, pixel values within an 80m x 80m grid cell were averaged for all 20 dates and added to the attribute table. In addition to that, the National Agriculture Imagery Program (NAIP) provided the access to the National Elevation Dataset which was used to calculate the slope then averaged within the fishnet grid squares and added to the attribute table as well.

Table 7 Spectral Indices

Vegetation Index	Formula	Reference
Normalized Difference Vegetation Index	$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$	(Rouse et al. 1973)
Green Normalized Difference Vegetation Index	$\frac{\rho_{NIR} - \rho_{Green}}{\rho_{NIR} + \rho_{Green}}$	(Gitelson, Kaufman, and Merzlyak 1996)
Soil Adjusted Vegetation Index	$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red} + L} * (1+L)$	(A. R. Huete 1988)
Modified Soil Adjusted Vegetation Index 2	$\frac{2\rho_{NIR} + 1 - \sqrt{(2\rho_{NIR} + 1)^2 - 8(\rho_{NIR} - \rho_{Red})}}{2}$	(Qi et al. 1994)
Normalized Difference Water Index	$\frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}}$	(S.K. 1996)

3.2.7 Yield Forecasting

The Python code (Appendix A – Planet Python Code) extracted all mean values for each index and bands red, blue and NIR along with the mean yield values for each grid cell making it suitable for running regression analysis to explain yield variability within the fields. Exploratory Regression in ArcGIS Pro helped reduce the number of potential explanatory variables for the yield model through several runs. Subsets of the explanatory variables included a selection of the Green, Red and Near Infrared bands while others used the different indices: NDVI, GNDVI, SAVI, MSAVI2 and NDWI. In addition, a subset included a combination of bands with spectral indices (Appendix B – Regression Models). From each run, the most significant variables became the new

subset for the next run. After three exploratory regression runs, the variables that had the highest correlations were used for input into the Ordinary Least Square (OLS) tool.

3.3 Results

3.3.1 Variation of Indices over Growing Season

All five indices used (NDVI, GNDVI, SAVI, MSAVI2, and NDWI) showed similar variation over the fields throughout the season (Figure 12, Figure 13 & Figure 14). However, the PlanetScope imagery gave a more detailed observation about what was happening in the fields along with determining critical dates and stages for the growing season due to its higher temporal resolution. The PlanetScope dataset identified specific dates regarding plant emergence, row closure and full maturity in addition to the dates with higher temperatures and increased evapotranspiration rates.

In both the Sentinel and Planet imagery for all the fields, NDVI and SAVI increase gradually during the season to reach a peak value of 0.9 and then decrease to around 0.8. However, the limitation of Sentinel-2A availability resulted in fewer data points between critical times over the growing season (Figure 12, Figure 13, & Figure 14). PlanetScope imagery gave a more detailed interpretation of the variation between the weeks and the fields. While the water index seems to vary inversely proportional to the other indices, it actually varies negatively proportional to them. The water index measures the water stress level in the crops so when the plant is healthy and showing high NDVI and SAVI values, it indicates that it is rich in water and consequently the water stress is low. This is why the plot for the NDWI decreases over the growing season rather than increase to show the progress of the crop cycle and the plant's response during different stages of the season such as emergence, full bloom and harvest.

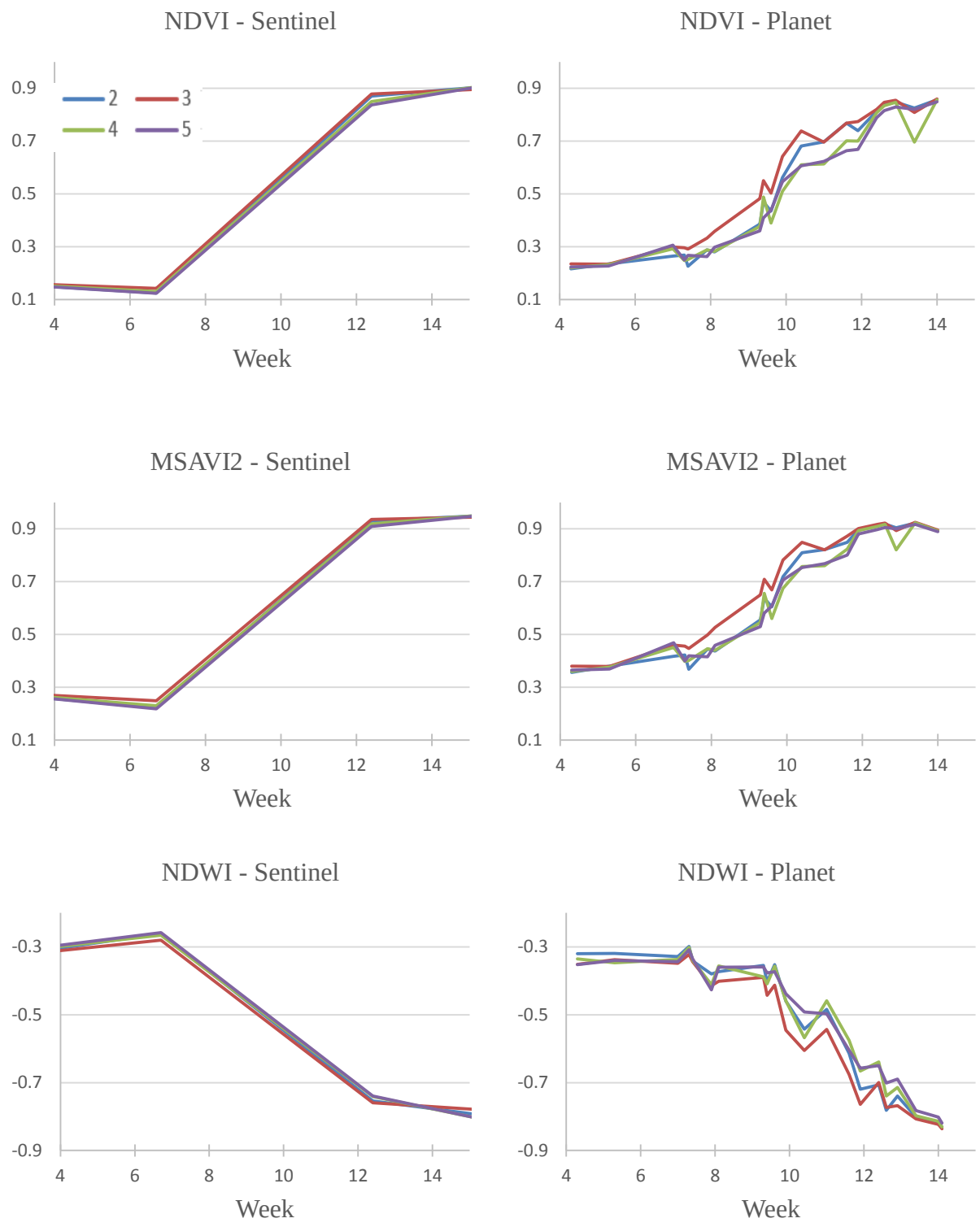

Figure 12 Idaho: Russet Variety - G1 Fields 2, 3, 4 & 5

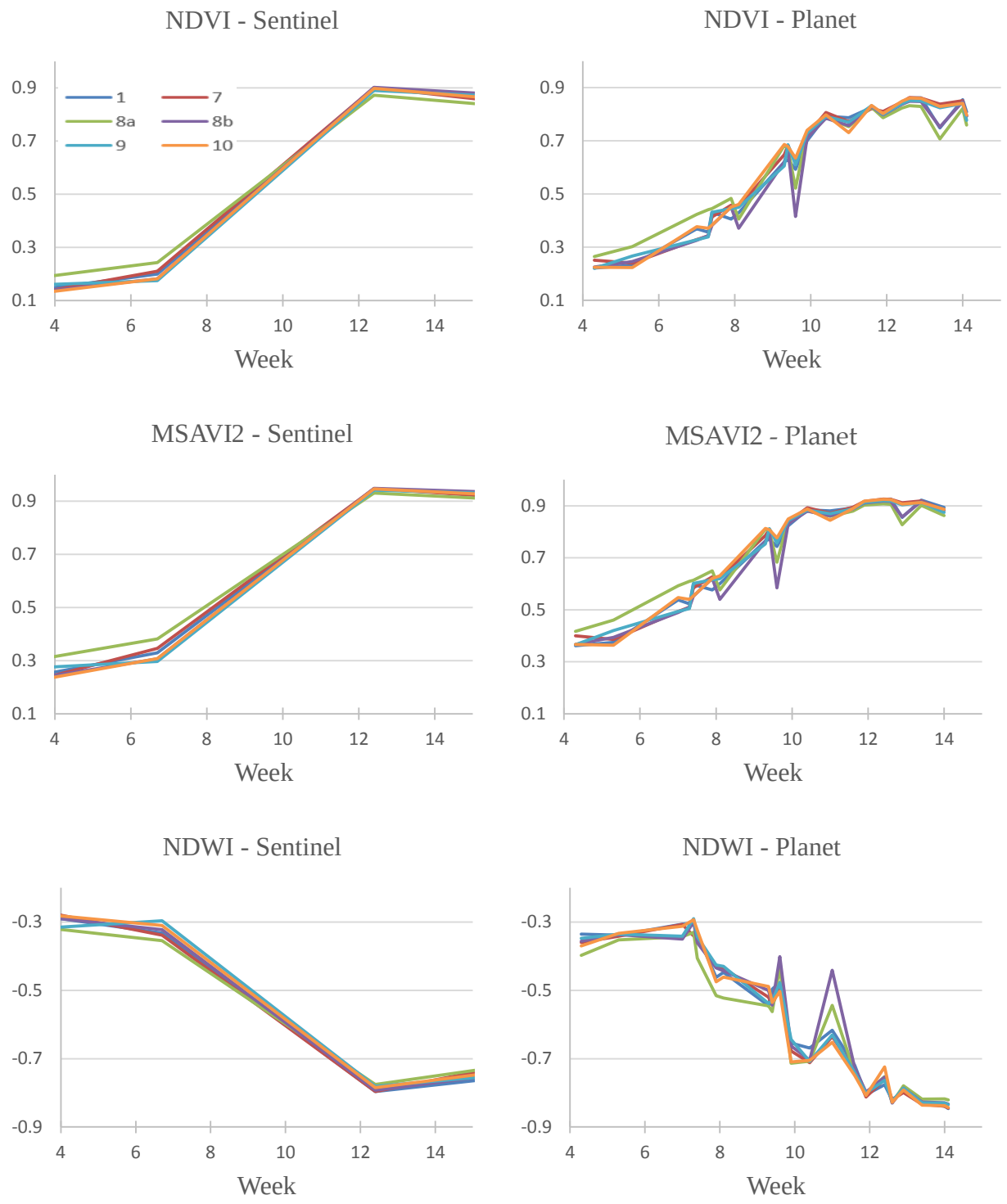

Figure 13 Idaho: Norkotah Variety - G2 Fields 1, 7, 8a, 8b, 9 & 10

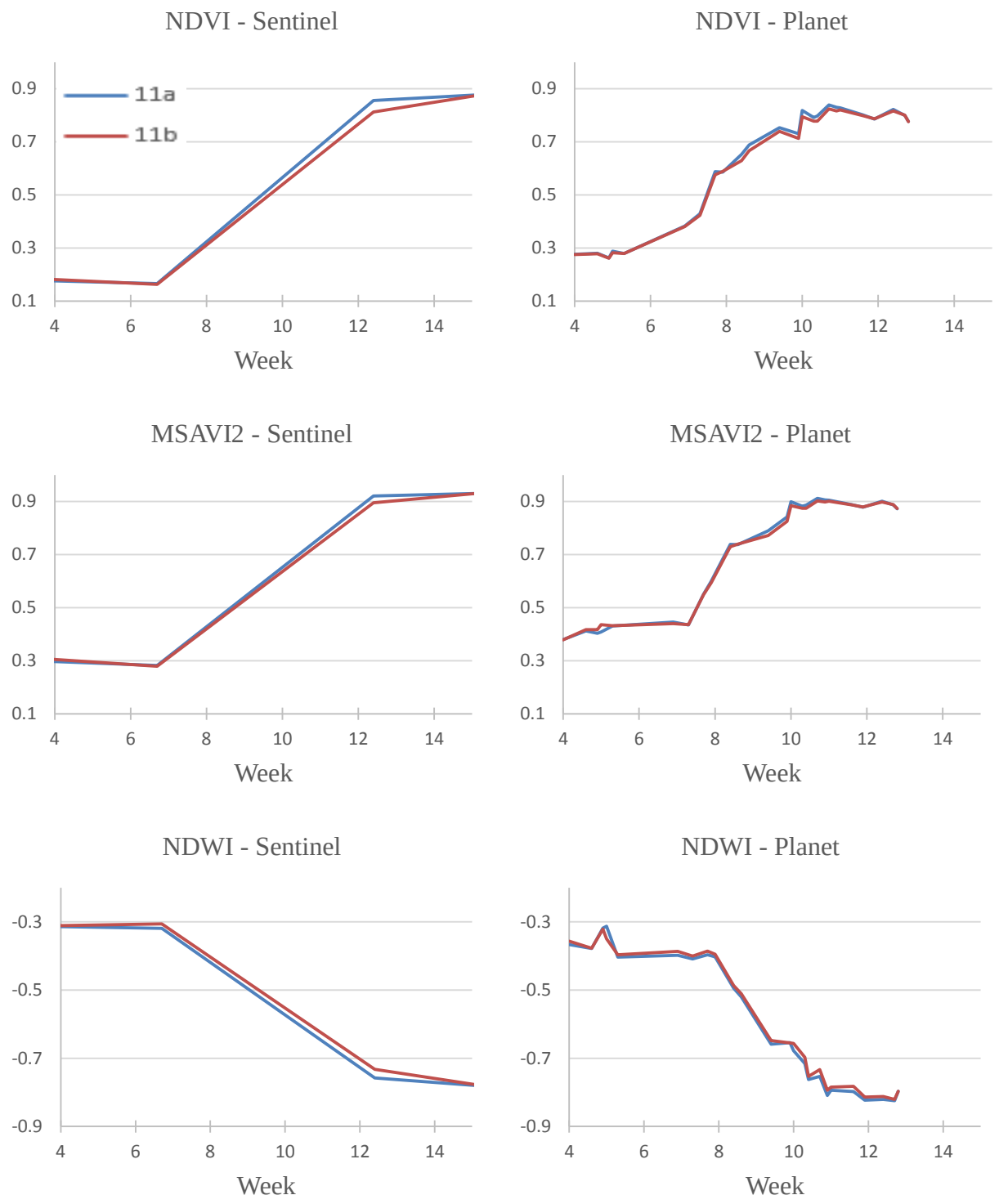

Figure 14 Idaho: Russet Variety - G3 Fields 11a & 11b

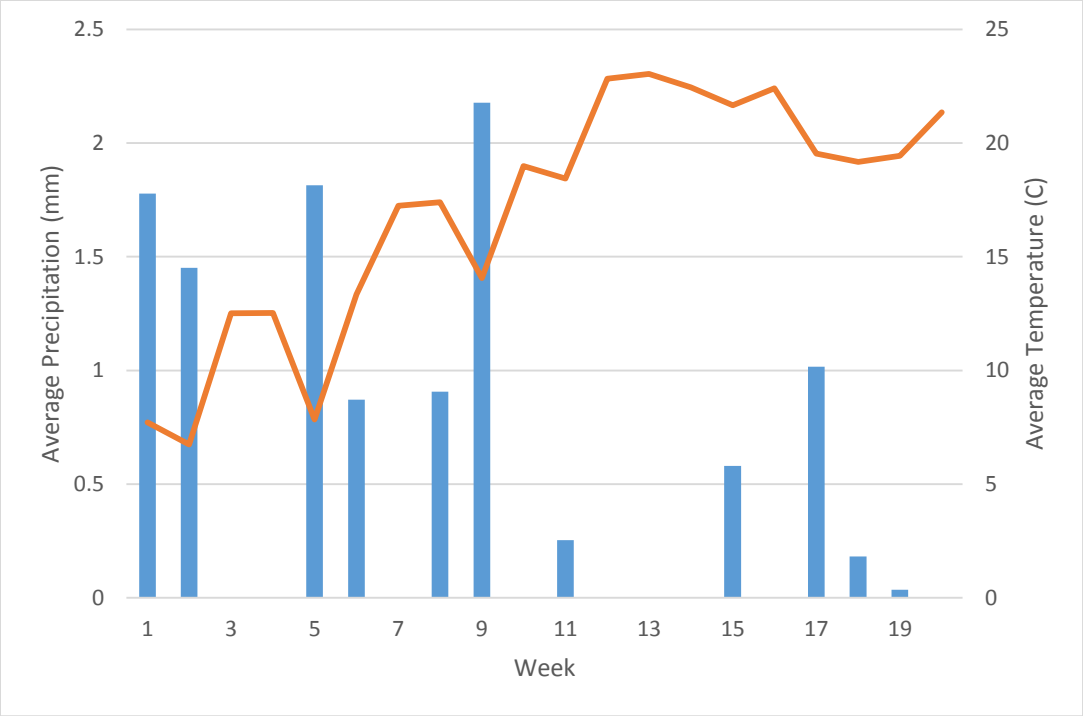

Figure 15 Weekly average precipitation (mm) and temperature (ºC) over the growing season of 2017.
Data source: https://www.usbr.gov/pn/agrimet/agrimetmap/terida.html

Potatoes are an irrigated crop and water is a very crucial element in the development of the crops. Thus, it is important to study precipitation and temperature records that relate to indices' and yield values. A better understanding of those weather variables provides vital information related to critical weeks in the growth stages to help the farmers with irrigation decisions. The daily precipitation and temperature over the growing is from the Idaho National Lab (INL) weather station located at 43.841683 N, -112.41825 W. The data was downloaded through the Agrimet website, weekly averages were calculated and plotted into the same graph (Figure 15).

3.3.2 Planet Data Yield Regression Model

The dependent variable used is the yield as it is what we are trying to explain through the regression model. For all three fields groups, the different vegetation and water indices allowed the explanation and prediction of yield values. Averaging SAVI and slope values from the most

critical week for each potato variety resulted in a best-fit model for yield forecasting with highest R^2 value and low VIF values.

3.3.2.1 Russet Fields (Early Season) Regression Model

The first group of fields (G1) includes fields 2, 3, 4 and 5 planted with Russet potato on April 15 and 16 (Table 6). The significant week for yield prediction was week 12 with the combination of average slope and SAVI as explanatory variables for yield (Table 8). With an R^2 value of 0.444, the model's residuals followed a normal distribution with a bell shape and a random distribution of residual versus predicted values (Figure 16).

Table 8 Russet fields regression model variables

Variable	Coefficient	StdError	t-Statistic	Probability	VIF
Intercept	133.967	33.678	3.978	0.000*	
Average Slope	-75.273	8.375	-8.988	0.000*	1.023
SAVI	344.912	30.164	11.435	0.000*	1.023

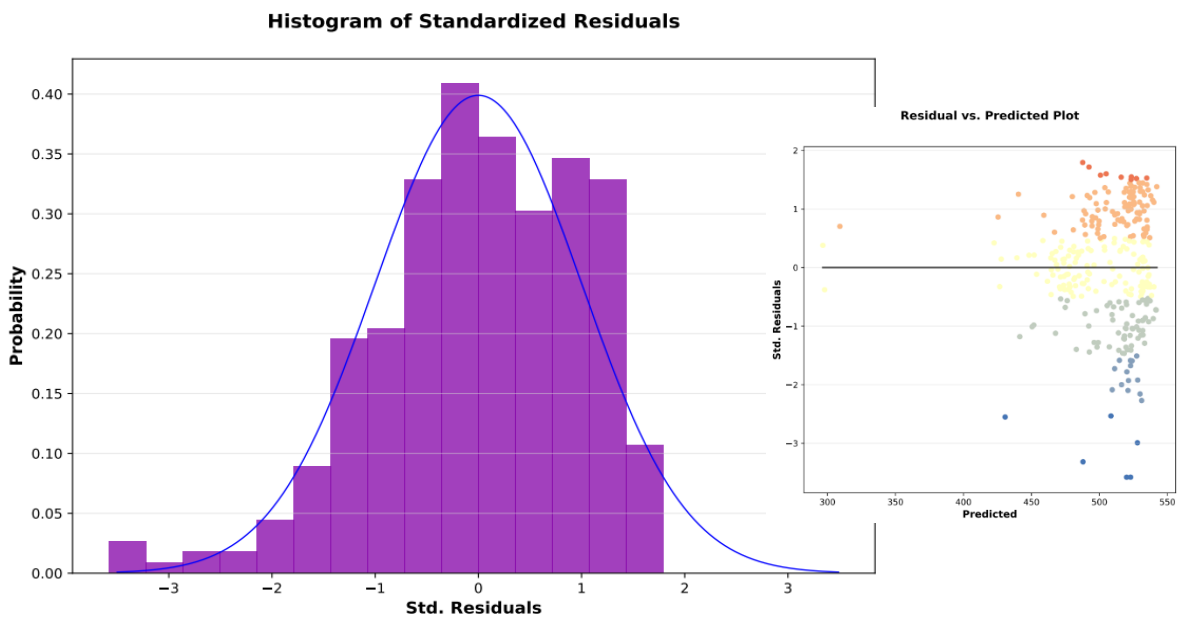

Figure 16 Summary statistics of Russet fields regression model

3.3.2.2 Norkotah Fields (Early Season) Regression Model

The second group of fields (G2) includes fields 1, 7, 8 (a & b), 9 and 10 planted with Norkotah potato variety on April 13 and 14 (Table 6). The significant week for this group of fields and potato variety yield prediction was week 10 using SAVI as the explanatory variable for yield (Table 9). With an R^2 value of 0.570, the model's residuals followed a normal distribution with a bell shape and a random distribution of the residual versus predicted values (Figure 17).

Table 9 Norkotah fields regression model variables

Variable	Coefficient	StdError	t-Statistic	Probability
Intercept	-193.982	40.188	-4.827	0.000*
SAVI	800.392	41.542	19.267	0.000*

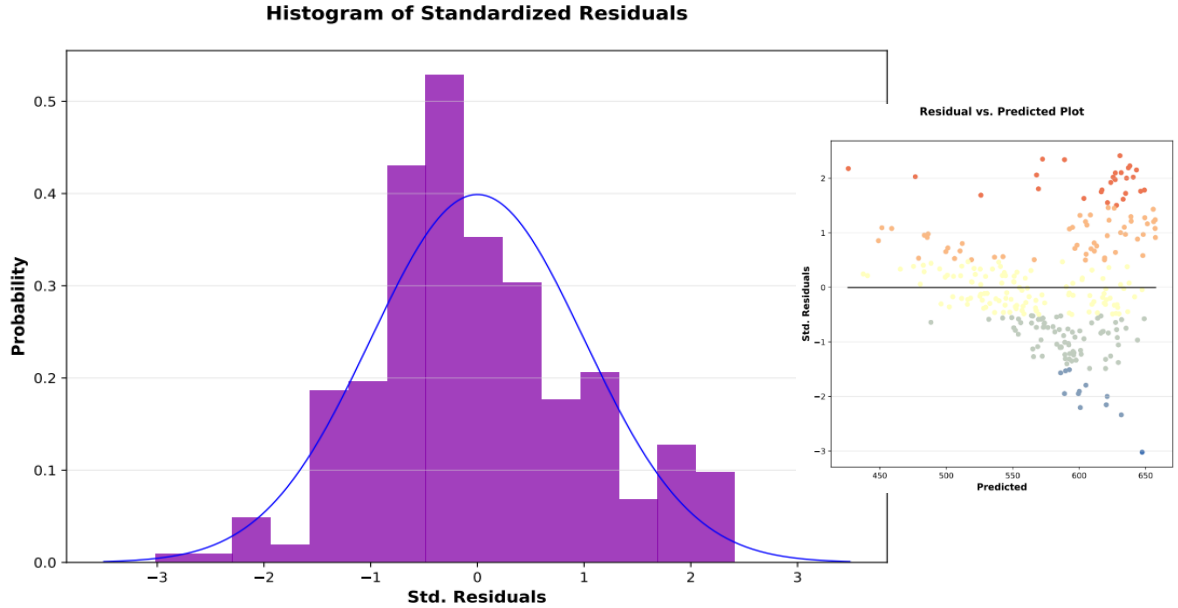

Figure 17 Summary statistics of Norkotah fields regression model

3.3.2.3 Russet Fields (Late Season) Regression Model

The last group of fields (G3) is fields 11 a and b planted with Russet potato later in the season on May 3rd (Table 6). The significant week for this group of fields and potato variety yield prediction was week 12 (3.3.2.1). The combination SAVI and slope are the explanatory variables

for yield (Table 10). With an R^2 value of 0.633, the model's residuals followed a normal distribution with a bell shape and a random distribution of residual versus predicted values (Figure 18).

Table 10 Russet field 11 regression model variables

Variable	Coefficient	StdError	t-Statistic	Probability	VIF
Intercept	296.297	118.158	2.508	0.017	
Average Slope	37.363	5.425	6.887	0.000*	1.001
SAVI	208.697	99.517	2.097	0.044*	1.001

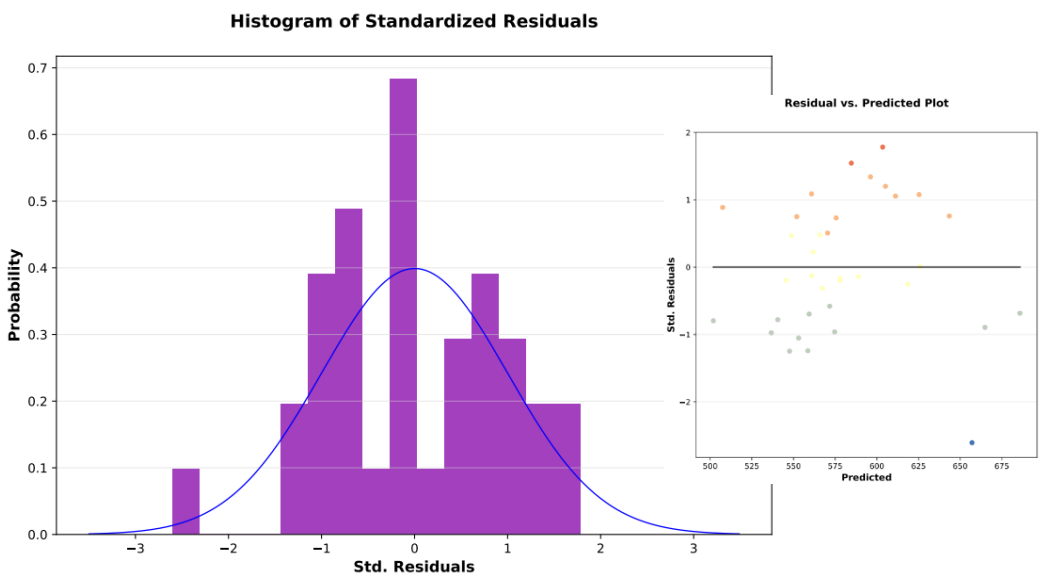

Figure 18 Summary statistics of Russet fields 11 regression model

3.4 Discussion

The graphs from the PlanetScope imagery indicated the different stages of growth for the potato plants. Despite the difference in potato variations and having a two week offset in maturity, all fields exhibit similar indices responses. As the plants started to emerge, the indices increased

gradually to show a slight peak in values at around week 4. A significant drop in the curves also appeared at week 9 (Figure 12, Figure 13 & Figure 14). This drop indicates high stress levels in the crops relating to highly recorded temperatures, and winds, which meant the crops needed extra water to compensate for the weather conditions (Figure 15).

With the Norkotah variety having an earlier maturity stage than the Russet crops by two weeks, the significant week for the Norkotah in both G1 and G3 despite the offset turned out to be week 10 as compared to week 12 on the Russet. All of the regression models proved that SAVI along with average slope best explain and predict yield (Appendix B – Regression Models). According to previous works and research, NDVI and SAVI are generally the best indices to explain yield (2.7), however, none of these studies use high temporal and spatial resolution as that of Planet neither do they incorporate elevation data into the regression model.

Utilizing the Planet data with the three meter resolution and almost a daily revisit time allows the capturing of high details within the fields as opposed to lower resolution data that results in merging pixels together and as a result falsely give higher model R^2 values. In addition, in all the regression models of all three groups of fields (G1, G2 & G3) showed that the average slope is a significant explanatory variable for yield. The slope has a direct impact on the crops since with higher slope values, the land is subjected to erosion and as a result impact the quality and types of soil within the fields. In addition, the microtopograohy impacts the level of amount of water reaching the plants. Using spectral indices along with taking into account the microtopography as it constantly shows as a significant variable.

Satellite imagery shows promise to help farmers identify problems within the field early on as well as indicating critical time points in the growing season that are specifically important for yield prediction. Despite the fact that various indices reveal different information about the

crops, the combination of the three bands is the most important for crop yield prediction. However, there are some gaps that can only be closed by higher resolution data and as a result better improve the yield prediction model. Even though satellite imagery technology is rapidly advancing, integrating UAS imagery and potentially hyperspectral datasets into the analyses would be the next step in this work for an improved outcome. In addition, it is important to work closely with the farmer and perform on-site validation of what the indices show.

Chapter 4 Utilizing Satellite Imagery for Precision Agriculture in Lebanon

4.1 Introduction

Lebanon (Figure 19) relies heavily on its agricultural sector with an agricultural area of approximately 2,730 km^2 out of a total 10,452 km^2 (El Gazzar 2015). The agricultural sector contributes 7% to the GDP while providing work to 15% of the Lebanese population (FITA 2008). Potato crops account for 56% of vegetable production in the country, mainly in the Bekaa Valley and North Lebanon states (Hatoum 2005b). Hence, it is important to monitor the health and progress of potato crops to improve production in order to revive the local market as growers are having a hard time selling their produce (المراحل أخطر في نمر :درباس المزارعين معاناة سلام إلى نقل شهيب 2015) and expanding the Lebanese economy by increasing potato exports. The Bekaa Valley, located in the center of Lebanon, is one of the largest agricultural regions in the country as the valley is 120 km (75 miles) long and averages 16 km (9.9 miles) across with potatoes as its main crop. Potatoes are an important irrigated crop vulnerable to water stress, pests, disease and other crop threats. Precision agriculture has the potential to help minimize such issues and improve crop yield by empowering farmers with timely scientific knowledge on crop conditions. This approach offers valuable guidance to farmers; however, in a country such as Lebanon, precision agriculture is still in its early stages of adoption. Lebanon recently teamed up with the Food and Agriculture Organization of the United Nations on May 16, 2016, to launch the Country Programming Framework from 2016 to 2019 (United Nations Food and Agriculture Organization 2016) to develop more sustainable practices to improve the agricultural sector and may include the concept of precision agriculture.

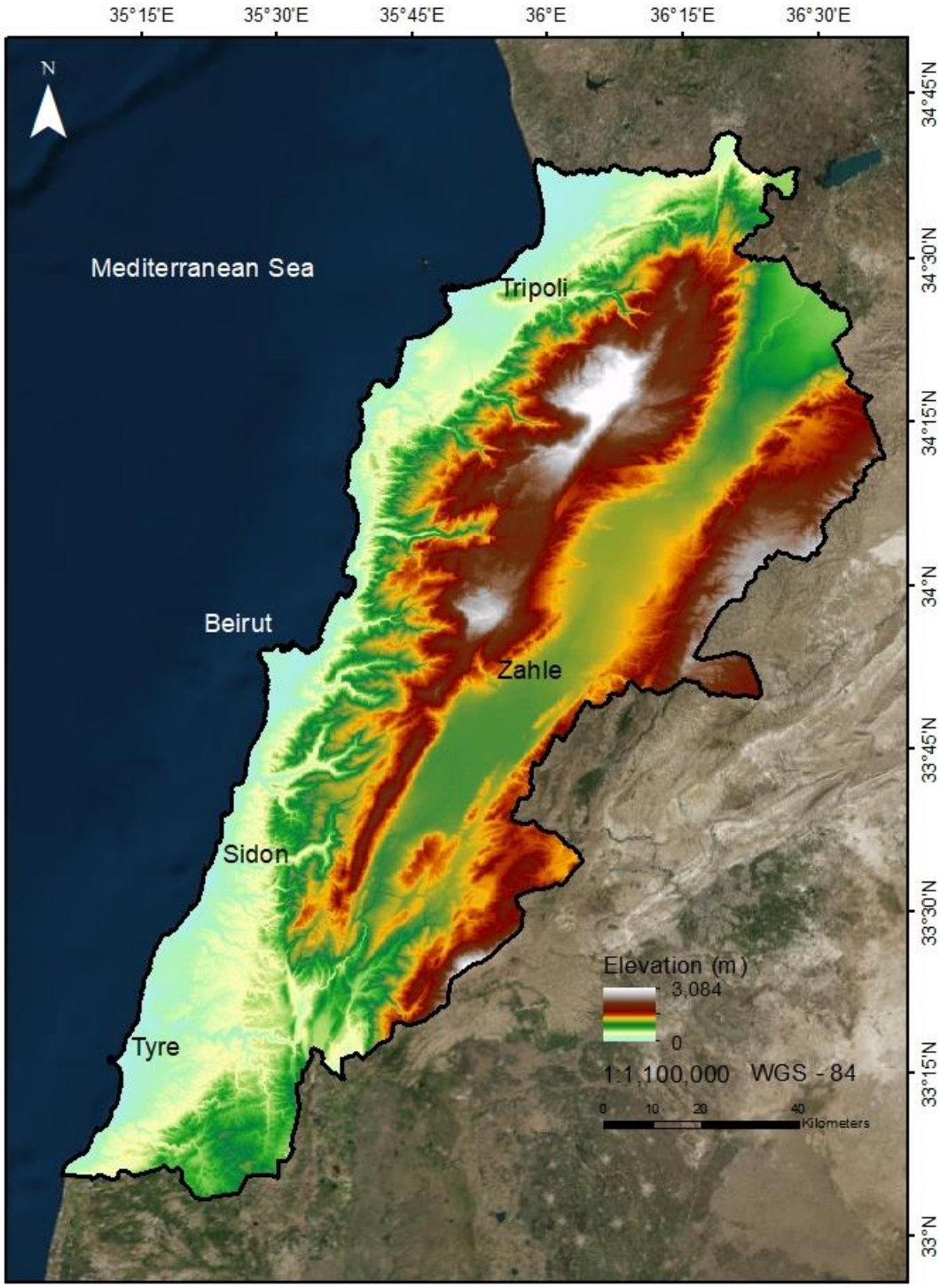

Figure 19 Lebanon Elevation Map. Elevation Dataset: Global Multi-resolution Terrain Elevation Data 2010 (GMTED 2010) – USGS EarthExplorer Website

This study aims to show the potential of precision agriculture through utilizing both open-access satellite imagery and high-resolution satellite imagery to monitor crops throughout the growing season and identify critical dates during the growing season with the goal to help farmers with important decision- making. The primary objective of this study is to submit this work to the Lebanese Ministry of Agriculture to demonstrate how satellite image analysis can assist farmers to respond to crop health issues.

Satellite imagery technology has improved steadily since the early 1970's, as reflected by increased sensor resolution and expanding applications of satellite imagery to cover precision agriculture applications (Mulla 2013). Not only has satellite imagery improved in quality, but they have also become more accessible to the public (Turner et al. 2015) through constantly updated datasets such as the Sentinel-2 mission by the ESA. Sentinel-2A, launched in June 2015, and more recently, Sentinel-2B launched in March 2017 which are two new multispectral imagers covering 13 spectral bands (443 nm – 2190 nm) at resolutions of 10 – 20 and 60 m. The Sentinel-2 program is filling a void for low-cost, medium resolution imaging with 5 day revisit times depending upon latitude, cloud cover and other factors, to assess plant health and vigor during growing seasons (Dash and Ogutu 2016). In addition, there is the RapidEye satellite constellation operated by Planet Labs that is freely available for university researchers. It was launched on August 29 of 2008 with 5 spectral bands (440 nm – 850 nm) at a resolution of 5 m. Both the Sentinel-2 mission and RapidEye data are partially aimed for agricultural uses making them an ideal source for researching precision agriculture.

Irrigation plays a major role in the health status of potato crops, and as the world's climate is constantly changing, it is important to analyze water stress and content in crops to manage water resources more efficiently and to assess what irrigation techniques are the most efficient.

According to research performed in Egypt, (Nahry, Ali, and Baroudy 2011), using satellite imagery proved to be very effective in increasing economic profitability and environmental sustainability respectively by 29.89% and by limiting fertilizers and irrigation to where it is only needed. Using multispectral satellite imagery such as Sentinel data, NDWI could be easily calculated to provide a better assessment of water stress in agricultural fields (Gao 1996).

The main vegetation indices used for monitoring crop health status are NDVI, GNDVI, SAVI and MSAVI2. Both NDVI and GNDVI have similar purposes in which they reflect how dense the vegetation by showing the health status of the dense leaf cover of the planted area. The main difference is that NDVI reduces noise and provides an approach for comparing change over periods of time (A. Huete et al. 2002) while GNDVI is more aimed towards the greenness of the plant due to its sensitivity to the chlorophyll content and is related to the Leaf Area Index (Candiago et al. 2015). Similarly, SAVI and MSAVI2 are vegetation indices less sensitive to bare soil in the pixel. While SAVI minimizes the effect of soil in vegetation, MSAVI2 gives a more accurate estimation of wide range vegetation cover (Liu et al. 2007).

This study focused on demonstrating the potential of Sentinel-2 and PlanetScope imagery to monitor crop progress over the growing season and identify important growth stages for yield prediction. Our hypothesis is that low vegetation or water indices during critical times over the growing season indicates problems in the field and will correspond to lower yield. Despite the offset in the season between Lebanon and Idaho, and the difference in potato varieties, the Idaho potato yield forecasting model is adjusted and used to predict yield in the Lebanon fields. This study demonstrates the importance of high spatial and temporal resolutions for more accurate results in depicting changing crop condition over the growing season.

4.2 Methods

4.2.1 Study Area

The study area is located in Tal Znoub in the southwestern part of the Bekaa Valley in Lebanon (Figure 20). It lies north of Quaroun Lake and is along the path of the Litani River. It is located at around 33.66 N latitude and 35.78 E longitude with an altitude of 872 m (5861 ft.) above mean sea level. Tal Znoub is located at 4 km north northeast of the city of Jeb Jannine, the capital of the West Bekaa.

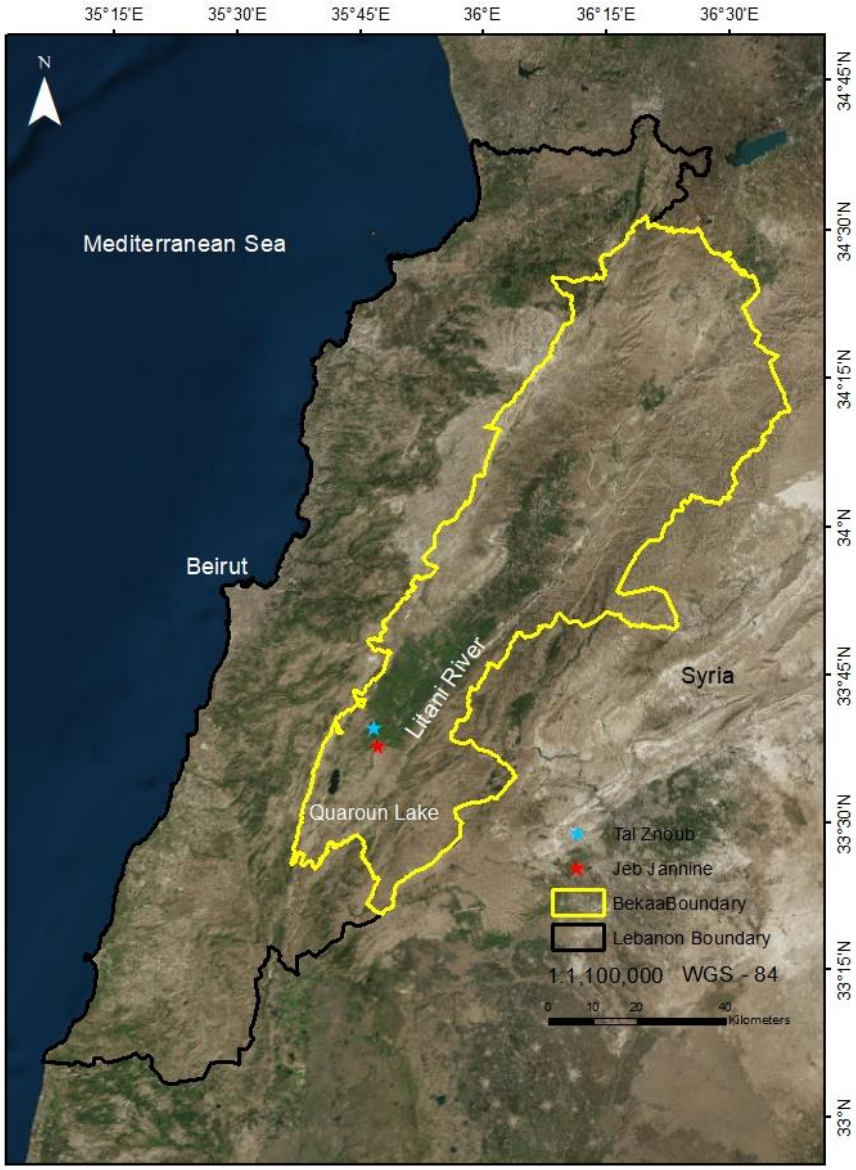

Figure 20 Location of Tal Znoub within the Bekaa Valley, Lebanon.

The overall area of the site is 462,264 m^2 made up of 16 sub fields as shown in Figure 21 and Table 11. Just like Lebanon overall, the study area has a moderate climate with the hottest months of the year between June through September, while the rainy season is between November and February. The area is known for growing mainly potatoes, vegetables, grapevines, and various grains including wheat. The Bekaa Valley is uniform when it comes to soil types mainly being *Lithic Leptosols* as the soil type comprising this study area in Tal Znoub.

Table 11 Field Areas in m^2

Field ID	**Area (m^2)**	**Field ID**	**Area (m^2)**
1	6,047	8	29,008
2	11,640	9	32,074
3	19,128	10	35,069
4	13,896	11	30,711
5	24,386	29	16,864
6	38,362	30	18,953
7	65,375	33	51,012
Total		392,525	

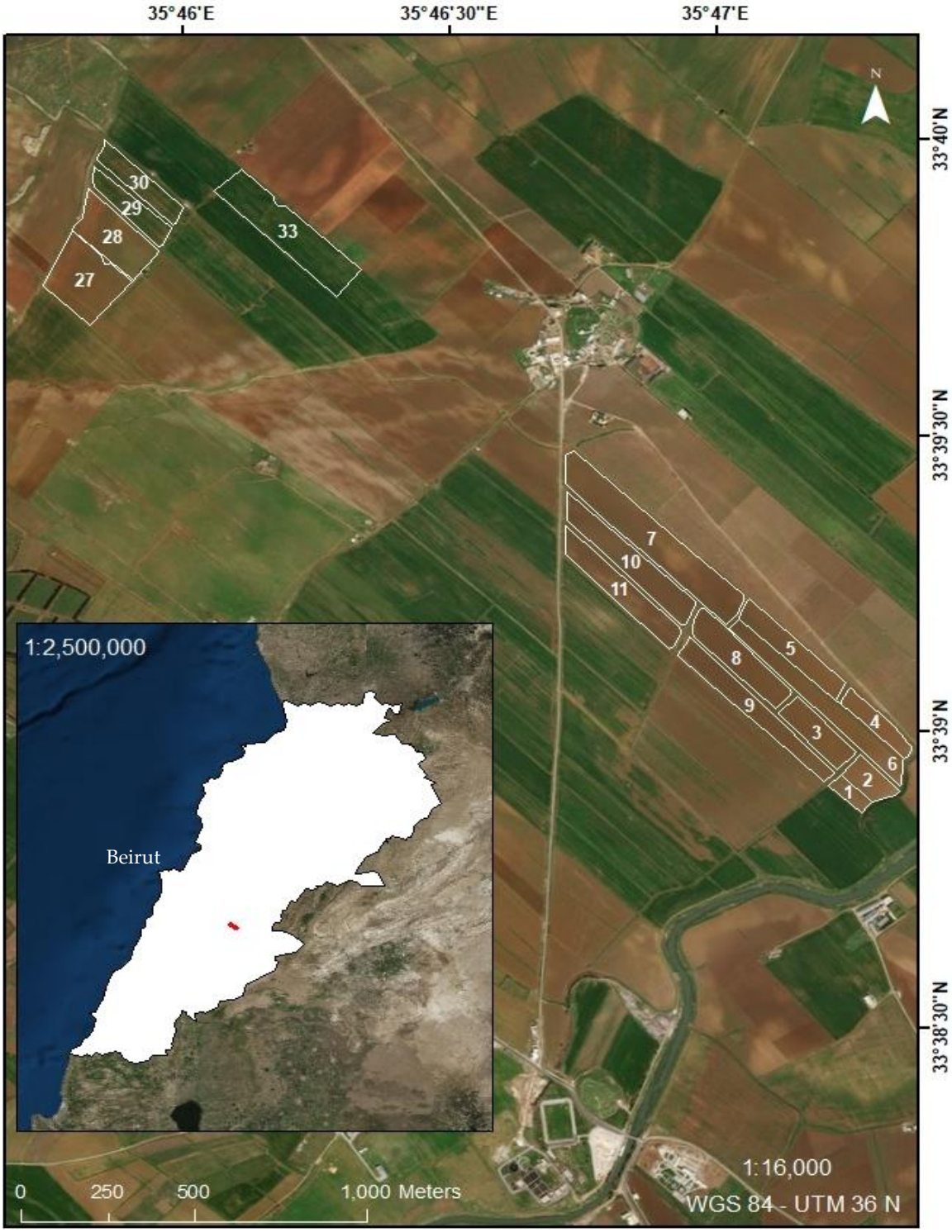

Figure 21 Study Area showing the location of 16 fields within Lebanon in the WGS 84 - UTM 36N coordinate system at a scale of 1:16,000
Source of inset map: ArcGIS Online: https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/LBN_Boundaries_2016/FeatureServer

4.2.2 Data Resources

The satellite imagery for this work composed of Sentinel-2A and PlanetScope data. The USGS Earth Explorer interface allowed open access to the Sentinel-2A imagery where the data is freely available to the public thanks to the ESA, operated by European governments. The PlanetScope imagery operated by Planet, a non-governmental privately owned commercial company, made accessible through the Planet Application Program Interface through the Education and Research Program license (Table 12). The PlanetScope imagery was the base for digitizing the field boundaries within the study area location due to the high resolution of the imagery at 5 m.

Table 12 Satellite Imagery Sources

Dataset	**Sentinel – 2A**	**PlanetScope**
Data Source	USGS Earth Explorer	Planet Application Program Interface
Data Operated by	European Space Agency	Planet
Type of Organization	Governmental	Private/Commercial
Data Access	Free/Open access	Research and Education License
Number of Scenes Obtained	16	68
Data Link URL	https://earthexplorer.usgs.gov/	https://www.planet.com/explorer/

4.2.3 Data Pre–Processing

The Sentinel–2A imagery was processed using ESA's SNAP software for atmospheric correction via the Sen2Cor Plugin. The software takes level – 1C Sentinel imagery metadata and individual bands to convert the imagery from Reflectance Values Top of Atmosphere (TOA) to Surface Reflectance Bottom of Atmosphere (BOA) correction required in order to run indices and perform image analyses (L. Congedo 2016). A Python code (Appendix A – Planet Python Code) processed the Planet imagery where it iterated through folders and subfolders to read each raster file with its corresponding metadata to obtain the surface reflectance product. The code extracted the single bands of Green, Red and Near Infrared and saved them individually from the original stacked tiff file. As a result, both Sentinel and Planet data were atmospherically corrected and ready for processing and analysis.

4.2.4 Data Processing

Indices used for analyzing crop health status are outlined in Table 13 and include: NDVI and SAVI (Candiago et al. 2015). NDVI is a normalized ratio of near-infrared and red bands that ranges between -1 and 1 where areas with green plants have values above 0 and the higher the value the more photosynthetic activity there is due to the energy absorption of plant canopies. SAVI values between -1 and 0.1 are most likely not vegetated. This index has the variable L in its equation which is related to the density of vegetation and is used as a canopy background adjustment factor based on soil brightness (Candiago et al. 2015). For this research, the estimated value of L was 0.5 as by the recommendation of A. R. Huete (1988) since the fields have an average vegetation densities.

After the correction of the Sentinel-2A scenes in SNAP, the needed bands (Near Infrared: Band 8, Red: Band 4 and Green: Band 3) were imported into Esri's ArcGIS Pro 2.2.2 for

processing vegetation indices. In order to increase efficiency, the various indices' formulas were integrated into a tool using model builder in Esri's ArcGIS Pro and the output raster datasets were saved into a specific geodatabase for data management purposes. Similarly, the Python code mosaicked Planet scenes of the same dates together then calculated the indices of interest and saved them in to a specified folder while renaming them using the format *DDMM*.

Table 13 Vegetation and Water Indices

Vegetation Index	**Formula**	**Reference**
Normalized Difference Vegetation Index	$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$	(Rouse et al. 1973)
Soil Adjusted Vegetation Index	$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red} + L} * (1+L)$	(A. R. Huete 1988)
Normalized Difference Water Index	$\frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}}$	(S.K. 1996)

4.2.5 Data Post–Processing

After the processing of all indices on the fields, "Zonal Statistics as Table" tool in ArcGIS Pro summarized the values obtained using the Python code for both Sentinel and Planet imagery with the fishnet over the fields as a grid. To maintain consistency in the outputs and to compare across both data sets, the grid size of the fishnet was fixed to 80 m by 80 m for the data summary. Using these grids, pixel values from individual bands along with indices' values were averaged within each grid square and added to the attribute table. The shapefile with the updated attribute table was the input for the "stats" function (Appendix A – Planet Python Code) that exported the data in an Excel spreadsheet format for easy access. The statistics function within the python code read the mean values and generated excel spreadsheets based on the resulting statistical tables for each variable per imagery dataset as needed.

4.2.6 Information from Grower

The Lebanese farmer supported the data analysis by providing information for the fields and the growing practices throughout the season (Table 14). The harvesting process involved manual extraction by digging up potato tubers while the crop canopy was still green. The green plant matter was removed and tossed into the field afterwards.

Table 14 2017 Growing season information

Planting Date	February 20th
Harvesting Date	Approximately June 25th
Potato Variety	Agria
Fertilizer/Pesticide Application Date	April 20th
Pesticide Re-application	Every 20 days
Overall Crop Yield	Between 3 ton to 5 ton per 1000m^2

Due to the geographical similarities between Idaho and Lebanon, introducing the results from the calculated indices over the growing season – particularly from the Planet data, we are able to apply the yield prediction model from Idaho (3.3.2) to predict the yield for the Lebanese potato fields. The major difference between the two study areas is the size of the fields. Idaho fields are much larger in area than those of Lebanon. The main reason for the small size of the Lebanon fields is that the majority of the fields are inherited from father to sons and as a result they get subdivided when passed from generation to another.

4.3 Results

4.3.1 Variation of Indices over Growing Season Sentinel-2 vs. Planet

The indices vary throughout the season in alignment with one another where as both NDVI and SAVI increase, NDWI decreases (Figure 22, Figure 23, Figure 24 & Figure 25). Despite the difference in resolution between the Planet and Sentinel data, they both show consistent results and it's hard to distinguish between the two datasets. The plots of the indices indicate an increase in index value corresponding to full maturity and row closure leading to high NDVI and SAVI values. The dips in the graphs on the other hand relate to the increase of temperature at different points throughout the growing season.

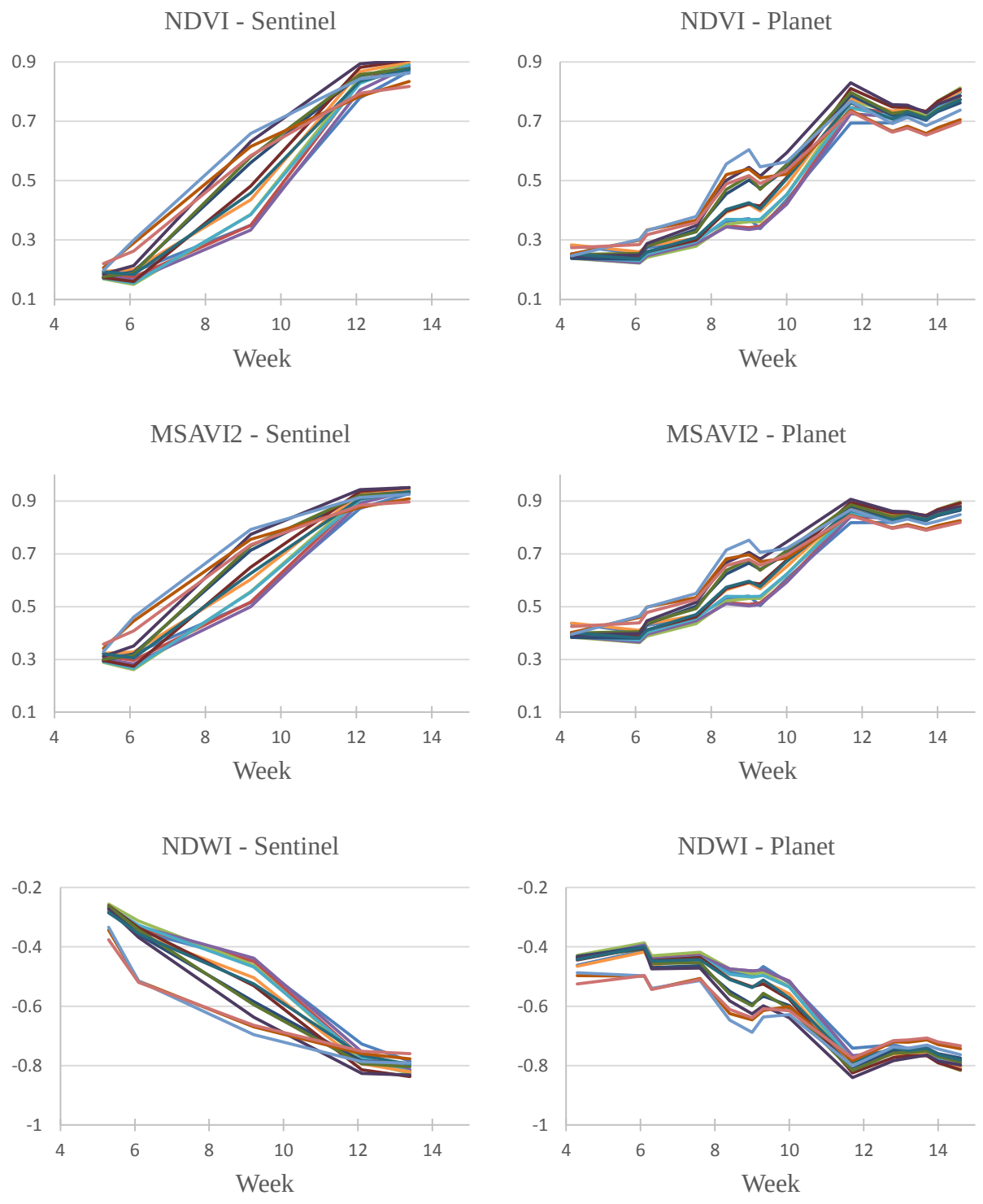

Figure 22 Variation of indices over the 2017 growing season

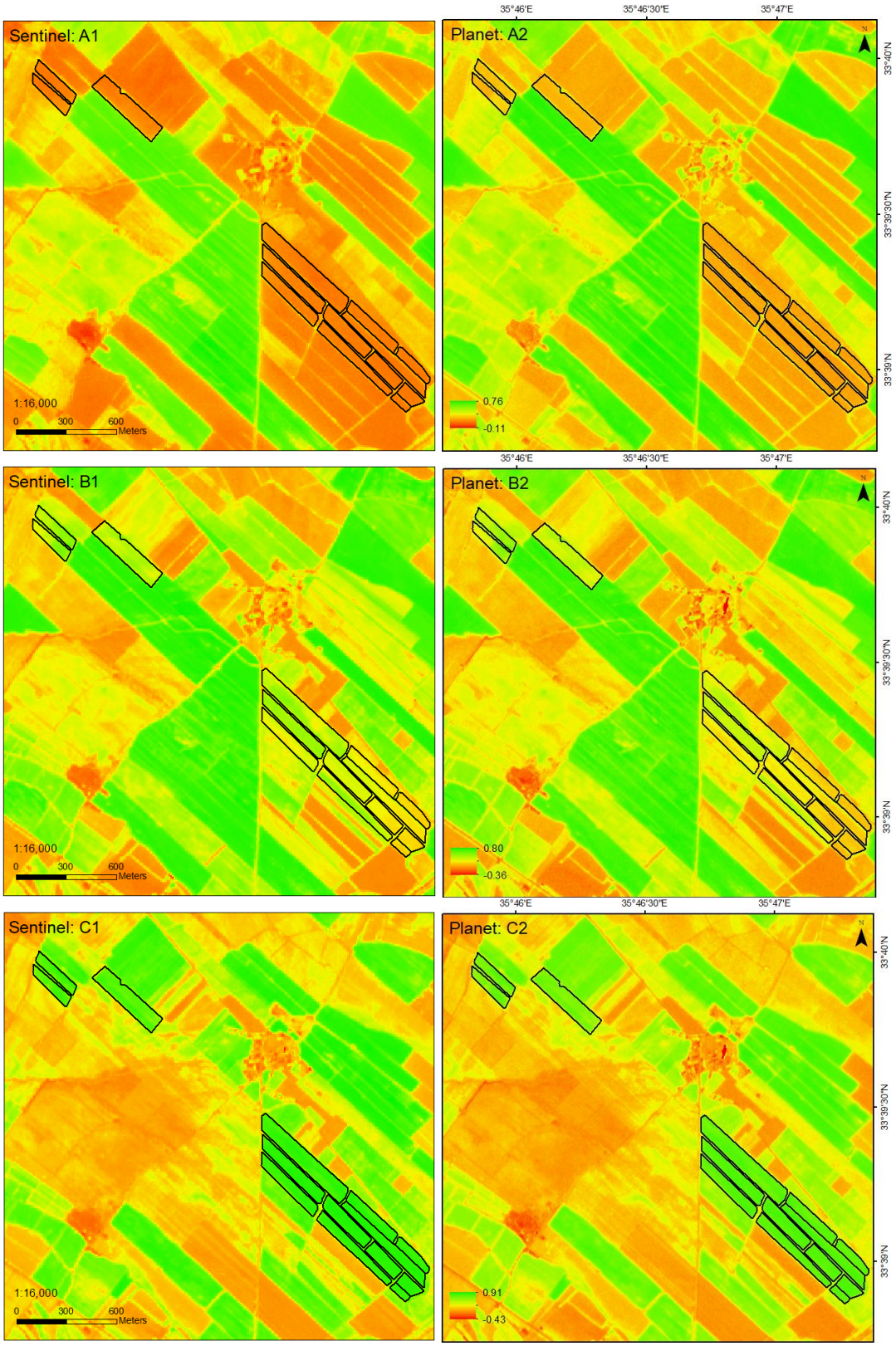

Figure 23 Variation of NDVI over the 2017 growing season Sentinel (A1: April 11, B1: May 11, C1: June10) vs. Planet (A2: April 17, B2; May 9, C2: June9)

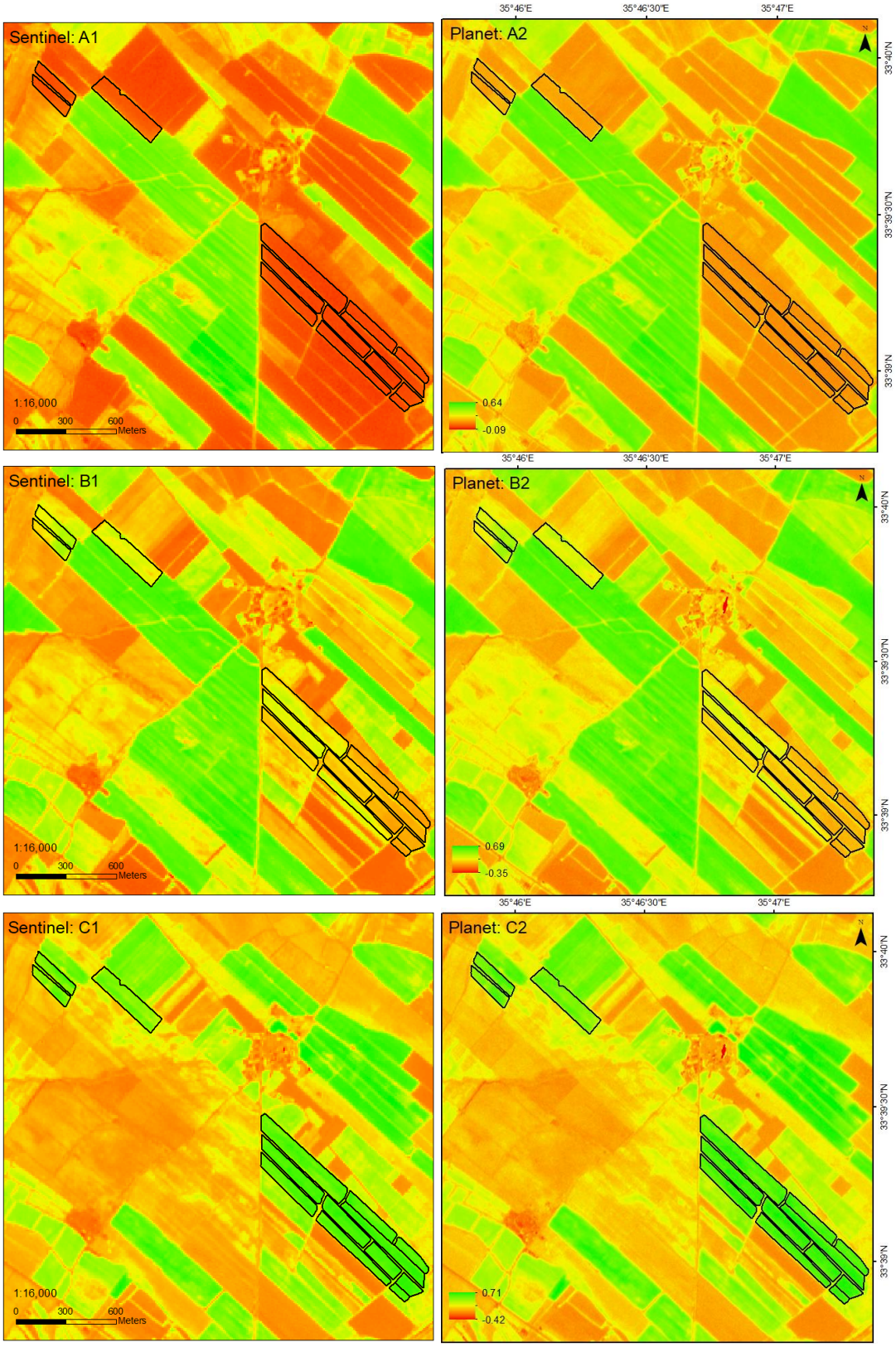

Figure 24 Variation of SAVI over the 2017 growing season Sentinel (A1: April 11, B1: May 11, C1: June10) vs. Planet (A2: April 17, B2; May 9, C2: June9)

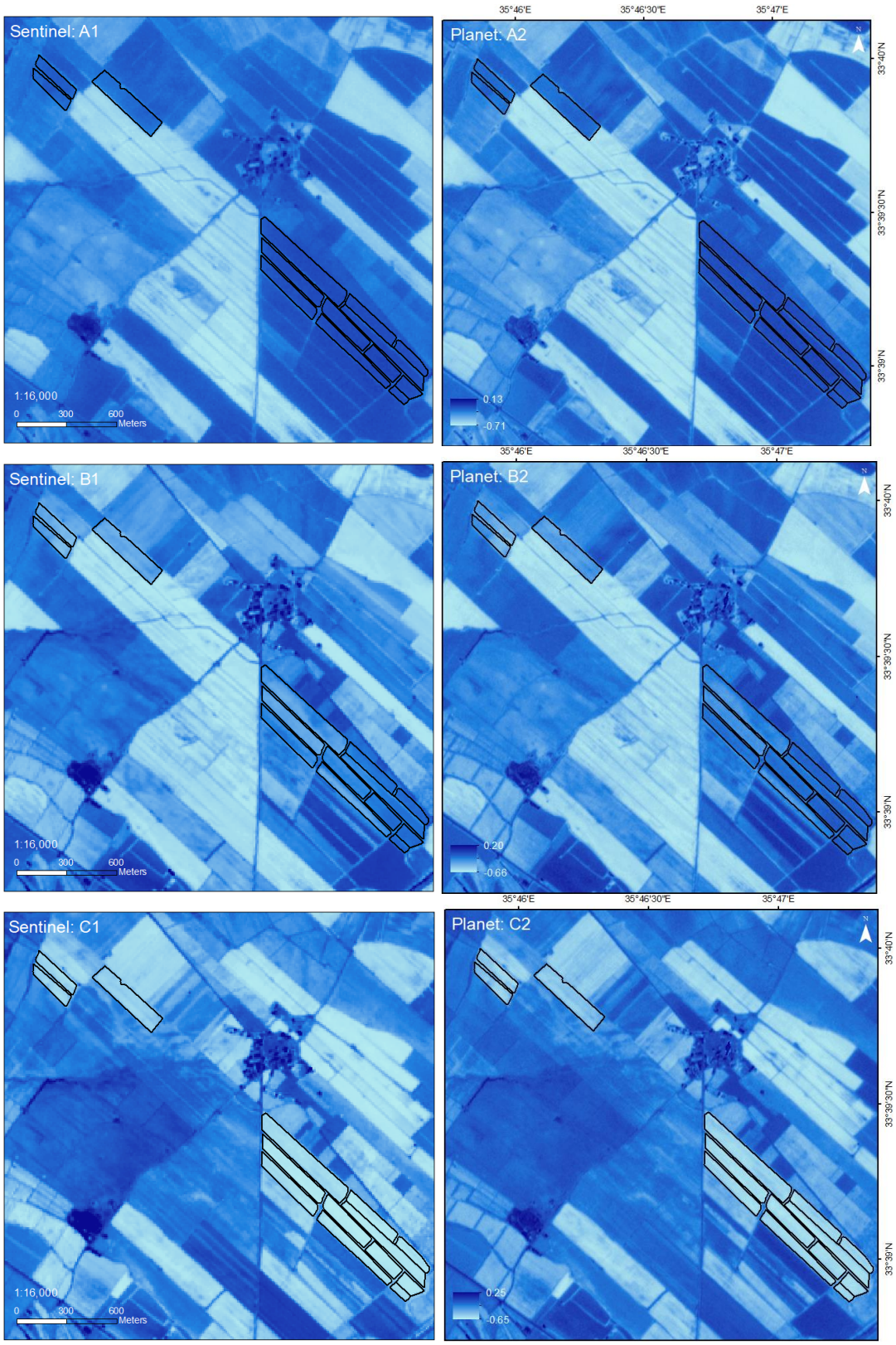

Figure 25 Variation of NDWI over the 2017 growing season Sentinel (A1: April 11, B1: May 11, C1: June10) vs. Planet (A2: April 17, B2; May 9, C2: June9)

4.3.2 Yield Forecasting

The regression model from the Idaho data (3.3.2.2) predicted the yield values for the Lebanon fields (Figure 26). Due to the difference in potato variety, the approach was to calculate yield using the three models (3.3.2.1), (3.3.2.2) and (3.3.2.3). Though the Idaho model used averaged values of the variables within a fishnet cell size of 80m x 80m, the fishnet used to predict yield in the Lebanon data was 32m x 32m. The reason for that choice is the fact that the Lebanese farmer gave an approximation of the yield to have had a range of 3 ton to 5 ton per 1,000 m^2. Hence, a 32m x 32m cell grid is the closest approximation to what the farmer provided.

Figure 26 Predicted yield values (A) and the corresponding slope values (B)

4.4 Discussion

Based on a visual interpretation, it is apparent that the fields can be subdivided into clusters with similar responses to the vegetation and water indices. Fields 1 through 5 showed consistency with each other, recording the highest NDVI and SAVI values and lowest water stress values. Those fields had similar predicted yield values and they correspond to the information from the farmer as he confirmed that they were the fields with higher yield values. This could be due to the geographic location of these fields as they are the closest to the Litani River, which could explain the higher indices values within these fields. As water is crucial for crop health and (3.3.2), the fields with higher proximity to the river showed better yield results. The farmer indicated that those fields have different soil type than the ones with lower yield values. In addition, the micro topography of the fields affected the yield values due to the variation in slope leading to uneven irrigation. The farmer has confirmed this as well and it is evident in the distribution of the predicted yield values within the fields. The farmer mentioned that there are three sprinklers every 100 meters and the yield values are clustered in a 100 meter neighborhood showing the potential positions of the installed sprinklers.

The NDVI plot from the Planet data (Figure 22) identifies the main dates for the potato growing season. The gradual increase in the NDVI values start from around week three which is when the leaves start developing and showing above the ground. There are two major peaks in the graph, the first peak corresponds to the row closure, meaning that the plants' leaves are fully developed closing any gaps between individual plants. This is the point in the season where the plants are at their healthiest state. This week is particularly crucial since as the leaves reach full development, the tubers start initiating and developing so the farmers use this indication to apply the first round of pesticides. According to the Lebanese farmer, after the first application during

week eight, they reapply pesticides almost every 20 days prior to harvest to ensure better yield and protect the crops from harmful insects or organisms. The second peak is evident at week twelve which is around the time of when the when the tubers are done filling up and represent full maturity of crops prior to harvest. Throughout the growing season, there are various dips in NDVI values due to increases in temperature (Figure 27) and corresponding increases in evapotranspiration rates resulting in a decrease in the NDVI values. One of the main issues the Lebanese farmers had to deal with in this growing season was the increase of temperatures above their average annual values. The three major dips in the growing season correspond to recorded temperatures of 24º C, 26º C and 29 C for May 12th, June 2nd and June 9th respectively. Due to the increase of temperatures during the critical week of tube filling (around week 10), the yield of the fields suffered and resulted in lower production than initially expected according the farmers.

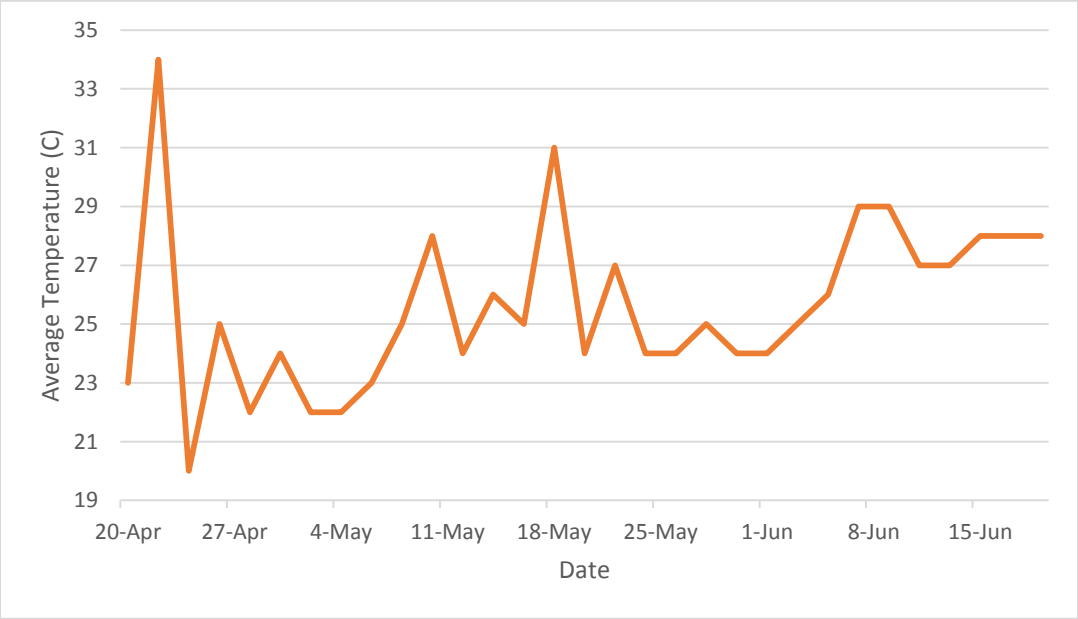

Figure 27 Average daily temperature 2017 growing season (°C)

Through the various indices processed for Sentinel-2 and the Planet data, satellite imagery demonstrated that it is a reliable source for analysis and for precision agriculture applications. All the indices showed compatibility with one another. The Sentinel-2 imagery depicted the rough main points that are critical during the growing season while PlanetScope delivered a more accurate representation due to the higher temporal and spatial resolution. Despite Sentinel showing the potential of open source data for crop monitoring and yield prediction, having a lower resolution data for small area fields leads to less accurate results. The reason being is that with lower spatial resolution, details are overlooked and responses from the crops get blurred and thus it becomes harder to depict important variation that occurs with fields. In addition, the lower temporal resolution of Sentinel with a return period of 5 – 10 days limits the data points that may be available especially if obscured by cloud coverage. Hence, with PlanetScope data and a 1 day return period, not only is the higher temporal resolution key to a denser dataset, but also the higher spatial resolution allows the delivery of higher detailed output.

This study highlights the potential of precision agriculture to support decision making for crop management practices throughout Lebanon. Though this research was limited to satellite imagery only, utilizing UAS data is another additional resource for crop management in Lebanon especially with smaller crop sizes. As important as it is to monitor crops' health status, it is equally vital to detect crop viruses and threats but with the current available satellite imagery resolution, that is not yet achievable. With UAS imagery, the resolution is capable of taking this work a step further. With the Lebanese farmers' openness to UAS technology, presenting this work to the Lebanese Ministry of Agriculture and outlining the potential of UAS, it would be more realistic to implement. With the support from the Ministry of Agriculture, proper arrangements with the Lebanese Army forces, one could obtain the needed permission and protection to fly UAS and

capture high-resolution aerial imagery. This would be combined with fieldwork using a spectrometer that allows the plotting of healthy versus sick plant signals to serve as a control to what is remotely observed. In addition, working closer with the farmers leads to more accurate and updated information regarding the fields, which was one of the major obstacles for this work due to poor communication with the farmer because of geographical barriers. With UAS data collection, a higher resolution elevation model generated from the data allows for better study of the micro topography of the fields that is directly related to the water intake of the plants.

In addition, this research demonstrated the potential of using satellite imagery to forecast yield. Though there is a difference between the Idaho and Lebanon fields regarding the potato varieties, offset in the planting dates and variation in microtopography, the preliminary result of the model aligned with the input from the farmer. The yield values themselves weren't accurate, which is expected due to the limitation of difference in variables between the two locations, however higher yield values showed up in the correct fields according to the Lebanese farmer. For further improvements, the model should take adjust to the different potato varieties and their corresponding growth cycle.

Chapter 5 Conclusion

Satellite imagery is a strong tool for monitoring crops and predicting yield. The biggest advantage of using the precision agriculture approach is the ability to take into consideration the variation within the field as opposed to the assumption that variables such as microtopography and soil types within the field are uniform. With each spectral band delivering specific information about the plants based on the reflectance signature, combining slope data with vegetation or water index values – particularly SAVI or NDVI allows the prediction of yield. NDVI provides information about the health status of the crops and can help with decision regarding adding nutrients and water to plants when needed based on the index value. Both vegetation and water indices provided important dates within the growing season and showed what week was the most critical for a better yield. The study areas in Idaho and Lebanon complemented one another; the Lebanon area had a denser dataset due to cloud cover being at minimum that gave a more detailed observation of crop variation over the growing season while Idaho has accurate yield data that allowed the development of a yield forecasting model that is applicable to Lebanon.

As with any scientific research, this work faced a few issues that put a limit on the processing and analysis done. One hurdle this work faced was the remote location of the Lebanon study area. Doing this work over a study area in Lebanon while being physically present in Idaho meant that we were unable to perform on-site validation for our observations. With the satellite imagery, though Sentinel-2 has a 5 day revisit period, the unavailability of Sentinel-2B meant that we were limited to Sentinel-2A with a 10 revisit period cutting the potential number of available scenes at least by half depending on cloud coverage. As for the Planet data, though we were able to obtain almost daily imagery over Lebanon, the Idaho study area had a much higher cloud coverage and drastically limited the number of available scenes for processing. This resulted in a

gap between the Lebanon and Idaho datasets in processing. Another limitation of having a study area in Lebanon and another in Idaho was the difference in potato varieties between the two. This resulted in a conflict with the yield prediction model especially that there was a lack of accurate yield data for Lebanon that would have allowed fine tuning the Idaho model to Lebanon for improved results.

This work demonstrated the ability of satellite imagery to assist growers with decision making as we were able to depict critical dates in the growing season that ensures a good yield based on the specific variety and planting date. The Sentinel imagery is free and easily accessible to the public, and utilizing an open source processing software such as QGIS, reduces the cost on the farmer's behalf for obtaining needed information. This data, and with the Sentinel-2B now available, has a 5 day revisit period which allows a detailed observation of indices' variation over the growing season especially with it having a 10 m resolution and 13 spectral bands giving a wider range for spectral analyses. Meanwhile, with Planet, it provides a higher temporal and spectral resolution for a finer detail level of processing. However, this data has only four spectral bands thus limiting the analyses. In addition, with a spatial resolution of 3m and a daily revisit time, it is not a free resource adding to the financial burden on the farmers especially those with smaller fields particularly in Lebanon. For precision agriculture, it is vital for the satellite imagery to combine high spatial and temporal resolution. The high spatial resolution is important to detect change at a smaller scale while the increased temporal resolution ensures that important dates during the growing season aren't overlooked.

In summary, this thesis research concluded the following major points:

- Satellite imagery is showing promise to help farmers identify problems within the field early on and indicate critical time points in the growing season that are indicators for yield prediction.
- In addition to how various vegetation indices reveal different information about the crops they allow the prediction of potato crop yield values.
- Slope is a significant variable for yield forecasting.
- Using previous yield data and utilizing high resolution satellite imagery, a yield prediction model is developed and can help the farmers with management decisions.
- To better predict yield, it is crucial to take into consideration the potato variety and its corresponding significant week.

Based on the yield information from Idaho, the current prediction models will serve as the building blocks for future models that are tailored to the potato varieties in Lebanon. Working closely with the Lebanese growers and obtaining numerical yield data will better validate the models and the results at hand. Moreover, the satellite imagery analyses indicated crucial times in the growing season for improving yield such as when to increase irrigation. In order for this information to be fully useful, it is essential to deliver it to the farmers for them to benefit from. The Lebanese Ministry of Agriculture is very involved with the farmers and constantly attempting to improve crop production, thus, presenting this work to them will help better spread the information on a larger scale and keep the farmers up to date with what precision agriculture has to offer them. Discussing with the Ministry of Agriculture about the importance of each critical week for specific potato varieties and its importance for yield prediction as well as best approaches to deliver the quantitative information to the growers in a qualitative form so that they could benefit

from it. In addition to that, using the broad workflow of this research work it could be applied to other crops based on the farmers' interests and needs. The Sentinel-2A along with Sentinel-2B provides a high temporal resolution needed for the yield prediction models and this data is freely available through ESA. Using the Sentinel data along with QGIS and SNAP for image processing – which both are open source, the cost of would be reduced and as a result help promote the project with the Lebanese officials and farmers. Another approach is developing a web application that the farmers can access and retrieve updated information regarding their fields by the use of their smartphones especially that almost everyone owns a smartphone and has internet access. This application will be tailored to the specific needs and interest of the farmer and their fields for most effectiveness.

Alongside satellite imagery, UAS platforms and sensors technologies are rapidly advancing, providing another valuable tool for precision agriculture applications. With the Lebanese growers expressing their openness to using this technology, using UAS is the next step for this work. Not only will it provide higher resolution imagery and elevation models, it is a faster more flexible approach without having to worry about cloud coverage limiting the data at hand. Although the political region around Lebanon is not currently stable, working with the Lebanese military will provide the needed security blanket to fly UAS missions for data collection. Utilizing UAS technology will improve the yield prediction model through the improved resolution and will allow the detection of anomalies within the fields that can indicate virus infected plants or irrigation issues. This will provide the farmer with more detailed information about what is going on in the field and help detect problems early and mitigate them.

References

Aboelghar, M, A.-R. Ali, and S Arafat. 2014. “Spectral Wheat Yield Prediction Modeling Using SPOT Satellite Imagery and Leaf Area Index.” *Arabian Journal of Geosciences* 7(2): 465–74. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84895074388&partnerID=40&md5=b2380777f682c3b4890ffd429333d2bf.

Abou-Jawdah, Yusuf, Hana Sobh, and Adib Saad. 2001. “Incidence of Potato Virus Diseases and Their Significance for a Seed Certification Program in Lebanon.” *Phytopathologia Mediterranea* 40(2): 113–18.

Actualitix. 2016. “Lebanon Potato Production.” http://en.actualitix.com/country/lbn/lebanon-potato-production.php.

Aditya Shastry, H.A. Sanjay and E. Bhanusree. 2017. “Prediction of Crop Yield Using Regression Techniques.” *International Journal of Soft Computing* 12(2): 96–102.

AgriMet. 2017. “Reclamation Managing Water in the West.” https://www.usbr.gov/pn/agrimet/agrimetmap/terida.html.

Al-Gaadi, Khalid A. et al. 2016. “Prediction of Potato Crop Yield Using Precision Agriculture Techniques.” *PLoS ONE* 11(9): 1–16. http://dx.doi.org/10.1371/journal.pone.0162219.

Alganci, Ugur, Mutlu Ozdogan, Elif Sertel, and Cankut Ormeci. 2014. “Estimating Maize and Cotton Yield in Southeastern Turkey with Integrated Use of Satellite Images, Meteorological Data and Digital Photographs.” *Field Crops Research* 157(January): 8–19. http://dx.doi.org/10.1016/j.fcr.2013.12.006.

An-Nahar Newspaper. 2015. “دراسات المزارعين معاناة سلام إلى نقل شهيب: المراحل أخطر في نمر.” *An-Nahar Newspaper*. http://newspaper.annahar.com/article/248665-المزارعين-معاناة-سلام-إلى-نقل-شهيب-دراسات-المراحل-أخطر-في-نمر.

Arnold. 2010. “The Electromagnetic Spectrum.” : 14–22. http://hosting.soonet.ca/eliris/remotesensing/bl130lec3.html.

Atzberger, Clement. 2013. “Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs.” *Remote Sensing* 5(2): 949–81.

Burke, Marshall, and David B. Lobell. 2017. “Satellite-Based Assessment of Yield Variation and Its Determinants in Smallholder African Systems.” *Proceedings of the National Academy of Sciences* 114(9): 2189–94. http://www.pnas.org/lookup/doi/10.1073/pnas.1616919114.

Campbell, James B., and Randolph H. Wynne. 2011. *Introduction to Remote Sensing*.

Candiago, Sebastian et al. 2015. “Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images.” *Remote Sensing* 7: 4026–47.

Choueiri, Elia et al. 2017. “Surveys of Potato-Growing Areas and Surface Water in Lebanon for Potato Brown and Ring Rot Pathogens.” *Phytopathologia Mediterranea* 56(1): 87–97.

Crum, Shannon. “Typical Spectral Response Characteristics of Green Vegetation.” www.rsunt.geo.ucsb.edu/rscc/vol2/lec2/2_2.html#21 (March 1, 2017).

DAR-IAURIF. 2005. NPMPLT *National Physical Master Plan of the Lebanese Territory*.

Darwish, T, T Atallah, N Khatib, and F Karam. 2004. “Fertigation and Conventional Potassium Application to Field Grown Potato in Lebanon: Perspective to Enhance Efficiency.” *IPI Regional workshop on potassium and fertigatoin development in West Asia and North Africa*: 1–13.

Dash, Jadunandan, and Booker O. Ogutu. 2016. “Recent Advances in Space-Born Optical Remote Sensing Systems for Monitoring Global Terrestrial Ecosystems.” *Progress in Physical Geography* 40(2): 322–51. http://journals.sagepub.com/doi/abs/10.1177/0309133316639403.

Dempewolf, Jan et al. 2014. “Wheat Yield Forecasting for Punjab Province from Vegetation Index Time Series and Historic Crop Statistics.” *Remote Sensing* 6: 9653–75.

FAO. 2014. *The Impact of the Syria Crisis on Agricultre, Food Security and Livelihoods in Lebanon*.

———. 2017. “Food and Agriculture Driving Action Across the 2030 Agenda for Sustainable Development.” : 40. http://www.fao.org/3/a-i7454e.pdf.

FITA. 2008. “Lebanon, Main Branches of Industry.” *The Federation of International Trade Associations*. http://www.fita.org/countries/lebanon.html.

Gale, Thomson. 2007. “Lebanon.” *Woldmark Encyclopedia of Nations*. http://www.encyclopedia.com/places/asia/lebanese-political-geography/lebanon.

Gao, Bo Cai. 1996. “NDWI - A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space.” *Remote Sensing of Environment* 58(3): 257–66.

El Gazzar, Hisham. 2015. “Potatoes and Leafy Green Vegetables: Value Chain Analysis.” *International Labour Organization*: 13–14.

Geipel, Jakob, Johanna Link, and Wilhelm Claupein. 2014. “Combined Spectral and Spatial Modeling of Corn Yield Based on Aerial Images and Crop Surface Models Acquired with an Unmanned Aircraft System.” *Remote Sensing* 6(11): 10335–55.

Gitelson, Anatoly A., Yoram J. Kaufman, and Mark N. Merzlyak. 1996. “Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS.” *Remote Sensing*.

Govender, M, K Chetty, and H Bulcock. 2007. “A Review of Hyperspectral Remote Sensing and Its Application in Vegetation and Water Resource Studies.” *waters Water SA* 33(2): 145–

51.

Grow, Ray. 1993. *Soil Survey of Fremont County, Idaho, Western Part*.

Guo, T., T. Kujirai, and T. Watanabe. 2012. "Mapping Crop Status from an Unmanned Aerial Vehicle for Precision Agriculture Applications." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXIX-B1: 485–90. http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXIX-B1/485/2012/.

Hassan, M. a., T. H. Pham, H. Shi, and J. Zheng. 2013. "Nematodes Threats to Global Food Security." *Acta Agriculturae Scandinavica, Section B - Soil & Plant Science* 63(5): 420–25. http://www.tandfonline.com/doi/abs/10.1080/09064710.2013.794858.

Hatoum, Talal. 2005a. "*مديريّة الدراسات والتنسيق*: 2005." لبنان في الزراعة 23. http://www.agriculture.gov.lb/Arabic/DataAndAgriStatistics/AgriInLebanon/Documents/Enq_2005.pdf (January 11, 2017).

———. 2005b. *لبنان في الزراعة 2005*. http://www.agriculture.gov.lb/Arabic/DataAndAgriStatistics/AgriInLebanon/Documents/Enq_2005.pdf.

Huete, A. et al. 2002. "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices." *Remote Sensing of Environment* 83(1–2): 195–213.

Huete, A. R. 1988. "A Soil-Adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment* 25(3): 295–309.

Hunt, E. Raymond, Craig S T Daughtry, Steven B. Mirsky, and W. Dean Hively. 2014. "Remote Sensing with Simulated Unmanned Aircraft Imagery for Precision Agriculture Applications." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(11): 4566–71.

Ibrahim, Said K., Halim Abi Saad, and Nisrin Moussa. 2004. "Potato Cyst Nematodes Globodera Spp. in Lebanon - Occurence and Distribution." *Lebanese Science Journal* 5(1): 25–36.

Johannsen, Chris J. 2010. "Advances in Remote Sensing for Agriculture." *Information Systems* 2(10): 2369–87.

Karaa, K, and F Karam. 2000. "Recent Trends towards Developing a Sustainable Irrigated Agriculture in the Bekaa Valley." *CIHEAM - Mediterranean Agronomic Institute of Bari* 31: 59–84.

Khot, Lav R. et al. 2016. "UAS Imaging-Based Decision Tools for Arid Winter Wheat and Irrigated Potato Production Management." *International Journal of Remote Sensing* 37(1): 125–37.

Knoema. 2015. "Lebanon - Potatoes." *The Agricultural Production*. https://knoema.com/FAOPRDSC2016R/production-statistics-crops-crops-processed.

Ktheien. 2008. “Lebanon.” *World Potato Atlas*. https://research.cip.cgiar.org/confluence/display/wpa/Lebanon.

Kussul, Nataliia et al. 2017. “SENTINEL-2 FOR AGRICULTURE NATIONAL DEMONSTRATION IN UKRAINE : RESULTS AND FURTHER STEPS.” In *IEEE*, Fort Worth: IEEE, 5842–45.

Kuwata, Kentaro, and Ryosuke Shibasaki. 2015. “Estimating Crop Yields with Deep Learning and Remotely Sensed Data.” In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, , 858–61. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7325900.

Lewin, Paul et al. 2011. “The Role of Agricultural Processing in Idaho ’ S Economy : Status and Potential.”

Liu, Zhan Yu, Jing Feng Huang, Xin Hong Wu, and Yong Ping Dong. 2007. “Comparison of Vegetation Indices and Red-Edge Parameters for Estimating Grassland Cover from Canopy Reflectance Data.” *Journal of Integrative Plant Biology* 49(3): 299–306.

Lobell, David B. 2013. “The Use of Satellite Data for Crop Yield Gap Analysis.” *Field Crops Research* 143: 56–64. http://dx.doi.org/10.1016/j.fcr.2012.08.008.

Mulla, David J. 2013. “Twenty Five Years of Remote Sensing in Precision Agriculture: Key Advances and Remaining Knowledge Gaps.” *Biosystems Engineering* 114(4): 358–71. http://dx.doi.org/10.1016/j.biosystemseng.2012.08.009.

Nahry, A. H El, R. R. Ali, and A. A El Baroudy. 2011. “An Approach for Precision Farming under Pivot Irrigation System Using Remote Sensing and GIS Techniques.” *Agricultural Water Management* 98(4): 517–31. http://dx.doi.org/10.1016/j.agwat.2010.09.012.

Prasad, Anup K., Lim Chai, Ramesh P. Singh, and Menas Kafatos. 2006. “Crop Yield Estimation Model for Iowa Using Remote Sensing and Surface Parameters.” *International Journal of Applied Earth Observation and Geoinformation* 8(1): 26–33.

Qi, J et al. 1994. “A Modified Soil Adjusted Vegetation Index.” 126: 119–26.

Quere, Bernard. 2009. “Potato Sector in Countries of Africa and the Middle East.”

Rad, Ciprian-Radu, Olimpiu Hancu, Ioana-Alexandra Takacs, and Gheorghe Olteanu. 2015. “Smart Monitoring of Potato Crop: A Cyber-Physical System Architecture Model in the Field of Precision Agriculture.” *Agriculture and Agricultural Science Procedia* 6: 73–79. http://linkinghub.elsevier.com/retrieve/pii/S2210784315001746.

Rembold, Felix, Clement Atzberger, Igor Savin, and Oscar Rojas. 2013. “Using Low Resolution Satellite Imagery for Yield Prediction and Yield Anomaly Detection.” *Remote Sensing* 5(4): 1704–33.

Rouse, J. W., R. H. Hass, J.A. Schell, and D.W. Deering. 1973. “Monitoring Vegetation Systems

in the Great Plains with ERTS." *Third Earth Resources Technology Satellite (ERTS) symposium* 1: 309–17.

Rud, Ronit et al. 2014. "Crop Water Stress Index Derived from Multi-Year Ground and Aerial Thermal Images as an Indicator of Potato Water Status." *Precision Agriculture* 15(3): 273–89.

S.K., McFeeters. 1996. "The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features." *International Journal of Remote Sensing* 17(7): 1425–32.

Satir, O.a, and S.b Berberoglu. 2016. "Crop Yield Prediction under Soil Salinity Using Satellite Derived Vegetation Indices." *Field Crops Research* 192: 134–43. http://www.agriculturejournals.cz/publicFiles/161769.pdf.

Sellam, V, and E Poovammal. 2016. "Prediction of Crop Yield Using Regression Analysis." *Indian Journal of Science and Technology* 9(October): 1–5.

Sitienei, Betty, Shem Juma, and Everline Opere. 2017. "On the Use of Regression Models to Predict Tea Crop Yield Responses to Climate Change: A Case of Nandi East, Sub-County of Nandi County, Kenya." *Climate* 5(3): 54. http://www.mdpi.com/2225-1154/5/3/54.

Sivarajan, Saravanan. 2011. "Estimating Yield of Irrigated Potatoes Using Aerial and Satellite Remote Sensing." Utah State University. http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=2046&context=etd.

Skakun, Sergii et al. 2016. "Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(8): 3712–19.

Song, Renzhong et al. 2016. "Evaluation of Landsat 8 Time Series Image Stacks for Predicting Yield and Yield Components of Winter Wheat." *IEEE*: 6300–6303.

Steinger, T., H. Gilliand, and T. Hebeisen. 2014. "Epidemiological Analysis of Risk Factors for the Spread of Potato Viruses in Switzerland." *Annals of Applied Biology* 164(2): 200–207.

Tadesse, Tsegaye et al. 2015. "Evaluating a Satellite-Based Seasonal Evapotranspiration Product and Identifying Its Relationship with Other Satellite-Derived Products and Crop Yield: A Case Study for Ethiopia." *International Journal of Applied Earth Observation and Geoinformation* 40: 39–54. http://dx.doi.org/10.1016/j.jag.2015.03.006.

Tiefelsdorf, Michael, and Barry Boots. 1997. "A Note on the Extremities of Local Moran's I and Their Impact on Global Moran's I." *Geographical Analysis* 29(3): 248–57.

Turner, W. et al. 2015. "Free and Open-Access Satellite Data Are Key to Biodiversity Conservation." *Biological Conservation* 182: 173–76. http://dx.doi.org/10.1016/j.biocon.2014.11.048.

United Nations Food and Agriculture Organization. 2016. *Lebanon Country Programming Framework 2016-2019*. http://www.fao.org/3/a-i5605e.pdf.

———. 2017. *The Future of Food and Agriculture – Trends and Challenges.* Rome.

Wojtowicz, Marek, Andrzej Wojtowics, and Jan Piekarczyk. 2016. "Application of Remote Sensing Methods in Agriculture." *Communication in Biometry and Crop Science* 11(1): 31–50.

Yang, Chenghai et al. 2013. "Using High-Resolution Airborne and Satellite Imagery to Assess Crop Growth and Yield Variability for Precision Agriculture." *Proceedings of the IEEE* 101(3): 582–92.

Yardımcı, N, H Çulal Kılıç, and Y Demir. 2015. "Detection of PVY, PVX, PVS, PVA, and PLRV on Different Potato Varieties in Turkey Using DAS-ELISA." *Journal of Agricultural Science and Technology* 17: 757–64.

You, Jiaxuan et al. 2017. "Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data." *Association for the Advancement of Articial Intelligence*.

Zhang, L, L Lei, and Dongmei Yan. 2010. "Comparison of Two Regression Models for Predicting Crop Yield." *IEEE*: 1521–24. http://ieeexplore.ieee.org.bdigital.udistrital.edu.co:8080/stamp/stamp.jsp?tp=&arnumber=5652764.

Zhang, Qin. 2016. *Precision Agriculture Technology for Crop Farming*. Prosser, Washington: Taylor & Francis Group.

"المراحل أخطر في نمر: درباس المزارعين معاناة سلام إلى نقل شهيب." 2015. *An-Nahar Newspaper*. http://newspaper.annahar.com/article/248665-درباس-المزارعين-معاناة-سلام-إلى-نقل-شهيب- (January 11, 2017).

Appendix A – Planet Python Code

```
#Programming for MS GIS Thesis Work
#By ABOU ALI, Hanan
#---------------------------------------------

#Purpose: read in PlanetScope satellite imagery, run various vegetation indices
#and compute zonal statistics based on areas of interest
#-----------------------------------------------------------------------------------
#-----------------------------------------------------------------------------------


#function: Stats
#arguments: none
#purpose: Calculate Zonal Statistics & Generate Shapefile & Excel Sheet
#-----------------------------------------------------------------------------------

def Stats():

    #import needed modules
    import geopandas as gpd
    import pandas as pd
    import glob
    import os
    from rasterstats import zonal_stats

    #set directories
    rsterDir=r'C:\Students\Hanan\Thesis_work\Data\Processing\Lebanon\LB_Planet\LB_Rasters'
    mydir=r'C:\Students\Hanan\Thesis_work\Data\Processing\Lebanon\LB_Planet\LB_Rasters'
    shpDir=r'C:\Students\Hanan\Thesis_work\Data\Processing\Lebanon\LB_Planet\shp'
    procDir=r'C:\Students\Hanan\Thesis_work\Data\Processing\Idaho\ID_Planet\GridStats'

    #change home directory
    os.chdir(shpDir)

    #read in fishnet shapefile
    for shp in glob.glob('*.shp'):
        print shp
        shpNmeList=shp.split('.')
        shpNme=shpNmeList[0]
        rasterList=[]
        gdf=gpd.read_file(shp)

        #create attribute table headers
        gdf['FID']=None
        counterShp=0
```

```
while counterShp<len(gdf):
    gdf.loc[counterShp,'FID']=shpNme+'_'+str(counterShp)
    counterShp+=1

#change directory to rasters
os.chdir(rsterDir)

#calculate and extract mean pixel values
stat= 'mean'

#populate attribute table with values
cols=[]
cols.append('FID')
#read in rasters & add to list
for rster in rasterList:
    rster=rster.split('.')
    rster=rster[0]
    cols.append(rster)

#define data frame for new shapefile
df=pd.DataFrame(columns=cols)
counter=0

#reading file
while counter<len(gdf):
    df.loc[counter,'FID']=gdf.loc[counter,'FID']
    counter+=1
for rster in rasterList:
    os.chdir(rsterDir)
    #calculate average pixel values
    stats=zonal_stats(gdf,rster,stat)
    #split name of raster variable to rename in attribute
    rster=rster.split('.')
    rster=rster[0]

    #populate new shapefile with mean values
    counter1=0
    while counter1<len(stats):
        stat=stats[counter1]
        fld=df.loc[counter1,'FID']
        counter2=0
        while counter2<len(df):
            if df.loc[counter2,'FID']==fld:
                df.loc[counter2,rster]=stat[stat]
            counter2+=1
        counter1+=1
```

```
        os.chdir(mydir)

        #export data as Excel spreadsheet
        df.to_csv(shpNme+'_'+statVar+'.csv')
        if statVar=='mean':
            for col in cols:
                if col!='FID':
                    df[col]=df[col].astype(float)

            gdfFinal=gdf.join(df.set_index('FID'),on='FID')

            #save final shapefile with Mean pixel values
            gdfFinal.to_file(shpNme+'_stats.shp',driver='ESRI Shapefile')

        os.chdir(shpDir)

#function: Indices
#arguments: rename, Green, Red & NIR (date of scene & corrected bands)
#purpose: Calculate Vegetation and Water Indices
#-----------------------------------------------------------------------------------------

def Indices(rename,Green,Red,NIR):

    import arcpy
    from arcpy import env

    arcpy.env.workspace = entry

    NDVI = ((NIR-Red)/(NIR+Red))
    NDVI.save("NDVI_"+rename+".tif")
    print("NDVI calculated")


    GNDVI = ((NIR-Green)/(NIR+Green))
    GNDVI.save("GNDVI_"+rename+".tif")
    print("GNDVI calculated")


    SAVI = (((NIR-Red)/(NIR+Red+0.5))*1.5)
    SAVI.save("SAVI_"+rename+".tif")
    print("SAVI calculated")


    MSAVI2 = ((2*NIR)+1-SquareRoot(Square((2*NIR)+1)-8*(NIR-Red)))/2
    MSAVI2.save("MSAVI_"+rename+".tif")
    print("MSAVI2 calculated")
```

```
    NDWI = ((Green-NIR)/(Green+NIR))
    NDWI.save("NDWI_"+rename+".tif")
    print("NDWI calculated")

    print("---------------------------")
    print("---------------------------")

    env.workspace = entry

    del rename
    del Green
    del Red
    del NIR


def main():

    import os
    import arcpy
    import re

    from arcpy import env
    from arcpy.sa import *

    #initial directory with all the subfolders and raw satellite imagery
    mydir = r"C:\Students\Hanan\Thesis_work\Data\4Band\Lebanon"
    #directory to store the individual corrected bands and indices
    rasters = r"C:\Students\Hanan\Thesis_work\Data\Processing\Lebanon\LB_Planet\LB_Rasters"

    env.workspace =  mydir
    env.overwriteOutput = False

    #-----------------------------------------------------------------------------------

    #define a list to store rasters (.tif) files
    P_4Band = []

    #define a list to store metadata (.xml) files
    P_Metadata = []

    #iterate through subfolders to find rasters & metadata files and add them to respective list
    for root,dirs,files in os.walk(mydir):
        for name in files:
            if name.endswith("MS.tif"):
```

```
            mypath = root+"\\"+name
            P_4Band.append(mypath)

            #-------testing code progress-----------
            print(P_4Band)
            print("---------------------------")

        elif name.endswith("metadata.xml"):
            mypath = root+"\\"+name
            P_Metadata.append(mypath)

            #-------testing code progress-----------
            print(P_Metadata)
            print("---------------------------")

#-------------------------------end of file iterating----------------------------------------
#-----------------------------------------------------------------------------------------------

#progress report purposes and ensure name match between raster & metadata file
print(P_4Band)
print("---------------------------")

print(P_Metadata)
print("---------------------------")

#list element index counter
j=0
temp1 = 1

#reading through xml files
for entry in P_Metadata:

    #read xml file & extract correction coefficients

    #initiate a variable to read lines from metadata file
    #coefficient locations based on PlanetScope product descriptions
    #open up the metadata file for reading
    doc = open(entry,"r")

    i = 0
    while i<208:

        #read line by line
        var = doc.readline()

        #correction coefficient of band 2 (Green)
```

```
    if i==183:
        temp = var
        #clean out unwanted stuff & extract only the coefficient value
        a_2=re.sub("<ps:reflectanceCoefficient>", "", temp)
        a_2=re.sub("</ps:reflectanceCoefficient>","",a_2)
        #-------testing code progress-----------
        print (i)
        print (a_2)

    #correction coefficient of band 3 (Red)
    elif i==192:
        temp = var
        #clean out unwanted stuff & extract only the coefficient value
        a_3=re.sub("<ps:reflectanceCoefficient>", "", temp)
        a_3=re.sub("</ps:reflectanceCoefficient>","",a_3)
        #-------testing code progress-----------
        print (i)
        print (a_3)

    #correction coefficient of band 4 (NIR)
    elif i==201:
        temp = var
        #clean out unwanted stuff & extract only the coefficient value
        a_4=re.sub("<ps:reflectanceCoefficient>", "", temp)
        a_4=re.sub("</ps:reflectanceCoefficient>","",a_4)
        #-------testing code progress-----------
        print (i)
        print (a_4)

    i+=1

print("---------------------------")

print(j)

print("---------------------------")
print("---------------------------")

#-------------------------------end of coeffcient reading-------------------------------------
#-----------------------------------------------------------------------------------------------

#to see which file its at
print(entry)

#rename date using format _DDMM
rename = entry[-43:-39]
```

```
a = rename[2:]
b = rename[:2]

rename = a+b

print(rename)

#pulll out the image file that corresponds to the xml file
x = P_4Band[j]

#to verify that it has the matching set
print(x)

arcpy.env.workspace = entry

#first run
if j==0:

    #do atmoshperic correction & indices' calculations here

    #extract & save Green band from raster ensuring it has the full path
    Green = Raster(os.path.join(x,"band_2"))
    Green.save("B02_"+rename+"_old.tif")
    #perform atmospheric correction for Green band & save
    Green_correct = Green*float(a_2)
    env.workspace = rasters
    Green_correct.save("B02_"+rename+".tif")
    #-------testing code progress-----------
    print("Green band corrected & saved")
    env.workspace = entry

    #extract & save Red band from raster ensuring it has the full path
    Red = Raster(os.path.join(x,"band_3"))
    Red.save("B03_"+rename+"_old.tif")
    #perform atmospheric correction for Red band & save
    Red_correct = Red*float(a_3)
    env.workspace = rasters
    Red_correct.save("B03_"+rename+".tif")
    #-------testing code progress-----------
    print("Red band corrected & saved")
    env.workspace = entry

    #extract & save NIR band from raster ensuring it has the full path
    NIR = Raster(os.path.join(x,"band_4"))
    NIR.save("B04_"+rename+"_old.tif")
```

```
    #perform atmospheric correction for NIR band & save
    NIR_correct = NIR*float(a_4)
    env.workspace = rasters
    NIR_correct.save("B04_"+rename+".tif")
    #-------testing code progress-----------
    print("NIR band corrected & saved")
    print("---------------------------")

    #------------------------------end of band correction----------------------------------------
    #-----------------------------------------------------------------------------------------------

    #assign corrected bands to variables for indicies' calculation
    Green = Green_correct
    Red = Red_correct
    NIR = NIR_correct

    #call Indices function
    Indices(rename,Green,Red,NIR)

    env.workspace = entry

#second run onward
if j!=0:

        #read the date of the entry & the one before it to see if same date but different scene
        c1 = P_Metadata[j-1]
        c2 = P_Metadata[j]

        rename1 = c1[-43:-39]
        rename2 = c2[-43:-39]

        a1 = rename1[2:]
        a2 = rename2[2:]

        b1 = rename1[:2]
        b2 = rename2[:2]

        rename1 = a1+b1
        rename2 = a2+b2

        if rename2 != rename1:

            Green = Raster(os.path.join(x,"band_2"))
            Green.save("B02_"+rename+"_old.tif")
            Green_correct = Green*float(a_2)
            env.workspace = rasters
```

```
            Green_correct.save("B02_"+rename+".tif")
            print("Green band corrected & saved")
            env.workspace = entry


            Red = Raster(os.path.join(x,"band_3"))
            Red.save("B03_"+rename+"_old.tif")
            Red_correct = Red*float(a_3)
            env.workspace = rasters
            Red_correct.save("B03_"+rename+".tif")
            print("Red band corrected & saved")
            env.workspace = entry


            NIR = Raster(os.path.join(x,"band_4"))
            NIR.save("B04_"+rename+"_old.tif")
            NIR_correct = NIR*float(a_4)
            env.workspace = rasters
            NIR_correct.save("B04_"+rename+".tif")
            print("NIR band corrected & saved")
            print("---------------------------")

            #assign corrected bands to variables for indicies' calculation
            Green = Green_correct
            Red = Red_correct
            NIR = NIR_correct

            #call Indices function
            Indices(rename,Green,Red,NIR)

            temp1 = 1
            env.workspace = entry

        elif rename2 == rename1:

        #if same date but different scene use different naming system using additional
parameter counter

        Green = Raster(os.path.join(x,"band_2"))
        Green.save("B02_"+rename+"_old.tif")
        Green_correct = Green*float(a_2)
        env.workspace = rasters
        Green_correct.save("B02_"+rename+"_"+str(temp1)+".tif")
        print("Green band corrected & saved")
        env.workspace = entry
```

```
            Red = Raster(os.path.join(x,"band_3"))
            Red.save("B03_"+rename+"_old.tif")
            Red_correct = Red*float(a_3)
            env.workspace = rasters
            Red_correct.save("B03_"+rename+"_"+str(temp1)+".tif")
            print("Red band corrected & saved")
            env.workspace = entry


            NIR = Raster(os.path.join(x,"band_4"))
            NIR.save("B04_"+rename+"_old.tif")
            NIR_correct = NIR*float(a_4)
            env.workspace = rasters
            NIR_correct.save("B04_"+rename+"_"+str(temp1)+".tif")
            print("NIR band corrected & saved")
            print("---------------------------")

            #assign corrected bands to variables for indicies' calculation
            Green = Green_correct
            Red = Red_correct
            NIR = NIR_correct

            #call Indices function
            Indices(rename,Green,Red,NIR)

            temp1+=1
            env.workspace = entry

    j+=1

    Stats()

if __name__=='__main__':
    main()
```

Appendix B – Regression Models

G1: Fields 2, 3, 4 and 5

Week	Variable	R2
	B2	0.246
	B3	0.248
	B4	0.237
	B2, B3 & B4	0.258
8	GNDVI	0.255
	MSAVI2	0.244
	NDVI	0.241
	NDWI	0.255
	SAVI	0.241

Week	Variable	R2
	B2	0.294
	B3	0.330
	B4	0.243
	B2, B3 & B4	0.353
10	GNDVI	0.321
	MSAVI2	0.327
	NDVI	0.325
	NDWI	0.321
	SAVI	0.325

Week	Variable	R2
	B2	0.278
	B3	0.313
	B4	0.386
	B2, B3 & B4	0.391
11	GNDVI	0.361
	MSAVI2	0.370
	NDVI	0.366
	NDWI	0.361
	SAVI	0.366

Week	Variable	R2
	B2	0.305
	B3	0.401
	B4	0.391
	B2, B3 & B4	0.454
12	GNDVI	0.437
	MSAVI2	0.445
	NDVI	0.444
	NDWI	0.437
	SAVI	0.444

Week	Variable	R2
	B2	0.214
	B3	0.344
	B4	0.564
	B2, B3 & B4	0.580
13	GNDVI	0.452
	MSAVI2	0.445
	NDVI	0.459
	NDWI	0.452
	SAVI	0.459

G2: Fields 1, 7, 8 (a & b), 9 and 10

Week	Variable	R2
	B2	0.225
	B3	0.229
	B4	0.241
	B2, B3 & B4	0.251
8	GNDVI	0.422
	MSAVI2	0.415
	NDVI	0.414
	NDWI	0.422
	SAVI	0.414

Week	Variable	R2
	B2	0.435
	B3	0.473
	B4	0.191
	B2, B3 & B4	0.558
10	GNDVI	0.558
	MSAVI2	0.545
	NDVI	0.570
	NDWI	0.558
	SAVI	0.570

Week	Variable	R2
	B2	0.338
	B3	0.221
	B4	0.201
	B2, B3 & B4	0.614
11	GNDVI	0.199
	MSAVI2	0.185
	NDVI	0.185
	NDWI	0.199
	SAVI	0.185

Week	Variable	R2
	B2	0.192
	B3	0.193
	B4	0.246
	B2, B3 & B4	0.337
12	GNDVI	0.202
	MSAVI2	0.214
	NDVI	0.216
	NDWI	0.202
	SAVI	0.216

Week	Variable	R2
	B2	0.251
	B3	0.274
	B4	0.221
	B2, B3 & B4	0.327
13	GNDVI	0.228
	MSAVI2	0.230
	NDVI	0.235
	NDWI	0.228
	SAVI	0.235

G3: Fields 11 a and b

Week	Variable	R2
	B2	0.626
	B3	0.631
	B4	0.589
	B2, B3 & B4	0.666
9	GNDVI	0.614
	MSAVI2	0.618
	NDVI	0.619
	NDWI	0.614
	SAVI	0.619

Week	Variable	R2
	B2	0.588
	B3	0.590
	B4	0.604
	B2, B3 & B4	0.642
10	GNDVI	0.601
	MSAVI2	0.610
	NDVI	0.607
	NDWI	0.601
	SAVI	0.607

Week	Variable	R2
	B2	0.597
	B3	0.612
	B4	0.624
	B2, B3 & B4	0.642
11	GNDVI	0.619
	MSAVI2	0.622
	NDVI	0.621
	NDWI	0.619
	SAVI	0.621

Week	Variable	R2
	B2	0.617
	B3	0.628
	B4	0.638
	B2, B3 & B4	0.639
12	GNDVI	0.632
	MSAVI2	0.633
	NDVI	0.633
	NDWI	0.632
	SAVI	0.633

Week	Variable	R2
	B2	0.616
	B3	0.627
	B4	0.624
	B2, B3 & B4	0.645
13	GNDVI	0.626
	MSAVI2	0.627
	NDVI	0.628
	NDWI	0.626
	SAVI	0.628