

INFORMING FIELD MANAGEMENT DECISIONS TO ENHANCE
ALFALFA SEED PRODUCTION USING REMOTE SENSING

by
Thomas V. Van Der Weide



A thesis
submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Geophysics
Boise State University

December 2019

© 2019

Thomas V. Van Der Weide

ALL RIGHTS RESERVED

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Thomas V. Van Der Weide

Thesis Title: Informing Field Management Decisions to Enhance Alfalfa Seed Production Using Remote Sensing

Date of Final Oral Examination: 9 August 2019

The following individuals read and discussed the dissertation submitted by student Thomas V. Van Der Weide, and they evaluated the students presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

T. Dylan Mikesell Ph.D. Chair, Supervisory Committee

Nancy F. Glenn Ph.D. Member, Supervisory Committee

Trevor Caughlin Ph.D. Member, Supervisory Committee

The final reading approval of the thesis was granted by T. Dylan Mikesell Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

DEDICATION

I dedicate this work to my late grandmother, Emily Marrufo Archuleta. Thank you for sharing your wisdom and always encouraging me to do my best.

ACKNOWLEDGMENT

I would like to thank the Idaho Global Entrepreneurial Mission for funding the project. Thank you to Dylan Mikesell and Nancy Glenn for their guidance and advice throughout this project. Thank you to my committee member Trevor Caughlin for his upbeat attitude and willingness to answer my questions. Thank you to Lucas Spaete and the rest of the Boise Center Aerospace Lab and also the Environmental Seismology Lab for their motivation, support, questions, and contributions throughout the last two years.

ABSTRACT

The development rate of alfalfa seed crop depends on both environmental conditions and management decisions. Crop management decisions, such as determining when to release pollinators to optimize pollination, can be informed by the identification of plant development stages from remote sensing data. I first identify what electromagnetic wavelengths are sensitive to alfalfa plant development stages using hyperspectral data. A Random Forest regression is used to determine the best Vegetation Index (VI) to monitor how much of the plant is covered in flower. The results indicate that Blue, Green, and Near-Infrared are the important electromagnetic wavelengths for the VI. Imagery collected throughout this study are converted into a VI time-series for analysis. The analysis involves using a state-space model to estimate the percentage of flower cover from observations. We found that a simple state-space model can be used to estimate, as well as predict, the flower cover percentage.

TABLE OF CONTENTS

DEDICATION	iv
ACKNOWLEDGMENT	v
ABSTRACT	vi
LIST OF FIGURES	xi
LIST OF TABLES	xx
LIST OF ABBREVIATIONS	xxi
LIST OF ABBREVIATIONS	xxii
1 INTRODUCTION	1
1.1 Alfalfa Background	1
1.2 Remote Sensing	3
1.3 Statistical Methods	6
1.4 Thesis Organization	7
2 DETERMINATION OF ELECTROMAGNETIC WAVELENGTHS SENSITIVE TO ALFALFA BLOOM	9
2.1 Summary	9

2.2	Introduction	10
2.2.1	Vegetation Indices	10
2.3	Field sites	11
2.3.1	Location	11
2.4	Methods	15
2.4.1	Data Collection	15
2.4.2	SamplePoint Procedure	18
2.4.3	Random Forest Regression	20
2.5	Results	21
2.5.1	SamplePoint	21
2.5.2	Regression	22
2.6	Discussion	23
2.7	Conclusions	25
3	ACQUISITION OF NIR-GREEN-BLUE DIGITAL IMAGERY FROM FIELD STATIONS FOR CROP MONITORING	27
3.1	Summary	27
3.2	Introduction	27
3.3	Field-Sensor Development	29
3.3.1	Sensors	32
3.3.2	Power Management	42
3.4	2018 Field Season Data Collection	43
3.4.1	Location	43
3.4.2	Data Collection	44
3.5	Results	46

3.5.1	Field Imagery	46
3.5.2	Temperature and Humidity	47
3.5.3	Solar Irradiation	49
3.6	Discussion	49
3.7	Conclusions	54
4	IMAGE PROCESSING AND CLASSIFICATION	55
4.1	Summary	55
4.2	2017 Data	56
4.2.1	Introduction	56
4.2.2	Methods	57
4.2.3	Results	62
4.2.4	Discussion	63
4.3	2018 Data	64
4.3.1	Introduction	64
4.3.2	Methods	67
4.3.3	Results	70
4.3.4	Discussion	71
4.4	Conclusions	74
5	PARTICLE FILTER MODELING TO PREDICT ALFALFA SEED CROP FLOWER PERCENTAGE USING VI TIME SERIES	96
5.1	Summary	96
5.2	Introduction	97
5.3	Methods	100

5.3.1	Process Model	100
5.3.2	Observation Model	104
5.3.3	Particle Filter	111
5.4	Results	112
5.5	Discussion	115
5.6	Conclusions	117
6	DISCUSSION, CONCLUSIONS, AND FUTURE WORK	121
6.1	Discussion and Conclusions	121
6.2	Future work	125
	REFERENCES	128
	APPENDICES	138
	APPENDIX A VEGETATION INDICES USED IN REGRESSION ANALYSIS	139

LIST OF FIGURES

2.1	The field and greenhouse locations for the 2016 data collection.	11
2.2	The data collection pattern that was used for the GH1 plots.	13
2.3	The data collection pattern that was used for the GHS plots.	14
2.4	Structure of the sampling locations within a field.	15
2.5	Example of the ASD FieldSpec HandHeld2 used for hyperspectral data collection.	17
2.6	This chart shows the contact probe hyperspectral signatures for buds, leaves, and flowers averaged to represent broadband wavelengths of Blue, Green, Red, RedEdge, and NIR. The Blue band has a center wavelength of 475 nm with a full width half max (FWHM) of 20 nm, Green has a center wavelength of 560 nm with a FWHM of 20 nm, Red has a center wavelength of 668 nm with a FWHM of 10 nm, Rededge has a center wavelength of 717 nm and FWHM of 10 nm, and NIR has a center wavelength of 840 nm with a FWHM of 40 nm.	18
2.7	Example of one of the images classified in SamplePoint image analysis software.	19

2.8	SamplePoint flower percentage cover for three groups in two fields where groups in Field 1 are shown in green and groups in Field 2 are shown in blue. A) Is the original SamplePoint data. B) Same data smoothed with an exponential weighted-average moving-window filter.	21
2.9	R^2 values for the RF regression on percentage flower cover for Field 2.	22
2.10	The pseudo-R2 results from the random forest regression at both Field 1 and Field 2.	23
2.11	The 30 most important variables from the second iteration of the random forest regression for A) Field 1 B) Field 2 C) GHS and D) GHL.	26
3.1	The solar panel wires are soldered directly to the voltage regulator, which is then soldered to wires connected to a micro USB adapter. The micro USB adapter connects to a female micro USB adapter which goes to a male D/C adapter that plugs into the portable USB power bank.	30
3.2	This diagram shows the circuit layout to charge the portable USB battery bank and power the Raspberry Pi 3B.	31
3.3	This image shows a RPi board and the 40-pin header, from Raspberry Pi Pin Map (2017).	33
3.4	The MAPIR Survey3W blue (475 nm), green (550 nm), and NIR (850 nm) spectral responses, from MAPIR Survey3 Camera (2017).	35
3.5	The RPi is powered by a USB battery bank and is connected to the Survey3 camera via a USB to mini USB cable and two header pins of a micro HDMI.	37

3.6	The multi-sensor unit in the field. This is the unit that uses the PVC tube enclosure. Triangular support arm and steel support cables are visible. This support system is used to emplace all field units, regardless of enclosure type.	38
3.7	Interior of the junction box enclosure. The USB pyranometer has extra cable that is coiled inside this enclosure.	39
3.8	Exterior of the junction box enclosure. A silicone sealant surrounds the camera lens to keep water out. The temperature/humidity sensor is zip-tied to the bottom of the enclosure to shield it from rain and direct sunlight, while remaining out of the image.	40
3.9	Overview of the fields with the locations of sensors A and B during the 2018 data collection.	44
3.10	Time-line for the data collected during the 2018 growing season. With the pollinator release date for Field 2 shown as a vertical orange line.	47
3.11	Preliminary results from Field 1B imagery. Vignette corrected TIFF images are converted into a Flower Vegetation Index (FVI), from which the mean value for each image is plotted through time over the growing season. The pollinator release date for Field 1 is indicated by the vertical orange line.	48

3.12	Comparison of temperature (top) and humidity (bottom) data recorded with the low-cost AM2302 compared to high-quality <i>Davis Vantage Pro2 Plus</i> continuous recording weather station.	50
3.13	Comparison of solar radiation data recorded with the Apogee SP-420 pyranometer compared to high-quality <i>Davis Vantage Pro2 Plus</i> continuous recording weather station.	51
4.1	The fields and greenhouse locations for the 2017 data collection. . . .	56
4.2	Micasense Rededge Imager Nomenclature shows that the RE band is in the middle and is used for the registration of the other bands (Base, 2015).	58
4.3	Spectral response curves for the Micasense Rededge camera (Tagle, 2017) with the color of the lines corresponding to the camera band in the table below.	59
4.4	Micasense Rededge image processing steps. All steps except those in red are implemented in Python.	75
4.5	Vignette correction model where the left image shows how far from the center (cx, cy) each pixel is, and the right image shows the value the original pixel is divided by.	76

4.6	This is an example of the Python script output used to collect the reflectance panel values. The Micasense Rededge bands are first calibrated and then aligned. The combined RGB image is displayed for a user to manually draw a rectangle (blue box) within the boundaries of the reflection panel. The average value for each band is then saved and applied to the images of the dataset.	77
4.7	Reflection panel values extracted using a custom Python script from over the course of the growing season for fields 1 and 2.	78
4.8	Original RGB composite image from the Micasense Rededge multi-spectral camera. Misaligned pixels are evidenced as blue, green, and red striping	79
4.9	Shifted cross-power spectrum for two of the Micasense Rededge camera bands RE and Blue. (a) Is the original modified Hanning Window. (b) Same data but with a hard-coded 1/6th maximum translation. While colors mean nothing without a scale, please excuse the lack of a color bar for now and it will be included later.	80
4.10	Zoomed in view of the shifted cross-power spectrum for two of the Micasense Rededge camera bands, RE and Blue. The center of the images is at pixel (640,480). While colors mean nothing without a scale, please excuse the lack of a color bar for now and it will be included later.	81

4.11	RGB composite image from the Micasense Rededge after image registration. Note the striping is now gone in most of the image and flowers are visible. Misalignment still exists in some areas (e.g. upper left corner).	82
4.12	Histograms for the image translation results.	83
4.13	Field 1 field VI progression after each of the processing steps is applied. The panels should be read top to bottom and left to right to see the progression of the image processing steps.	84
4.13	Field 2 field VI progression after each of the processing steps is applied. The panels should be read top to bottom and left to right to see the progression of the image processing steps.	85
4.14	The fields locations for the 2018 data collection.	86
4.15	MAPIR Survey3W used in this experiment.	87
4.16	Spectral response for the MAPIR Survey3W.	88
4.17	Setup for the vignette correction data collection.	89
4.18	Example of the vignette correction to the NIR channel at 12:00 on 11 June 2018: original NIR image (left); vignette NIR correction function (middle); corrected NIR image (right).	89
4.19	Examples of the resulting classified images from the SVM trained classifier.	90
4.20	Comparison between the two VI used in the image classification for Field1A. The panels should be read top to bottom and left to right to see the progression of the image processing steps.	91

4.20	Comparison between the two VI used in the image classification for Field1B. The panels should be read top to bottom and left to right to see the progression of the image processing steps.	92
4.20	Comparison between the two VI used in the image classification for Field2A. The panels should be read top to bottom and left to right to see the progression of the image processing steps.	93
4.20	Comparison between the two VI used in the image classification for Field2B. The panels should be read top to bottom and left to right to see the progression of the image processing steps.	94
4.21	All stations from the two fields are classified using a trained SVM classifier. The percentage cover is extracted from each classification image and smoothed using a Savgol Filter with a window size 1/3 of the length of the dataset and a polynomial order of degree 3.	95
5.1	Percentage flower cover with time from the 2016 (green lines), 2017 (red lines), and 2018 (blue lines) growing seasons. The 2016 and 2017 PFC estimates come from SamplePoint image analysis, while the 2018 PFC estimates are from the SVM classification (Chapter 4).	102
5.2	Percentage flower cover with time at two alfalfa seed crop fields during the 2017 growing season.	104
5.3	GNDVI values over days after cutback from all the locations sampled during this study. Each data set is collected using a different sensor, which likely explains the variations between years. As in the process model example, the 2016 data are shown as green lines, 2017 data as red lines, and the 2018 data as green lines.	106

5.4	GNDVI values over PFC estimates from all the locations sampled during this study. Each data set is collected using a different sensor, which likely explains the variations between years. As in the process model example, the 2016 data are shown as green lines, 2017 data as red lines, and the 2018 data as green lines.	107
5.5	Vegetation Index values over percentage flower from two fields during the 2017 growing season.	108
5.6	The 30 most important variables from the random forest regression on all 2017 data.	110
5.7	The best performing model fit is shown as a blue line with the 95% confidence interval range shown in grey.	118
5.8	Field 1: sites 1, 2, 3 (left column, top to bottom). Field 2: sites 1, 2, 3 (right column, top to bottom). The process and observation models are passed into a particle filter algorithm along with the observations from individual field sites. The prediction line (blue) is estimated by combining the process model with the most current observation, in this case a GNDVI value. The observations are subset to only include the first 10 observation values (green line). The observation subset is used to forecast state estimates for the remainder of the data collection period (orange line). The accuracy of the forecasted predicted percentage flower cover values (orange) is given by the R^2 and $RMSE$ value. . .	119

5.9 Field 1: sites 1, 2, 3 (left column, top to bottom). Field 2: sites 1, 2, 3 (right column, top to bottom). The process and observation models are passed into a particle filter algorithm along with the observations from individual field sites. The prediction line (blue) is estimated by combining the process model with the most current observation, in this case a GNDVI value. The accuracy of the predicted percentage flower cover values (blue) is given by the R^2 and $RMSE$ value. 120

LIST OF TABLES

2.1	Buttons used in the SamplePoint percentage cover classification. . . .	19
2.2	A list of the best performing VIs that are common between both fields.	23
3.1	Breakdown of the cost of components to build this system.	32
4.1	Date ranges the field cameras collected data.	65
4.2	Radiometric coefficients used to calibrate the MAPIR Survey3W camera.	69
4.3	ROI separability for each class at every field. Any value above 1.9 is considered reliable separability between classes with a maximum value of 2 (Jeffreys, 1946).	71
4.4	User and Producer accuracy for each class at every field. A perfect classification accuracy score is 1.	72
5.1	Leave-one-out cross-validation model comparison of the three process models.	113
5.2	Leave-one-out cross-validation model comparison of the three observation models.	113
5.3	Summary of the variable estimates for the process model.	113
5.4	Summary of the variable estimates for the observation model.	113
5.5	Summary of the R-squared and RMSE values for the estimated states and the predicted states.	114

5.6	Summary of the R-squared and RMSE values for the predicted states of an accelerated dataset.	115
A.1	Vegetation indices used in the Random Forest Regression. Band ab- breviations are N =NIR, R =Red, G =Green, B =Blue.	140

LIST OF ABBREVIATIONS

ASD	Analytical Spectral Devices, Inc
elpd	Expected Log Point-wise Predictive Density
FOV	Field of View
FVI	Alfalfa Flower Vegetation Index
GDD	Growing Degree Day
GNDVI	Green Normalized Difference Vegetation Index
GSD	Ground Sample Distance
MSE	Mean Squared Error
NIR	Near-Infrared
pdf	Probability Density Function
PF	Particle Filter
PFC	Percentage Flower Cover
PWM	Pulse Width Modulation
RE	Rededge
RF	Random forest
RFR	Random forest regression
RGB	Red-Green-Blue

RMSE Root Mean Square Error

ROI Region of Interest

RPi Raspberry Pi

SVM support vector machine

T/H Temperature/Humidity

VI Vegetation Index

VNIR Visible and Near-infrared

CHAPTER 1: INTRODUCTION

1.1 Alfalfa Background

Alfalfa (scientific name *Medicago sativa L.*), commonly known as Lucerne, is an important forage crop in many countries around the world and is used heavily in feeding programs of various livestock, especially as beef and dairy fodder. Alfalfa was the fourth most valuable field crop in the United States in 2017 and is the most widely cultivated perennial forage crop in the world (NASS, 2017). The United States harvested 16.9 million acres of alfalfa for hay in 2016 (NASS, 2017) and has historically had more area of alfalfa production (in million ha) than any other country in the world (Yuegao & Cash, 2010). In 2015 the United States produced approximately 50.5 million pounds of alfalfa seed (NASS, 2017) and was also the largest exporter of alfalfa seeds for sowing, representing 25.5% of the global export market (USDA, 2015). The demand for high-quality alfalfa seed is expected to increase as developing countries are projected to have an annual growth in meat production of 2.4% and annual growth in milk production of 2.5% by 2030. The growth in livestock production will create a demand for more high-quality forage crops, which must be backed by high-quality forage seeds.

The planting of alfalfa for forage requires between 15-25 pounds of seed per acre,

with approximately 200,000 seeds per pound (CFAC, 2004). This high demand for quality alfalfa seeds puts an emphasis on getting optimal seed yields from alfalfa seed crops. In order to get the largest seed yields, pollinators should be introduced to the field one week prior to peak bloom conditions (Husman, 2015). Making informed management decisions, such as when to release pollinators, increases the crop and seed productivity and thus improves farm profitability (Mulla, 2013). Making these management decisions requires analysis of data related to the real-time plant growth (i.e. the phenology).

Alfalfa is a perennial herbaceous legume that grows from a semi-woody base or crown. The crown sends up many leafy multibranched stems 2 to 4 feet high; with each stem terminating in a raceme or cluster of 10 to 100 purple florets. The 1/2 inch long florets begin opening at the base of the 1 to 4 inch long raceme and take about a week to progress to the tip of the raceme. The florets can open any time of day and remain open for about a week before wilting. However, once pollinated the flower wilts within a few hours (Mcgregor, 1976). Each flower has 5 petals, with the lower two petals modified into a 'keel' that encloses the reproductive organs. Pollinators must force their way between the keel petals to access nectar and pollen. When a flower visitor 'trips' the flower it gets struck in the head by the pistil and stamens, which deposits pollen. A flower must be tripped for fertilization to take place and a seed pod to form.

The highest seed yields come from sparse stands that flower during the warmest part of the season, but there are many factors at play such as proper agronomic care, sufficient pollination, freedom from harmful insects and diseases, and proper seed-harvesting methods (Mcgregor, 1976). The preferred pollinator for alfalfa is

the alfalfa leaf cutter bee. The bee's effectiveness can be increased by timing their development with the beginning of bloom. However, it can be difficult to anticipate bloom one week prior to peak bloom, which is when the bees' emergence must be initiated (Husman, 2015). Balancing the quantity and timing of bee release is critical for successful pollination as well as the bees' own health and reproduction (NSERC-CANPOLIN Canadian Pollination Initiative, 2006).

The knowledge of the current stages of bloom for alfalfa seed crop fields, with a high spatial and temporal resolution, can be a powerful tool for practical management purposes. This project explores the use of remote sensing technologies to gain knowledge of the current stage of bloom development, as well as forecasting bloom percentages at alfalfa seed crop fields in southwestern Idaho.

1.2 Remote Sensing

Remote sensing, which refers to non-contact measurements of radiation reflected from an object, includes platforms for making measurements such as satellite, aerial, and handheld or tractor mounted sensors. These platforms can be used to collect data for a variety of applications, including precision agriculture (Mulla, 2013). Well-validated remotely sensed data sets are more cost and time effective compared to sampling the same area using field-based approaches (Weng, 2002). Moreover, when several well calibrated satellite images are used in time-series, the phenological stages of crops, including the floral cycle, can be effectively mapped (Ge *et al.*, 2006). The floral cycle refers to the duration of the blossoming period and flowering intensity, which includes the fractional coverage of flowering buds within a single tree, plant, or vegetation community (McIntosh, 2002). Landmann *et al.* (2015) used time series hyperspectral remote sensing to determine the ability to monitor the floral cycle

of flowering plant groups in African savannas. Previous research has shown that spectral signatures from the visible to shortwave infrared can be used to monitor plant phenology development (Sakamoto *et al.*, 2005; Fan *et al.*, 2015), but until now little work has gone into identifying specific spectral signatures of alfalfa seed crop phenology. In this thesis, electromagnetic wavelengths and vegetation indices sensitive to alfalfa bloom progression are identified by applying random forest regression to hyperspectral and Red-Green-Blue (RGB) color imagery collected during the 2016 alfalfa growing season in Idaho.

A vegetation index (VI) is a mathematical combination or transformation of spectral wavelengths that accentuates the spectral properties of plants (Xue & Su, 2017). A VI can be used to distinguish between vegetation and soil, differentiate between species of plants, provide estimates of plant biomass or leaf area index (Viña *et al.*, 2011), estimate phenology (Atkinson *et al.*, 2012), and even track flowering plant species (Landmann *et al.*, 2015). The spectral changes related to phenology development can be emphasized using sequential images, transformed into a vegetation index, and ordered into a time-series (Sakamoto *et al.*, 2005).

Publicly-available satellite data can be used to observe alfalfa seed crop fields throughout a growing season. For example, the Sentinel-2 satellite constellation has four spectral bands centered at: Blue (490nm), Green (50 nm), Red (665 nm), and NIR (842 nm) that are imaged at 10m resolution. The spectral information from satellites can be used to characterize spatial variability within a field (Lee *et al.*, 2010). While the Sentinel-2 (A&B) satellites provide spectral data coverage once every five days globally, this good temporal-resolution can be limited by cloud or smoke cover. Complimentary data can be collected using "near-surface" remote sensing im-

agery to improve alfalfa phenology monitoring with high temporal data collection. Many studies have demonstrated the usefulness of proximal remote sensing to provide quantitative information about crop conditions (Mulla, 2013; Nagai *et al.*, 2016; Marino & Alvino, 2014). For example, Richardson *et al.* (2009) show how near-surface imaging sensors can be used to document changes in the phenology of forest canopies. When Hufkens *et al.* (2012) compared near-surface and satellite remote sensing-based observations of vegetation phenology, they found “significant agreement between phenological time series and metrics derived from these two data sources. However, issues of scale and representation strongly influence the relationship between near-surface and satellite remote sensing measures of phenology.” Their results show that the two sensing platforms can complement each other, but that the correlation is strongly dependent on the camera field of view (FOV). Care was taken to ensure that the near-surface cameras have as wide a FOV as possible and a similar spectral response as the Sentinel-2 satellites to make integration between the two platforms as easy as possible.

In this study, the relationship between remote sensing measurements and the onset of alfalfa flowers is determined using state-of-the-art signal processing techniques, an ensemble learning method for regression, and Bayesian statistical inference. The results from determining this relationship can be used in a state-space modelling approach where the time-series spectral development of alfalfa can be used to give a near-real time analysis of the current phenological state, as well as predict when bloom will occur (De Bernardis *et al.*, 2016b).

1.3 Statistical Methods

In statistical modeling, regression analysis is a statistical process to estimate the relationships among variables. Random forest regression (RFR) is an ensemble learning method for classification or regression that operates by constructing a multitude of decision trees, using a randomly selected subset of training samples and variables, and outputting the class that is the mean prediction of the individual trees. Belgiu & Drgu (2016) determine that a RFR classifier can successfully handle high data dimensionality and multicollinearity, being both fast and insensitive to overfitting. RFR is thus well suited to be used with hyperspectral imagery. A RFR performed with data collected during the 2016 growing season identifies significant wavelengths, or wavelength combinations, for tracking alfalfa bloom development.

The significant wavelengths, along with ground validation data, are used in a Bayesian regression model to fit mathematical approximations for the 'standard' behavior of the field, known as a process model, and the relationship between observed measurements and the process (observation model). The mathematical approximations benefit from fitting a Bayesian multilevel model which can account for variations among groups (field locations) within the data (McElreath, 2015). Using a Bayesian framework also allows us to take advantage of prior data for several parameters in the model. A logistic function (Tong & Vendettuoli, 2017) is used as the process model to represent the standard bloom trajectory at an alfalfa seed crop field. A Michaelis-Menten function (Michaelis *et al.*, 1913) is used to relate the observation data to the ground validations because it is characterized by a steep rise that gradually flattens into a plateau. The process and observation models are then passed into a particle filter framework that is able to infer the state of bloom at alfalfa seed crop fields using

observation data.

A Particle Filter Approach for Real-Time Estimation of Crop Phenological States Using Time Series of NDVI Images by De Bernardis *et al.* (2016b) illustrates the potential for using a particle filter (PF) approach with rice parcels, although the methodology is transferable for a near-real time estimation and prediction of alfalfa bloom. While the PF is often used in robotics and computer vision problems to estimate state variables (e.g. Durrant-Whyte & Bailey, 2006), its application in remote sensing is still fairly novel. The particle filter is a dynamic approach that combines information from a process model and observation data (provided in this research by the remote sensing data). We can then take the estimated state based on time from the process model and combine it with the observation data to get a near-real time estimate of percentage flower.

1.4 Thesis Organization

This thesis contains four chapters related to the monitoring of alfalfa seed crops as they progress through a flowering cycle. The first study demonstrates the potential of tracking the development of bloom at alfalfa seed crop fields using remote sensing technologies (Chapter 2). A RFR using data collected over the 2016 growing season identifies electromagnetic wavelengths, or combination of wavelengths sensitive to alfalfa bloom. In Chapter 3, the three most significant electromagnetic wavelengths are used to select a camera to include in a near-surface sensor. This chapter covers the development of a near-surface sensor deployed in a field during the 2018 growing season to monitor the spectral signature of alfalfa as it progresses through a flowering cycle. Chapter 4 details how the 2017 and 2018 field imagery are processed in order to get a calibrated time-series and illustrates the effects of calibration. The

image calibration process includes a vignette correction, which was defined by the manufacturer for the 2017 data and determined in a laboratory experiment for the 2018 imagery, and a radiometric calibration where digital numbers are converted into units of radiance. The calibrated field imagery is then transformed into a VI time-series. The vegetation indices from the field imagery serve as observation data in the bloom prediction model presented in Chapter 5. This chapter covers the use of a state-space model to predict the on-set of bloom at alfalfa seed crop fields. The state space model uses a prediction model to estimate the percentage of flower cover and an observation model to update the predictions. The prediction model is defined as the mathematical approximation of the expected behavior for a field where the onset of flower is a function of time. For the observation model, a vegetation index is used as an indication of the crops percentage flower cover. The observation model relates the percentage flower cover with the observed VI value. As a proof of concept, the bloom prediction model is used to forecast alfalfa seed crop field bloom percentage 5 and 10 days in advance. In the final chapter, the results of this research are discussed and reflections are made about shortcomings of this research, as well as about future research in this direction.

CHAPTER 2: DETERMINATION OF ELECTROMAGNETIC WAVELENGTHS SENSITIVE TO ALFALFA BLOOM

2.1 Summary

The development of flowers at alfalfa seed crop fields has an influence on the spectral signature that can be detected by satellites. In order to make remote sensing technology accessible to growers, the three most significant electromagnetic wavelengths for monitoring bloom are determined because that is what is widely available in consumer grade electronics. The determination of electromagnetic wavelengths that indicate bloom is determined by a random forest regression using data collected over the 2016 growing season. The 2016 dataset includes spectral data measurements and Red-Green-Blue (RGB) imagery. A high resolution spectroradiometer is used to collect spectral data measurements at 15 locations within 4 different growing environments in southwestern Idaho and are compared to the percentage cover results from SamplePoint photo analysis software using the RGB images. The percentage cover and the hyperspectral signature progression is used in a random forest regression to determine the significant electromagnetic wavelengths sensitive to percentage

flower cover. The three most significant wavelengths, including wavelength combinations, determined by the random forest regression are Blue, Green, and Near-infrared (NIR) and are used in a phenocam at ground level during the 2018 growing season, the results of which are presented in Chapter 3.

2.2 Introduction

In this chapter we use a random forest (RF) regression to analyze spectral data measurements collected at the two fields and two greenhouses during the 2016 growing season. RGB imagery is used in SamplePoint image analysis software to classify the data into seven categories, and the random forest regression (RFR) determines the spectral wavelengths or vegetation indices (VI) that are sensitive to alfalfa flowers. We begin with a description of the field sites and then describe the methods used to determine the significant wavelengths for tracking bloom development. Finally, we present the results.

2.2.1 Vegetation Indices

A vegetation index is a mathematical combination or transformation of spectral wavelengths that accentuates the spectral properties of plants. A VI can be used to distinguish between vegetation and soil, differentiate between species of plants, provide estimates of plant biomass or leaf area index (Viña *et al.*, 2011), estimate phenology (Atkinson *et al.*, 2012), and even track flowering plant species (Landmann *et al.*, 2015). We test different VIs in the random forest regression, as well as individual electromagnetic wavelengths.

2.3 Field sites

2.3.1 Location

We conducted the 2016 study at two fields and two greenhouse locations in the Nampa and Parma agricultural region (Figure 2.1) in southwestern Idaho, with the two greenhouses being located very close to each other. The 2016 data set was collected throughout the growing season from June 6th to August 22nd and included around 40 days worth of RGB and spectral data measurements taken. Greenhouse #1 (GHL) was imaged at six locations while Greenhouse #2 (GHS) and the two fields were imaged at three locations.

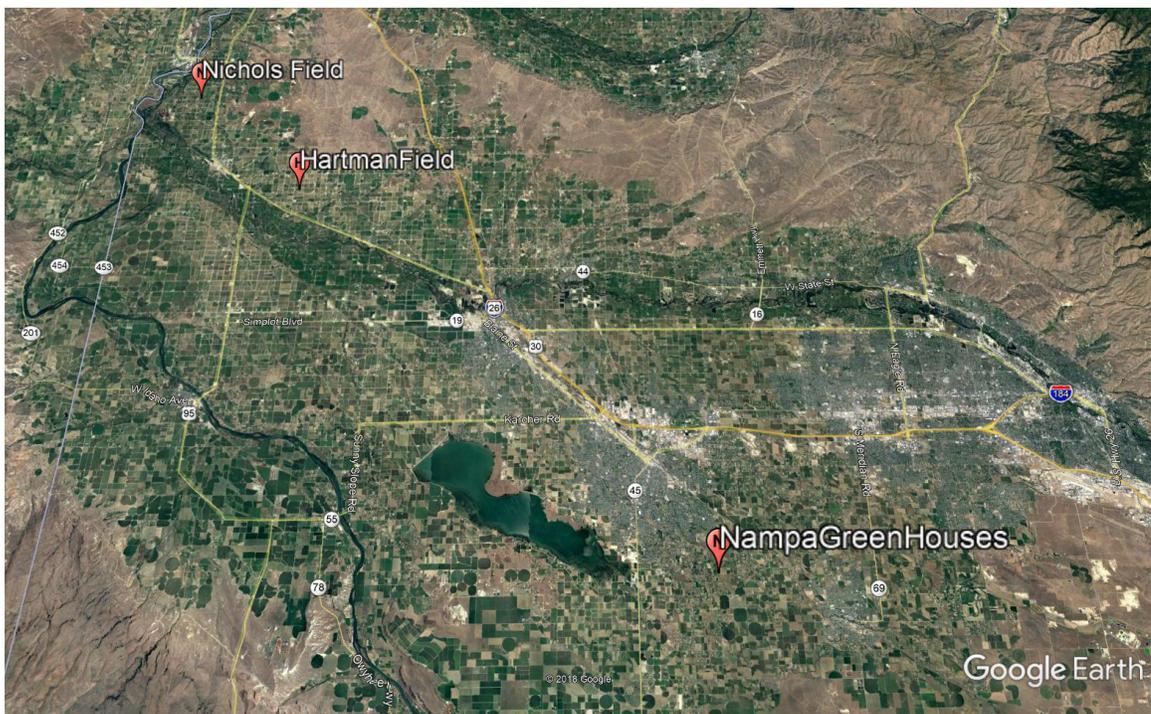


Figure 2.1: The field and greenhouse locations for the 2016 data collection.

GHL

The larger of the two greenhouses, GHL, consisted of a steel cage placed over a research plot at the S&W Seed Co. Research Facility with netting placed over the entire cage to control bee access to plants. It contained 12 trays of plants with 3 groups on each tray. Each group was considered a sampling point giving us 36 sample plots total. Figure 2.2 illustrates the data collection pattern that was used for the GHL plots. The numbers from the figure are used at the end of the file names for each data type and are used as the ID number, e.g. an ASD hyperspectral measurement taken on June 15 at location 1 would have the file name 0615G00001.asd. RGB images were collected daily for all 36 groups in the greenhouse, but ASD spectra are only collected for the center plant for the middle 2 rows for all 3 columns of trays as shown in blue numbers on Figure 2.2. This strategy minimizes any potential edge effects. The spectral data measurements are collected a low and high distance spectra for each group, giving 12 spectral signatures for the 6 groups. This procedure was followed from June 8th until August 23rd, unless pesticide application interfered, providing a total of 52 days of imagery.

GHS

The smaller of the two greenhouses, GHS, consists of an uncovered research plot at the S&W Seed Co. Research Facility. It consists of 2 trays, one tray with 1 group and the other with 2 groups, giving a total of 3 groups with approximately 40 plants each. Figure 2.3 illustrates the data collection pattern that was used for the GHS plots. These numbers are used in the file naming scheme and as group ID number in the same format as GHL data. One RGB image was collected daily for each group

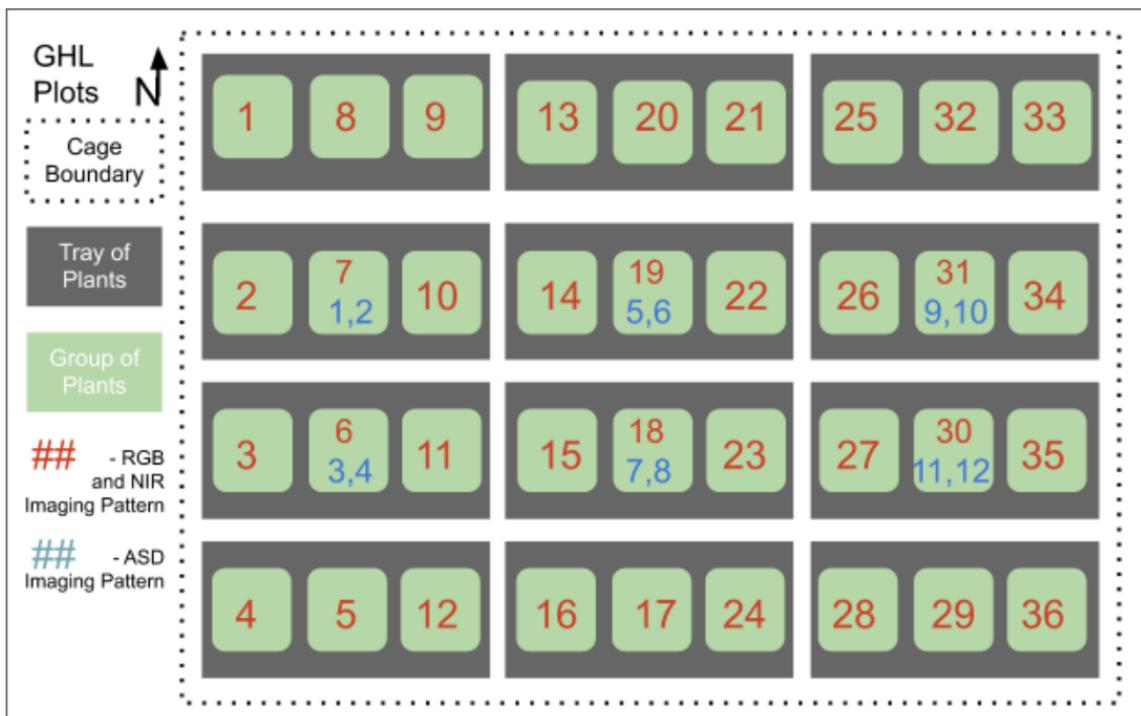


Figure 2.2: The data collection pattern that was used for the GHL plots.

and 10 ASD samples are collected at the same location for each group. The plants in this plot are started later than those in GHL and this procedure was followed from July 30th until Oct 10th, unless pesticide application interfered, providing a total of 49 days of imagery.

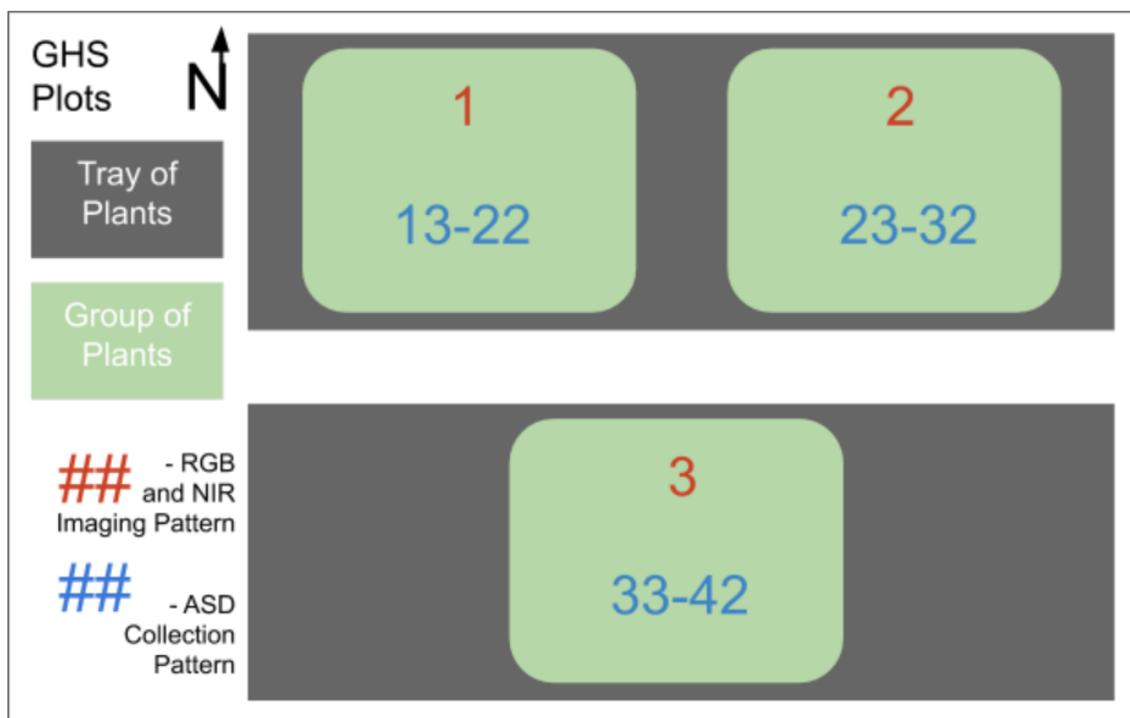


Figure 2.3: The data collection pattern that was used for the GHS plots.

Fields

Two fields, Field 1 and Field 2, located in Nampa and Parma, ID are selected for imaging during the 2016 growing season. Three locations are sampled per field in the same configuration as shown in Figure 2.4. Point 1 is 30 paces into the field to avoid any edge effects and so each point is as representative of the field as possible. Each point is marked with a small flag and care is taken to avoid disturbing the area. At

each point an RGB image is collected from 1m height with a consistent orientation so the camera is parallel with the rows. Ten scans are taken at each point with the high resolution spectroradiometer (See Section 2.4.1). This procedure is followed every day on both fields from June 7th until July 8th, unless pesticide application interfered. Once the fields are close to fully seeded, sampling is reduced to every 4 days from July 8th to August 10th.

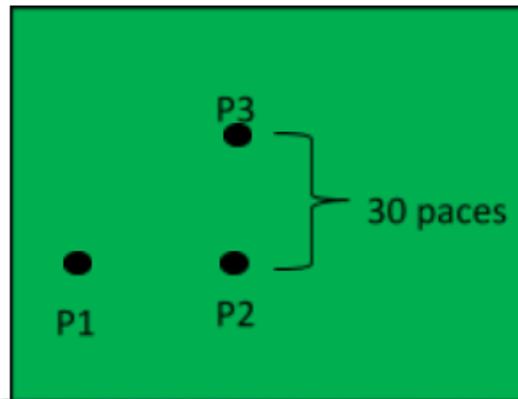


Figure 2.4: Structure of the sampling locations within a field.

2.4 Methods

2.4.1 Data Collection

Hyperspectral Data

Spectral data measurements are taken with a tripod-mounted ASD FieldSpec Hand-Held 2 (HH2) spectroradiometer (Figure 2.5), which has a spectral range over the Visible and Near-infrared (VNIR) spectrum. The ASD FieldSpec HH2 has a wavelength range from 325 to 1075 nm with a wavelength accuracy of ± 1 nm and a spectral resolution of < 3 nm at 700 nm. It has a 25° Field of View (FOV), which

when taking measurements at 140 cm height translates to an area of 60 cm diameter being collected. Each signature is the spectrum average of 10 samples. This is the recommended default setting according to the HH2 User Manual, which states “the signal-to-noise ratio improves with the square root of the number of scans used in the averaging”.

Prior to data collection the ASD HandHeld 2 was calibrated using a dark current calibration and a white reference panel, where white means the panel diffusely reflects nearly 100% of the incident light throughout the spectral range (Panalytical, 2017). The white panel is made of Spectralon, and while it is treated as having reflectance values of 1 at all wavelengths, the actual reflectance is slightly less than 1. ASD ViewSpec Pro software is used to process the raw reflectance values into absolute reflectance values, where the small discrepancies in spectral reflection values of the Spectralon material are accounted for.

The 1 nm wavelength hyperspectral wavelengths are then combined into a representation of Red-Green-Blue-NIR bands that span a similar range as the Micasense Rededge camera used during the 2017 growing season (See Chapter 4). It should be noted that no additional noise was added to account for differences in sensor calibration so this is not a straight comparison of how the multispectral camera would respond. The bands are combined into vegetation indices (VI) (Appendex A) and passed all of these parameters as response variables in the regression.

Contact Probe

In addition to the 2016 field hyperspectral data, an active source ASD contact probe attachment is used with an ASD Fieldspec Spectroradiometer to collect pure spectral



Figure 2.5: Example of the ASD FieldSpec HandHeld2 used for hyperspectral data collection.

signatures for leaves, buds, flowers, and seed pods. The samples are collected on 2 August 2016 from the GHL plots and 14 spectral signatures of each class are saved. The signatures from each class are averaged and then used to represent how the spectral library looks from a broadband camera.

A visual inspection of the contact probe data (Figure 2.6) demonstrates that the spectral signature of buds is very similar to leaves and the spectral signature by itself would be very difficult to use to differentiate buds from vegetation. It is because of this similarity in the spectral signature between buds and vegetation that we focus on tracking flowers, which has a noticeably different spectral signature than the vegetation.

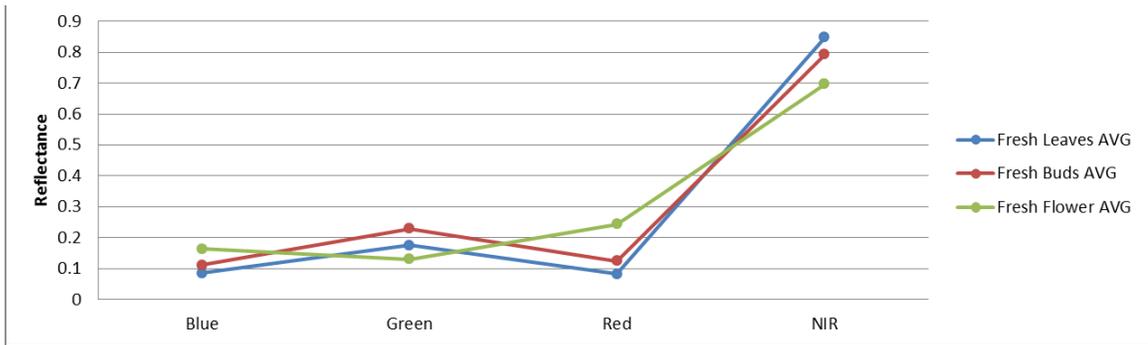


Figure 2.6: This chart shows the contact probe hyperspectral signatures for buds, leaves, and flowers averaged to represent broadband wavelengths of Blue, Green, Red, RedEdge, and NIR. The Blue band has a center wavelength of 475 nm with a full width half max (FWHM) of 20 nm, Green has a center wavelength of 560 nm with a FWHM of 20 nm, Red has a center wavelength of 668 nm with a FWHM of 10 nm, Rededge has a center wavelength of 717 nm and FWHM of 10 nm, and NIR has a center wavelength of 840 nm with a FWHM of 40 nm.

2.4.2 SamplePoint Procedure

SamplePoint photo analysis (Booth *et al.*, 2006) software was used to classify 100 pixels from the RGB imagery into percentage cover for 6 categories: *soil*, *vegetation*, *bud*, *flower*, *seed*, *unknown*, and *other* (Figure 2.7). The first step when using SamplePoint is to setup a database. This is where all the JPG images to be classified will be stored. Next, a custom button file is created with the classes shown in Table 2.1. Once this is completed we begin the image classification. We first select a grid size where we use the default setting of 10x10 or 100 points. Once the points have been classified they are saved to a .XLS file. Once the image set has been classified, the data are summarized by creating a statistics file. The statistics file contains a summary of how many pixels of each class in each image are classified and are referred to as *percentage cover*.



Figure 2.7: Example of one of the images classified in SamplePoint image analysis software.

Name - key	Definition
Soil - s	This is both the dark (wet) and light (dry) soil
Veg - v	Anything alfalfa that is not a raceme.
Buds - b	This refers to a green raceme, not purple, and the entire raceme
Flower - f	This refers to a purple raceme
Seed Pod - w	A fresh, green seedpod.
Unknown - x	Something unidentifiable because of the image quality
Other - z	Misc item in the image

Table 2.1: Buttons used in the SamplePoint percentage cover classification.

2.4.3 Random Forest Regression

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships among variables. The goal in using regression analysis is to determine which three wavelengths, or combination of wavelengths, best tracks bloom development. The RGB imagery and hyperspectral data from Field 1, Field 2, GHS, and GHL are used in a random forest regression to determine the significant wavelengths for tracking alfalfa bloom.

Random forest is an ensemble method for classification or regression that operates by constructing a multitude of decision trees, using a randomly selected subset of training samples and variables, and outputting the class that is the mean prediction of the individual trees. Belgiu & Druş (2016) determine that a RF classifier can successfully handle high data dimensionality and multicollinearity, being both fast and insensitive to overfitting. RF is thus well suited to be used with hyperspectral data. Because only about 2/3 of the data is used in each decision tree, the remaining 1/3 are used in an internal cross-validation technique (Breiman, 2001) that returns variable importance in terms of increasing Mean Squared Error (MSE) and Increasing Node Purity. The MSE is used to determine which variables are the most important. A "pseudo R-squared" value is returned for each regression and is used to indicate how well the regression performed. The formula for the pseudo R^2 value is $R^2 = 1 = mse/Var(y)$, where mse is the mean square errors, and $Var(y)$ is the variance of the response vector (Breiman, 2001).

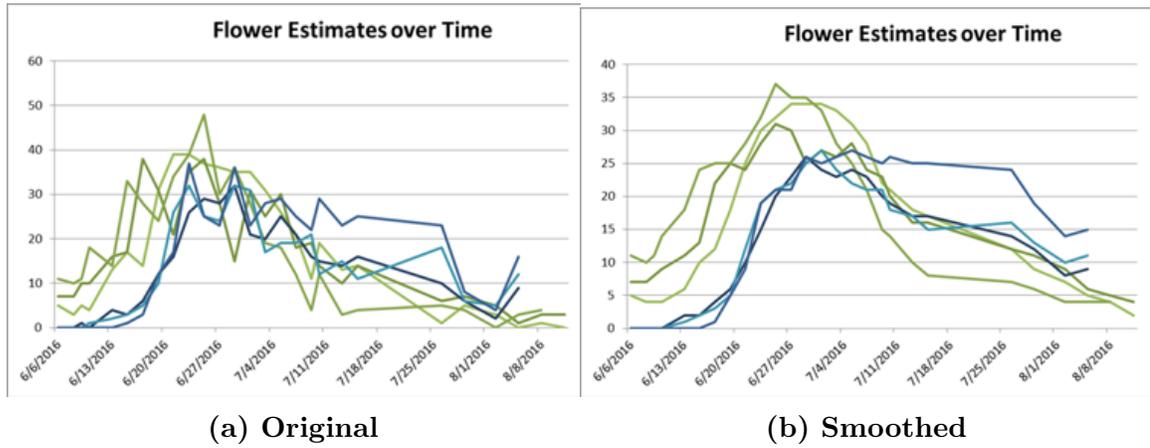


Figure 2.8: SamplePoint flower percentage cover for three groups in two fields where groups in Field 1 are shown in green and groups in Field 2 are shown in blue. A) Is the original SamplePoint data. B) Same data smoothed with an exponential weighted-average moving-window filter.

2.5 Results

2.5.1 SamplePoint

The SamplePoint analysis was done by an undergraduate in the Boise Center Aerospace Laboratory in the Summer 2017. The SamplePoint summary values for percentage cover vary quite a bit from day to day but show a definite overall trend where there are no flowers until after a certain date when the flower cover estimate increases until it levels off and eventually decreases. The percentage cover data is read into a pandas dataframe in Python and smoothed using an exponential weighted-average moving-window filter with a span of 5 (Figure 2.8). The smoothed percentage cover are used as the prediction variables in the regression model.

2.5.2 Regression

The RF regression was run in the free software R (R Core Team, 2013). The regression was run for four of the six classes: *soil*, *veg*, *bud*, and *flower* individually. The regression was set to run with 1000 trees, although Figure 2.9 shows that the error converged on a result with less than 200 trees. A second iteration of the RF regression was run with only the 40 most important wavelengths, bands, and VIs from the previous regression, where importance is defined as the mean decrease in accuracy.

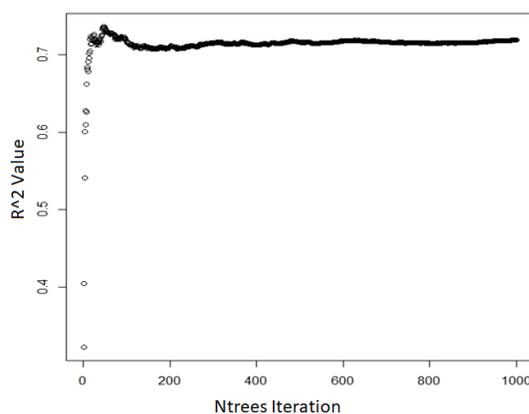


Figure 2.9: R^2 values for the RF regression on percentage flower cover for Field 2.

The accuracy of the regression at Field 1 and 2 for *soil*, *vegetation*, *bud*, and *flower* are shown in Figure 2.10. The R^2 values for each class are used as a goodness of fit indicator, where higher R^2 values mean the regression is better able to fit the data. These results show that *soil* and *flower* are the two most distinct classes to monitor, with *soil* having R^2 values of 0.48 and 0.84 and *flower* having R^2 values of 0.67 and 0.80 at Field 1 and 2, respectively. The second iteration where only the 40 most important wavelengths, bands, and VIs from the previous regression are used improved the R^2 values by 2-6% depending on the class and field. The R^2 values

for flower at GHS and GHL is 0.58 for both. The second iteration improved the R^2 values by 2% resulting in an R-squared value of 0.60.

<u>Soil Pseudo R2</u>	<u>Vegetation R2</u>
➤Hartman: 0.48	➤Hartman: 0.71
➤Nichols: 0.84	➤Nichols: 0.64
<u>Bud Pseudo R2</u>	<u>Flower Pseudo R2</u>
➤Hartman: 0.32	➤Hartman: 0.67
➤Nichols: 0.36	➤Nichols: 0.80
2 nd Iteration	2 nd Iteration
➤Hartman: 0.38	➤Hartman: 0.72
➤Nichols: 0.38	➤Nichols: 0.82

Figure 2.10: The pseudo-R2 results from the random forest regression at both Field 1 and Field 2.

The 30 most important variables estimated by out-of-bag cross-validation (Breiman, 2001) from the RF regression on flowers for both Field 1, Field 2, GHS, and GHL are organized by increasing Mean Square Error in descending order as shown in Figure 2.11. The vegetation indices that are common to the top 10 between both fields are displayed in Table 2.2.

VI	Equation	Bands
GNDVI	$(\text{NIR} - \text{Green}) / (\text{NIR} + \text{Green})$	Green, NIR
FVI	$(\text{Blue}/\text{Green}) * (\text{NIR}/\text{Green})$	Blue, Green, NIR
CVI	$(\text{NIR}/\text{Green}) * (\text{Red}/\text{Green})$	Green, NIR
GRVI	$\text{NIR} / \text{Green}$	Green, NIR
GIPVI	$\text{NIR} / (\text{NIR} + \text{Green})$	Green, NIR

Table 2.2: A list of the best performing VIs that are common between both fields.

2.6 Discussion

When attempting to predict the onset of bloom at an alfalfa seed crop field, it would be ideal to track the development of buds as these are the precursor for the emer-

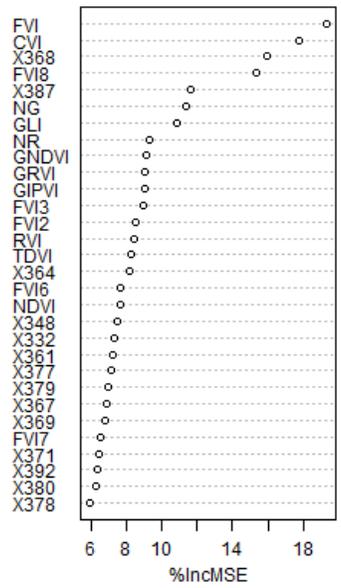
gence of flowers. However, the comparison of spectral signatures between buds and vegetation with a spectroradiometer shows the spectral difference is not sufficient to accurately differentiate between the two. This is further confirmed by the low R^2 values for buds (Figure 2.10). While buds and vegetation could likely be differentiated using spatial information in an object-based image classification method, this was not pursued because it would not be possible from satellite based observations due to the buds/vegetation being much smaller than the satellite spatial resolution. In the following chapters, the development of flowers and our ability to monitor alfalfa bloom using different electromagnetic wavelengths and wavelength combinations is explored.

The analysis of the relationship between percentage flower cover and hyperspectral data showed that it is possible to track the development of flowers using a spectral signature with an R-squared value as high as 0.82. The results of the regression show that many narrow band vegetation indices provide better percentage flower estimates than most of the individual wavelengths and bands. The top performing wavelengths are almost exclusively under 400 nm in the ultraviolet spectral range. While this is an interesting finding, those wavelengths would not be useful from satellite acquired imagery due to the significant atmospheric scattering at wavelengths below 400 nm and thus fraught with noise. The best performing VI are compared to determine which three bands are most commonly present (Table 2.2). The most significant VI for the highest R^2 valued location is the Green Normalized Difference Vegetation Index (GNDVI). This VI uses only two bands, Green and NIR. The best performing VI at the other field is the Alfalfa Flower Vegetation Index (FVI), which utilizes Blue, Green, and NIR. The field VIs are emphasized because they have the higher

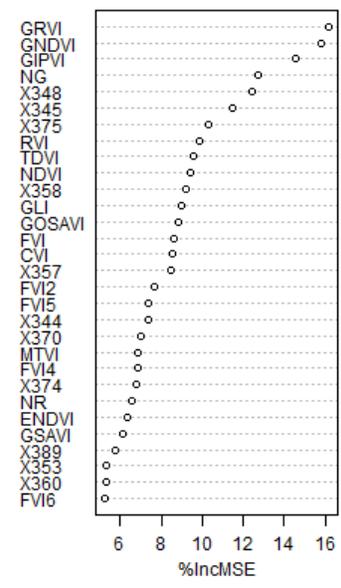
R-squared values. GNDVI and FVI both give the best results tracking at one field, but each VI is also significant at the other field and FVI was also the second most significant VI at both greenhouses (Figure 2.11).

2.7 Conclusions

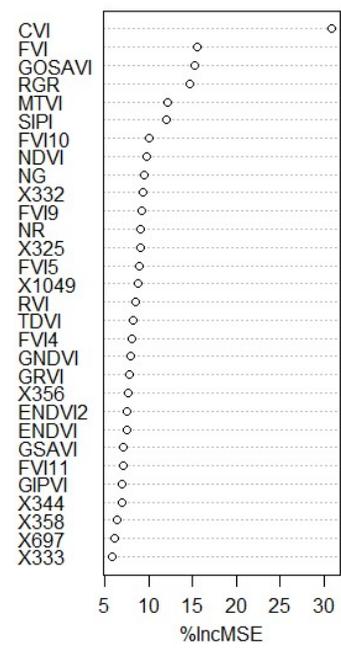
This study demonstrates the potential of tracking the development of bloom at alfalfa seed crop fields using only the spectral signatures and indicates that future work is warranted. The data tested identifies Blue, Green, and NIR as the three most significant bands for tracking the development of alfalfa seed crop flowers and provides a basis in the sensor development for the 2018 growing season.



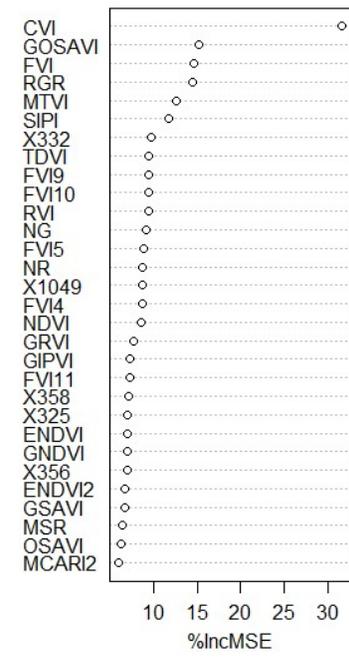
(a) Field 2



(b) Field 1



(c) GHS



(d) GHL

Figure 2.11: The 30 most important variables from the second iteration of the random forest regression for A) Field 1 B) Field 2 C) GHS and D) GHL.

CHAPTER 3: ACQUISITION OF NIR-GREEN-BLUE DIGITAL IMAGERY FROM FIELD STATIONS FOR CROP MONITORING

3.1 Summary

Near-surface remote sensing using radiometric instruments has great potential to improve phenological monitoring because automated observations can be made at high temporal frequency (Richardson *et al.*, 2009). This chapter focuses on the development of a near-surface sensor system to monitor the phenology of alfalfa as it progresses through a flowering cycle. The sensor collects temperature, humidity, solar irradiation, and digital images. Four sensors each with a camera that collects three bands (blue, green, and near-infrared) are deployed at two alfalfa seed crop fields located in southwestern Idaho during the 2018 growing season. The sensor provides daily RAW format imagery from three meters height.

3.2 Introduction

This project explores remote sensing instrumentation built for and used at alfalfa seed crop fields. The data collected will be used in a prediction model (Chapter 5) to provide insight about the alfalfa bloom cycle. The emergence of flowers at alfalfa fields is

dependent on environmental factors including temperature, photoperiod, and possibly soil moisture. Temperature is a first-order controlling variable in many crop growth models (Hodges, 1991) and is a primary determinant of alfalfa growth and development (Pearson & Hunt, 1972; Sharratt *et al.*, 1989). Ben-Younes (1992) determined that the development of alfalfa cultivars is hastened by warmer temperatures and shows there is a strong relationship between growing degree days (GDD) and alfalfa growth stage. Noland *et al.* (2018) performed a similar study using remote sensing methods to determine when to harvest alfalfa forage crops and including GDD as a variable increased prediction accuracy by up to 17%. The temperature-mediated phenological development suggests that including temperature as a covariate in the prediction model should improve the estimation results and so temperature measurements are collected at each field sensor to be integrated with the spectral measurements. Alfalfa is a long-day plant meaning the time to flowering decreases as the photoperiod is lengthened (Major *et al.*, 1990). The local solar irradiance is measured using a pyranometer to include the photoperiod response as a variable in the prediction model. Several researchers and agronomists suggest that mild water stress may increase plant development rate (Ottman & Putnam, 2017), however other researchers disagree (Halim *et al.*, 1989) and more quantitative data is needed (Hodges, 1991). Soil moisture data is collected by the field management company along side one field station and a soil moisture sensor may be incorporated into the field sensor in following years. The estimation of alfalfa bloom onset should include a spectral time-series, temperature, photoperiod, and possibly soil moisture. Targeted wavelengths along with the other environmental variables are collected using an automated low-cost sensor; enabling the potential for a real-time management response

to field conditions. This chapter includes details about the sensor development, the data collection process, and the data collected.

3.3 Field-Sensor Development

The automation of near-surface image collection used to monitor plant phenology development over time has numerous and significant advantages to manual collection practices. First, automating the data collection process means fewer workers on the ground and less hours spent commuting. The data can be more continuous and offers greater consistency in both the frame of reference and timing. Moreover, this collection approach can be incorporated into a larger work flow for near-real time assessment of vegetation status. The sensor we develop includes a MAPIR camera with blue, green, and NIR bands, a temperature-humidity sensor, a pyranometer, and a USB cellular modem to send daily data to a server.

A reliable power source is used to power the electronics at the remote field locations. A 20 W solar panel with an output of 17.1 V at 1.17 A is used. The solar panel is oriented south and connected to a compact voltage regulator. The Pololu synchronous switching step-down regulator measures 1.8×1.8 cm and delivers a typical continuous output current of up to 2.5 A at 5 V. This regulator features reverse voltage protection and has a maximum input voltage of 36 V, putting the solar panel input voltage above the 90% conversion efficiency for the regulator. The 5 V output of the regulator charges a portable USB battery bank and is connected by soldering the wires of a male micro USB connector directly to the output terminals on the regulator board (Figure 3.1). This provides power to the Raspberry Pi 3B (RPi), which runs the multi-sensor system (Figure 3.2).

The RPi board has a 40-pin header (Figure 3.3) that allows for sensors to be added.



Figure 3.1: The solar panel wires are soldered directly to the voltage regulator, which is then soldered to wires connected to a micro USB adapter. The micro USB adapter connects to a female micro USB adapter which goes to a male D/C adapter that plugs into the portable USB power bank.

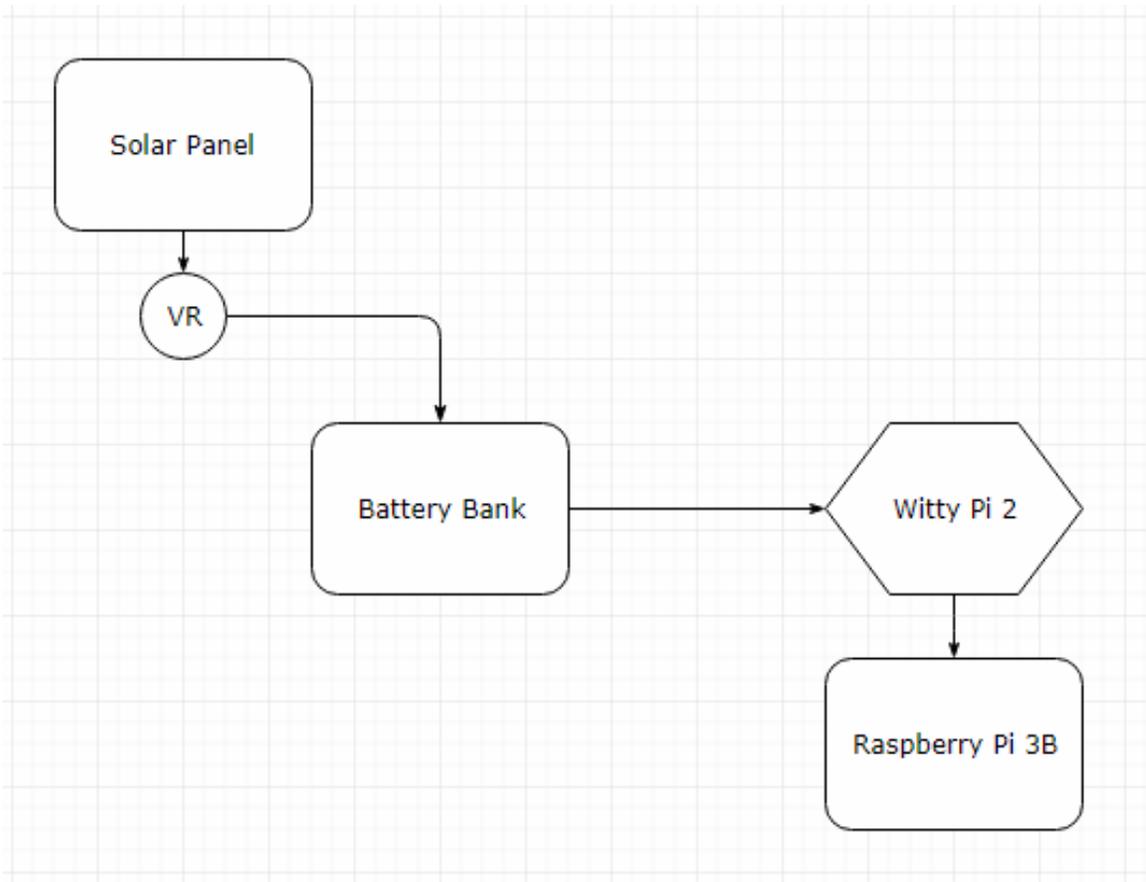


Figure 3.2: This diagram shows the circuit layout to charge the portable USB battery bank and power the Raspberry Pi 3B.

Component	Cost
20 W Solar Panel	\$62
Pololu Voltage Regulator	\$8
USB Battery Bank	\$69
Witty Pi 2	\$24
Raspberry Pi 3B	\$35
AM2302 T/H Sensor	\$15
Apogee SP-420 Pyranometer	\$312
MAPIR camera	\$410
Huawei E303 3G Wireless Modem	\$27
USB Y-connector	\$5
Electronics housing	\$35
32 GB Micro SD Card	\$12
64 GB Micro SD Card	\$20

Table 3.1: Breakdown of the cost of components to build this system.

The RPi controls the data acquisition and all of the peripheral sensor attachments such as a Temperature/Humidity (T/H) sensor, a MAPIR camera, and a USB cell modem. The controlling Python script is called via a bash script every time the RPi boots. A Witty Pi 2 board is attached to the RPi to keep the battery bank alive (see Section 3.3.1) and control the RPi boot sequence. The Witty Pi 2 board connects to the RPi via the 40-pin header and provides a *dummy-load* to keep most battery banks *On*. The Witty Pi 2 also provides a real-time clock.

3.3.1 Sensors

An AM2302 (wired DHT22) temperature-humidity sensor is connected to the RPi via the 40-pin header with the Yellow wire – Signal (Pin 13, GPIO 27), Black wire – Ground (Pin 6), and Red Wire – Power (Pin 2). This sensor can take measurements every 2 seconds and is good for humidity readings between 0-100% with a 2-5% accuracy and for temperatures ranging from -40° to 80°C with $\pm 0.5^{\circ}\text{C}$ accuracy. This sensor is compact with a body size of $2.7 \times 5.9 \times 1.35$ cm and fits under a small

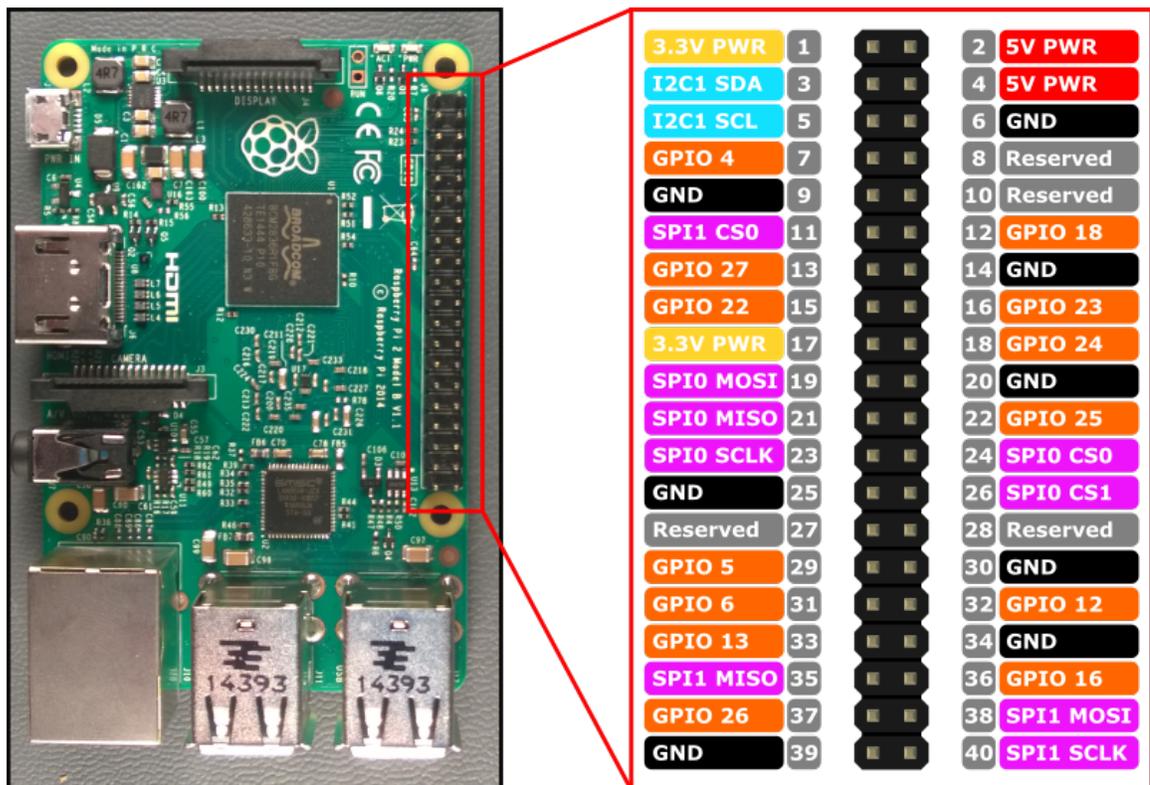


Figure 3.3: This image shows a RPi board and the 40-pin header, from Raspberry Pi Pin Map (2017).

cover attached to the main sensor enclosure.

In order to detect phenology changes by the spectral response of the field camera images a radiometric calibration is needed. Calibration is defined as the process of quantitatively defining the system responses to known, controlled signal inputs (Morain & Zanoni, 2004). In this case, the system response is the incoming electromagnetic radiation for the field camera and the controlled signal inputs is the solar irradiance provided by the pyranometer. An Apogee SP-420 Pyranometer is setup above the field camera enclosure and weighed 90 g with a body size of 2.4×3.3 cm height. It has a 180° field of view (FOV) with a resolution of 0.1 W/m^2 and measures a spectral range of 360-1120 nm with less than 2% long-term drift and less than 1% non-linearity up to 1750 W/m^2 . This pyranometer is connected to the RPi via a USB port.

The most important component in this multi-sensor field unit is the camera. Based on the previous field study, the camera needs to be robust enough to leave out in a field throughout the growing season, have consistent settings, take time-lapse imagery, and to collect spectral data in the blue, green, and NIR bands. Because the unit is setup a few meters above the field surface, the camera should have a low distortion, wide-angle lens to maximize the area accurately imaged. The MAPIR Survey3W camera meets all of these requirements. The Survey3W is a modified GitUp G3 action camera that can endure the weather conditions at an agricultural field during the growing season when in a proper enclosure. It is easy to fit this camera into an enclosure as it measures $5.9 \times 4.15 \times 3.6$ cm and weighs 50 g.

The camera settings are configured by MAPIR before being sent to the customer for optimal midday imaging and remain consistent unless changed. The Survey3W

also offers Pulse Width Modulation (PWM) triggering via the micro HDMI port for easy control of when to collect images. The Survey3W has a lens that provides well-defined spectral responses and can be used at close range. The 87° horizontal FOV (19 mm) Extreme Low Distortion (Non-Fisheye) Glass lens, which when mounted at 3 m height, provides an image covering an area of approximately 0.9×1.2 m with a resolution at ground surface of ≈ 0.14 cm/pixel. This camera captures three narrow spectral bands in the blue, green, and NIR regions (Figure 3.4).

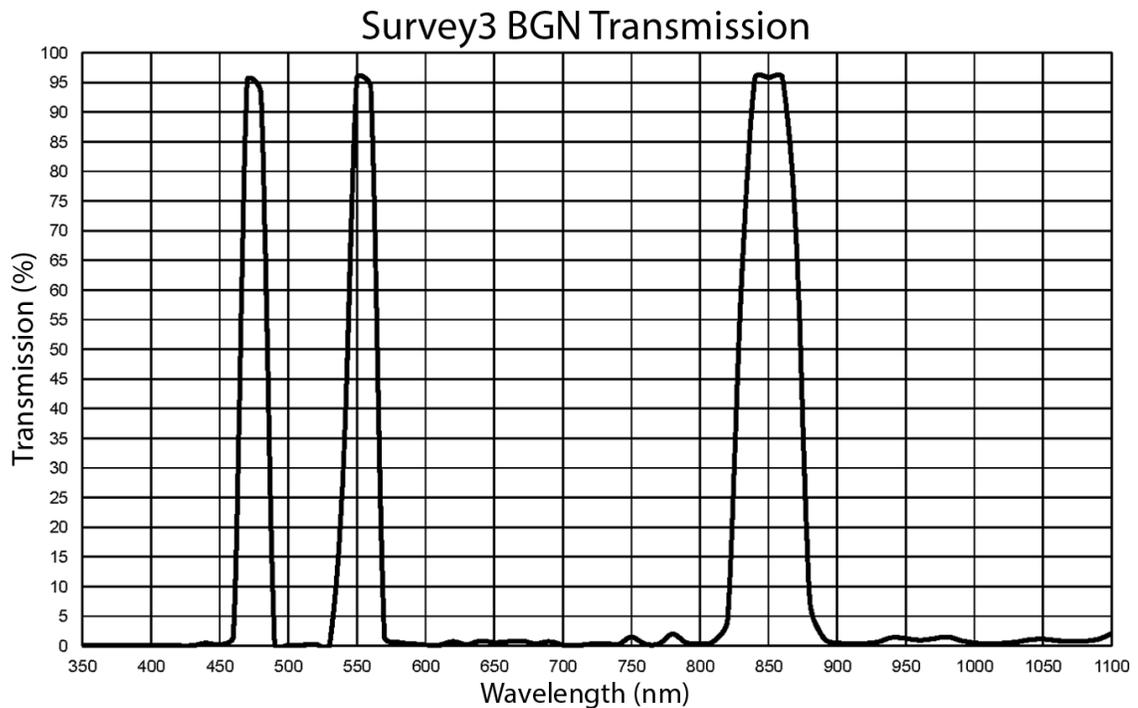


Figure 3.4: The MAPIR Survey3W blue (475 nm), green (550 nm), and NIR (850 nm) spectral responses, from MAPIR Survey3 Camera (2017).

The RPi is connected to the Survey3W camera using one USB port and two header pins (Figure 3.5). The USB cable goes from the RPi to the camera’s mini USB port to supply power when the RPi boots. The camera has a setting to turn on when power is supplied and this is enabled. The MAPIR PWM trigger cable is

three 20-gauge wires with the red wire (power) connected to Pin 19, the black wire (ground) connected to Pin 4, and the white wire (PWM signal) connected to Pin 2 on a male micro USB adapter. These wires are connected to the RPi 40-pin header in the following manner: the red wire is not used, the black wire is connected to Pin 34, and the white wire is connected to Pin 32, GPIO 12. The male micro USB is plugged into the Survey3 camera and the Python library RPi.GPIO (Croston, 2012) is used to control the PWM output.

The PWM output tells the camera to take an image and to enter Media Transfer Mode, which allows the camera to be mounted as a drive on the RPi. Once the camera is mounted as a new drive, the images are transferred and backed up to the RPi micro SD card. Once the images are backed up on the RPi, the camera drive is unmounted and media transfer mode is disabled. This is important because the Survey3W draws 0.2 A of the total 2 A available and is setup to automatically turn off after 3 minutes of no activity, but only after the media transfer mode is disabled will the camera turn off. After image backup, data is transferred via a cellular modem.

The Huawei E303 3G Wireless Modem draws up to 1 A when sending data. Because the USB ports are rated for 0.5 A, a USB Y-connector is used to supply enough power to modem. All of these components are housed in two different enclosures depending on which battery bank is used. The first two sensor units used the 22,000 mAh RAVPower battery bank, which has dimensions $16.5 \times 2 \times 6.6$ cm and weighs 405 g. These units are housed in 16 cm diameter PVC tubing enclosure (Figure 3.6). Dividers are 3D printed and set inside the PVC tubes to keep the components such as the battery, RPi, and camera separated. The camera lens is set up to point out the bottom of the enclosure through a snugly fit hole, the lens is then sealed from the

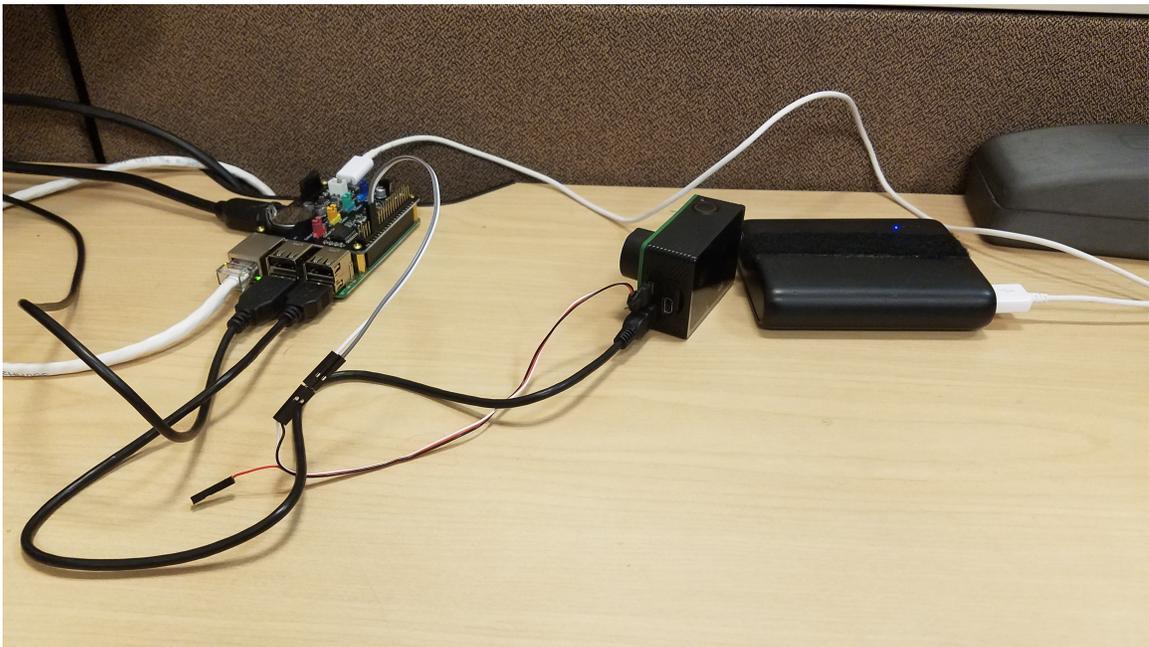


Figure 3.5: The RPi is powered by a USB battery bank and is connected to the Survey3 camera via a USB to mini USB cable and two header pins of a micro HDMI.

outside with silicone sealant to help prevent water infiltration. A second enclosure is used for the Voltaic V44 12,000 mAh battery, which is $10.4 \times 10.7 \times 1.9$ cm and weighs 312 g because it is shorter in length but longer in width than the other battery pack and would not fit into the PVC enclosures. This enclosure is a $0.30 \times 0.30 \times 0.15$ m weatherproof junction box (Figures 3.7 and 3.8). Once again the camera lens is setup to point out the bottom of the enclosure through a snugly fit hole; the lens is then sealed from the outside with silicone sealant to help prevent water infiltration. Four small holes are drilled on the top to fit two U-bolts through. These are then secured using a nut inside the box to tighten the enclosure to the outdoor structure and sealed from the outside with silicone sealant.



Figure 3.6: The multi-sensor unit in the field. This is the unit that uses the PVC tube enclosure. Triangular support arm and steel support cables are visible. This support system is used to emplace all field units, regardless of enclosure type.

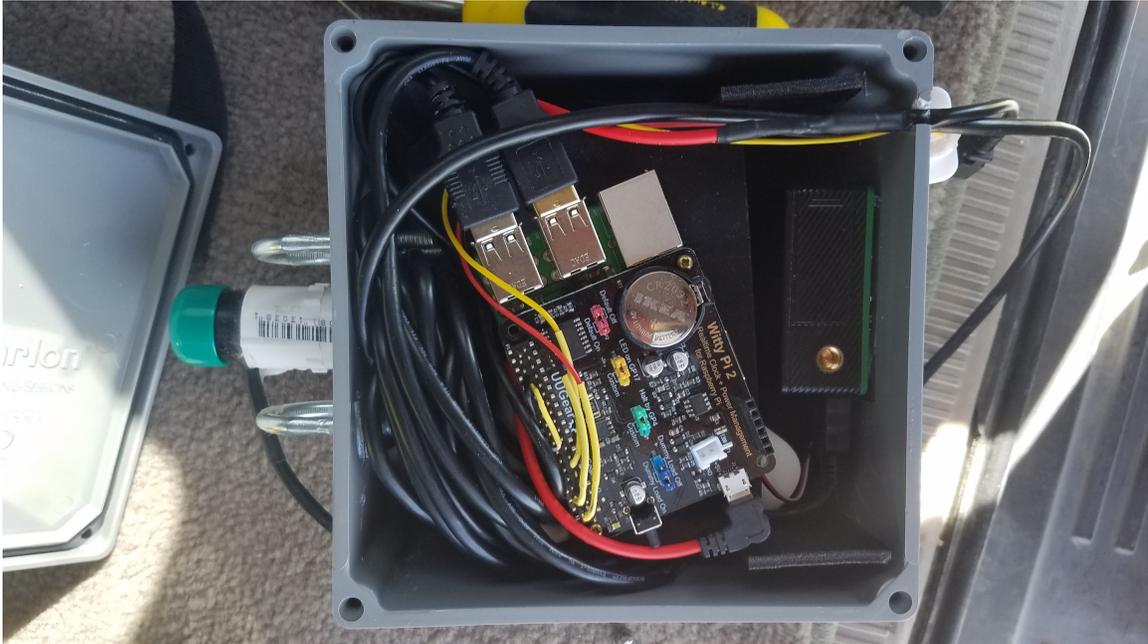


Figure 3.7: Interior of the junction box enclosure. The USB pyranometer has extra cable that is coiled inside this enclosure.

Mounting Hardware

The structure that holds the enclosures consists of a 10 cm diameter by 3 m tall galvanized steel fence post connected at a right angle to a 2.5 cm diameter by 1.5 m long steel tube. The 3 m tall post serves as the base for this structure and went over a 1.83 m tall T-post that is hammered with a fence post driver into the ground. To keep the post level, three 0.76 m T-posts are hammered into the ground at approximately 1.5 m from the post and 120° from each other. A steel cable is run from the smaller T-posts to a Zinc-Plated Turnbuckle Eye/Eye, which is then connected to a Galvanized Steel Tension Band wrapped around the fence post. The turnbuckle is used as a tensioner to keep the steel cables tight. The smaller steel tube acted as a boom to get the camera far enough away so the main fence post would not be in the image. This



Figure 3.8: Exterior of the junction box enclosure. A silicone sealant surrounds the camera lens to keep water out. The temperature/humidity sensor is zip-tied to the bottom of the enclosure to shield it from rain and direct sunlight, while remaining out of the image.

is connected to the main post using a Galvanized Pipe Grip Tie and secured to the grip tie using U-bolts. Another tension band is put on the main post a few feet below the boom, connected to a 1.22 m steel post, which is connected to the boom using a steel fence bracket to create a triangle support frame for the sensor unit (Figure 3.6).

Battery tests

The first rechargeable battery bank tested is a RAVPower 22,000 mAh power bank. This bank can output 2.4 A per USB port and unlike some USB battery banks, it is able to both output current and be charged at the same time. As with most portable USB battery banks, it has an *Auto-off* feature that will turn the power supply off when not enough current is being drawn. To counter this behavior, a Witty Pi 2 is used to provide a *dummy-load*. When the system is connected as shown in Figure 3.2 and tested outdoors under normal sun conditions, only two of the four battery banks are able to successfully remain on. The two working battery banks are used, however the two that would not stay on are replaced by a different battery bank. The Voltaic V44 12,000 mAh battery pack features an "Always-on" mode and is designed to be connected to a solar panel for time-lapse projects and outputs 2 A at 5 V. The voltaic battery bank has a couple of down sides for this application, such as being slightly under powered. This battery bank powers the RPi, which has a recommended 2.5 A at 5 V power supply, but because the RPi is often well below full CPU usage the Voltaic power output is sufficient. Another downside is that the recommended input is 2 A at 5 V, while the voltage regulator output is 2.5 A. Moreover, the Voltaic system requires a D/C plug to charge, so the provided charging cables are modified – the micro USB end of the charging cable is replaced with a 5.5×2.1 mm D/C power

in adapter. The different battery banks both perform well when test charging under full sun conditions.

3.3.2 Power Management

Precautions are taken to minimize battery usage and ensure proper power cycling, such as connecting a Witty Pi 2 power management board to the RPi. The RPi's recommended power supply is 2.5 A at 5 V, but the Voltaic battery is only able to provide 2 A, which is thus considered our maximum possible power draw. When idle, the RPi draws 260 mA and at full CPU usage draws 730 mA (Geerling, 2017). The RPi also powers all of the peripheral sensor components, which each have their own power draw. The AM2302 T/H sensor uses a max current of 2.5 mA while taking readings. The MAPIR Survey3W camera consumes 0.2 A while turned on. The Apogee SP-420 pyranometer has a 2.1 mA current draw when recording (Apogee Instruments, Inc, 2018). The USB modem is the most power consuming peripheral at upwards of 1 A when sending data. The equation below shows that the maximum power draw remains below the maximum that the battery bank can provide. We note that the RPi is never close to the maximum CPU usage when performing tasks associated with the field unit, and the camera always shuts off before sending data over the USB cellular modem.

$$\begin{aligned} \text{MaxPower} &= 730mA + 2.5mA + 0.2A + 2.5mA + 1A \\ &= 1.935A \end{aligned}$$

Considering the above configurations, at maximum power consumption the 12,000 mAh

battery bank will last six hours. The system needs to be powered on for long enough to take an image, transfer it to the RPi, log the T/H and solar irradiance data, and once a day send the data back to our server. The Witty Pi 2 power management board ensures the system is fully shut off after a certain amount of time and that it turns back on as defined by the programmed schedule. Overall, the system is *On* for a maximum of 90 minutes a day. This means the battery pack should last 4 days without being recharged. There is a continuous power draw even when the system is turned *Off*. When completely *Off* the RPi draws 20-30 mA Geerling (2017); however, with the Witty Pi 2 board installed the power draw without a *dummy-load* is 1 mA and with a *dummy-load* is 15 mA. With the *Always-on* Voltaic battery pack there is no need for the *dummy-load* and the 1 mA vampire consumption is negligible.

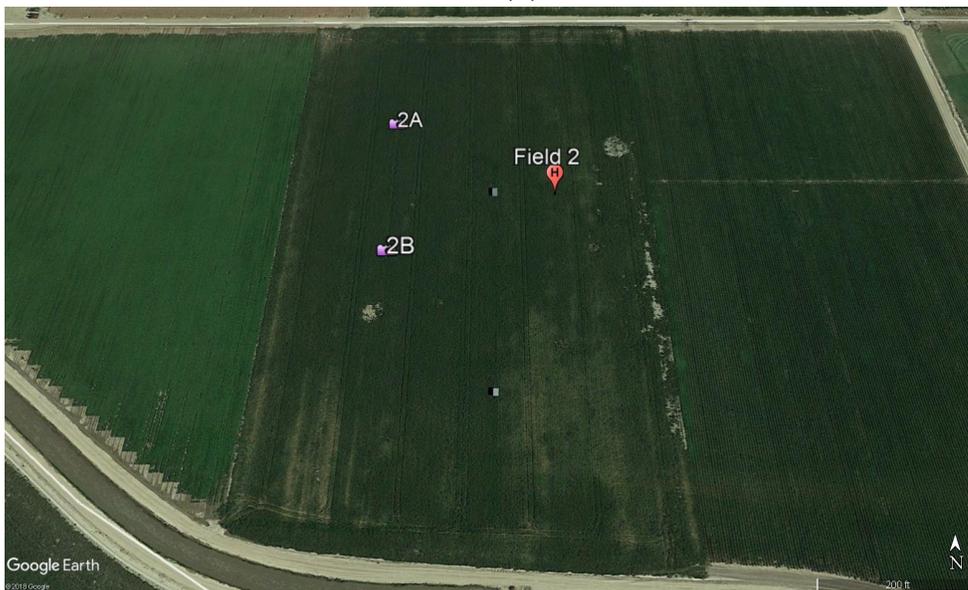
3.4 2018 Field Season Data Collection

3.4.1 Location

The sensors are setup during the 2018 growing season at two alfalfa seed crop fields (Figure 4.1) located in Southwestern Idaho in the Nampa and Parma agricultural region after their final cutback, which is when the entire field is cut back and then allowed to grow for the rest of the season. The first field, referred to as Field 1, is cutback on 1 May 2018 and the second field, referred to as Field 2, is cutback on 8 May 2018. The first sensors (1A and 2A) begin collecting data on 14 May 2018, with the second sensors (1B and 2B) starting on 7 June 2018. The sensor naming convention is a combination of the field number and the first deployment (A) or second deployment (B) and their locations within each field is shown as a camera icon in Figure 3.9.



(a) Field 1



(b) Field 2

Figure 3.9: Overview of the fields with the locations of sensors A and B during the 2018 data collection.

3.4.2 Data Collection

For this study, the RPi is turned *On* and *Off* at seven pre-determined times throughout the day and performed different tasks depending on the time of day. The Sur-

vey3W is sent a PWM signal to take an image five times a day, everyday at exactly 10:00, 11:00, 12:00, 13:00, and 14:00 hours. To ensure that the sensor is working when deployed in a distant location, a USB cell modem connection is used to send data to a research server hosted by Boise State University. Everyday at 12:00 the Huawei E303 3G Wireless Modem is used with a GSM cellular data plan to send the RAW and JPG images taken at 12:00, along with the log file, the Temperature/Humidity data file, and the solar illumination data file, back to our research server. The log file provided information about whether the other images are successfully transferred and stored on the RPi or if any errors during the data collection the process. Having the system connected to the internet via a cellular data plan, even without a static IP, allows the Python control code to be updated if/when errors are detected. The RPi is also booted up at 15:00 hours, but instead of collecting an image, it uses the cellular data connection to check if there is a newer version of the Python control code on the research server. It then ensures that the 12:00 imagery, the log file, the T/H data file, and pyranometer data file are all backed up on the research server.

Two sensor units (A and B) are deployed at each alfalfa seed crop fields (Fields 1 and 2) in southwestern Idaho during the 2018 growing season. Sensor units A used the PVC tube enclosure and sensor units B used the junction box enclosure. The station locations are determined mainly to keep the sensors out of the way of field machinery and away from the field edges. Spectral ground-truth data is collected throughout the growing season using a field spectroradiometer (ASD FieldSpec 4). This high-resolution spectroradiometer collects data from 350 to 2500 nm with a spectral resolution of 3 nm at 700 nm and 8 nm at 1400 to 2100 nm. This sensor is radiometrically calibrated and the data are used for sensor validation of the field

cameras. Data is collected with the ASD FieldSpec 4 ten times throughout the 2018 growing season. Figure 3.10 displays the data collection time-line, as well as the pollinator release date at each field.

The final cutback, which is when the entire field is cut to a certain height and then allowed to grow for the rest of the season, for Field 2 is on 11 May 2018. The first sensor unit at this field began collecting data on 15 May 2018 with the second sensor setup on 7 June 2018. The pollinators are released at this field on 17 June 2018, and a large storm hit the field on 18 June 2018. This storm is the first real test of the sensor enclosures and causes the first sensors deployed at each field to go down, as is shown in red on Figure 3.10. The first version of the sensor enclosures, which uses the 16 cm diameter PVC tubing, are not water tight and let water in during the storm. The sensor at Field 1A is re deployed a week later on 26 June 2018, but the second sensor from Field 2A needs a replacement Witty Pi 2 board and is redeployed in a junction box enclosure 1 July 2018. From then on, the sensors collected data uninterrupted until the day before seed harvest on 23 August 2018.

3.5 Results

3.5.1 Field Imagery

The Field 1B sensor collected three-band imagery continuously throughout this study and the images collected at 12:00 are used as an example for the data collected. After the vignette correction (see Section 4.3.2) the three-band images are converted into a single band Vegetation Index that comes from the 2016 field data analysis (see Section 2.6). For this example we will focus on the Flower Vegetation Index (FVI) given by equation 3.1). When the vignette correction is applied to the TIFF images

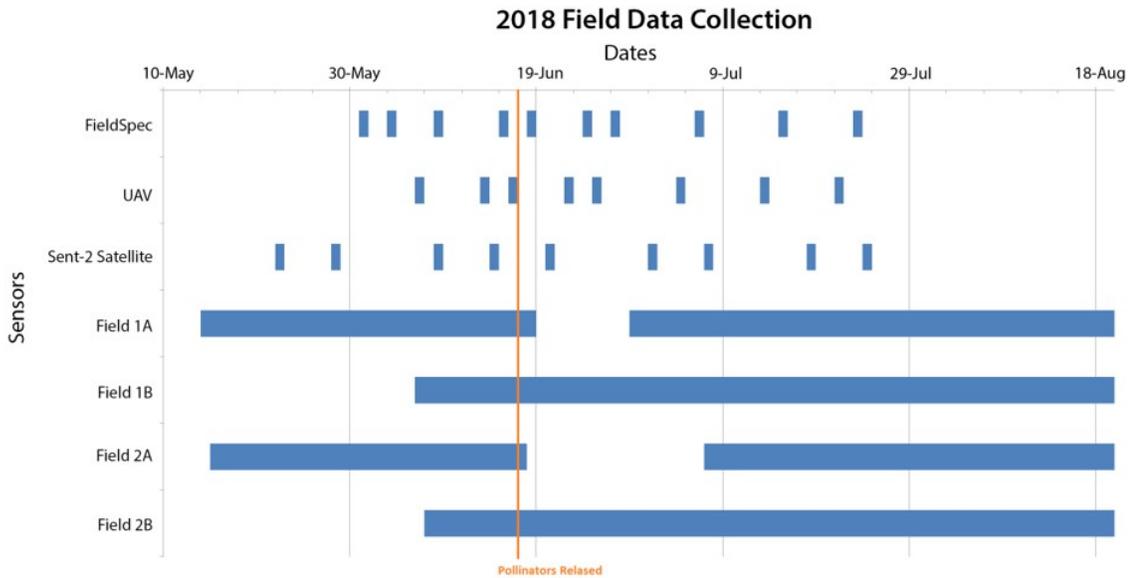


Figure 3.10: Time-line for the data collected during the 2018 growing season. With the pollinator release date for Field 2 shown as a vertical orange line.

a more accurate vegetation indices can be extracted. We computed the FVI at every pixel in the image and then computed the mean value for the image. The results from the growing season are presented in Figure 3.11.

$$FVI = \left(\frac{Blue}{Green}\right) * \left(\frac{NIR}{Green}\right) \quad (3.1)$$

3.5.2 Temperature and Humidity

The low-cost AM2302 temperature and humidity sensor data are compared to data collected at a nearby Davis Vantage Pro2 Plus weather station. A visual comparison of these two sensors outputs is presented in Figure 3.12. Both the temperature and humidity data track the continuously recorded Pro2 Plus data. This indicates that the low-cost AM2302 can accurately record environmental data critical to alfalfa seed

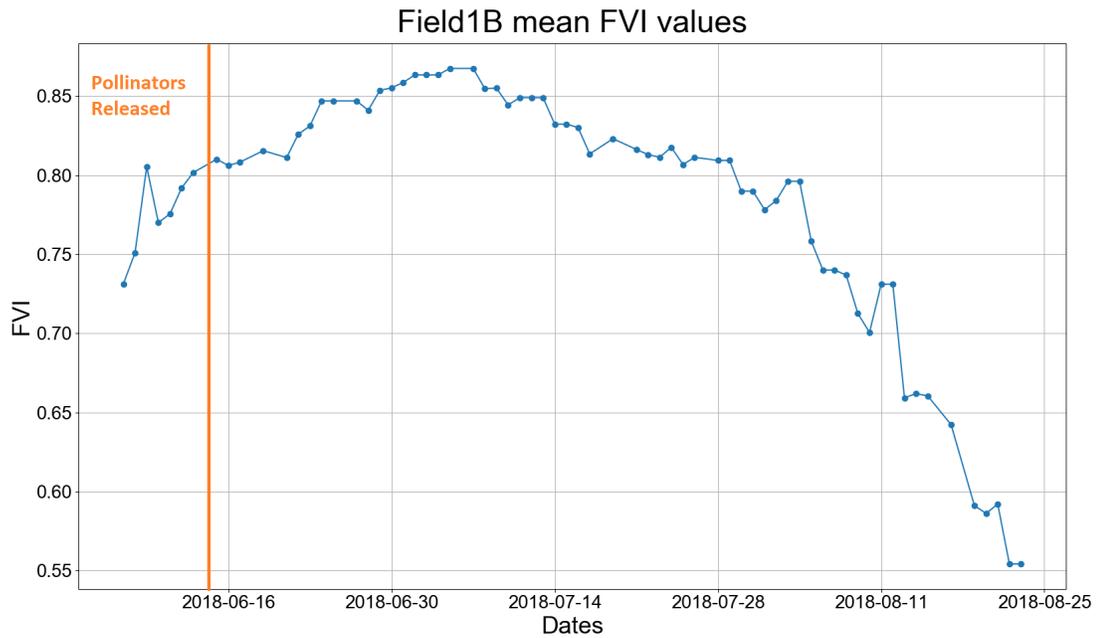


Figure 3.11: Preliminary results from Field 1B imagery. Vignette corrected TIFF images are converted into a Flower Vegetation Index (FVI), from which the mean value for each image is plotted through time over the growing season. The pollinator release date for Field 1 is indicated by the vertical orange line.

crop growth and development.

3.5.3 Solar Irradiation

Solar irradiation information is collected every second for one minute before and after the field camera takes a picture. This data is then averaged into a single data point to represent the incoming solar irradiation at the time the image is collected. A nearby Davis Vantage Pro2 Plus weather station collected continuous solar irradiation data. The Davis precision pyranometer and Apogee SP-420 pyranometer comparison is shown in Figure 3.13. We observe that the trends in the daily variations align; however, there appears to be a constant shift between the two data sets. The different readings between the two sensors is likely due to differences in the wavelengths they measure – Apogee [360–1120] and Davis [400–1100] nm.

3.6 Discussion

The sensors that were enclosed in a junction box collected data during the entire growing season while the sensors enclosed in the PVC tubing malfunctioned during important growing days due to water damage caused by a storm. The data collection results presented in Figure 3.11 show the FVI starting at a value of 0.72 with a maximum value of 0.87. The goal for this sensor is to observe the FVI starting at a low value and increase as more flowers begin to bloom, and this high starting value is likely because the sensor was not deployed early enough in the season. We also observe a dip in the FVI values when pollinators are released on 15 June 2018, which could be an indicator that the alfalfa flowers have been pollinated. The FVI pattern suggests that phenology development is captured in the spectral signatures and is a good start for further analysis and flower development monitoring.

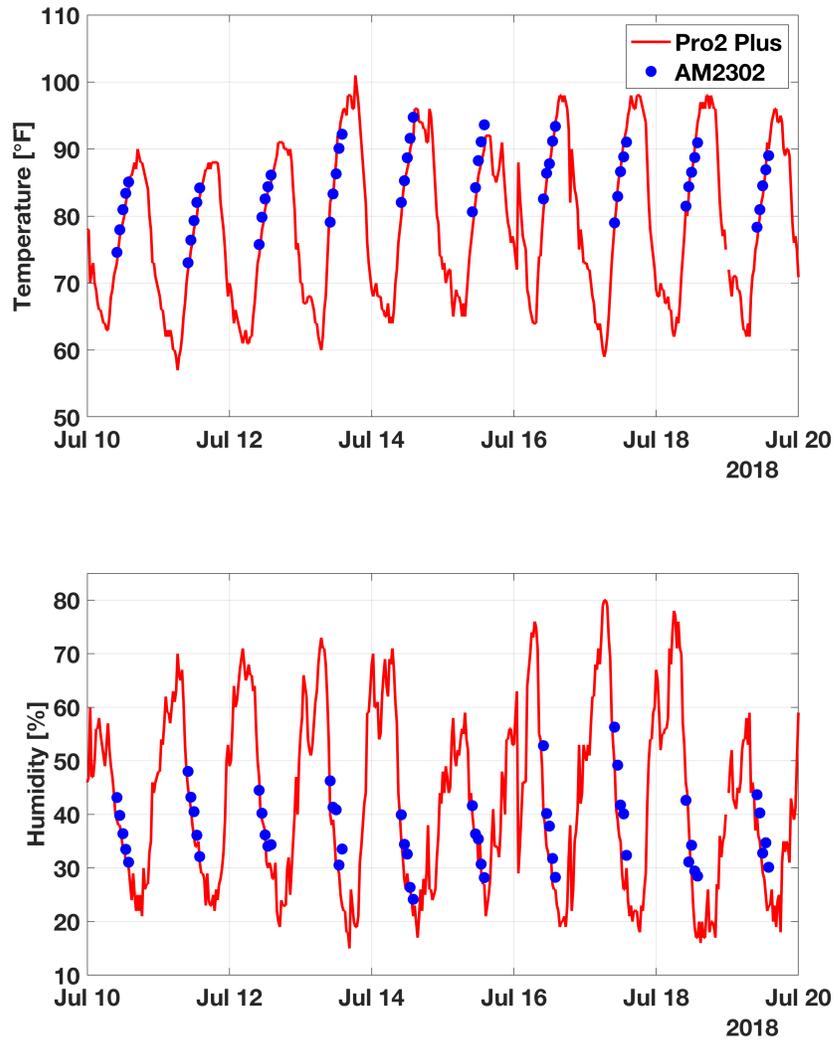


Figure 3.12: Comparison of temperature (top) and humidity (bottom) data recorded with the low-cost AM2302 compared to high-quality *Davis Vantage Pro2 Plus* continuous recording weather station.

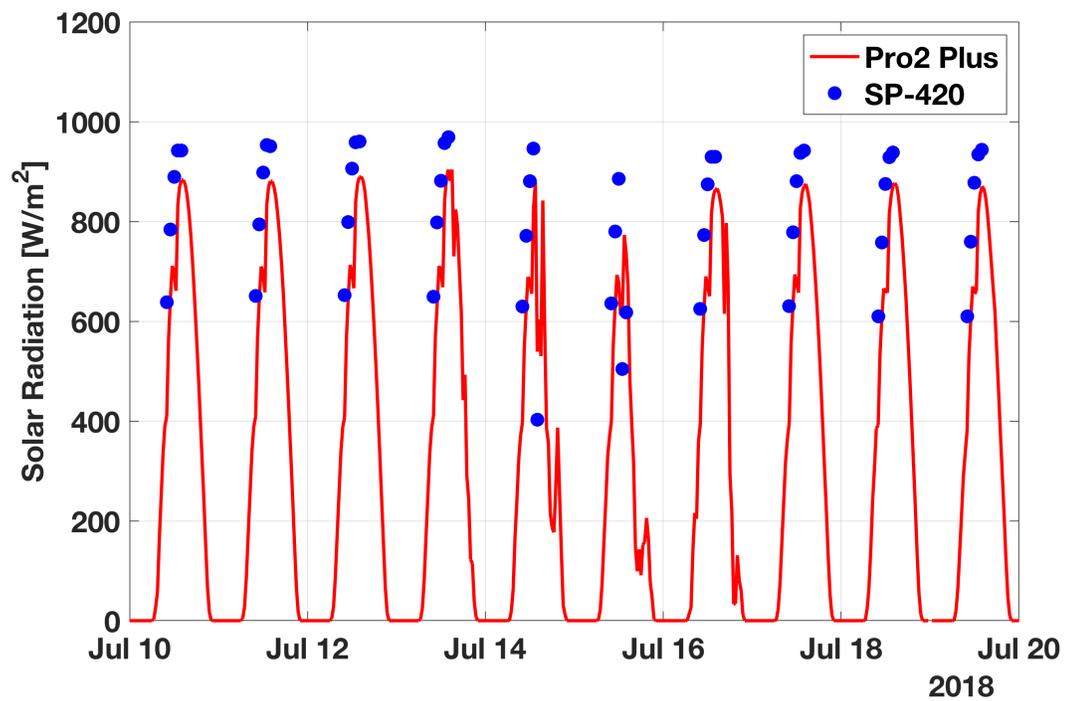


Figure 3.13: Comparison of solar radiation data recorded with the Apogee SP-420 pyranometer compared to high-quality *Davis Vantage Pro2 Plus* continuous recording weather station.

Recommended changes to the field sensor include a soil moisture sensor and a small RGB camera. An important thing to consider is that while the NIR-modified camera collects the spectral wavelengths of interest, it does not readily facilitate manual point classification of soil, vegetation, buds, flowers, and seeds in the imagery. Point classification is vital to performing any type of regression or quantifying how well the imagery captures phenological development. In order to quantify the quality of data collected during the 2018 growing season, the false color images need to be classified into percentage cover, which will be done manually or by using a supervised or unsupervised classification algorithm (see Chap 4). An RGB image that coincides with the B-G-NIR image would make the classification process straightforward and provide more accurate classification.

The Survey3W camera saves JPG images with 8 bits per channel, but can also save RAW format images with 12 bits per channel. A RAW image file contains the greatest amount of information, but most importantly the RAW imagery stores the original sensor response allowing for greater control of the image quality. Analyzing the original response can lead to better results (Verhoeven, 2010) when using spectral changes over time to infer something about the area being monitored. The downsides include a larger file size, which takes longer to save and transfer images, larger disk space requirements, and preprocessing including converting the images to TIFF format and applying the vignette correction. The ability to customize the spectra collected by switching camera models, a wide non-distortion lens, and camera control with a PWM signal make this camera a good choice for budget minded near-surface studies. But for accurate analysis the vignette correction needs to be determined by the user. This may or may not be something the user wishes to do,

and other, more expensive cameras may offer on-board vignette corrected data or higher bit resolution.

In a simplified process, the amount of incoming solar irradiation is directly related to how much is reflected off of the surface and captured by the sensor. A pyranometer is used to measure the incoming solar irradiation to perform a radiometric calibration for the field camera images. Much of the noise in Figure 3.11 can be attributed to changes in the incoming solar radiation when the images are collected. However, a pyranometer is not an appropriate tool to use for this as it collected a single value for the entire spectrum from 350 nm to 1100 nm. Ideally, this would be replaced with a down-welling light sensor that is sensitive to each of the same bands that the camera is collecting.

Finally, a larger than required solar panel is used because it allowed the panel to have less than optimal output and still provide an adequate power supply. It is also readily available from previous field work. The field sensor system has very low power requirements and a smaller solar panel can be used to charge the battery banks. The Voltaic V44 battery bank has a lower capacity than the other battery bank, and is rated for a slower charge rate with a lower output; however, the *Always-On* feature more than made up for these disadvantages.

The first two field sensors deployed (1A and 2A) were taken out by a storm right as the pollinators were released on the fields. This timing was less than ideal but they still managed to capture the majority of the bloom up period. The second set of sensors deployed (2A and 2B) did not begin collected data until just about a week before pollinator release. This dataset did not capture the full bloom up period but was able to continue collecting after the pollinators were released.

This type of field sensor system can be used to monitor a variety of vegetation types by simply switching the camera to collect whatever three spectral bands are required. One limitation of this system is that it only collects three spectral bands, however with some ingenuity or a different camera this could be overcome. A major advantage of this low-cost system is that the base design allows for additional components to be added as needed.

3.7 Conclusions

The four field stations were setup in time to capture the majority of the bloom up period during the 2018 growing season. The four stations saved a total of 48.3 GB including a total of 1079 images taken and properly saved. These images are processed (see Chapter 4) before use in a bloom prediction model (see Chapter 5).

CHAPTER 4:

IMAGE PROCESSING AND CLASSIFICATION

4.1 Summary

The images collected during the 2017 and 2018 growing seasons both require a vignette correction and radiometric calibration before use in a statistical analysis. The 2017 growing season used a Micasense Rededge multispectral camera. This camera requires a band alignment due to use below the recommended height, along with a vignette correction and radiometric calibration.

During the 2018 growing season a MAPIR Survey3W NGB camera collects RAW+JPG imagery. This is a single lens camera which fixed the band alignment issue from the previous year, but the imagery still requires a vignette correction and a radiometric calibration before use in a image classification or as a Vegetation Index (VI). The calibrated false-color images recorded by this sensor are not conducive to SamplePoint classification and so a Support Vector Machine (SVM) supervised image classification estimates the percentage cover of flower throughout the growing season. By performing these image processing techniques, we ensure that we have the best data possible for statistical analysis.

4.2 2017 Data

4.2.1 Introduction

The 2017 imagery is collected during the growing season at two alfalfa seed crop fields from June 7th to August 10th (40 days) and 47 days of imagery at the greenhouse from July 7th to August 23rd. The fields and greenhouse are located in Southwestern Idaho in the Nampa and Parma agricultural region (Figure 4.1). The two fields are imaged at 3 locations and the greenhouse has 6 cohorts with 6 plots each. A digital RGB camera is used to collect data for SamplePoint image classification. A Micasense Rededge multispectral camera is used to collect the spectral progression of alfalfa seed crop progression through a floral cycle.

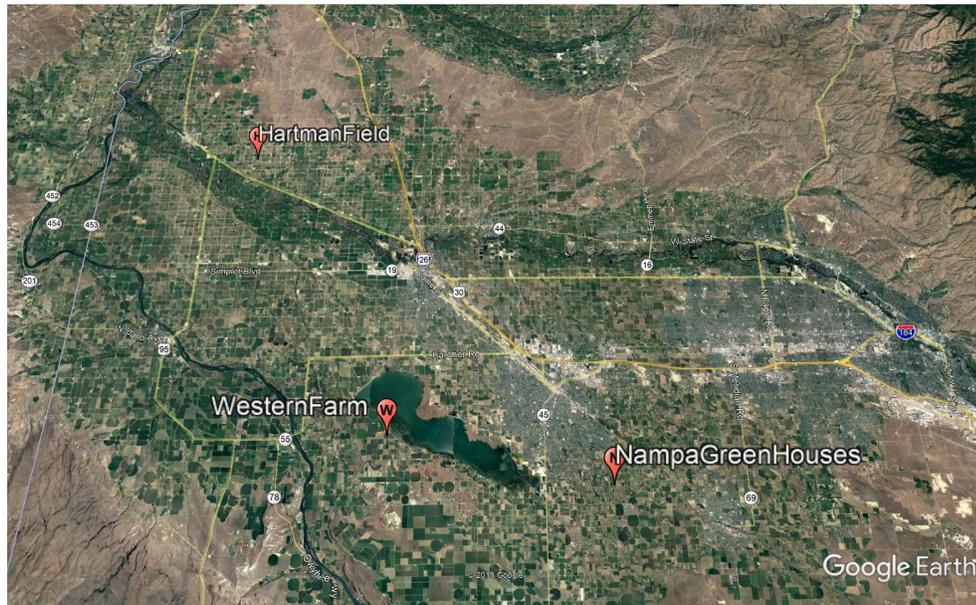


Figure 4.1: The fields and greenhouse locations for the 2017 data collection.

The Micasense Rededge camera (Figure 4.2) is 12.1 cm long x 6.6 cm wide x

4.6 cm deep and weighs 150 grams (Base, 2015). It is a multispectral camera that simultaneously captures five discrete bands. Each of the five lenses collects a different narrow-band electromagnetic spectra including: Blue, Green, Red, Rededge (RE), and Near-infrared (NIR) (Figure 4.3). The narrow bandwidths at specific wavelengths collect precise and quantitative spectral information about the crop in the image. The focal length of each lens is 5.5 mm, and the Field of View (FOV) is 47.2° with a resolution of 1280 x 960 pixels. The linear dimension of a single pixels projection on the ground is the Ground Sample Distance (GSD), a measure of the spatial resolution of an image (Comer *et al.*, 1998). The Micasense camera is setup at approximately 2 m height, giving a GSD of 0.14 cm/pixel. We now present the processing that we did to prepare the Micasense imagery for classification analysis.

4.2.2 Methods

In order for the imagery to be comparable to other sensors the images are converted into reflectance using a radiometric calibration. Calibration includes (1) converting the 16-bit digital numbers into values of radiance with units of $\mathbf{W/m^2/sr/nm}$, (2) applying a radial vignette model to correct for the fall-off in light sensitivity in pixels further from the center of the image with the manufacturer provided values, and (3) performing a reflectance calibration using the reflectance panel. The reflectance panel is tested by the manufacturer to determine the reflectance across the spectrum of light captured by the Micasense camera and should compensate for the lighting conditions at the time of image capture by taking a picture of the panel before data collection and then applying the reflectance conversion in post-processing.

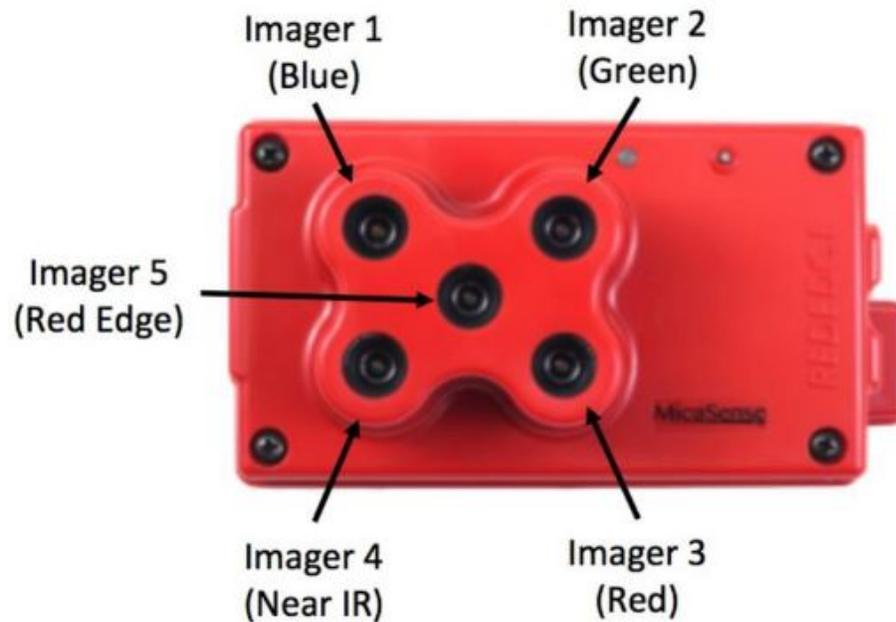
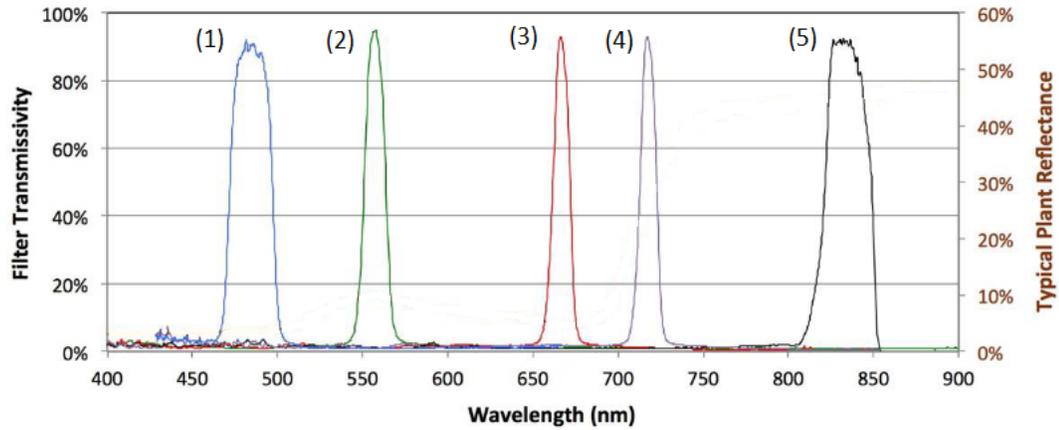


Figure 4.2: Micasense Rededge Imager Nomenclature shows that the RE band is in the middle and is used for the registration of the other bands (Base, 2015).

Image Calibration

The Micasense Rededge Image Processing workflow (Figure 4.4) is implemented using the Python programming language following the manufacturers guidelines. “The RedEdge radiometric calibration converts the raw pixel values of an image into absolute spectral radiance values. It compensates for sensor black-level, the sensitivity of the sensor, sensor gain and exposure settings, and lens vignette effects. All the parameters used in the model are read from the XMP metadata inside the TIFF file” (Base, 2019). First the images are normalized to have a value between 0 and 1 by dividing the raw pixel value by 2^N where N is the bit value, so for a 16-bit TIFF image each pixel is divided by 65536.



Band Number	Band Name	Center Wavelength (nm)	Bandwidth FWHM (nm)
1	Blue	475	20
2	Green	560	20
3	Red	668	10
4	Near IR	840	40
5	Red Edge	717	10

Figure 4.3: Spectral response curves for the Micasense Rededge camera (Tagle, 2017) with the color of the lines corresponding to the camera band in the table below.

Next the vignette correction model (Figure 4.5) is applied to the normalized image arrays (Base, 2019). The six polynomial coefficients k_n and the vignette center of the image c_x and c_y are read in from the metadata and applied in the following formulas:

$$r = \sqrt{(x - c_x)^2 + (y - c_x)^2},$$

$$k = 1 + k_0 * r + k_1 * r^2 + k_2 * r^3 + k_3 * r^4 + k_4 * r^5 + k_5 * r^6,$$

$$I_{corrected}(x, y) = \frac{I(x,y)}{k},$$

where

r is the distance of the pixel from the vignette center in units of pixels,

(x,y) are the coordinates of the pixel being corrected,

k is the correction factor by which the raw pixel value should be divided to correct for sensor falloff,

$I(x,y)$ is the original intensity, and

$I_{corrected}(x, y)$ is the corrected intensity.

The pixel values are then converted into radiance. First the black level value is defined as 4600 and then normalized by dividing by 65536. The gain, the three calibration coefficients (a_n), and exposure time are extracted from the metadata and used in the following formula:

$$L = V(x, y) * \left(\frac{a_1}{g}\right) * \left(\frac{(p-pbl)}{t_e+(a_2*y)+(a_3*t_e*y)}\right),$$

$$V(x, y) = 1/k$$

where

p is the normalized raw pixel value,

p_{BL} is the normalized black level value,

a_1, a_2, a_3 are the radiometric calibration coefficients,

$V(x, y)$ is the vignette polynomial function for pixel location,

t_e is the image exposure time,

g is the sensor gain setting,

x, y are the pixel column and row number, respectively, and

L is the spectral radiance in **$W/m^2/sr/nm$** .

Finally, the calibrated reflectance panel is used to convert the images into reflectance. An automatic panel detection method would not reliably detect the panel in the image and so the values are manually extracted using a custom Python script. The script first performs the radiometric calibration and image registration. It then combines the RGB bands into a color image to display. The user is then able to click in the center of the reflection panel where a box that is 80 x 80 pixels is drawn with the mouse click as the center of the blue outline as shown in Figure 4.6. If the box is well within the panel, close the figure window and the mean values for each band from the boxed area are saved (Figure 4.7) and later applied to the rest of the image set using the following transfer function:

$$F_i = \frac{p_i}{avg(L_i)},$$

where

F_i is the reflectance calibration factor for band i ,

p_i is the average reflectance of the calibrated reflectance panel for the i -th band, and

$avg(L_i)$ is the average value of the radiance for the pixels inside the panel for band i .

“This factor can be used for the i -th band to convert all radiance values to reflectance by simply multiplying the radiance values of any image by the factor F_i . This same process can be applied to each of the 5 bands independently to convert the images for each band to units of reflectance” (Base, 2017).

Image Registration

Because the multi-lens Micasense was used at 2 m height, image registration to fix the slight offset between each band in an image is required. Image registration is

the operation of aligning two or more images of the same scene. A Discrete Fourier Transform based technique brings the separate images from each band into alignment. In this method a phase correlation determines the translational movement on the order of 1 pixel registration accuracy (Srinivasa Reddy & Chatterji, 1996; Averbuch & Keller, 2002).

After the images are converted into radiance they are processed in a Python library called `Imgreg_DFT (IRD)` (Gohlke & Tyc, 2018), which uses a phase correlation to shift one image to match the other. The initial band mismatch is shown in Figure 4.8 as an RGB composite. The Red Edge band is the template to match because it is in the center of the other 4 bands. The `IRD.translate` function determines how much to shift each image. This function is modified from the default by increasing the Hanning window filter, which further tapers the shifted cross-power spectrum. Because the images are all going to have a very small shift due to the fixed lens locations, this is later hard-coded to 1/6th of the shifted cross-power spectrum smallest axis in both directions as shown in Figure 4.9.

4.2.3 Results

The Shifted Cross-Power Spectrum shown in Figure 4.10 has a definitive peak where the images match. The location of the maximum value determines how much to shift the Blue image. The RGB composite after image registration is shown in Figure 4.11.

The translation function is applied to all the images in the dataset. The distribution of X and Y axis shift values for the blue band compared to the RE band is shown in Figure 4.12. Here we can see that each axis has a normal distribution of determined shift values. The X-axis has an average shift value of 6 pixels and the Y-axis has an average shift value of 24 pixels. This difference in the shift amount is also

shown in Figure 4.10, where the peak is always fairly close to the X-axis and slightly further away on the Y-axis from the center of the image. Figure 4.13 demonstrates how the image processing techniques impacts the VI progression over the course of the growing season at the two 2017 fields.

4.2.4 Discussion

Converting the Raw TIFF images into calibrated radiance images gives the VI time-series a more definitive trend and corrects for variations due to environmental conditions (e.g. cloudy vs. clear day). The use of a calibrated VI time-series changes the shape of the curve into what is likely a more reproducible signal.

The results of the image registration provided a consistent method to translate each band into better alignment with the RE image as shown by the normal distribution of the histograms for the X and Y image offsets. Because the lens are fixed in their relationship with each other, I would expect the translations to remain constant, meaning for large datasets it may be possible to find the mean translation value after a certain number of images. These translation values can then manually be applied to translate each image without having to run a fast fourier transform for each band for each image. This would greatly speed up the image processing workflow. Using the multi-lens camera at such a low height also caused lens distortion to be prevalent and is likely the cause for much of the uncertainty in the image registration process.

Future studies could consider using photogrammetry software to model the amount of lens distortion present, and once the lens distortion is quantified, each band could be further corrected before the initial alignment. Instead the image registration is further improved upon by breaking the images up into smaller blocks and translating those blocks, keeping track of the translations and carefully recombining. The

block registration and recombination method is successful in improving the registration results; however, it is not applied to the final VI time-series due to processing time constraints. A visual inspection of Figure 4.13 shows that the image registration process did not significantly alter the VI time-series progression.

The average reflection panel values varied from day to day. Whether these variations are caused entirely by solar illumination is questionable. The images of the reflection panel are fairly consistent over the course of the growing season; however, there are occasionally miscellaneous objects in the image, such as blue jeans or an orange tape measure which could have influenced the reflection of the panel. Over the course of the growing season the panel also became quite dusty and a consistent procedure for cleaning the panel is not documented.

The reflection panel values are applied to the aligned radiance images to convert them into values of reflectance. This should be the most reliable source of data for the field collection campaign because it is calibrated to the current solar illumination conditions. While some of the noise from the aligned radiance trend seems to have been suppressed, it is apparent that using the reflection panel did not significantly alter the VI trend. The cause of the large amount of variability in the day-to-day reflection panel measurements remains unknown.

4.3 2018 Data

4.3.1 Introduction

The 2018 data set is collected at the same fields as in 2017 located in Southwestern Idaho in the Nampa and Parma agricultural region (Figure 4.14). Greenhouse data is not collected during this growing season. Each field is equipped with two MAPIR

Survey3W BGN cameras in a custom field sensor enclosure, and RAW + JPG images are recorded throughout the growing season as shown in Table 4.1.

Dates	Images
Field1A 5/14 - 8/23	48
Field1B 6/7 - 8/23	61
Field2A 5/15 - 8/23	45
Field2B 6/7 - 8/22	77

Table 4.1: Date ranges the field cameras collected data.

Although the JPEG file format is a commonly used file format, the automatic application of amplitude corrections such as white balance and gamma corrections, and compression changes the spectral intensity values in an unknown manner and should not be used for scientific applications. The RAW format images have a sensor falloff caused by the lens that is corrected by using an anti-vignetting method. The vignette correction is applied to all images over the growing season. A pseudo-radiometric calibration is then applied with calibration coefficient values extracted from the official MAPIR Camera Control code. MAPIR did not comment when requested on the origin of the calibration coefficients and claimed the values for gain and offset did not have any units, meaning any radiometric calibration done with these values will not have units of $\mathbf{W/m^2/sr/nm}$.

The MAPIR Survey3W camera (Figure 4.15) is 5.9 cm long x 4.15 cm wide x 3.6 cm deep and weighs 50 grams without a battery (Camera, 2016). It is a modified GitUP G3 action camera that has a spectral filter in front of the Sony Exmor R IMX117 sensor to collect Blue, Green, and NIR bands (Figure 4.16). It has an extreme

low distortion lens with a Field of View (FOV) of 87° and a resolution of 4000 x 3000 pixels. The ground-based cameras are positioned at approximately 3 m height giving a GSD of 0.14 cm/pixel. The narrow bandwidths at specific wavelengths collected precise spectral information about the crop being imaged. However, in order for this spectral information to be useful it needs to be calibrated, which includes applying an anti-vignetting correction (Yu, 2004) to correct for the fall-off in light sensitivity in pixels further from the center of the image, as well as converting the 12-bit digital numbers into values of pseudo-radiance.

The MAPIR Survey3W camera is the only camera used during this growing season; therefore, the false-color images need to provide the information to determine the flower progression at the field as well as the VI time-series. The images after processing remain very dark, making it difficult for an observer to tell what is in the image. Image processing techniques are used to brighten the image and emphasize the bloom for classification. Even with a brightened image the flowers are not always getting classified properly in the SamplePoint photo analysis (Booth *et al.*, 2006), leading us to perform a supervised classification on the images. A supervised classifier is chosen because we are given control over the classes of interest and the trained classifier can be applied to any image with the same number of bands. “These classifiers are able to learn the characteristics of target classes from training samples and to identify these learned characteristics in the unclassified data” (Belgiu & Dru, 2016). Training classes are created using ENVI software program (Exelis Visual Information Solutions, Boulder, Colorado) and the classification is run in the companies coding environment, ENVI IDL, in order to automate the process of image classification as much as possible.

4.3.2 Methods

At the end of the growing season the cameras are collected from the fields and the RAW + JPG images are saved directly to a desktop computer from the cameras SD card. A curated data set is created by removing any image files that did not meet the naming convention. The remaining RAW and JPG files are read into Python where the metadata from the JPG image is extracted, although this step is optional as only the information used in the file naming convention is saved. The MAPIR Camera Control code provided the steps to unpack the RAW file where it is read into a numpy array where the bit data are unpacked, reshaped, split into two arrays by reading in alternating bytes, flipping every other bit, repacking as a 16-bit number, and finally read back in from buffer as an 16-bit unsigned integer array. Once the 4000 x 3000 array is read back in, the byte data is debayered (Bayer & Rochester, 1976) using the OpenCV library and a RG2RGB debayering algorithm. This data is then saved as a TIFF image file. The vignette correction is then applied to this TIFF image.

Vignette Correction

Each camera has unique lens properties and the center of the lens will vary. Therefore, each camera needs an independent correction function. In order to correct for the sensor fall off, we imaged an evenly lit white surface where each pixel is expected to have the same value. This is used to estimate the correction factors that we could apply on a pixel-by-pixel basis. The second method from the (Yu, 2004) paper is used to find a 2-D hypercosine function to fit to the correction factors. A uniform illumination source is setup in a dark room to avoid ambient light interference. The uniform illumination comes by diffusing the light source to create *soft* lighting conditions. The

light source is placed behind an umbrella diffuser which is 5 cm away from a white nylon ripstop material. Each camera is then placed on a tripod so that the lens is parallel with the diffusion material and level with the light source (Figure 4.17). Care is taken to collect images that are evenly illuminated so the difference in pixel values from the center to the edges of the image are only due to the optical vignetting of the lens. The cameras are controlled via wi-fi and each took approximately 120 images of this light source through the diffusion materials.

The correction factor is calculated by averaging all the images, separating this average image into the respective spectral bands, and taking each pixel value and dividing it by the maximum value of that band (Yu, 2004). The distribution of the correction factors is then approximated by a single correction function in the form of a 2-D hypercosine function (Yu, 2004), and the images are multiplied by this correction function.

Radiometric Calibration

A radiometric calibration is required to quantitatively use remotely sensed data. An optical sensor calibration is achieved using known gain and offset coefficients to convert digital numbers into at-sensor radiance (Pompilio *et al.*, 2018). The MAPIR Survey3W gain and offset values are found in the MAPIR Camera Control code (Table 4.2), although they do not have the typical units of $\mathbf{W/m^2/sr/nm}$. The images are normalized to have values between zero and one using the global minimum and maximum from the entire growing season, this is known as a relative radiometric normalization and is a commonly used method when processing remotely sensed time-series images (Chen, 2018). Next each image is multiplied by 65,535 to take on

the range of a 16-bit unsigned integer image. Once the radiometric calibration where the gain and offset is applied to each band, the images can be used in a supervised classification algorithm and used to study variations in the VI over time.

Slope	Intercept
Red	
6.96	-0.086
Green	
1.89	-0.049
Blue	
2.74	-0.038

Table 4.2: Radiometric coefficients used to calibrate the MAPIR Survey3W camera.

Image Brightening for classification

It is apparent when viewing the images in SamplePoint that the images are too dark for manual classification. When the images are viewed in ENVI 5.5 the images look bright with buds easily visible. The ENVI software automatically applies a brightness adjustment, contrast stretch, and image sharpening. Each image is opened in ENVI IDL and then saved as an 8-bit JPG image as it would appear on the users screen. These brightened images are then used to create Regions of Interest (ROI) for a supervised classification.

Image Classification

The image taken on 6/15/2018 at 1300 hours is opened in ENVI 5.5 for each field because this is the day before pollinators are released at Field 1. These images are used to create an ROI set with care taken to ensure the number of pixels and regions are

similar between classes and that each class has over 500 pixels. Five classes are used in the classification: *shadow*, *soil*, *vegetation*, *maybe flower*, and *flower*. The *flower* class consisted of the very obvious flowers, for example where a stem is extending over the soil and the flower shape and color are very obvious, while the *maybe flower* class is comprised of less obvious flowers that are mixed in with the vegetation. This ROI set is then used to train a SVM supervised classifier in ENVI IDL 8.7. The classifier is trained separately on two arrays of the calibrated TIFF image plus a VI band, one including a band for GNDVI and the other including a band for FVI. Each trained classifier is saved to a .epo file so it could be reused and applied to other images. The trained SVM classifier is then applied to all the images from that field station and this is done for each field station. A comparison of the classified images for Field1B, hour 1300 , FVI and GNDVI can be seen in Figure 4.19. The percentage cover for each class from the classified image is calculated and then saved to a csv file, where it is later read into Python for plotting. The percentage flower is fairly noisy so a Savgol Filter is applied with a window size $1/3$ of the length of the dataset and a polynomial order of degree 3.

4.3.3 Results

There were 1079 images collected and properly saved at the four field stations during the 2018 field season. All these images are pre-processed and classified in the same manner. Each processing step has an effect on the spectral band ratios for each pixel in the image, which in turn influenced the resulting VI. The VI progression for the JPG, TIFF, vignette corrected, and radiometrically calibrated images from the field stations are shown in Figure 4.20.

The radiometrically calibrated and vignette corrected TIFFs are used in a SVM

supervised image classification. Before classification each image has a fourth VI band added to the image. Both GNDVI or FVI are used with a classification run for each VI separately. The classification images are broken down into percentage cover for the five classes that are classified. However, based on the ROI Separability measurement (Richards & Jia, 2006) the *maybe flower* and *flower* class has low separability (Table 4.3), indicating that they are at least similar in their spectral ratios and potentially all flowers. Because of this, the *maybe flower* and *flower* classes are combined into a total flower class with the minimum total flower value subtracted. The resulting percentage cover for flower is shown in Figure 4.21. The *maybe flower* and *flower* class also has the lowest user and producer accuracy results (Table 4.4) because the two classes are often confused with each other.

Classes	Field1A	Field1B	Field2A	Field2B
Maybe Flower and Flower	0.16	1.09	0.83	0.16
Vegetation and MaybeFlower	1.32	1.56	1.76	1.69
Flower and Vegetation	1.39	1.81	1.64	1.91
Flower and Soil	1.63	1.99	1.97	1.91
MaybeFlower and Soil	1.75	1.99	1.99	1.91
Vegetation and Soil	1.95	1.99	1.99	1.99
Shadow and Soil	1.97	1.91	1.99	1.99
Shadow and MaybeFlower	1.99	1.99	1.99	1.99
Shadow and Flower	1.99	2.00	1.98	2.00
Shadow and Vegetation	1.99	1.99	1.92	1.99

Table 4.3: ROI separability for each class at every field. Any value above 1.9 is considered reliable separability between classes with a maximum value of 2 (Jeffreys, 1946).

4.3.4 Discussion

The 2018 MapIR Survey3W images are used to get both the VI time-series progression as well as the percentage cover. The vignette correction and radiometric calibration

Field1A	Shadow	Flower	Veg	maybe Flower	Soil
FVI					
Producer accuracy:	1.00	0.88	0.95	0.77	1.00
User accuracy:	0.99	0.77	0.92	0.70	1.00000
GNDVI					
Producer accuracy:	1.00	0.87	0.98	0.83	1.00
User accuracy:	0.99	0.74	0.94	0.79	1.00
Field1B	Shadow	Flower	Veg	maybe Flower	Soil
FVI					
Producer accuracy:	1.00	0.89	0.97	0.83	1.00
User accuracy:	1.00	0.87	0.95	0.89	0.99
GNDVI					
Producer accuracy:	1.00	0.88	0.95	0.77	1.00
User accuracy:	1.00	0.83	0.95	0.83	0.99
Field2A	Shadow	Flower	Veg	maybe Flower	Soil
FVI					
Producer accuracy:	1.00	0.79	0.98	0.83	0.98
User accuracy:	1.00	0.81	0.89	0.89	0.99
GNDVI					
Producer accuracy:	0.99	0.73	1.00	0.90	0.95
User accuracy:	1.00	0.88	0.90	0.74	1.00
Field2B	Shadow	Flower	Veg	maybe Flower	Soil
FVI					
Producer accuracy:	1.00	0.36	0.99	0.75	1.00
User accuracy:	1.00	0.57	0.99	0.56	1.00
GNDVI					
Producer accuracy:	1.00	0.50	0.99	0.84	1.00
User accuracy:	1.00	0.77	0.99	0.56	1.00

Table 4.4: User and Producer accuracy for each class at every field. A perfect classification accuracy score is 1.

to get the images ready to use in a VI time-series and classification were successful. The a 2-D hyperbolic cosine function is fit to the correction factors for the vignette correction. This has improved results compared to directly applying the correction factors because the imaged screen has dust speckles that made certain pixels darker than expected. The function is an appropriate estimation of the sensor fall off and worked even with the noise from the data collection process.

The VI time-series is created for every field (Figure 4.20) for each image file type and for two Vegetation Indices, FVI and GNDVI. When looking at a single field and VI, the curves from one image type to the next all have different values; however, they tend to have the same shape, with the JPG images being the noisiest. The curves between TIFF, vignette corrected images, and radiance images all maintain the same shape. The radiance images could be further calibrated by using a grey calibration card located within each image. While this wouldn't be a reflectance calibration this card could be used to normalize the images from day to day and would account for variations in pixel values due to differences in solar illumination. The grey cards used in this study quickly became unusable due to physical weathering of the card. Using a shielded grey card for normalization would likely decrease the noise in the VI time-series. When comparing the different VI time-series progression for a single field the curves all have a similar shape as the equivalent image type. The VI progression for the same image type between different fields shows not only very similar trends in the curve, but also values that are very close in magnitude. This is a promising result indicating the sensors are collecting a similar signature, and that the differences should be mostly caused by differences in location. The time-series progression for each VI are very similar in shape; however, the values differ because the GNDVI is

normalized whereas the FVI is not.

Using a SVM supervised classification method has good results in terms of the user and producer accuracy. A visual inspection of the classification images also shows reasonable results with the metal post often being misclassified as either soil or flower. The similarity in spectral signatures between *maybe flower* and *flower* gave the lowest lowest ROI separability score and user/producer classification accuracy at every field station. This is expected and demonstrates that the ROIs used as *maybe flower* are often actual flowers. To get the total flower percentage cover, the *maybe flower* and *flower* class are combined, and the minimum % flower value for each station is subtracted. The subtraction of the minimum value of total flower from the % flower cover variable essentially removes the metal post from the % flower cover. While the SVM classification seemed to perform well, it estimated an earlier bloom up period than in previous years as estimated by SamplePoint classification (Chapter 5). The difference in percentage flower cover estimates made the 2018 data set incompatible with the time-series from the other years. In order to make the 2018 data set comparable an RGB camera needs to be included in the sensor to allow for accurate ground truthing.

4.4 Conclusions

The image processing techniques used for each of the cameras is successful in providing greater accuracy for the VI time-series analysis. There is a definitive trend in both the 2017 and 2018 VI time-series. The SVM classification performed well on the MAPIR Survey3W camera imagery to obtain a percentage flower cover value.

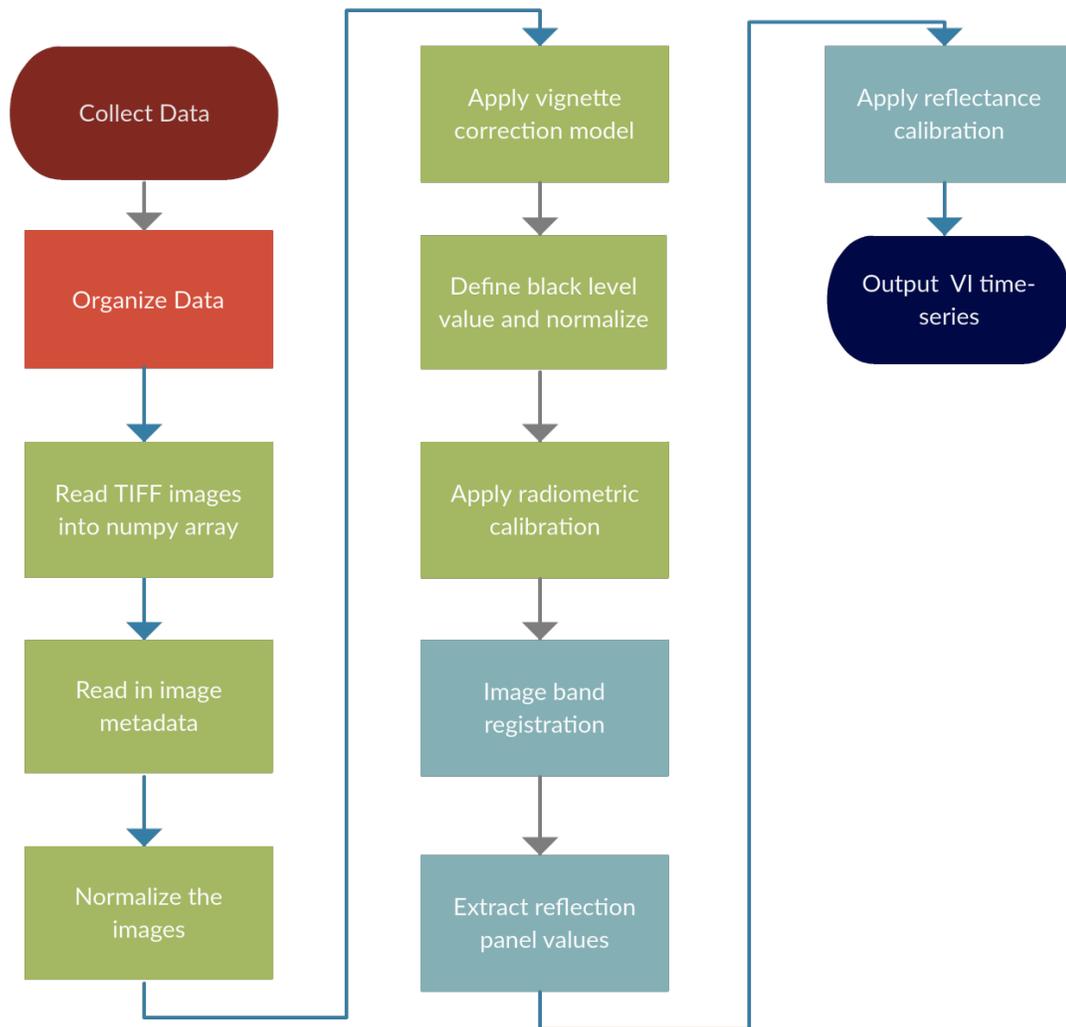


Figure 4.4: Micasense Rededge image processing steps. All steps except those in red are implemented in Python.

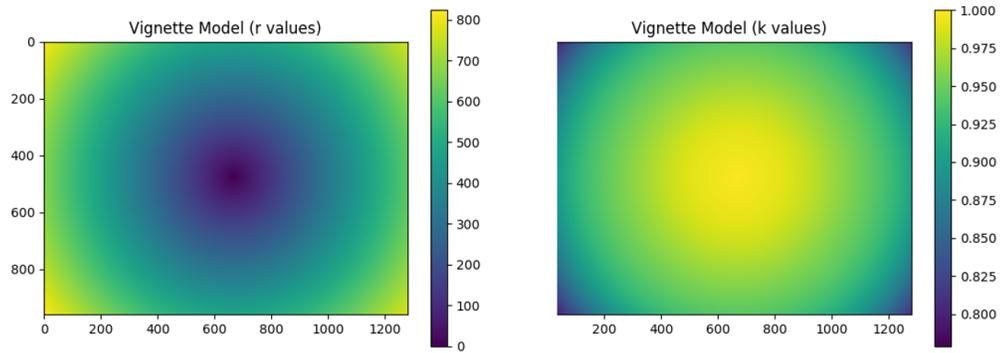
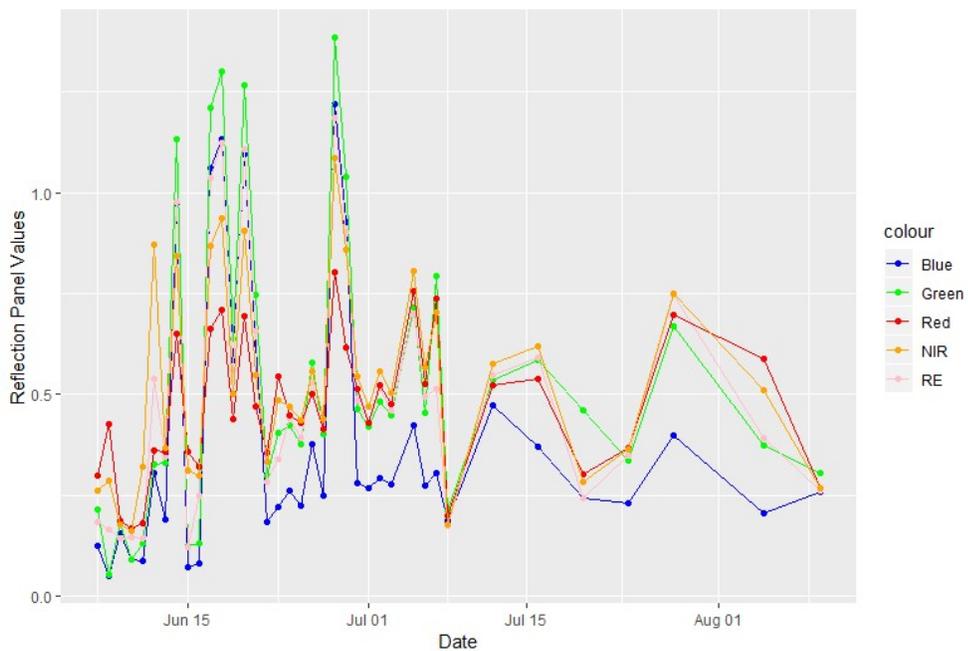


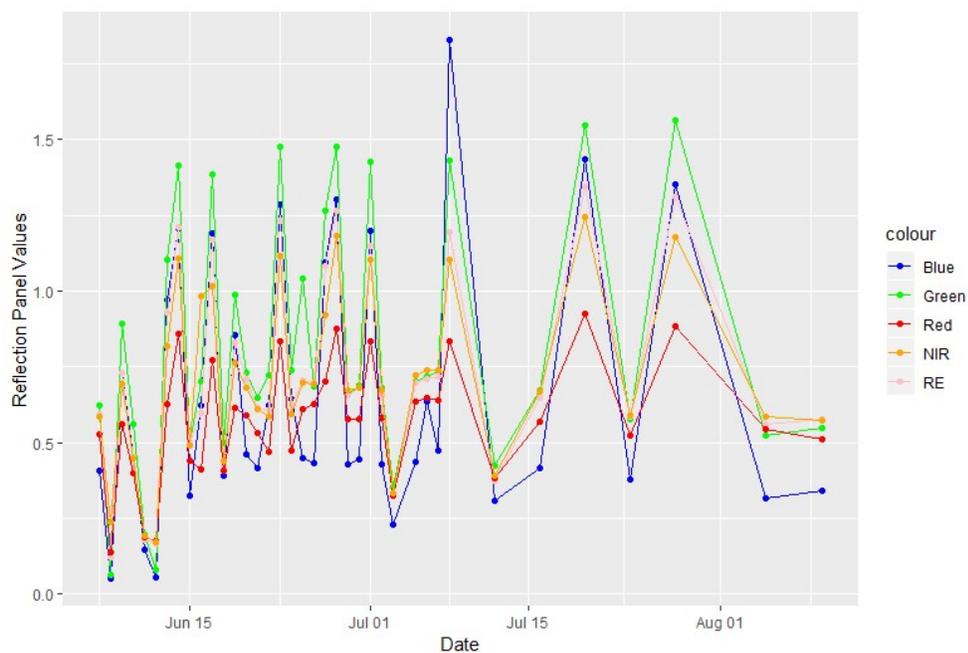
Figure 4.5: Vignette correction model where the left image shows how far from the center (cx , cy) each pixel is, and the right image shows the value the original pixel is divided by.



Figure 4.6: This is an example of the Python script output used to collect the reflectance panel values. The Micasense Rededge bands are first calibrated and then aligned. The combined RGB image is displayed for a user to manually draw a rectangle (blue box) within the boundaries of the reflection panel. The average value for each band is then saved and applied to the images of the dataset.



(a) Field 1 Reflection Panel Values



(b) Field 2 Reflection Panel Values

Figure 4.7: Reflection panel values extracted using a custom Python script from over the course of the growing season for fields 1 and 2.

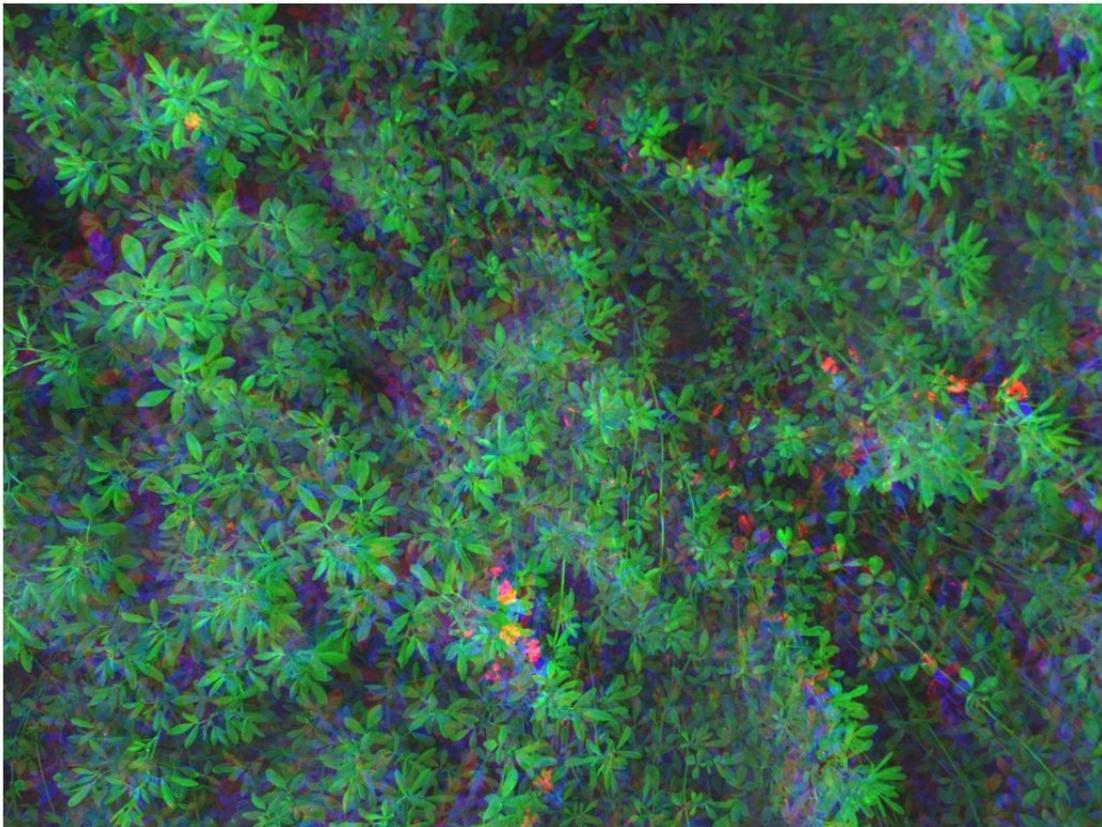


Figure 4.8: Original RGB composite image from the Micasense Rededge multispectral camera. Misaligned pixels are evidenced as blue, green, and red striping

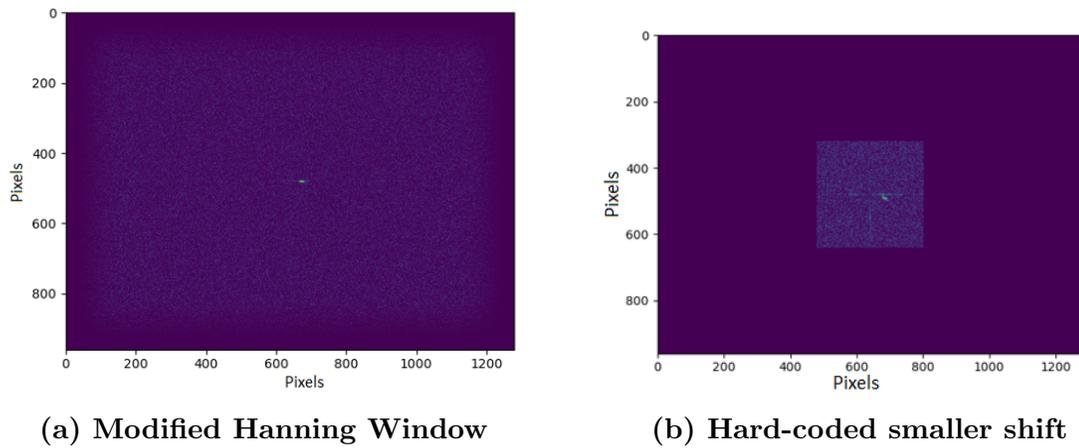


Figure 4.9: Shifted cross-power spectrum for two of the Micasense Red-edge camera bands RE and Blue. (a) Is the original modified Hanning Window. (b) Same data but with a hard-coded 1/6th maximum translation. While colors mean nothing without a scale, please excuse the lack of a color bar for now and it will be included later.

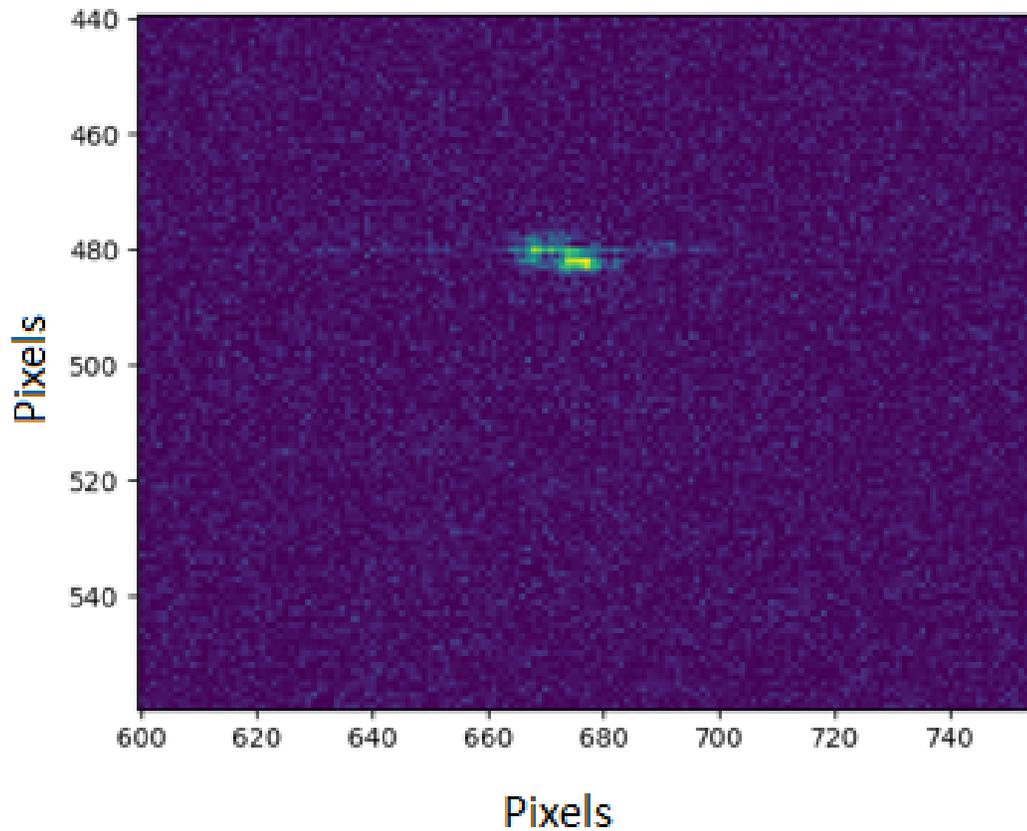
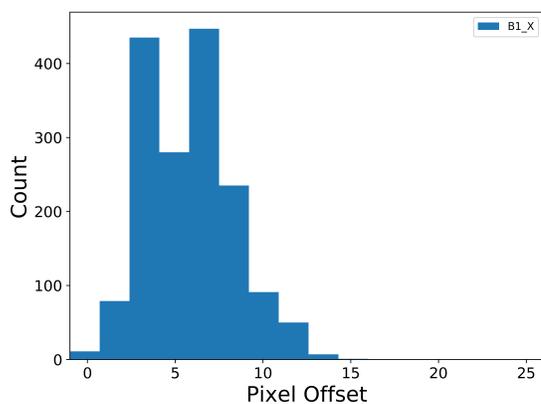


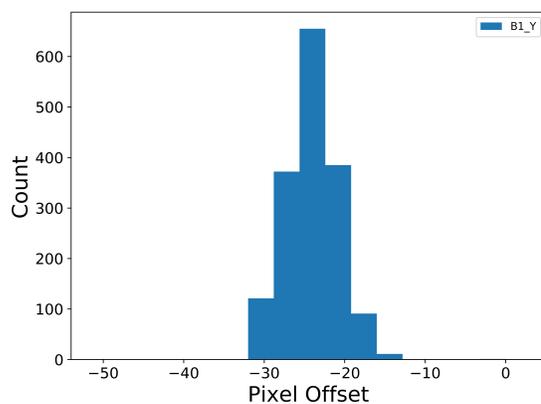
Figure 4.10: Zoomed in view of the shifted cross-power spectrum for two of the Micasense Rededge camera bands, RE and Blue. The center of the images is at pixel (640,480). While colors mean nothing without a scale, please excuse the lack of a color bar for now and it will be included later.



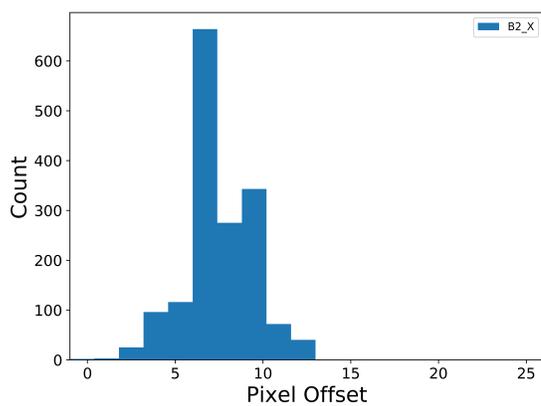
Figure 4.11: RGB composite image from the Micasense Rededge after image registration. Note the striping is now gone in most of the image and flowers are visible. Misalignment still exists in some areas (e.g. upper left corner).



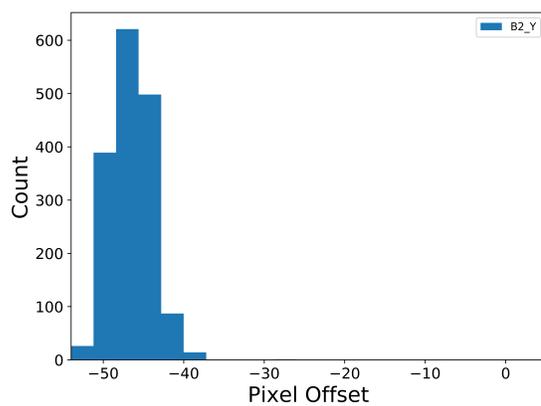
(a) Band 1 X-Axis offsets



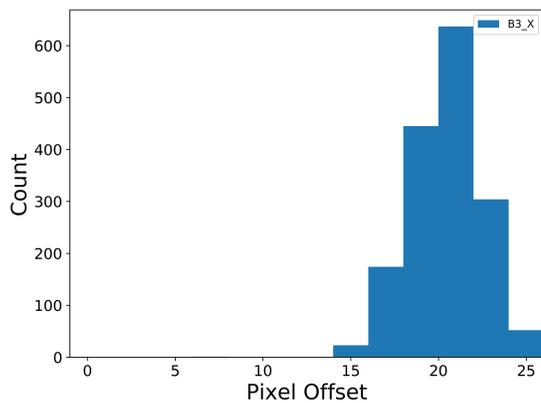
(b) Band 1 Y-Axis offsets



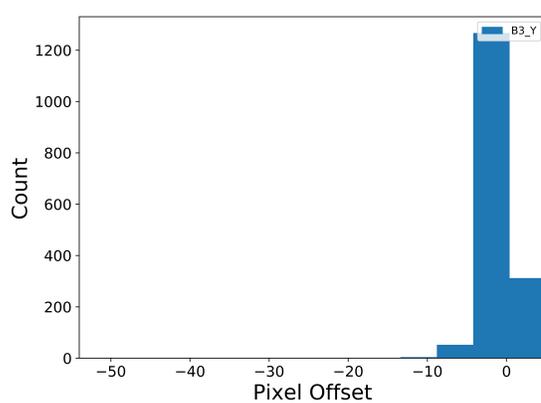
(c) Band 2 X-Axis offsets



(d) Band 2 Y-Axis offsets

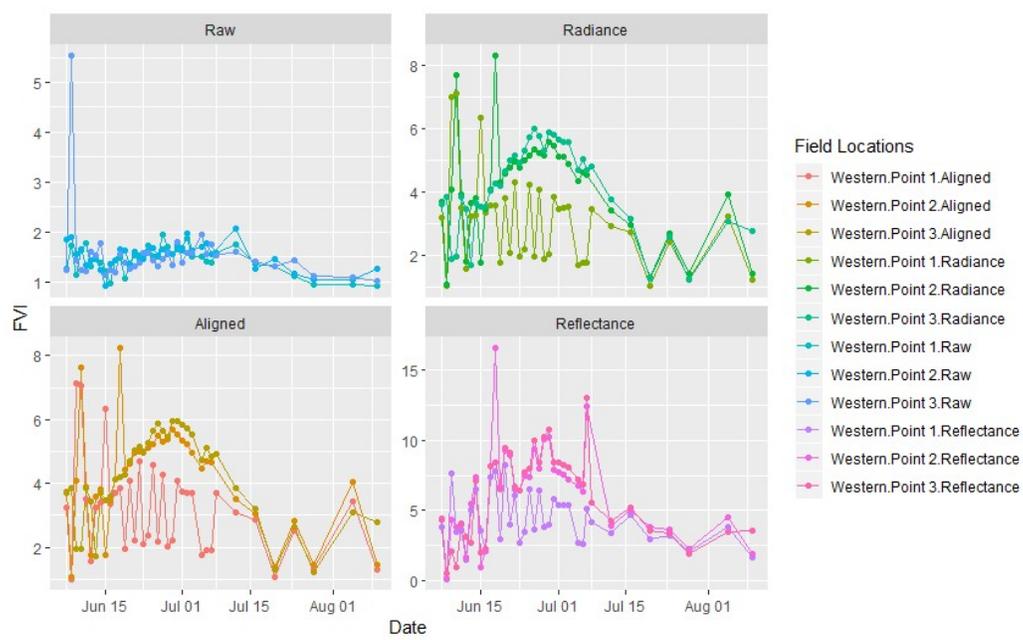


(e) Band 3 X-Axis offsets

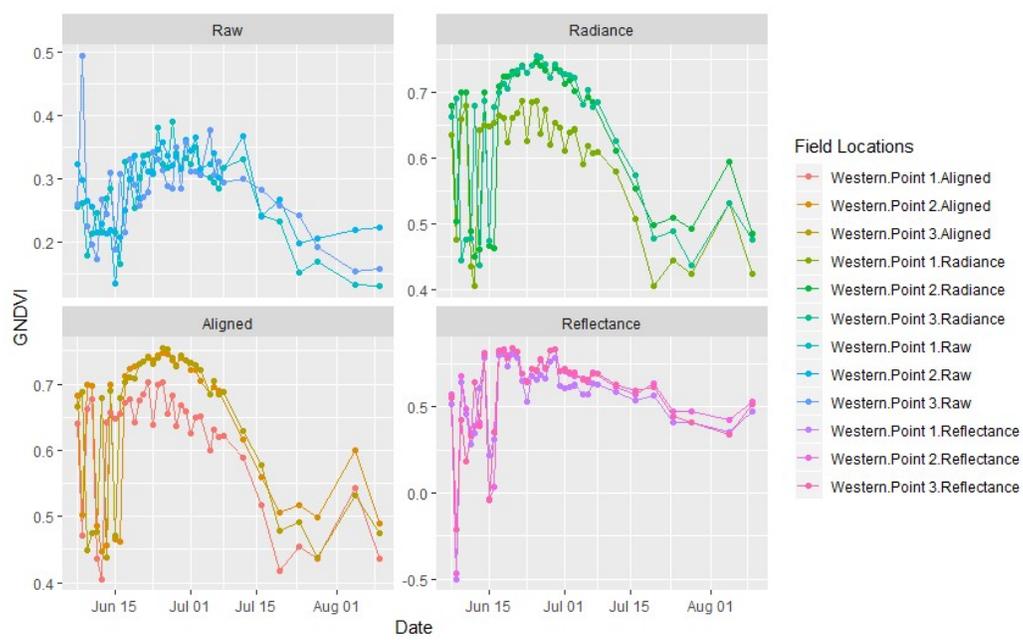


(f) Band 3 Y-Axis offsets

Figure 4.12: Histograms for the image translation results.

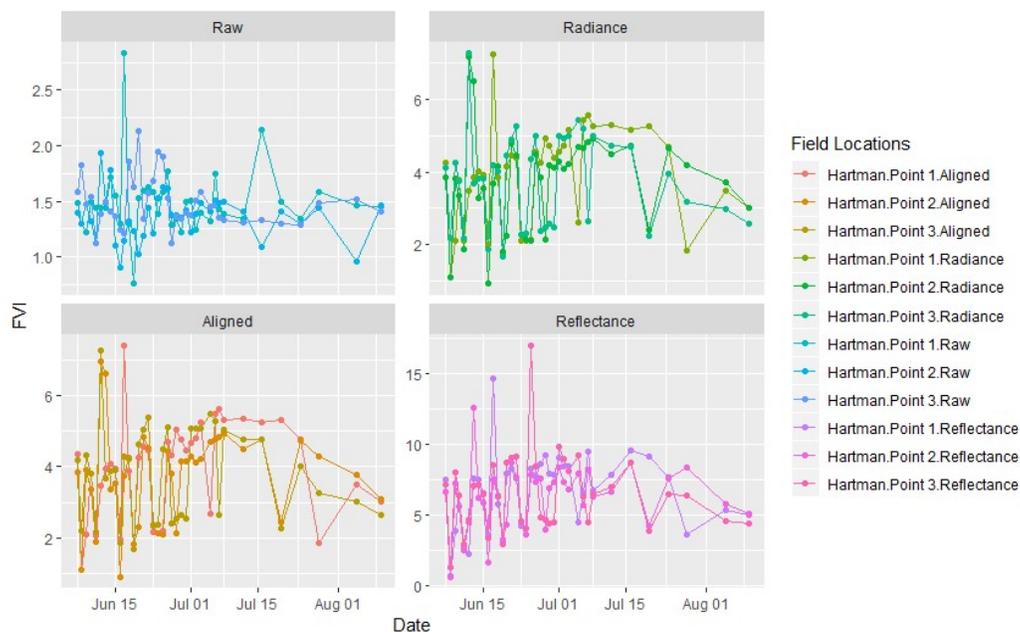


(a) Field 1 FVI

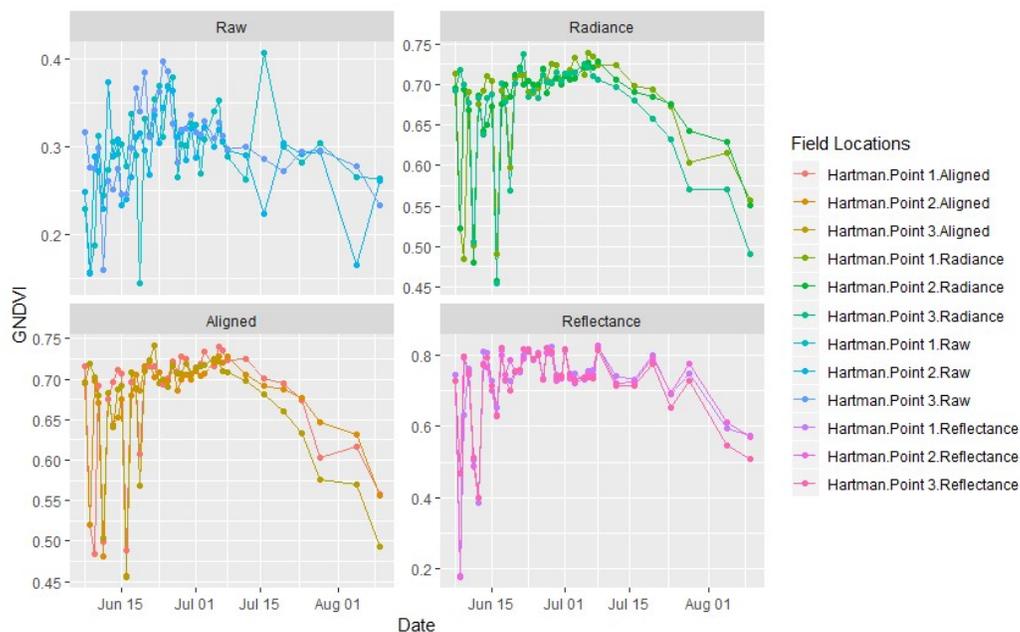


(b) Field 1 GNDVI

Figure 4.13: Field 1 field VI progression after each of the processing steps is applied. The panels should be read top to bottom and left to right to see the progression of the image processing steps.



(c) Field 2 FVI



(d) Field 2 GNDVI

Figure 4.13: Field 2 field VI progression after each of the processing steps is applied. The panels should be read top to bottom and left to right to see the progression of the image processing steps.

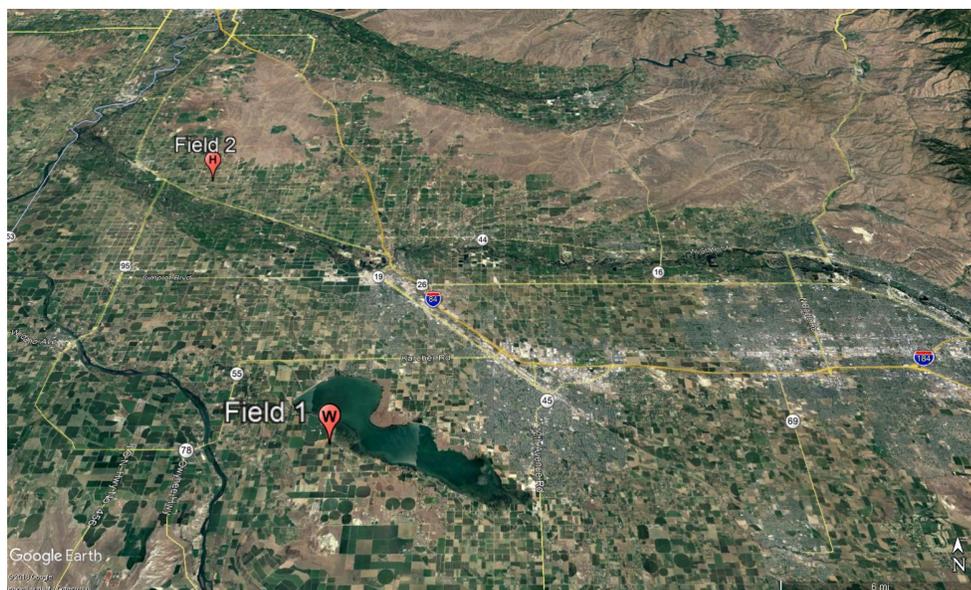


Figure 4.14: The fields locations for the 2018 data collection.



Figure 4.15: MAPIR Survey3W used in this experiment.

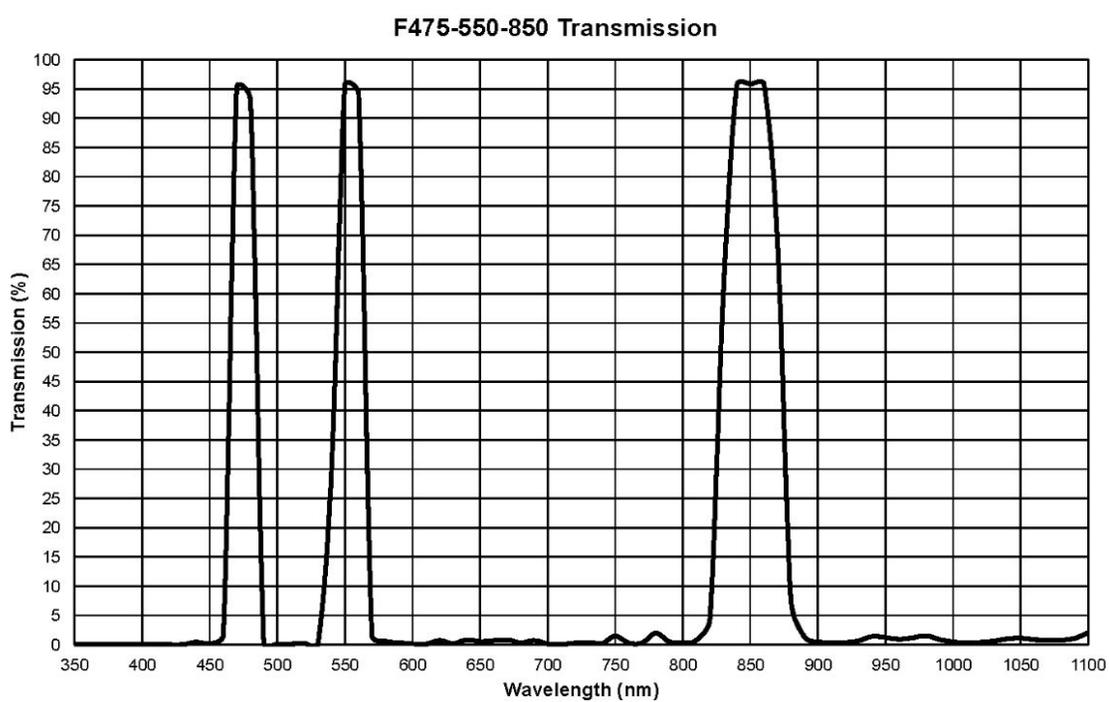


Figure 4.16: Spectral response for the MAPIR Survey3W.



Figure 4.17: Setup for the vignette correction data collection.

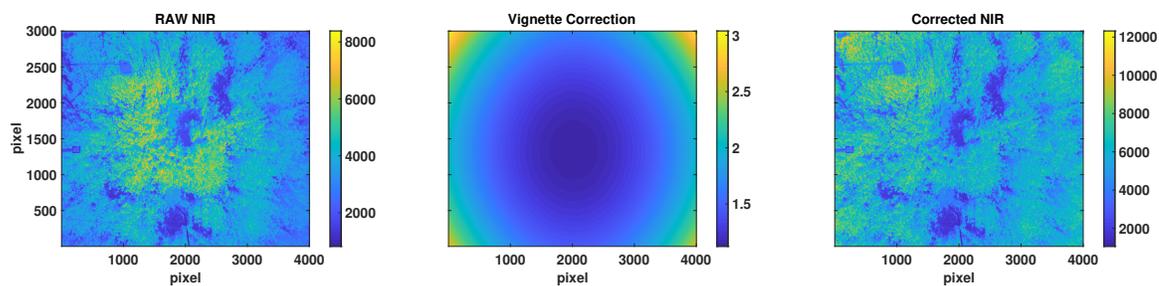
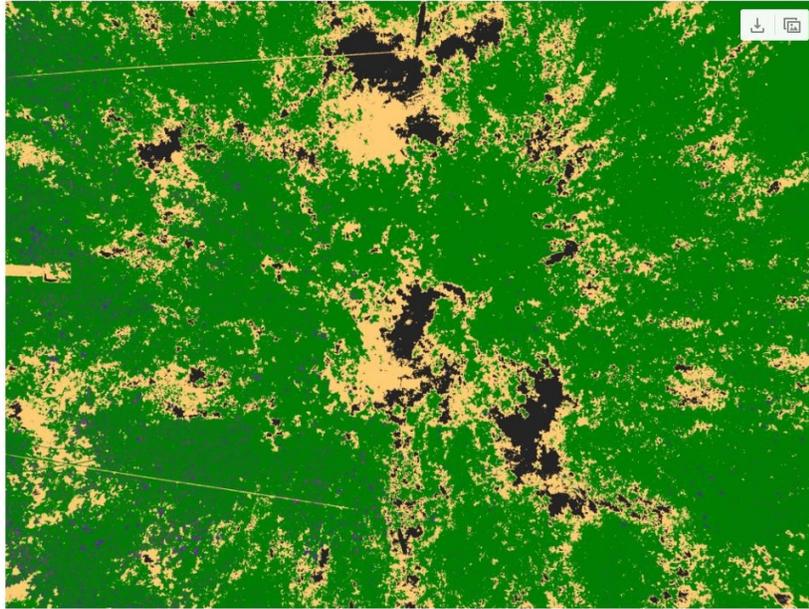
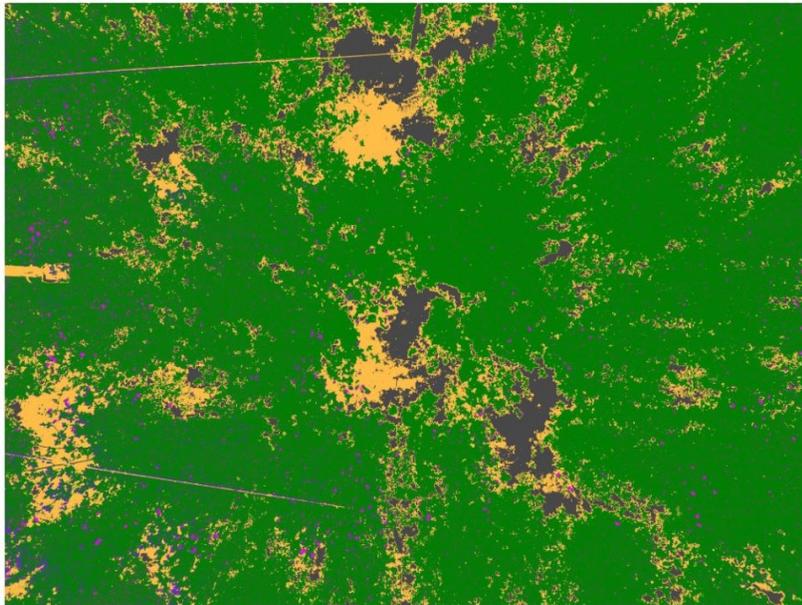


Figure 4.18: Example of the vignette correction to the NIR channel at 12:00 on 11 June 2018: original NIR image (left); vignette NIR correction function (middle); corrected NIR image (right).

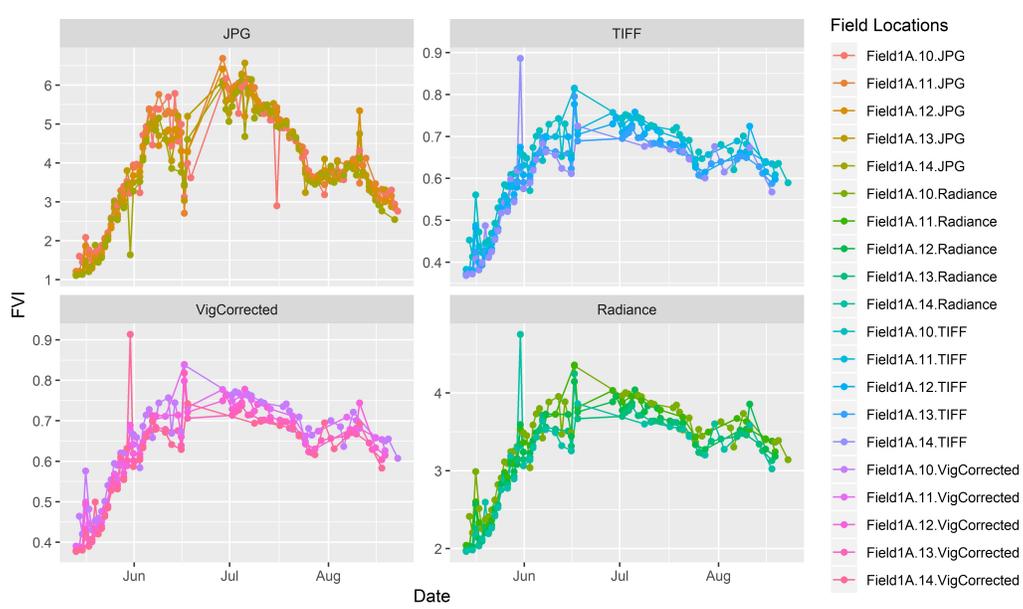


(a) Field1B 6/15/2018 1300 FVI Classification

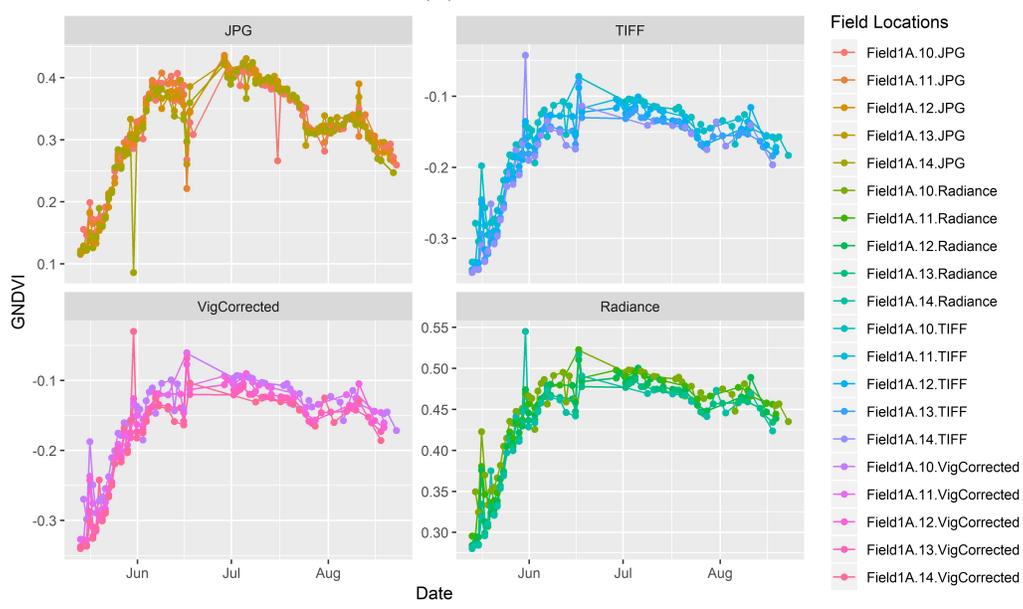


(b) Field1B 6/15/2018 1300 GNDVI Classification

Figure 4.19: Examples of the resulting classified images from the SVM trained classifier.

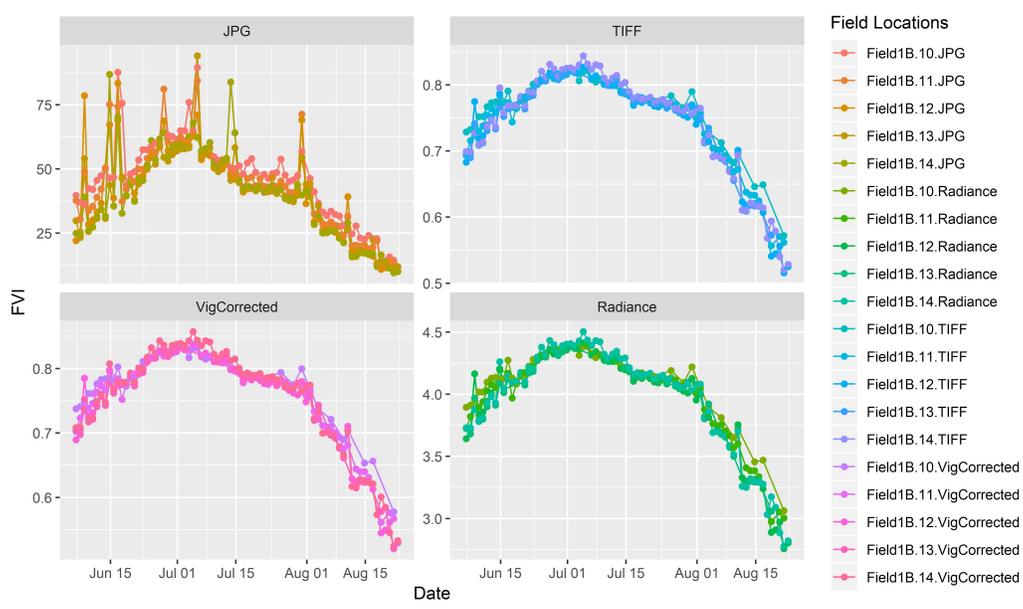


(a) 2018 Field1A FVI

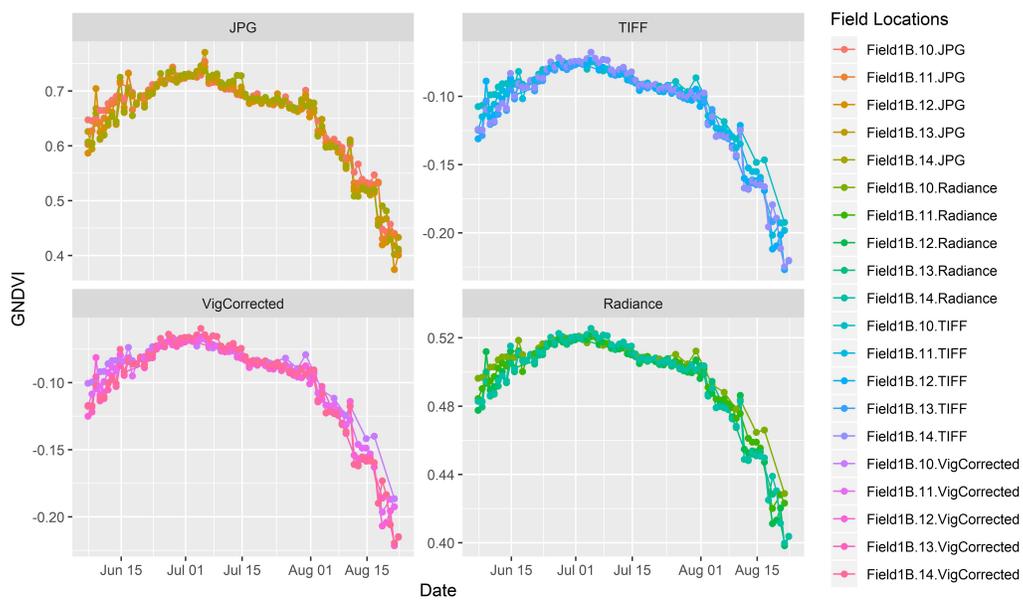


(b) 2018 Field1A GNDVI

Figure 4.20: Comparison between the two VI used in the image classification for Field1A. The panels should be read top to bottom and left to right to see the progression of the image processing steps.

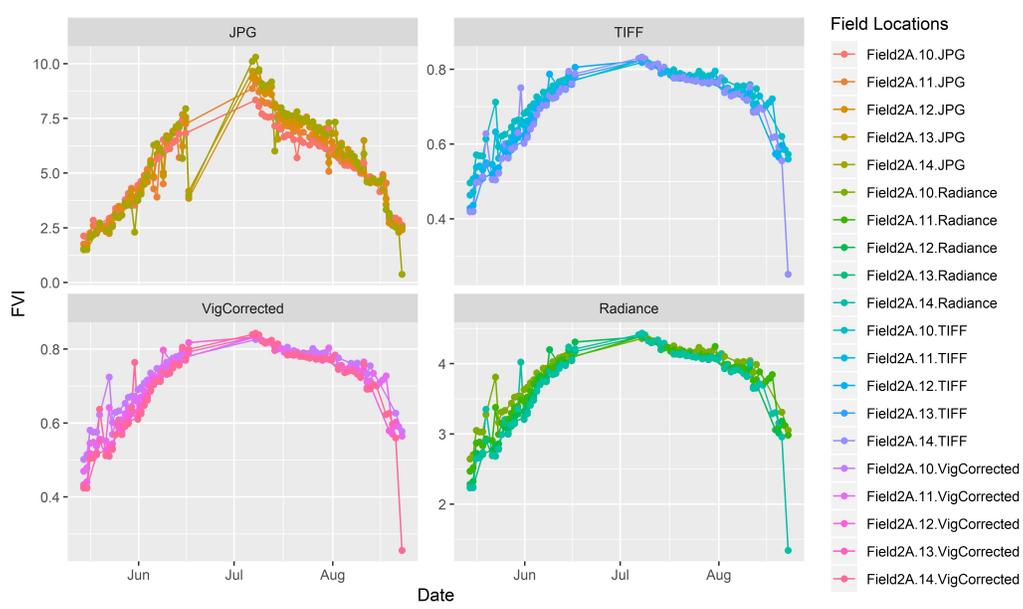


(c) 2018 Field1B FVI

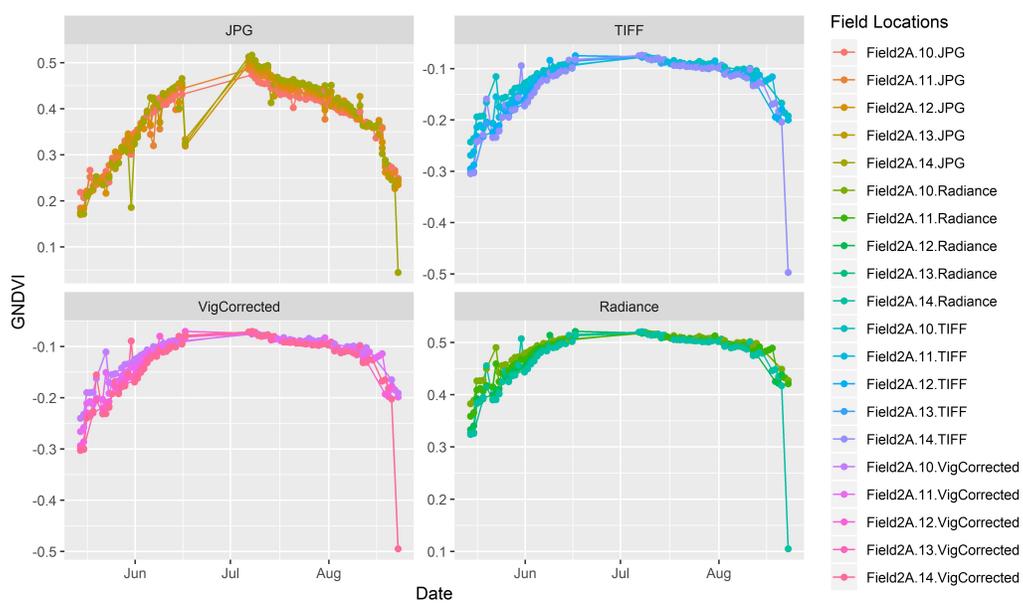


(d) 2018 Field1B GNDVI

Figure 4.20: Comparison between the two VI used in the image classification for Field1B. The panels should be read top to bottom and left to right to see the progression of the image processing steps.

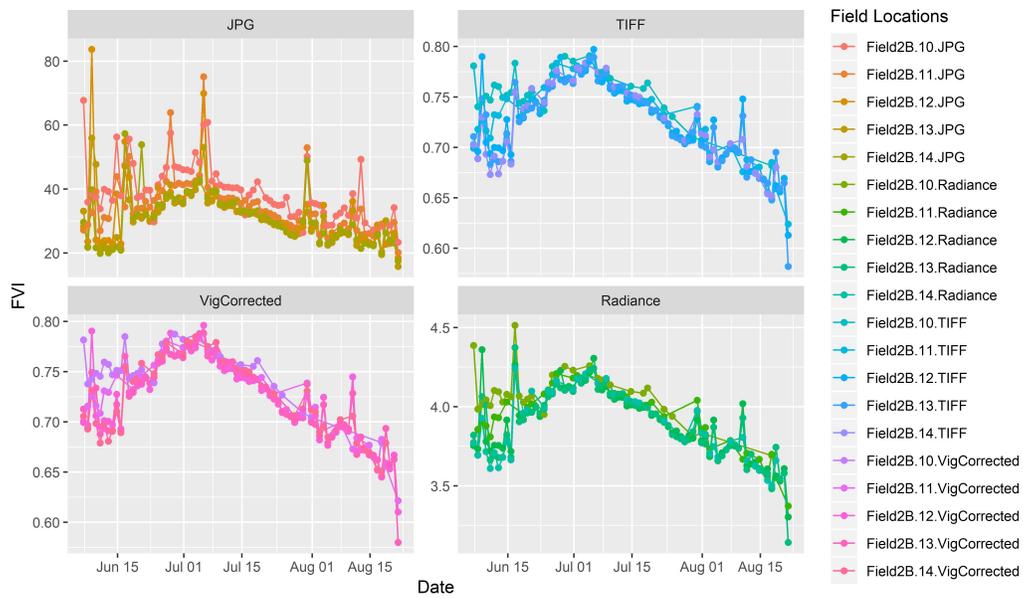


(e) 2018 Field2A FVI

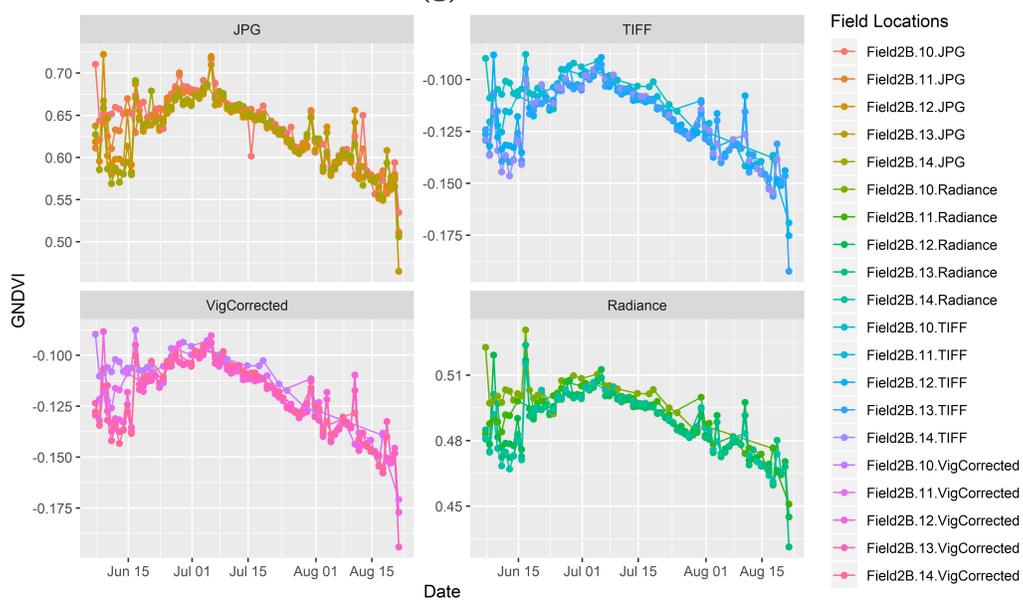


(f) 2018 Field2A GNDVI

Figure 4.20: Comparison between the two VI used in the image classification for Field2A. The panels should be read top to bottom and left to right to see the progression of the image processing steps.

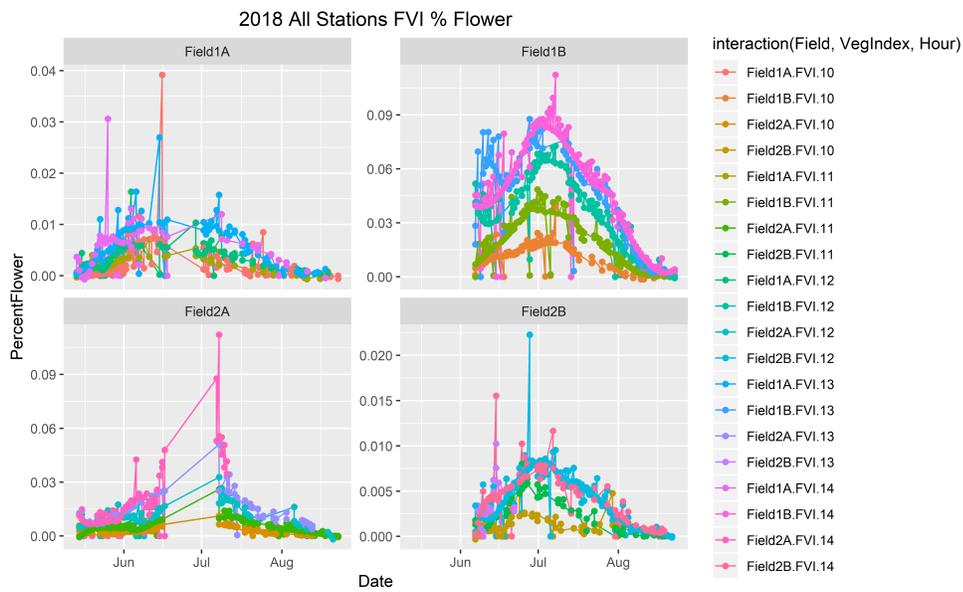


(g) 2018 Field2B FVI

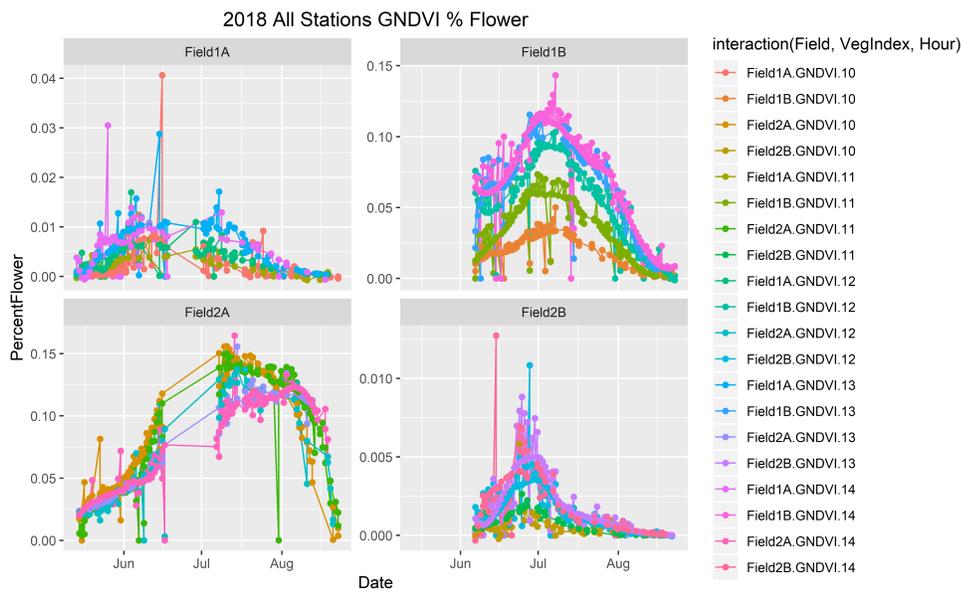


(h) 2018 Field2B GNDVI

Figure 4.20: Comparison between the two VI used in the image classification for Field2B. The panels should be read top to bottom and left to right to see the progression of the image processing steps.



(a) FVI Percentage Cover



(b) GNDVI Percentage Cover

Figure 4.21: All stations from the two fields are classified using a trained SVM classifier. The percentage cover is extracted from each classification image and smoothed using a Savgol Filter with a window size 1/3 of the length of the dataset and a polynomial order of degree 3.

CHAPTER 5: PARTICLE FILTER MODELING TO PREDICT ALFALFA SEED CROP FLOWER PERCENTAGE USING VI TIME SERIES

5.1 Summary

In this chapter, a particle filter modeling approach is used to estimate and predict values for percentage flower cover (PFC) based on observation data. This approach uses GNDVI observation values to update a process model estimates for PFC and to predict future PFC values. A subset of the data discussed in the previous chapters is used to create process and observation models for the expected growth of alfalfa seed crops. These models, along with the observation data from each field is input into a particle filter (PF). The results of the PF, using data from the final cutback through peak bloom conditions show the potential to achieve a highly accurate estimation with R^2 as high as 0.95. The particle filter is also used to predict the PFC out to 20 days using only the first 10 observations, with R^2 values as high as 0.92. Testing the PF on an accelerated dataset that the process model would not be able to accurately predict resulted in an R^2 value as high as 0.96. This procedure is robust enough to be used with noisy observations and can be expanded to include other observation

measurements (e.g. temperature, soil moisture, growing degree days, etc.). The results from this analysis demonstrate a proof of concept for the estimation and prediction of PFC at alfalfa seed crop fields using remote sensing observations and state-space modeling.

5.2 Introduction

Knowing the percentage flower cover of an alfalfa seed crop field can be used to predict when optimal bloom conditions will occur and thus when to release the pollinators. Therefore, the ability to predict phenology in this way would be a very powerful precision agriculture tool. A state-space approach is used to model the bloom progression at alfalfa seed crop fields. In order to analyze and make inferences about this system, at least two models are required: a model describing the evolution of the state with time (the process model) and a model relating the (noisy) observation measurements to the state (the observation model) (Arulampalam *et al.*, 2002). The process and observation models we create in this chapter come from the 2016 (Chapter 2), 2017 (Chapter 4), and 2018 (Chapter 4) data.

In order to predict PFC, a realistic model of the bloom development must be established. The validation data (i.e. ground truth data) for bloom development is derived from the high-resolution digital images taken at two alfalfa seed crop fields during the 2016, 2017, and 2018 growing seasons. An estimate of PFC comes from SamplePoint image analysis done on these high-resolution images (Chapter 2, 4). We approximate the trend in PCF over time using a logistic equation (Equation 5.1), which is then transformed into a state-space representation and discretized to a formula referred to as the process model. This non-linear equation is suitable for many biologic systems and is chosen as the process model for several reasons:

- the model has a 'S' or sigmoidal shape that fits the expected bloom behavior, such as starting at zero percent flower cover until a certain date where bloom begins, until the plant is in full bloom where the percentage flower cover levels off,
- the model parameters have a straightforward interpretation, including a lower asymptote representing the starting percentage cover, an upper asymptote representing the typical maximum percentage cover, an inflection point representing how many days after cutback the field begins to flower, and a scaling factor related to the steepness of the curve at the inflection point (Drummond, 2017; Tong & Vendettuoli, 2017),
- the model is a simple representation of how the field is expected to behave, and the percentage flower estimate can be updated from field measurements (observations), and
- the PFC estimate is an indirect counting method that includes (process) error, such as flowers hidden from the camera beneath a leaf, that can be accounted for using a Bayesian model fit.

Incorporating field observations obtained from remote sensing technologies provides extra information to gain more accurate estimates of the PFC than the process model can provide. Observation data from the 2016 and 2017 growing seasons are transformed into a vegetation index (VI), as detailed in Chapter 2, and compared to the PFC estimates, which is referred to as the observation model. A Michaelis-Menten function (Eqn 5.4) is used to fit the observation data because it is characterized by a steep rise that gradually flattens into a plateau. There are only two parameters to be

fit in this function, V_m and K , where V_m is the maximum rate of VI change per PFC and K is the PFC value where the rate of change is equal to $V_m/2$. The observation and process models are then input into a state-space model.

The state-space representation, also known as the “time-domain approach”, is a mathematical model of a physical system as a set of input, output, and state variables. While state-space modeling has not been widely applied to remote-sensing data, crop monitoring is an appropriate application because agricultural crops are typically systems that evolve over time. While this is not entirely true for alfalfa’s indeterminate growth pattern, the crop does generally go through various phenological stages and the use of growing degree days (GDD) has successfully been used to estimate alfalfa phenology development (Ben-Younes, 1992; Munk, 1999), where GDD is a measure of the amount of heat needed for the plants to grow and develop (Munk, 1999). While many statistical methods have used remote sensing data to estimate different phenological states (Sakamoto *et al.*, 2005; Fan *et al.*, 2015; Wu *et al.*, 2014), such as greenup, maturity, and senescence, these methods typically require the complete time-series to obtain estimates for the dates and stages (White & Nemani, 2006). A Kalman filter can be used for a real-time solution to estimate the state (Vicente-Guijalba *et al.*, 2015), but is limited to linear systems. De Bernardis *et al.* (2014) demonstrate that a particle filter is a more reliable method to infer phenological stages compared to a Kalman filter. In this study a PF, also known as a sequential Monte Carlo approach, is used to combine the process and observation models to provide a single state estimate. The PF is a nonlinear recursive Bayesian filter where the posterior probability density function (pdf) is represented by a set of particles. The particles represent tentative states in the process model and concentrate around

the most probable state (De Bernardis *et al.*, 2016a). If the number of particles is sufficiently large then the particle distribution approximates the true posterior pdf (Arulampalam *et al.*, 2002). When observation data is available, the pdf given by the model is combined with the pdf of the observation to estimate the most likely state (De Bernardis *et al.*, 2014). In this chapter, the potential for a near-real time estimate of percentage bloom, in which observation data is combined with a process model is tested using a particle filter.

5.3 Methods

5.3.1 Process Model

The process model is a sigmoid function

$$x(t) = A + \frac{B}{1 + e^{-r*(t-t_0)}}, \quad (5.1)$$

where t is time measured in days after cutback, $x(t)$ is the bloom percentage at time t , A is the lower asymptote, $A + B$ is the upper asymptote, t_0 is the position of the inflection point, and r is the scaling factor for the logistic curve. We fit this model to the PFC data to estimate the coefficients and create the process model; however, we need to be careful we use high-quality data when creating this model so that we are able to make accurate predictions with the PF. The state-space formulation (Eq 5.2) is obtained by taking the derivatives of Equation 5.1.

$$\dot{x} = \frac{r}{B} * (x(t) - A) * (B - x(t) + A), \quad (5.2)$$

The state-space formulation is then discretized to get a computational solution given by Equation 5.3.

$$x_{k+1} = x_k + r * \Delta t * (x_k - A) * \frac{(B - x_k + A)}{B}, \quad (5.3)$$

where k has integer values, $t_k = k * \Delta t$, x_k is the bloom percentage at time t_k , and x_{k+1} is the next bloom percentage value.

Data Selection

When comparing the 2016, 2017, and 2018 data for PFC over time as shown in Figure 5.1, it is apparent that the SVM classification of the 2018 data is significantly different than the SamplePoint estimates from 2016 and 2017. The 2018 data set is not used in this initial analysis because based on seed production agronomists expertise, the 2018 environmental conditions should have caused a later bloom up period than in 2017, while the SVM classification of the 2018 data has an earlier bloom up. The 2016 PFC estimates, shown in green in Figure 5.1, also have an earlier bloom up period than the 2017 data. This is likely because the greenhouse experiments, which constitute 9 of the 12 plots monitored during the 2016 growing season, are planted much later in the season. The late cutback dates make for warmer growing conditions and likely faster growth. While the change in growth rate should also be captured by the observation measurements, the sensor used in 2016 had significantly different responses, as discussed in Section 5.3.2, than in 2017.

The logistic model is used to fit 2017 PFC data from Field 1 and Field 2 because these are the data we have the most confidence in given what we know from agronomy and how alfalfa plants normally grow. The PFC data are smoothed using a Savgol

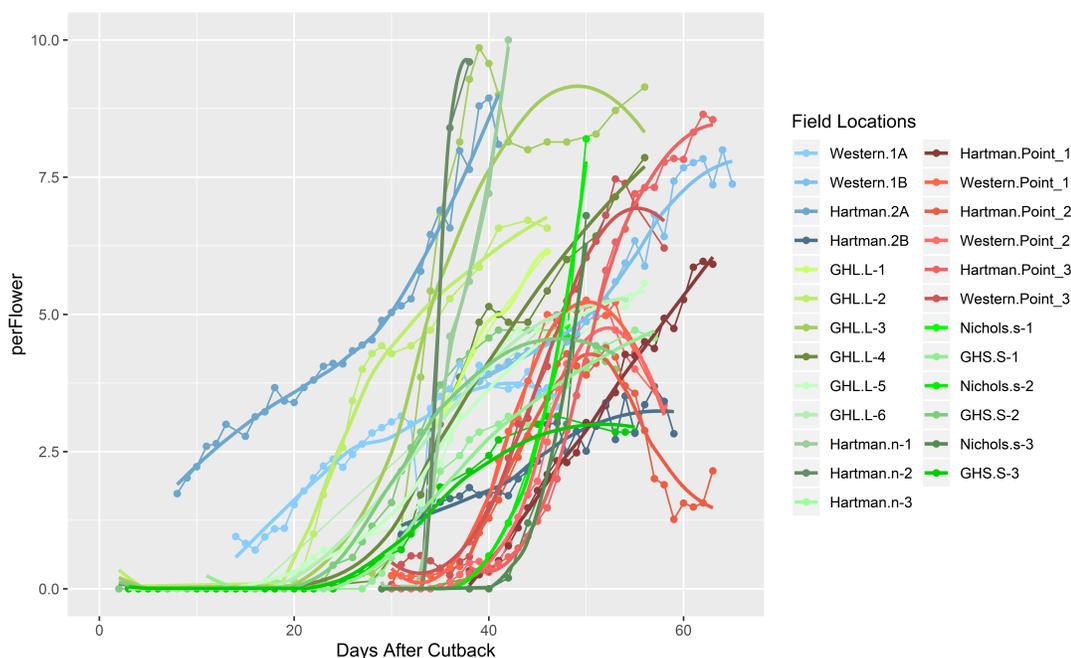


Figure 5.1: Percentage flower cover with time from the 2016 (green lines), 2017 (red lines), and 2018 (blue lines) growing seasons. The 2016 and 2017 PFC estimates come from SamplePoint image analysis, while the 2018 PFC estimates are from the SVM classification (Chapter 4).

filter with polynomial order of 3 and a window 1/3rd the length of the data collected at each site. The data is processed to include an integer value for days after cutback that is calculated for each field using the date minus the date of cutback, plus 1. In order to properly fit a sigmoidal function, the fall off in PFC (e.g. decreasing PFC days in Figure 5.1) is not included by truncating the data on July 5th, which is four weeks after the pollinators are released and is before the PFC fall off for most field sites monitored (Figure 5.2).

Model Fitting

Equation 5.1 is fit using a multi-level Bayesian model from the R package BRMS (Brkner, 2018). Before fitting the model using BRMS, a nonlinear least-squares estimate of the parameters A , B , t_0 , and r were obtained using the R package NPLR. These estimates provide the starting range for the priors to be used in the BRMS function where the r parameter is also defined to have a minimum value of zero. Three mixed effect models are run in BRMS to find the best fit for the four parameters. The first model uses a no-pooling approach, where all of the fields and locations are grouped together. The second model includes a group-level effect that differentiates between the two fields, using a parameter called *par1*; in this case each field is given a unique identifier. The *inits* parameter for the second model is set to zero instead of generating random initial values to minimize any divergent transitions after the *warmup* period as suggested in the BRMS documentation. The third model includes a group-level effect that differentiates between each field site using a parameter called *par2*, where each site is given a unique identifier. The models are run with eight chains and 2000 iterations with a warm-up period set to 1000. The convergence of the Bayesian model fit is supported by the Gelman-Rubin diagnostic term (Gelman & Rubin, 1992) and a visual inspection of chain mixing. After model fitting, the predictive accuracy is compared using a Bayesian leave-one-out cross-validation method, where the out-of-sample predictive fit is given by the expected log point-wise predictive density (*elpd*) (Vehtari *et al.*, 2017). The larger the *elpd*, the better the out-of-sample predictive performance of the model. The best-fit model parameters are extracted and used as the prediction function in the particle filter algorithm.

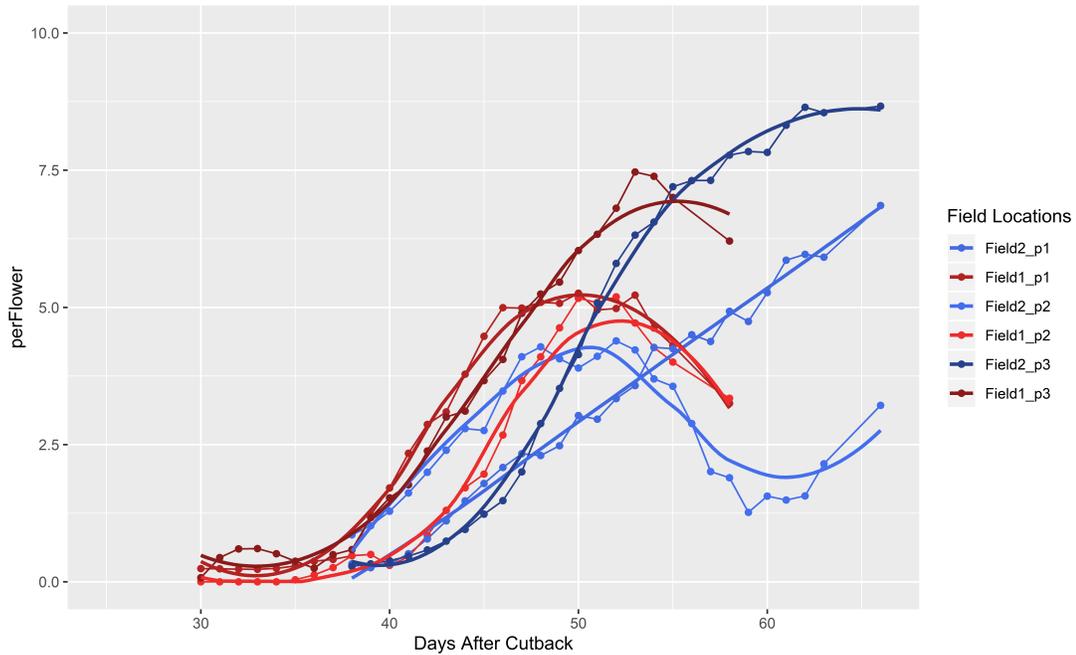


Figure 5.2: Percentage flower cover with time at two alfalfa seed crop fields during the 2017 growing season.

5.3.2 Observation Model

The VI over PFC data are used to fit the Michaelis-Menten function (Michaelis *et al.*, 1913):

$$VI = V_m * flower / (K + flower), \quad (5.4)$$

where VI is the vegetation index value, V_m is the maximum rate of increase, K is the PFC at which the growth rate is half the maximum ($V_m/2$), and $flower$ is the PFC estimate. As with the process model, we need to first determine which data to use to estimate the parameters in this observation model so that the particle filter provides accurate predictions.

Data Selection

GNDVI over time is shown in Figure 5.3 for the reader to get a feel for how the observation measurements progressed throughout the 2016, 2017, and 2018 growing seasons. The measurements from each year are taken with a different sensor and are processed at various levels of calibration. The 2016 data, shown in green, is taken from the ASD handheld 2 spectroradiometer from approximately 1.5 m height and are processed to absolute reflectance (Chapter 2). The 2017 data, shown in red, are collected with a Micasense Rededge multi-spectral camera from approximately 2 m height and are processed to units of radiance (Chapter 4). The 2018 data, shown in blue, are collected with a MapIR Survey3 camera from approximately 2 meters height and are processed to have units of pseudo-radiance (Chapter 4). The purple lines are from the Sentinel-2 satellite imagery, calibrated to top of atmosphere reflectance, taken during the 2017 and 2018 growing season and are shown for visual comparison. The difference between years is significant and likely a result of the sensors, their optical field of view, and the imaging height as opposed to actual field conditions. These different data collection methods complicate the observation model, so instead one type of data is selected.

Figure 5.4 shows the VI observation data as a function of PFC, with data from the 2016 (green lines), 2017 (red lines), and 2018 (blue lines) growing seasons shown. The 2017 data are used in the observation model. The 2018 observations are not used due to the uncertainty in the ground validation estimates as explained in the previous section. The 2016 data are also not used because they do not show a definitive relationship between the VI value and the PFC estimate. The 2017 model shows a rapid increase in GNDVI values as bloom begins and then levels off around 4% flower

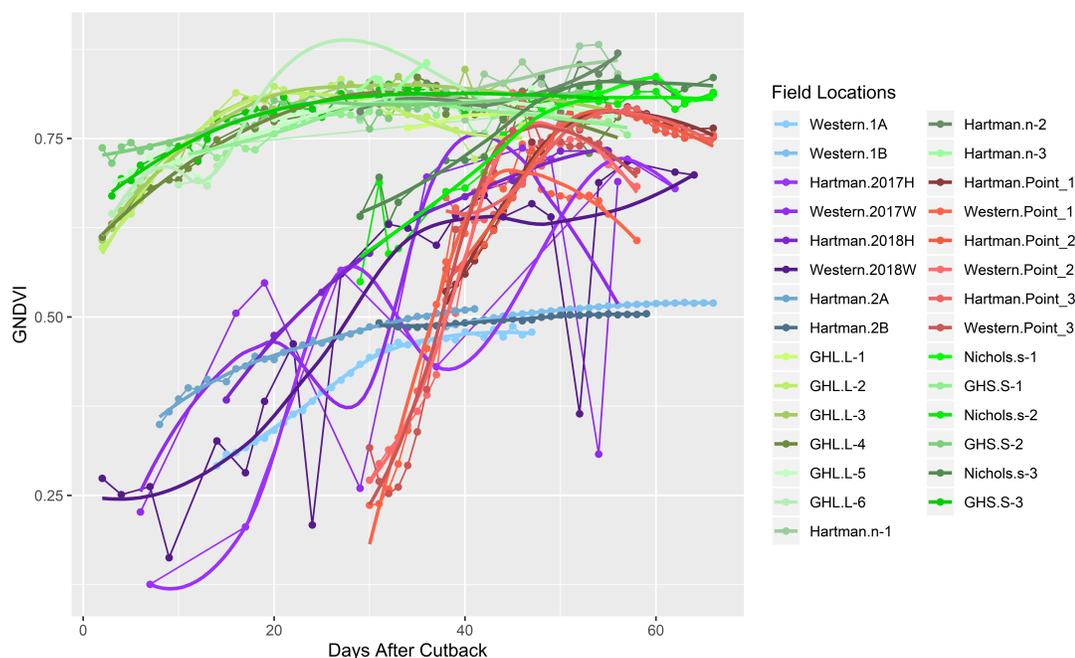


Figure 5.3: GNDVI values over days after cutback from all the locations sampled during this study. Each data set is collected using a different sensor, which likely explains the variations between years. As in the process model example, the 2016 data are shown as green lines, 2017 data as red lines, and the 2018 data as purple lines.

cover. This is better shown in Figure 5.5, where the 2017 data is split up by field and both GNDVI and FVI values are compared. Based on these images the observation model is fit to the GNDVI values, because the values from different locations have less variation.

Vegetation Index Check

We check to make sure the most suitable vegetation index is being used for the observation model. The random forest regression (RFR), as described in Chapter 2, is performed on the 2017 data using the multi-spectral images instead of the 2016

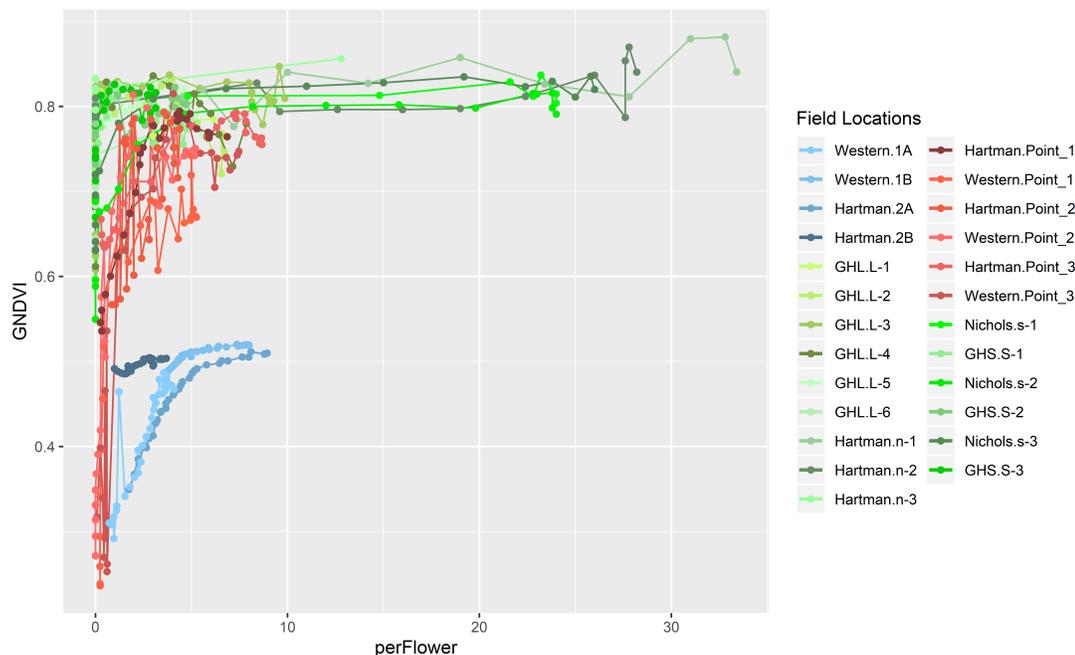
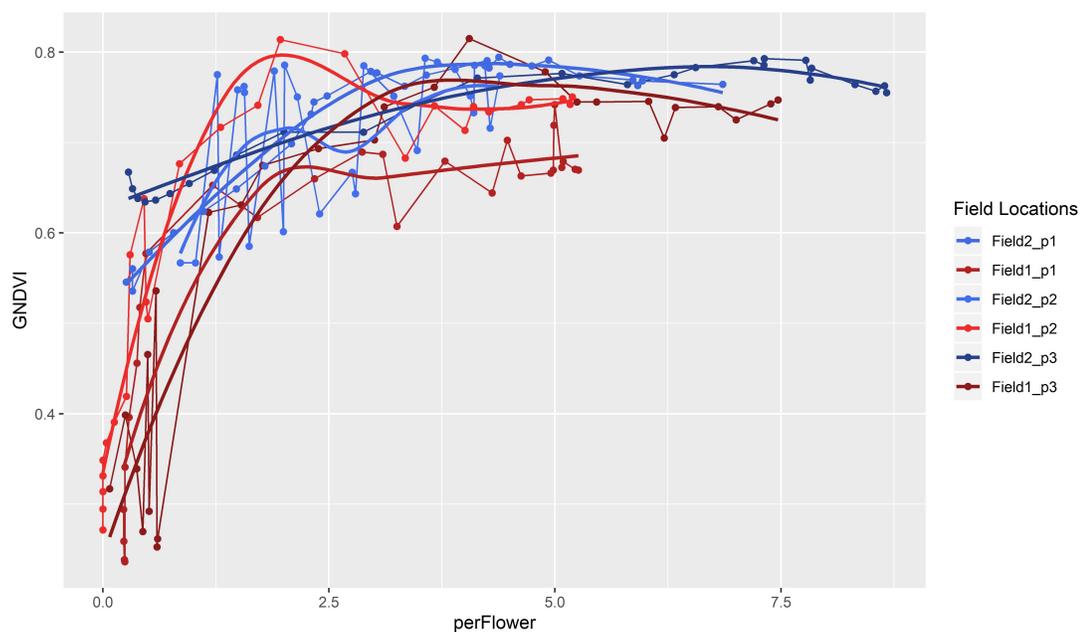
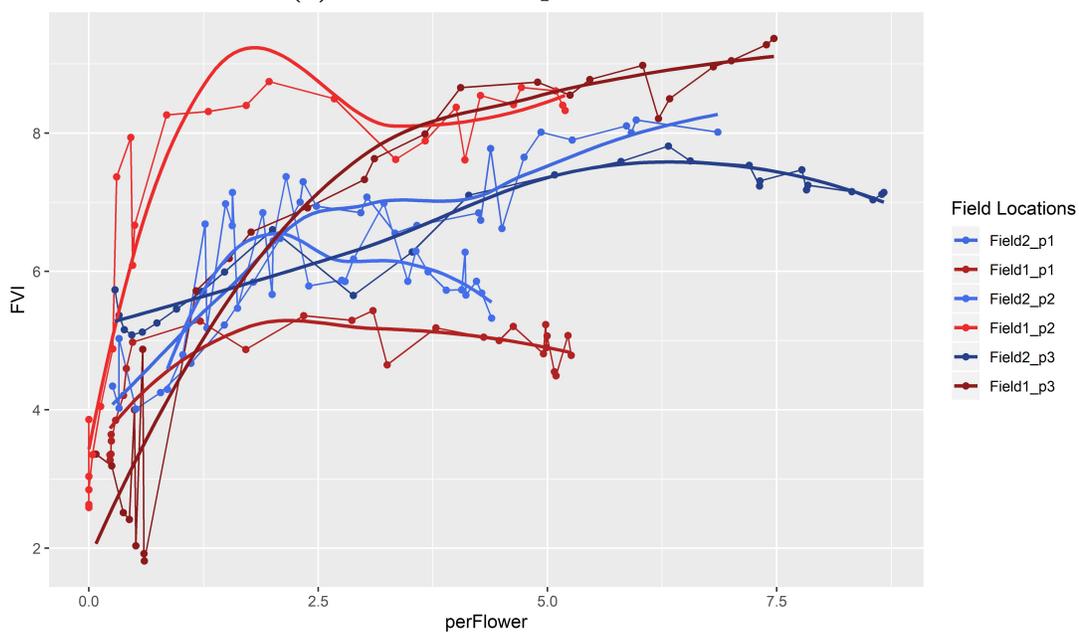


Figure 5.4: GNDVI values over PFC estimates from all the locations sampled during this study. Each data set is collected using a different sensor, which likely explains the variations between years. As in the process model example, the 2016 data are shown as green lines, 2017 data as red lines, and the 2018 data as green lines.

hyperspectral data. The multi-spectral bands are first transformed into the same VI used in the initial regression as listed in Appendix A. The individual bands and VI are used as the response variables in the regression. The analysis returned a “pseudo R-squared” value, indicating how well the regression performed, of 0.82. This means the RFR performed on the multi-spectral data did nearly as well as the best performing RFR on the 2016 hyperspectral data. The mean-squared error (MSE) is used to determine which variables are the most important. The 30 most important variables estimated by out-of-bag cross-validation (Breiman, 2001) from the RFR on PFC for both Field 1 and Field 2 using all three locations at each field is shown in Figure 5.6.



(a) GNDVI over percent flower



(b) FVI over percent flower

Figure 5.5: Vegetation Index values over percentage flower from two fields during the 2017 growing season.

These results indicate that FVI and GNDVI are both in the top 5 predictors for PFC and are appropriate to use as the observation model.

Model Fitting

The 2017 GNDVI over PFC from Field 1 and Field 2 are used to fit the Michaelis-Menten function in Equation 5.4. This non-linear function is fit using a multilevel Bayesian model using the R package BRMS (Brkner, 2018). Before fitting the model using BRMS, a nonlinear least-squares estimate of the parameters were obtained using the R package NLS, these estimates provide the starting range for the priors for V_m and K to be used in the BRMS function. Three mixed effect models are run in BRMS to find the best-fit parameters. The first model uses a no-pooling approach where all of the fields and locations are grouped together. The second model includes a group-level effect that differentiates between the two fields, using a parameter called *par1*, where each field is given a unique identifier. The third model includes a group-level effect that differentiates between each field site using a parameter called *par2*, where each site is given a unique identifier. The *inits* parameter for the third model is set to zero instead of generating random initial values to minimize any divergent transitions after the *warmup* period as suggested in the BRMS documentation. The models are run with eight chains and 2000 iterations with a warm-up period set to 1000. The convergence of the Bayesian model fit is supported by the Gelman-Rubin diagnostic term (R-hat) and a visual inspection of chain mixing. After model fitting, the predictive accuracy is compared using a Bayesian leave-one-out cross-validation method where the out-of-sample predictive fit is given by the (*elpd*) (Vehtari *et al.*, 2017). The best-fit model parameters are then extracted and used as the observation

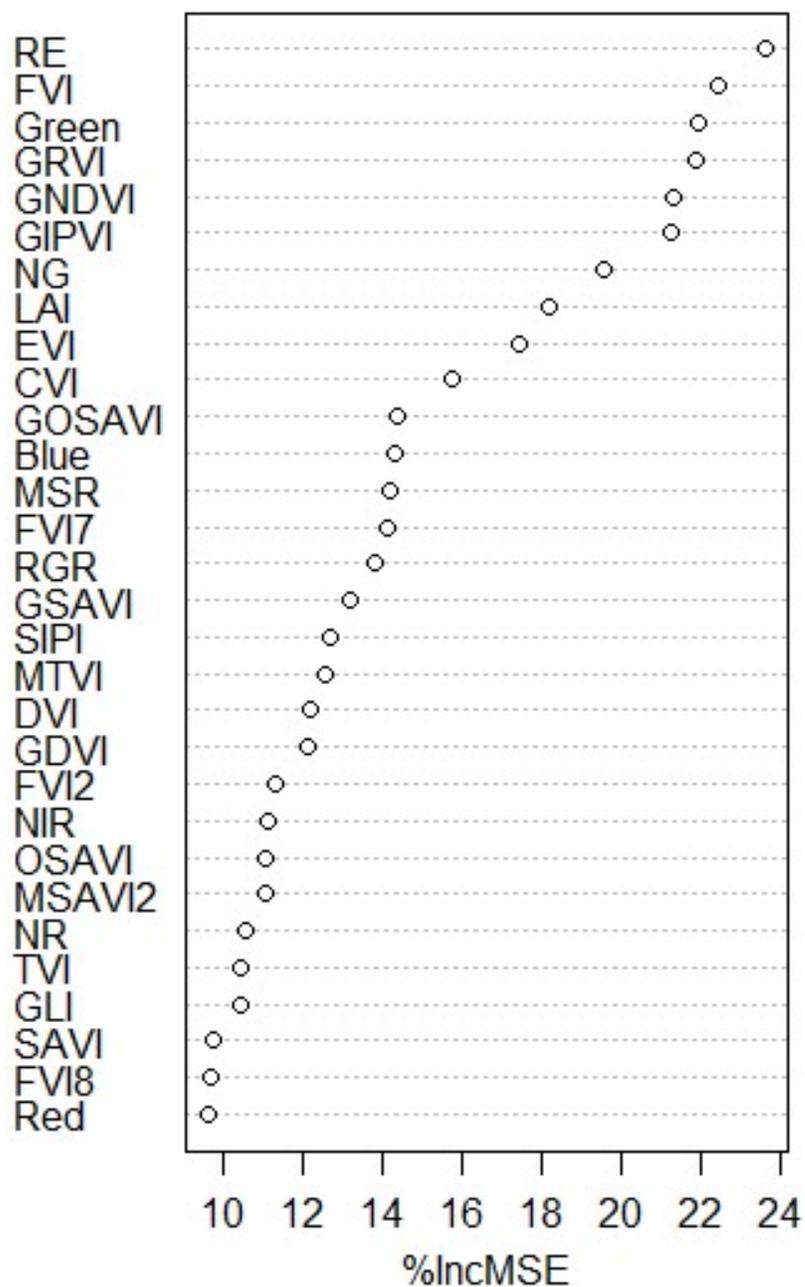


Figure 5.6: The 30 most important variables from the random forest regression on all 2017 data.

model in the particle filter algorithm.

5.3.3 Particle Filter

The PF is implemented in the free software R (R Core Team, 2013) as described in Crassidis & Junkins (2012) book *Optimal estimation of dynamic systems*. The PF is initiated with four components known as initial state estimates: the error covariance of the initial state estimate, the measurement noise, and the process noise. We set the initial state estimate to zero, the error of this initial state to 0.2, the residual standard deviation, σ , from the observation model as the measurement noise, and the residual standard deviation, σ , from the process model as the process noise. The PF is implemented to run with 1000 particles and passed the process model, observation model, and a vector with the observations, GNDVI values in this case. The observation vector is padded with *NaN* values to make the length of the vector represent days after cutback. The PF is then able to reconstruct the PFC curve using the observation values and the expected values from the process model. The model fits are evaluated using an R^2 value that compares the true PFC to the estimated PFC and *RMSE* that measures the differences in percent flower between the true PFC and estimated PFC. Next, the predictive capabilities of the PF are tested. This is done by running the PF model over only the first 10 observations and then calling the PF forecast method to make predictions for the remainder of the data collection period. The forecast uses the last observation to estimate the next state, and the state estimate from that measurement is also included in the forecasted percentage flower cover values. Finally, the model is tested on a simulation of a faster than expected crop growth where the fields percentage flower cover is accelerated by 20%. In this case, the process model will not follow the changes and the bloom percentage

estimates need to be compensated by the observations.

5.4 Results

Model Fitting

All of six of the Bayesian models successfully converged with R-hat values of 1 and the chains show visual evidence of mixing. The leave-one-out cross-validation technique used for model comparisons returns both an *elpd* and the *standard error of difference* as shown in Table 5.1 for the process models and Table 5.2 for the observation models. These models are arranged in descending order according to the expected out-of-sample predictive accuracy, where the best model (highest *elpd*) is shown with a value of 0.0 and the other values are relative to that model. The third process model performs the best with an estimated difference in *elpd* of -169.0 and -189.7 with a standard error of 13.4 and 14.2, respectively. The third observation model also performs the best with an estimated difference in *elpd* of -13.8 and -24.5 with a standard error (se) of 6.1 and 8.3, respectively. In both model validations, the difference in *elpd* is greater than twice the estimated standard error, indicating that the third process and observation models are expected to have significantly better predictive performance than the other two models (Vehtari *et al.*, 2017). The best performing process model and observation model fit and their 95% confidence interval are shown in Figure 5.7 to provide a visual display of how well the regression performed. A summary of the variable fits for the process and observation models are presented in Table 5.3 and Table 5.4, respectively. The confidence intervals are used to determine how much variability there can be for each parameter.

Model comparisons	<i>elpd_diff</i>	<i>se_diff</i>
ProcessModel3	0.0	0.0
ProcessModelI2	-169.0	13.4
ProcessModel	-189.7	14.2

Table 5.1: Leave-one-out cross-validation model comparison of the three process models.

Model comparisons	<i>elpd_diff</i>	<i>se_diff</i>
ObservationModel3	0.0	0.0
ObservationModelI2	-13.8	6.1
ObservationModel	-24.5	8.3

Table 5.2: Leave-one-out cross-validation model comparison of the three observation models.

	Estimate	Est.Error	L-95% CI	U-95% CI	Eff.Sample	Rhat
<i>A</i> _Intercept	0.11	0.06	0.00	0.24	741	1.00
<i>B</i> _Intercept	6.25	0.75	4.76	7.73	357	1.00
<i>t</i> ₀ _Intercept	45.95	1.94	42.32	50.0	281	1.00
σ _Intercept	0.22	0.02	0.22	0.29	634	1.00

Table 5.3: Summary of the variable estimates for the process model.

	Estimate	Est.Error	L-95% CI	U-95% CI	Eff.Sample	Rhat
<i>K</i> _Intercept	0.24	0.09	0.04	0.43	110	1.02
<i>V</i> _{<i>m</i>} _Intercept	0.78	0.02	0.75	0.82	262	1.00

Table 5.4: Summary of the variable estimates for the observation model.

State Estimate			Forecasted Estimates	
Field-Site	R^2	$RMSE$	R^2	$RMSE$
F1-S1	0.903	0.67	0.626	1.227
F1-S2	0.923	0.59	0.807	1.0
F1-S3	0.951	1.3	0.898	1.5
F1-S1	0.937	1.1	0.92	1.6
F2-S2	0.761	0.86	0.788	0.66
F2-S3	0.922	1.4	0.809	1.72

Table 5.5: Summary of the R-squared and RMSE values for the estimated states and the predicted states.

Particle Filter

The particle filter reconstruction of the time-series using the process model and the full set of GNDVI observations from each individual field location sampled from the 2017 data had R^2 values ranging from 0.761 to 0.951, with a root mean square error (RMSE) ranging from 0.59 to 1.4 PFC. The results from the model fitting for each field location are given in Table 5.5 and shown in Figure 5.8 where the true PFC is shown in gray and the PF estimated PFC is shown in blue.

The results from using the PF to predict PFC are shown in Figure 5.8, where the model reconstruction using only a subset of the data is shown in green, and the forecasted percentage flower cover values are shown in orange. The forecasted (predicted) R^2 values range from 0.626 to 0.92, with RMSE values ranging from 0.66 to 1.72 percent flower over the forecasted range.

The results from using the PF model to predict the accelerated dataset are shown in Figure 5.9. The predicted R^2 values range from 0.131 to 0.961, with RMSE values ranging from 0.72 to 2.6 percent (Table 5.6).

Accelerated State Estimate		
Field-Site	R^2	$RMSE$
F1-S1	0.918	0.98
F1-S2	0.921	1.0
F1-S3	0.961	0.75
F1-S1	0.89	1.0
F2-S2	0.574	0.72
F2-S3	0.131	2.6

Table 5.6: Summary of the R-squared and RMSE values for the predicted states of an accelerated dataset.

5.5 Discussion

One of the advantages of using a Bayesian statistical method to estimate model fit is that we can do multi-level data fitting and get group-level estimates. The process model results show that accounting for site specific variability significantly improves our model fit as shown by a much larger *elpd_diff* than *se_diff*. The method used to find the best-fit process model can easily be expanded to work for more complex multi-level and multi-variate models. The current model estimates could further be improved by combining both the process and observation models into the same model, which would properly propagate estimates of the parameters between the models. The incorporation of control variables in the process model, such as growing degree days, has the potential to further improve the prediction step because it is an influential variable for alfalfa growth and development (Ben-Younes, 1992). The process model may also benefit from fitting a different function, such as one that is also able to track the fall-off in percentage flower cover. The variations in bloom onset timing from field to field can be caused by different alfalfa varieties, management practices, and environmental conditions. The variation due to crop variety can be

accounted for by collecting more ground validation data and fitting the cultivars as another parameter in the multi-level process model. The field-level variations, such as management practices and environmental conditions, can also be accounted for by adding new measurements into the observation model.

The observation model has a very quick initial rise in VI value as bloom up begins, but the VI value reaches a maximum around 4% flower cover. A VI that has a unique value throughout the bloom percentage range would be more ideal for an observation model. A surprising result of the RFR done on the 2017 data is that the Red Edge (RE) band is the most significant predictor. This band was not included in the 2016 regression (Chapter 2) because the Sentinel-2 satellite does not collect that band at a 10 m spatial resolution. Accordingly, none of the vegetation indices tested include the RE band. A band combination that is more sensitive to flowers could be determined using a RFR by including VIs with the RE band.

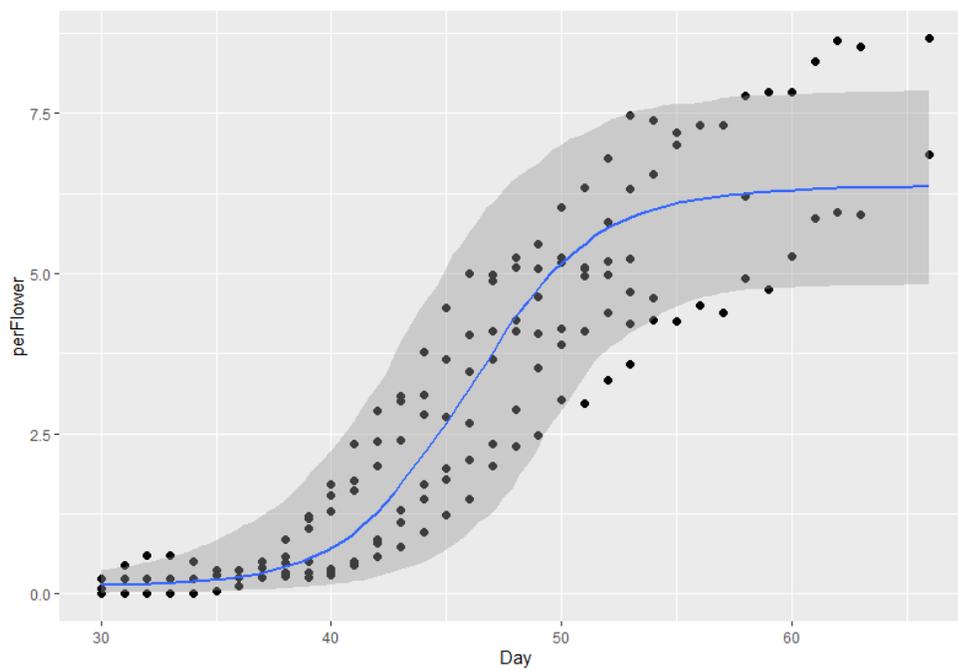
The Sentinel-2 data was not used in the model fitting because there were only 2 days of data collection during the 2017 growing season that coincided with satellite imagery and due to a lack of ground validation during the 2018 growing season. The satellite imagery provides us with a VI value over the entire field, however ground validation points for percentage flower cover are needed to relate the VI value to PFC.

The PF did a good job of estimating the PFC using the process model and GNDVI observations. However, the results of the PF prediction for most fields was not yet sufficiently accurate to base field management decisions as the predictions lowest RMSE is 0.66% PFC which is a large estimation error compared to the fields maximum PFC values which ranged from 4.5 to 7.5%. As discussed above, using a process model that allows for the fall off in PFC could improve the predictions. The PF had similar

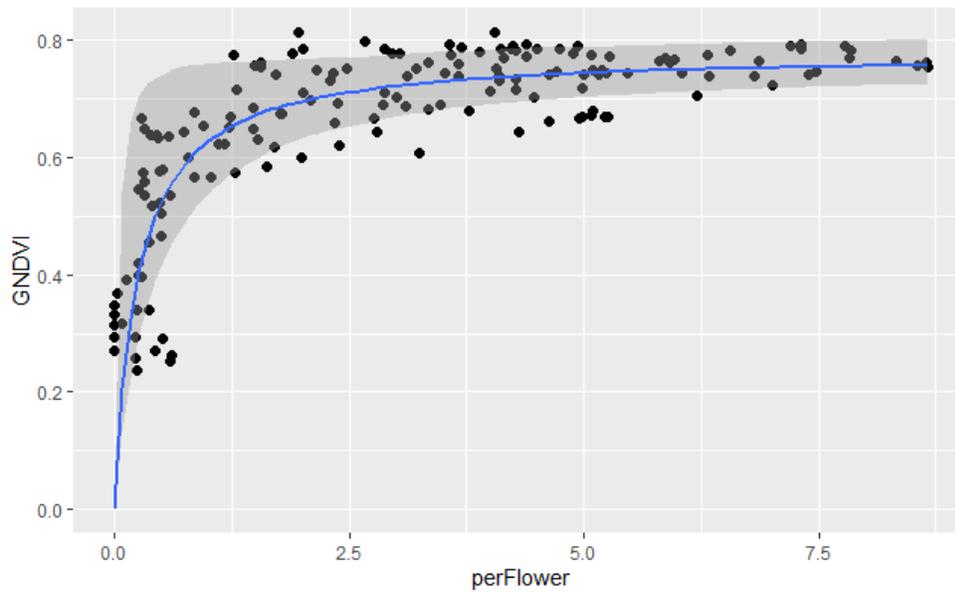
accuracy to the original dataset when estimating PFC on an accelerated dataset. The results indicate that the VI observations are related to the bloom percentage on the plants and that a varying crop growth rate can be accounted for using only observations. Considering the RFR showed the spectral signatures only explain up to 82% of the variability in PFC, the results were good for a simple model and show that further analysis is justified. The careful incorporation of other measurements, such as temperature and plant water stress, which are also known to influence the onset of bloom (Mueller, 2008), would likely improve the estimation and prediction results. Finally, future work to incorporate satellite-borne VI observations into the prediction model could greatly increase the usefulness of this method because once a reliable observation model is established, the publicly available satellite measurements can observe alfalfa seed crop fields throughout a growing season on a regional scale.

5.6 Conclusions

The variability in data collection methods between years complicated the observation model building procedure. Confidence in the process and observation models is obtained by focusing on a single year (2017) of data because it limited the influence of different data collection techniques and sensors. Despite the simplicity of the process and observation models, much of the variability in the PFC from site to site was accounted for using observations to update the system in a PF approach. The PF method is demonstrated to be capable of providing accurate estimates of PFC when applied to the individual 2017 field data with R^2 values as high as 0.95. The results also highlight the near-real time prediction potential of this method where 10 days worth of observations are used to forecast the percentage flower cover of the next 20 days with R^2 values as high as 0.92.



(a) Process Model fit



(b) Observation Model fit

Figure 5.7: The best performing model fit is shown as a blue line with the 95% confidence interval range shown in grey.

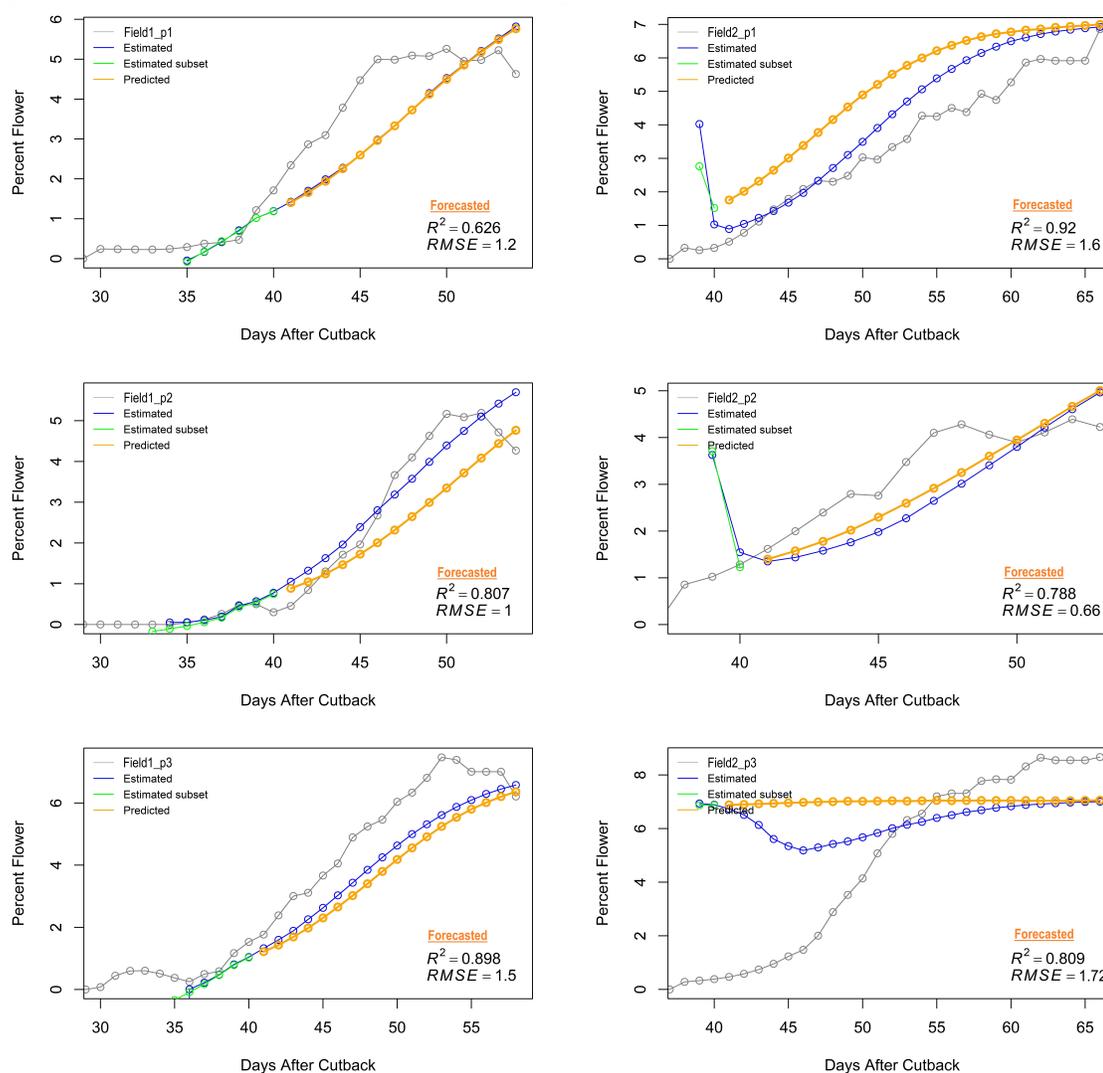


Figure 5.8: Field 1: sites 1, 2, 3 (left column, top to bottom). Field 2: sites 1, 2, 3 (right column, top to bottom). The process and observation models are passed into a particle filter algorithm along with the observations from individual field sites. The prediction line (blue) is estimated by combining the process model with the most current observation, in this case a GNDVI value. The observations are subset to only include the first 10 observation values (green line). The observation subset is used to forecast state estimates for the remainder of the data collection period (orange line). The accuracy of the forecasted predicted percentage flower cover values (orange) is given by the R^2 and $RMSE$ value.

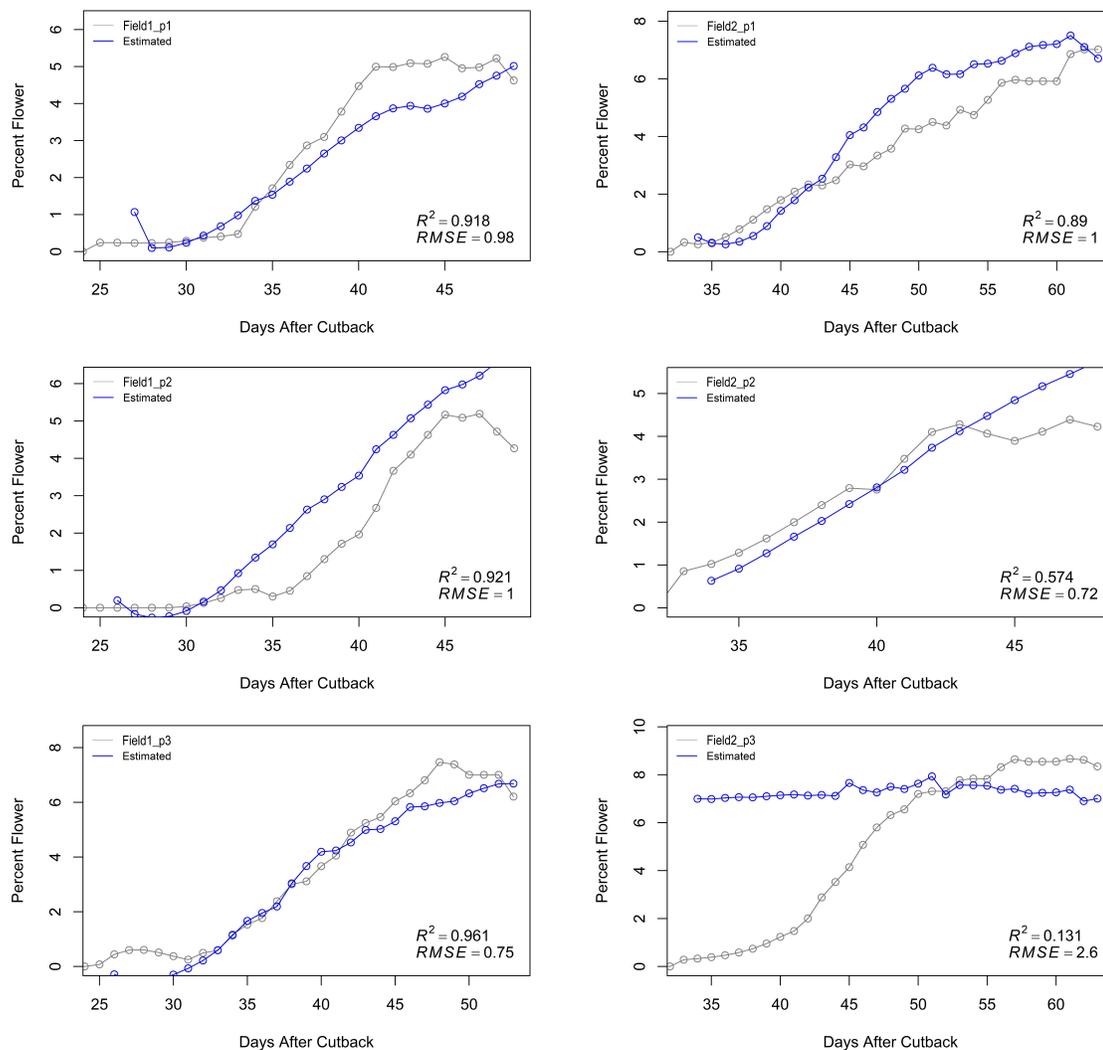


Figure 5.9: Field 1: sites 1, 2, 3 (left column, top to bottom). Field 2: sites 1, 2, 3 (right column, top to bottom). The process and observation models are passed into a particle filter algorithm along with the observations from individual field sites. The prediction line (blue) is estimated by combining the process model with the most current observation, in this case a GNDVI value. The accuracy of the predicted percentage flower cover values (blue) is given by the R^2 and $RMSE$ value.

CHAPTER 6:

DISCUSSION, CONCLUSIONS, AND FUTURE WORK

6.1 Discussion and Conclusions

This thesis contains five chapters related to the monitoring of alfalfa seed crop bloom. Data supporting this research was collected at four different locations throughout southwestern Idaho in 2016, 2017, and 2018. In Chapter 1 the reader was introduced to the biology of alfalfa, the potential of fields to produce larger quantities of seed by ideal timing of pollinator release, and the increasing demand for quality seed. A background on related remote sensing technologies is provided and how these measurements can be used to improve seed yield and quality through statistical inference is explained.

Chapter 2 covers the determination of electromagnetic bands sensitive to alfalfa bloom. This initial study forms the basis for the other chapters and sets the groundwork for if and how well spectral signatures can be used to monitor alfalfa seed crops phenological progression. To make this determination, hyperspectral and RGB imagery were collected at four locations during the 2016 growing season. A percentage flower cover (PFC) estimate was determined by manual classification of the RGB images using Sample Point image analysis software. The hyperspectral data were

combined into broadband wavelengths, similar to the 10 m spatial resolution bands on the Sentinel-2 satellites, and these wide bands were further transformed into Vegetation Indices (VI). The PFC and spectral data were used as the prediction and response variables in a random forest regression (RFR) to determine the most significant bands. This regression showed that tracking buds using the spectral signature alone was not possible as the R^2 value reached a maximum of 38%. This low value is likely explained by the similarities in spectral signatures between the buds and vegetation. Our focus then turned toward monitoring the alfalfa flowers, which had an R^2 value of up to 82% for a single VI. Examination of the most significant variables, as determined by increasing mean squared error, shows that typically a VI performed better than individual wavelengths, but also that many wavelengths below 400 nm were significant. The wavelengths below 400 nm were not included in further research due to the inability for satellites to make accurate measurements at these wavelengths due to atmospheric scattering. The 30 most significant VIs for tracking flower progression were further examined in order to determine which three individual bands did the best job to track flower development. The three bands determined from this RFR were Blue, Green, and NIR. These bands are used in a custom field sensor during the 2018 growing season.

Chapter 3 discusses the development of field station sensors for the acquisition of NIR-green-blue digital imagery for alfalfa seed crop monitoring. From a development perspective, two different sensor enclosure designs were tested with a reliable, storm-proof way to collect future data at agricultural fields or any remote outdoor location being determined. Each field station included a Raspberry Pi (RPi) and control board, a temperature and humidity sensor, a wireless cellular modem and data plan, a

battery bank and solar panel, a MapIR Survey3W camera, a gray calibration card, and one station at each field had a pyranometer. The RPi sensor functioned throughout the growing season and has the potential to expand with more periphery sensors. We found the temperature and humidity sensors to be reliable and accurate, as compared to a high-quality commercial weather monitoring system. Having a wireless data connection on the system was a valuable addition for troubleshooting and further backing up the data. The voltaic battery bank performed well and provided plenty of power and storage with the use of a solar panel for this application. The MapIR Survey3W camera has excellent specifications and the compact size made it easy to house. The RAW image data from the camera required extra processing in order to get the best signal, and while a relative radiometric calibration was used, these values are not comparable to other sensors. The gray card that was going to be used for image calibration quickly deteriorated once at the field and a method for shielding the card when it is not being imaged would greatly enhance its usefulness. The pyranometers were successful in collecting data; however, the single value output from a broad spectrum measurement was not useful for individual band calibration. Instead, incoming solar radiation measurements should be taken at each specific band the camera is collecting. This would allow for an accurate reflectance calibration. Overall, a more all-inclusive camera such as the Tetracam Auk 3 would make data collection, processing, and interpretation easier.

Chapter 4 details the image processing and classification done on the 2017 and 2018 imagery. The 2017 imagery was collected with a multi-spectral Micasense Red-edge camera. Although this camera suffered from misaligned band images due to being used below the recommended height, band registration using a discrete fourier

transform technique was able to bring the images mostly into alignment. The remaining misalignment was due to lens distortion and the image registration could be further improved in the future by applying a lens distortion correction before alignment. The Micasense camera has a reliable sensor whose digital numbers were able to be converted into units of radiance. The reflection panel values were manually extracted but were not used in the final analysis due to large daily variations. The 2018 data were collected using a custom field sensor described in Chapter 3. The camera used was a MapIR Survey3W which recorded RAW imagery. These images were converted into 12-bit TIFFs, which suffered from a position-dependent light intensity falloff remnant called vignetting. An anti-vignetting correction factor was applied to each pixel based on laboratory experiments of the lens vignette. After this each image was normalized and converted into a pseudo-radiance value so the individual bands had more meaningful relationships. These pseudo-radiance images were used to create a VI time-series. The images were also used in a supervised image classification algorithm to get PFC. The 2018 field sensors lack of an RGB camera made Sample Point image classification on the false color images difficult; thus a Support Vector Machine (SVM) classification was done in an attempt to get PFC over the growing season. While the classification seemed to work, having high user and producer accuracies, as well as a definitive trend for the onset of flower, the comparison with Sample Point data from the other years showed an earlier onset of flower which was not realistic. The 2017 Micasense multispectral imagery, along with the PFC from Sample Point image analysis done on the RGB images at the same locations proved to be the most reliable data for the prediction analysis.

Chapter 5 introduces the use of a particle filter modeling approach to estimate

and predict alfalfa seed crop flower percentage cover values using VI time series. In order to use this state-space modeling approach, two mathematical approximations, one for the expected behavior of the field, known as a process model, and the other for the observational relationship to this behavior, known as an observation model, is required. A subset of the data discussed in the previous chapters is used to create these process and observation models for the expected growth of alfalfa seed crops. A logistic function is transformed into its state-space representation and discretized before being used as the process model and a Michaelis-Menten function is used for the observation model. The data is first fit to their respective functions using a non-linear least squares estimate. These estimates of the parameters are used as priors in a nonlinear multilevel Bayesian model, which has the advantage of being able to separate out group-level effects. The process and observation models, along with the observation data from a single field is input into a particle filter. The results of the PF, using data from the final cutback through peak bloom conditions show the potential to achieve an accurate estimation with R^2 as high as 0.97. The particle filter is also used to predict the PFC out to 20 days using only the first 10 observations, with R^2 values as high as 0.91. This analysis demonstrates a proof of concept for the estimation and prediction of PFC at alfalfa seed crop fields using remote sensing observations and state-space modeling.

6.2 Future work

For future work it is worth running the random forest regression again including the Rededge band and associated VI on the 2016 and 2017 data.

Regarding the custom field sensors, if they are to be used again in future field studies they would benefit from the following upgrades:

- include a RGB camera,
- include a soil-moisture sensor,
- some protection for the gray calibration card, and
- use the Tetracam Auk 3 camera with down-welling light sensor.

More work that can be done includes changing the function used in the process model to account for the fall-off in PFC, testing the use of growing degree days instead of days after cutback, and further validating the models in future campaigns. The lack of reliable ground validation during the 2018 growing season made the data collected unusable. More data should be collected to validate the findings from the 2017 data set. Images and measurements taken with high-temporal sampling rates from in-situ stations located within a field can be used for ground truthing and data points for the the process and observation models. The ability of a UAV to capture the variability within an entire field can be compared to satellite imagery in order to determine the viability of using freely available satellite measurements in the bloom prediction model. Using a multivariate regression to including both the process and observation models in the same model fit would ensure the uncertainty in parameter estimation is correctly propagated. The observation model could be expanded upon by including Sentinel-2 satellite imagery and the particle filter could then be tested using the combined observation model as well as only the Sentinel-2 observations.

The results from this study provide support for future efforts using state-space models and remote sensing platforms for making observations. While the simple bloom prediction model performed well; the greatest immediate value may come from modeling inter-field variability. Providing an alfalfa seed crop field manager with a

map of where bloom is likely to occur first would allow for more strategic placement of pollinator housing. This would likely decrease the quantity of pollinators needed to adequately pollinate a field and improve pollinator health.

REFERENCES

- Apogee Instruments, Inc. 2018 (05). *Apogee SP-420 Manual*. Tech. rept. Apogee Instruments, Inc.
- Arulampalam, M. Sanjeev, Maskell, Simon, Gordon, Neil, & Clapp, Tim. 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, **50**(2), 174–188.
- Atkinson, Peter M., Jeganathan, C., Dash, Jadu, & Atzberger, Clement. 2012. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*, **123**, 400–417.
- Averbuch, A., & Keller, Y. 2002 (May). FFT based image registration. *Pages IV–3608–IV–3611 of: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4.
- Base, MicaSense Knowledge. 2015. *MicaSense RedEdge Multispectral Camera User Manual*.
- Base, MicaSense Knowledge. 2017. *Use of Calibrated Reflectance Panels For RedEdge Data*.
- Base, MicaSense Knowledge. 2019. *RedEdge Camera Radiometric Calibration Model*.

- Bayer, Bryce E, & Rochester, N Y. 1976. *A sensing array for color imaging includes individual*. Tech. rept. United States Patent.
- Belgiu, Mariana, & Drgu, Lucian. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, **114**(apr), 24–31.
- Ben-Younes, Mongi. 1992. *Modeling the Temperature-Mediated Phenological Development of Alfalfa (Medicago sativa L.)*. Doctor of Philosophy, Oregon State University.
- Booth, D Terrance, Cox, Samuel E, & Berryman, Robert D. 2006. POINT SAMPLING DIGITAL IMAGERY WITH 'SAMPLEPOINT'. *Environmental Monitoring and Assessment*, **123**, 97–108.
- Breiman, Leo. 2001. *Random Forests*. Tech. rept. University of California, Berkeley.
- Brkner, Paul-Christian. 2018. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, **10**(1), 395–411.
- Camera, MAPIR. 2016. *MAPIR Survey3 Camera Datasheet*.
- CFAC. 2004. *Commodity Fact Sheet: Alfalfa*. Tech. rept. California Foundation for Agriculture in the Classroom.
- Chen, Yeppei. 2018. Improved relative radiometric normalization method of remote sensing images for change detection. *Journal of Applied Remote Sensing*, **12**(04), 1.

- Comer, Robert, Kinn, Gerry, Light, Don, & Mondello, Charles. 1998. Talking Digital. *Photogrammetric Engineering & Remote Sensing*, 1139–1142.
- Crassidis, J., & Junkins, J. 2012. *Optimal Estimation of Dynamic Systems*. Chapman and Hall/CRC.
- Croston, Ben. 2012. *RPi.GPIO: A module to control Raspberry Pi GPIO channels*.
- De Bernardis, Caleb, Vicente-Guijalba, Fernando, Martinez-Marin, Tomas, & Lopez-Sanchez, Juan M. 2016a. Contribution to Real-Time Estimation of Crop Phenological States in a Dynamical Framework Based on NDVI Time Series: Data Fusion with SAR and Temperature. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **9**(8), 3512–3523.
- De Bernardis, Caleb, Vicente-Guijalba, Fernando, Martinez-Marin, Tomas, & Lopez-Sanchez, Juan. 2016b. Particle Filter Approach for Real-Time Estimation of Crop Phenological States Using Time Series of NDVI Images. *Remote Sensing*, **8**(7), 610.
- De Bernardis, Caleb G, Vicente-Guijalba, Fernando, Martinez-Marin, Tomas, & Lopez-Sanchez, Juan M. 2014. Estimation of Key Dates and Stages in Rice Crops Using Dual-Polarization SAR Time Series and a Particle Filtering Approach. *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*.
- Drummond, James. 2017. *Four Parameter Logistic Regression*.
- Durrant-Whyte, Hugh, & Bailey, Tim. 2006. *Simultaneous Localisation and Mapping*

- (SLAM): *Part I The Essential Algorithms*. Tech. rept. Australian Centre for Field Robotics.
- Fan, Hui, Fu, Xiaohua, Zhang, Zheng, & Wu, Qiong. 2015. Phenology-based vegetation index differencing for mapping of rubber plantations using landsat OLI data. *Remote Sensing*, **7**(5), 6041–6058.
- Ge, Shaokui, Everitt, James, Carruthers, Raymond, Gong, Peng, & Anderson, Gerald. 2006. HYPERSPECTRAL CHARACTERISTICS OF CANOPY COMPONENTS AND STRUCTURE FOR PHENOLOGICAL ASSESSMENT OF AN INVASIVE WEED. *Environmental Monitoring and Assessment*, **120**, 109–126.
- Geerling, Jeff. 2017. *Power Consumption Benchmarks — Raspberry Pi Dramble*.
- Gelman, Andrew, & Rubin, Donald B. 1992. *Inference from Iterative Simulation Using Multiple Sequences Inference from Iterative Simulation Using Multiple Sequences**. Tech. rept. University of California, Berkeley.
- Gohlke, Christoph, & Tyc, Matej. 2018. *imregdft Documentation*.
- Halim, R, Buxton, D, Hattendorf, M, & Carlson, R. 1989. Water-Stress Effects on Alfalfa Forage Quality After Adjustment for Maturity Differences. *American Society of Agronomy*, **81**.
- Hodges, Tom. 1991. *Temperature and water stress effects on phenology*. CRC Press, Inc.
- Hufkens, Koen, Friedl, Mark, Sonnentag, Oliver, Braswell, Bobby H., Milliman, Thomas, & Richardson, Andrew D. 2012. Linking near-surface and satellite remote

- sensing measurements of deciduous broadleaf forest phenology. *Remote Sensing of Environment*, **117**, 307–321.
- Husman, Stephen H. 2015. *Growing Alfalfa for Seed in Arizona*.
- Jeffreys, Harold. 1946. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and physical sciences*, **186** **1007**, 453–61.
- Landmann, Tobias, Piironen, Rami, Makori, David M., Abdel-Rahman, Elfatih M., Makau, Sospeter, Pellikka, Petri, & Raina, Suresh K. 2015. Application of hyper-spectral remote sensing for flower mapping in African savannas. *Remote Sensing of Environment*.
- Lee, W S, Alchanatis, V, Yang, C, Hirafuji, M, Moshou, D, & Li, C. 2010. Computers and Electronics in Agriculture Sensing technologies for precision specialty crop production. *Computers and Electronics in Agriculture*, **74**, 2–33.
- Major, D. J., Hanna, M. R., & Beasley, B. W. 1990. Photoperiod response characteristics of alfalfa (*Medicago sativa* L.) cultivars. *Canadian Journal of Plant Science*, **71**(1), 87–93.
- MAPIR Survey3 Camera. 2017. *Survey3W Camera - NIR+Green+Blue (NGB, ENDVI) - MAPIR CAMERA*.
- Marino, Stefano, & Alvino, Arturo. 2014. Proximal sensing and vegetation indices for site-specific evaluation on an irrigated crop tomato. *European Journal of Remote Sensing*, **47**(47), 271–283.

- McElreath, Richard. 2015. *Statistical Rethinking*. CRC Press.
- Mcgregor, S E. 1976. Insect Pollination Of Cultivated Crop Plants Introduction: Economics of Plant Pollination Flowering and Fruiting of Plants Hybrid Vigor in Plants and its Relationship to Insect Pollination Wild Bees and Wild Bee Culture Wild Flowers and Crop Pollination P. *The First and Only Virtual Beekeeping Book*.
- McIntosh, Margrit E. 2002. Plant size, breeding system, and limits to reproductive success in two sister species of *Ferocactus* (Cactaceae). *Plant Ecology*, **162**(2), 273–288.
- Michaelis, Von L, Maud Menten, Miss L, Goody, Roger S, & Johnson, Kenneth A. 1913. *The Kinetics of Invertase Action*. Tech. rept. Max-Planck Institute of Molecular Physiology, Otto-Hahn-Strasse.
- Morain, SA, & Zanoni, Vicki. 2004. Joint ISPRS/CEOS-WGCV task force on radiometric and geometric calibration. *International Archives of Photogrammetry and Remote Sensing*, **35**(01).
- Mueller, Shannon C. 2008. PRODUCING QUALITY ALFALFA SEED FOR THE FORAGE INDUSTRY. *In: Alfalfa Symposium Proceedings*.
- Mulla, David J. 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*.
- Munk, Gail. 1999. *Using Growing Degree Days for Alfalfa Production*. Tech. rept. University of Nevada, Reno.
- Nagai, Shin, Ichie, Tomoaki, Yoneyama, Aogu, Kobayashi, Hideki, Inoue, Tomoharu, Ishii, Reiichiro, Suzuki, Rikie, & Itioka, Takao. 2016. Usability of time-lapse digital

- camera images to detect characteristics of tree phenology in a tropical rainforest. *Ecological Informatics*, **32**, 91–106.
- NASS. 2017. *AGRICULTURAL STATISTICS 2017*. Tech. rept. UNITED STATES DEPARTMENT OF AGRICULTURE NATIONAL AGRICULTURAL STATISTICS SERVICE.
- Noland, Reagan L., Wells, M. Scott, Coulter, Jeffrey A., Tiede, Tyler, Baker, John M., Martinson, Krishona L., & Sheaffer, Craig C. 2018. Estimating alfalfa yield and nutritive value using remote sensing and air temperature. *Field Crops Research*, **222**(April), 189–196.
- NSERC-CANPOLIN Canadian Pollination Initiative. 2006. *Medicago sativa. Best Management Practices for Pollination in Ontario Crops, 2006–2006*.
- Ottman, Michael, & Putnam, Daniel. 2017. Deficit Irrigation with Alfalfa. *Page 4 of: Western Alfalfa & Forage Symposium*.
- Panalytical, Malvern. 2017. *FieldSpec HandHeld 2 Spectroradiometer User Manual*. Tech. rept. September. Malvern Panalytical.
- Pearson, C J, & Hunt, L A. 1972. Effects of temperature on primary growth of alfalfa. *Canadian Journal of Plant Science*, **52**, 1007–1015.
- Pompilio, Loredana, Marinangeli, Lucia, Amitrano, Luigi, Pacci, Giacomo, D'andrea, Salvatore, Iacullo, Salvatore, & Monaco, Enrico. 2018. Application of the empirical line method (ELM) to calibrate the airborne Daedalus-CZCS scanner. *European Journal of Remote Sensing*, **51**(1), 33–46.

- R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raspberry Pi Pin Map. 2017. *Raspberry Pi 2 & 3 Pin Mappings - Windows IoT* — *Microsoft Docs*.
- Richards, John A, & Jia, Xiuping. 2006. *Remote Sensing Digital Image Analysis*. 4 edn. Springer.
- Richardson, Andrew D., Braswell, Bobby H., Hollinger, David Y., Jenkins, Julian P., & Ollinger, Scott V. 2009. Near-surface remote sensing of spatial and temporal variation in canopy phenology. *Ecological Applications*, **19**(6), 1417–1428.
- Sakamoto, Toshihiro, Yokozawa, Masayuki, Toritani, Hitoshi, Shibayama, Michio, Ishitsuka, Naoki, & Ohno, Hiroyuki. 2005. A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, **96**(3-4), 366–374.
- Sharratt, B.S., Sheaffer, C.C., & Baker, D.G. 1989. Base temperature for the application of the growing-degree-day model to field-grown alfalfa. *Field Crops Research*, **21**(2), 95 – 102.
- Srinivasa Reddy, B., & Chatterji, B. N. 1996. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, **5**(8), 1266–1271.
- Tagle, Casapia. 2017. *Study of radiometric variations in Unmanned Aerial Vehicle remote sensing imagery for vegetation mapping*. Ph.D. thesis, Lund University.

- Tong, Christopher, & Vendettuoli, Marie. 2017. *Parameterizations of the four-parameter logistic curve in software for estimating relative potency*. Tech. rept. USDA Center for Veterinary Biologics.
- USDA. 2015. United States Department of Agriculture National Agricultural Statistics Service. *2015 Crop Production Summary*.
- Vehtari, Aki, Gelman, Andrew, Gabry, Jonah, & Gelman, Andrew. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*, **27**, 1413–1432.
- Verhoeven, G. J. J. 2010. It's all about the format – unleashing the power of RAW aerial photography. *International Journal of Remote Sensing*, **31**(8), 2009–2042.
- Vicente-Guijalba, Fernando, Martinez-Marin, Tomas, & Lopez-Sanchez, Juan M. 2015. Dynamical approach for real-time monitoring of agricultural crops. *IEEE Transactions on Geoscience and Remote Sensing*, **53**(6), 3278–3293.
- Viña, Andrés, Gitelson, Anatoly A., Nguy-Robertson, Anthony L., & Peng, Yi. 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, **115**(12), 3468–3478.
- Weng, Qihao. 2002. Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. *Journal of Environmental Management*, **64**, 273–284.
- White, Michael A., & Nemani, Ramakrishna R. 2006. Real-time monitoring and short-term forecasting of land surface phenology. *Remote Sensing of Environment*, **104**(1), 43 – 49.

- Wu, Chaoyang, Gonsamo, Alemu, Gough, Christopher M., Chen, Jing M., & Xu, Shiguang. 2014. Modeling growing season phenology in North American forests using seasonal mean vegetation indices from MODIS. *Remote Sensing of Environment*, **147**, 79 – 88.
- Xue, Jinru, & Su, Baofeng. 2017. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, **2017**(may), 1–17.
- Yu, Wonpil. 2004. Practical anti-vignetting methods for digital cameras. *IEEE Transactions on Consumer Electronics*.
- Yuegao, Hu, & Cash, Dennis R. 2010. Global Status and Development Trends of Alfalfa. *In: ALFALFA MANAGEMENT GUIDE FOR NINGXIA*.

Appendices

APPENDIX A:
VEGETATION INDICES USED IN
REGRESSION ANALYSIS

Vegetation Indices	
NDVI	$(N - R)/(N + R)$
EVI	$2.5(N - R)/(N + 6R - 7.5B + 1)$
LAI	$3.618(EVI) - 0.118$
CVI	$(N/G)(R/G)$
MSR	$((N - R) - 1)/((\sqrt{N}/\sqrt{R}) + 1)$
OSAVI	$(N - R)/(N + R + 0.16)$
TDVI	$\sqrt{0.5 + ((N - R)/(N + R))}$
GNDVI	$(N - G)/(N + G)$
GLI	$(2G - R - B)/(2G + R + B)$
NG	$G/(N + R + G)$
NR	$R/(N + R + G)$
RVI	N/R
GRVI	N/G
DVI	$N - R$
GDVI	$N - G$
SAVI	$(1.5(N - R))/(N + R + 0.5)$
GSAVI	$(1.5(N - G))/(N + G + 0.5)$
GOSAVI	$(N - G)/(N + G + 0.16)$
MSAVI2	$0.5(2(N + 1) - \sqrt{(2N + 2)^2 - 8(N - R)})$
TVI	$0.5(120(N - G) - 200(R - G))$
MTVI	$1.2(1.2(N - R) - 2.5(N - G))$
RDVI	$(N - R)/(\sqrt{N} + \sqrt{R})$
MCARI2	$1.5(2.5(N - R) - 1.3(N - G))/\sqrt{(2N + 1)^2 - (6N - 5\sqrt{R})} - 0.5$
ENDVI	$((N + G) - 2B)/((N + G) + 2B)$
GIPVI	$N/(N + G)$
Flower Specific	
FVI	$(B/G)(N/G)$
FVI2	$(B/G)(R/G)$
FVI3	$(B - G)/(B + G)$
FVI4	$(N - B)/(N + B)$
FVI5	N/B
FVI6	B/G
FVI7	$N - B$
FVI8	$B - G$
FVI9	$B/(N + R + B)$
FVI10	$B/(N + B)$
FVI11	$B/(N + G + B)$
RGR	R/G
SIPI	$(N - B)/(N - R)$

Table A.1: Vegetation indices used in the Random Forest Regression.
Band abbreviations are N =NIR, R =Red, G =Green, B =Blue.