# NORTHWEST NAZARENE UNIVERSITY

Analysis of Spectral Reflectance and Separability of Vegetation for FireMAP.

THESIS Submitted to the Department of Mathematics and Computer Science in partial fulfillment of the requirements for the degree of BACHELOR OF ARTS

Mikhail Bowerman 2017

# THESIS Submitted to the Department of Mathematics and Computer Science in partial fulfillment of the requirements for the degree of BACHELOR OF ARTS

# Mikhail Bowerman 2017

Analysis of Spectral Reflectance and Separability of Vegetation for FireMAP.

Author:

What Mikhail Bowerman

Approved:

Wet How we

Dale Hamilton, Professor of Computer Science Department of Mathematics and Computer Science, Faculty Advisor

Approved: <u>Und</u> Eva Strand, Research Assistant Professor Department of Forest, Rangeland, and Fire Sciences, Second Reader

Approved: <u>Barry Myers</u> Barry Myers, Ph.D., Chair

Department of Mathematics and Computer Science

# ABSTRACT

Analysis of Spectral Reflectance and Separability of Vegetation for FireMAP.

BOWERMAN, MIKHAIL(Department of Mathematics and Computer Science),

HAMILTON, DALE (Department of Mathematics and Computer Science)

Wildland fires can be destructive to properties and dangerous to people in close proximity, with the cost of some large fires exceeding \$1 billion. They are a threat to the economy, property, and the public safety. Wildfires are however an essential component for the ecology of many vegetation types and it is important to understand when fires are beneficial and when they are destructive. The goal of the Fire Monitoring and Assessment Platform (FireMAP) is to provide fire managers with the tools and knowledge for acquiring, analyzing, and managing hyper-resolution imagery to map burn severity in a faster, safer, and more affordable manner than is currently possible. This will allow for quicker and more educated decisions on how to proceed with recovery after a wildland fire.

The FireMAP Spectral Analysis focuses on vegetation common to Idaho and the Pacific Northwest as well as ash from post-burn sites. This effort investigated whether spectral reflectance can be utilized to differentiate between classes of vegetation and ash. Following spectral and statistical analysis, spectral separability of classes of ash and vegetation was discovered in the visible light range of the electromagnetic spectrum. With this information, fire severity and extent can be determined from hyper resolution imagery using machine learning classifiers focusing on the visible light spectrum.

#### Acknowledgements

First, I would love to thank my professor, advisor, and mentor, Dale Hamilton, for giving me the opportunity to work on this project and attend multiple conferences where I could present our research. I would also like to acknowledge the helping hands that several other NNU faculty offered throughout the project: Dr. Barry Myers, Dr. Jerry Harris, Dr. Jason Colwell and Dr. John Cossel. Also, a thank you to Sarah Hurt from Deer Flat Wildlife Refuge for her help with collection of vegetative samples and spectroscopy. Last, but not least, I must give my appreciation to my fellow students that participated in the FireMAP research project alongside me.

# Table of Contents

Title Page	i
Signature Page	ii
Abstract	iii
Acknowledgements	iv
List of Figures	vi
Overview	1
Background	2
Collecting the Samples	3
Spectral Analysis	5
Statistical Analysis	8
Building the Database	
Future Work	14
References	16
Appendix A	
Appendix B	

# List of Figures

Figure 1 - Sample Reflectance Curve	6
Figure 2 - Mean Class Reflectance	7
Figure 3 - T-test Curves	9
Figure 4 - Database Entity Relationship Diagram	
Figure 5 - Database Sample	
Figure 6 - Example .csv File	
Figure 7 - Example Unnecssary Data from .csv File	

# Overview

The purpose of this research is to determine if different types or classes of vegetation and ash produce varying spectral results when analyzed with a spectrometer. The vegetation samples in question are all native to Idaho and other parts of the Northwest. Using statistical analysis, the goal is to calculate if and where these spectral differences occur. This research is contributing to a larger project, FireMAP, with the goal of mapping extent and severity of wildfires. To map wildland fire severity, it is necessary to classify post-fire imagery into high severity, low severity, surface and canopy vegetation. These classes are white ash and black char, shrubs, herbaceous, coniferous, and deciduous. Using the machine learning algorithms to classify pixels based on their spectral reflectiveness requires establishing enough of a difference between the classes. This ensures that the classifiers will be able to differentiate between the classes based on each class' spectral signature. After statistical analysis of the spectral data from each class, there was a detectable difference in the spectral reflectance between each class. This enables the machine learning classifiers to differentiate between the classes.

#### Background

Because of the major threat that wildfires pose to areas in the western United States, wildland fire managers need a way to quickly and accurately understand the situation. The ultimate purpose of the FireMAP project is to equip managers with decision-making tools. What kind of vegetation was burned? What kind of vegetation surrounded the fire? FireMAP helps answer these questions by mapping post-fire effects, allowing recovery teams to determine the severity of large area fires. This also enables them to make decisions on how to proceed with rehabilitation of the burned area, as well as provides them with information on how to respond to an active fire.

This specific part of the FireMAP project was quite distinct from the rest. Most other aspects of FireMAP dealt with coding and testing the system, but data collection was the first step. The data analysis was essentially the foundation upon which the machine learning classifier was built. However, it needed the proper parameters and data so that it could work the way it was intended. The goal was to ascertain this parametrical data to enable the classifiers to extract useful information from the data.

#### **Collecting the Samples**

The first order of business when starting up the research was knowing what vegetation in Idaho was native to the area or adequately common throughout the region that it would be important to have information about it. This was important because it was deemed necessary to have spectral data on vegetation that a wildland fire would be likely to encounter. This process took longer than planned. Along with several days of creating a list of ideal plant species to gather information on, there were meetings with local ecologists who offered input on what should be collected. Because of the numerous species of vegetation on campus, it was important to have a map of the campus and supplemental list of vegetative species. With this information, it became much easier to understand what trees on campus would be viable samples, due to their common occurrence in the Pacific Northwest. Several trees across campus are not native to Idaho or even the northwestern United States (US); many were brought in from other parts of the country or even outside of the country. Only abundant species of the northwest were really of interest. This stage would have taken much less time if there was adequate spectral data for a variety of species readily available. Unfortunately, there were no sufficient spectral libraries (databases with spectroscopic information) for ash and vegetation common to the northwest.

Most vegetation samples were collected within Idaho, typically on the campus of NNU or at the nearby Deer Flat National Wildlife Refuge (DFNWR). Because of the time it takes to get out to the refuge and a lack of knowledge of shrubs and grasses, it was in the best interest of the project to ask for help from personnel on the Deer Flat NWR. Sarah Hurt, a Field Botanist at the DFNWR, graciously offered her services to the project. She knew the layout of the refuge and the types of vegetation around it. Over 40 vegetation samples were collected from DFNWR, close to half of the total samples collected. Additional samples were also collected across

montane regions of southern Idaho, from the mesic Payette River watershed to the xeric Owyhee Mountains.

Any vegetative samples collected outside of Idaho were found in Oregon. A few of them come from Banks, some were collected just east of Pendleton in the Blue Mountains, and the rest were found around Bend. The locations from which the vegetative samples were obtained are representative of the mesic and xeric ecosystems prevalent across the Pacific Northwest.

Ash was another key component to this research. It was important that ash samples were collected so that the spectral library would contain information for post-burn analytics. Both black char and white ash samples were collected from multiple burn sites. Black char indicates where the fire did not completely combust, while white ash is indicative of higher burn severity (Hudak et al., 2013).

Along with collecting and recording the vegetation or ash, it was imperative that the location of collection be recorded as well. Attributing samples with latitude and longitude, it was possible to have an exact location of where that vegetation or ash was collected. The collection was more than just grabbing a leaf or two. For trees, it was necessary to break, tear, or cut off a significant limb with enough needles or leaves for multiple samples. For grasses and shrubs, it was often a matter of just taking a handful of the vegetation and pulling it out of the ground. Collecting black char was fairly easy, as there were large deposits of charred vegetation all throughout a burned area. White ash was often rarer and more difficult to collect because it was so powdery, and it was too easy to accidentally collect black char with a white ash sample. It was important to take a sample of the vegetation that would be visible from directly above. The reason for this will be explained further on.

#### **Spectral Analysis**

After a piece of vegetation was cut and collected, the specimen was spectrally analyzed within 24 hours. Waiting any longer would likely lead to invalid results because of the loss of water by the leaves or needles. Water retention was very important for the retention of chlorophyll (Richardson, 2002). Coniferous samples had a longer viability period; the needles retain water much longer than the leaves of deciduous samples. It was discovered, however, that the viability period could be extended by at least 24 more hours if the sample was kept in damp, cool conditions after being cut. Samples were often held in a cooler with ice, and it worked quite well.

The purpose of the spectral analysis was to understand what range in the electromagnetic spectrum different classes of vegetation would produce different spectral results. It quickly became clear that the help of Dr. Jerry Harris would be needed to run the spectrometer in the Advanced Chemistry lab. It is a Cary 100 UV-Vis Spectrometer by Agilent Technologies. The spectral range of this spectrometer was 190 to 900 nanometers (nm) with a resolution of one nm. There was concern that this would not cover an adequate spectral extent; it would have been ideal to have a spectrometer with a range of about 300 to 1400 nm. This covers much more of the infrared spectrum, which initially seemed to be most important because it has generally shown promise for vegetative identification (Van Aardt, 2000). Fortunately, the UV-Vis spectrometer worked quite well, as subsequent analysis showed spectral separability between the classes within the spectral extent of the spectrometer.

Each sample of vegetation was analyzed three separate times (specimens) for the sake of consistency and accuracy. Each sample had to be calibrated to a baseline curve (zero reflectance). **Figure 1** is an example of the results from a Boxelder sample collected on NNU's

campus. The spectral values are recorded at each wavelength and saved automatically in a .csv file. It was then possible to create a line graph to visually represent the spectral curves for each sample.



Figure 1 - Sample Reflectance Curve

After several weeks of nonstop spectral analysis, there was finally enough sample data to create a respectable spectral library. The samples' spectral results, such as found in **Figure 1**, were averaged among their respective vegetative or ash class. There were approximately 15 to 20 samples of vegetation for each class (shrub, herbaceous, conifer, and deciduous). This created six separate average spectral curves, one for each class, including the black char and white ash classes. **Figure 2** is the visual representation of these calculations.



As can be seen, there are six separate curves on this graph, one for each of the classes of interest. Black char and white ash stand out among the six classes. Black char has little reflectance, regardless of the wavelength. White ash takes a dip in reflectance around 250 nm, but begins to increase at a steady rate after 400 nm. The vegetative reflectance curves all have very similar peaks, valleys, and plateaus, with the deciduous and coniferous curves falling almost on top of each other across the spectrum. The shrub and herbaceous curves stay a bit above the deciduous and coniferous curves, but still take the same rises and dips. All the vegetative reflectance curves rise significantly around 700 nm. This is Red Edge, close to entering the Near-Infrared spectrum.

It became quickly clear that vegetation was distinguishable from black char between about 350 and 900 nm and from white ash between 200 and 900 nm. White ash and vegetative reflectance values intersected in the Red Edge, but this did not lead to any issues in further analysis. The tricky part was distinguishing between the vegetative classes. This was to be expected. Looking at **Figure 2**, it was noticed that most of the fluctuations in vegetative reflectance occurred between 400 and 700 nm. Interestingly enough, this separation happens in the visible light range of the spectrum. Theories started forming, but this was purely speculation. Statistical proof was needed to solidify any theories about how to proceed.

#### **Statistical Analysis**

In order to calculate spectral separability, a statistical T-test had to be used. This T-test calculates a probability as to whether samples from two different populations (or classes) are statistically different at a given wavelength. With the spectrometer's range from 190 to 900 nm and a measurement taken every five nm, there were 143 measurements taken for each specimen, their values

#### $R_{190}, R_{195}, R_{200}, \dots R_{895}, R_{900}$

For each R-value, the T-test was used to calculate the difference in mean between the two classes. In a T-test, a null hypothesis is required. In this case, the null hypothesis states that there is not a significant enough difference between the reflectance values of Class 1 and Class 2, that any differences are only due to a matter of chance. A P-value is returned from each calculation, essentially determining this chance. A higher P-value means a greater probability that the difference is a matter of chance. The goal was to reject the null hypothesis, or confirming that there is a significant difference. This required low P-values, or low probabilities of chance differences. In order to do this, a significance level had to be chosen and compared to the P-values for each wavelength. With the help of Dr. Jason Colwell from Northwest Nazarene University, a significance level of 0.1 was originally chosen. Anywhere that the P-values fell

below this significance level showed where differences in reflectance values were not a matter of chance.



Figure 3 - T-test Curves

**Figure 3** displays each T-test's P-value curve. Six tests were performed between the six classes. T-tests between the canopy fuels (conifer/deciduous) and ground fuels (shrub/herb) are not displayed; their results were inconclusive. Every curve in **Figure 3** is below the significance level 0.1 between 450 nm and 700 nm, the visible light range of the spectrum. However, because 0.1 is a rather high significance level (0.05 and 0.01 are much preferred in the scientific community), it was appropriate to drop the significance level. For the vegetative T-tests, the significance level was reduced to 0.05. This narrowed the results of acceptable P-values to the range of approximately 590 nm to 700 nm for the tests where only vegetative classes are compared. The significance level was reduced even further for the T-tests comparing black char

to another class. At 0.01, this new significance level displayed the most accurate P-values. The spectral ranges in which black char and vegetation are separable are about 420 to 650 nm and then 700 to 900 nm. Black char and white ash are spectrally separable across the board. In fact, their T-test curve is practically invisible because it rests so close to the x-axis in **Figure 3**.

This is a very good sign because differences are detectable, especially in the visible light spectrum for every T-test. Normal cameras are going to be just fine for taking pictures of an area of land for the classifier. Currently, there is not a demand for any UV or Infrared cameras. However, the curves are all dropping quite rapidly just beyond 700 nm, excluding the "Black/Veg" and "Black/White" T-tests because they are already so low. In fact, two of these curves drop back below 0.01 right at 900 nm. The top two T-test curves do not reach that point in this spectral range, but there is a chance that, beyond 900 nm, they drop back below the 0.05 significance level, or possibly even below 0.01. If this is true, it would mean that all canopy fuels would be spectrally separable from ground fuels. The only way to test this theory, however, is with a spectrometer with a spectral range extending beyond 900 nm, which is near the beginning of the Infrared light range. Unfortunately, NNU is not in possession of such a spectrometer. This opens the door for future work.

#### **Building the Database**

After getting most of the samples analyzed, storing the data became an urgent task. Unfortunately, due to some unforeseen technical problems and a mid-summer power outage, connecting to the CS server was impossible, let alone creating a database. So, instead of creating a DBMS (database management system) through the preferred MySQL Workbench, the DBMS was built in Microsoft Access. This was less than ideal for many reasons, one of which was the fact that Access has less functionality than desired, especially for this project. But it needed to be

done; the data needed to be stored somewhere if only temporarily. **Figure 4** is the Entity Relationship Diagram of the database created in Access.



Figure 4 - Database Entity Relationship Diagram

There are three tables in this database: *Cutting*, *Sample*, and *Spectrum*. The way in which the Access database is set up, it allows users to navigate rather easily to any data they are looking for. Every cutting contains three samples, and each sample has 710 wavelength and reflectance values. **Figure 5** is a visual representation of this layout. As can be seen, it is a sort of hierarchy in terms of data location. A user can decide which wavelength(s) of which sample(s) of which cutting(s) he/she wants to view and then follow that path in reverse. This was the easiest way to store the data with Access. Now *Cutting* contains all of the metadata, or information about the data, such as Species or Cut Date.

2	_	Сι	utll	D 🚽	t	Spec	ies		l
	曱	1				A	sh		
	L	Z		Sa	mpleID	→ Cli	ck to Add	-	
			Ŧ			1			
				$\angle$	Wavelen	gth 🚽	Reflecta	nce	*
						900	54	.644	73
						899	54	.855	88
						898	54	.980	93
						897	54	.384	18
						896	55	.241	.83
						895	54	.759	59
						894	55	.118	33
						893	5	4.73	58

Figure 5 - Database Sample

Importing the data from the original .csv files into Access was a bit of a nuisance at first. Fortunately, it did not take much time to figure out a functional set of steps that made the files easy to import and without information that was not needed for the database. This was essentially an ETL (Extract Transform Load) process (Coronel, 2015). The data from each .csv file was extracted, manipulated, and loaded in to the Access database using multiple MySQL importation queries. After getting each file imported, another set of queries were needed to add the data from the file into the necessary tables of the DBMS. These queries are listed in Appendix B.

**Figure 6** is an example of what a typical .csv file contained. This is from a sample of Subalpine Fir. **Figure 6** focuses on the data that was relevant to this project, specifically the sample IDs and their respective wavelength reflectance values. The *Baseline 100%T* and *Baseline0%T* headings are the calibration curves mentioned in the Spectral Analysis section of this paper. Their reflectance values are also contained in the .csv file. These are examples of data that is not necessary to the FireMAP project, so they needed to be removed during importation. **Figure 7** is a snippet from the same .csv file. This figure represents all of the extra information that did not need to be inserted into the Access database. There were hundreds of lines of information like this, but this information would not be important to the users, so it was not

included.

```
Baseline 100%T,,Baseline 0%T,,Sample 1,,Sample 2,,Sample 3,,
Wavelength (nm),%R,Wavelength (nm),%R,Wavelength (nm),%R,Wavelength (nm),%R,Wavelength (nm),%R,
900,58.11428833,900,0.01005071681,900,51.70055008,900,53.31423187,900,51.09118271,
899,58.60682678,899,0.008380273357,899,51.36390305,899,51.91990662,899,51.65079117,
898,58.29543304,898,0.009659480304,898,52.07881165,898,52.72284698,898,52.03866196,
897,57.70018005,897,0.01211153902,897,52.17168808,897,53.33103561,897,52.1198616,
896,58.13626099,896,0.01434976701,896,51.6951561,896,53.51199341,896,51.80083084,
895,57.76392746,895,0.004459221847,895,51.80823135,895,52.84288025,895,51.81542587,
```

Figure 6 - Example .csv File

Baseline 100%T,					
Baseline 100%T					
Collection Time: 8/24/2016 11:37:17 AM					
Operator Name :					
Scan Software Version: 4.20(468)					
Parameter List :					
Instrument	Cary 100				
Instrument Version	12.00				
Start (nm)	900.00				
Stop (nm)	190.00				
X Mode	Nanometers				
Y Mode	%R				
UV-Vis Scan Rate (nm/min)	300.000				
UV-Vis Data Interval (nm)	1.000				
UV-Vis Ave. Time (sec)	0.200				
UV-Vis SBW (nm)	2.0				
Beam Mode	Double				

Figure 7 - Example Unnecssary Data from .csv File

Building the DBMS was very much a learning experience. It offered a brief yet pleasant aspect to this part of the project. While most of the summer was spent with data collection and analytics, creating the DBMS provided a means of data storage. Currently, access to the DBMS is limited only to those associated with the FireMAP project, but it can be requested by contacting NNU's Department of Mathematics and Computer Science.

#### **Future Work**

A great deal was accomplished last summer, but there is still much to be done. Specifically, three main tasks were completed. First, the collection of data needs to grow. This means continuing to go out into the field and obtaining more vegetation and ash samples. The spectral library is already the most comprehensive in the northwest, but why stop there? There is no harm in expanding it; in fact, growing would be very beneficial for NNU's credentials and for fine-tuning the statistical analysis portion. The more samples involved in the statistical analysis, the more accurate the T-tests can be. Along with collection of sample data, metadata should be stored as well. A goal in FireMAP is to create an app that will enable metadata acquisition, including geographic coordinates, collection date, and even an image of the collection site. This application will interface with the database so that the metadata can quickly and easily be stored.

Second, with more samples being collected, someone will need to input the data into the DBMS. While this task is not time consuming, it is very important that it be done correctly so as not to compromise the rest of the data. Data security and integrity are incredibly important when dealing with data entry and management. This part does not necessarily need to be done by the same person involved with data collection. Ideally, someone experienced with databases should be the one to work on this task.

Third, as mentioned earlier, Microsoft Access is not the ideal software to use when building a DBMS. It would be convenient to have the data stored on a SQL server. Someone will need to convert Access data so that the database can be easily migrated onto the SQL server. Storing the data here will be much more convenient for the purposes of this project. It is a multiuser platform so multiple people can access and manipulate data at the same time. Because of the nature of FireMAP, the extra functionality of a SQL server will be greatly beneficial.

Even with the completion of these tasks, it is more than likely that new tasks and goals will arise. A project like this is never finished; there is always more information to add, more tools to implement, new features to include. It is an ongoing project that can continually be perfected.

### References

- Coronel, C., & Morris, S. (2015). Database systems: design, implementation, and management. Australia: Course Technology Cengage Learning.
- Harris Geospatial Solutions. (2017). Plant Foliage. Retrieved June 20, 2016, from <a href="http://m.harrisgeospatial.com/docs/plantfoliage.html">http://m.harrisgeospatial.com/docs/plantfoliage.html</a>
- Hudak, A. T., Ottmar, R. D., Vihnanek, R. E., Brewer, N. W., Smith, A. M., & Morgan, P. (2013). The relationship of post-fire white ash cover to surface fuel consumption. International Journal of Wildland Fire, 22(6), 780-785. doi:10.1071/wf12150
- Mohler, R. L., & Goodin, D. G. (2012). Identifying a suitable combination of classification technique and bandwidth(s) for burned area mapping in tallgrass prairie with MODIS imagery. International Journal of Applied Earth Observation and Geoinformation, 14(1), 103-111.
- Richardson, A. D., & Berlyn, G. P. (2002). Changes in foliar spectral reflectance and chlorophyll fluorescence of four temperate species following branch cutting. Tree Physiology, 22(7), 499-506.
- Stat Trek. (n.d.). Hypothesis Test: Difference Between Means. Retrieved July 19, 2016, from http://stattrek.com/hypothesis-test/difference-in-means.aspx?tutorial=ap
- UC Davis. Spectral Properties of Leaves and Plants [Powerpoint slides]. Retrieved June 17, 2016, from <u>http://cstars.metro.ucdavis.edu/files/3613/4419/0702/Lecture\_3-Leaves\_Plants.pdf</u>
- Ustin, S. L., & Santos, M. J. (2010) Spectral Identification of Native and Non-Native Plant Species. University of California, Davis, n/a.

Van Aardt, J. A. N. (2000). Spectral separability among six southern tree species (Dissertation). Virginia Polytechnic Institute and State University.

### Appendix A

Standard Error.

Compute the standard error (SE) of the sampling distribution. SE = sqrt[ (s12/n1) + (s22/n2) ]

where s1 is the standard deviation of sample 1, s2 is the standard deviation of sample 2, n1 is the size of sample 1, and n2 is the size of sample 2.

Degrees of Freedom.

```
The degrees of freedom (DF) is:

DF = (s12/n1 + s22/n2)2 / \{ [(s12 / n1)2 / (n1 - 1)] + [(s22 / n2)2 / (n2 - 1)] \}
```

If DF does not compute to an integer, round it off to the nearest whole number. Some texts suggest that the degrees of freedom can be approximated by the smaller of n1 - 1 and n2 - 1; but the above formula gives better results.

Test Statistic.

The test statistic is a t statistic (t) defined by the following equation. t = [(x1 - x2) - d] / SE

where x1 is the mean of sample 1, x2 is the mean of sample 2, d is the hypothesized difference between population means, and SE is the standard error.

# **Appendix B**

Inserting Metadata from .csv File into Access DBMS.

```
INSERT INTO Cutting (CutID, Species, [Life Form], [Cut Date], Location,
Latitude, Longitude, [Test Date], [Spectrometer User])
SELECT [Vegetation Metadata - Sheet1].ID, [Vegetation Metadata -
Sheet1].[Cutting ID], [Vegetation Metadata - Sheet1].[Life Form],
[Vegetation Metadata - Sheet1].[Cut Date], [Vegetation Metadata -
Sheet1].Location, [Vegetation Metadata - Sheet1].Latitude,
[Vegetation Metadata - Sheet1].Longitude, [Vegetation Metadata -
Sheet1].[Test Date], [Vegetation Metadata - Sheet1].[Spectrometer
User]
FROM [Vegetation Metadata - Sheet1];
```

Removing Unnecessary Data from Imported .csv File.

```
DELETE *
FROM Conversion
WHERE Conversion.Wavelength Is Null;
```

#### Creating a New Sample for a Cutting.

```
INSERT INTO Sample ( CutID, SampleID )
SELECT Cutting.CutID, 1
FROM Cutting;
```

#### Adding Reflectance Data into *Spectrum* for each *Sample*.

```
INSERT INTO Spectrum (CutID, Wavelength, [Reflectance Value], SampleID)
SELECT [Conversion].CutID, [Conversion].Wavelength,
       [Conversion].Reflectance1, 1
FROM Conversion;
```