Summer 8-22-2020

Airborne Hyperspectral Data Application in Stress Detection of Blueberry Fields and Ash Trees

Catherine Chan
University of Maine, catherine.chan@maine.edu

AIRBORNE HYPERSPECTRAL DATA APPLICATION IN STRESS DETECTION OF BLUEBERRY FIELDS AND ASH TREES

By
Catherine Chan
B.S. Boston College, 2014

A THESIS
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
(in Forest Resources)

The Graduate School
The University of Maine
August 2020

Advisory Committee:

Peter Nelson, Assistant Professor, Biological Sciences & Environmental Studies, Co-Advisor
Daniel Hayes, Assistant Professor, Geospatial Analysis & Remote Sensing, Co-Advisor
Ryan Hanavan, Forest Entomologist - U.S. Forest Service
Aaron Weiskittel, Professor of Forest Biometrics & Modeling
Shawn Fraver, Associate Professor, Forest Ecology

AIRBORNE HYPERSPECTRAL DATA APPLICATION IN STRESS DETECTION OF BLUEBERRY FIELDS AND ASH TREES

By Catherine Chan

Co-advisors: Drs. Peter Nelson & Daniel Hayes

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Forest Resources)
August 2020

Water management and irrigation practices are persistent challenges for many agricultural systems. Changing seasonal and weather patterns impose a greater need for understanding crop deficiencies and excesses (e.g. water, sunlight, nutrients) for optimal growth while allocating proper resources for prompt response. The wild blueberry industry is at heightened susceptibility due to its unique growing conditions and uncultivated nature. Early detection of stress in agricultural fields can prompt management responses to mitigate detrimental conditions including drought and disease. Remote sensing has provided timely and reliable information covering large spatial extents, while novel applications in hyperspectral data and imaging spectroscopy have shown potential in early stress detection. We assess airborne spectral data accompanied by ground sampled water potential over three developmental stages of wild blueberries to accurately detect water content.

Airborne scans of spectral data were collected three times throughout the 2019 summer in Deblois, Maine. Data were collected over two adjacent fields, one irrigated and one non-irrigated. Ground sampled data were collected in tandem to the UAV collection. The ground sampled data over the irrigated and non-irrigated fields guided digital sampling from the imagery to act as training for our models. Using methods in machine learning and statistical analysis, we related hyperspectral reflectance measurements to different water potential levels in blueberry plant leaves to decipher vegetation signals both spatially and temporally through utilizing the capacity of imaging spectroscopy.

Models were developed to determine irrigation status and water potential. Seven models were assessed in this study with four used to process six hyperspectral cube images for analysis. These images were classified as irrigated or non-irrigated and estimated water potential levels. Our global water potential model had an R^2 of 0.62. Models for the water potential predictions were verified with a validation dataset.

Forest insect and disease pests have a significant impact on the well-being of individual trees and forest stands, affecting ecosystem processes and potentially human health. Dispersing through 35 states within only 17 years (USDA, 2020), the effect of emerald ash borer (*Agrilus Planipennis Fairmaire*) (EAB) in the United States has been particularly severe and devastating. Early detection of stress in forests can prompt management responses to mitigate detrimental conditions including drought and disease as well as pest outbreaks. Remote sensing has provided timely and reliable information covering large spatial extents, while novel applications in hyperspectral data and imaging spectroscopy have shown potential in early stress detection. We build on previous work by assessing airborne spectral data, and health classifications of EAB infested ash trees in aims to accurately detect stress.

Airborne scans of spectral data were collected within three days in late July 2019 over three sites in southern New Hampshire. Ground sampled data were collected in November 2019 and include sampled ash classified on a scale of 1-5 (1=healthy, no major branch morality, 5=dead). The ground sampled data of different health classifications guided digital sampling from the imagery to act as training and validation for our models. Using methods in machine learning and statistical analysis, we related reflectance measurements to different classifications of ash tree health to understand tree stress signals while utilizing the capacity of remote sensing.

Models were developed to classify health in ash trees impacted by EAB. The first entailed a shadow classifier, followed by one for health. Eighteen cube images contained ground sampled data and were processed with the two models, then further buffered. Pixel classification for each buffer sample was calculated. The health classifier model was used on a validation test set and had an prediction accuracy of 76.1%.

DEDICATION

Dedicated to Vivian

1995-2019

ACKNOWLEDGEMENTS

I would like to thank my advisors, Peter Nelson and Daniel Hayes, for giving me the opportunity to pursue a new life path and passion. Their consistent patience, encouragement, and unique advising styles have been invaluable to my experience and learnings.

The assistance of Kevaughan Smith has been instrumental to this project. His contributions from data collection to model formation heavily comprise this thesis. Additionally, I am grateful for his support as a peer, and found camaraderie in our partnership.

Data for the blueberry component of this thesis was collected from the lab of YongJiang Zhang. The "blueberry team" was very accommodating and kind in their assistance to this thesis while completing their own project objectives.

Ground data and field logistics for the ash component of this thesis was made possible through Ryan Hanavan. I also thank him for his guidance and support as a committee member.

I would like to express my gratitude to Shawn Fraver and Aaron Weiskittel for serving on my committee. I have valued their inquiries and input as well as their teachings as professors in the School of Forest Resources.

I lastly give thanks to the Maine Economic Initiative Fund, Small Campus Initiative managed by Ryan Low, along with the Maine Research Reinvestment Fund with organization through Jason Charland of the Office of Research Development. Without these support opportunities, this research would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

AIRBORNE HYPERSPECTRAL DATA APPLICATION IN ESTIMATING WATER POTENTIAL OF WILD BLUEBERRY FIELDS

1.1. Introduction

Maine is the top producer of wild blueberries (*Vaccinium angustifolium* Ait.) (also referred to as lowbush) in the world (Yarborough, 2004), acting as the state's 4th highest agricultural revenue generator at 47.2 million dollars (Bertone, 2017). Lowbush blueberries require unique growing conditions including acidic and infertile soils, unsuitable for many other crop types (Glass, 2005). This distinctive environment increases vulnerability due to its uncommon land makeup, further amplified by climate change factors. In recent years, wild blueberry production has faced challenges relating to drought, freezing, and pathogens (Whittle, 2018). As a result, forecasting land conditions and taking prompt mitigative action, including water management, have become increasingly necessitated. To assist farmers in water management practices, we analyzed reflectance rates of blueberry crops throughout different developmental stages to detect patterns of water potential temporally and spatially.

The application and use of hyperspectral imaging technology has made great advances in the past decade. Hyperspectral imaging is a method in detecting and classifying objects based on the light reflectance rate (or spectral signature) on the electromagnetic spectrum. Objects project different signatures at different wavelengths as a result of their composition. Its current functions range from detecting human tissue damage (Leavesley, 2016) to aiding agricultural quality control (Nguyen-Do-Trong, 2018). One of its most valued benefits is its ability to collect

information in a non-destructive manner as it does not require direct contact with the scanned object (Nguyen-Do-Trong, 2018). In this project, hyperspectral data was used to classify the type of irrigation methods being implemented in blueberry crop areas, and determine water potential.

Our goal was to use hyperspectral imaging processes to associate reflectance measurements to a categorical response (irrigated or non-irrigated) as well as to water potential as a continuous variable, to inform mitigation options for water management in wild blueberry fields. We planned to achieve this through:

1. Collecting airborne data in Downeast Maine on an irrigated and non-irrigated field over three developmental stages.
2. Acquire water potential measurements of samples in each field.
3. Generate classified maps of irrigated/non-irrigated areas and water potential in order to determine locations of low or high water content.

The objectives aim to configure methods and processes that will allow for more convenient detection. The process of configuration involves combining spatial measurements, ground data, as well as approaches in technology, processing, and computation (Nelson, 2018). Using its ability to unobtrusively capture and process different absorption and reflectance rates, this method in imaging spectroscopy can be applied to Maine's wild blueberry industry, as well as its other agricultural sectors. We seek to classify water potential in wild blueberry fields as a means to offer new techniques in crop health monitoring.

1.2. Methods

1.2.1. Study site

Our research was conducted on commercial blueberry fields in Deblois, Maine owned by Jasper Wyman and Sons, a corporation specializing in frozen fruit. These crop fields tend to contain a number of different species clones, growing in sections within a particular field (Agriculture and Agri-Food Canada, 2008). Due to this pattern, a number of plots have been selected from the area in order to accurately assess the locations and tendencies of water potential. The site includes two separate fields, an irrigated and non-irrigated, each approximately 16 hectares in area. These two fields were selected to act as a comparison between two current water management practices in a single area. These fields serve other various research projects, but are not significantly impacted in relation to our goals.

1.2.2. Workflow overview

The project workflow is outlined in figure 1. Input data included the ground data (irrigated/non-irrigated classes and water potential) and manual digitizations from the imagery which trained the predictor models. Model process details are outlined in figure 3. The model was then applied to cube images for classification and prediction which were then used to verify the models.

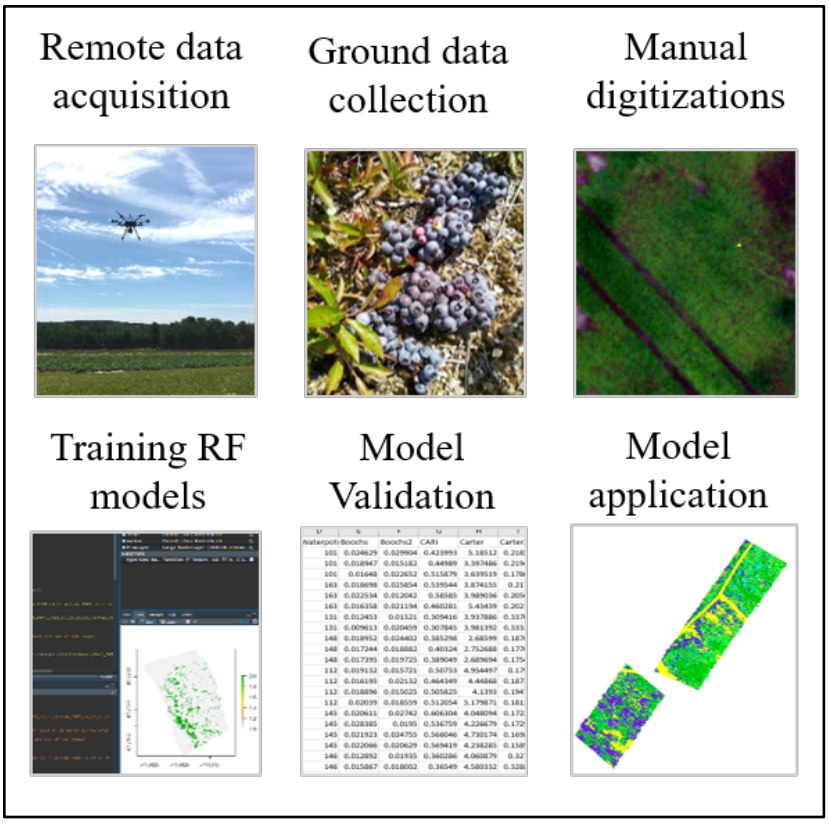

Figure 1.1. Broad process flow from data collection to model verification.

1.2.3. Image collection

Our imaging spectrometer is a Micro A-Series Sensor by Headwall Photonics. The device is attached to a DJI Matrice 600 Pro unoccupied aerial vehicle (UAV), and operates as a line-scanning (pushbroom) instrument. The sensor captures the visible and near-infrared portion of the electromagnetic spectrum from 400 nm to 1000 nm and collects 324 spectral bands. The number of flight lines, flying height, and flight speeds were determined on site size, takeoff distance from scanning area, land topography, and specific daylight conditions. Data collection with the UAV was conducted between 10:00AM and 2:00PM local time (period where sunlight is overhead in the northeastern region of the United States) to avoid shadows in the data. Image

collection was limited to these timeframes, battery capacity, and the amount of time that we had access to the field sites.

1.2.4. Ground sampling

Measurements were taken over the two fields three times in the 2019 spring and summer which were peak bloom (June 7), green fruit (July 3), and color break (July 25). These represent some of the development stages of wild blueberries throughout the summer growing season (Schilder, 2015). We chose to collect data over the different stages to determine temporal variation in water conditions and spectral response over the seasonal cycle of blueberry growth. Derived data could exhibit the characteristics in how water potential and intake varies throughout the season.

We used 20 ground sample measurements on each field that were selected by a random sampling design and on the basis of clone morphological distinction. The samples entailed a branch of blueberry leaves attached to the stem. These were gathered as the drone captured image data approximately between 10:00AM-2:00PM EDT. The samples were stored in plastic ziploc bags in coolers and measured approximately two hours later for water potential. The measurements were taken using a leaf pressure chamber in a climate controlled laboratory space in Wyman's facilities.

Other measurements conducted by the collaborating lab include total leaf weight, SPAD (measure of chlorophyll), anthocyanin, leaf area, the number of leaves on the sampled branch, dry leaf weight, plant height, height of winter damage, total number of buds, injured buds, soil water potential, soil temperature, if mummy berry disease (caused by fungal pathogen *Monilinia vaccinii-corymbosi*) (Annis, 2004) was present as well as severity, and any other notes such as

tip-dieback. One clone group (Baxter-B11) on the non-irrigated field experienced winter damage in half the area, leading the group to take two samples in order to detect any differences.

1.2.5. Image data & sampling

Imagery was processed using the Headwall application Spectral View. A white tarp taken in the imagery and a spectralon panel were both considered as the white reference for image processing, however the spectralon provided more accurate spectra reflectances and greater consistency. Processing entailed transforming imagery from its raw form, to radiance, reflectance, and then orthorectification. The ground sampling distance was approximately 10x10 cm per pixel.

Using the collected imagery, we delineated pixels of ground sampled blueberry leaves through the ENVI software program (version 5.5 64-bit). These delineations were used to extract pixels as image samples from the imagery data in the R programming environment.

Training pixels in the categorical model were digitized in large samples. With a binary response of irrigated and non-irrigated uninfluenced by the water potential ground reference data, we digitized large polygon areas within each field to gain a larger training size. Four samples from two cube images of each field were digitized for each stage. Figure 2 shows an example of the distinction between classification and regression model sampling.

The water potential digitizations were guided by the water potential ground reference samples. Four to eight pixels were digitized around the coordinates of where samples were collected. Pixels that were in clouded or shadowed images were discarded. Two samples were removed from the first scan date, and eight from both the second and third.

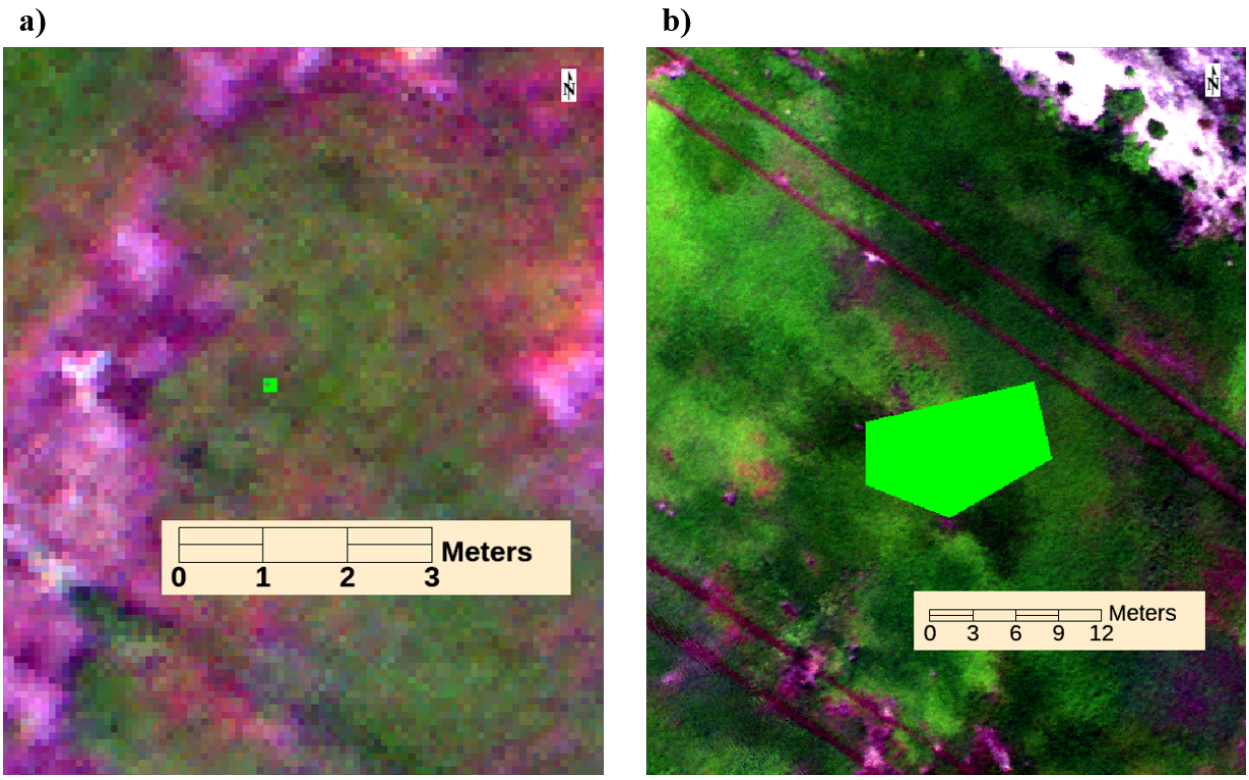

Figure 1.2. (a) Image of a digitized sample for a continuous response (water potential).

(b) Image of a digitized sample for a categorical response (irrigated/non-irrigated).

1.2.6. Model development

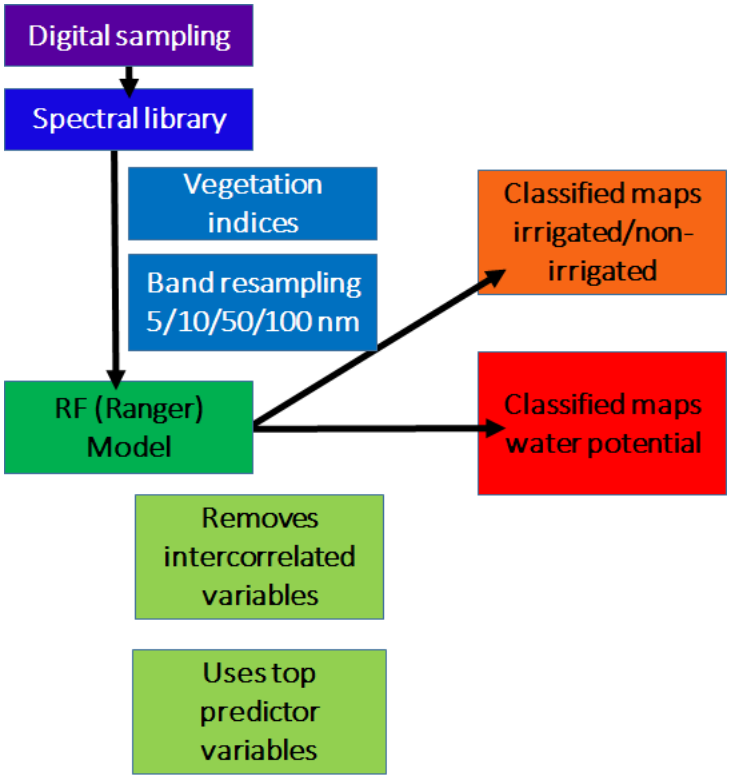

Figure 1.3. Model outline process beginning with image sampling and outputs of classified images of both irrigated/non-irrigated, and water potential.

Variables were calculated from reflectance values to use as predictors in our model which will be referred to as spectral indices. One of these methods in deriving the variables included resampling at 5, 10, 50, and 100 nm. These values were selected to show a range of resampling sizes and determine what significance they had, if any. The other method was calculating vegetation indices with the 'vegindex' function from package 'hsdar'. The resampling and vegetation indices totaled to 260 variables. A random forest model was trained with the calibration data using the 'Ranger' package to determine the relationship between the health classification to the predictors. Ranger was selected rather than the 'Random Forest' package due

to its functionality in handling large spatial data. The Ranger package was designed to improve statistical analysis using random forests with modest computing power (Wright, 2015).

In developing the model, our process produces accuracy rates at the different number of important variables utilized. We selected the model that utilized a lower number of variables to produce a lower error rate.

We developed models that used the extracted pixel samples and water potential measurements as inputs to classify unsampled areas of imagery (Maschler, 2018). The product of the models would be generated maps of irrigated or non-irrigated areas as well as water potential classifications. Models were built upon those developed by the Nelson lab (Nelson, 2020). We also planned to calculate vegetation health indices that may assist in understanding what particular stresses are taking place (Zarco-Tejada, 2018).

1.3. Results

1.3.1. Data overview

The data set used in this study includes 23 hectares of imagery over the irrigated field and 16 over the non-irrigated field for each of the three stages. This included 48 cube images for each scan stage with a spatial resolution of approximately 10 cm. Ground referenced water potential measurements for 20 samples on each field were provided by the collaborating team.

1.3.2. Model

The errors for removing intercorrelated predictor variables were calculated at 0.9, 0.93, 0.96, and 0.99 pair-wise absolute correlation cutoff levels. Removing those at a 0.99 cutoff produced the highest accuracy rates.

Numerous models were developed with the digitized samples used as training data. A model was created for each field stage with one for a binary classification (irrigated or non-irrigated), and one for water potential. Additionally, a model for water potential was trained with combined samples from all three stages. Table 1 outlines the model information and error rates.

Each model's sample size is the number of digitized pixels. The categorical models had an out-of-bag (OOB) error of 25.2% for the first stage, followed by similar error rates of around 17% for the last two stages. The stage-specific (or local) models for water potential had an R^2 ranging from 0.437 to 0.499. The global model utilized all pixel samples as training data and resulted in an error of 0.554. To further analyze, predictions were also conducted for all four of these water potential models.

Table 1.1. Model information used in image classification with each collection stage having a separate model for classification, and one global model for regression.

	Irrigated/Non-Irrigated Categorical			**Water Potential-local Continuous**			**Water Potential Global**
	Peak bloom	Green fruit	Color break	Peak bloom	Green fruit	Color break	
Sample size	139	115	99	139	115	99	353
Independent variables	25	25	5	25	25	25	20
OOB prediction error	25.2%	17.4%	17.2%	320	679	754	709
R squared (OOB)				0.115	0.458	0.437	0.554

1.3.3. Prediction

To further analyze the performance of the continuous models, predictions using calibration and validation samples were conducted. Each local water potential model used a split of 70% calibration and 30% validation from the total sample set, randomly chosen from each field. The split was chosen due to the small sample sizes with peak bloom having 139, green fruit 115, and color break 99. The global model prediction, however, used an 80% and 20% split of the total sample data. The sample size in the global model was three times that of the local models. This split ratio for calibration data was increased because it improved the model's R^2 while still maintaining a validation size of about 70. Table 2 outlines each model's OOB error, R^2, and calculated RMSE.

Table 1.2. Outline of global and local regression model statistics on predicting water potential.

	Local Regression Prediction			**Global Regression Prediction**
	Peak bloom	Green fruit	Color break	
OOB prediction error (MSE)	237	533	556	616
R squared (OOB)	0.554	0.563	0.564	0.617
Calculated RMSE	29.8	36.9	45.2	46.4

The table shows that all three local models had very similar R^2 values although calculated RMSEs were dissimilar and increased with each model. The global model had a comparable RMSE to the third local model but also had the highest R^2. As a result of the prediction, we decided to use only the global water potential model for our image classifications.

1.3.4. Variable importance

Each model was developed through particular predictor variables and a certain number of those top predictors. These predictor variables, or spectral derivatives, consist of vegetation indices and resampled bands. Models of differing numbers (in multiples of five) of top predictors were generated in the process along with accuracy rates, however the one with a lower number of variables with a comparable lower error rate was selected in order to improve efficiency but maintain efficacy.

Figure 4 is a plot of the 35 most important predictor variables with the importance levels while table 3 lists the top 20 predictors that were used in the model. Figures 5 and 6 show these variables for each model of local classification and regression. These plots show the diminishing importance of each variable and how despite there being a level of relevancy, inclusion of certain variables does not improve model accuracy. The local regression plots were included to show the similarities and differences among these, as well as in comparison to the global model.

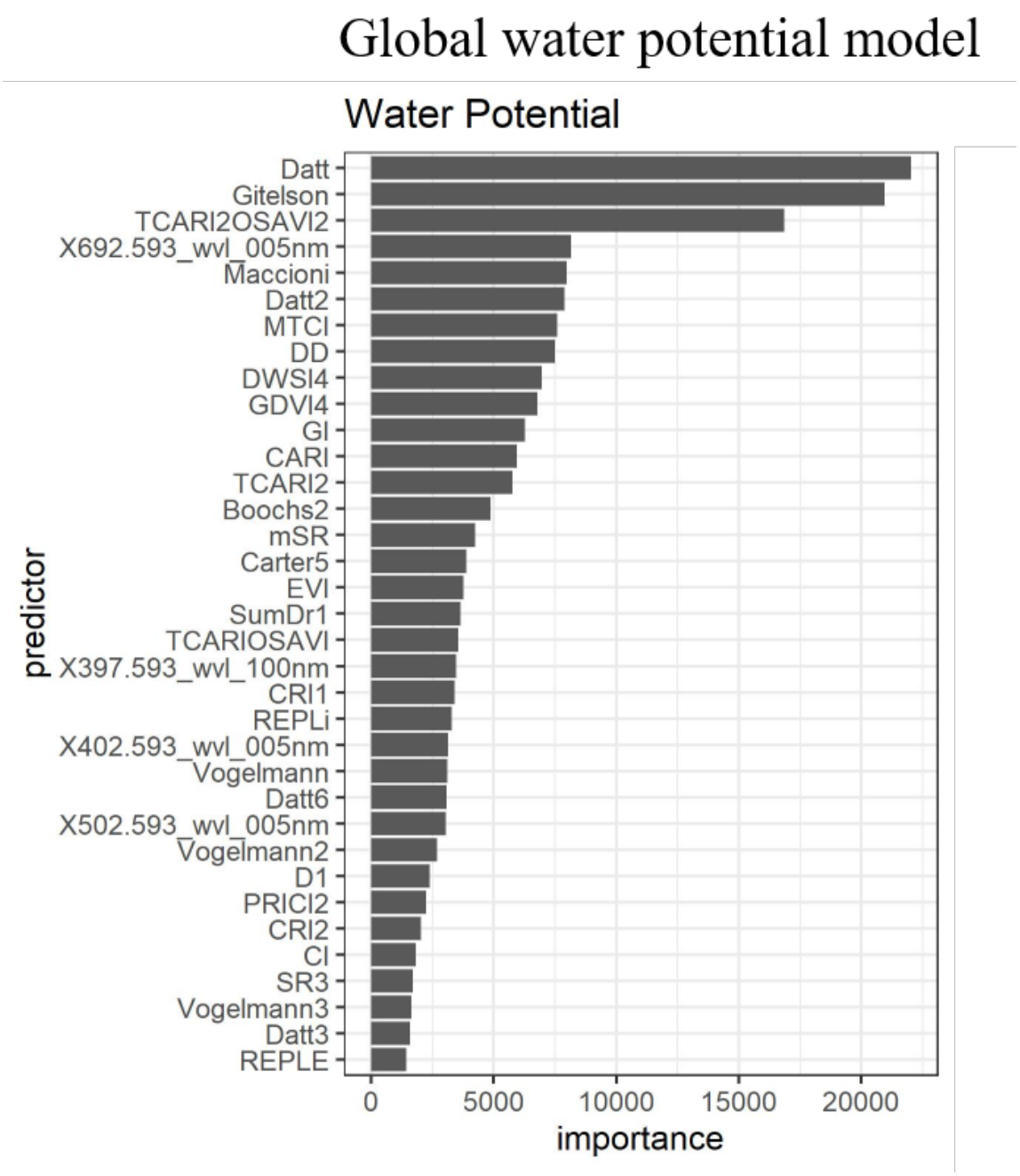

Figure 1.4. Displays the 35 most important predictor variables in the water potential model, the first 20 of which were included in the selected classifier, and corresponding importance level.

Table 1.3. Top 20 predictor variables used in the water potential model with description and band formula.

	Abbreviation	Name	Formula
1	Datt	'Chlorophyll & height'	$(R_{749} - R_{720}) - (R_{701} - R_{672})$
2	Gitelson	'Chlorophyll'	$1/R_{700}$
3	TCARI2OSAVI2	Transformed Chlorophyll Absorption Ratio 2 / Optimized Soil Adjusted Vegetation Index 2	$(3 * ((R_{750} - R_{705}) - 0.2 * (R_{750} - R_{550}) * (R_{750}/R_{705}))) / ((1 + 0.16) * (R_{750}-R_{705})/(R_{750} + R_{705} + 0.16))$
4	X692.593_wvl_005nm		'Bandpass 692.593 resampled at 5 nm'
5	Maccioni	'Chlorophyll'	$(R_{780} - R_{710})/(R_{780} - R_{680})$
6	Datt2	'Chlorophyll & height'	R_{850}/R_{710}
7	MTCI	MERIS Terrestrial Chlorophyll Index	$(R_{754} - R_{709})/(R_{709} - R_{681})$
8	DD	Double Difference Index	$(R_{749} - R_{720}) - (R_{701} - R_{672})$
9	DWSI4	Disease water stress index 4	R_{550}/R_{680}
10	GDVI4	Green Difference Vegetation Index 4	$(R^4_{800} - R^4_{680})/(R^4_{800} + R^4_{680})$
11	GI	Greenness Index	R_{554}/R_{677}
12	CARI	Chlorophyll Absorption Ration Index	$R_{700} * abs(a * {}_{670} + R_{670} + b)/R_{670} * (\alpha^2 + 1)^{0.5}$ $\alpha = (R_{700} - R_{550})/150$ $b = R_{550} - (550 * \alpha)$
13	TCARI2	Transformed Chlorophyll Absorption Ratio 2	(3 * ((R750 - R705) - 0.2 * (R750 -R550) * (R750/R705)))
14	Boochs2	Single Band 703 Boochs	D_{703}
15	mSR	modified Simple Ratio	$(R_{800}–R_{445})/(R_{680}–R_{445})$
16	Ctr5	Carter 5	R_{695}/R_{670}
17	EVI	Enhanced Vegetation Index	$2.5 * ((R_{800} - R_{670})/(R_{800} - (6 * R_{670}) - (7.5 * R_{475}) + 1))$
18	SumDr1	'LAI & % green cover'	$\sum_{i=626}^{795} D1i$
19	TCARIOSAVI	Transformed Chlorophyll Absorption Ratio / Optimized Soil Adjusted Vegetation Index	$(3 * ((R_{700} - R_{670}) - 0.2 * (R_{700} - R_{550}) * (R_{700}/R_{670}))) / ((1 + 0.16) * (R_{800}-R_{670})/(R_{800} + R_{670} + 0.16))$
20	X397.593_wvl_100_nm		'Bandpass 397.593 resampled at 100 nm'

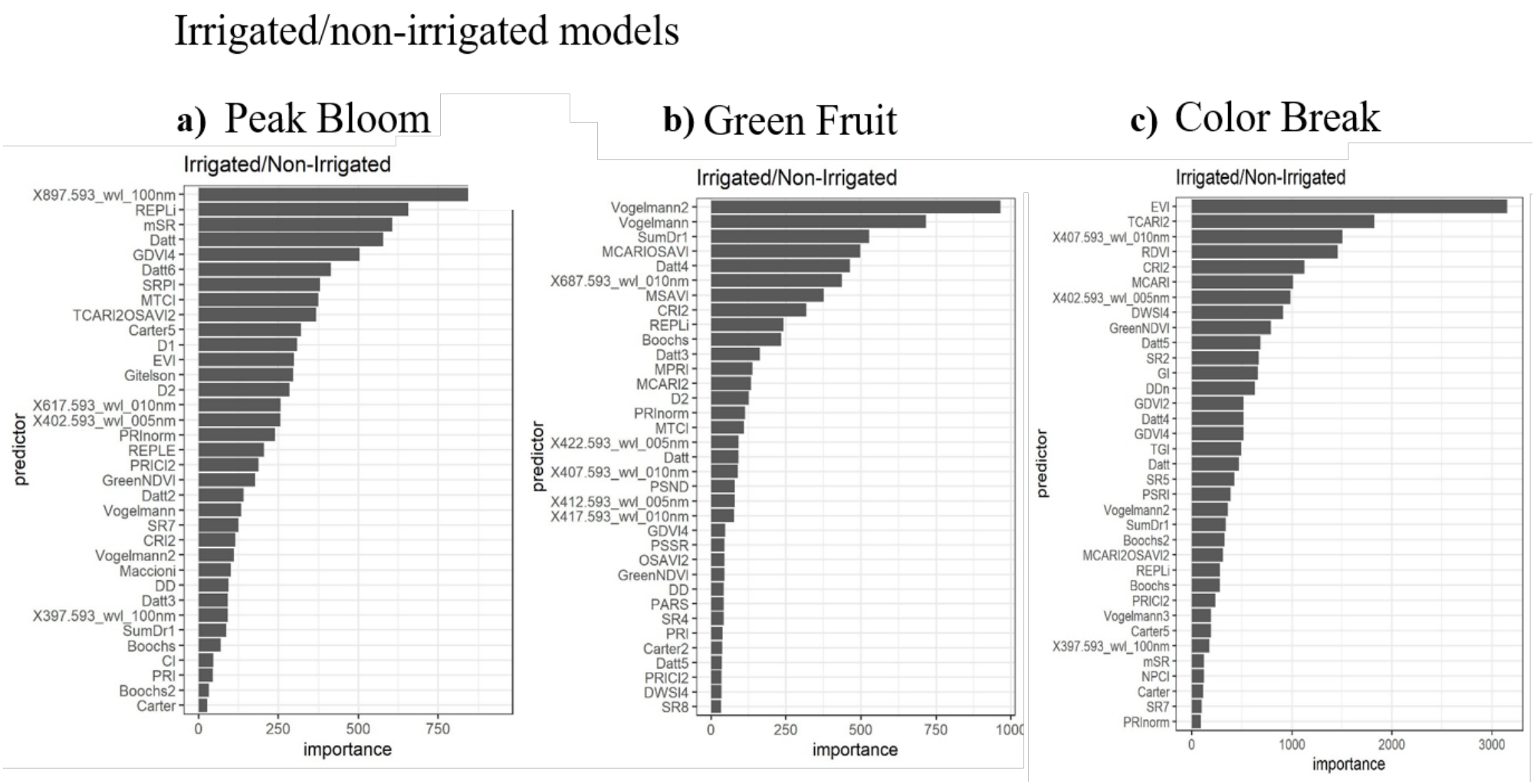

Figure 1.5. Plot of the 35 most important predictor variables in the local models and corresponding importance levels.

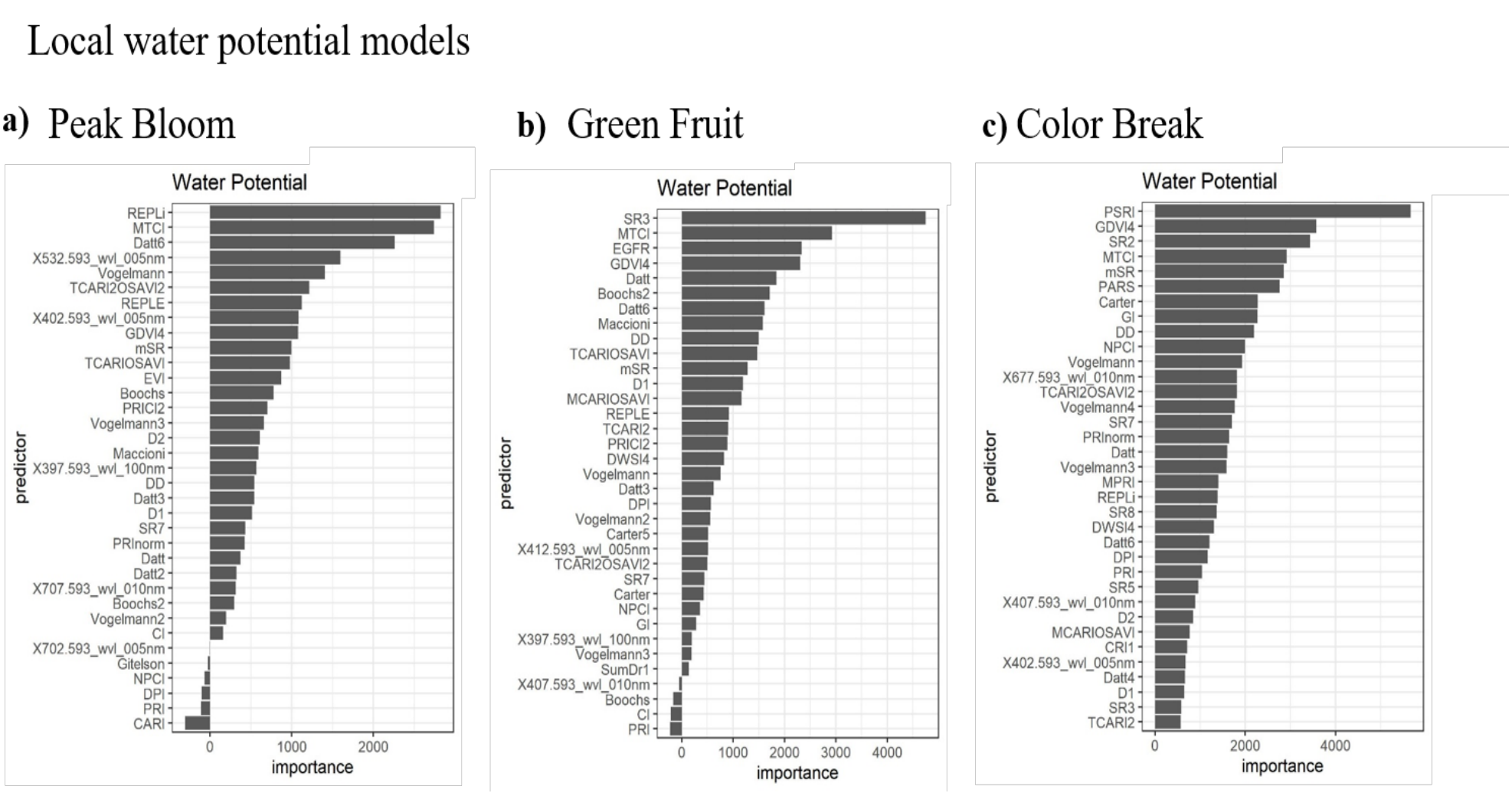

Figure 1.6. Plot of the 35 most important predictor variables in the water potential models and corresponding importance level.

Most top predictor variables for each of the models focus on assessing chlorophyll levels. Some however include elements of soil adjustment such as TCARI2OSAVI2 and REPLi. Most of the vegetation indices and relevant bands range between 650 nm to 800 nm on the electromagnetic spectrum, or in the visible to near infrared region.

The top predictors of the 3 categorical models showed very distinct variable selection results with no model containing a shared top five variable. The first shared predictor was CRI2 as the fifth for color break and the eighth of green fruit. This predictor was the 24th for peak bloom. Green fruit and color break had predictors focused on chlorophyll while the top predictor of peak bloom was bandpass X897.593 resampled at 100 nm. The region in the far near infrared, or around 900 nm to 1000 nm, are understood to be indicative of water content levels in vegetation (Roberts, 2016), however, these findings may not have been as evident in this study due to omittance of that region as a result of noise. The second predictor was REPLi, used in monitoring wheat canopies using biomass, water content, and underlying soil characteristics (Baret, 1988). The distinction between predictor variables in the local models of classifying irrigated and non-irrigated may be a result of the variation in growing patterns that could be less of a factor in assessing water potential.

In all three individual models for water potential, MTCI was a high predictor. Medium Resolution Imaging Spectrometer (MERIS) terrestrial chlorophyll index measures chlorophyll in the red edge region distinguishing it from other red edge position indices by its sensitivity to higher levels of chlorophyll (Dash, 2004). Overall, predictors in the local water potential models can be characterized as less variable than those of the categorical models, and can also be compared against the global water potential model.

The top five predictors in the global model all involve bands in the 700 nm to 800 nm region, a range associated with chlorophyll measures. The top predictor being the Datt index, is one that was developed to estimate chlorophyll content, particularly in higher plants or tree canopies (Datt, 1999). The Giltelson index is a measure of chlorophyll fluorescence which is proportional to actual chlorophyll content (Gitelson, 1999). The third most important variable was TOCARI2OSAVI2 which incorporates soil adjustment in chlorophyll measurements and has been recommended for agricultural applications (Rondeaux, 1996). These top three predictors had a significant level of importance over the following ones, whereas the three local models had a more even dispersal over the top variables. The most important global indices are likely a result of the consolidation of all three collection stages. Peak bloom, green fruit, and color break were different stages of blueberry development and the integration of all three results in predictors that will account for the variability. The model predictors are composed of reflectance values at different portions of the electromagnetic spectrum that can also be represented by the spectral signatures.

1.3.5. Spectral signatures

Various spectra of blueberry samples and pixels were analyzed and was mainly done so to assess radiometric calibration. The critical analysis, however, was using these spectra in development of the spectral variables, being the calculations of the vegetation indices and band resampling values. Nonetheless, the signatures can showcase information of interest. Figures 7, 8, and 9 plot the signatures of four blueberry samples (two from the irrigated field, and two from the non-irrigated), over each development stage which are plotted in different combinations.

As can be seen in figure 8 (a), the spectra of the irrigated field tend to show higher reflectances across all wavelengths as the season progresses, but is much more evident between the first and last two stages. The non-irrigated field can be seen in figure 8 (b), as having less of a discrepancy between all three stages, however there is fluctuation between the last two stages where green fruit at times has higher reflectance rates than color break. It can be seen in figure 9, that generally for each stage, the irrigated spectra had lower reflectance rates than that of the non-irrigated although there are points in the first stage where the spectra of the irrigated field surpass the non-irrigated. These sample spectra display the collected data in a raw form, but the transformations of these is what the model is developed from.

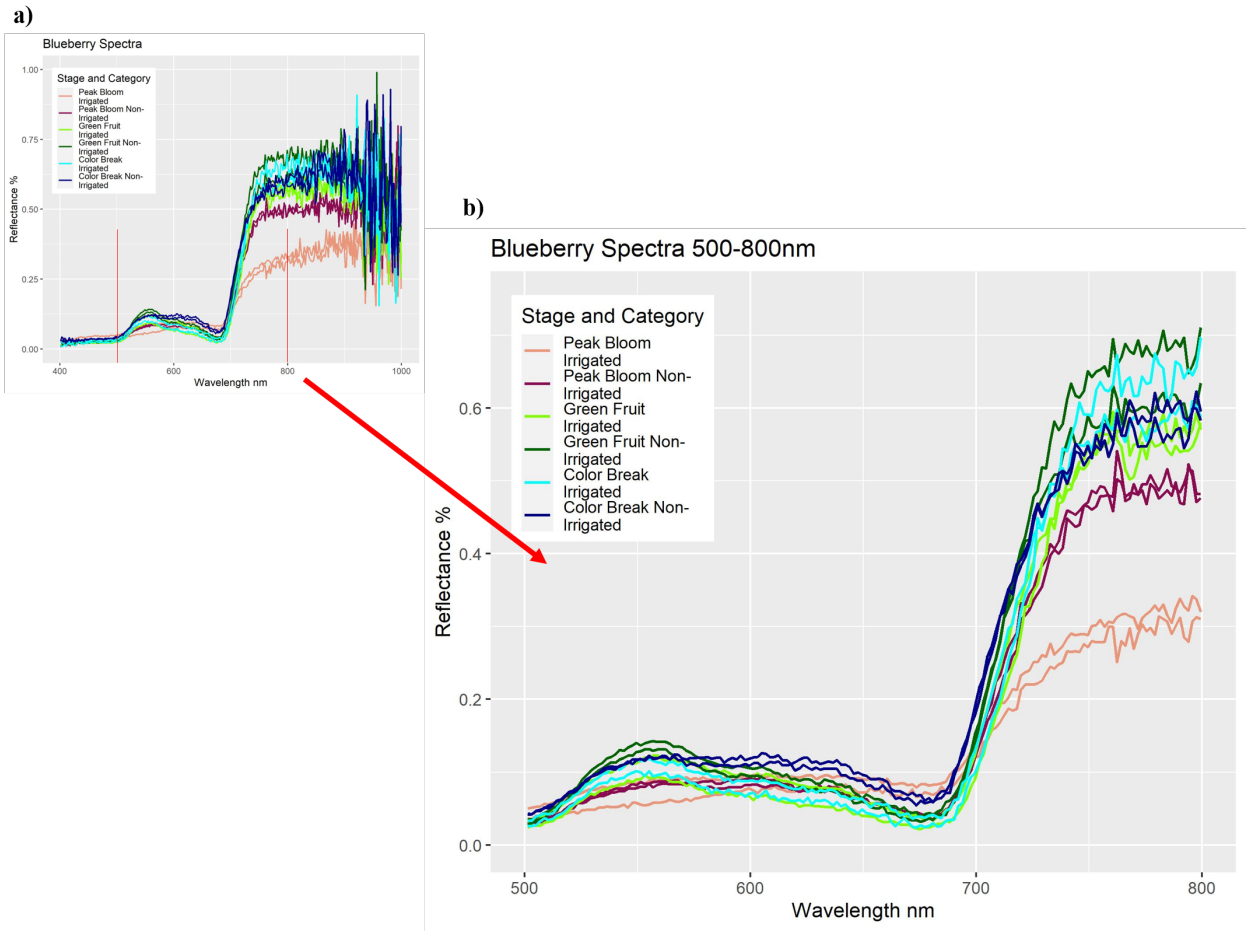

Figure 1.7. (a) 12 spectral plots of four blueberry samples from all three stages. **(b)** The spectrum focused on 500-800 nm wavelengths.

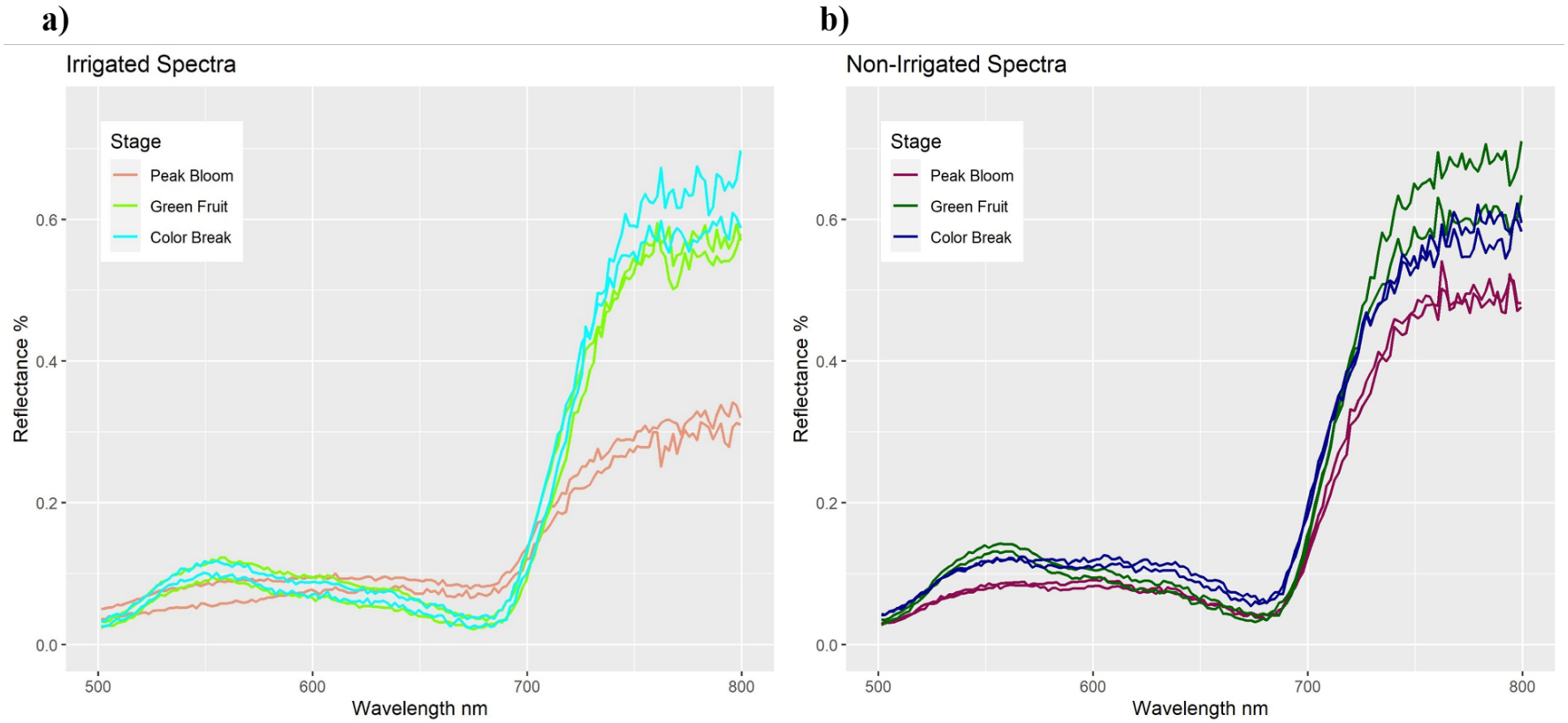

Figure 1.8. (a) Spectral plots of six blueberry samples of the irrigated field from all three stages. **(b)** Spectral plots of six blueberry samples of the non-irrigated field from all three stages.

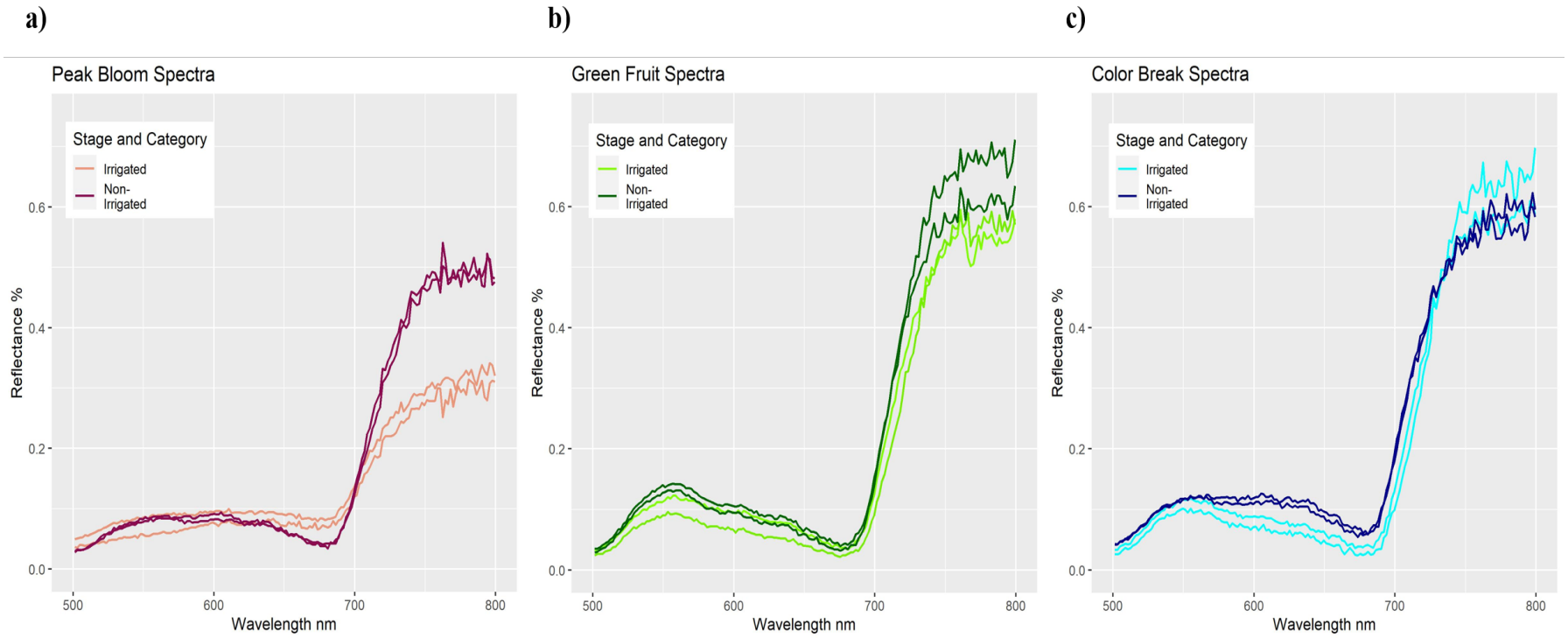

Figure 1.9. (a) Spectral plots of four blueberry samples from peak bloom. **(b)** Spectral plots of the same samples from green fruit. **(c)** Spectral plots of the same samples from color break.

1.3.6. Classification

Cube images of the same area that were well-illuminated for all three stages were classified for analysis which amounted to 12 total. Most of the images were not classified due to time constraints and cloud shadows producing poor illumination. Though many are functional, they were not used for this study.

Figures 10, 11, and 12 each represent a separate stage, or date, the image was taken with an irrigated image on top, and non-irrigated image on bottom. They first display a true color image, followed by a classified image being irrigated or non-irrigated, and then a predicted image of water potential. The cube images from the different stages represent the same area and have overlapping extents, but they are not exact.

All categorical images were classified accurately in that a majority of each image was classified as it actually was. The water potential models show a large portion of the irrigated field with lower water potential throughout all three stages, and the non-irrigated field with more sections of higher water potential.

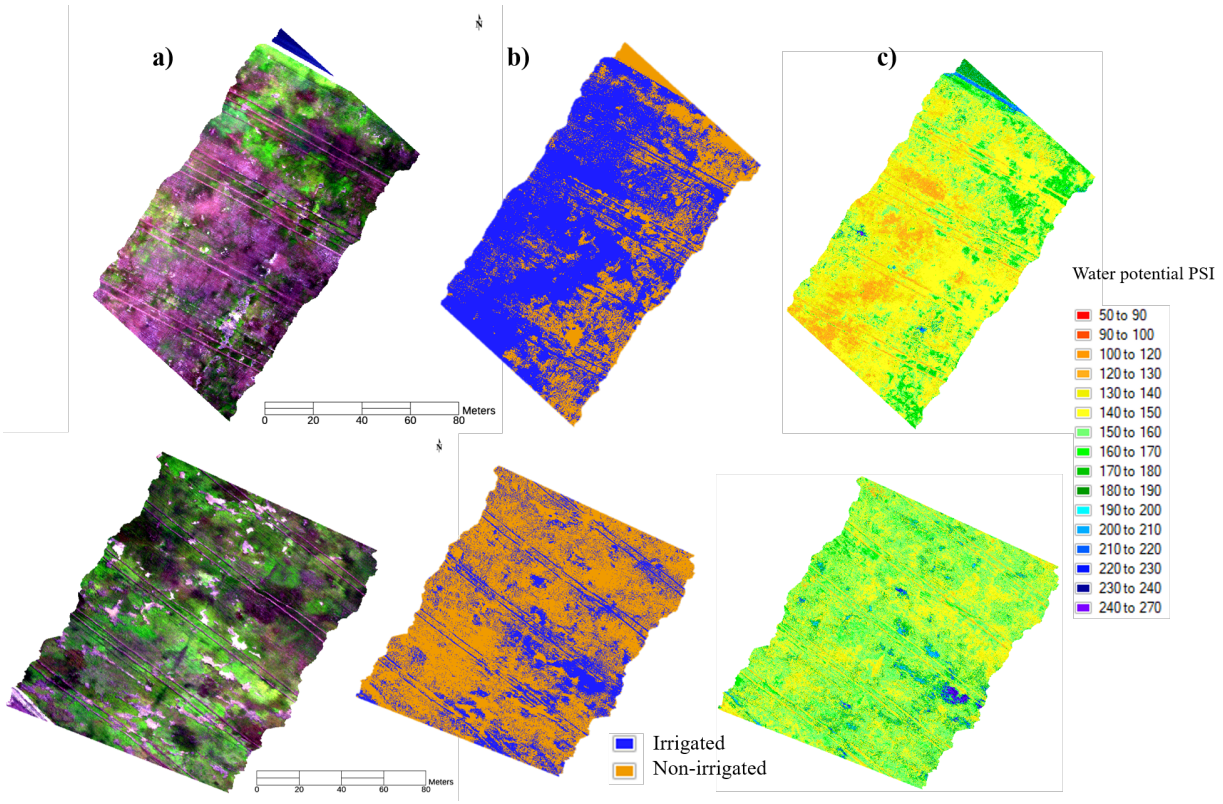

Figure 1.10. Images of processed data cubes over peak bloom **(a)** True color image of area. **(b)** Classified data cubes of irrigated and non-irrigated areas. **(c)** Processed data cubes of water potential.

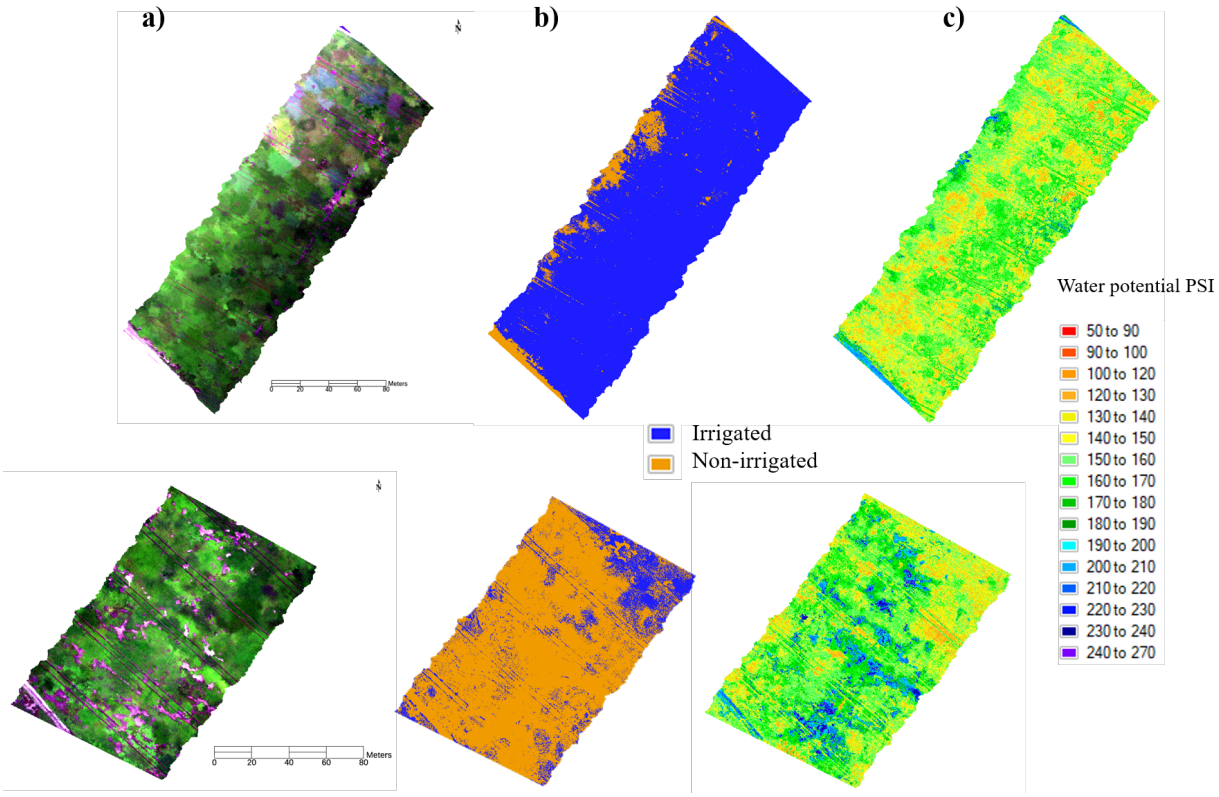

Figure 1.11. Images of processed data cubes over green fruit **(a)** True color image of area. **(b)** Classified data cubes of irrigated and non-irrigated areas. **(c)** Processed data cubes of water potential.

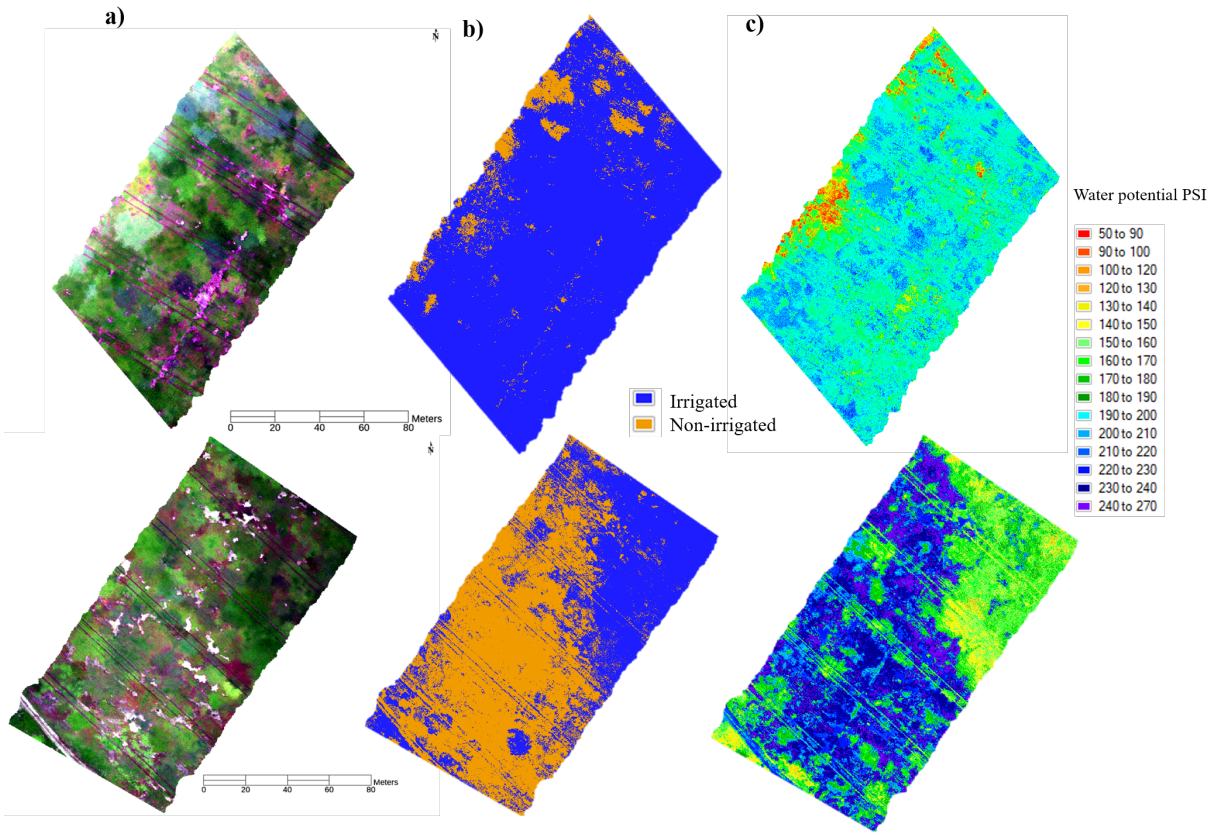

Figure 1.12. Images of processed data cubes over color break **(a)** True color image of area. **(b)** Classified data cubes of irrigated and non-irrigated areas. **(c)** Processed data cubes of water potential.

1.4. Discussion

This study utilized machine learning and remote imaging spectroscopy to predict water stress in wild blueberry fields. The spectral indices produced using the Ranger package had shown effectual in classifying hyperspectral cube images, where the indices used had logical premise and emphasized the sensitivity spectral bands hold through the index transformations. The results presented similarities and differences within our models and cast light on the methods that can produce the desired outcome.

The results displayed temporal distinctions within the blueberry development stages. The model statistics and other outputs showed contrasts particularly between the first stage and the next two. These differences are supported in that the peak bloom stage was more distinct in phenology from the other two, where leaf area was much smaller and ground soil still exposed. This may be a factor in the soil adjusted indices that were important in the first blueberry scanning. Dissimilarly, the other two stages had much more vegetation coverage. With this, the water potential classification showed similar between the first and second stage. This distinguishes the processes in how the predictive models perform.

The categorical models classified all processed data cubes accurately. All irrigated images were mainly classified as that, and the same result occurred with the non-irrigated. The predictor variables for the models were different from one another most likely as a result of it being local and used only samples from the day it was taken. The water potential models, however, used the ground sampled water potential values from all collection days. These factors produce different results and were evident in the predicted images where water potential was shown as higher in all non-irrigated fields. This raises notice to what information is most beneficial.

An objective of this study was to associate reflectance measurements with a categorical treatment (irrigated or non-irrigated) in order to make predictions. This was achieved, however it prompts questioning in how constructive the answer actually is. The local models were separate for the irrigated and non-irrigated classification due to the relativity in deeming a field irrigated or not irrigated throughout a changing season. What is classified as irrigated at the beginning of the season may not be classified the same towards the end. The irrigated/non-irrigated is a more

convenient method in classification, but finding the utility in it may be the challenge. A more objective approach is predicting and calculating water potential which provides greater value.

Measuring water potential and how it would impact irrigation practices is a more appropriate pursuit in providing farmers and landowners greater utility. The global water potential model had an R^2 of 0.62 against the validation dataset. Adding more samples in water potential would assist in predictions and from more than a few select dates. Another addition in strengthening the model and prediction process would be including other types of variables from other functions or calculations. Spectral band derivatives and extensions of those, another common method in determining characteristics of vegetation (Thorp, 2017), were not heavily incorporated in this process aside from a few vegetation indices that include derivatives in the index formula. It may also be helpful in reducing the resampling predictors as it may not have had a significant effect on the predictors, especially as the number of band resampling increased.

1.5. Conclusion

The goal of this project was to use hyperspectral imaging processes in detecting measurements of water content practically for the benefit of the agricultural industry. This was mainly achieved through the development of our models, although there are a number of actions that could strengthen the outcomes. Models will be improved if larger training datasets are used. This is particularly relevant if other blueberry fields or varying crops are to be measured, such as fields in other areas. The process of streamlining by means of data collection, processing, and efficient programming, is also critical to provide timely and accurate data that can be practically used by industry workers, which would require greater computing capacity as well as data

management. With continued input additions and modifications, our methods can assist in improved agricultural practices.

Maine's economy is heavily based on its natural resources and agricultural industries with the blueberry industry being one of the top five revenue-generating crops. New technologies and methods such as those pertaining to precision agriculture are becoming more widely used (McBratney, 2015), necessitating the adoption of new methods in order to maintain standards for a competitive economy. With greater sampling and in-depth studies, the application of hyperspectral imaging methods is becoming a more viable option.

CHAPTER 2

AIRBORNE HYPERSPECTRAL DATA APPLICATION IN STRESS DETECTION OF ASH TREES

2.1. Introduction

Ash trees (*Fraxinus* spp.) have a commercial value of 320 million dollars in Maine and are used in the production of snowshoes, canoe paddles, baseball bats, and other products. Black ash plays a large cultural role in Wabanaki folklore as well as the nation's use in basket-weaving (Blackmore, 2019; Neuman, 2010; Ritter, 2017). The tree species however is under threat by the emerald ash borer (*Agrilus Planipennis Fairmaire*) (EAB), an invasive species from Asia. EAB is presumed to have migrated to the United States through wood packaging or shipping crates **(**Canadian Food Inspection Agency, 2014; Siegert, 2014). The insect bores the ash tree, oftentimes preferring those that are in preexisting stress conditions, and lays its eggs in bark crevices. When the larvae emerge, they feed on the tree's sapwood depleting the tree's water supply and forming serpentine galleries which can girdle the tree with increased EAB population density (McCullough, 2017; Wang, 2010). The spread occurs through adult EAB flight of up to 1.6 km but additionally through infested firewood transported by humans unknowingly (Cappaert, 2005). The beetle was detected for the first time in the state of Maine in both southern and northern regions in 2018 (Sampson, 2018), prompting the state to take action in quarantining and monitoring.

Understanding the spatial distribution and respective health conditions of ash trees is critical in early EAB detection. Infestations lead to mortality within three to five years and the retardation of this process can stop the spread with swift action (Sampson, 2018). Climate

change will also contribute to unpredictable and variable weather patterns, producing unforeseen disease and pest outbreaks (Lawrence & Labus, 2003). This constitutes a greater need for preparedness and innovative solutions in this sector. With threats from pathogens, hyperspectral imaging can detect areas of infestation and assist in preventing dissemination.

The application and use of hyperspectral imaging technology has made great advances in the past decade. Hyperspectral imaging is a method in detecting and classifying objects or conditions based on the light reflectance rate (or spectral signature) on the electromagnetic spectrum. Objects project particular signature values at different wavelengths based on their composition (Adão, 2017). Its current functions range from detecting human tissue damage (Leavesley, 2016) to aiding agricultural quality control (Nguyen-Do-Trong, 2018). One of its most valued benefits is its ability to collect information in a non-destructive manner as it does not require direct contact with the scanned object (Nguyen-Do-Trong, 2018). In this project, hyperspectral data was used to detect stress in ash trees, a rapidly declining resource in the United States (Tallamy, 2016).

Knowledge of EAB and its migration patterns are relatively well understood, and preventative measures have and continue to be taken by state and local municipalities. Nonetheless, having the ability in measuring stress of ash quickly and with high spatial capacity is critical in monitoring and conducting action. Remote sensing technology provides a targeted advantage in addressing these objectives. Hyperspectral research has been applied to vegetation stress, however the capability in determining specific disease or pest infestation is still a further endeavor (Lawrence & Labus, 2003). This project aims to link health classifications of ash trees impacted by EAB to airborne hyperspectral data, investigating the potential remote imaging spectroscopy holds in identifying targeted information.

Our overall goal was to use hyperspectral imaging processes to classify stress while associating the characteristics to emerald ash borer infestation to inform mitigation options for Maine forest management. We planned to achieve this through:

1. Collecting airborne and ground measurements over infested sites in New Hampshire.
2. Sample ash trees within those sites including health classifications encompassing the classification spectrum.
3. Generate predictor maps of ash health in order to determine locations of infestation, susceptible areas of infestation, and potential solutions.

These objectives seek to configure methods and processes that will make detection more efficient and reliable. The process of configuration involves combining spatial and spectral measurements, ground data, as well as new approaches in technology, processing, and computation (Nelson, 2018). Using its ability to unobtrusively capture and process different absorption and reflectance rates, hyperspectral imaging can be applied to Maine's ash trees.

2.2. Methods

2.2.1. Study sites

Our study sites include three infested areas in southern New Hampshire where EAB was detected in 2013 (Kovacs, 2010). The areas include ash trees of varying stages of infestation and stress levels. We used these sites to conduct our study across the health spectrum for classification. Having representation of different health classifications from infestation can also

show how EAB progression develops, especially in the early stages, potentially offering new techniques in monitoring and early detection.

2.2.2. Workflow overview

The project workflow is outlined in figure 1. Input data included the ground data (health classes) and manual digitizations from the imagery which trained the predictor models. Model process details are outlined in figure 2. The classifier was verified through a validation test set. The model was then applied to cube images for classification, and further buffered and extracted.

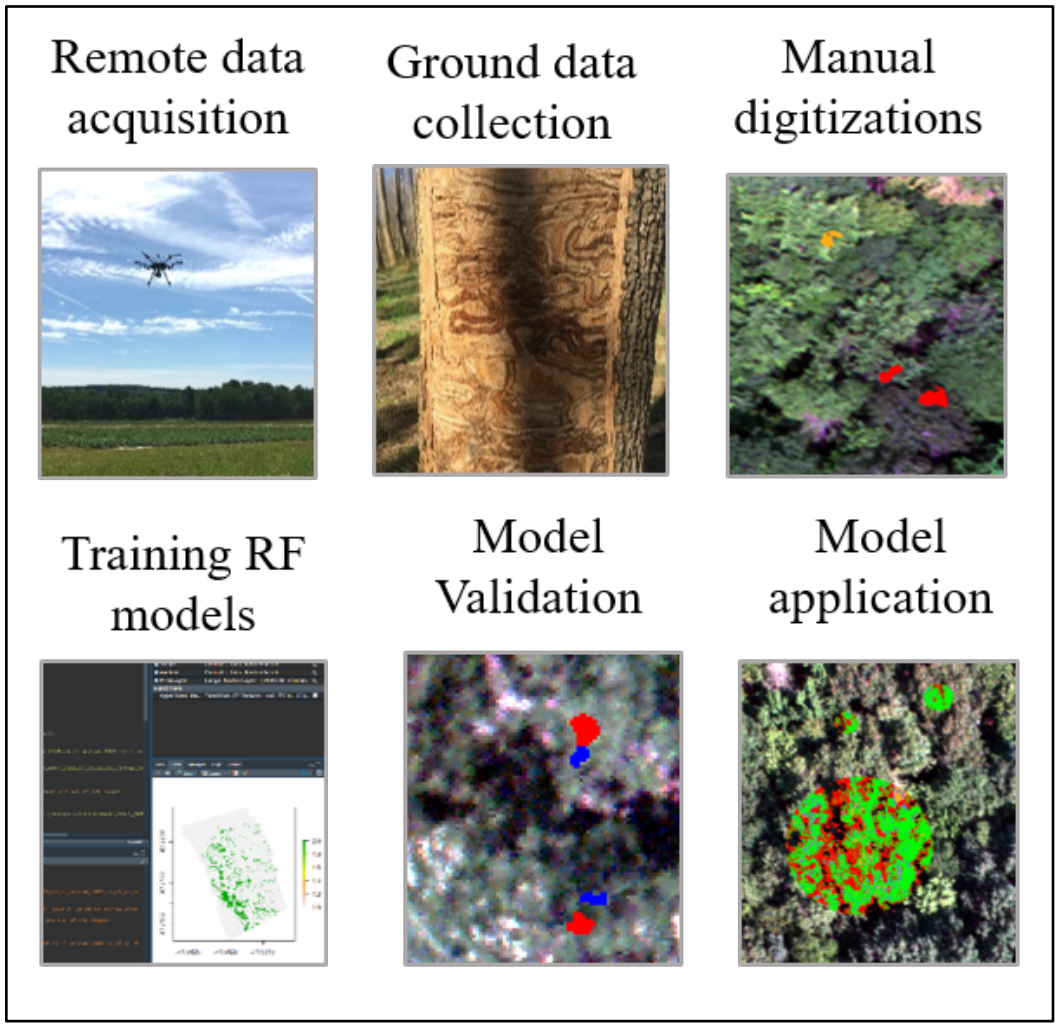

Figure 2.1. Broad process flow from data collection to model verification.

2.2.3. Image collection

Our imaging spectrometer is a Micro A-Series Sensor by Headwall Photonics. The device is attached to a DJI Matrice 600 Pro unoccupied aerial vehicle (UAV), and operates as a line-scanning (pushbroom) instrument. The sensor captures the visible and near-infrared portion of the electromagnetic spectrum from 400 nm to 1000 nm and collects 324 spectral bands. The number of flight lines, flying height, and flight speeds were determined on site size, takeoff distance from scanning area, land topography, and specific daylight conditions. Data collection with the UAV was conducted on July 26, 27, and 28 of 2019, between 10:00AM and 2:00PM local time (period where sunlight is overhead in the northeastern region of the United States) to avoid shadows in the data. Image collection was limited to these timeframes, battery capacity, and the amount of time that we had access to the field sites.

2.2.4. Ground sampling

Ground sample measurements of individual ash trees were taken in November 2019. These include the health classification ratings (visually determined), crown width, diameter at breast height (DBH), and height. We used the health assessment approach reported by Pontius et al. (2008). This procedure was developed through various tree decline rating systems. It entails crown vigor ratings of 1-5 where 1=healthy (no major branch morality, 2=light decline (10-25% of crown damaged), 3=moderate decline (26-50% of crown damaged), 4=severe decline (>50% of crown damaged), and 5=dead (Pontius, 2008). Our study targeted all ash species but our sampling included only white (*Fraxinus americana*) and green (*Fraxinus pennsylvanica*). Coordinate locations of alternative species in the surrounding areas of sampled trees were gathered as reference for canopy cover in the acquired imagery.

2.2.5. Image data & sampling

Imagery was processed using the Headwall application Spectral View. A white tarp taken in the imagery and a spectralon panel were both considered as the white reference for image processing, however the spectralon provided more accurate spectra reflectances and greater consistency. Processing entailed transforming imagery from its raw form, to radiance, reflectance, and then orthorectification.

Using the collected imagery, we delineated the crown pixels of ground sampled trees through the ENVI software program (version 5.5 64-bit). These delineations were used to extract illuminated pixels as image samples from the imagery in the R programming environment. Additional sunlit pixels from the same crowns were also delineated to use as validation data. Although vigor ratings were assigned to trees in November during the leaf-off season while the digitized samples were taken of leaf-on crowns, we determined that this would not impose any issues. We assigned appropriate classes to the leaf-off ash and the comparison of the two types of sampling information would be of the same tree.

2.2.6. Model development

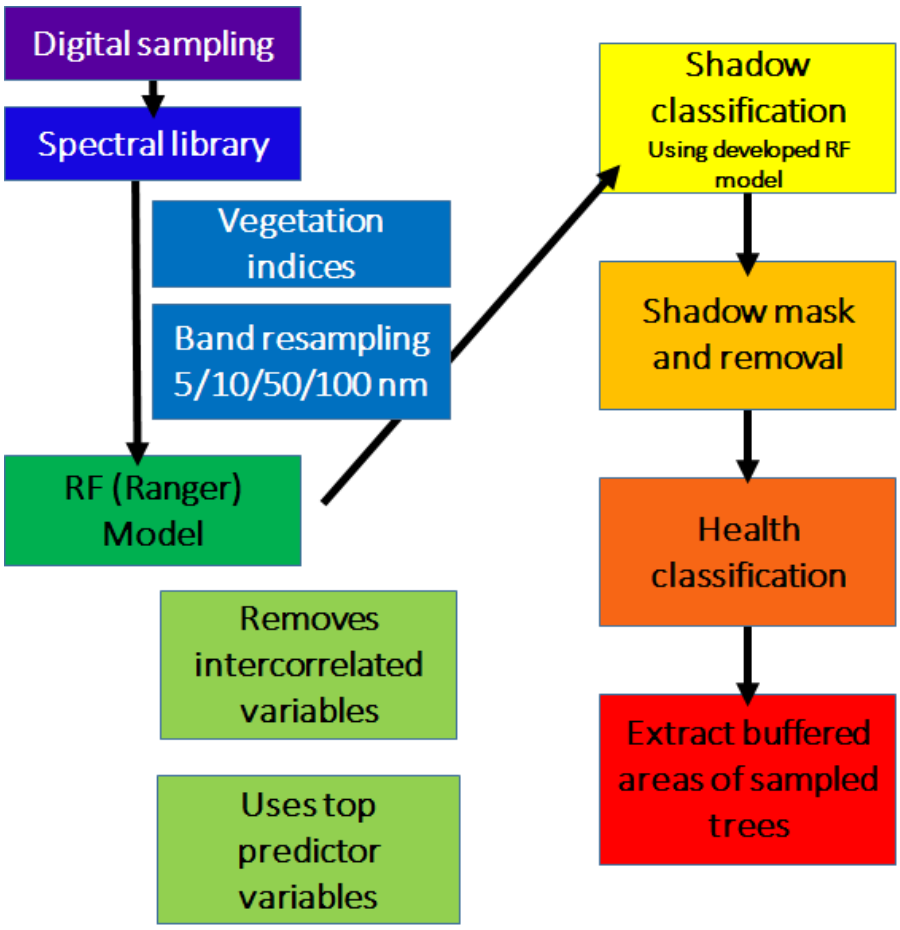

Figure 2.2. Model outline process beginning with image sampling and outputs of classified pixels of sampled trees.

Variables were calculated from reflectance values to use as predictors in our model which will be referred to as spectral indices. One of these methods in deriving the variables included resampling at 5, 10, 50, and 100 nm. These values were selected to show a range of resampling sizes and determine what significance they had, if any. The other method was calculating vegetation indices with the 'vegindex' function from package 'hsdar'. The resampling and vegetation indices totaled to 260 variables. A random forest model was trained with the

calibration data using the Ranger package to determine the relationship between the health classification to the predictors. Ranger was selected rather than the Random Forest package due to its functionality in handling large spatial data. The Ranger package was designed to improve statistical analysis using random forests with modest computing power (Wright, 2015).

In developing the model, our process produces accuracy rates at the different number of important variables utilized. We selected the model that utilized a lower number of variables to produce a lower error rate.

Our aim was to develop models that would use the extracted pixel samples, health ratings, and ground scans as inputs to classify unsampled areas of imagery (Maschler, 2018). The product of these models would be classified crowns of predicted ash tree health based on the five vigor levels. Models were built upon existing frameworks developed through the Nelson lab (Nelson 2020). We also calculated vegetation health indices that may assist in understanding what specific stresses are taking place (Zarco-Tejada, 2018). We also further differentiated trees in declining health by causes due to either EAB infestation or other stress factors.

2.2.7. Masking

Our models were applied to whole data cube images, although buffers that contain the training pixels were ultimately extracted from the classified images. This was done to avoid errors in classification processing using nearest neighbors methods. The first model classified areas of shadows which were then masked out from the original data cube. Shadow was differentiated as it is the only category type that would be present in a buffered crown sample. Efforts in distinguishing other types such as water, grass, shrub, soil, rock as well as between tree

species were developed and considered, however forgone due to time constraints and some computing power.

2.2.8. Buffers

After applying the shadow predictor, mask, and then health classification, we extracted the buffer samples. The buffer was applied to each ground sampled tree and given a diameter of the ground measured crown width which was measured at the maximum length. This may have resulted in buffer areas larger than the actual tree crown and encompassed other trees. The extractions were performed in the R environment similar to the extraction process of the digitized samples.

2.2.9. Model verification

Our crown buffers were verified with an accuracy assessment. For each crown, the number of pixels correctly classified was divided by the total number of pixels in that crown. Another prediction test was conducted using the validation pixels. The model was applied on the validation data and assessed on accuracy. A confusion matrix was generated for each correctly predicted class.

2.3. Results

2.3.1. Data overview

Imagery was collected over the 3 sites and covered approximately 152,827 m^2 of ground area. This included 67 cube images with a spatial resolution of approximately 10x10 cm. A total of 60 ash trees were sampled, and coordinates of 31 other tree species were taken as reference.

No analysis on the results in differences between green and white ash was conducted and though these contrasts may be present, understanding these differences was not a goal of this study. Table 4 outlines the number of ash trees sampled, each tree id, and pixel classification accuracy.

2.3.2. Model

A model categorizing different types of cover was originally created and tested. These included the referenced trees, ash trees, shadow, shrub, grass, water, rock, and soil. The results however were flawed for certain categories, particularly involving the discrimination among trees, shrub, and grass. Differentiating tree species also proved to be a challenge. We therefore applied the model to the entire data cubes where a shadow mask was then applied. The ash health model was then run over the masked data and then further buffered around the sampled ash tree points.

The shadow model classified pixels that were either considered “shadow” or “other”. The input training data however consisted only of digitized shadow samples and ash tree pixels. The last step of the process would be to extract the buffers of sampled trees which would only contain shadow or ash, therefore, other categories were not included in the shadow mask process.

The selected health classification model uses the twenty most important variables and has a prediction error of 8.89%. The shadow model uses the ten most important variables and had a prediction error of 0.05%. The model information can be viewed in table 1.

Table 2.1. Model information for shadow and health classification.

Ash Tree Model Information	
Shadow	Health
Predictions: 9936 Num.trees: 1000 Number of independent variables: 10 Mtry: 3 Min.node.size: 1 Prediction.error: 0.0005032206 Class values: 1 2 Tree type: classification Splitrule: gini	Predictions: 6941 Num.trees: 1000 Number of independent variables: 20 Mtry: 4 Min.node.size: 1 Prediction.error: 0.08889209 Class values: 2 4 3 1 Tree type: classification Splitrule: gini

2.3.3. Variable importance

Each model was developed through particular predictor variables and a certain number of those top predictors. These predictor variables, or spectral derivatives, consist of vegetation indices and resampled bands. Models of differing numbers (in multiples of five) of top predictors were generated in the process along with accuracy rates, however the one with a lower number of variables with a comparable lower error rate was selected in order to improve efficiency but maintain efficacy.

Figure 3 is a plot of the 35 most important predictor variables with the importance levels for the shadow model while table 2 lists the top ten predictors that were used in the model. Figure 4 and table 3 show the same information on the health classification model and used the twenty most important variables. The importance plots show the diminishing utility of each variable and how despite there being a level of relevancy, inclusion of certain variables does not improve model accuracy.

Predictors between the shadow and health classifier are similar although the shadow model utilizes less variables. Figures 3 and 4 also show that the shadow variable importance levels have a steady importance decline, whereas in the health model, the top 2 variables have very high importance, followed by a noticeable level drop in the third and fourth. This shows that the health model was much more influenced by particular predictors. For both models, predictors of importance mainly encompass wavelengths within 650 nm and 800 nm in the upper regions of the visible portion and the near infrared region, but also included wavelength values as low as 400 and as high as 850. Many of these indices are geared towards measuring chlorophyll, green pigment, and leaf area index (Haboudane, 2002; Viña, 2011). The one predictor that was within the top five of both models, was the enhanced vegetation index (EVI), which was developed for application over a multitude of landscapes and vegetation types (Huete, 1997).

The top two predictor variables had very high importance. The first was the double difference index (DD) and second was SumDr1, or fully written out as first-order derivative green vegetation index derived using zero baseline. These both relate to measuring chlorophyll (Elvidge & Chen 1995; le Maire, François, & Dufrêne, 2004), however the DD index is more pronounced in accounting for the change in chlorophyll. SumDr1 is based on leaf area index (LAI) and percent green cover, however this index has greater capability in removing background noise than others such as normalized difference vegetation index (NDVI) or ratio vegetation index (RVI) (Elvidge & Chen, 1995). The important variables are mainly focused vegetation indices with some variation in detail, such as the fourth variable RARS, which uses the absorption bands of chlorophyll a and b, and those of carotenoids (Chappelle 1992).

The vegetation indices showed greater importance than the resampled bands themselves. Two of the twenty most important variables were resampled with the shadow classifier having

one. This predictor, however, was the most important variable in the model which was a 10 nm resampled predictor at a similar region of the resampled band predictor for the health classifier. The spectra of the ash samples also show the relevant areas of the electromagnetic spectrum in deciphering ash health. The model predictors are composed of reflectance values at different portions of the electromagnetic spectrum that can be represented by the spectral signatures.

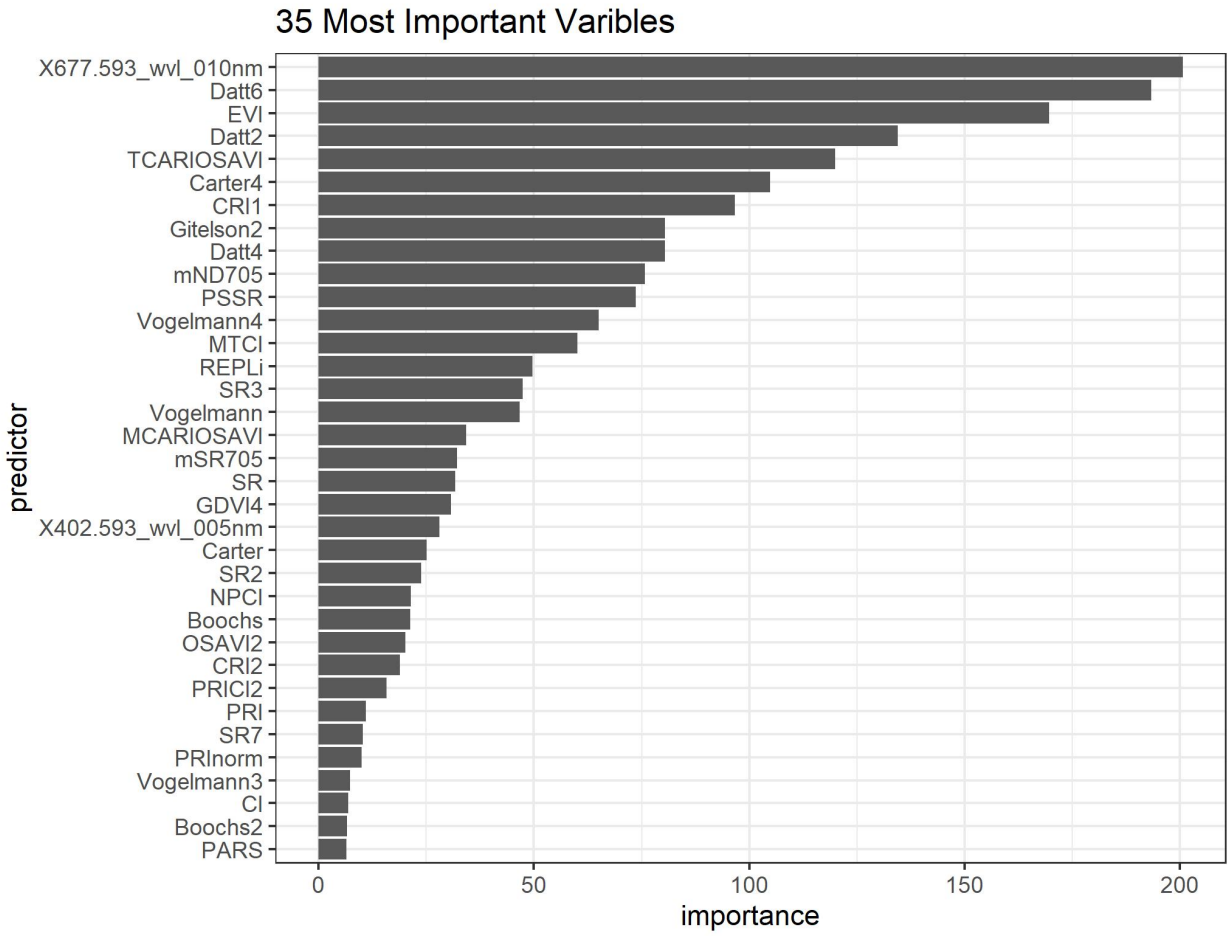

Figure 2.3. Displays the 35 most important predictor variables in the shadow model, the first 10 of which were included in the selected classifier, and corresponding importance level.

Table 2.2. 10 most important variables in the shadow model including name or description, and band formula.

	Abbreviation	**Name**	**Formula**
1	X677.593_wvl_010nm		'Bandpass 677.593 resampled at 10 nm'
2	Datt6	'Chlorophyll & carotenoids'	$R_{860}/(R_{550} * R_{708})$
3	EVI	Enhanced Vegetation Index	$2.5 * ((R_{800} - R_{670})/(R_{800} - (6 * R_{670}) - (7.5 * R_{475}) + 1))$
4	Datt2	'Chlorophyll & height'	R_{850}/R_{710}
5	TCARIOSAVI	Transformed Chlorophyll Absorption Ratio / Optimized Soil Adjusted Vegetation Index	$(3 * ((R_{700} - R_{670}) - 0.2 * (R_{700} - R_{550}) * (R_{700}/R_{670})))/ ((1 + 0.16) * (R_{800}-R_{670})/(R_{800} + R_{670} + 0.16))$
6	Ctr4	Carter 4	R_{710}/R_{760}
7	CRI1	Carotenoid Reflectance Index 1	$1/R_{515} - 1/R_{550}$
8	Gitelson2	'Chlorophyll & height'	$(R_{750} - R_{800})/(R_{695} - R_{740}) - 1$
9	Datt4	'Chlorophyll & carotenoids'	$R_{672}/(R_{550} * R_{708})$
10	mND705	Modified NDVI 2	$(R_{750} - R_{705})/(R_{750} + R_{705} -2 * R_{445})$

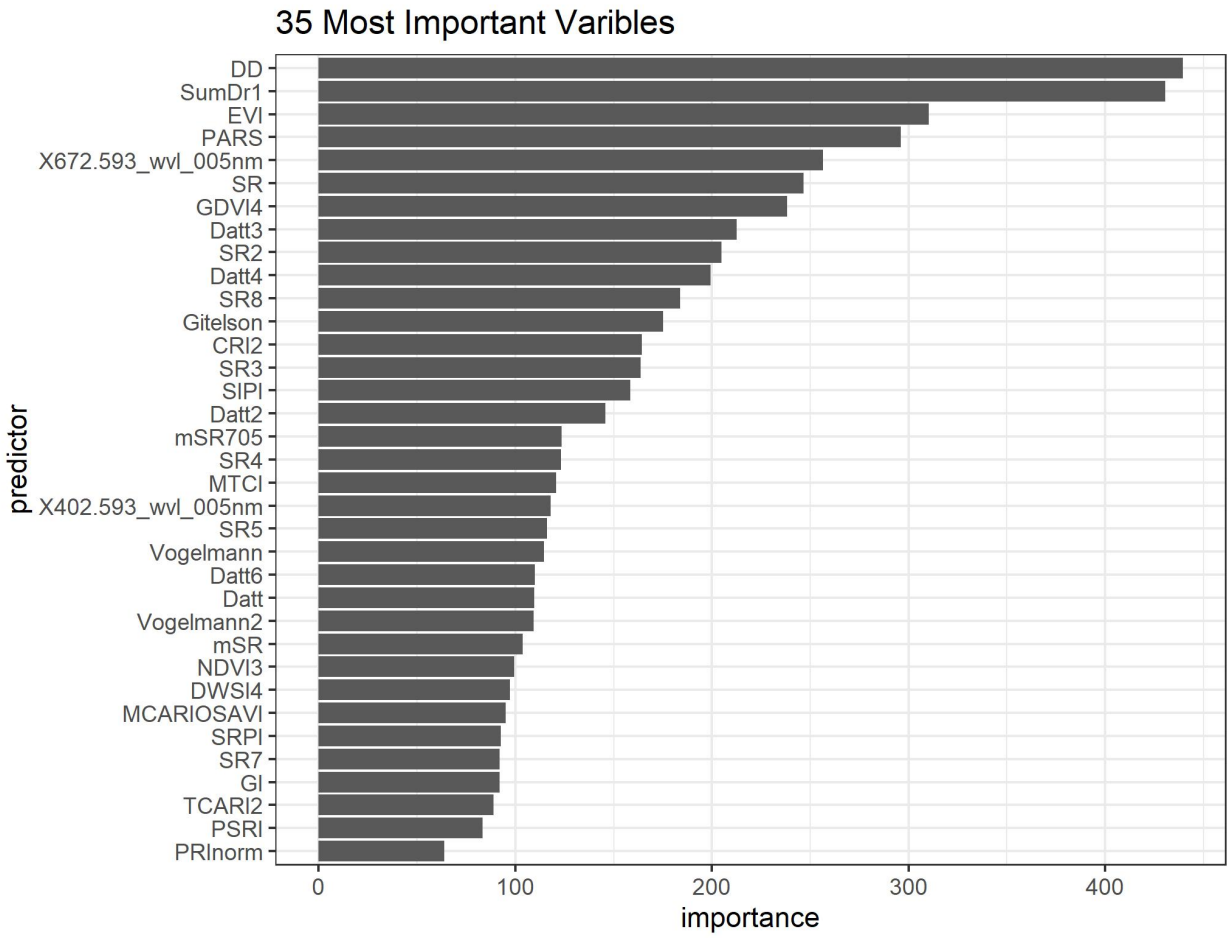

Figure 2.4. Displays the 35 most important predictor variables in the health classification model, the first 20 of which were included in the selected classifier, and corresponding importance level.

Table 2.3. 20 most important variables in the health model including name or description, and band formula.

	Abbreviation	**Name**	**Formula**
1	DD	Double Difference Index	$(R_{749} - R_{720}) - (R_{701} - R_{672})$
2	SumDr1	'LAI & % green cover'	$\sum_{i=626}^{795} D1i$
3	EVI	Enhanced Vegetation Index	$2.5 * ((R_{800} - R_{670})/(R_{800} - (6 * R_{670}) - (7.5 * R_{475}) + 1))$
4	RARS	Ratio Analysis of Reflectance Spectra	R_{746}/R_{513}
5	X672.593_wvl_005_nm		'Bandpass 672.593 resampled at 5 nm'
6	SR	Simple Ratio	R_{800}/R_{680}
7	GDVI4	Green Difference Vegetation Index 4	$(R^4_{800} - R^4_{680})/(R^4_{800} + R^4_{680})$
8	Datt3	'Chlorophyll & height'	D_{754}/D_{704}
9	SR2	Simple Ratio 2	R_{752}/R_{690}
10	Datt4	'Chlorophyll & carotenoids'	$R_{672}/(R_{550} * R_{708})$
11	SR8	Simple Ratio 8	R_{515}/R_{550}
12	Gitelson	'Chlorophyll'	$1/R_{700}$
13	CRI2	Carotenoid Reflectance Index 2	$1/R_{515} - 1/R_{770}$
14	SR3	Simple Ratio 3	R_{750}/ R_{550}
15	SIPI	Structure Intensive Pigment Index	$(R_{800} - R_{445})/(R_{800} - R_{680})$
16	Datt2	'Chlorophyll & height'	R_{850}/R_{710}
17	mSR705	modified Simple Ratio 705	$(R_{750} - R_{445})/(R_{705}-R_{445})$
18	SR4	Simple Ratio 4	R_{700}/ R_{670}
19	MTCI	MERIS Terrestrial Chlorophyll Index	$(R_{754} - R_{709})/(R_{709} - R_{681})$
20	X402.593_wvl_005nm		'Bandpass 402.593 resampled at 5 nm'

2.3.4. Spectral signatures

Various spectra of tree samples and pixels were analyzed and was mainly done so to assess radiometric calibration. The critical analysis, however, was using these spectra in development of the spectral variables, being the calculations of the vegetation indices and band

resampling values. Nonetheless, the signatures can showcase information of interest. Figure 5 shows a signature plot of eight ash samples, with two of each health class.

The spectra of lower classes (healthier) generally have higher reflectance rates compared to those of the less healthy classes. The spectra also tend to have more noise as it progresses into the end of the near infrared region, a result of the spectrometer.

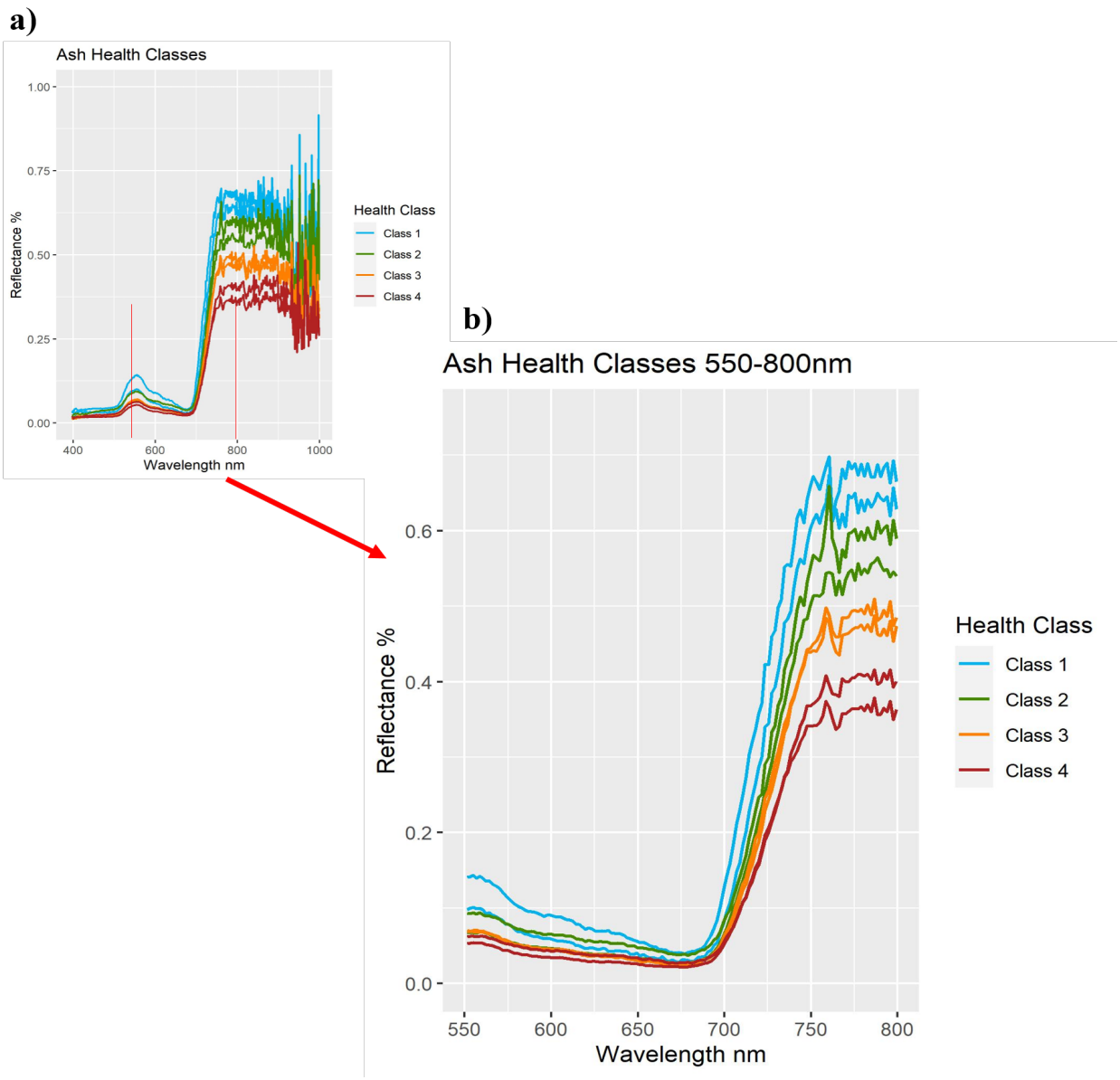

Figure 2.5. (a) Plots of eight ash spectra, two of class 1 (healthy) as light blue, two of class 2 as green, two of class 3 as orange, and two of class 4 (unhealthy) as red. **(b)** The spectrum focused on 550-800 nm wavelengths.

2.3.5. Classification

All cube images that contained samples were classified in order to obtain the buffer classifications. In total there were 18 processed cubes with 59 ash samples. Figure 6 shows a true color scene on the Hooksett site along with the shadow and health classification. The fourth image shows buffered samples (with shadows removed) overlaid on the true color image. Figure 7 similarly shows the overlaid buffers on the NHTI site.

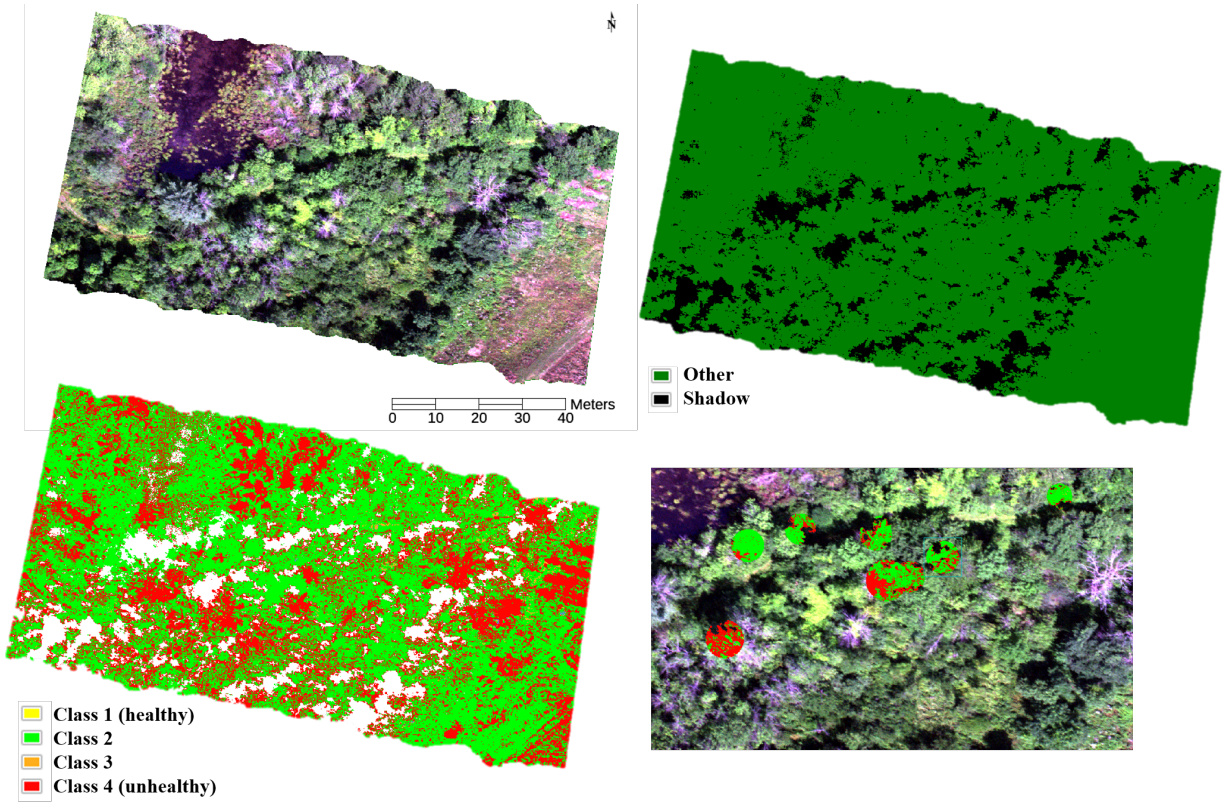

Figure 2.6. (a) A color scene on the Hooksett site **(b)** The classified scene for shadow **(c)** The shadow-masked scene classified for health. **(d)** A cropped image within the scene over the buffered tree samples.

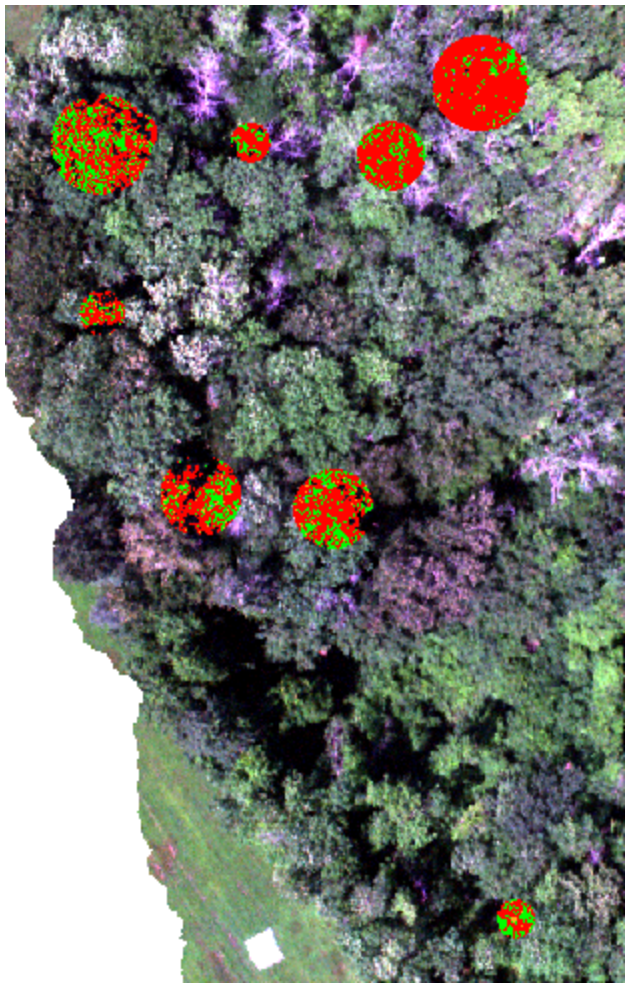

Figure 2.7. Cropped image over the NHTI site with 8 classified crown buffers

2.3.6. Crown prediction

Pixel classifications were calculated for each tree sample. To determine accuracy, the number of correctly classified pixels was divided by the total number of classified pixels, excluding shadows. Table 4 shows the individual tree id and its accuracy.

Table 2.4. Table of each buffer sample tree and its calculated accuracy (number of correctly classified pixels over the total).

	Tree Id	Accuracy
1	1_WA	0.041
2	11_GA	0.570
3	12_WA	0.893
4	12D_WA	0.600
5	13_WA	0.774
6	14_WA	0.787
7	14.5_WA	0.782
8	15_WA	0.720
9	16_WA	0.629
10	17_WA	0.637
11	18_WA	0.750
12	19_WA	0.702
13	20_WA	0.724
14	22_WA	0.629
15	23_WA	0.715
16	24_WA	0.786
17	25_WA	0.661
18	31_WA	0.034
19	101_GA	0.881
20	102_GA	0.719
21	102.5_WA	0.678
22	103_GA	0.472
23	104_GA	0.721
24	105_GA	0.712
25	106_GA	0.723
26	107_GA	0.894
27	107.5_GA	0.227
28	108_GA	0.695
29	109_GA	0.799
30	109D_GA	0.754

	Tree Id	Accuracy
31	110_GA	0.829
32	111_GA	0.860
33	112_GA	0.493
34	113_GA	0.875
35	116_GA	0.738
36	118_GA	0.742
37	119_GA	0.790
38	120_GA	0.381
39	201_WA	0.800
40	201D_WA	0.537
41	202_WA	0.583
42	203_WA	0.636
43	208.5_WA	0.891
44	208.5D_WA	0.900
45	209_WA	0.411
46	217_WA	0.340
47	219_WA	0.497
48	226_WA	0.160
49	227_WA	0.463
50	228_WA	0.456
51	228D_WA	0.209
52	230_WA	0.359
53	230D_WA	0.096
54	232_WA	0.345
55	232D_WA	0.166
56	234_WA	0.366
57	235_WA	0.189
58	237_WA	0.252
59	237D_WA	0.208

2.3.7. Validation test

The model was applied to a test validation dataset of pixels that were sampled on the same tree crowns of the calibration ash samples. The validation pixels were separate from those used to train the model. The prediction had an overall accuracy of 76%.

The classifier performed most accurately on those pixels of class 4 and less accurately on those pixels of class 1. The model evidently classifies those categories containing high calibration sample sizes more accurately than those of smaller. The highest sample was of class 4, followed by 2, 3, and then 1. The total number of class 1 ash trees was only 3.

Table 2.5. Error matrix of health class predictions on validation dataset.

	Predicted				
	1 (healthy)	2	3	4 (unhealthy)	
1 (healthy)	113	106	76	1	
2	14	877	18	65	
3	68	25	248	106	
4 (unhealthy)	2	88	53	738	
					76.1%

2.4. Discussion

This study utilized machine learning and remote imaging spectroscopy to classify tree health in ash trees. The spectral indices produced using the Ranger package had shown effectual in making predictions on the hyperspectral cube images. The indices used had logical premise and emphasized the sensitivity spectral bands hold through the index transformations. A possible addition in the process would be including other types of indices from other functions or

calculations. Derivatives and extensions of those were not heavily incorporated in this process aside from the vegetation indices that entailed the calculation. It may also be helpful in reducing the resampling predictors as it may not have had a significant effect on the predictors, especially as the number of band resampling increased.

In analyzing the classified buffer samples, it could be observed that near-shadow pixels, or outlines of the shadows, were frequently classified as class 4 (unhealthy). Further analysis could be conducted in assessing the characteristics of semi-shadowed pixels and how this potential misclassification could be avoided. This shows that it may be advantageous to use an object-based approach in a future study. This however can prove challenging as it is much more time-consuming in relating to both ground sampling and remote collection. In cases where ash trees are dispersed, processing can require high computing power. Images can be cropped and then processed, however cropping also takes time and manual effort.

Another consideration in the image classification process would be to add in a model to differentiate tree species. If this were effectively included, whole cube images could be processed which might negate the buffer process. The ability to at least decipher hardwood from softwood could be very conducive to the purpose of this project. This does impose challenges as species discrimination models are still being developed (Apostol, 2020; Kishore, 2020). It would also require a great deal more time and effort in the process and may be worthwhile to consider alternative model strengthening techniques.

The models developed in this study could be improved with larger and more consistent sample sizes. Accuracies of the buffered samples were variable and there was a large number that had a low accuracy score. The predicted pixel validation set also had variable accuracies due

to inconsistent sample sizes despite the pixel approach providing much higher counts. With understanding the limitations from sample size and uneven training data, the models still performed moderately well. The processes developed in this project can be utilized for additional studies in vegetation health detection.

A next step for these processes in further development could be to incorporate very recently infested sites in Maine using some destructive ash sampling to link EAB larval density to crown health. This would provide additions in other health classifications, assuming they are relatively healthy from no EAB infestation. Additionally, this could show if there are characteristics differentiating trees in declining health from borer infestation apart from other health factors. Pooling these data together will exhibit trends of infestation and may reveal early indications.

2.5. Conclusion

The goal of this project was to use hyperspectral imaging processes to practically classify stress through associating ground collected information to spectral data. Our collected information was impacted by EAB and factors into the overarching purpose of assisting in the provision of mitigation options for Maine land management. This was mainly achieved through the development of our models, although there are a number of actions that could strengthen the outcomes. Models may be improved if an object-based approach were taken. This would require a larger training dataset where each sample would be a whole crown rather than a pixel. In either case, a larger and more evenly dispersed data input will reinforce any model. The process of streamlining by means of data collection, processing, and efficient programming, is also critical to provide timely and accurate data that can be practically used by land managers (or other

interested groups), which would require greater computing capacity as well as data management. With continued input additions and modifications, our methods can assist in better tree monitoring practices.

Monitoring the health of forests is critical in land management and understanding, but there are limited convenient and rapid methods in doing so. Most of these still currently involve numerous field crews with variable protocols, time, resources, and repeated updating. With advances in remote sensing techniques paired with imaging spectroscopy, accurate and timely data can be provided.

BIBLIOGRAPHY

Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R., & Sousa, J. J. (2017). Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. Remote Sensing, 9(11), 1110.

Adjorlolo, C., Mutanga, O., Cho, M. A., & Ismail, R. (2013). Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification. International Journal of Applied Earth Observation and Geoinformation, 21, 535-544.

Agriculture and Agri-Food Canada. (2008). Crop Profile for Wild Blueberry in Canada. Retrieved from http://publications.gc.ca/collections/collection_2009/agr/A118-10-1-2008E.pdf

Altieri, M. A., & Nicholls, C. I. (2003). Soil fertility management and insect pests: harmonizing soil and plant health in agroecosystems. Soil and Tillage Research, 72(2), 203-211.

Annis, S. L., Slemmons, C. R., Hildebrand, P. D., & Delbridge, R. W. (2013, June). An internet-served forecast system for mummy berry disease in Maine lowbush blueberry fields using weather stations with cellular telemetry. In Phytopathology(Vol. 103, No. 6, pp. 8-8). 3340 PILOT KNOB ROAD, ST PAUL, MN 55121 USA: AMER PHYTOPATHOLOGICAL SOC.

Annis, S. L., & Stubbs, C. S. (2004). Stem and leaf diseases and their effects on yield in maine lowbush blueberry fields. Small Fruits Review, 3(1-2), 159-167. doi:10.1300/J301v03n01_16

Apostol, B., Petrila, M., Lorenţ, A., Ciceu, A., Gancz, V., & Badea, O. (2020). Species discrimination and individual tree detection for predicting main dendrometric characteristics in mixed temperate forests by use of airborne laser scanning and ultra-high-resolution imagery. The Science of the Total Environment, 698, 134074. doi:10.1016/j.scitotenv.2019.134074

Baret, F., Guyot, G., Begue, A., Maurel, P., & Podaire, A. (1988). Complementarity of middle-infrared with visible and near-infrared reflectance for monitoring wheat canopies. Remote Sensing of Environment, 26(3), 213-225. doi:10.1016/0034-4257(88)90078-8

Bertone, R. (2017). Maine's Top 10 Ag Products. Retrieved from https://www.farmflavor.com/maine/maines-top-10-ag-products-infographic/

Blackmore, W. (2019, November 25). As insect invaders approach, researchers use a combination of indigenous knowledge and Western forestry science to save a valuable tradition. The Verge. Retrieved from https://www.theverge.com/2019/11/25/20976144/emerald-ash-borer-baskets-wabanaki-invasive-pest-maine-tradition-trees

Bojović, B., & Marković, A. (2009). Correlation between nitrogen and chlorophyll content in wheat (Triticum aestivum L.). Kragujevac Journal of Science, 31, 69-74.

Brino, A. (2016). What does climate change mean for The County. Retrieved from https://bangordailynews.com/2016/06/23/news/aroostook/what-does-climate-change-mean-for-the-county/

Canadian Food Inspection Agency. (2014). Emerald Ash Borer - Questions and Answers. Retrieved from http://www.inspection.gc.ca/plants/plant-pests-invasive-species/insects/emerald-ash-borer/faq/eng/1337355937903/1337356019017

Cappaert, D., McCullough, D. G., Poland, T. M., & Siegert, N. W. (2005). Emerald ash borer in North America: a research and regulatory challenge. *American Entomologist. 51 (3): 152-165.*, *51*(3).

Chappelle, E. W., Kim, M. S., & McMurtrey III, J. E. (1992). Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. Remote sensing of environment, 39(3), 239-247.

Cox, S. (2002). Information technology: the global key to precision agriculture and sustainability. Computers and electronics in agriculture, 36(2-3), 93-111.

Dash, J., & Curran, P. J. (2010;2004;). The MERIS terrestrial chlorophyll index. International Journal of Remote Sensing, 25(23), 5403-5413. doi:10.1080/0143116042000274015

Datt, B. (1999). A new reflectance index for remote sensing of chlorophyll content in higher plants: tests using Eucalyptus leaves. Journal of Plant Physiology, 154(1), 30-36.

Elvidge, C. D., & Chen, Z. (1995). Comparison of broad-band and narrow-band red and near-infrared vegetation indices. Remote sensing of environment, 54(1), 38-48.

FOR/Maine. (2018). Forest Opportunity Road Map/Maine Report. Retrieved from https://formaine.org/wp-content/uploads/2018/09/FORMaine_Report_DL.pdf

Gitelson, A. A., Buschmann, C., & Lichtenthaler, H. K. (1999). The chlorophyll fluorescence ratio F735/ F700 as an accurate measure of the chlorophyll content in plants. Remote Sensing of Environment, 69(3), 296-302. doi:10.1016/S0034-4257(99)00023-1

Gitelson, A. A., & Merzlyak, M. N. (1997). Remote estimation of chlorophyll content in higher plant leaves. International Journal of Remote Sensing, 18(12), 2691-2697.

Glass, V. M., Percival, D. C., & Proctor, J. T. A. (2005). Tolerance of lowbush blueberries (Vaccinium angustifolium Ait.) to drought stress. II. Leaf gas exchange, stem water potential and dry matter partitioning. Canadian journal of plant science, 85(4), 919-927.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote sensing of environment, 81(2-3), 416-426.

Hepler, P. R., & Yarborough, D. E. (1991). Natural variability in yield of lowbush blueberries. Hortscience, 26(3), 245-246. doi:10.21273/HORTSCI.26.3.245

Herms, D. A., & McCullough, D. G. (2014). Emerald ash borer invasion of north america: History, biology, ecology, impacts, and management. Annual Review of Entomology, 59(1), 13-30. doi:10.1146/annurev-ento-011613-162051

Hernández-Clemente, R., Navarro-Cerrillo, R. M., & Zarco-Tejada, P. J. (2012). Carotenoid content estimation in a heterogeneous conifer forest using narrow-band indices and PROSPECT+ DART simulations. Remote Sensing of Environment, 127, 298-315.

Huete, A. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sensing of Environment, 59(3), 440-451. doi:10.1016/s0034-4257(96)00112-5

Kishore, B. S. P. C., Kumar, A., Saikia, P., Lele, N. V., Pandey, A. C., Srivastava, P., ... & Khan, M. L. (2020). Major forests and plant species discrimination in Mudumalai forests region using airborne hyperspectral sensing. Journal of Asia-Pacific Biodiversity.

Kovacs, K. F., Haight, R. G., McCullough, D. G., Mercader, R. J., Siegert, N. W., & Liebhold, A. M. (2010). Cost of potential emerald ash borer damage in U.S. communities, 2009–2019.Ecological Economics, 69(3), 569-578. doi:10.1016/j.ecolecon.2009.09.004

Lawrence, R., & Labus, M. (2003). Early detection of douglas-fir beetle infestation with subcanopy resolution hyperspectral imagery. Western Journal of Applied Forestry, 18(3), 202-206. doi:10.1093/wjaf/18.3.202

Leavesley, S. J., Walters, M., Lopez, C., Baker, T., Favreau, P. F., Rich, T. C., . . . Boudreaux, C. W. (2016). Hyperspectral imaging fluorescence excitation scanning for colon cancer detection. Journal of Biomedical Optics, 21(10), 104003.

Le Maire, G., Francois, C., & Dufrene, E. (2004). Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. Remote sensing of environment, 89(1), 1-28.

Maine Forest Service. (2018). Forest & Shade Tree Insect & Disease Conditions for Maine. Retrieved from https://www.maine.gov/dacf/mfs/publications/condition_reports.html

Maccioni, A., Agati, G., & Mazzinghi, P. (2001). New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. Journal of Photochemistry and Photobiology B: Biology, 61(1-2), 52-61.

Maschler, J., Atzberger, C., & Immitzer, M. (2018). Individual tree crown segmentation and classification of 13 tree species using airborne hyperspectral data. Remote Sensing, 10(8), 1218. doi:10.3390/rs10081218

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. (2005). Future directions of precision agriculture. Precision agriculture, 6(1), 7-23.

McCullough, D. (2017). Emerald Ash Borer Information Network FAQ. Retrieved from http://www.emeraldashborer.info/faq.php

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia plantarum, 106(1), 135-141.

Nelson, P.R. LECOSPEC, (2020), Github Repository, https://github.com/nelsopet/lecospec

Nelson, P., & Thompson, N. (2018). Visible and infrared imaging spectroscopy for high resolution mapping and health assessment of Maine's forest and agricultural resources. 2018 Maine Economic Improvement Funds (MEIF) Small Campus Initiative (SCI) Research Grant.

Neuman, L. K. (2010). Basketry as Economic Enterprise and Cultural Revitalization: The Case of the Wabanaki Tribes of Maine. Wicazo Sa Review 25(2), 89-106. University of Minnesota Press. Retrieved December 19, 2018, from Project MUSE database.

Nguyen-Do-Trong, N., Dusabumuremyi, J. C., & Saeys, W. (2018). Cross-polarized VNIR hyperspectral reflectance imaging for non-destructive quality evaluation of dried banana slices, drying process monitoring and control. Journal of Food Engineering, 238, 85-94. doi:10.1016/j.jfoodeng.2018.06.013.

Oumar, Z., Mutanga, O., & Ismail, R. (2013). Predicting Thaumastocoris peregrinus damage using narrow band normalized indices and hyperspectral indices using field spectra resampled to the Hyperion sensor. International Journal of Applied Earth Observation and Geoinformation, 21, 113-121.

Panta, G., Rieger, M., & Rowland, L. (2001). Effect of cold and drought stress on blueberry dehydrin accumulation. The Journal of Horticultural Science and Biotechnology, 76(5), 549-556. doi:10.1080/14620316.2001.11511409

Pontius, J., Hanavan, R. P., Hallett, R. A., Cook, B. D., & Corp, L. A. (2017). High spatial resolution spectral unmixing for mapping ash species across a complex urban environment. *Remote Sensing of Environment*, *199*, 360-369.

Pontius, J., Martin, M., Plourde, L., & Hallett, R. (2008). Ash decline assessment in emerald ash borer-infested regions: A test of tree-level, hyperspectral technologies. *Remote Sensing of Environment*, *112*(5), 2665-2676.

Ritter, M. (2017). Scientists say ash tree species on brink of extinction in eastern U.S. Retrieved from https://www.pressherald.com/2017/09/14/scientists-say-ash-tree-species-on-brink-of-extinction-in-eastern-u-s/

Roberts, D. A., Roth, K. L., & Perroy, R. L. (2016). 14 hyperspectral vegetation indices. Hyperspectral remote sensing of vegetation., 309.

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. Remote Sensing of Environment, 55(2), 95-107. doi:10.1016/0034-4257(95)00186-7

Sampson, W. (2018). Emerald ash borer found in Maine. Retrieved from https://www.woodworkingnetwork.com/news/woodworking-industry-news/emerald-ash-borer-found-maine

Schilder, A. (2015, April). Growth Stages. Michigan State University Extension. https://www.canr.msu.edu/blueberries/growing_blueberries/growth-stages

Siegert N.W., McCullough D.G., Liebhold A.M., Telewski F.W. Dendrochronological reconstruction of the epicentre and early spread of emerald ash borer in North America. Divers. Distrib. 2014;20:847–858. doi: 10.1111/ddi.12212

Tallamy, Douglas. Bringing Nature Home: How Native Plants Sustain Wildlife in Our Gardens. 2007. Timber Press. ISBN-10:0881929921

Thorp, K. R., Wang, G., Bronson, K. F., Badaruddin, M., & Mon, J. (2017). Hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield. Computers and Electronics in Agriculture, 136, 1-12. doi:10.1016/j.compag.2017.02.024

USDA. (2020, June 2). Emerald Ash Borer. United States Department of Agriculture, Animal and Plant Health Inspection Service. https://www.aphis.usda.gov/aphis/ourfocus/planthealth/plant-pest-and-disease-programs/pests-and-diseases/emerald-ash-borer

Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. Remote Sensing of Environment, 115(12), 3468-3478.

Wang, X. Y., Yang, Z. Q., Gould, J. R., Zhang, Y. N., Liu, G. J., & Liu, E. S. (2010). The biology and ecology of the emerald ash borer, Agrilus planipennis, in China. *Journal of insect science*, *10*(1).

Whittle, P. (2018). Hard times as Maine wild blueberry industry in decline. Retrieved from https://www.timesrecord.com/articles/maine-1/hard-times-as-maine-wild-blueberry-industry-in-decline/

Wright, M. N., & Ziegler, A. (2015). ranger: A fast implementation of random forests for high dimensional data in C++ and R. arXiv preprint arXiv:1508.04409.

Yarborough, D. E. (2004). Factors contributing to the increase in productivity in the wild blueberry industry. Small Fruits Review, 3(1-2), 33-43. doi:10.1300/J301v03n01_05

Zarco-Tejada, P. J., Camino, C., Beck, P. S. A., Calderon, R., Hornero, A., Hernández-Clemente, R., ... & Gonzalez-Dugo, V. (2018). Previsual symptoms of Xylella fastidiosa infection revealed in spectral plant-trait alterations. Nature Plants, 4(7), 432.

Zarco-Tejada, P. J., Ustin, S. L., & Whiting, M. L. (2005). Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery. Agronomy Journal, 97(3), 641-653.

BIOGRAPHY OF THE AUTHOR

Catherine Chan was born in Lincoln, Maine, on May 29th, 1992. She was raised in Lincoln, Maine and attended the district public schools. She graduated from Boston College in 2014 with a degree in Marketing and minor in Environmental Studies. Catherine is a candidate for the Master of Science degree in Forest Resources from the University of Maine in August 2020.