12-2019

Spatially-explicit Snap Bean Flowering and Disease Prediction Using Imaging Spectroscopy from Unmanned Aerial Systems

Ethan W. Hughes
ewh5693@rit.edu

Spatially-explicit Snap Bean Flowering and Disease Prediction Using Imaging Spectroscopy from Unmanned Aerial Systems

by

Ethan W. Hughes

B.S., Widener University, Chester, PA, 2014

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology

December, 2019

Signature of the Author _____

Accepted by _____
Coordinator, M.S. Degree Program Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

M.S. DEGREE THESIS

The M.S. Degree Thesis of Ethan W. Hughes
has been examined and approved by the
thesis committee as satisfactory for the
thesis required for the
M.S. degree in Imaging Science

Dr. Jan van Aardt, Thesis Advisor

Dr. Carl Salvaggio

Dr. Emmett Ientilucci

Date

THESIS RELEASE PERMISSION
ROCHESTER INSTITUTE OF TECHNOLOGY
CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE

Title of Thesis:

Spatially-explicit Snap Bean Flowering and Disease Prediction Using Imaging Spectroscopy from Unmanned Aerial Systems

I, Ethan W. Hughes, hereby grant permission to Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature _____

Date

Acknowledgments

I'd like to thank those who have helped me along my journey through higher education.

"Practice the things that are hard. That's how you get better."

- Rob Ickes

Spatially-explicit Snap Bean Flowering and Disease Prediction Using Imaging Spectroscopy from Unmanned Aerial Systems

Publication No. _____

Ethan W. Hughes, M.S.
Rochester Institute of Technology, 2019

Supervisor: Dr. Jan van Aardt

Abstract

Sclerotinia sclerotiorum, or white mold, is a fungus that infects the flowers of snap bean plants and causes a subsequent reduction in snap bean pods, which adversely impacts yield. Timing the application of white mold fungicide thus is essential to preventing the disease, and is most effective when applied during the flowering stage. However, most of the flowers are located beneath the canopy, i.e., hidden by foliage, which makes spectral detection of flowering via the leaf/canopy spectra paramount. The overarching objectives of this research therefore are to i) identify spectral signatures for the onset of flowering to optimally time the application of fungicide, ii) investigate spectral characteristics prior to white mold onset in snap beans, and iii) eventually link the location of white mold with biophysical (spectral and structural) metrics to create a spatially-explicit probabilistic risk model for the appearance of white mold in snap bean fields. Spectral angle mapper (SAM) and ratio and thresholding (RT) were used to detect pure vegetation pixels, toward creating the flowering detection models. The pure pixels then were used with a single feature logistic regression (SFLR) to identify wavelengths, spectral ratio indices, and normalized difference indices that best separated the flowering classes. Features with the largest c-index were used to train a support vector machine (SVM) and were then applied to imagery from a different growing

season to evaluate model robustness. This research found that single wavelength features in the red (600-700 nm, with a peak at 680 nm) discriminated and predicted flowering up to two weeks before visible flowering occurred, with c-index values above 90%. Structural metrics, such as leaf area index (LAI), have been proven to correlate with white mold incidence, so linear and multivariate regressions were used to ingest spatial- and spectral-related features, derived from the imaging spectroscopy data, and predict ground truth LAI data. These features included raw spectral reflectance values, pixel density, normalized difference index (NDVI), green normalized difference index (GNDVI), and the enhanced vegetation index (EVI). Indicators in the green and red-red edge portion of the spectrum exhibited coefficients of determination (CoD) greater than 0.7. The spatial and spectral indices had CoDs and root mean squared errors (RMSE) ranging from 0.422-0.565 and 0.817-0.942, respectively. The top 28 features were used in a multivariate regression to predict LAI and the results showed a maximum adjusted CoD of 0.849, with an RMSE of 0.390. Future work should include raw reflectance values, LAI correlated spectral features, as well as auxiliary in-field measurements (degree days, average rainfall, average temperature) in the creation of a white mold risk model.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Snap beans represent the fifth largest crop in the United States in terms of acres planted [3], and New York State is the second largest producer of fresh market and processing snap beans [4]. However, like many board-acre crops, snap beans are susceptible to a variety of biological stressors, with the most detrimental disease being the fungus *Sclerotinia sclerotiorum*, also known as white mold. White mold infects the flowers of the snap beans, thereby causing premature pod abscission and contamination from mycelia, which has a significant impact on crop yield. In fact, Shah et al. (2002) and McCreary et al. (2016) found snap beans to be among the most susceptible to white mold among the bean varieties [5][6]. White mold control is centered around protecting the flowers of the snap beans, since the white mold spores enter the plants via the open flowers [5][6]. Some options for control include widening crop row spacing and decreasing plant populations (density), the use of nitrogen fertilized irrigation, and selection of upright snap bean varieties. However, when used alone, these methods have proven to be insufficient in preventing infection and "method intensity/level" are inversely proportional to yield [7][8][9][10][11]. Other pro-active methods for white mold control in-

clude the application of prophylactic fungicide when 10% of the individual bean plants have at least one flower; however, research by Lehner (2017) cites sub-optimal timing of the fungicide as the leading cause of white mold contamination [12]. To this end, select research has been devoted to create risk models for bean crops and associated fungal diseases, focused mostly on more advanced statistical techniques or modeling approaches.

Multivariate analysis, for example, has been used to create white mold risk models for soy and snap bean fields [13][14]. Environmental factors, such as consecutive days with rain and average temperature, have shown strong correlations with white mold severity [14]. McDonald and Boland (2004) discussed a New York-based disease forecasting system for white mold in snap beans, based on canopy density, soil moisture, and precipitation [15]. Logistic Regression [16] also has been applied to predict risk of white mold in soy bean [17], grey leaf spot [18], and Stewart's wilt on corn [19]. For example, Harikrishnan and del Río (2008) used climate variables and logistic regression modeling to predict white mold risk on dry bean in North Dakota [20]. The authors obtained a true positive ratio and accuracy of 0.79 and 0.91, respectively, for white mold predicted-observed incidences. However, environmental models like the above-mentioned are constrained heavily by the deployment of accurate in-field weather monitoring equipment [21], and the outputs of the risk models are on a per-field basis. Our intent thus was to develop spatially-explicit, remote sensing spectral-based indicators of white mold risk that could

be incorporated in such risk models, toward wall-to-wall decision support information products, which would enable judicious and cost-effective white mold management strategies. The first step in such a remote sensing approach entails accurate detection of the flowering bean plants.

The flowers of the snap beans are primarily located beneath the canopy of the plant; therefore, spectral detection of snap bean flowering at the canopy level is of primary interest. Research by Sankaran et al. (2010) and Mahlein (2013) concluded that spectroscopy-based disease signatures exist, with the near-infrared (NIR) spectral regions being particularly useful [22][23]. Multivariate analysis and machine learning techniques have been used to classify diseased plants before using spectral data sets. Examples include Rumpf et al. (2010) [24], who used ground-based spectroradiometers and support vector machines (SVM) to classify *Cercospora* leaf spot, powdery mildew, and leaf rust on sugar beets with accuracies that ranged from 85.7-96.5%, and Delalieux et al. (2007), who used a single-feature logistic regression (SFLR) over a wavelength range from 350-2500 nm to detect biotic stress (*Venturia inaequalis*) in apple trees. Authors of the latter study found c-index values greater than 0.8 when classifying diseased plants, at stages well before the infection became visually identifiable (one week after infection). Early work by Hughes et al. [2] used a combination of SFLR and SVM to create probability maps for snap bean flowering. This research thus aims to extend past work on snap bean flowering detection towards prediction.

The second step toward development of a comprehensive white risk model entails an assessment of canopy structure, i) via spectral features that are correlated with leaf area index (LAI) or ii) direct assessment of plant structure, e.g., structure-from-motion (SfM) and light detection and ranging (lidar) approaches. Here we focus on the first approach, i.e., spectral-based assessment of canopy structure and complexity, with a number of previous studies that have shown the potential of such an approach. For example, Durbha et al. (2007) used data from the multi-angle imaging spectroradiometer (MISR) instrument, along with ground truth LAI measurements of wheat, to establish regressions between LAI (dependent variable) and the NIR and red bands. R^2 values ranged from 0.83-0.96, with root mean square errors (RMSE) between 0.20-0.99 [25]. Examples in forest ecosystems abound, with Berterretche et al. (2005) using ordinary least-squares regression to investigate relationships between reflectance values from Landsat ETM+ data and ground truth LAI from boreal forests. Results showed R^2 values of 0.52 and 0.47 for bands 5 (1.55-1.75 μm) and 7 (2.08-2.35 μm), respectively [26]. Bands 3 and 4 (0.63-0.69 μm and 0.76-0.90 μm, respectively) from Landsat-5 TM data were used by Chen et al. (1996) to establish linear regressions between normalized difference vegetation index (NDVI) and ground-truth LAI for boreal conifer forests [27]. Results yielded R^2 values of 0.5 and 0.42 for late spring and mid-summer data, respectively. Research by Colombo et al. (2003), for a more expansive set of crops, used satellite data from IKONOS to find regressions between spectral vegetation indices and LAI for different types of vegetation (forests,

plantations, vineyards, soy beans, and corn). R^2 values varied between 0.6-0.8 for the individual vegetation types and 0.33 across all types of vegetation [28]. This result by Colombo et al. (2003) lends itself to the idea that spectral-to-LAI regressions could be crop specific. The aim of our research therefore is to use linear and multivariate regression techniques to find relationships between raw reflectance values, spectral vegetation indices, and spatial features to estimate LAI in snap beans towards a spatially-explicit mold risk model.

The specific, overarching objectives of this research were to i) identify spectral signatures for the onset of flowering to optimally time the application of fungicide, ii) investigate spectral characteristics prior to white mold onset in snap beans, and iii) link the location of white mold with biophysical (spectral and structural) metrics to eventually create a spatially-explicit probabilistic risk model for the appearance of white mold in snap bean fields. The next chapter (Chapter 2) focuses on the discrimination and prediction of flowering in snap beans, as well as a first step toward relating spectral features to structural metrics, such as LAI. The study conclusions are presented in Chapter 3. Each chapter will reference the past studies, described above, in order to contextualize our results. We hypothesized that detection of flowering occurrence will be possible using a select, downsampled set of spectral variables, even at the plant canopy-level, and also that white mold infection may cause distinct physiological impacts in impacted plants, with such impacts allowing for discrimination across visible and NIR wavelengths. If successful, such a

flower detection algorithm, canopy structure assessment, and eventual mold prediction approach could contribute to more judicious application of preventative measures, resulting in reduced environmental impacts, increased yields, and optimized profits at the farm-level.

Chapter 2

Toward development of a UAS-based white mold risk model for snap beans: Flowering detection/prediction and plant canopy structure assessment

Snap beans are susceptible to a detrimental fungus called *Sclerotinia sclerotiorum*, or white mold, which infects the flowers of the snap beans [5][6]. Disease prevention is centered around protecting the flowers of the snap beans while the plants are flowering, and ensuring that the within-plant incubation environment is not conducive to mold growth. Prevention methods, related to manipulation of plant canopy structure, such as widening crop row spacing, decreasing plant density, using nitrogen fertilized irrigation, and the selection of upright snap bean varieties, have been found to be inadequate in disease prevention and reduce crop yield [7][8][9][10][11]. Other methods for white mold control include the use of fungicide. White mold can infect the flowers of snap beans within four days of flowering; therefore, research by Lehner (2017) cites sub-optimal timing of the fungicide application as the leading cause of white mold [12]. To this end, creating/implementing methods to predict and create risk models for disease (white mold in particular) is focused on using

statistical, multivariate, and machine learning based techniques.

Some research utilizes multivariate analysis to create white mold risk models for soy and snap bean fields using environmental factors, such as consecutive days with rain, average temperature, canopy density, and soil moisture[13][14][15]. Linear models such as logistic regression [16] also has been implemented to create predictive white mold models for soy bean [17], grey leaf spot [18], and Stewart's wilt on corn [19]. Research by Harikrishnan and del Río (2008), for example, used a combination of climate variables and logistic regression to predict incidence of white mold on dry bean in North Dakota [20]. A true positive ratio and accuracy of 0.79 and 0.91, respectively, was obtained for white mold predicted-observed incidences by the authors. These results, while promising, are restricted by the need for accurate in-field weather monitoring systems [21], and the predictive results from the models are on a per-field basis. Our intention was to create a remote sensing spectral-based solution for white mold risk that is spatially-explicit.

Research in spectroscopy-based remote sensing of vegetation shows that disease signatures exist, with the near-infrared (NIR) spectral regions being particularly useful [22][23]. Many researchers use ground based spectroradiometers to collect spectroscopy data for disease prediction. Such an example is work performed by Rumpf et al. (2010) who used support vector machines (SVM) to classify *Cercospora* leaf spot, powdery mildew, and leaf rust on sugar beets with accuracies that ranged from 85.7-96.5% [24]. Specifically, research

by Delalieux et al. (2007), who used a single feature logistic regression (SFLR) over a wavelength range from 350-2500 nm to detect biotic stress (*Venturia inaequalis*) in apple trees is of particular interest for this work. Delalieux found c-index values greater than 0.8 when predicting diseased plants a week after infection, but before physical sign of infection were visible.

Research has also shown that there is a strong correlation between LAI and white mold incidence [29]. Estimating bio-metrics related to vegetation from spectral data is a proven technique [30][31][32]. Rouse et al. have shown correlations greater than 89% between spectral features and moisture content, percent green estimate, and green biomass. Research centered around the spectral estimation of LAI specifically utilizes imagery from space-borne spectroradiometers. Research by Durbha et al. (2007) used data from the multi-angle imaging spectroradiometer (MISR) instrument along with ground truth LAI measurements of wheat to find relationships between NIR and red bands with LAI. R^2 values ranged from 0.83-0.96 with root mean squared errors (RMSE) between 0.20-0.99 [25]. Landsat ETM+ data was used by Berterretche et al. (2005) along with ground truth LAI from boreal forests to find an ordinary least squares regression to go from spectral to LAI space. Results showed R^2 values of 0.52 and 0.47 for bands 5 (1.55-1.75 μm) and 7 (2.08-2.35 μm), respectively [26]. Landsat-5 TM data (bands 3 and 4; 0.63-0.69 μm and 0.76-0.90 μm, respectively) was used by Chen et al. (1996) to find linear regressions between normalized difference vegetation index (NDVI)

and ground truth LAI for boreal conifer forests [27]. R^2 values of 0.5 and 0.42 were achieved for late spring and mid-summer data, respectively. Satellite data from IKONOS was used by Colombo et al. (2003) to find regressions between spectral vegetation indices and LAI for different types of vegetation (forests, plantations, vineyards, soy beans, and corn). Results for R^2 values varied between 0.6-0.8 for individual types of vegetation and 0.33 across combined vegetation types [28]. Colombo et al. concluded that spectral-to-LAI regressions should be crop specific. With that in mind, our research uses linear and multivariate regression techniques to find relationships between raw reflectance values, spectral vegetation indices, and spatial features to estimate LAI in snap beans towards a spatially-explicit mold risk model.

The methods in this chapter extends earlier work by Hughes et al. [2] and uses a combination of SFLR, linear and multivariate regressions, and SVM to create discriminating and predictive probability maps for snap bean flowering at the canopy level as well as spectral estimates for LAI for the eventual inclusion in a white mold risk model.

2.1 Data Collection

A DJI Matrice-600 UAS, mounted with a Headwall Photonics Nano imaging spectrometer (272 bands; 400-1000 nm; FWHM slit image of 6 nm), was used to collect the imaging spectroscopy data of the snap beans. High frequency flights were flown at the New York State Agricultural Experiment

Station, operated and monitored by Cornell University, during seasonal periods when portions of the snap bean fields were flowering. Snap bean plants were planted at three different periods during the growing season, starting the first week of June (early) and with mid and late plantings starting two and four weeks after, respectively. This allowed for different flowering stages to be captured during a single UAS flight. 22 flights on 10 different days at $42°52'26.4"N\ 77°01'37.2"W$, and 22 flights on eight different days at $42°52'26.4"N\ 77°01'48.0"W$ were flown throughout the 2017 and 2018 season, respectively. The flights were flown with ground sampling distances (GSD; spatial resolution) that ranged between 1.25-4.5 cm. Ground truth flowering data from Cornell was collected on the same day as the UAS flights. The time difference between ground truth assessment and UAS flights was less than a day.

we used an AccuPAR PAR/LAI Ceptometer Model Lp-80 to collect in-field estimates of LAI. These ground truth LAI measurements are based on how much of the incident sunlight is transmitted through the canopy, with known or estimated parameters, such as the fraction of photosynthetically active radiation (PAR), leaf absorbtivity, and zenith sun angle used to estimate final LAI values [33][34][35]. ** Ethan, describe the AccuPar protocol in more technical detail ** Ground truth LAI and imaging spectroscopy data were collected on August 2, 6, and 23, 2018 at the New York State Agricultural Experiment Station, operated by Cornell University ($42°52'26.4"N\ 77°01'48.0"W$). Each

flight captured the spectra of all three plantings (early, mid, and late) with GSDs ranging from 2-3 cm. Each planting had ten plots, thus resulting in 90 plots of ground truth LAI for use in our spectra to LAI regressions. The ground truth LAI collection protocol included taking 8 LAI measurements with the Lp-80 and averaging those LAI measurements to obtain an average LAI for the plot. Plot average spectra and ground truth LAI were used to create regression to go from some feature space (i.e. raw reflectance, normalized difference vegetation index, enhanced vegetation index, etc.) to LAI space.

Figure 2.1 and 2.2 below show the analysis workflow, which will be described in detail in the following sections.

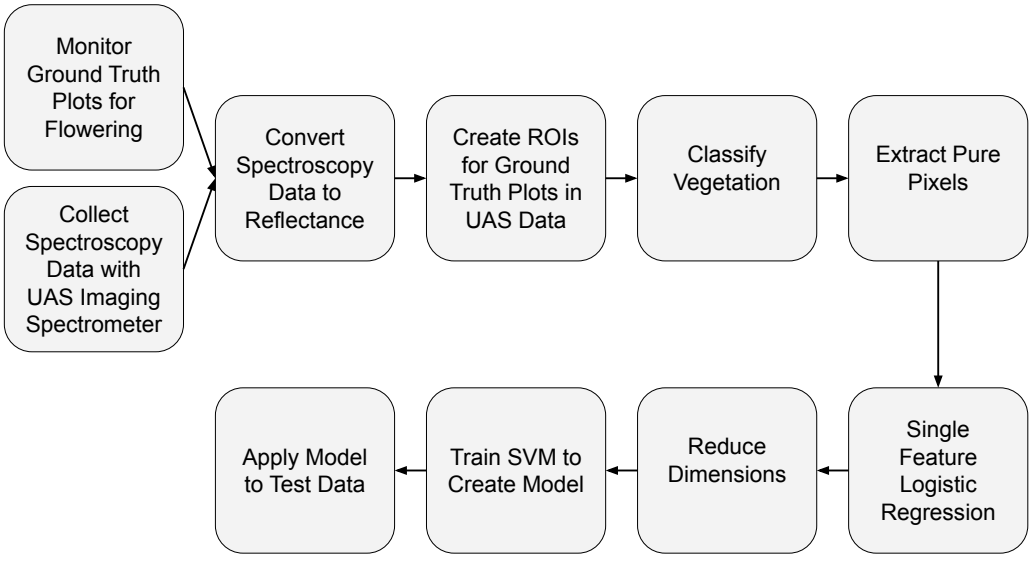


Figure 2.1: Data Collection/Processing Flow Chart for Flowering

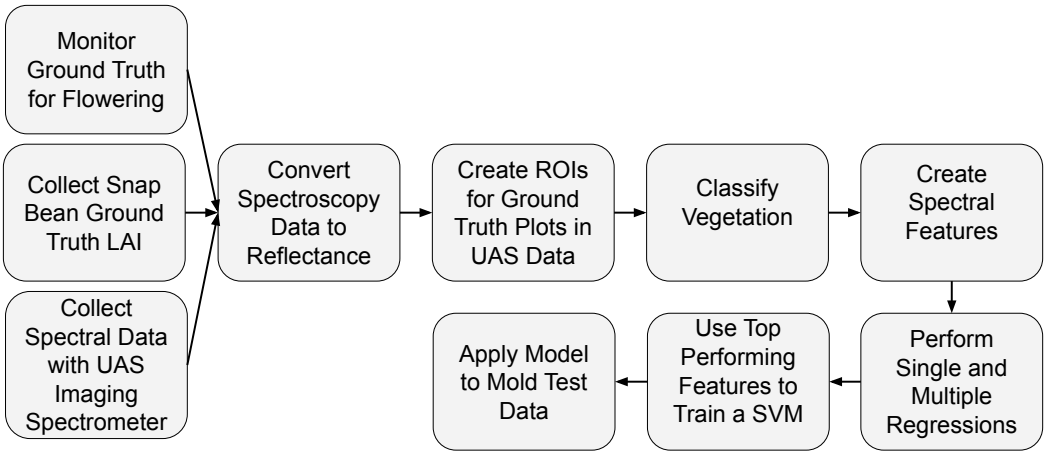


Figure 2.2: LAI Data Collection/Processing Flow Chart for Spectral-based LAI Regression

2.2 Conversion to Reflectance

Light and dark calibration panels were placed in the field during the flights to convert the raw spectroscopy imagery from digital counts to reflectance. This was done by using a SVC hand-held spectroradiometer (334-2508 nm; 2 nm resampled bandwidths) to collect ground truth reflectance spectra of the light and dark calibration panels and the two-point empirical line method of calibration (ELM) [36]. Converting the imagery into reflectance is paramount, because it removes the dependence on the illumination conditions (sun angle, clouds, etc.), which makes it possible to compare spectral data from different days, irrespective of such varying illumination conditions [37].

2.3 Classifying Vegetation

2.3.1 Vegetation vs. Background Classification

Another important step in the data processing pipeline is to separate the vegetative spectra from that of the background (soil). One way to do this is to use the spectral angle mapper (SAM) approach [38]. SAM is a supervised classification technique that computes the cosine between a sample spectrum and that of a reference. Classification is then made through the application of an angular threshold, i.e., the user-specified threshold is used to classify pixels in terms of their multi-dimensional spectral angle, relative to the reference pixel. SAM is relatively insensitive to changes in illumination, as well as phenological variability, because the algorithm compares the direction of the vectors [38][39]. Reference reflectance spectra of snap beans were extracted from the imagery and used to separate vegetative pixels from soil using SAM.

Equation 2.1 below shows the formulation for SAM, where S is the sample spectrum and R is the reference spectrum. $1 - cos\theta$ was used in this work, therefore a smaller threshold is more strict (i.e., the sample and reference spectra are more similar). For this work a threshold of $\sim$0.01 was selected for flowering and $\sim$0.025 for leaf area index regressions using Otsu's method. Otsu's method chooses a threshold (the angle) that maximizes inter-class variance, in this case the variance between vegetation and the ground.[40].

$$\cos\theta = \frac{S \cdot R}{\|S\| \cdot \|R\|} \tag{2.1}$$

2.3.2 Extracting Pure Vegetation Pixels for Flowering

Once the locations of vegetation-only spectra have been identified, we need to locate pixels in the snap bean canopy that contain pure, or high-signal, vegetative spectra. The goal is to find vegetative pixels that are not self-shadowed, mixed with the background ("mixed pixels"), or contain leaves that are not perpendicular to the sensor. Healthy vegetation absorbs red energy for photosynthesis, while being highly reflective in the near-infrared, due to inter-cellular leaf structure [41]. Therefore, we used a traditional ratio approach between a near-infrared (NIR) and a red band to filter out the "spectrally impure" pixels in the canopy and locate the healthy, sunlit, and spectrally pure vegetation pixels. This method is called Ratio and Thresholding (RT), where RT computes the average and standard deviation of the ratios of the vegetative pixels and then returns the pixels that contain ratios greater than the average plus the standard deviation, as modulated by a user-defined constant. For this work a multiplier of zero was selected (which means the threshold is equal to the average of all the ratios calculated for the vegetative pixels). This multiplier was used because it classified pixels at the center of the canopy, while preserving the most samples. Pseudocode for RT can be found in Figure 2.3.

Algorithm 1: Ratio and Thresholding

```
image = array with shape samples x features
numerator and denominator are input wavelengths
for i samples do
   ratio[i] = image[i, numerator] / image[i, denominator]
end for
avg = mean(ratio)
std = std(ratio)
threshold = avg + multiplier · std
indices vegetation = where(ratio > threshold)
indices background = where(ratio < threshold)
vegetation = image[indices vegetation, :]
background = image[indices background, :]
return vegetation, background
```

Figure 2.3: This figure contains psuedocode for Ratio and Thresholding. This method was used for finding pure pixels for flowering.

2.4 Identifying Spectral Features for Flowering

The next step in the analysis is to identify key spectral features that will best separate the binary classes (flowering vs. not-flowering, will flower vs. not-flowering). Single spectral features, e.g., individual wavelengths, were fit using a logistic regression [1][2] to accomplish this. In other words, single spectral features could be the raw reflectance values at a single wavelength, ratio indices (RIs), or normalized difference indices (NDIs). The RIs and NDIs used in this work have been found to exhibit correlations with chlorophyll content and water stress [1], and are listed in Table 2.1. RI simply refers to the ratioing of reflectance values at two different wavelengths. NDI is similar to RI, except that the reflectance values at the two different wavelengths

are subtracted in the numerator and normalized by their sum in the denominator; this approach has been shown to be more robust against illumination changes, while also standardizing the index value between 0 and 1. These metrics arguably are simplistic in nature, but have been shown to be effective in identifying spectral-biophysical correlations, since they typically highlight inverse reflectance behavior for two wavelength regions, e.g., the normalized difference vegetation index (NDVI) and the photochemical reflectance index (PRI) [42][43][44][45].

$$RI = \frac{R_{\lambda_n}}{R_{\lambda_m}} \tag{2.2}$$

$$NDI = \frac{R_{\lambda_n} - R_{\lambda_m}}{R_{\lambda_n} + R_{\lambda_m}} \tag{2.3}$$

where n and m denote the different wavelengths.

The discriminating performance of each feature was found by calculating the c-index, the false positive (classification) rate (FPR), and the false (classification) negative rate (FNR). The c-index is equal to the area under the receiver operator curve (ROC), where a c-index score of 0.5 implies random discriminating performance, and a c-index of 1 is considered perfect class separation. Specifically, an acceptable model will have a c-index of 0.8 or more [46]. The FPR is the number of false positives, divided by the sum of the number of false positives and true negatives. The FNR is the number of

false negatives, divided by the sum of the number of false negatives and true positives. These metrics are indicative of errors of commission and omission, respectively [47].

$$FPR = \frac{FP}{FP + TN} \tag{2.4}$$

$$FNR = \frac{FN}{FN + TP} \tag{2.5}$$

Table 2.1: The wavelengths below were used to create ratio indices (RI) and normalized difference indices (NDI), which subsequently were used as features in the single feature logistic regression (SFLR) analysis. This table is adapted from Delalieux et al. (2009)[1], and was used previously by Hughes et al. (2018) [2].

Ratio Indices	
Numerator R_{λ_n} (nm)	Denominator R_{λ_m} (nm)
430	680
440	690
550	800
605	760
672	550
675	700
695	420
695	670
695	760
710	760
740	720
750	550
750	705
750	710
800	550
800	635
800	680

(a)

Normalized Difference Indices	
Numerator R_{λ_n} (nm)	Denominator R_{λ_m} (nm)
415	435
680	430
750	660
750	705
750	445
800	635
800	680

(b)

2.5 Spectral Vegetation Features for LAI

The next steps involved the creation of features that we can feed to our regressions and subsequent models in addition to single reflectance spectra, following the identification of pure vegetation pixels. A range of established vegetation indices were used to augment raw reflectance, single-wavelength analyses. Canonical normalized difference vegetation index (NDVI) was selected as one such an index candidate, since it is highly correlated with green biomass and vegetation moisture content [30].

$$NDVI = \frac{R_{NIR} - R_{Red}}{R_{NIR} + R_{Red}} \tag{2.6}$$

Where R_{NIR} and R_{Red} are the down-sampled reflectance spectra from 800-1000 nm and 600-700 nm, respectively. Green normalized difference vegetation index (GNDVI) is similar to NDVI, except that it has proven to be more sensitive and highly correlated to chlorophyll content [31].

$$GNDVI = \frac{R_{NIR} - R_{Green}}{R_{NIR} + R_{Green}} \tag{2.7}$$

Where R_{NIR} and R_{Green} are the reflectance spectra at 750 nm and 550 nm, respectively. The enhanced vegetation index (EVI) was developed to optimize the vegetation signal in cases with high biomass, which is an area where NDVI has been shown to saturate (reduced sensitivity at high biomass and LAI levels) [32].

$$EVI = 2.5 \cdot \frac{R_{NIR} - R_{Red}}{R_{NIR} + 6 \cdot R_{Red} - 7.5 \cdot R_{Blue} + 1} \tag{2.8}$$

Where R_{NIR}, R_{Red}, and R_{Blue} are down-sampled reflectance spectra from 841-876 nm, 620-670 nm, and 459-479 nm, respectively. Finally, we decided on a feature related to 2D green leaf area, which we calculated by finding the density of classified vegetative pixels per region of interest. The assumption was that regions with more vegetation-per-area could have a micro-climate that is more conducive to mold occurrence [29].

$$Pixel\ Density = \frac{Classified\,Veg.\,Pixels}{Total\ Number\ of\ Pixels} \tag{2.9}$$

2.6 Flowering Model Development

It is important in machine learning to ensure that the size of the training and testing sets are close to equal, thereby avoiding the challenges associated with over- or underfitting [48]. For this research, the training and testing data sets were created from a random sample of the full set and were split 50-50. This resulted in testing and training class sizes on the order of 4,700 samples for flowering. Features with the largest c-index subsequently were used to train a support vector machine (SVM) with a radial basis function (rbf) [49]; a SVM finds a (multi-dimensional) hyper-plane that maximizes the distance between classes [50][51][52][53].

The discriminating flowering model was created using data from July 19th, 2018 (3 cm GSD), where the snap beans from the early-planting were all visibly flowering, and the snap beans from the late-planting were not flowering, and would not flower until August 15th, 2018. The model then was applied to data from the early- and late-planting from July 19th, 2017 (3 cm GSD), where the early-planting was flowering at 100 % and the late-planting was flowering at a 0% level. The predictive flowering model, on the other hand, was created using data from July 19th, 2018, where the snap beans from the mid-planting were not visibly flowering and would not do so until August 2nd, 2018 (i.e., two weeks later), and snap beans from the late-planting wouldn't be visibly flowering until August 15th (i.e., four weeks later). This latter model then was applied to mid- and late-planting data from July 19th, 2017, where the mid-planting was at 28% flowering (not 100% flowering until two weeks later) and the late-planting was at a 0% flowering level (not 100% flowering until four weeks later). The models were applied to data from the previous year to evaluate feature and model robustness, or prediction stability over time, across the two growing years.

2.7 Leaf Area Index Regression

We implemented a traditional least-squares regression [54] between each single reflectance feature and the ground truth LAI, in order to identify how well each feature performed at predicting LAI. The coefficient of determination

(CoD) and root mean squared error (RMSE) were used to determine how well the least-squares regression fit the data. It should be noted that because LAI is a dimensionless quantity the RMSE is also dimensionless.

We subsequently used the reflectance spectra with the top 24 highest CoD, along with pixel density, NDVI, GNDVI, and EVI to fit a multivariate regression model. 28 features were selected because adjusted R^2 values decreased rapidly after 30 features. A multivariate regression model with the least absolute shrinkage and selection operator (LASSO) was used because it encourages sparse models (fewer output parameters) and is widely used in remote sensing literature related to forestry and vegetation [55][56][57]. LASSO is a linear model that is equivalent to fitting N number of independent regressions with each feature having its own L_1 penalty [58], where L_1 regularization adds a penalty term equal to the absolute value of the magnitude of the coefficients. This approach limits the size of the coefficients [59]. Alpha is a parameter that you define in the application of LASSO that is multiplied by the L_1 penalty; for example, an alpha of 0 would result in LASSO being an ordinary least-squares fitting approach.

As is common with multiple regression, the CoD, or R^2, must be adjusted to measure the portion of the variance in the dependent variable explained by the independent variables where over-fitting is penalized. This implies that the adjusted CoD typically will be smaller than the standard CoD.

$$Adjusted\,CoD = 1 - (1 - R^2)\frac{n-1}{n-p-1} \tag{2.10}$$

Where R^2 is the CoD, n is the sample size, and p is the number of predictor variables.

2.8 Results and Discussion for Flowering

The c-index, FPR, and FNR from the SFLR for wavelengths (raw reflectance), RI, and NDI features for flowering discrimination can be found in Figures 2.5-2.7. Figure 2.5a shows c-indexes well above 0.8 in the red portion of the spectrum (600-700 nm) with a peak around 680 nm. This region is especially important in biophysical assessment of vegetation, since it is heavily impacted by plant vigor (photosynthesis) and inter-cellular structure of leaves; specifically, one often can observe a shift to longer wavelengths of the red egde position for healthy vegetation, and an opposite shift for stressed/unhealthy plants [60][61]. Some RI and NDI features yielded c-index values close to 1.0, but these models, when applied to data from the previous year, were sensitive to noise. We attributed this behavior to the compounding of variance from each feature in the RI or NDI (uncertainties add in quadrature)[62], and interaction with the model's decision or classification hyper-plane.

The c-index values for the flowering predictive SFLR (Figures 2.8-2.10) look very similar to that of the discrimination model (peak class separability from 600-700 nm with a peak at 680 nm), aside from small changes in the

NIR and green portions of the spectrum. This implies that the snap beans have already undergone the phenological change to initiate flowering, even though the flowers are not yet visible; it should be noted that the flowers only visibly emerged two weeks later. Figure 2.4 shows that the average early and mid plant canopy level spectra are almost the same especially at reflectance values before 760 nm, and that the snap beans become more absorptive in the red ($\sim$ 680 nm) and reflective in the NIR (730-1000 nm) when the snap beans prepare to flower. The phenological change in the snap bean effectively causes an increase in the red edge slope (680-740 nm). This implies that one is able to predict the onset of flowering two weeks in advance, using the selected spectral features. We believe the small increase in the green and small decrease in the NIR spectral regions in Figure 2.8a, approximately 2% and 5%, respectively, are due to variance induced during the conversion to reflectance and also inherent canopy spectral variability, even for a set of arguably pure canopy-level pixels [36][63]. These results bode well for using a UAS-based, spectral analysis approach to determine within-field crop locations for directed fungicide applications, targeted on flowering locations.

The wavelength feature with the largest wavelength c-index (680 nm) was used to train a SVM with a rbf kernel and then applied to data from July 19th, 2017, to create spatially-explicit flowering probability maps or "heat maps". A probability map was used, instead of a traditional classification map, to highlight the fact that the model works best at the center of the

canopy, where mixed and self-shadowed pixels are less abundant. Again, only a single feature was used for each model. This is due to the practicality (cost-effective, operational simplicity) and availability of multispectral UAS platforms, i.e., sensors with a much reduced set of wavelength bands when compared to imaging spectroradiometers. Imaging spectroradiometer systems spectrally oversample the area of interest in a general sense, which implies a large set of independent wavelength variables that are collected at relatively high cost; both these characteristics effectively are addressed in multispectral, limited-band, application-focused systems. To that end using the SFLR c-index results could be a useful approach via which to select the width of bandpass filters when engineering multispectral, applied solutions. A future study could perform a spectral convolution (downsampling of the spectroscopy data) with band center at the peak (680 nm) with a bandpass width of $\sim$40 nm to investigate the applicability of a multi-spectral solution.

Figure 2.11 shows the performance of the discrimination model on test images from the 2018 growing season, where 100% (Figure 2.11a) and 0% (Figure 2.11b) of the snap beans are flowering, respectively. The model shows that the probability of flowering is close to 1.0 at the center of the canopy in validated flowering instances and zero for flowering and not flowering test images.

Figure 2.12 shows the performance of the predictive model on on test images, where the snap beans will be flowering in two weeks (Figure 2.12a)

and flowering in 4 weeks (Figure 2.12b), respectively; note that neither of the images exhibited visible flowering. The model is in effect querying whether or not the snap beans will flower in two weeks. Here the results are similar to the discrimination model, where probability at the center of the canopy is close to 1.0 and zero for flowering in two weeks, for Figures 2.12a and 2.12b), respectively. It should be noted that the test images in Figures 2.11-2.12 were not used in the training or testing of the models.

The models related to actual and early flowering detection are promising and important. Snap bean growers in effect can use an operational multispectral sensor, with defined spectral (wavelength) features, to accurately identify within-field locations where pro-active fungicide applications should be focused. Such an approach arguably will contribute to optimized disease management - management interventions can be optimized in terms of application levels, locations, and reduction in any potential groundwater or stream impacts.

While we believe we are observing a phenological change, it is possible for chlorophyll and inter-cellular structure to change due to environmental or physciological impacts, other than those associated with flowering. Environmental stressors can impact red-edge behavior such as red edge slope and position ** Ethan - citation? ** . More data therefore should be collected from different field locations to ensure the robustness of the model. Other future work towards model robustness should include the implementation of indepen-

dent calibration panels for verification of reflectance data fidelity. Uncertainty in the conversion from raw digital counts to reflectance will negatively affect the interaction between classes and the hyper-plane of our model.

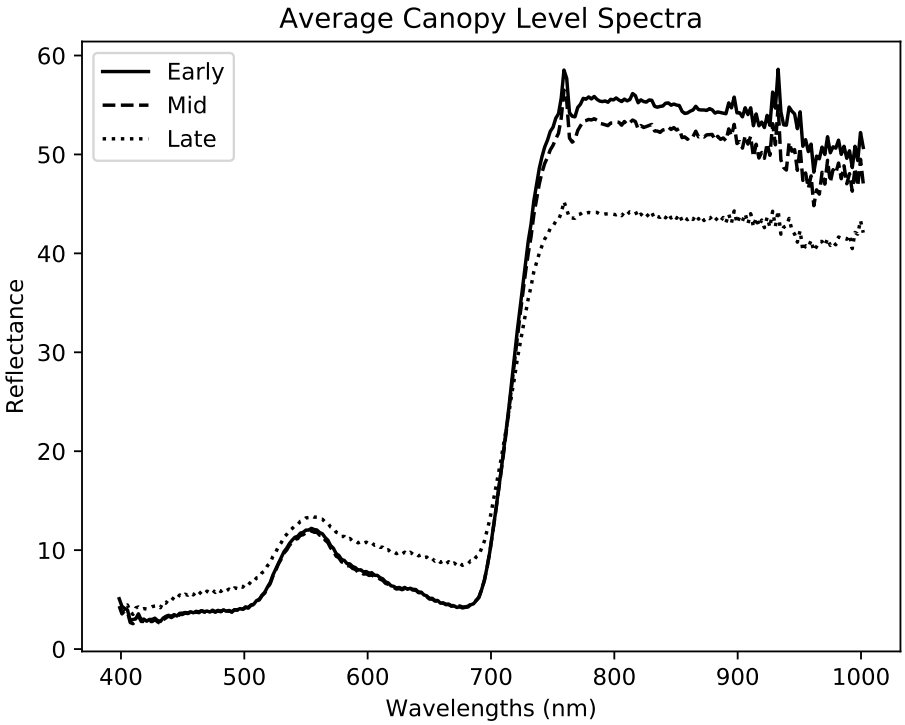
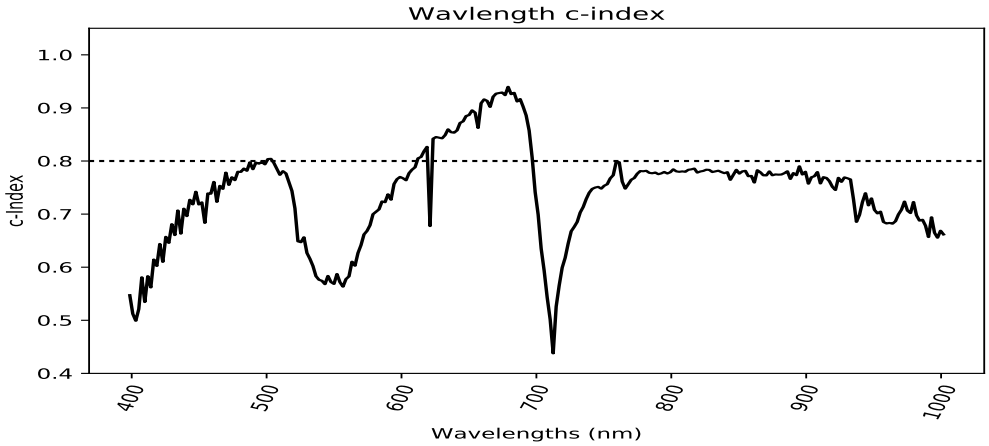
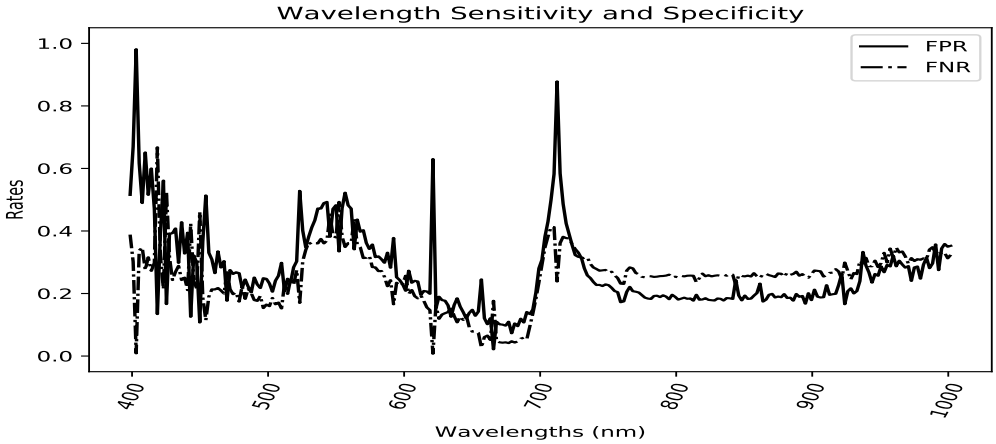


Figure 2.4: Canopy Level Spectra for Early, Mid, and Late Plantings

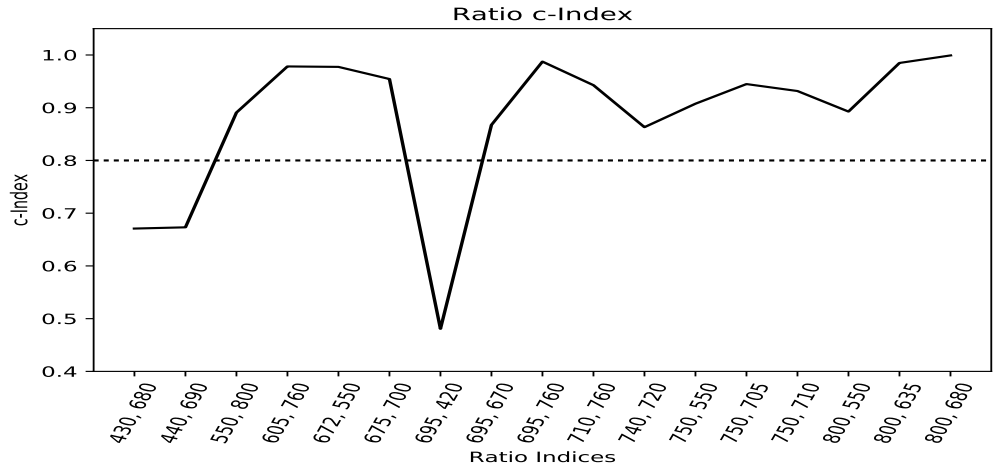


(a) c-index from wavelength single feature logistic regression for flowering discrimination; note the discrimination levels in the 600-700 nm spectral region, i.e., the red absorption trough and adjacent reflectance rise in the red edge region, toward 700 nm

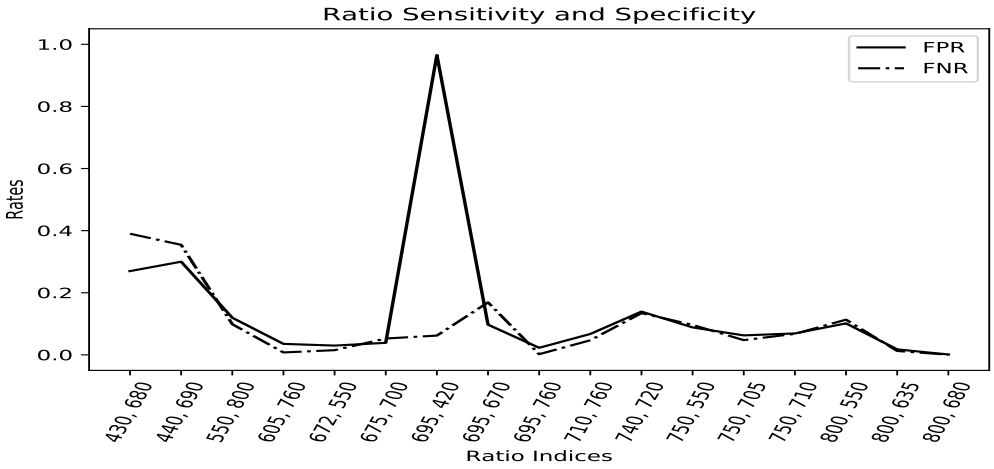


(b) False positive and false negative rates for wavelength single feature logistic regression for flowering discrimination; again note the associated lower false prediction rates in the 600-700 nm spectral region (see Figure 2.5a above)

Figure 2.5: This figure shows the c-index, false positive, and false negative rates for each of the wavelengths used to separate the classes flowering vs. not-flowering. Features in the NIR red edge have the largest c-index and lowest false positive and false negative rates. These regions are known to be impacted by photosynthesis and the inter-cellular structure of leaves.

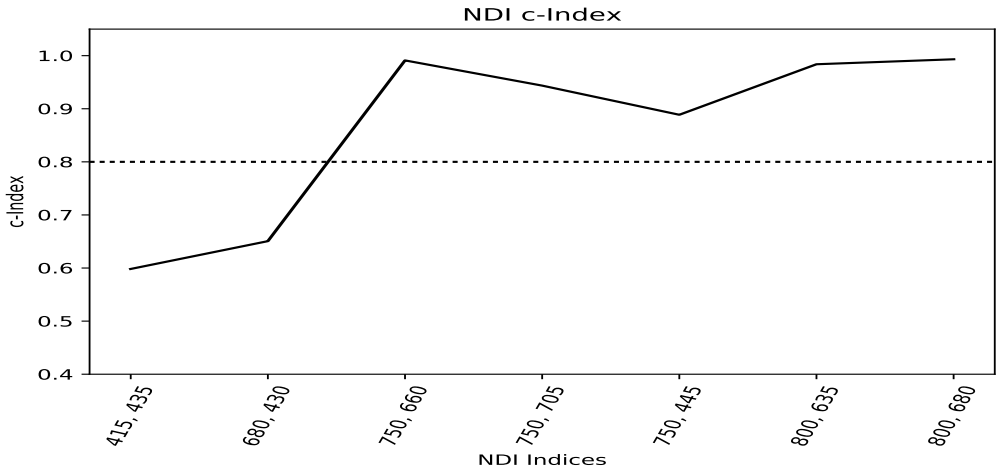


(a) c-index for Ratio Index Single Feature Logistic Regression for Flowering Discrimination; the combination of green and red (visible) and NIR wavelengths seem especially effective in discriminating flowering from non-flowering plants

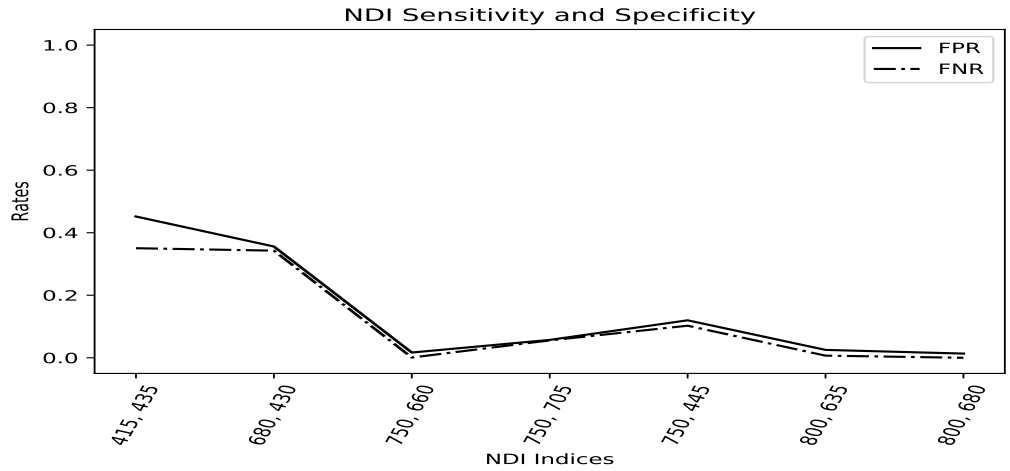


(b) False Positive and False Negative Rates for Ratio Index Single Feature Logistic Regression for Flowering Discrimination

Figure 2.6: This figure shows the c-index, false positive, and false negative rates for each of the ratio indices (RIs) used in the single feature logistic regression to separate the classes flowering vs. not-flowering. Ratio indices with wavelengths in the green, red, red edge, and NIR have the largest c-index and lowest false positive and false negative rates. This is due to photosynthesis absorption (red spectral region) and the inter-cellular structure (NIR spectral region) of the leaves.

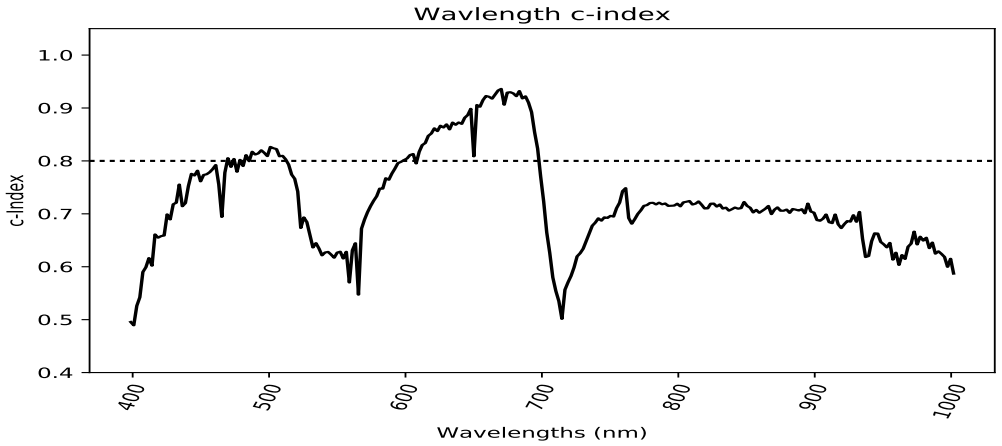


(a) c-index for Normalized Difference Index Single Feature Logistic Regression for Flowering Discrimination

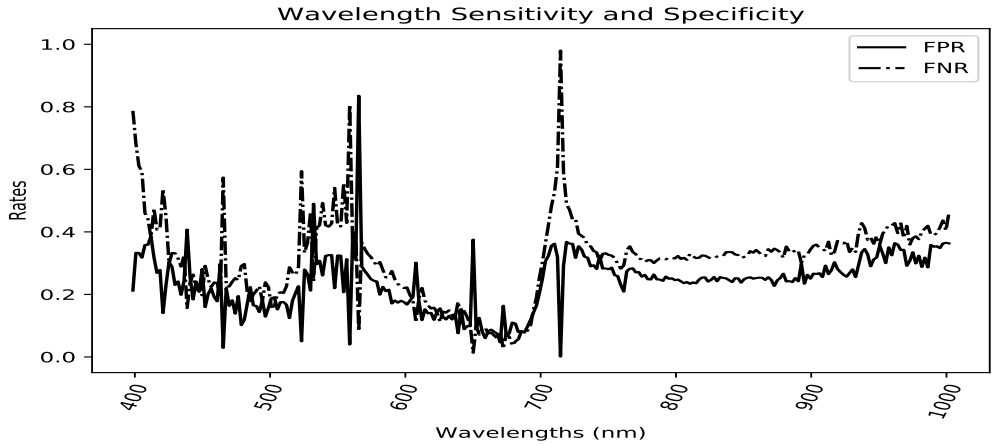


(b) False Positive and False Negative Rates for Normalized Difference Index Single Feature Logistic Regression for Flowering Discrimination

Figure 2.7: This figure shows the c-index, false positive, and false negative rates for each of the normalized difference indices (NDIs) used in the single feature logistic regression to separate the classes flowering/not flowering. Indices with wavelengths in the red to red edge, and the broader NIR had the largest c-index and lowest false positive and false negative rates.

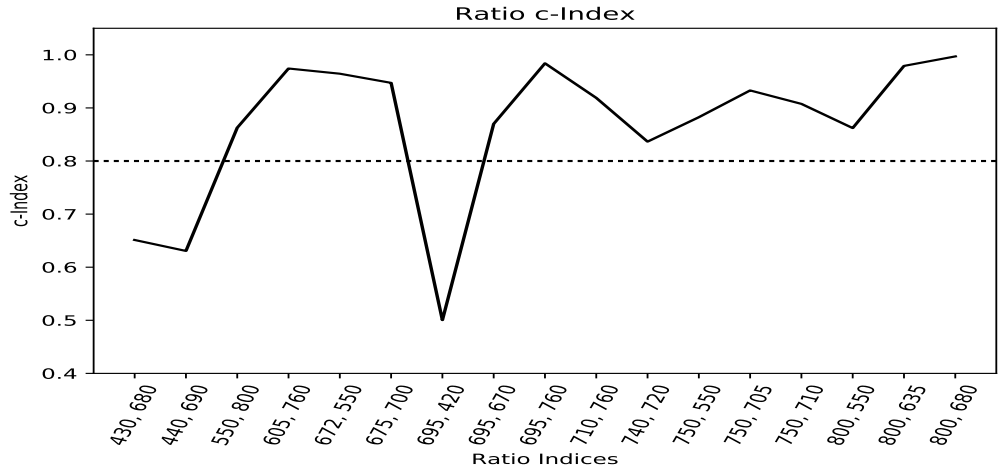


(a) c-index for Wavelength Single Feature Logistic Regression for Flowering Prediction

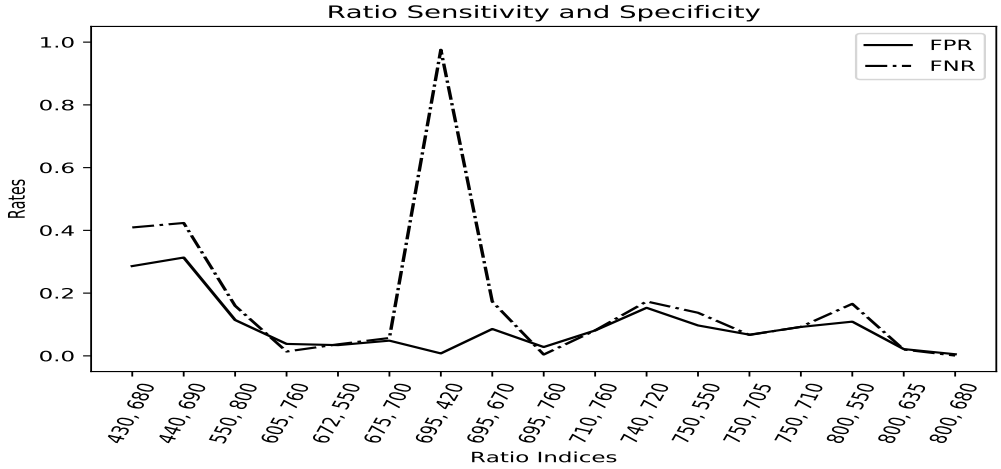


(b) False Positive and False Negative Rates for Wavelength Single Feature Logistic Regression for Flowering Prediction; the red, red edge, and NIR spectral regions again exhibit the best performance for flowering discrimination

Figure 2.8: This figure shows the c-index, false positive, and false negative rates for each of the wavelengths used to separate the classes flowering in two weeks/flowering in four weeks. Features in the NIR red edge region have the largest c-index and lowest false positive and false negative rates. The values are very similar to the discrimination model shown in Figure 2.5. We believe the snap beans have already undergone the phenological change and notable differences in the NIR and green c-indexes are due to the variance in conversion to reflectance and the canopy.

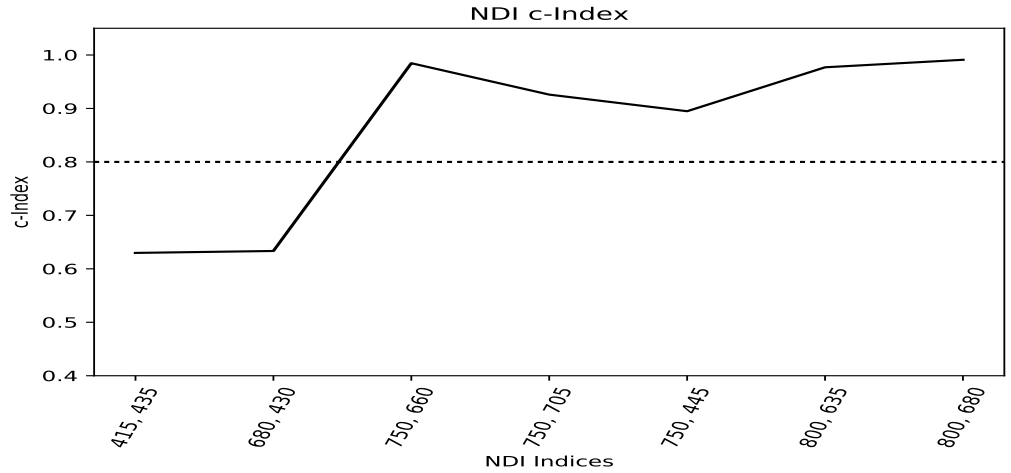


(a) c-index for Ratio Index Single Feature Logistic Regression for Flowering **Prediction**; note that similar spectral regions were identified as was the case for flowering discrimination, with the inclusion of select blue-green spectral regions

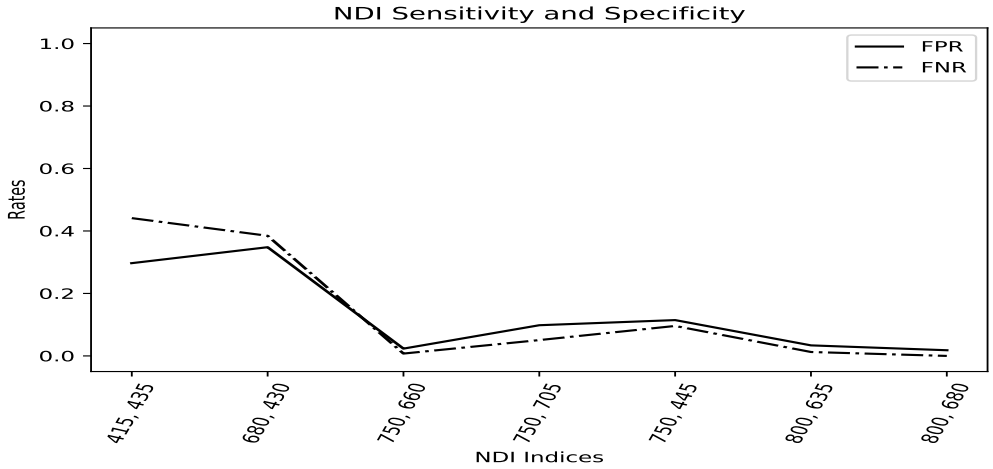


(b) False Positive and False Negative Rates for Ratio Index Single Feature Logistic Regression for Flowering Prediction

Figure 2.9: This figure shows the c-index, false positive, and false negative rates for each of the ratio indices (RIs) used to separate the classes flowering in two weeks/flowering in four weeks. Features that contain wavelengths in the NIR red edge region and the broader NIR have the largest c-index and lowest false positive and false negative rates. These results look similar to that of the discrimination model in Figure 2.6. Differences include a small increase in features that contain green wavelengths.

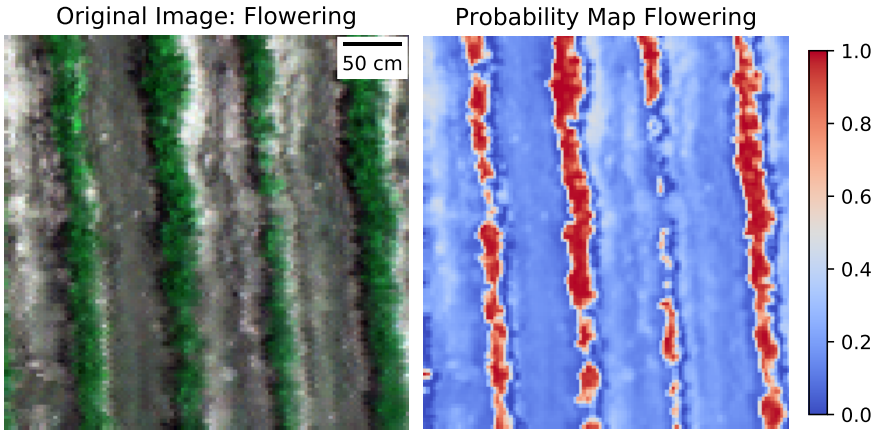


(a) c-index for Normalized Difference Index Single Feature Logistic Regression for Flowering Prediction

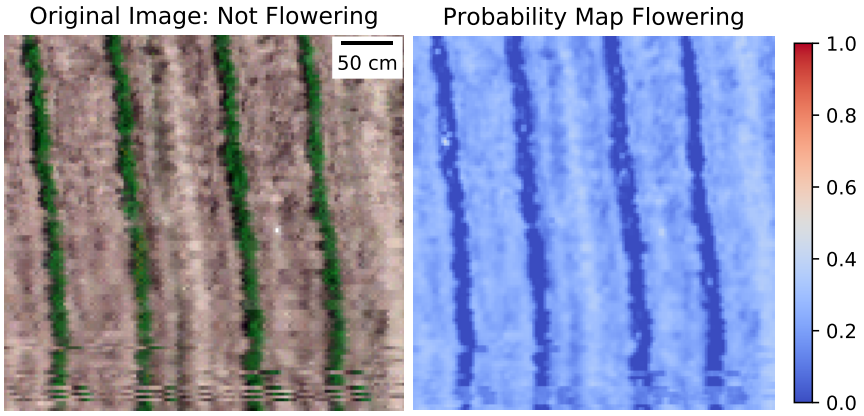


(b) False Positive and False Negative Rates for Normalized Difference Index Single Feature Logistic Regression for Flowering Prediction

Figure 2.10: This figure shows the c-index, false positive, and false negative rates for each of the normalized difference indices (NDIs) used to separate the classes flowering in two weeks/flowering in four weeks. Features that contain wavelengths in the NIR red edge and broader NIR regions have the largest c-index and lowest false positive and false negative rates. These results look similar to that of the discrimination model in Figure 2.7. Differences include a small increase in features that contain blue wavelengths.

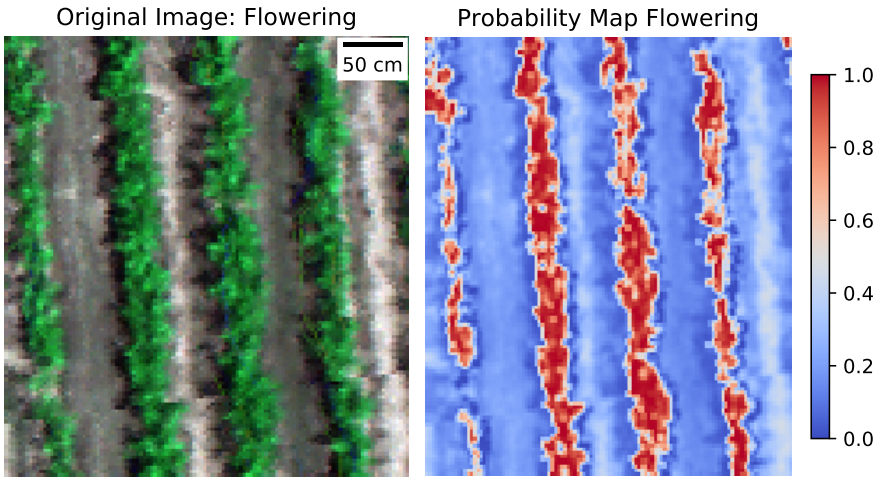


(a) Discriminating flowering model on a flowering test image

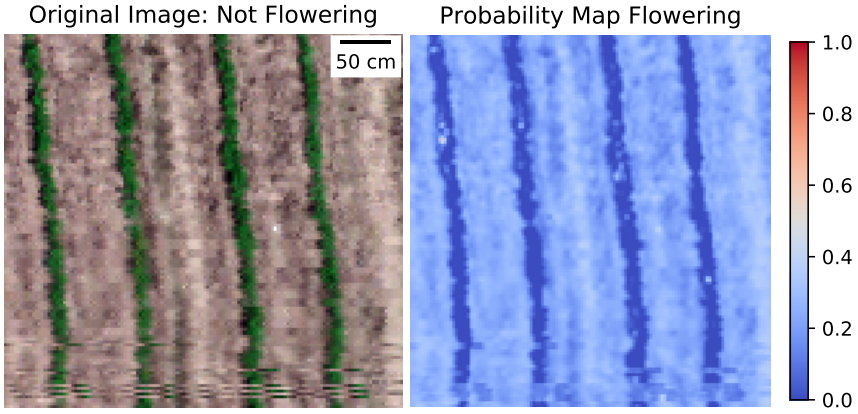


(b) Discriminating model on a not-flowering test image

Figure 2.11: This figure shows the results of the flowering discrimination model using a single NIR red edge feature on flowering and not-flowering test images. The model yields probabilities of close to 1.0 and zero at the center of the canopies for flowering and not-flowering images, respectively.



(a) Predictive flowering model on an image that will flower in two weeks



(b) Predictive flowering model on an image that will flower in four weeks

Figure 2.12: This figure shows the results of the flowering predictive model using a single NIR red edge feature on images that will flower in two a) and four b) weeks. The model yields probabilities of close to 1.0 and zero at the center of the canopies for flowering at two and flowering at four weeks, respectively.

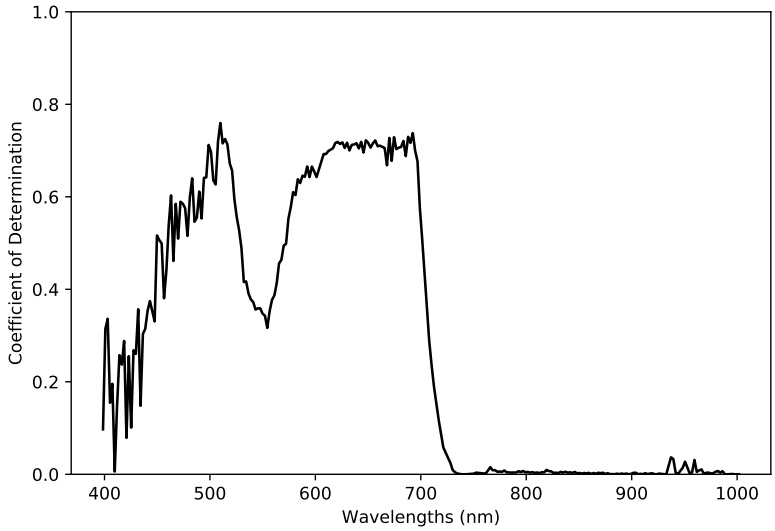
2.9 Results and Discussion for LAI Spectral Regression

Figure 2.13a shows that the CoDs from the single reflectance feature least-squares regression exhibited values larger than 0.7 in the red-red edge (∼600-690 nm) and green (∼507-518 nm) spectral regions, with a maximum CoD at 510 nm, equal to 0.76. These wavelengths are important indicators of plant physiology in terms of chlorophyll absorption and inter-cellular structure absorption/reflectance features (biomass, vigor, etc.) [30][31][60][61]. Figure 2.13b, in turn, shows the data and fit of the regression between reflectance spectra at 509.872 nm and ground truth LAI. The trend shows that as the reflectance at 509.872 nm increases, the LAI decreases. This feature has a RMSE of 0.61. Figures 2.14a-2.15b show the data and fit between pixel density, EVI, NDVI, and GNDVI. CoDs and RMSE values ranged from 0.42-0.57 and 0.817-0.942, respectively. These features, along with the top 24 single reflectance features, were used in the subsequent LASSO regression.

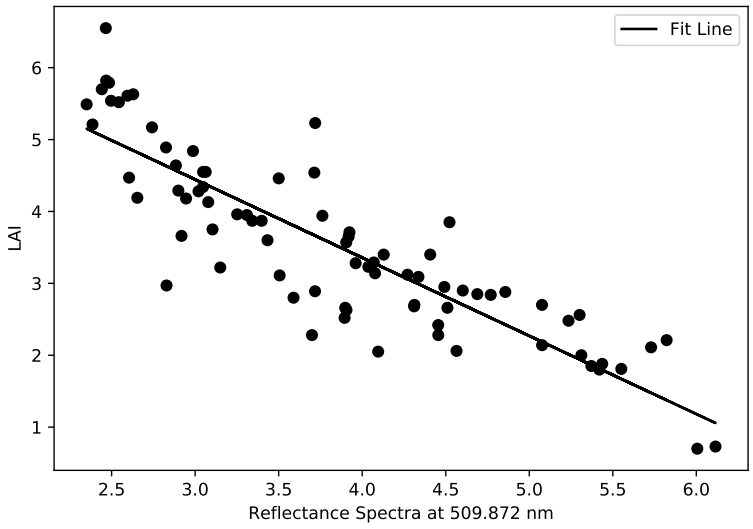
Figure 2.16 shows the results for the CoD and adjusted CoD as a function of alpha level (0.0001-2, by 0.0005 increments). Figure 2.17a and Figure 2.17b show how the number of explanatory features changes with alpha, and how the number of features impacts the RMSE. We observed that a small alpha yields the largest adjusted CoD (0.85), using all 28 features, and the smallest RMSE (0.39). As alpha increases, the adjusted CoD decreases, the number of features decreases, and the RMSE increases. Figure 2.16 shows a rapid drop in performance before 0.01 alpha, while the number of features is

reduced from 28 to 5; from there the decrease in performance is more gradual.

Future work should include the use these LAI correlated features along with other auxiliary environmental variables (degree days, average rain fall, temperature, etc.) in the creation of a white mold risk model.

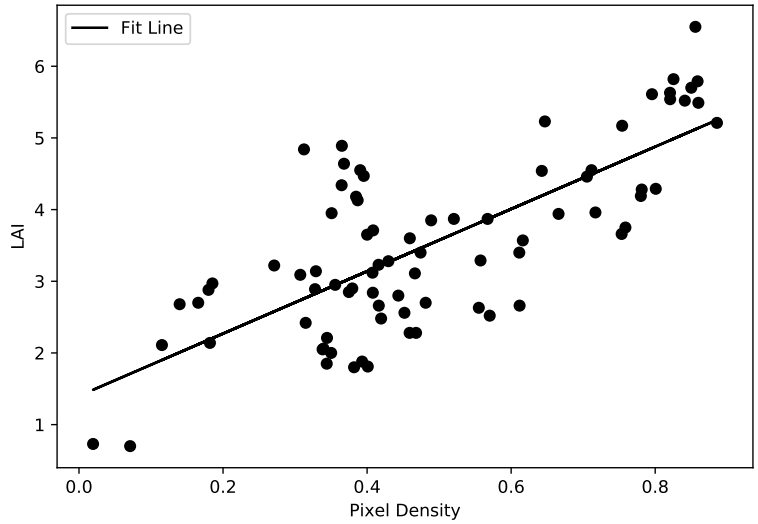


(a) CoD from the linear regression of single reflectance spectra and ground truth LAI

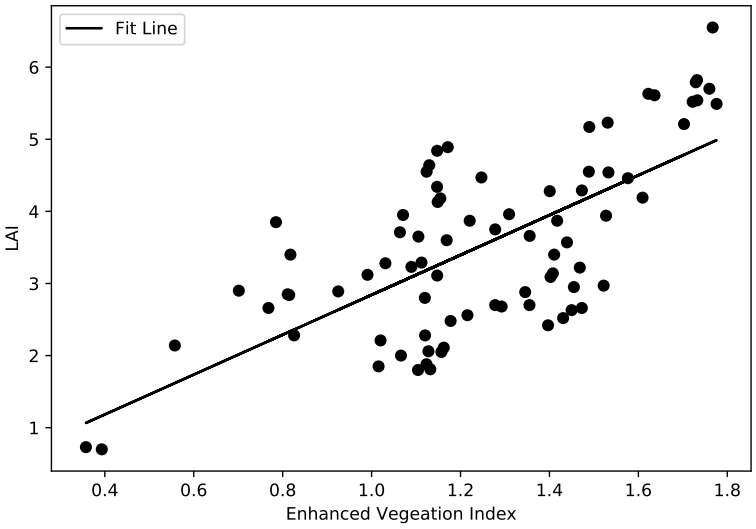


(b) Regression for spectra at 509.872 nm and ground truth LAI

Figure 2.13: Figure 2.13a shows the resulting CoD from fitting single reflectance spectra to ground truth LAI with 509.872 nm having a CoD and RMSE of 0.76 and 0.609, respectively. Figure 2.13b shows the model fit and residual distribution around the fitted regression.

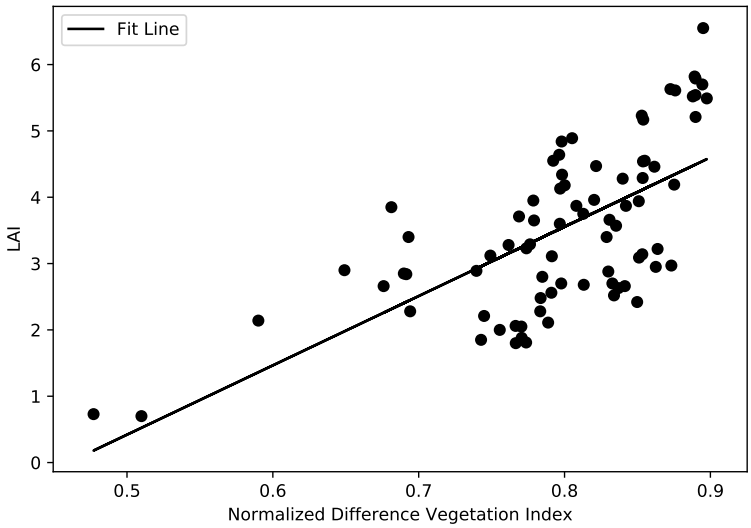


(a) Regression for pixel density and ground truth LAI

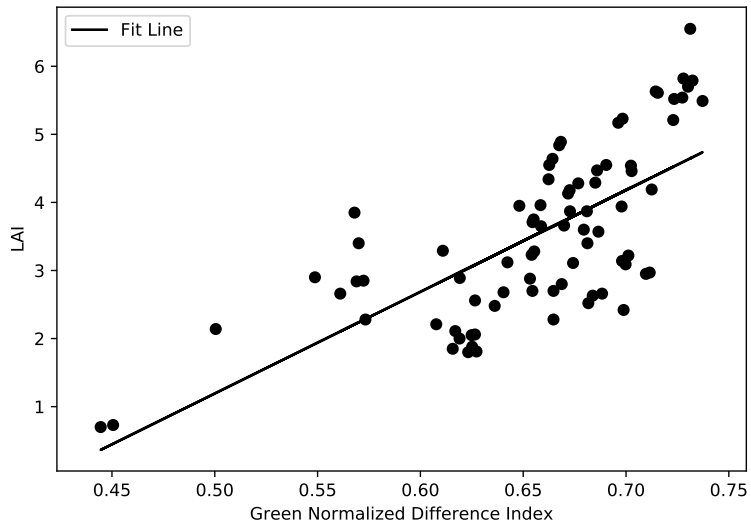


(b) Regression for EVI and ground truh LAI

Figure 2.14: Figure 2.14a shows the data and fit when using pixel density as a predictor of LAI, with CoD and RMSE values of 0.57 and 0.817, respectively. Figure 2.14b, on the other hand, shows the data and fit for EVI as a predictor of LAI, with CoD and RMSE values of 0.48 and 0.898, respectively.



(a) Regression for NDVI and ground truth LAI



(b) Regression for GNDI and ground truth LAI

Figure 2.15: Figure 2.15a shows the data and fit of NDVI as a predictor of LAI, with a CoD and RMSE of 0.42 and 0.942, respectively. Figure 2.15b shows the data and fit of GNDVI as a predictor feature for LAI. with a CoD and RMSE of 0.50 and 0.879, respectively.

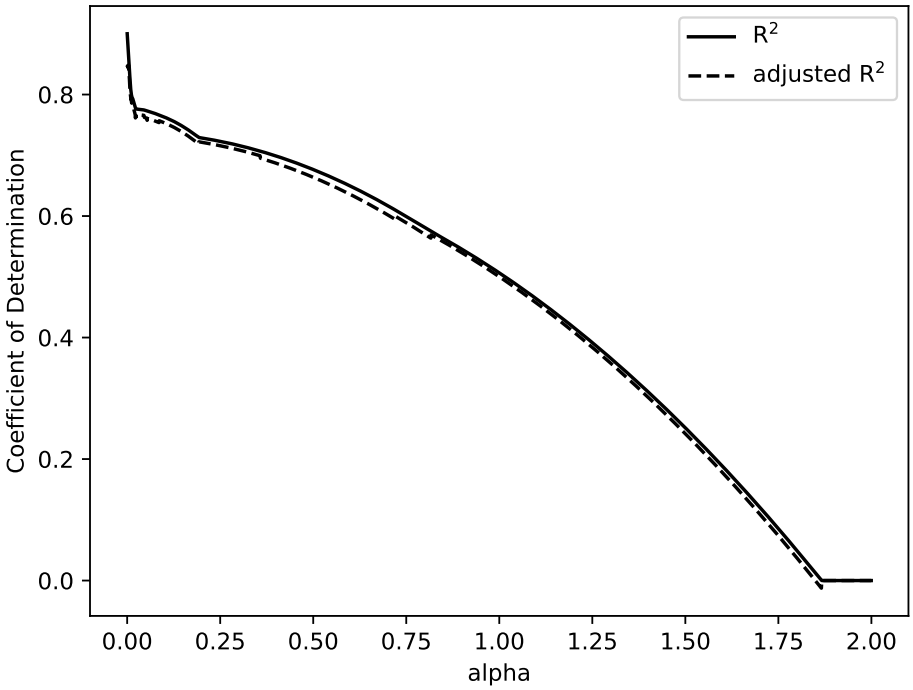
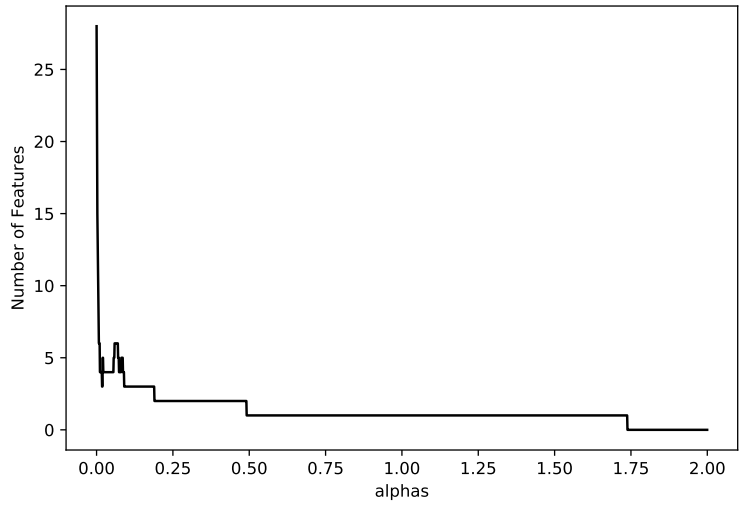
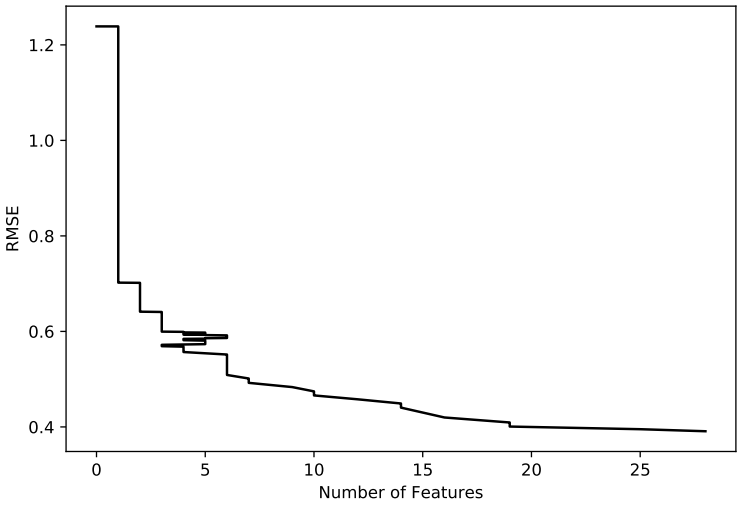


Figure 2.16: The CoD and adjusted CoD from the LASSO regression as a function of the alpha model parameter. As alpha increases, the CoD and adjusted coefficient of determination decrease. With alphas around 0.001-0.01, we see CoDs and adjusted CoDs near 0.8, with a maximum adjusted CoD of 0.85.



(a) The number of explanatory features from the LASSO regression, as a function of alpha.



(b) The RMSE of the LASSO regression as a function of the number of explanatory features.

Figure 2.17: Figure 2.17a shows that as alpha increases, the number of explanatory features decreases, while Figure 2.17b shows that as the number of features increases, the RMSES decreases. We can see the LASSO model with 28 features yields a low RMSE of 0.39.

Chapter 3

Conclusions

It is well know that *Sclerotinia sclerotiorum*, or white mold, has a significant impact on snap bean yield, with especially the flowering crop stage being sensitive to mold inoculation, while dense canopy structures are conducive to subsequent mold growth. The timing of fungicide application, i.e., to coincide with the snap bean flowering stage, thus is essential to limiting disease impacts, along with an assessment of plant canopy structure, to serve as an indication of mold risk after inoculation. In fact, previous studies have shown that mold occurrence is tightly coupled to denser canopy structures [29], i.e., the closed-canopy structures are conducive to white mold incubation. This in turn implies that i) accurate detection of actual flowering occurrence or even more ideal, predicting the onset of flowering, would be beneficial to optimized management of white mold, and that ii) image-based structural assessment could contribute to eventual mold risk model development. Our objectives for this work therefore were to identify spectral signatures for the onset of flowering in snap beans and use the same imagery to assess leaf area index (LAI), an indicator of plant canopy structural complexity. We used existing (spectral angle mapper, SAM; published indexes) and new methods (ratio thresholding,

RT) to filter the canopy for pure, sunlit pixels. A combination of SAM and RT were used to identify pure pixels in order to create predictive models for flowering. We were able to not only discriminate, but also predict flowering at two weeks prior to flowering onset, using single red wavelength features, with c-index values above 90%, using these methods. These wavelength/spectral features are known for their physiological coupling to plant health, photosynthetic activity, and inter-cellular structure [41][42][43][44][45].

Our structural models models, in turn, were based on both direct, raw reflectance features, as well as spectral indices that have been shown to be correlated to plant structure, e.g., leaf area index (LAI). Plant-level leaf structure/layering (LAI) were measured in situ and then modeled by performing linear and multivariate regressions between spatial and spectral features. Our work found that a single reflectance feature at 510 nm yielded a CoD of 0.76 and an RMSE of 0.609. Other features, including pixel density, EVI, NDVI, and GNDVI had CoD and RMSE values that ranged between 0.42-0.57 and 0.817-0.942, respectively. We then included 28 spatial and spectral features in a multivariate regression to predict LAI,and obtained an adjusted CoD of 0.85 and RMSE of 0.390. This arguably represents successful outcomes for both objectives: We could accurately detect and predict flowering onset and also assess canopy structural complexity. Both of these outcomes eventually can be used to develop a comprehensive mold risk model, although this final effort is beyond the scope of this thesis.

Future work could include a spectral convolution study to investigate the robustness of a multi-spectral solution (band center at 680 nm, bandpass width ∼40 nm), the placement of independent calibration panels in the field to assess the fidelity of the conversion to reflectance, and the incorporation of spectral reflectance values, LAI correlated spectral features, and auxiliary environmental factors (degree days, average rainfall, average temperature, etc.) into a single white mold risk model.

The key take-aways from this study was that highly accurate actual and/or predictive (two week in advance) flowering maps can be created for pro-active white mold management in snap beans, along with proven, accurate assessments of plant canopy structure, for improved risk modeling. These results bode well for eventual implementation toward more directed, judicious application of fungicide, which in turn will have significant impacts in terms of a reduced environmental footprint and optimization of yield/profit. It remains essential, however, that UAS-based solutions to pest management in crops focus on operational solutions, i.e., solutions that are distilled from expensive, research-grade equipment to accessible, low-risk, and cost-effective automated platforms.

Appendices

Appendix A

Mold Detection and Prediction Using Imaging Spectroscopy

A.1 Introduction to Mold Detection and Prediction

This appendix contains an additional exploration of whether or not we can differentiate diseased from healthy plants, i.e., we wanted to take the methods outlined in this thesis and apply them directly to the spectral detection and prediction of white mold. The mold trials contained 40 plots (see Figure A.1), 20 inoculated with the white mold disease (highlighted in yellow) and 20 control plots (not inoculated; highlighted in blue). The plots were 10 feet long and one row wide. These plots were monitored by Cornell University collaborators and were used in a fungicide efficacy experiment. That implies that plants were inoculated with mold and then treated with fungicide, to evaluate how the fungicide affected white mold incidence at the end of the season (August 15th, 2018). Another challenge to our mold detection study was that only eight of the 20 plots inoculated with mold exhibited white mold incidence greater than 50%. This significantly reduced the size of the training/testing sets. While this experiment wasn't designed with spectral detection and prediction of disease in mind, we wanted to determine if the methods outlined in

this thesis, which were effective at predicting flowering onset, could also detect and predict disease.

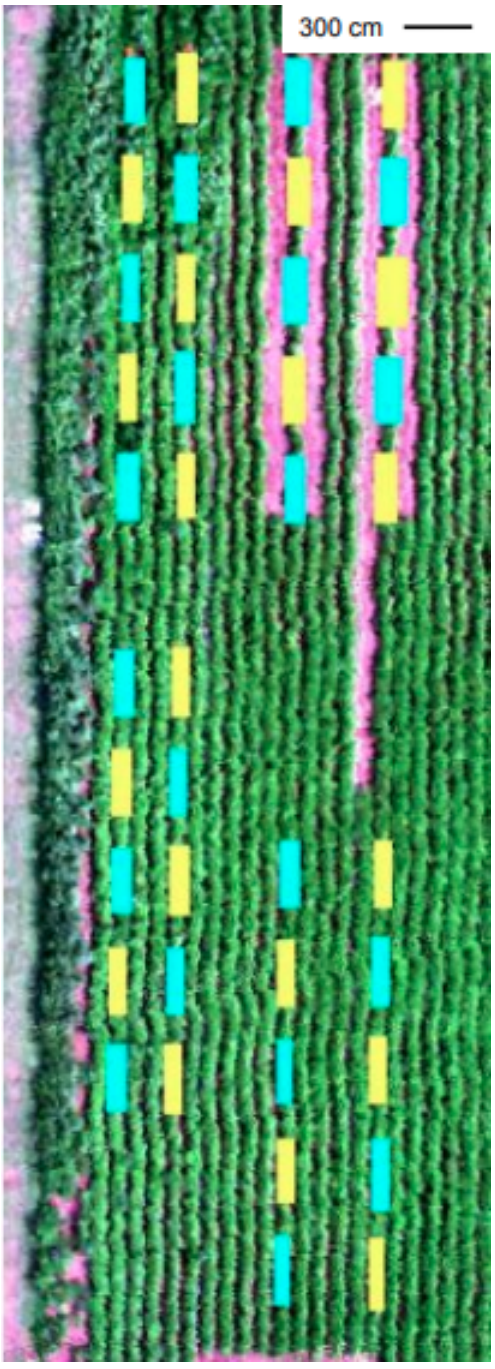


Figure A.1: Location of the Mold Trials 2018 Season. Yellow and blue regions indicate inoculated and uninoculated with mold, respectively.

A.2 Detection of Pure Pixels

SAM and RT should not be used to detect pure pixels of diseased plants that contain mold, since the selected vegetation spectra may not approximate that of a healthy plant anymore. The vegetation classification and pure pixel extraction therefore is performed in a single step. To accomplish this, we assume that pixels with the largest broadband spectral power, i.e., when we integrate the reflectance spectrum over the wavelength range and then divide it by the number of wavelengths, represent healthy plants. The result is a number that represents the average reflective power (ARP) of each pixel. The "brightest" pixels, or the pixels with the most power, are the most representative of the pure pixel spectra. Next, similar to the RT approach, ARP extracts all the pixels that are above the average, plus the modulated standard deviation. A multiplier of zero again was selected for ARP, since it resulted in a threshold that selected center-of-canopy pixels, while preserving the most samples. Psuedocode for ARP can be found below in Figure A.2.

A.3 Features and Model Development for Mold Detection and Prediction

Raw reflectance, RI, and NDI features (from Figure 2.1) were used along with a SFLR [46] to separate the classes that contain or do not contain mold, and will/won't contain mold in the future, i.e., for the predictive step. The top performing feature was used along with a SVM [49][50][51][52][53]

Algorithm 2: Average Reflective Power

```
image = array with shape samples x features
start and end refer to the wavelengths you want to keep
n = feature length of shortened image
short = image[:, start:end]
for i samples do
   arp[i] = sum(short[i, :]) / n
end for
avg = mean(arp)
std = std(arp)
threshold = avg + multiplier · std
indices vegetation = where(arp > threshold)
indices background = where(arp < threshold)
vegetation = image[indices vegetation, :]
background = image[indices background, :]
return vegetation, background
```

Figure A.2: This figure contains psuedocode Average Reflective Power. These methods were used for finding pure pixels for mold detection.

to discriminate mold. Top LAI-correlated features, from our spectra-to-LAI study also were used along with a SVM to predict mold.

The discriminating mold model was created using data collected on August 15th, 2018 (3 cm GSD). Two plots, where greater than 80% of the plants exhibited white mold, and two plots, where the plants had 0% mold, were used to create the model. The model then was applied to a plot where 78% of the snap beans contained mold, and a plot where 0% of the plants contained mold. It should be noted that our ground truth dataset contained significantly fewer samples than was the case for the flowering model development; case-in-point, we were able to use 8x more training data for flowering ($\sim$4,700 samples) than for mold detection ($\sim$700 samples). This could result in poorer model performance, over-and-above the complexities related to biophysical changes in vegetative growth, associated with either flowering and mold events. Finally, our expectation was that mold would cause a change in canopy-level reflectances, when compared to each stage's counterpart, i.e., non-mold.

The top N features from the LASSO regression (LAI-correlated features) then were used to train and test a SVM with a radial basis function [49] on data from August 6, 2018 in the mid planting; this was done in order to create the predictive white mold occurrence model. Two plots with white mold incidence above 76% and two control plots (0% white mold incidence) were used to train the model. The model was then applied to test plots with white mold incidence of 78.3% and 0%, respectively. The reason that August

6 is an important date, is based ont he fact that 100% of the snap beans were flowering in the mid-planting and on August 15, 2018, some of these snap beans were diseased with mold. We wanted to evaluate whether the features correlated with LAI can predict the occurrence of the disease, when applied to flowering snap beans.

A.4 Results and Discussion for Mold Discrimination and Prediction

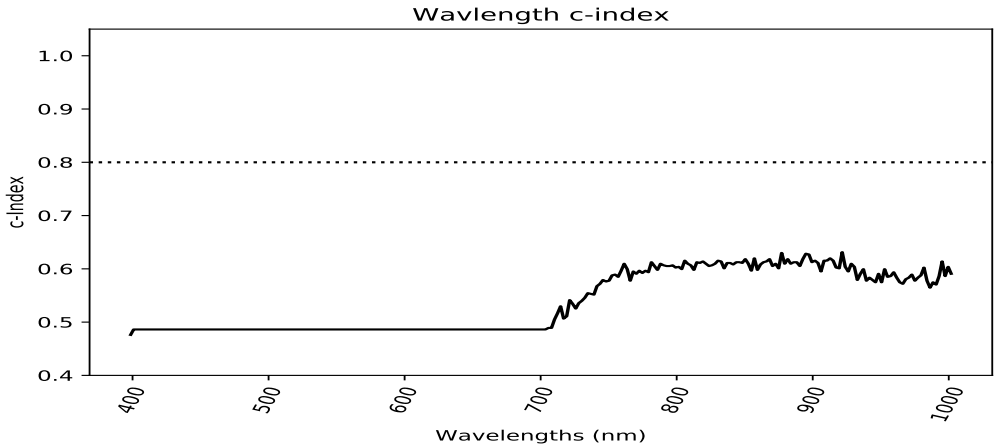
All of the c-indexes for mold discrimination were below 0.8; in fact, the values were closer to 0.5, which implies random/chance discrimination. The sensitivity and specificity plots show that the SFLR exhibited either high false positive rates (FPR) or false negative rates (FNR). This is indicative of models that are largely guessing whether or not the snap bean plant contains mold or not. In Figure 8b we see that, for wavelengths between 400-700 nm, the model has a false negative rate of approximately 1.0, which means that the number of true positives is zero, i.e., the model is consistently guessing no-mold, even if this is invalid. The results from the SFLR for mold can be found below in Figures A.3-A.5. This result was contrary to our hypothesis, i.e., that white mold would have a distinct physiological impact on plants, which would be observable via canopy-level spectral indicators. We attributed this outcome to high within-group spectral variability, although this needs to be verified in future research.

Figure A.6 contains test images where the mold discrimination model was applied. The results show that the model cannot distinguish between snap beans that contain no mold and regions where 78% of the plants exhibited mold. This is an unexpected result, in that it seems that imaging spectroscopy of the snap bean canopy, collected via an UAS platform, is unable to differentiate between healthy and diseased plants. This must imply that the spectral variability induced by white mold are small, relative to the variance in the canopy. This outcome is supported by Vigier et al. (2004), who attempted to separate soy beans diseased with white mold one month before the harvest, using ground-based spectroradiometers at a height of $2.25m$ above the canopy. Their results show that there is a larger difference in canopy spectra between growing seasons, than between diseased and healthy beans [64]. Machado et al. (2015) also used a ground-based spectroradiometer to measure the reflectance of snap bean leaves diseased with white mold in a lab setting. The results showed that, even under ideal laboratory conditions, the difference between plants with a disease severity index (DSI) of 0 and 50 was far less than a 3% reflectance change [65]. Finally, Martinez-Martinez et al. (2018) concluded that the spectrum of beans diseased with angular leaf spot could not be discriminated from healthy beans with a ground-based spectroradiometer from heights greater than $1m$ [66]. Although such studies corroborate our findings, it arguably is of limited, other than academic importance; i.e., once white mold is observed in the snap bean crop, the crop yield already will have suffered significantly.

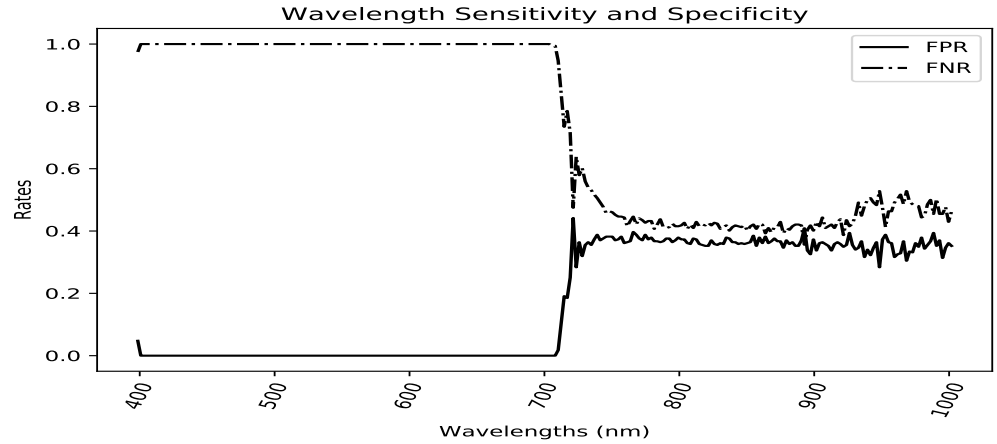
Using 28 LAI correlated features (CoD=0.85, RMSE=0.39), we trained a SVM with a radial basis function to predict white mold. Training data consisted of two images that contained 76% white mold and two images that contained 0% white mold. The model then was applied to test images that contained 78.3% mold and 0% mold. The results from the applied model can be found below in Figure A.7a and A.7b. The results of the applied model both yield average canopy probabilities for mold of less than 50%. This means the hyper-plane that was fit using the training data was unable to separate the two classes. This was attributed to the limited amount of mold vs. no mold ground truth data (i.e., only three plots with white mold incidence greater than 75%). The mold/no mold class sizes are significantly smaller than the flowering data sets, which resulted in more successful models. As stated before, plots were inoculated with mold and then treated with select fungicides. This potentially is another factor that could be negatively affecting our results, since the white mold incidence is suppressed via fungicide applications and as such did not occur naturally. Stated differently, if we were trying to identify naturally predictive features for mold, but are either forcing or encouraging the mold to occur, then the selected features may not be representative at all.

We concluded that the disease signatures we are trying to identify are variable, relative to the variance of the canopy, and obviously more variable than the flower discrimination work outlined in Chapter 2, which was also corroborated by other research [64][65][66]. Therefore, future research should

include the use of larger data sets with snap beans naturally diseased with white mold (i.e., not inculcated with disease and then treated with fungicide). This aid in the development of a SVM hyper-plane model to separate predictive classes. Averaging spectra over the canopy also decreases our ability to detect/predict small signatures, so an automated system with smaller GSDs, such as a terrestrial imaging platform, could increase our ability to detect/predict disease in the field.

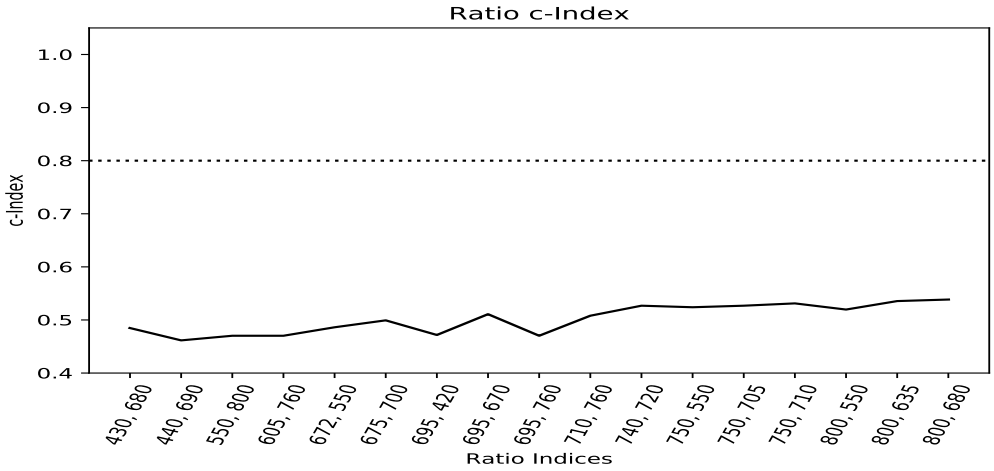


(a) c-index for Wavelength Single Feature Logistic Regression for Mold Discrimination; note that no spectral features exceed the 0.8 c-index detection level

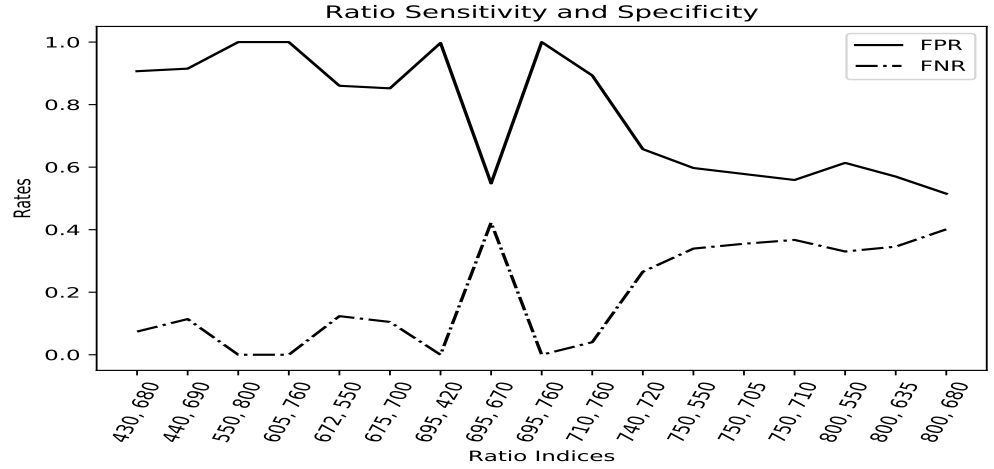


(b) False Positive and False Negative Rates for Wavelength Single Feature Logistic Regression for Mold Discrimination

Figure A.3: This figure shows the c-index, false positive, and false negative rates for each of the wavelengths used to separate the classes mold vs. no-mold. Features that contain wavelengths in the NIR region have the largest c-index and lowest false positive and false negative rates. However, the results are close to random/chance discrimination. The results in b) show that from wavelengths in the 400-700 nm region, the model has a false negative rate of almost 1.0 and a false positive rate of almost zero. This means the model consistently guesses no-mold.

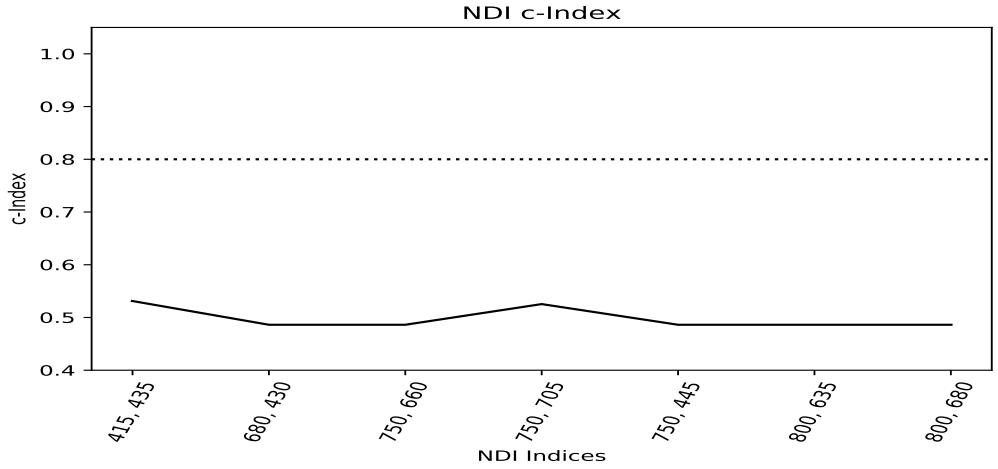


(a) c-index for Ratio Index Single Feature Logistic Regression for Mold Discrimination; again there are no spectral features, ratio indices in this case, that exceed the 0.8 c-index detection threshold

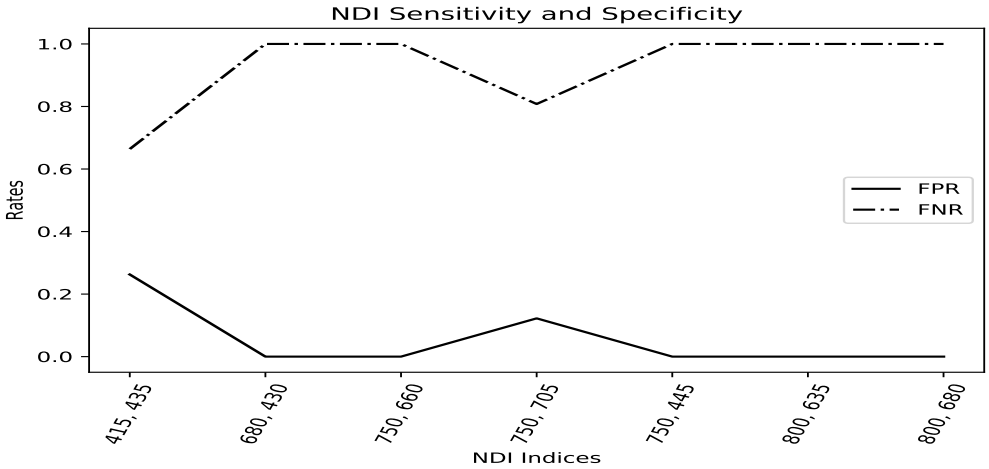


(b) False Positive and False Negative Rates for Ratio Index Single Feature Logistic Regression for Mold Discrimination

Figure A.4: This figure shows the c-index, false positive, and false negative rates for each of the ratio indices used to separate the classes mold vs. no-mold. All the features have a c-index less than 0.6 (close to random/chance discrimination). The results in b) show that features have large false positive rates and low false negative rates. This means the model is biased towards consistently guessing that mold is present.

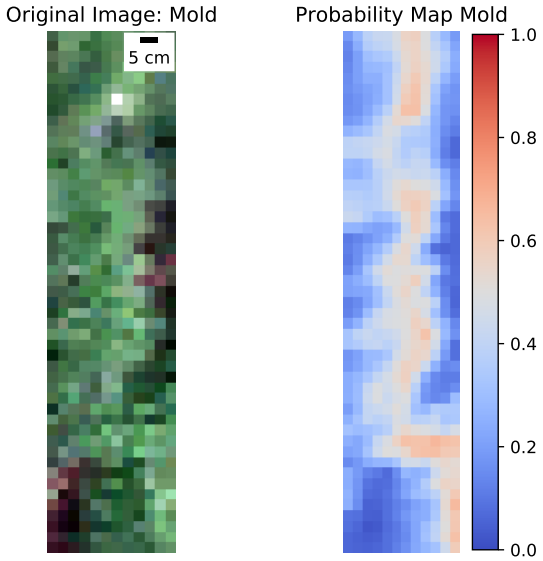


(a) c-index for Normalized Difference Index Single Feature Logistic Regression for Mold Discrimination

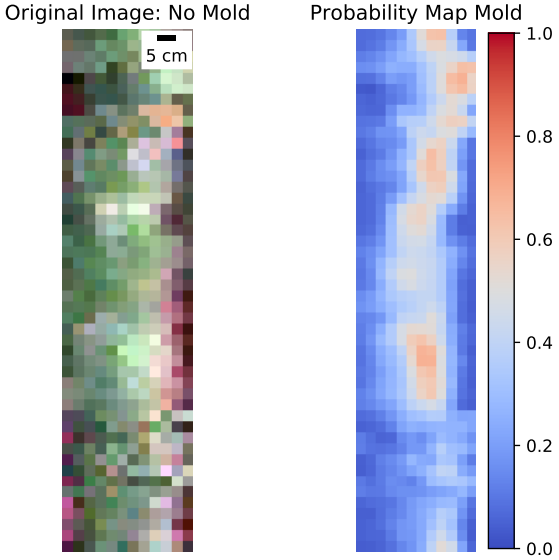


(b) False Positive and False Negative Rates for Normalized Difference Index Single Feature Logistic Regression for Mold Discrimination

Figure A.5: This figure shows the c-index, false positive, and false negative rates for each of the normalized difference indices (NDIs) used to separate the classes mold vs. no-mold. All the features have a c-index around 0.5 (random/chance discrimination). The results in b) show that features have large false negative rates and low false positive rates, which implies that the model is biased towards guessing no-mold.

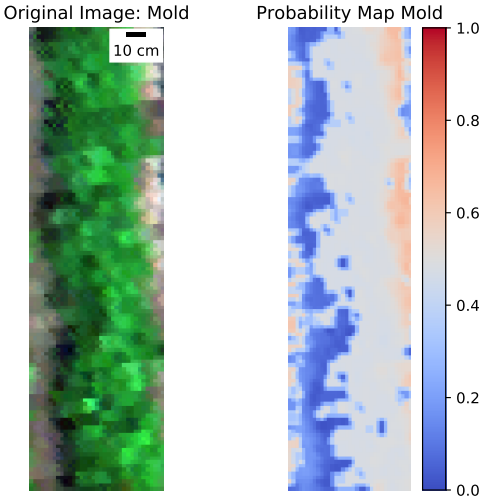


(a) Discriminating mold model on a mold test image

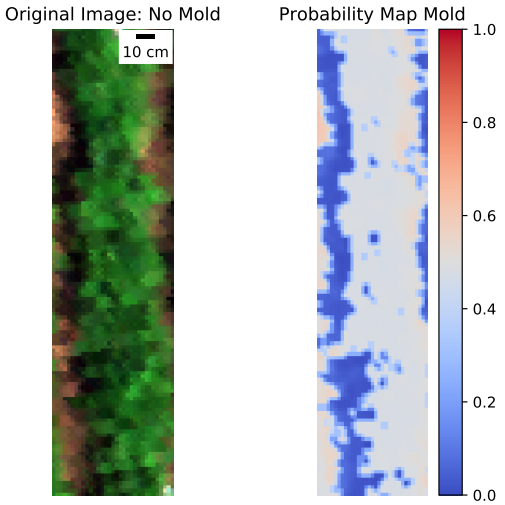


(b) Discriminating mold model on an image with no mold present

Figure A.6: This figure shows the results of the mold discrimination model using a single NIR feature on images that have mold a) and no-mold b). The model yields probabilities of close to 0.5-0.6 for mold and no-mold examples. It is evident that the model cannot reliably discriminate between the two classes.



(a) The outcome for 28 LAI-correlated features, when used to fit a SVM to predict mold and applied to an image containing 76.2% white mold incidence.



(b) the result for 28 LAI-correlated features, when used fit a SVM to predict mold and applied to an image containing 0% white mold incidence.

Figure A.7: Figures A.7a and A.7b show the probability maps from the SVM trained on LAI-correlated features to predict mold nine days before occurrence. The test images in Figures A.7a and A.7b contains 78.3% and 0% white mold incidence, respectively. The figures show that the average probability for mold over both canopies is approximately 48%.

A.5 White Mold Detection/Prediction Conclusions

We found that applying the successful flowering detection and prediction methods to mold discrimination and prediction was unsuccessful. The spectral and estimated structural metrics were not conducive to disease detection (contains mold/does not contain mold) and prediction (will contain mold/ will not contain mold). We believe that the low sample size for diseased snap beans, the concurrent fungicide efficacy trials, the inherent spectral variance of the canopy, and spectral averaging over the canopy pixels led to the diminished success of our ability to detect and predict mold occurrence. Future work should include an increased number of mold monitored plots, naturally occurring white mold incidence (no fungicide applications), and potentially a terrestrial-based method for imaging spectroscopy data acquisition, which would reduce the GSD and averaging over canopy pixels.

Bibliography

[1] S. Delalieux, B. Somers, W. W. Verstraeten, J. van Aardt, W. Keulemans, and P. Coppin, "Hyperspectral indices to diagnose leaf biotic stress of apple plants, considering leaf phenology," *International Journal of Remote Sensing*, April 2009.

[2] E. W. Hughes, S. J. Pethybridge, J. R. Kikkert, C. Salvaggio, and J. van Aardt, "Snap bean flowering detection from uas imaging spectroscopy," *14th Internation Conference on Precision Agriculture*, 2018.

[3] U. N. A. S. Service, "Census of agriculture," *U.S. Dep. Agric. Nat. Agric. Stat. Serv. Online*, 2012.

[4] U. N. A. S. Service, "The 2014-2015 new york annual bulletin," *U.S. Dep. Agric. Nat. Agric. Stat. Serv. Online*, 2014.

[5] D. A. Shah, H. R. Dillard, and A. C. Cobb, "Alternatives to vinclozolin (ronilan) for controlling gray and white mold on snap bean pods in new york," *Plant Health Prog. [online]*, 2002.

[6] C. M. McCreary, D. Depuydt, R. J. Vyn, and C. L. Gillard, "Fungicide efficacy of dry bean white mold [(sclerotinia sclerotiorum (lib.) de bary,

causal organism] and economic analysis at moderate to high disease pressure," *Crop Prot.*, vol. 82, pp. 75–81, 2016.

[7] J. R. Steadman, D. P. Coyne, and G. E. Cook, "Reduction of severity of white mold disease on great northern beans by wider row spacing and determinate growth habit," *Plant Dis. Rep.*, vol. 57, pp. 1070–1071, 1973.

[8] B. L. Blad, J. R. Steadman, and A. Weiss, "Canopy structure and irrigation influence white mold disease and microclimate of dry edible beans," *Phytopathology*, vol. 68, pp. 1431–1437, 1978.

[9] C. R. Grau and V. L. Radke, "Effects of cultivars and cultural practices on sclerotinia stem rot of soybean," *Plant Dis.*, vol. 68, pp. 56–58, 1984.

[10] C. R. Grau and G. L. Hartman, "Sclerotinia stem rot," *In G. I. Hartman, J. B. Sinclair, and J. C. Rupe (eds.). Compendium of soybean diseases*, vol. 4th, pp. 46–48, 1999.

[11] A. J. Peltier, C. A. Bradley, M. I. Chilvers, D. K. Malvick, D. S. Nueller, K. A. Wise, and P. D. Esker, "Biology, yield loss and control of sclerotinia stem rot of soybean," *J. Integr. Pest Mngmt*, vol. 3, p. 2012, 2013. DOI: http://dx.doi.org/10.1603/IPM11033.

[12] M. Lehner, J. Kikkert, B. Gugino, and S. Pethybridge, "Fungicide sensitivity and efficacy of boscalid, fluazinam and thiophanate-methyl for

white mold control in snap bean," *Plant Disease*, vol. 101, pp. 1253–1258, 2017.

[13] A. L. Mila, A. L. Carriquiry, J. Zhao, and X. B. Yang, "Impact of management practices on prevalence of soybean sclerotinia stem rot in the north-central united states and on farmers' decision under uncertainty," *Plant Dis*, vol. 87, pp. 1048–1058, 2003.

[14] S. J. Jones, D. H. Gent, S. J. Pethybridge, and F. S. Hay, "Site-specific risk factors of white mould epidemics in bean (phaseolus vulgaris) in tasmania, australia," *N. Z. J. Crop Hort. Sci.*, vol. 40, pp. 147–159, 2012.

[15] M. R. McDonald and G. J. Boland, "Forecasting diseases caused by sclerotinia spp. in eastern canada: fact or fiction?," *Can. J. Plant Pathol.*, vol. 26, pp. 480–488, 2004.

[16] D. Hosmer and S. Lemeshow, "Applied logistic regression," *John Wiley & Sons, New York*, 1989.

[17] A. L. Mila, A. L. Carriquiry, and X. B. Yang, "Logistic regression modeling of prevalence of soybean sclerotinia stem rot in the north-central united states," *Phytopathology*, vol. 94, pp. 102–110, 2004.

[18] P. A. Paul and G. P. Munkvold, "A model-based approach to preplanting risk assessment for gray leaf spot of maize," *Phytopathology*, vol. 94, pp. 11350–1357, 2004.

[19] P. D. Esker, J. Harri, P. M. Dixon, and F. W. Nutter, "Comparison of models for forecasting of stewart's wilt of corn in iowa," *Plant Dis.*, vol. 90, pp. 1353–1357, 2006.

[20] R. Harikrishnan and L. E. del Rio, "A logistic regression model for predicting risk of white mold incidence on dry bean in north dakota," *Plant Dis.*, vol. 92, pp. 42–46, 2008.

[21] R. D. Magarey, D. M. Bochert, J. S. Engle, M. Colunga-Garcia, F. H. Koch, and D. Yemshanov, "Risk maps for targeting exotic plant pest detection programs in the united states," *EPPO Bull*, vol. 41, pp. 46–56, 2011.

[22] S. Sankaran, A. Mishra, R. Ehsani, and C. Davis, "A review of advanced techniques for detecting plant diseases," *Comp. Electr. Agr.*, vol. 72, pp. 1–13, 2010.

[23] A. K. Mahlein, T. Rumpf, H. W. D. P. Welke, L. Plumer, U. Steiner, and E. C. Oerke, "Development of spectral indices for detecting and identifying plant diseases," *Rem. Sens. Env.*, vol. 128, pp. 21–30, 2013.

[24] T.Rumpf, A. Mahlein, U. Steiner, E.Oerke, H. Dehne, and L.Plümera, "Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance," *Computers and Electronics in Agriculture*, vol. 74, pp. 91–99, October 2010. Full text available.

[25] S. S. Durbha, R. L. King, and N. H. Younan, "Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer," *Remote Sensing of Environment*, vol. 107, no. 1, pp. 348 – 361, 2007. Multi-angle Imaging SpectroRadiometer (MISR) Special Issue.

[26] M. Berterretche, A. T. Hudak, W. B. Cohen, T. K. Maiersperger, S. T. Gower, and J. Dungan, "Comparison of regression and geostatistical methods for mapping leaf area index (lai) with landsat etm+ data over a boreal forest," *Remote Sensing of Environment*, vol. 96, no. 1, pp. 49 – 61, 2005.

[27] J. M. Chen and J. Cihlar, "Retrieving leaf area index of boreal conifer forests using landsat tm images," *Remote Sensing of Environment*, vol. 55, no. 2, pp. 153 – 162, 1996.

[28] R. Colombo, D. Bellingeri, D. Fasolini, and C. M. Marino, "Retrieval of leaf area index in different vegetation types using high resolution satellite data," *Remote Sensing of Environment*, vol. 86, no. 1, pp. 120 – 131, 2003.

[29] D. Shah, H. Dillard, and S. Pethybridge, "Identification of factors associated with white mould in snap bean using tree-based methods," *Plant Pathology*, 08 2019.

[30] J. W. Rouse, R. H. Haas, J. A. Well, and D. W. Deering, "Monitoring vegetation systems in the great plains with erts," *NASA. Goddard Space*

Flight Center 3d ERTS-1 Symp., vol. 1, pp. 309–317, Jan. 1974.

[31] A. A. Gitelson and M. N. Merzlyak, "Remote sensing of chlorophyll concentration in higher plant leaves," *Advances in Space Research*, vol. 22, pp. 689–692, 1998.

[32] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the radiometric and biophysical performance of the modis vegetation indices," *Remote Sensing of Environment*, vol. 83, pp. 195–213, 2002.

[33] J. Norman and P. Jarvis, "Photosynthesis in sitka spruce (picea sitchensis (bong.) carr.): V. radiation penetration theory and a test case," *Journal of Applied Ecology*, vol. 12, pp. 839–878, 1975.

[34] J. Goudriaan, "The bare bones of leaf angle distribution in radiation models for canopy photosynthesis and energy exchange," *Agricultural and Forest Meteorology*, vol. 43, pp. 155–169, 1988.

[35] G. Campbell, "Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution," *Agricultural and Forest Meteorology*, vol. 36, pp. 317–321, 1986.

[36] G. Smith and E. Milton, "The use of the empirical line method to calibrate remotely sensed data to reflectance," *International Journal of Remote Sensing*, vol. 20, pp. 2653–2662, 1999.

[37] M. Morana, R. Bryant, K. Thome, W.Ni, Y. Nouvellon, M. Gonzalez-Dugo, J. Qi, and T. Clarke, "A refined empirical line approach for reflectance factor retrieval fromlandsat-5 tm and landsat-7 etm+," *Remote Sensing of Enviornment*, vol. 78, pp. 71–82, 2001.

[38] P. E. Dennison, K. Q. Halligan, and D. A. Roberts, "A comparison of error metrics and constraints for multiple endmemberspectral mixture analysis and spectral angle mapper," *Remote Sensing of Enviornment*, vol. 93, pp. 359–367, 2004.

[39] H. Yagoub, A. H. Belbachir, and N. Benabadji, "Detection and mapping vegetation cover based on the spectral angle mapper algorithm using noaa avhrr data," *Advances in Space Research*, vol. 53, pp. 1686–1693, June 2014.

[40] H. Zhuang, K. Deng, H. Fan, and M. Yu, "Strategies combining spectral angle mapper and change vector analysis to unsupervised change detection in multispectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 681–685, 2016.

[41] M. R. Slaton, E. R. H. Jr., and W. K. Smith, "Estimating near-infrared leaf reflectance from leaf structural characteristics," *American Journal of Botany*, vol. 88, p. 278–284, 2001.

[42] J. A. Gamon, C. B. Field, M. L. Goulden, K. L. Griffin, A. E. Hartley, G. Joel, J. Peñuelas, and R. Valentini, "Relationships between ndvi,

canopy structure, and photosynthesis in three californian vegetation types," *Ecological Applications*, vol. 5, pp. 28–41, 1995.

[43] J. A. Gamon, L. Serrano, and J. S. Surfus, "The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels," *Oecologia*, vol. 112, pp. 492–501, 1997.

[44] D. Sims and J. Gamon, "Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages," *Remote Sensing of Environment*, vol. 81, pp. 337–354, 08 2002.

[45] D. M. Simms, T. W. Waine, J. C. Taylor, and G. R. Juniper, "The application of time-series modis ndvi profiles for the acquisition of crop information across afghanistan," *International Journal of Remote Sensing*, vol. 35, pp. 6234–6254, 2004.

[46] S. Delalieux, J. van Aardt, W. Keulemans, E. Schrevens, and P. Coppin, "Detection of biotic stress (venturia inaequalis) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications," *European Journal of Agronomy*, July 2007.

[47] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.

[48] G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, S. Huang, M. Brooks, M. J. Lee, and H. Asadi, "Peering into the black box of artificial intelligence: Evaluation metrics of machine learning methods," *American Journal of Roentgenology*, vol. 212, pp. 38–43, 2019.

[49] B. Scholkopf, K.-K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with gaussian kernels to radial basis function classifiers," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2758–2765, Nov 1997.

[50] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, August 2004.

[51] L. Su, M. J. Chopping, A. Rango, J. V.Martonchik, and D. P. Peters, "Support vector machines for recognition of semi-arid vegetation types using misr multi-angle imagery," *Remote Sensing of Environment*, vol. 107, pp. 299–311, 2007.

[52] A. C. Braun, U. Weidner, and S. Hinz, "Support vector machines for vegetation classification," *Photogrammetrie - Fernerkundung - Geoinformation*, vol. 2010, pp. 273–281, 2010.

[53] D. Liang, Q. Guan, W. Huang, L. Huang, and Y. Guijun, "Remote sensing inversion of leaf area index based on support vector machine regression

in winter wheat," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 29, pp. 117–123, 2013.

[54] D. Lambert, J. Lowenberg-Deboer, and R. Bongiovanni, "A comparison of four spatial regression models for yield monitor data: A case study from argentina," *Precision Agriculture*, vol. 5, p. 579–600, 2004.

[55] D. C. Lazaridis, J. Verbesselt, and A. P. Robinson, "Penalized regression techniques for prediction: a case study for predicting tree mortality using remotely sensed vegetation indicesthis article is one of a selection of papers from extending forest inventory and monitoring over space and time.," *Canadian Journal of Forest Research*, vol. 41, no. 1, pp. 24–34, 2011.

[56] M. A. Stapanian, J. Mack, J. V. Adams, B. Gara, and M. Micacchion, "Disturbance metrics predict a wetland vegetation index of biotic integrity," *Ecological Indicators*, vol. 24, pp. 120–126, 2013.

[57] J. Verrelst, G. Camps-Valls, J. Muñoz-Marí, J. P. Rivera, F. Veroustraete, J. G. Clevers, and J. Moreno, "Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 108, pp. 273–290, 2015.

[58] S. Kim, K.-A. Sohn, and E. P. Xing, "A multivariate regression approach to association analysis of a quantitative trait network," *Bioinformatics*, vol. 25, pp. 204–212, 2009.

[59] Y. Zhang, F. Ma, and Y. Wang, "Forecasting crude oil prices with a large set of predictors: Can lasso select powerful predictors?," *Journal of Empirical Finance*, vol. 54, pp. 97–117, 2019.

[60] J. E. Vogelmann, B. N. Rock, and D. M. Moss, "Red edge spectral measurements from sugar maple leaves," *International Journal of Remote Sensing*, vol. 14, pp. 1563–1575, 1993.

[61] P. J. Curran, W. R. Windham, and H. L. Gholtz, "Exploring the relationship between reflectance red edge and chlorophyll concentration in slash pine leaves," *Tree Physiology*, vol. 15, pp. 203–206, 1995.

[62] J. R. Taylor, *An Introduction to Error Analysis*. Sausalito, California: University Science Books, 1997.

[63] C.-H. Xiao, S.-K. Li, K.-R. Wang, Y.-L. Lu, J.-H. Bai, R.-Z. Xie, S.-J. Gao, X.-J. Li, and H.-Z. Tan, "The response of canopy direction reflectance spectrum for the wheat vertical leaf distributing," *Crop Modeling and Decision Support*, pp. 77–85, 2009.

[64] B. Vigier, E. Pattey, and I. Strachan, "Narrowband vegetation indexes and detection of disease damage in soybeans," *Geoscience and Remote Sensing Letters, IEEE*, vol. 1, pp. 255 – 259, 11 2004.

[65] M. L. Machado, F. de A. C. Pinto, T. J. de Paula Junior, D. M. de Queiroz, and O. de A. T. Cerqueira, "White mold detection in common beans

through leaf reflectance spectroscopy," *Eng. Agríc.*, vol. 35, 2015.

[66] V. Martínez-Martínez, J. Gomez-Gil, M. L. Machado, and F. A. C. Pinto, "Leaf and canopy reflectance spectrometry applied to the estimation of angular leaf spot disease severity of common bean crops," *PLoS ONE*, vol. 13, 2018.