

**PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Eric Jonathan Reichard

Entitled
Chemometrics Applied to the Discrimination of Synthetic Fibers by Microspectrophotometry

For the degree of Master of Science

Is approved by the final examining committee:

John V. Goodpaster

Chair

Robert Minto

Rajesh Sardar

Jay A. Siegel

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): John V. Goodpaster

Approved by: Christine Picard

Head of the Graduate Program

3/22/2013

Date

CHEMOMETRICS APPLIED TO THE DISCRIMINATION OF SYNTHETIC
FIBERS BY MICROSPECTROPHOTOMETRY

A Thesis

Submitted to the Faculty

of

Purdue University

by

Eric Jonathan Reichard

In Partial Fulfillment of the
Requirements for the Degree
of
Master of Science

May 2013

Purdue University

Indianapolis, Indiana

For the love of my life, Tina

ACKNOWLEDGMENTS

I would first like to acknowledge and thank my advisor and mentor, Dr. John V. Goodpaster for giving me the opportunity to accomplish my academic goals. I would also like to thank Dr. Stephen L. Morgan from the University of South Carolina for introducing me to forensics and being a collaborator with my research while here at IUPUI. I would like to give special thanks to Dana Bors, Wil Kranz, and Maria Diez for acquiring spectra for the external validation. My research was supported by Award No. 2010-DN-BX-K220 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect those of the Department of Justice. Finally, I would like to thank all my family, friends, professors, and the members of the Goodpaster group for all the advice and assistance during my studies.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ABBREVIATIONS	ix
ABSTRACT.....	xi
CHAPTER 1. INTRODUCTION TO FIBERS AND CHEMOMETRICS	1
1.1. Textile Fibers.....	1
1.1.1. Natural and Manufactured Textile Fibers.....	1
1.1.2. Fiber Dyes.....	3
1.1.3. Forensic Fiber Analysis.....	4
1.2. Chemometrics	6
1.2.1. Application of Chemometrics to Forensic Science.....	6
1.2.2. Preprocessing Techniques	8
1.2.3. Agglomerative Hierarchical Clustering.....	10
1.2.4. Principal Component Analysis	12
1.2.5. Discriminant Analysis.....	15
CHAPTER 2. CHEMOMETRIC ANALYSIS OF BLUE ACRYLIC VISIBLE SPECTRA.....	17
2.1. Introduction and Purpose	17
2.2. Materials and Methods	18
2.2.1. Materials	18
2.2.2. Instrumental Analysis.....	20
2.2.3. Data Analysis	20
2.3. Results and Discussion	21
2.3.1. Training Set.....	21
2.3.2. External Validation	32
2.4. Conclusions.....	35
CHAPTER 3. MICROSPECTROPHOTOMETRIC ANALYSIS OF YELLOW POLYESTER FIBER DYE LOADINGS WITH UTILIZATION OF CHEMOMETRIC TECHNIQUES.....	37
3.1. Introduction and Purpose	37

	Page
3.2. Materials and Methods	37
3.2.1. Materials	37
3.2.2. Instrumental Analysis	39
3.2.3. Data Analysis	39
3.3. Results and Discussion	41
3.3.1. Calibration Plots	43
3.3.2. Training Set.....	44
3.3.3. External Validation	54
3.3.4. Pair-Wise Comparisons	56
3.4. Conclusions	57
CHAPTER 4. LIMITATIONS AND FUTURE WORK.....	58
4.1. Limitations	58
4.2. Future Work.....	59
LIST OF REFERENCES.....	61
APPENDIX. ADDITIONAL FIBER FIGURES	69
A.1. Blue Acrylic Fibers	69
A.1.1. Training Set Exemplar Spectra.....	69
A.1.2. External Validation Exemplar Spectra	75
A.2. Yellow Polyester Fibers	81
A.2.1. Calibration Plots	81
A.2.2. Training Set Exemplar Spectra.....	84
A.2.3. External Validation Exemplar Spectra	89
A.2.4. PCA Projections of Pair-Wise Comparison Data	94

LIST OF TABLES

Table	Page
Table 2.1. Images of the representative exemplars	18
Table 2.2. Naming system used for eleven exemplars along with their dye compositions.....	19
Table 2.3. Cross-validation confusion matrix of the training set.....	31
Table 2.4. Confusion matrix of the prediction set	33
Table 3.1. Exemplars A-E with respective dye loadings in weight percent and images	38
Table 3.2. Exemplars F-J with respective dye loadings in weight percent and images	38
Table 3.3. Calibration curve results for three unknowns using a statistical and non-statistical approach.....	43
Table 3.4. Cross-validation confusion matrix of the training set	52
Table 3.5. Cross-validation confusion matrix of classes generated from dendrogram	54
Table 3.6. External validation results.....	55
Table 3.7. Pair-wise comparison results.....	56

LIST OF FIGURES

Figure	Page
Figure 2.1. Representative spectra of the eleven blue acrylic fibers	22
Figure 2.2. AHC dendrogram showing the three classes of averaged fibers	23
Figure 2.3. Normalized central objects plot of the three AHC classes	24
Figure 2.4. Projections of the data in the first and second principal component.....	25
Figure 2.5. Projections of the data in the first and third principal component.....	26
Figure 2.6. Projections of the data in the second and third principal component.....	26
Figure 2.7. Factor loadings plot of the first two principal components.....	27
Figure 2.8. Regions of high correlation (factor loadings) superimposed over the fiber spectra	28
Figure 2.9. Projections of the data in the first two canonical variates.....	29
Figure 2.10. Projections of the data in the first and third canonical variate	29
Figure 2.11. Projections of the data in the second and third canonical variate	30
Figure 2.12. Projections of the data in the first two canonical variates of the prediction set	34
Figure 2.13. Projections of the data in the first and third canonical variate of the prediction set	34
Figure 2.14. Projections of the data in the second and third canonical variate of the prediction set	35
Figure 3.1. Fiber spectra with adjusted absorbance values for A) background subtracted and normalized data and B) background subtracted only data	42
Figure 3.2. Magnified view of the shift of the absorbance maximum as dye loading increases	45
Figure 3.3. AHC dendrogram of the ten exemplars from the training set	46

Figure	Page
Figure 3.4. AHC central objects plot of the three classes.....	47
Figure 3.5. Projections of the data in the first two principal components	48
Figure 3.6. Projections of the data in the first two principal components of the class data	48
Figure 3.7. Factor loadings plot of the first two principal components.....	49
Figure 3.8. Regions of high correlation (factor loadings) superimposed over the three classes of fiber spectra from AHC	50
Figure 3.9. Projections of the data in the first two canonical variates.....	51
Figure 3.10. Projections of the data in the first two canonical variates of the classes generated from the dendrogram.....	53

LIST OF ABBREVIATIONS

(w/w)	weight percent
X	magnification
AA	acrylamide
AHC	agglomerative hierarchical clustering
ATR	attenuated total reflectance
AUC	area under the curve
CE	capillary electrophoresis
CI	colour index
cm ⁻¹	wavenumber
CV	canonical variate
DA	discriminant analysis
DHC	divisive hierarchical clustering
FTIR	Fourier transform infrared spectroscopy
IR-MALDESI	infrared matrix-assisted laser desorption electrospray ionization
LC-MS	liquid chromatography-mass spectrometry
LDA	linear discriminant analysis
MA	methyl acrylate
MMA	methyl methacrylate
MSP	microspectrophotometry
NAS	National Academy of Sciences
nm	nanometer
PAN	polyacrylonitrile
PBT	polybutylene terephthalate
PC	principal component

PCA	principal component analysis
PCR	principal component regression
PEN	polyethylene naphthalate
PET	polyethylene terephthalate
PLM	polarized light microscopy
PPT	polytrimethylene terephthalate
QDA	quadratic discriminant analysis
ROC	receiver operating characteristic
SIMCA	soft independent modeling of class analogy
Std. Dev.	standard deviation
SWGMAAT	Scientific Working Group on Materials Analysis
TLC	thin layer chromatography
TOF-SIMS	time-of-flight-secondary ion mass spectrometry
VA	vinyl acetate

ABSTRACT

Reichard, Eric Jonathan. M.S., Purdue University, May 2013. Chemometrics Applied to the Discrimination of Synthetic Fibers by Microspectrophotometry. Major Professor: John V. Goodpaster.

Microspectrophotometry is a quick, accurate, and reproducible method to compare colored fibers for forensic purposes. The use of chemometric techniques applied to spectroscopic data can provide valuable discriminatory information especially when looking at a complex dataset. Differentiating a group of samples by employing chemometric analysis increases the evidential value of fiber comparisons by decreasing the probability of false association. The aims of this research were to (1) evaluate the chemometric procedure on a data set consisting of blue acrylic fibers and (2) accurately discriminate between yellow polyester fibers with the same dye composition but different dye loadings along with introducing a multivariate calibration approach to determine the dye concentration of fibers. In the first study, background subtracted and normalized visible spectra from eleven blue acrylic exemplars dyed with varying compositions of dyes were discriminated from one another using agglomerative hierarchical clustering (AHC), principal component analysis (PCA), and discriminant analysis (DA). AHC and PCA results agreed showing similar spectra clustering close to one another. DA analysis indicated a total classification accuracy of approximately 93% with only two of the eleven exemplars confused with one another. This was expected because two exemplars consisted of the same dye compositions. An external validation of the data set was performed and showed consistent results, which validated the

model produced from the training set. In the second study, background subtracted and normalized visible spectra from ten yellow polyester exemplars dyed with different concentrations of the same dye ranging from 0.1-3.5% (w/w), were analyzed by the same techniques. Three classes of fibers with a classification accuracy of approximately 96% were found representing low, medium, and high dye loadings. Exemplars with similar dye loadings were able to be readily discriminated in some cases based on a classification accuracy of 90% or higher and a receiver operating characteristic area under the curve score of 0.9 or greater. Calibration curves based upon a proximity matrix of dye loadings between 0.1-0.75% (w/w) were developed that provided better accuracy and precision to that of a traditional approach.

CHAPTER 1. INTRODUCTION TO FIBERS AND CHEMOMETRICS

1.1. Textile Fibers

The Locard Exchange Principle states that when two objects come into contact, there is always a transfer of material.¹ This principle is especially relevant to trace evidence such as textile fibers. Fibers can be exchanged between two individuals, between an individual and an object, and between two objects. This exchange can either occur as a direct transfer or an indirect transfer. Fiber persistence is another important factor, which will determine whether or not a fiber will be found after a transfer. There are numerous factors that will determine the number of fibers lost and the rate of loss. Studies have shown that the initial rate of fiber loss is rapid. For example, in some studies, 18 percent or less of fibers remained after only two hours.² It is seen that transfer and persistence of fibers are two key factors that will determine the significance of fiber associations.

1.1.1. Natural and Manufactured Textile Fibers

A textile fiber is a unit of matter that has a length that is at least 100 times its diameter that forms the basic element of fabrics.³ Fibers can be classified as either natural or man-made. A natural fiber exists in a largely unaltered state and can come from a plant, animal, or mineral. Plant fibers can originate from the seed, stem, or leaf. The most common plant fibers include cotton, jute, flax, hemp, and sisal.³ Animal fibers are typically made from animal hairs, therefore, are made up of proteins. There are three main types of hair produced by animals: whiskers, guard, and fur. Guard hairs are the most useful when identifying the species of animal. Some examples of animal fibers include wool,

camel, and rabbit. It is important to note that silk, which is produced by the silkworm (*B. mori*), is considered an animal fiber, but it consists of fibroin fiber proteins instead of keratin fiber proteins like that of fur bearing mammals.⁴ The most common mineral fibers are asbestos. Examples of mineral fibers include chrysotile, amosite, and crocidolite.

In contrast, a man-made fiber is created from raw materials that are either natural or chemical based. Manufactured fibers made from natural materials are classified as cellulosic and manufactured fibers made from chemical polymers are classified as synthetic. Cellulosic fibers are made from regenerated or derivative cellulosic polymers like cotton or wood. Examples include acetate and rayon. Synthetic fibers consist of multiple monomers covalently linked to one another. Examples of synthetic fibers include polyester, nylon, and acrylic.

Polyester and acrylic fibers are two of the most widely produced textiles. Both polyester and acrylic fibers were used in this study and will be discussed further. Polyester is comprised of any long chain polymer composed of at least 85% by weight of an ester of a substituted aromatic carboxylic acid.³ Polyester comes in many forms, but the most successful and popular form is the polyethylene terephthalate (PET) fiber. It is composed of ester links of aliphatic (ethylene glycol) and aromatic (terephthalic acid) groups. Other common polyester fibers include polytrimethylene terephthalate (PPT), polybutylene terephthalate (PBT), and polyethylene naphthalate (PEN). Acrylic, also referred to as polyacrylonitrile (PAN), is comprised of any long chain polymer composed of at least 85% by weight of acrylonitrile units.³ The other 15% or less is comprised of methyl acrylate (MA), methyl methacrylate (MMA), acrylamide (AA), and/or vinyl acetate (VA) to create a copolymer. These monomers are added to the acrylonitrile backbone in order to improve the dyeability of the fiber.

1.1.2. Fiber Dyes

Dyeing is the process of imparting color to a textile fiber, which can provide discriminating characteristics for qualitative comparison purposes. Dyes are molecules that contain chromophores and auxochromes.⁵ A chromophore is a simple unsaturated group attached to benzene or fused benzene rings. There are two groups of chromophores, one containing π -bonds next to σ -bonds (double and triple bonds) and another containing non-bonding n-electrons (azo groups, cyano groups, carbonyl groups). Auxochromes, which increase the depth of the color and allow the dye molecule to bond to a fiber, are basic salt-forming groups like hydroxyl groups and amino groups. The dye produces a color in the visible region of the electromagnetic spectrum due to the arrangement of the π -electrons and n-electrons in its chromophores.⁵ These locations of high electron density decrease the gap between the ground state and excited states to allow for energy transitions within the visible region.

Fiber dyes can be classified according to their method of application, chemical class, or the type of fiber they are applied to. There are nine general dye classes: acid, basic, azoic, direct, disperse, metallized, reactive, sulfur, and vat.^{6,7} Acid dyes are applied under acidic conditions. Negatively charged functional groups on the dye molecule form ionic bonds with positively charged functional groups on the fiber substrate. Typical fiber substrates that are treated with acid dyes include wool, silk, polyamide, and polyacrylonitrile. Basic dyes are also applied under acidic conditions. In this case, however, the cationic dye forms an ionic bond with the anionic fiber functional groups. These dyes are applied to polyacrylonitrile, polyester, polyamide, and polypropylene. Azoic dyes are applied to cotton and viscose via coupling between a stabilized diazonium salt and a coupling component like naphthol.⁷ Direct dyes are mostly applied under slightly alkaline conditions to cellulosic fibers by direct incorporation in the presence of heat and an electrolyte. Disperse dyes are insoluble in water and are directly incorporated into polyester, polyacrylonitrile, polyamide, polypropylene, and acetate/triacetate fibers. High temperatures or the presence

of a carrier is needed to apply the dye, which is held onto the fiber via weak van der Waals forces and hydrogen bonding.⁷ Metallized dyes form metal complexes through the reaction of a mordant (metal) that is either applied before, after, or at the same time as the dye.⁶ Fibers that are dyed with metallized dyes include wool and polypropylene. Reactive dyes are applied to cotton, wool, and polyamide fibers. They react chemically to form covalent bonds with functional groups on the fiber. Sulfur dyes are applied to cellulosic fibers. The dye is chemically altered by a reducing agent into a soluble form where it penetrates the fiber. Once incorporated into the fiber, the soluble dye oxidizes back into its insoluble form. Vat dyes utilize a similar process to that of sulfur dyes where a reducing agent is used to form the soluble form and oxidation occurs within the fiber to form the original insoluble dye.⁶

1.1.3. Forensic Fiber Analysis

More often than not a forensic fiber examiner is requested to compare a known and questioned fiber to determine if the questioned fiber could have come from the known source. Textile fibers can be compared based on their macroscopic and microscopic characteristics, optical characteristics, chemical composition, and color.^{1,8-10} There are a variety of techniques that rely on microscopy, spectroscopy, chromatography, and mass spectrometry that the examiner can utilize in order to make a comparison.

Techniques used for fiber type comparisons can include stereomicroscopy, polarized light microscopy (PLM)¹¹, Fourier transform-infrared spectroscopy (FT-IR)^{12,13}, Raman spectroscopy¹⁴, and pyrolysis gas chromatography coupled with mass spectrometry¹⁵. Stereomicroscopy is primarily used to locate and recover fibers of interest. A stereomicroscope can also be used to identify certain natural fibers like cotton. PLM is primarily used for synthetic fibers and utilizes polarized light to characterize those fibers based on their optical characteristics like refractive index, birefringence, and sign of elongation. FT-IR can determine the chemical composition of a fiber based on

different vibrations of its functional groups when exposed to infrared light. Raman spectroscopy is considered a complement to FT-IR. This technique uses inelastic light scattering to characterize functional groups on the fiber. Raman spectroscopy has the advantage of characterizing not only the fiber polymer, but also the dye applied to that fiber.¹⁴ Pyrolysis gas chromatography coupled with mass spectrometry is used in some cases to determine the type of synthetic fiber, however, this technique can suffer from irreproducible results.¹⁶

Although the comparison of fiber polymers has discriminating power, the color of the fiber, which is attributed to the dye applied, can be the most important characteristic when comparing two fibers. Techniques used for fiber dye and color comparisons can include thin-layer chromatography (TLC)^{17,18}, UV-visible microspectrophotometry (MSP)⁵, liquid chromatography-mass spectrometry (LC-MS)^{19,20}, and capillary electrophoresis (CE)²¹. These techniques, except for MSP, require some sort of extraction of the dye from the fiber. An examiner will try to avoid these techniques or utilize them last due to their destructive nature. TLC, LC-MS, and CE also require correct extraction and separation solvents and methods in order to identify the dye(s) depending on the fiber and/or dye in question. This can cause difficulties especially when the sample is limited in amount. Recent research has been conducted to solve the problem of extracting thus destroying fiber evidence. Zhou et al.²² developed a method for dye identification utilizing time-of-flight-secondary ion mass spectrometry (TOF-SIMS). This method shows promise, but requires long sample preparation times and has only been optimized for acid dyes on nylon fibers.

UV-visible microspectrophotometry is a quick, accurate, reproducible, and non-destructive technique used by forensic fiber analysts to examine the color of dyed fibers. Humans are able to perceive color, however, color measurements between individuals is subjective. Other factors can influence color like lighting conditions and the phenomenon called metamerism. Metamerism occurs when two fibers are dyed with different dyes or combinations of dyes, but the perceived

color of the two fibers is the same. Visual differences can be seen between two metameric pairs of fabric under different lighting conditions, however, metameric pairs of single fibers cannot be visually discriminated, thus MSP is a vital technique in the fiber color analysis scheme. A microspectrophotometer is composed of two parts: a microscope and a spectrometer.¹⁰ The microscope gathers light from the sample and the spectrometer measures the change in light intensity as a function of wavelength. MSP can discriminate between two colored fibers that are visually similar based upon the different chromophores in the dye's molecular structure. Research in color comparisons with MSP have been conducted and show the viability of this technique.^{1,23-25} There are limitations to this technique, however. Resultant spectra tend to be broad and limited in features, although the first derivatives of the spectra can be taken to ascertain more information for comparative purposes.²⁶ First derivatives, however, can magnify the noise in the spectra, which could lead to harder interpretation. Quantitative analysis of the dye(s) applied to the fiber is also limited. For this reason, microspectrophotometric analysis of dyed fibers is used primarily for comparison purposes. Finally, lightly dyed fibers and darkly dyed fibers create issues due to the limits of the detector.

1.2. Chemometrics

1.2.1. Application of Chemometrics to Forensic Science

Forensic scientists are familiar with statistics that utilize one variable. An example of this would be comparing a known and unknown glass fragment based upon their refractive indices. Until recently, however, the use of multivariate statistics has been overlooked. Multivariate statistics, also known as chemometrics when applied to chemical data (e.g., spectra or chromatograms), is a form of statistics that utilizes multiple variables to describe complex datasets. Forensic scientists are often tasked with identifying patterns as well as

interpreting any differences between spectra. Currently, this is carried out by visual inspection and comparison by the examiner. The problem arises when more than three variables (dimensions) are used as with a collection of absorbance spectra, which are often contain hundreds or thousands of wavelengths. Although a trained examiner can locate the presence or absence of major peaks, subtle differences within the complex data set can be virtually impossible to find. This especially holds true when there are numerous samples to be compared.

Chemometrics has the ability to identify patterns and groupings from large complex datasets more accurately than visual examination alone. It can also investigate the dependence among variables, make predictions, and be used for hypothesis testing.²⁷ Chemometrics does this by extracting information from large data sets, which in turn allows for easier interpretation. It is important to note that multiple replicate samples must be acquired to obtain a valid conclusion from the data set. Since its emergence into the forensic science arena it has been applied to a number of sample types, including accelerants, document examination, drug analysis, fibers, inks, glass, gunpowder, paint, soil, and condom lubricants.²⁷

Visual comparison of trace evidence can be quite subjective. There is no statistical basis for the conclusions reached by the examiner. This is a concern for crime laboratories due to the issues of reliability and relevance of scientific evidence raised in the case of *Daubert v. Merrell Dow Pharmaceuticals*.²⁸ Chemometric analysis of multivariate data often found in trace evidence could help meet the *Daubert* requirements.²⁷ The use of chemometric techniques can also address two recommendations laid out in the National Academy of Sciences (NAS) report. Chemometrics could alleviate the issues of accuracy, reliability, and validity in trace evidence analysis (Recommendation 3) and assist in research on sources of human error in trace evidence analysis (Recommendation 5).²⁹

There are many multivariate techniques that could be applied to spectroscopic data. The three techniques utilized for this study were agglomerative hierarchical clustering (AHC), principal component analysis (PCA), and discriminant analysis (DA). Hierarchical clustering algorithms were created in the 1950's. The theory behind PCA was established by Pearson in 1901, however, the algorithm to compute principal components (PCs) was not introduced until 1933 by Hotelling due to the lack of machine computing.²⁷ Discriminant Analysis was first derived by Fisher in 1936.

1.2.2. Preprocessing Techniques

Preprocessing is simply defined as any mathematical manipulation of the data prior to multivariate statistical analysis.³⁰ Preprocessing the data before multivariate statistical analysis is often required to remove or reduce random or systematic sources of variation in the data set. This allows for easier interpretation of the data. Improper techniques applied to the data could remove important variation, so care must be taken when choosing the appropriate technique. There are two ways the data can be preprocessed before analysis: sample preprocessing or variable preprocessing. Sample preprocessing operates on one sample at a time over all variables. Variable preprocessing operates on one variable at a time over all samples. There are numerous methods used to preprocess the data, however, only background (baseline) correction and normalization will be discussed for sample methods and mean centering will be discussed for variable methods due to their use in the study.

Background correction reduces or eliminates a constant or systematically varying background within the data.^{27,30} There are various ways to background correct. One method, called the explicit modeling approach, involves subtracting a fitted model for a trend present in the baseline. Every spectra can be written as a function of variable number, where the function is equal to the sum of the signal of interest plus the baseline.³⁰ When the baseline has an offset baseline feature (i.e. horizontal line), one number can express the baseline, thus subtraction of

that number from the signal would remove the baseline. When a linearly sloping baseline is present, two or more points that only contain baseline information can be used to estimate a line.³⁰ To remove the sloping baseline, the estimated line is subtracted from the sample vector. Polynomials of higher magnitudes can be estimated using this approach depending on the shape of the baseline. Another method of removing the baseline takes the derivative of the spectra with respect to variable number. This approach is quite useful because it is not essential to select points that only contain baseline information.³⁰ Taking the first derivative is essentially the same as subtracting out an offset baseline via the explicit model approach.³⁰ Taking consecutive derivatives will remove all higher order baseline shapes. The four most common methods to determine the derivatives are the running simple difference, the running mean difference, the Gorry algorithm, and the Savitzky-Golay algorithm. The methods of Gorry and Savitzky-Golay are preferred because taking the derivative of a sample vector tends to propagate noise.²⁷

Normalizing the data usually comes after background correction. It removes systematic variations associated with sample size, concentration, amount of sample, and instrument response.²⁷ This is accomplished by dividing each variable of the sample by a constant. There are three common approaches to calculating a constant: normalizing to unit area, normalizing to unit length, and normalizing to maximum intensity.^{27,30} Normalizing to unit area is achieved by dividing each variable in the sample by the sum of the absolute value of all variables in that sample. The second approach, normalizing to unit length, is achieved by dividing each variable by the square root of the sum of squares of all the variables in each sample. The final approach divides each variable by the maximum value in the sample so that the maximum intensity is equal to 1.

Mean centering the data processes each variable at a time over all the samples. In simplest terms, mean centering repositions the centroid of the data set to the origin of the coordinate system by subtracting out the mean value of each variable over all the samples.³¹ This prevents data points away from the

centroid from having more influence than data points closer to the original origin. Mean centering is not always appropriate, but for principal component analysis, mean centering is recommended.³⁰

1.2.3. Agglomerative Hierarchical Clustering

Hierarchical clustering is a form of cluster analysis and is considered unsupervised because there is no prior knowledge of the underlying groupings in the data. It is performed to classify individual samples into groups or clusters based on their distances from each other. There are two types of hierarchical clustering techniques that can be employed: divisive hierarchical clustering (DHC) and agglomerative hierarchical clustering (AHC). DHC starts with all the samples in a single cluster. The single cluster is split into two smaller clusters and those clusters are then split until each sample forms its own cluster. This technique is uncommon because it is computationally demanding.³² AHC starts with each sample as its own cluster. Similar samples are clustered together until a single cluster is formed. This form of hierarchical clustering is more common and was utilized in this research. A visual representation of the clusters or groups is presented as a two dimensional plot called a dendrogram. The dendrogram, often expressed as a hierarchical tree, has the samples on the vertical axis and the dissimilarity or similarity distance on the horizontal axis. Branches, visualized as horizontal lines, represent the clusters and nodes. Vertical lines represent when two clusters are linked together.³³ A truncation line is often established to determine the significant clusters in the dendrogram. This line is determined either by the analyst or by more objective criteria.

As stated above, the interpoint distances between samples must be calculated in order to cluster similar samples together. Distance can be calculated in terms of similarity or dissimilarity. The most common type of

distance is Euclidean distance, which is based on the Pythagorean Theorem. It is the geometric distance in multidimensional space and is represented in Equation 1.1.^{33,34} Equation 1.1 is expressed in the matrix format.

$$d_{(x,y)} = [(x - y)'(x - y)]^{1/2} \quad \text{Equation 1.1}$$

The distance between points is expressed as $d_{(x,y)}$ and $(x - y)'$ is the transpose of the matrix $(x - y)$. The smaller the distance between samples the more similar the samples are to each other. Another common distance measurement is Manhattan distance. Manhattan distance is slightly different than Euclidean distance in that the sides of the triangle are summed to determine distance rather than the length of the hypotenuse. This method diminishes the effects of outliers and will always be slightly larger than that of the measurement from Euclidean distance.^{32,34} Manhattan distance is presented in Equation 1.2.³³

$$d(x,y) = \sum_i |x_i - y_i| \quad \text{Equation 1.2}$$

The correlation coefficient between samples can also be used as a distance measure. This method computes the cosine of the angle between two samples to determine their similarity. A correlation coefficient of 1 implies the two samples are very similar. This method is often used for the comparison of infrared and mass spectroscopy data.³² The last distance measurement to be discussed is Mahalanobis distance. This method is very similar to Euclidean distance except for it takes into account that some variables may be correlated. The inverse of the variance-covariance matrix is utilized as a scaling factor, which can be seen in Equation 1.3.³⁴

$$d(x,y) = [(x - y)'C^{-1}(x - y)]^{1/2} \quad \text{Equation 1.3}$$

The Mahalanobis distance is also employed in discriminant analysis when predicting the group membership of new samples, which will be discussed in

Section 1.2.5. This method is not always appropriate, especially when the number of variables exceeds the number of samples because the inverse of the variance-covariance matrix cannot be calculated.³⁴

Once the distances between samples are determined, various aggregation methods are employed to link clusters together. The most common method is single linkage. This method links clusters based on the distance between the two closest samples within each cluster. Another method, which is the opposite of single linkage, is complete linkage. Complete linkage links clusters together based on the distance between the two furthest samples within each cluster.^{30,33} The last aggregation method discussed is Ward's Method. This method utilizes an analysis of variance approach by determining the error sum of squares between any two clusters and linking the two clusters that have the least sum of squares.³³ Every possible pair of clusters that can be joined must be considered during each step. The error sum of squares is determined by measuring the total sum of squared deviations of every sample from the mean of the cluster.³⁵ Other aggregation methods can be employed like weighted and unweighted pair-group average linkage, centroid linkage, or median linkage.³³

Overall, AHC is an appropriate method when trying to determine the similarity or dissimilarity between samples in a data set. The dendrogram can provide insight into how the samples cluster as well as outlier detection. However, AHC cannot determine what variables influence certain clusters. A technique like principal component analysis (PCA) can be employed to determine such relationships. Cluster analysis, along with AHC, have been applied to inks³⁶, photocopy and printer toners³⁷, glass³⁸, soils^{39,40}, polymers^{41,42}, paint⁴³, fibers⁴⁴, hair dyes⁴⁵, and electrical tapes^{46,47}.

1.2.4. Principal Component Analysis

Principal component analysis is the most widely used multivariate technique. It is also considered an unsupervised technique because it does not require knowledge of the groupings in the data set. The purpose of PCA is to

reduce the dimensionality of the data by concentrating the total amount of variance into a smaller number of latent variables, which are linear combinations of the original variables.^{27,32} This technique provides a visual representation of the groupings of data along with information on the contributions of the original variables to the latent variables. These new variables are called principal components (PCs), and explain all or most of the total variance. Principal components are orthogonal to one another and represent directions of maximum variation in the data set.³⁰ The total number of PCs is equal to the number of samples or variables, whichever is smaller.³⁰ The first PC explains the greatest amount of variance, while successive PCs explain decreasing amounts of variance. Important PCs with eigenvalues representing systematic variation (signal) are retained, while insignificant PCs are eliminated, thus reducing the inherent dimensionality of the data.²⁷ The eigenvalues are the variances explained by the PCs and sum to the total variance.

A two or three dimensional scatter plot can be constructed to visualize the inherent groupings of the data if sufficient variation can be explained. This plot is called an observations plot, and it utilizes the first two or three PCs to plot the factor scores from one PC against the factor scores of another. These factor scores are the new coordinates determined by the PCs.²⁷ Groupings are determined based on the relative locations of each sample to one another. Samples close to one another are considered more similar than samples further away.

The contributions of the original variables to each new PC can also be represented as a factor loadings plot. The extent to which a variable contributes to a PC depends on the relative orientation in coordinate space of the PC and variable axes.³⁰ The factor loading is determined by taking the cosine of the angle between the variable axis and the PC axis. Factor loadings can range

from -1 to +1, where a factor loading of -1 indicates a strong negative correlation and a factor loading of +1 indicates a strong positive correlation between the variable and PC. A factor loading of 0 signifies no correlation between the two variables.

There are three common methods for determining the adequate amount of PCs to retain for further analyses like discriminant analysis. The first method is to simply determine a percentage of the total variability, usually 95 percent, to be retained. In this example, enough PCs would be retained so that they represent 95 percent of the total variance. Another method is the Kaiser Criterion. This criterion, proposed by Kaiser in 1960, only retains PCs with eigenvalues above 1.³³ Any eigenvalue below 1 would explain less variance than an original variable. This method often retains too many PCs. The last method, which was employed in this research, is the scree test, proposed by Cattell in 1966. The scree test utilizes a scree plot, which provides a visual representation of the decreasing variation in each principal component by plotting the eigenvalues against each principal component.^{27,33} A sudden break in the plot indicates the number of significant PCs to retain to the left of the plot. Any PC to the right of that break is considered noise. However, this method sometimes retains too few PCs for subsequent analysis.

Overall, PCA is a great dimensionality reduction technique that creates linear combinations of the original variables. Inherent groupings of the data can be visualized if a significant portion of the variation in the data set can be explained in two or three PCs. This technique also explains what original variables contributed the most to the new PC axes via factor loadings. PCA has been applied to the analysis of accelerants^{48,49}, fibers^{44,50,51}, gun shot residue⁵², documents⁵³, hair dyes⁴⁵, glass^{38,54}, inks^{36,55}, electrical tape^{46,47}, paints and clear coats^{56,57}, soils⁴⁰, and lubricants⁵⁸.

1.2.5. Discriminant Analysis

Unlike AHC and PCA, discriminant analysis (DA) is a supervised technique, which means prior knowledge of the group memberships in the data set is required. The purpose of DA is to visualize groupings in the data in two- or three-dimensions and to predict the group membership of new samples.³³ DA is similar to PCA in that new axes called canonical variates (CVs) are created by taking linear combinations of the original variables. Instead of determining directions of greatest variation like in PCA, CVs are constructed to maximize discrimination between groups of samples. This is achieved by maximizing a new criterion called the Fisher ratio. It is defined as the ratio of between group variance to within group variance.^{27,34} These new axes can then be plotted against each other to visualize the groupings of the data.

After new discriminant scores are established from the calculated CVs, the samples are classified into a group based on their Mahalanobis distance to the centroid of a particular group. The sample is classified into the group, which gives the smallest Mahalanobis distance. It is important to note that in order to compute the Mahalanobis distance, the number of samples must be greater than the number of variables because otherwise the inverse of the variance-covariance matrix cannot be calculated. PCA is usually performed before DA in order to reduce the number of variables.²⁷

The accuracy of the classification scheme can be estimated via various cross-validation methods. The resubstitution method utilizes the entire data set as a training set and creates a classification model based on the known class membership of each sample. The class membership of every sample in the data set is then predicted based on that model by its Mahalanobis distance. This method tends to overestimate the classification accuracy because resubstitution uses the same data to construct the classification model and estimate its accuracy.²⁷ The second method is the hold-out method. This method partitions the data set into a training and test set. The training set is used to construct the classification model and the test set is used for prediction of classification. The

method provides an unbiased way of estimating the error, however, this requires a large data set, which may not always be possible.²⁷ The third method, which was utilized in this research, is the leave-one-out cross validation method. In this method, the classification model is created using all but one of the samples. The left out sample is then added into the model and is classified to a group. This step is repeated for every sample until all samples in the dataset are classified. The results from the three methods can be placed in a table called a classification or confusion matrix. The classification accuracies of each group along with the total classification accuracy of all the groups are displayed.

Overall, DA is an appropriate method to discriminate between groups of samples and predict group memberships of new samples. DA can be performed on two groups or multiple groups depending on the data set, however, the number of canonical variates cannot exceed the number of groups minus 1 or the number of variables.²⁷ The most common way of determining group membership is by calculating the Mahalanobis distance of each sample to the centroid of each group and placing that sample into the group with the smallest distance. DA has been applied to the analysis of lubricants⁵⁸, fibers and dyes^{50,59}, paints and clear coats^{56,57}, gun shot residue⁵², fuel oils and asphalts⁶⁰, glass⁶¹, bacteria⁶², electrical tapes^{46,47}, inks^{36,63}, soils⁴⁰, and gasoline⁶⁴.

CHAPTER 2. CHEMOMETRIC ANALYSIS OF BLUE ACRYLIC VISIBLE SPECTRA

2.1. Introduction and Purpose

The purpose of this study was to use a chemometric approach to discriminate visible spectra of blue acrylic fibers with similar dye compositions. The traditional means of discriminating between two fibers is to visually compare the fibers' spectra acquired from a microspectrophotometer. An examiner compares the position of the peak maxima, peak width, and peak intensity to determine if significant differences are exhibited between two fibers.⁶⁵ The use of chemometrics can provide a better understanding of subtle differences within the data that an examiner could not see with the naked eye. This study is by no means trying to eliminate human judgment in favor of chemometrics for fiber color comparisons. It is, however, a good complement to the established protocols utilized in crime laboratories across the United States and can further assist the examiner in difficult fiber comparisons. Research by Cochran et al. has shown that dye identification can be achieved by direct analysis of a textile fabric utilizing an infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI) source for mass spectrometry.⁶⁶ Microspectrophotometric analysis would be all but obsolete due to the discriminating power of this new technique, however, it is not optimized for single fiber analysis, it can be destructive to the sample depending on the fiber polymer, and only a few dyes yielded acceptable spectral results.

2.2. Materials and Methods

2.2.1. Materials

Eleven blue acrylic fibers with a bilobal cross-section were provided by Dr. Stephen L. Morgan from the University of South Carolina. Table 2.1 provides images of each exemplar taken via the MSP utilized for this study. These eleven exemplars were dyed with varying compositions of basic dye formulations. Table 2.2 provides the naming system employed for this study along with the CI names of the dyes used for each exemplar. The exemplars had different diameters and dye concentrations were not known.

Table 2.1. Images of the representative exemplars











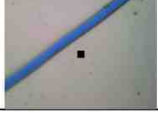
Exemplar	Image	Exemplar	Image
A		G	
B		H	
C		I	
D		J	
E		K	
F			

Table 2.2. Naming system used for eleven exemplars along with their dye compositions

EXEMPLAR	FIBER ID/DYE	Blue 3	Blue 41	Blue 60	Blue 147	Red 18	Red 29	Red 46	Yellow 21	Yellow 28	Yellow 29
A	FID 86	✓				✓				✓	
B	FID 87		✓					✓		✓	✓
C	FID 88		✓					✓		✓	
D	FID 91		✓				✓		✓		
E	FID 95				✓		✓			✓	
F	FID 98	✓			✓						
G	FID 99				✓			✓		✓	
H	FID 112	✓				✓				✓	
I	FID 113		✓				✓			✓	
J	FID 114		✓			✓				✓	
K	FID 145	✓						✓		✓	



2.2.2. Instrumental Analysis

Four fibers from each exemplar were removed and mounted on glass microscope slides using Permout (Fischer Scientific, Fairlawn, NJ) mounting media. Two additional fibers from each exemplar were removed and mounted in Permout for an external validation. Standard MSP protocols outlined by the Scientific Working Group on Materials Analysis: Fiber Subgroup (SWGMAF) were followed for data collection.⁶⁵ A CRAIC QDI 2000 microspectrophotometer (Craic Technologies, San Dimas, CA) was used in transmitted light mode at a total magnification of 150X. The microscope was calibrated by Kohler illumination and the spectrophotometer was calibrated by NIST traceable standards before each use of the instrument. Autoset optimization, a dark scan, and a reference scan were performed before each sample scan. Fifty scans were taken at a resolution factor of five for each sample spectrum as absorbance values. The wavelength range utilized was from 400 nm to 800 nm. Five spectra were obtained at different locations along each fiber to account for intra-fiber variation. A total of 20 spectra were collected for each exemplar. A different analyst used the same parameters, as above, for the two additional fibers. A total of 10 spectra were obtained for each exemplar for the external validation.

2.2.3. Data Analysis

Before use of statistical techniques, the data was preprocessed to eliminate systematic and random noise. Background subtraction was performed on each spectrum by subtracting the minimum absorbance value from all absorbance values for each sample. This eliminated the effects of scattered light and brought the baseline down to zero. Next, each spectrum was normalized to unit vector length by dividing each absorbance value by the square root of the sum of squares of all absorbance values for each sample. Normalizing the data accounted for variability due to varying fiber diameters and dye concentration.

All chemometric techniques were performed by use of XLSTAT Pro (Addinsoft, Paris, France), an add-in software program for Microsoft Excel. For

AHC, the average of the five spectra for each fiber were used to produce a readable dendrogram. The proximity measure utilized was the Euclidean distance and the aggregation method used for clustering samples was Ward's Method. The truncation line was automatically determined by a histogram of node positions. For PCA, all the spectra were utilized instead of the averaged spectra. The algorithm used was Pearson (n). A factor loadings plot and observations plots were generated from the first three principal components. Six principal components were retained for DA based upon a Scree plot. For DA, all spectra were utilized instead of the averaged spectra.

For the external validation, only PCA and DA were performed. PCA was performed to reduce the amount of variables being subjected to DA. DA was performed to predict the class memberships of the external validation samples. Otherwise, the same conditions were employed as with the training dataset.

2.3. Results and Discussion

2.3.1. Training Set

The eleven blue acrylic exemplars exhibited slight visible differences in hue and saturation (see Table 2.1). While differences in color could be seen, human judgment of color is subjective. More information was obtained by looking at the representative spectra of the eleven exemplars in Figure 2.1. A noticeable difference in the behavior of the peaks was seen. The absorbance maximum of all eleven exemplars ranged from 600 nm to 650 nm. Several exemplar spectra contained a second peak, while others exhibited shoulders along their absorbance maximum. Exemplars A and H had overlapping spectra, which was expected due to identical dye compositions (see Table 2.2).

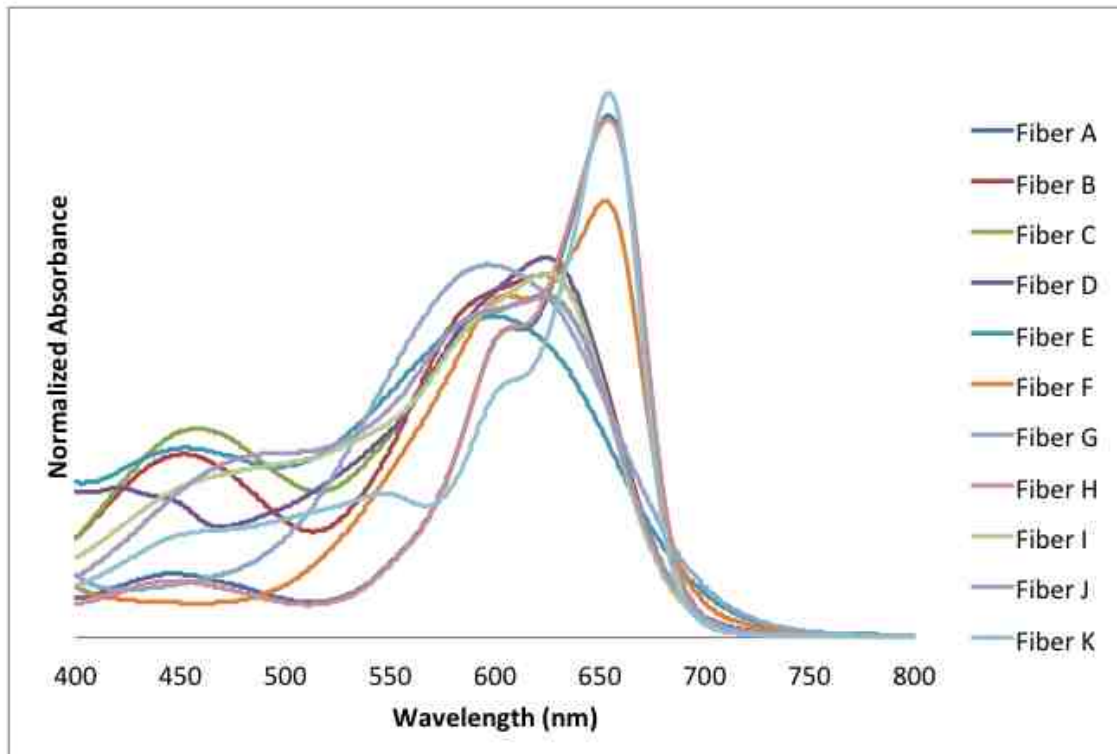


Figure 2.1. Representative spectra of the eleven blue acrylic fibers

Differences in the fibers' spectra could be seen; however, utilizing multivariate statistics can provide more information on the variation of the spectra thus providing a better understanding of the groupings of the data as well as the areas of the spectra that provide the most variation. The AHC dendrogram of the eleven exemplars is shown in Figure 2.2. The dendrogram provided a visualization of the groupings of spectra based on their distances from each other in coordinate space. From where the truncation line was set, three distinct classes were indicated.

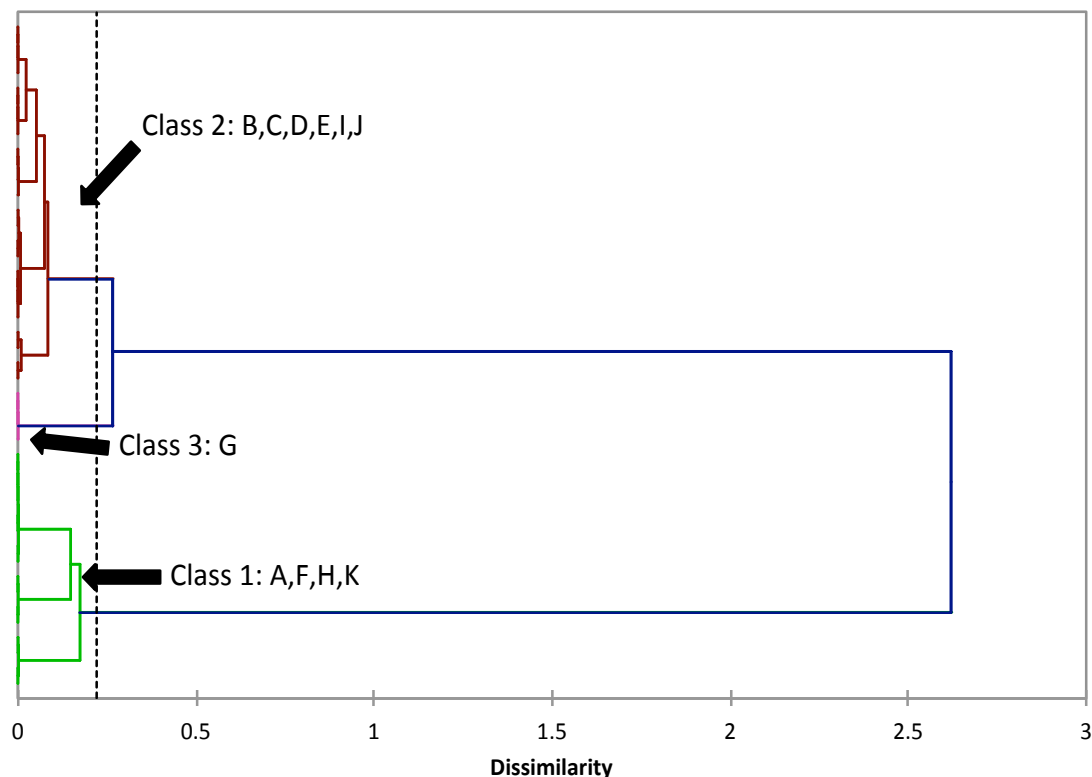


Figure 2.2. AHC dendrogram showing the three classes of averaged fibers

Class 1 was made up of Exemplars A, F, H, and K. Exemplars A and H were the most related followed by Exemplar K then Exemplar F. Replicates of Exemplars A and H were confused with each other. All other replicates for each exemplar were clustered together in the dendrogram.

Class 2 was made up of Exemplars B, C, D, E, I, and J. All the replicates of each exemplar were clustered together. Exemplars I and J were the most similar to each other followed by Exemplars B and C. Exemplar D was more similar to Exemplars B and C rather than Exemplars I and J. Exemplar E was the most dissimilar when compared to the others in Class 2.

Class 3 only contained Exemplar G, which is considered unique thus allowing discrimination of it from all others in the dataset. Figure 2.3 shows the spectra that were most similar to the average spectra for each class. Exemplars A4, I4, and G3 represent the average spectra for Classes 1, 2, and 3, respectively. Apparent differences in spectral features can be seen. Class 1 has

two peaks with a shoulder on one of the peaks. Class 2 has a peak at a lower absorbance and slightly blue-shifted when compared to Class 1 along with a minor peak at a lower wavelength. Class 3 has a single peak with a slightly higher absorbance to that of Class 2, but lower than Class 1. Differences in spectral features of the exemplars within each class are not as obvious except for Exemplar G since it was clustered by itself.

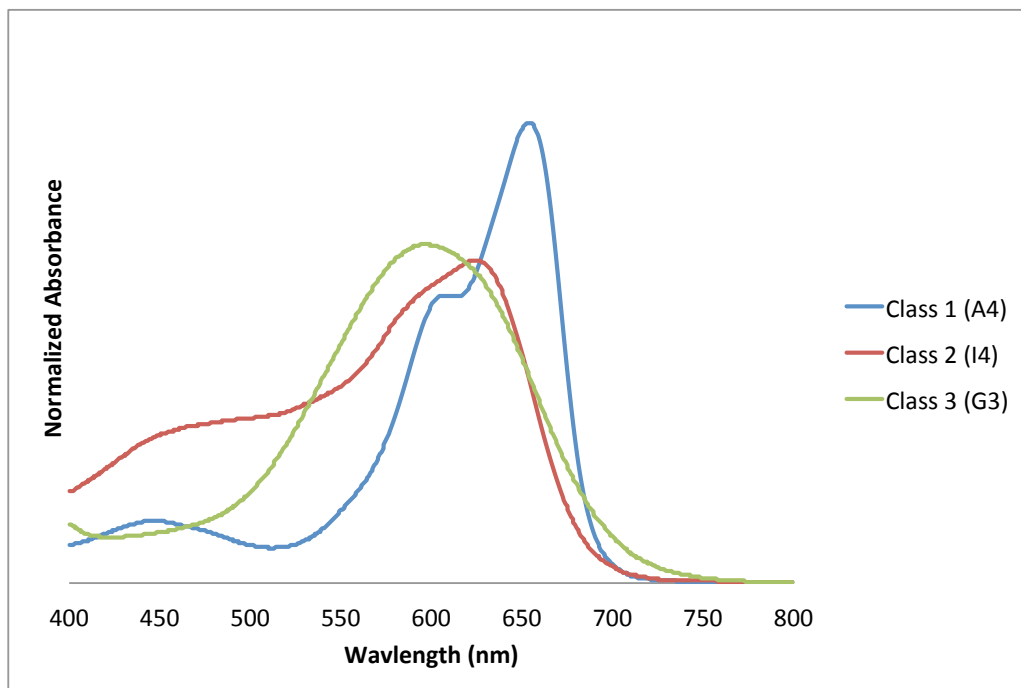


Figure 2.3. Normalized central objects plot of the three AHC classes

PCA was utilized to better understand groupings along with potential outliers and what areas of the spectra provided the most variations within the data. PCA was performed on every spectrum for each exemplar. From the observations plot, the first two principal components captured 75.69% of the total variance (see Figure 2.4). Class 1 exemplars were grouped on the right side of the plot. Exemplars within Class 1 were all grouped separately except for Exemplars A and H. Exemplar F was farthest away from the others in that class, which is similar to what was seen in the dendrogram. There was a potential

outlier for Exemplar A, but after reanalyzing the data without A4-5, no significant change in the groupings of the data was seen. Spectrum A4-5 was left in for further analysis. Class 2 exemplars were grouped on the lower left side of the plot. Exemplar E is grouped by itself, while the other exemplars are overlapping one another. Generating observation plots of PC2 vs. PC3 and PC1 vs. PC3, provides a three dimensional view of the data (see Figures 2.5 and 2.6). Utilizing these plots presents a better understanding of the separation within Class 2. Separation between exemplars can be seen, however, Exemplars I and J exhibit slight overlap. A three dimensional view does not allow for complete separation of the exemplars in Class 2; approximately 15% of the variation is unaccounted when utilizing only three PCs. Higher dimensions might provide more information. Finally Exemplar G was separated from all other exemplars, which is consistent with the dendrogram.

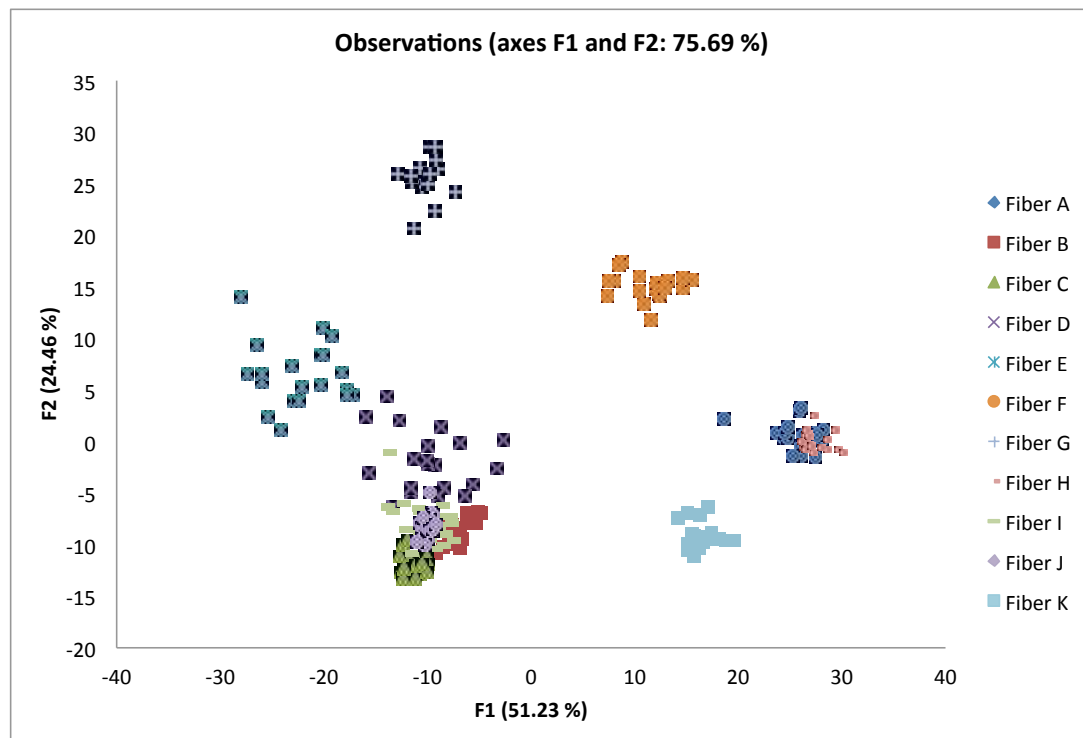


Figure 2.4. Projections of the data in the first and second principal component

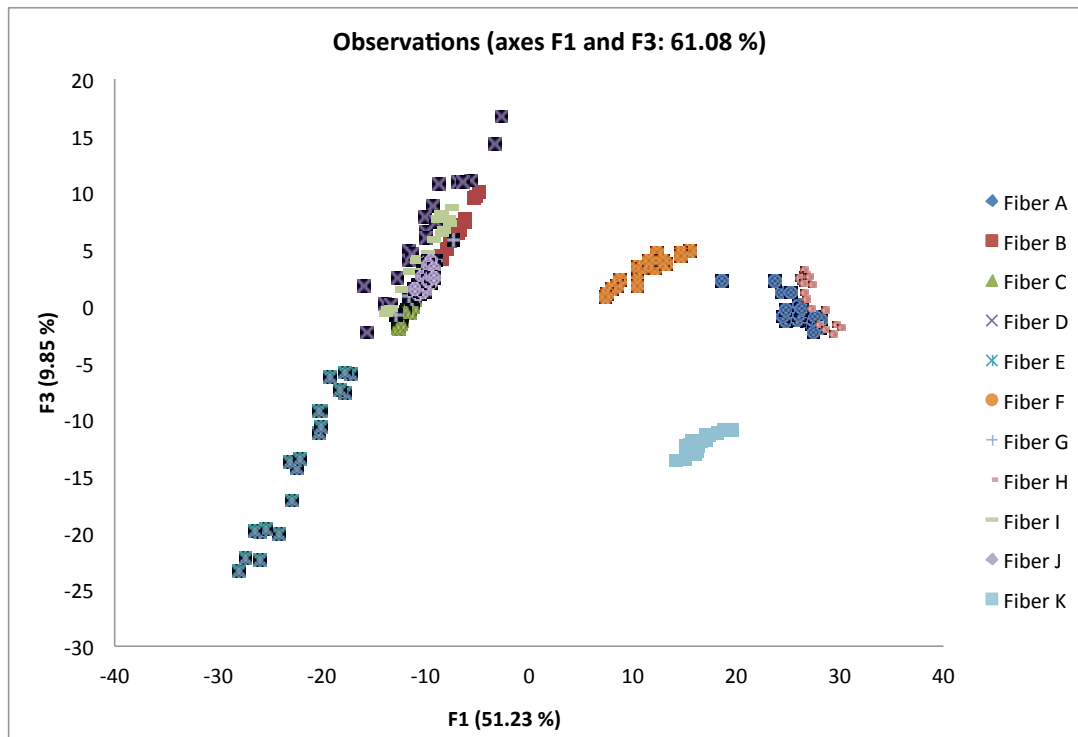


Figure 2.5. Projections of the data in the first and third principal component

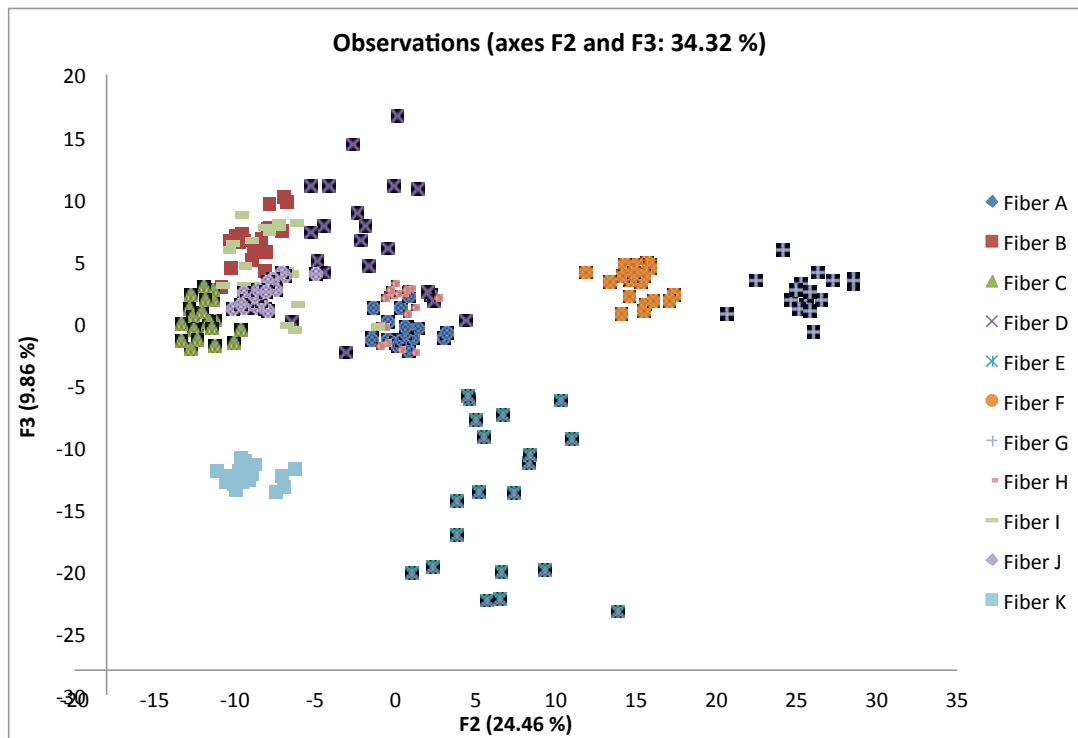


Figure 2.6. Projections of the data in the second and third principal component

A factor loadings plot, seen in Figure 2.7, plots the cosine of the angles between the original variables and principal components. This plot was used to determine spectral regions, where large variations in the data occur. This plot also expresses how the exemplars were separated on the PCA observations plot.

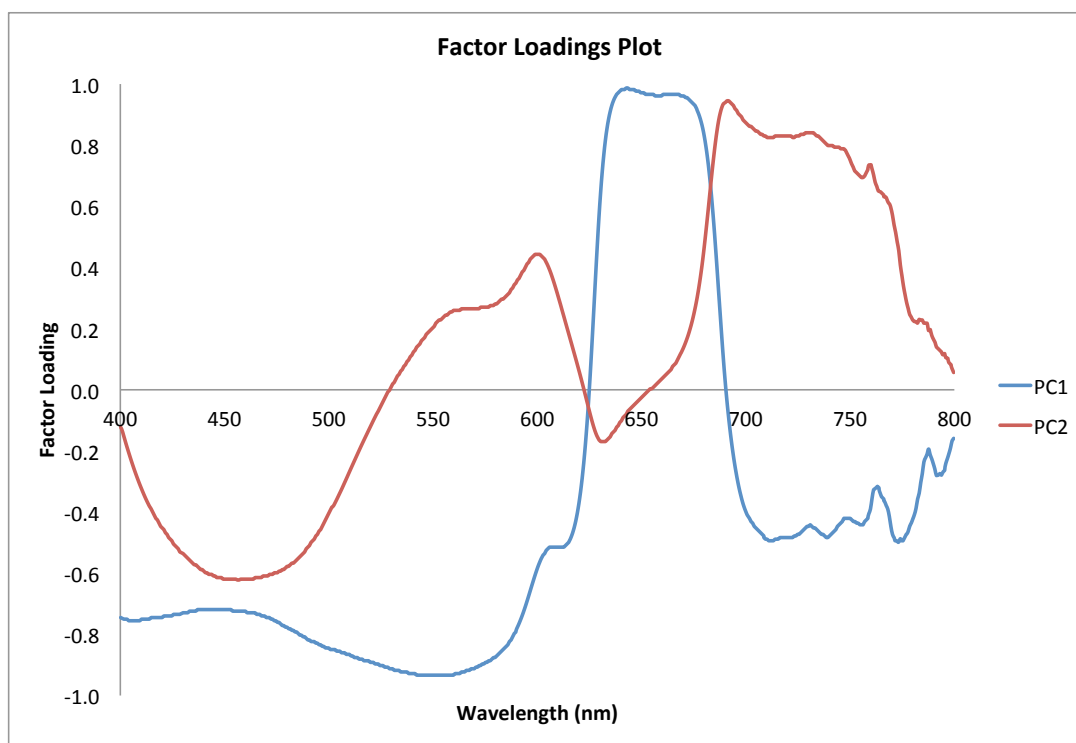


Figure 2.7. Factor loadings plot of the first two principal components

PC1 showed positive correlations between 640-675 nm and negative correlations between 500-580 nm. PC2 showed positive correlations between 690-725 nm and negative correlations between 450-475 nm. Figure 2.8 overlays these regions with the representative spectra. Class 1 was separated from the other classes based on the positive region of PC1 where the absorbance maximum of Class 1 exemplars is seen. Exemplars within Class 1 were separated based on the positive and negative regions of PC2. The trailing edges of their respective curves in the positive region have different slopes. The negative region of PC2 encompasses a peak for Exemplars A and H, which

cannot be said for Exemplars F and K. Class 2 exemplars are separated from the other classes based on the negative regions of PC1 and PC2. The negative region of PC1 exhibited differences along the leading edges of the spectral curves of the exemplars. Differences in the negative region of PC2 were seen based on the presence of a peak. As discussed above, utilizing three PCs instead of two further separated exemplars within this class. Class 3, which consisted of Exemplar G, was separated from the other classes based off of the positive region of PC2 and the positive region (not shown) of PC3. The positive region of PC2 consisted of the trailing edge of its spectral curve. The absorbance maximum of Exemplar G consisted of the positive region of PC3.

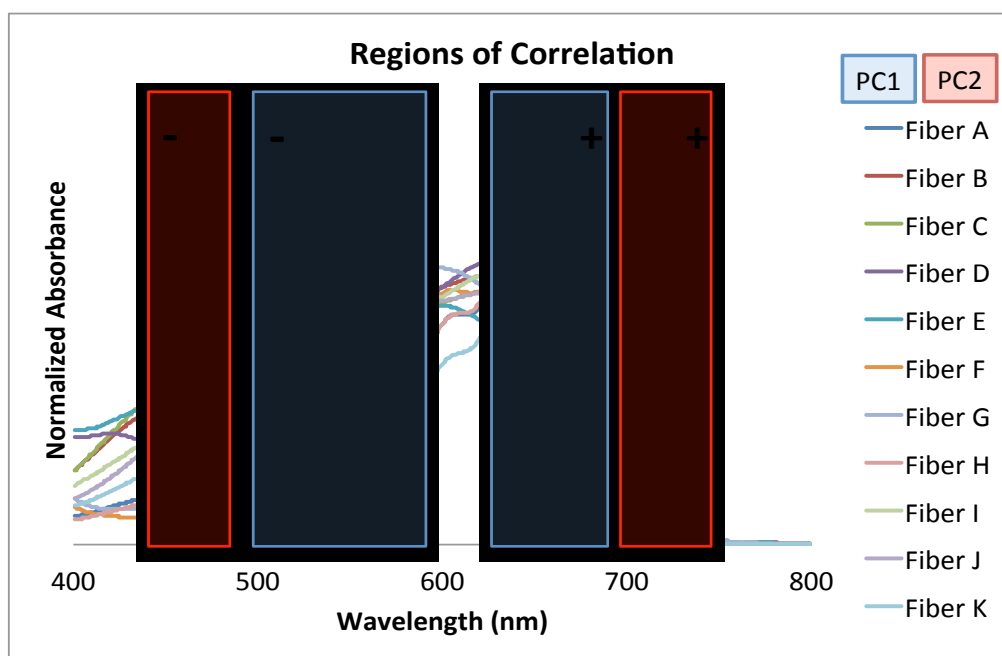


Figure 2.8. Regions of high correlation (factor loadings) superimposed over the fiber spectra

A Scree plot was used to determine the number of principal components to retain for DA and it was determined that the first six PCs were meaningful. In this analysis, each exemplar was considered to be its own class. The results from the first six PCs are shown in Figure 2.9, with 88.39% of the total variance accounted for in the first two canonical variates.

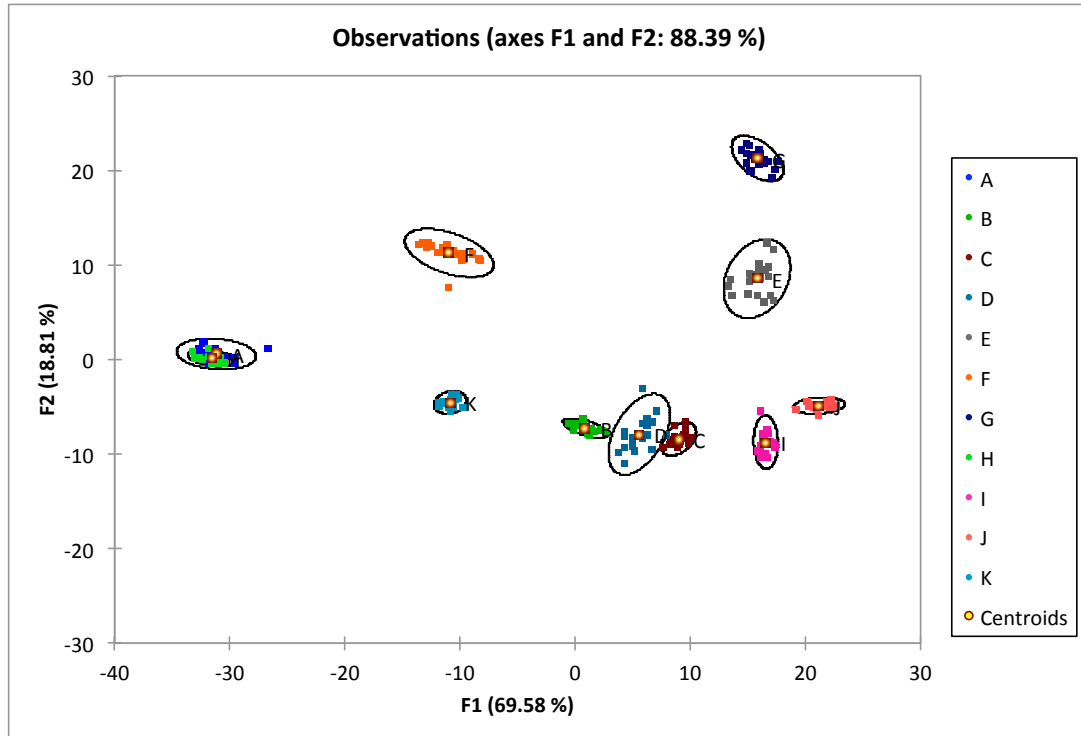


Figure 2.9. Projections of the data in the first two canonical variates

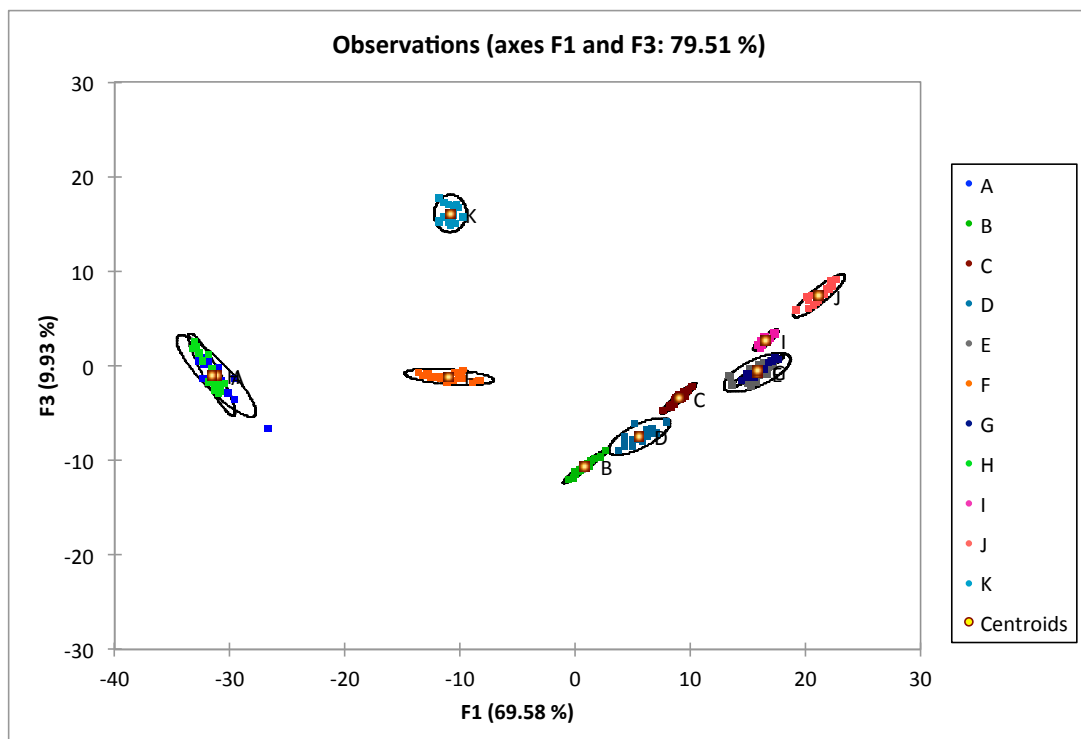


Figure 2.10. Projections of the data in the first and third canonical variate

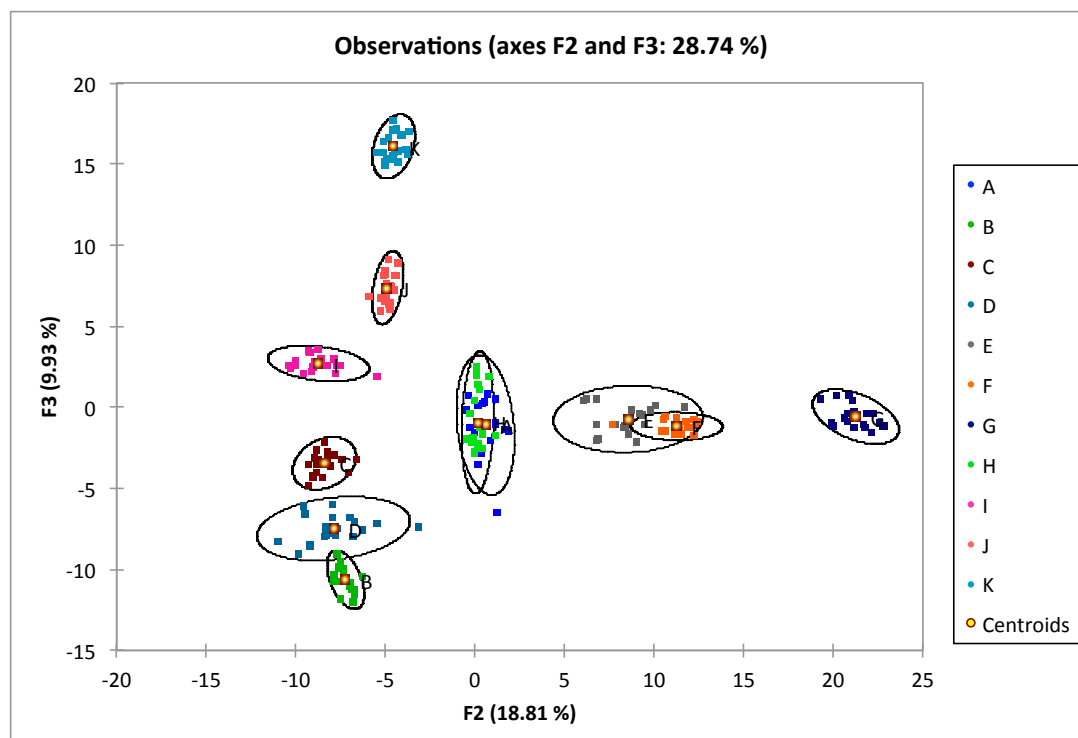


Figure 2.11. Projections of the data in the second and third canonical variate

There is overlap seen between Exemplars A and H and slight overlap between C and D. All other exemplars are separated into their own classes. Creating observation plots of CV1 vs. CV3 and CV2 vs. CV3, provides a three dimensional view of the data. After examining Figures 2.10 and 2.11, it becomes apparent that Exemplars C and D do not overlap, however, Exemplars A and H do not visually separate into two classes. A “leave-one-out” cross validation was performed to determine the classification accuracy of the model produced by DA. The generated confusion matrix, seen in Table 2.3, produced a classification accuracy of 92.73%. All Exemplars except A and H were classified correctly 100% of the time. Eight spectra from A and H were misclassified, which led to a 60% classification accuracy for both. This was expected since both exemplars had the exact same dye compositions.

Table 2.3. Cross-validation confusion matrix of the training set

Predicted ----- Actual	A	B	C	D	E	F	G	H	I	J	K	Total	% correct
A	12	0	0	0	0	0	0	8	0	0	0	20	60 %
B	0	20	0	0	0	0	0	0	0	0	0	20	100 %
C	0	0	20	0	0	0	0	0	0	0	0	20	100 %
D	0	0	0	20	0	0	0	0	0	0	0	20	100 %
E	0	0	0	0	20	0	0	0	0	0	0	20	100 %
F	0	0	0	0	0	20	0	0	0	0	0	20	100 %
G	0	0	0	0	0	0	20	0	0	0	0	20	100 %
H	8	0	0	0	0	0	0	12	0	0	0	20	60 %
I	0	0	0	0	0	0	0	0	20	0	0	20	100 %
J	0	0	0	0	0	0	0	0	0	20	0	20	100 %
K	0	0	0	0	0	0	0	0	0	0	20	20	100 %
Total	20	20	20	20	20	20	20	20	20	20	20	220	92.73%

The PCA results agreed with the AHC results, however, PCA separated nine out of the eleven exemplars, whereas AHC created three classes that clustered exemplars based on similar spectral features. The DA results created a model that best discriminated the exemplars from one another and provided error rates to all the classified spectra.

2.3.2. External Validation

An external validation on the training set was performed using two additional fibers from each exemplar. PCA was performed to determine the number of PCs to retain for subsequent DA. DA then predicted the class membership of the additional fibers spectra. Table 2.4 shows the confusion matrix of the external validation. Overall, the performance of the predicted model was excellent, with a total classification accuracy of 90.91%. Exemplars A and H were misclassified with each other and had classification accuracies of 70% and 80%, respectively. In this case, the accuracies are higher than expected. Examining the probability of each spectrum to the predicted class, however, shows probabilities of membership between 50% and 60% between the two classes. Two spectra from Exemplar E were misclassified as Exemplar G and three spectra from Exemplar I were misclassified as Exemplar J. Observation plots, seen in Figures 2.12, 2.13, and 2.14, of CV1, CV2, and CV3 show where the predicted fiber spectra were placed on the coordinate system. Utilizing all three dimensions shows that the two-Exemplar E misclassifications are closer in distance to the centroid of Exemplar G. The same can be said for the three Exemplar I misclassifications. An obvious drift of the misclassified spectra from their respective class centroid was seen. This could be due to the lack of homogeneity of the fibers within each exemplar. A study on the homogeneity of fibers within and between garments should be conducted.

Table 2.4. Confusion matrix of the prediction set

Predicted ----- Actual	A	B	C	D	E	F	G	H	I	J	K	Total	% correct
SA	7	0	0	0	0	0	0	3	0	0	0	10	70 %
SB	0	10	0	0	0	0	0	0	0	0	0	10	100 %
SC	0	0	10	0	0	0	0	0	0	0	0	10	100 %
SD	0	0	0	10	0	0	0	0	0	0	0	10	100 %
SE	0	0	0	0	8	0	2	0	0	0	0	10	80 %
SF	0	0	0	0	0	10	0	0	0	0	0	10	100 %
SG	0	0	0	0	0	0	10	0	0	0	0	10	100 %
SH	2	0	0	0	0	0	0	8	0	0	0	10	80 %
SI	0	0	0	0	0	0	0	0	7	3	0	10	70 %
SJ	0	0	0	0	0	0	0	0	0	10	0	10	100 %
SK	0	0	0	0	0	0	0	0	0	0	10	10	100 %
Total	9	10	10	10	8	10	12	11	7	13	10	110	90.91%

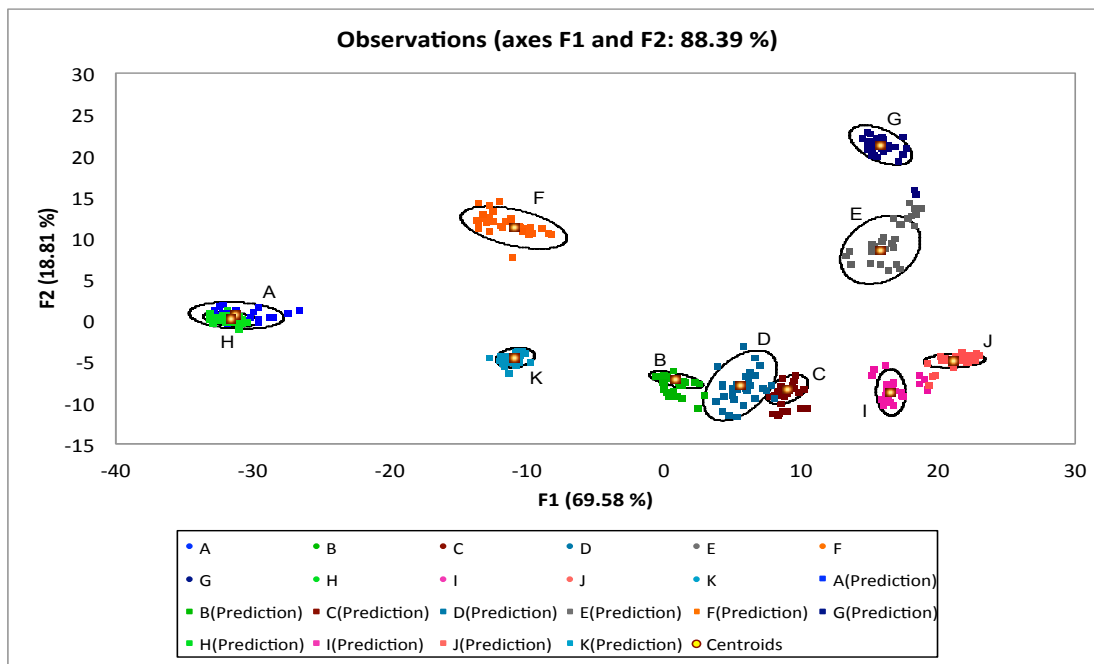


Figure 2.12. Projections of the data in the first two canonical variates of the prediction set

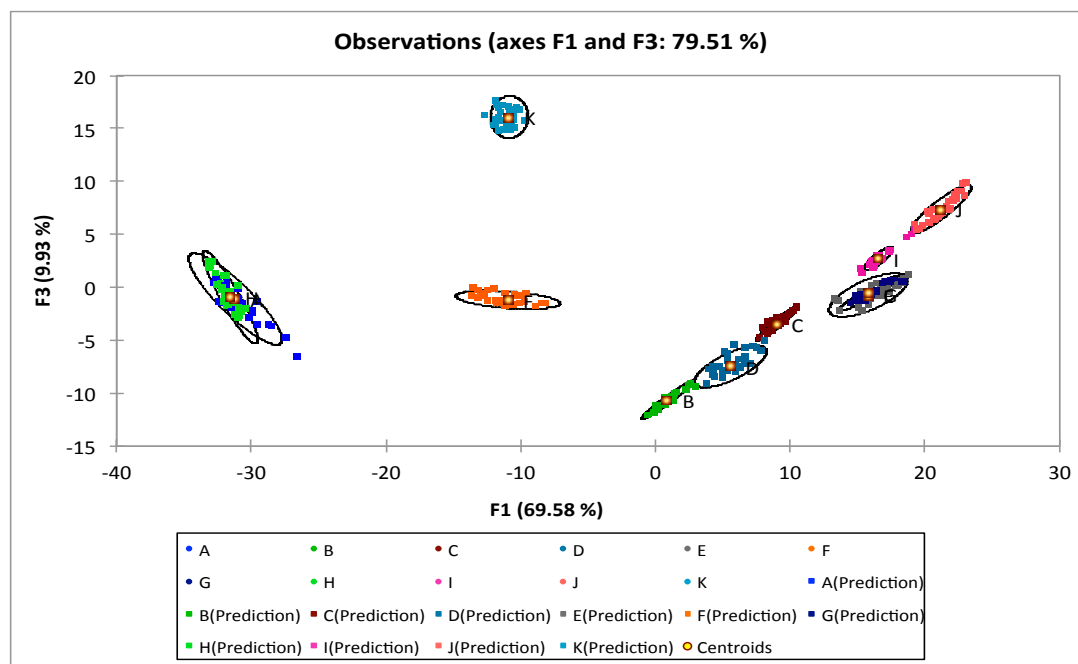


Figure 2.13. Projections of the data in the first and third canonical variate of the prediction set

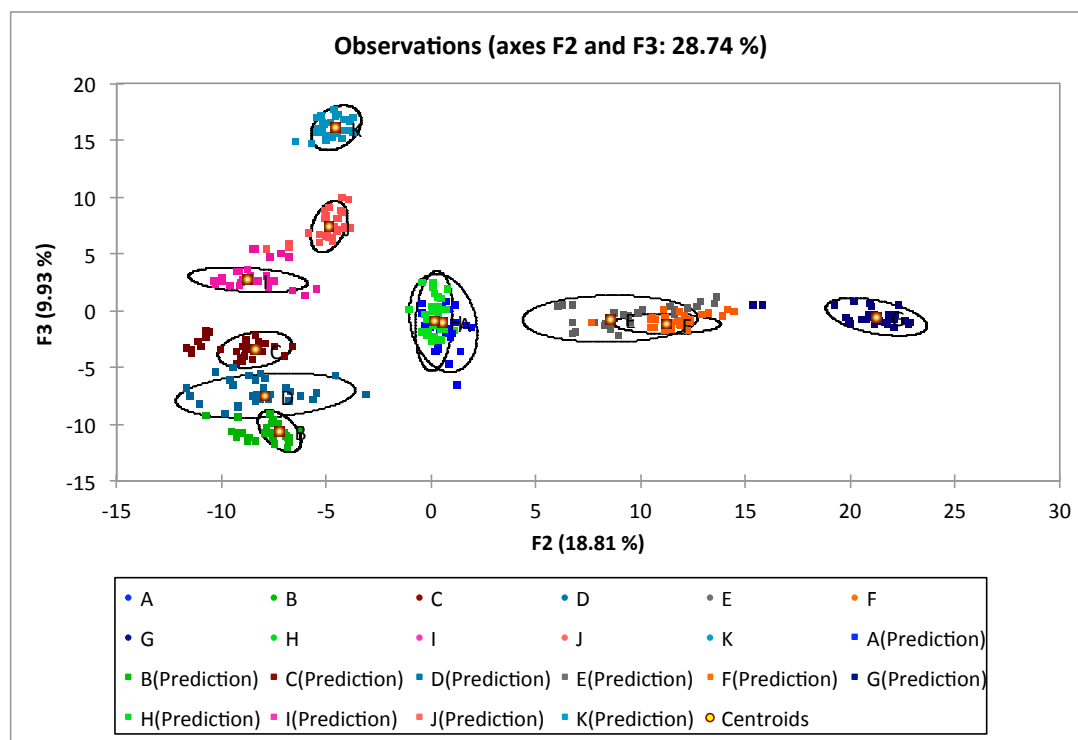


Figure 2.14. Projections of the data in the second and third canonical variate of the prediction set

2.4. Conclusions

A comparison of this dataset utilizing AHC, PCA, and DA resulted in several findings. AHC provided an initial understanding of the groupings of spectra. Three distinct classes were formed, with spectra in each class having similar spectral features due to the dye compositions of the exemplars. Groupings of replicate spectra within each class clustered well together thus possibility indicating further structure within the dataset. PCA was utilized to determine if such groupings existed along with regions of high correlation between the original variables and principal components. Class 1 exemplars (A,F,H,K) were separated along PC1 and PC2. Class 2 exemplars (B,C,D,E,J,I) overlapped when visualizing with two PCs. When three PCs were used separation between exemplars in this class could be seen. Class 3, which

consisted of Exemplar G, was grouped by itself using two PCs. PC1 exhibited high correlations around 500-580 nm and 640-675 nm. PC2 exhibited high correlations around 450-475 nm and 690-725 nm. The exemplars showed high classification accuracies using both cross-validation and external validation. Caution should be taken when interpreting the data because two exemplars showed misclassifications in the external validation and not the cross-validation. Overall, nine of the eleven exemplars were discriminated using AHC, PCA, and DA.

CHAPTER 3. MICROSPECTROPHOTOMETRIC ANALYSIS OF YELLOW POLYESTER FIBER DYE LOADINGS WITH UTILIZATION OF CHEMOMETRIC TECHNIQUES

3.1. Introduction and Purpose

The purpose of this study was to discriminate between different yellow polyester fiber visible spectra based solely upon their dye loadings using chemometric analysis. A dye loading is the concentration of a dye, usually in weight percent, applied to a fiber. Research has shown that visually similar yellow polyester fibers with different dye compositions can be discriminated based on their UV-visible spectra.⁵⁹ However, it has not been determined if fibers dyed with the same dye, but different dye loadings, can be reliably discriminated by their visible spectra alone using a chemometric approach. Being able to discriminate between dye loadings would provide a higher level of discriminating power to forensic fiber examiners in their everyday casework. This study addresses issues brought about by the National Academy of Sciences (NAS) report for strengthening forensic science in the United States.²⁹ For example, the use of chemometric analysis would address issues with the accuracy and reliability of observer interpretations of spectroscopic data.

3.2. Materials and Methods

3.2.1. Materials

Ten yellow polyester exemplars dyed with various amounts of Dianix Yellow 5-6G (Disperse Yellow 114) dye and round cross-section were supplied by Dr. Stephen L. Morgan from the University of South Carolina. Table 3.1 and

Table 3.2 show the amount of dye applied, in weight percent, to each exemplar, the naming system employed for the study, and an image of the fiber. All fibers in the study were of the same fiber type, cross-sectional shape, and dye.

Table 3.1. Exemplars A-E with respective dye loadings in weight percent and images











Fiber ID	Naming System	Dye Loading % (w/w)	Image
674	A	0.10	
675	B	0.20	
676	C	0.40	
677	D	0.50	
678	E	0.75	

Table 3.2. Exemplars F-J with respective dye loadings in weight percent and images

Fiber ID	Naming System	Dye Loading % (w/w)	Image
679	F	1.50	
680	G	2.00	
681	H	2.50	
682	I	3.00	
683	J	3.50	

3.2.2. Instrumental Analysis

Preliminary measurements using a Leica DM EP PLM (Leica Microsystems, Buffalo Grove, IL) and a Perkin Elmer Spectrum One FT-IR spectrometer with a universal ATR sampling accessory (Perkin Elmer, Waltham, MA) were performed in order to confirm that the dataset consisted of polyester fibers with a round cross-section. For PLM, Cargille oils (R.P. Cargille Laboratories, Cedar Grove, NJ) were used to determine refractive indices and Michel-Levy charts to determine birefringence values. For FT-IR, sixteen scans with a resolution of 4.00 cm^{-1} were utilized. All spectra were acquired with a wavenumber range of $4000\text{ to }650\text{ cm}^{-1}$.

Following the guidelines of the Scientific Working Group on Materials Analysis: Fiber Subgroup (SWGMAF)⁶⁵, ten fibers from each exemplar were removed and mounted on glass microscope slides using Permount (Fischer Scientific, Fairlawn, NJ) mounting media. A CRAIC QDI 2000 microspectrophotometer (Craic Technologies, Sam Dimas, CA) was used in transmitted light mode at a total magnification of 150X. Calibration of the spectrometer by use of NIST traceable standards was performed before each use along with Kohler illumination for the microscope. Autoset optimization, a dark scan, and a reference scan were acquired prior to each sample scan. Fifty scans were taken at a resolution factor of five for each sample spectrum. Five spectra were taken at different locations along each fiber to account for intra-fiber variation. A total of 50 spectra were collected for each exemplar. This provided enough information on the dye loading variation within each exemplar. All data was collected in the UV-visible wavelength range (350-800 nm).

3.2.3. Data Analysis

Preprocessing techniques were employed before subjecting the data to statistical treatment. Background subtraction was performed on each spectrum by subtracting the minimum absorbance value for each sample from all absorbance values. Background subtraction was performed to eliminate the

effects of scattered light and bring the baseline for each spectrum down to zero. Next, each background-subtracted spectrum was normalized to unit vector length by dividing each absorbance value by the square root of the sum of squares of all absorbance values. Normalization was used to account for differences in pathlengths due to varying fiber diameters. In theory, normalizing would also remove any concentration effects, but this may not always be the case as our results show otherwise (see Section 3.3). All subsequent data analyses were performed on the background subtracted and normalized data.

Calibration plots based on the absorbance maximum (425 nm) and the Euclidean distance between an exemplar and a blank were generated on the fibers with dye loading percentages from 0.1-0.75%. Three calibrators (0.2%, 0.4%, and 0.5%) were left out of the curve one at a time and considered an unknown to determine the efficiency of the calibration curve.

All chemometric techniques were performed by use of XLSTAT Pro (AddinSoft), an add-in software program for Microsoft Excel. For agglomerative hierarchical clustering, the five spectra for each fiber were averaged. This was done to produce a readable dendrogram. The proximity between two samples was measured by Euclidean distance and the aggregation method used for grouping samples was Ward's Method. The truncation line was set just higher than the most dissimilar exemplar's replicates. Values were determined by locating the node where all the replicates for each exemplar met. From that truncation line, exemplars were placed into classes.

For principle component analysis all scans were utilized instead of the averaged scans. The PCA algorithm used was Pearson (n), which utilizes a correlation matrix of the original variables to produce principal components. A factor loadings plot and observations plot were generated from the first two principal components. Three PCs were retained for subsequent discriminant analysis based upon a Scree plot.

For discriminant analysis all scans were utilized instead of the averaged scans. A Box test was performed to determine if the covariance matrices were

unequal. From that test it was determined if subsequent analysis of data should utilize quadratic discriminant analysis (QDA) or linear discriminant analysis (LDA). The first three PCs were used as variables instead of the original variables because the number of variables must be less than the number of samples when using DA.

An external validation was performed on three new fibers from each exemplar in the dataset. The same conditions were used as with the training dataset. Only PCA and DA were performed on these fibers. PCA was performed to obtain factor scores of the validation set and subsequent DA was performed to predict where those new samples would be placed in the training set model.

Finally, PCA and DA were performed on pair-wise comparisons of the original dataset. Receiver operating characteristic (ROC) curves were generated to determine the performance of the model created by DA. More specifically the area under the curve was calculated to determine the performance of the model. ROC curves are generated by plotting the true positive and false positive rates associated with the model. The same pretreatments and conditions were used as with the training dataset. Dyed exemplars with dye loadings closest to each other were compared to discriminate one from another (B-C, C-D, D-E, etc.).

3.3. Results and Discussion

Ten yellow polyester fibers with round cross-sections and the same dye composition were analyzed using multivariate statistical techniques to evaluate whether dye loadings could be used to discriminate them. Visual examination of the fibers showed a slight difference in saturation between exemplars.

Discrimination of similar dye loadings was not apparent, however. A preliminary examination of the exemplars by use of PLM and ATR-FTIR was conducted to confirm the identity of the fibers and to determine if the dye loading had any effect on the analysis. PLM was used to determine diameters, refractive indices, birefringence values, and signs of elongation. All exemplars had similar optical

characteristics and dye loadings did not affect the results other than visual differences based on saturation. IR spectra were subjected to chemometric analysis and resulted in no significant differences between spectra of different fiber dye loadings. Library searches of the spectra were conducted, and they were highly correlated to exemplars of polyester.

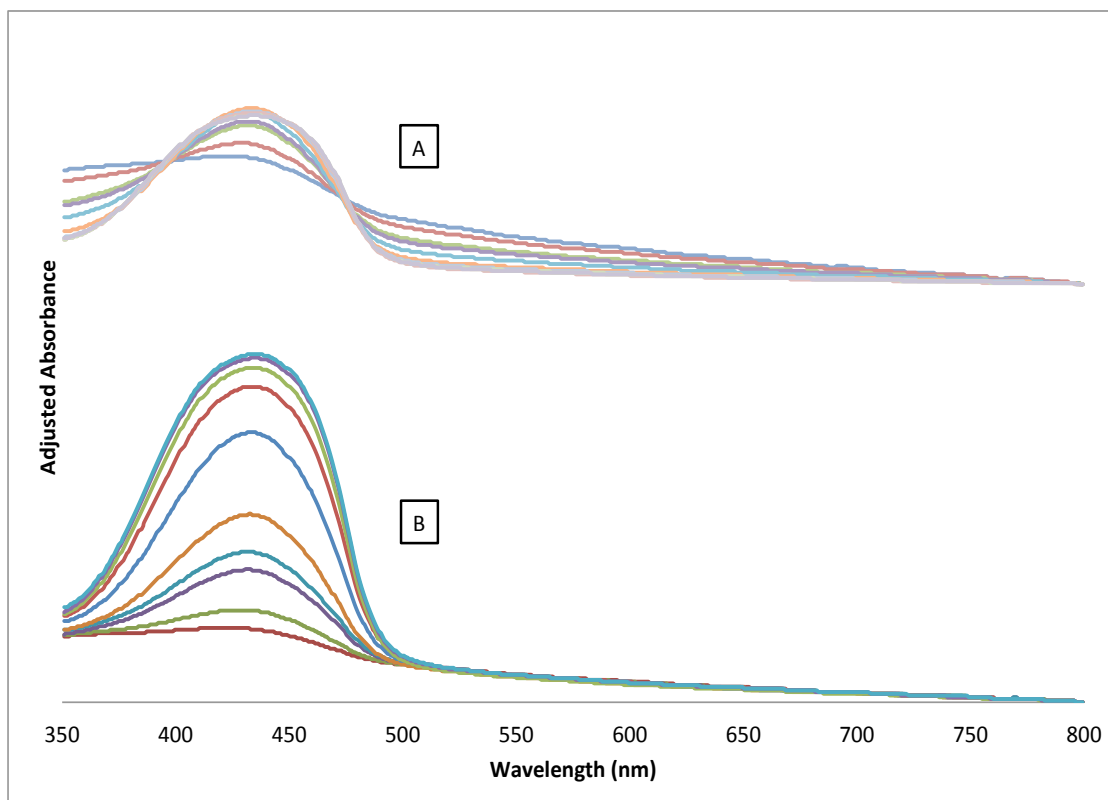


Figure 3.1. Fiber spectra with adjusted absorbance values for A) background subtracted and normalized data and B) background subtracted only data

Normalization of the spectra was appropriate due to the different diameters of the fibers. Normalizing spectral data should also eliminate concentration differences, but in this case it does not affect the results due to the superposition of the spectrum on the underlying background, which consisted of the fiber polymer and possibly the delustering agent. This background arises

due to increased light scatter at short wavelengths, which leads to a background that gradually increases. Regardless, normalizing retains the overall shape of the spectra (see Figure 3.1) and still allows for differentiation.

3.3.1. Calibration Plots

The absorbance maximum of the spectra was plotted against the exemplars' dye loading percentage. Proximity matrices were also created and the generated distances between exemplars and blanks were plotted against the exemplars' dye loadings. Both plots exhibited a curve with a linear portion at lower dye loadings that produced a negative deviation at higher dye loadings. This negative deviation can be attributed to several reasons, including the use of polychromatic light from the MSP source. In particular, Beer's law is considered invalid if polychromatic light is used as a light source at higher concentrations.⁶⁷ The linear portion of the curve ends at a dye loading of 0.75%. Therefore, calibration plots were created from 0.1% - 0.75% dye loading (see Appendix). This assertion complements previous work that characterized different pigments loaded into polypropylene fibers.⁶⁸

Table 3.3. Calibration curve results for three unknowns using a statistical and non-statistical approach

Statistical Treatment			
Expected Dye Loading	Calculated Dye Loading	Std. Dev.	R²
0.20%	0.20%	0.058	0.981
0.40%	0.45%	0.038	0.992
0.50%	0.53%	0.050	0.986
Non-statistical Treatment			
Expected Dye Loading	Calculated Dye Loading	Std. Dev.	R²
0.20%	0.23%	0.122	0.913
0.40%	0.50%	0.094	0.951
0.50%	0.57%	0.119	0.924

The results from Table 3.3 indicate that statistical treatment of the data provides better results than simply using the absorbance maximum. Using proximity matrices, which utilize every variable, are more accurate and precise than using one absorbance value for generating a calibration curve. This is particularly true in this case as the absorbance maximum shifts to a higher wavelength as dye concentration increases for this dataset.

3.3.2. Training Set

The ten exemplars from Table 3.1 and Table 3.2 were subjected to AHC, PCA, and DA after being background subtracted and normalized. Exemplars A and B exhibit a slight difference in their spectral shape because of the low dye loading on those fibers. The remaining exemplars exhibited the same general spectral curve, but at different absorbance maximum values. The absorbance maximum slightly shifted to higher wavelengths as dye loading increased as seen in Figure 3.2.

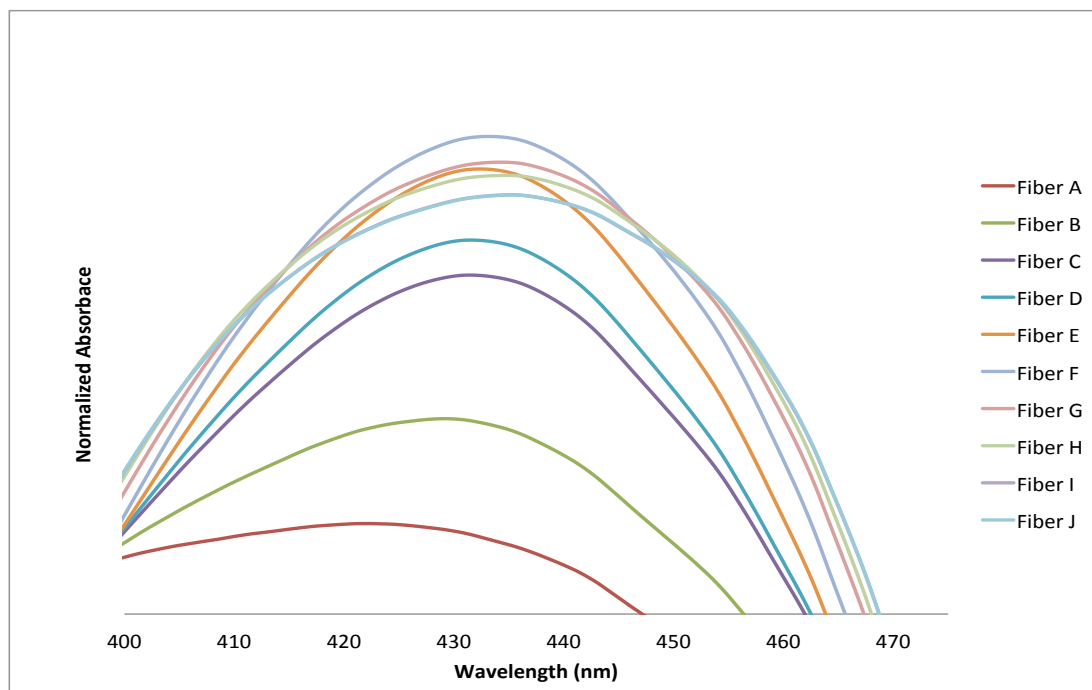


Figure 3.2. Magnified view of the shift of the absorbance maximum as dye loading increases

AHC was performed in order to detect any outliers and classes of fibers. The intra-fiber spectra were averaged in order to visualize the classes in the dendrogram. This brought the replicates down from fifty to ten. The dendrogram in Figure 3.3 shows how the data was grouped. The truncation line was set just higher than the highest level of dissimilarity between replicates. In this case, the replicates of Exemplars C and D were determined to be the most dissimilar by locating the node where all the replicates for each exemplar met. Everything then to the right of the dendrogram was grouped into classes. From the truncation line, three distinct classes of data were produced, representing low, medium, and high dye loadings.

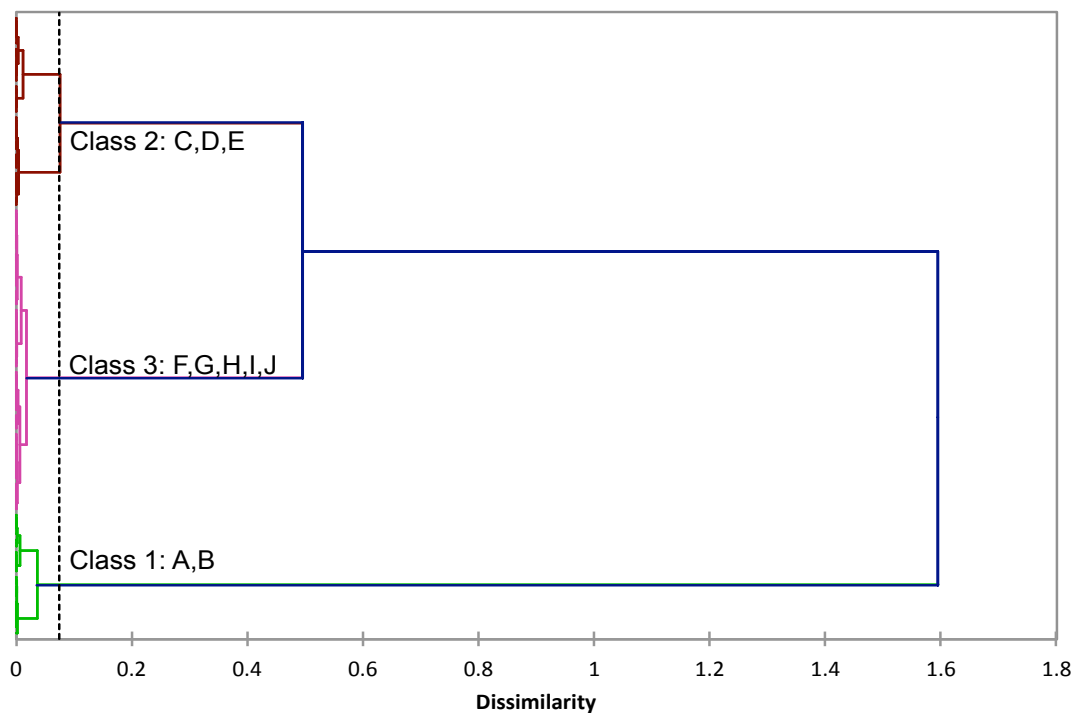


Figure 3.3. AHC dendrogram of the ten exemplars from the training set

Class 1 are low dye loadings and include Exemplars A and B. Class 2 are medium dye loadings and includes Exemplars C, D, and E. Class 3 are high dye loadings and includes Exemplars F, G, H, I, and J. Class 1 exemplars are further separated into their own respective groups because of their high reproducibility. Class 2 and Class 3 exemplars show low reproducibility and are intermixed in their respective classes. Figure 3.4 represents the spectra that were most similar to the average spectra for the three classes. Fibers B6, D3, and I1 represent the average spectral shape of all the other fibers clustered in Classes 1, 2, and 3, respectively. Each class exhibits different spectral features.

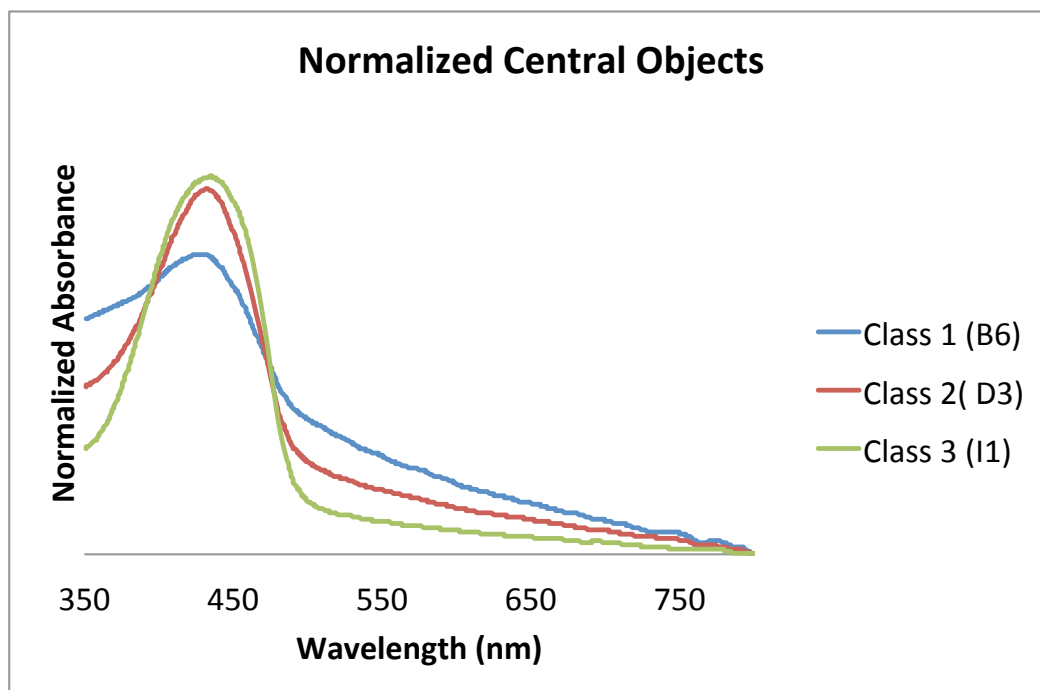


Figure 3.4. AHC central objects plot of the three classes

PCA was performed on every spectrum for each exemplar. Observation plots of the individual exemplars and the three classes captured 92.59% of the total variance in two PCs. Figure 3.5 shows the ten exemplars spread out from low to high dye loadings. PC1 separates out Exemplars A-E while PC2 slightly separates Exemplars F-J. Figure 3.6 shows the observation plot of the three classes produced from AHC. PC1 separates out Classes 1, 2, and 3, while PC2 separates samples within each class. From the results of PCA and AHC, it is apparent that three classes representing low, medium, and high dye loading can be seen.

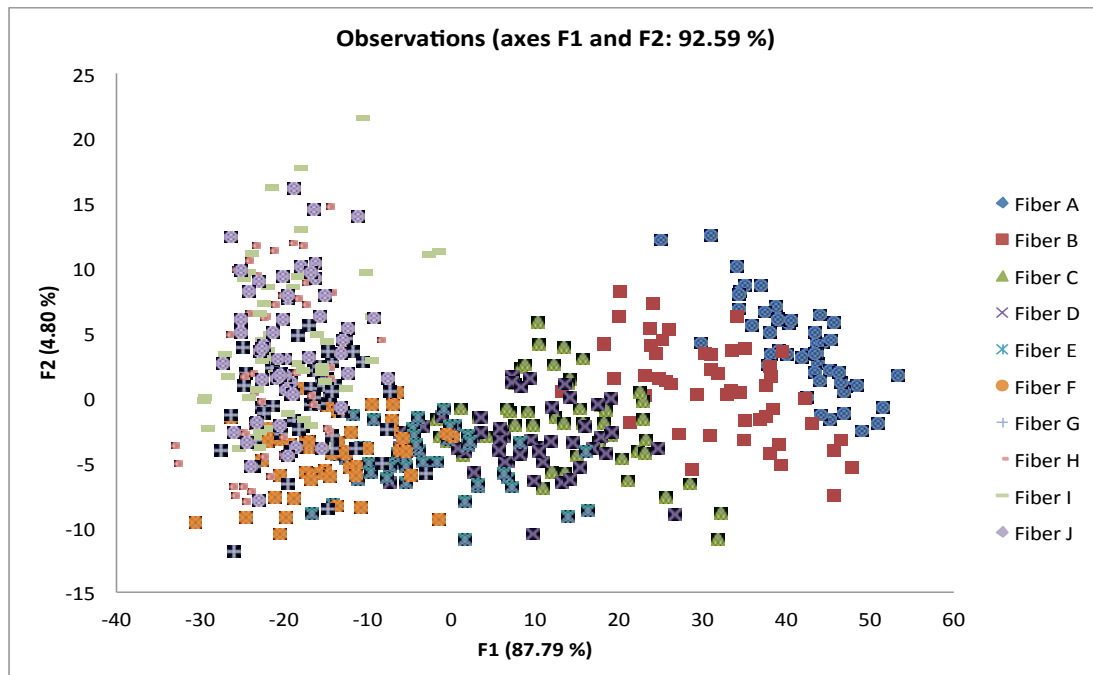


Figure 3.5. Projections of the data in the first two principal components

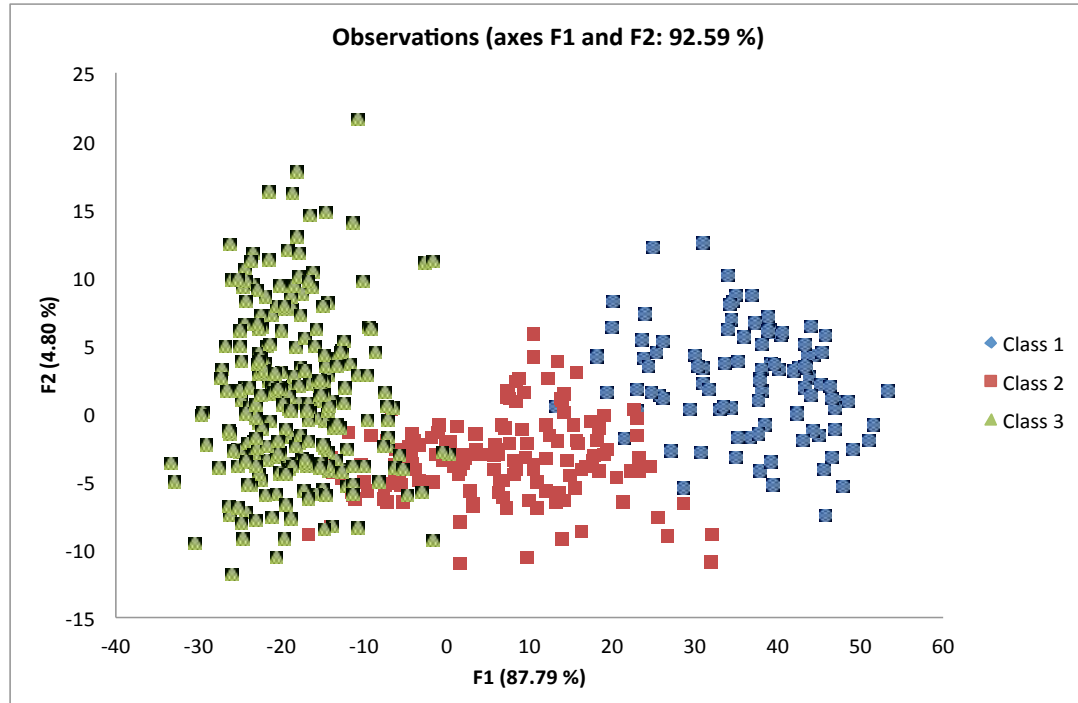


Figure 3.6. Projections of the data in the first two principal components of the class data

A factor loadings plot, seen in Figure 3.7, was also created to understand the relationship between the original wavelengths and the new principal components. The factor loading was determined by finding the cosine of the angle between the principal component and the original variable. A factor loading approaching 1 or -1 corresponds to positive and negative correlations, respectively, between the original wavelengths and the PCs.

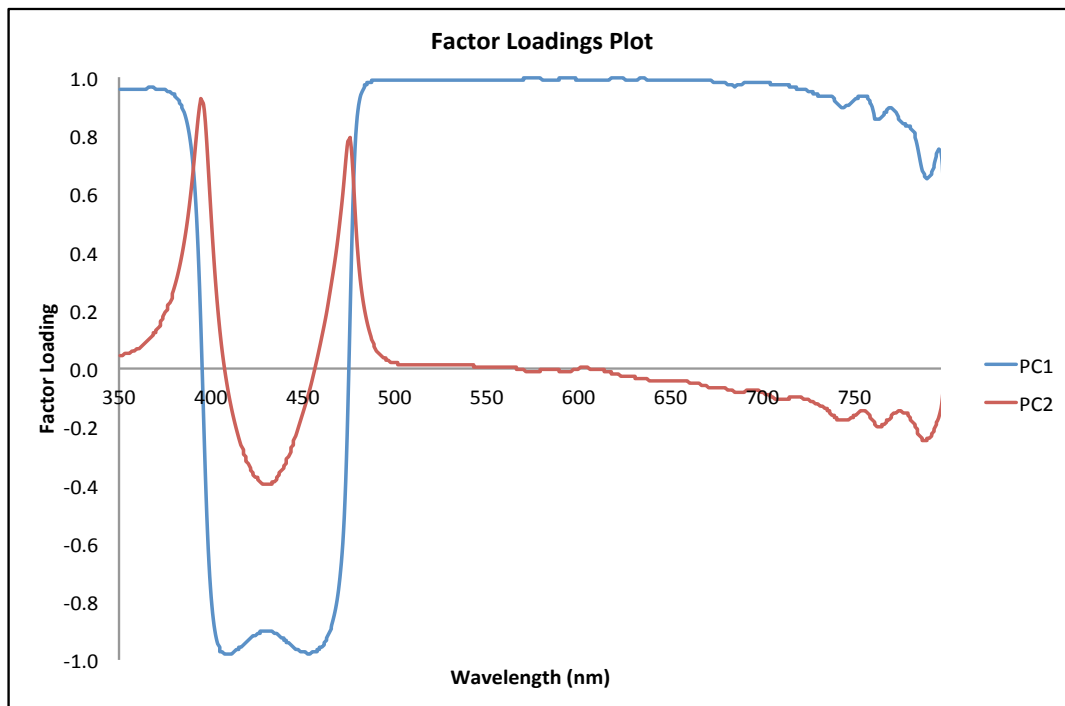


Figure 3.7. Factor loadings plot of the first two principal components

PC1 exhibited a strong negative correlation around the absorbance maximum (400-475 nm) of the spectra and strong positive correlations between 350-375 nm and 500-700 nm. PC2 exhibited positive correlations around 395 nm and 475 nm. The importance of these regions explains how the exemplars were separated on the observations plot. Figure 3.8 overlays the areas of positive and negative correlations of PC1 and PC2 onto the three classes of fiber spectra created from AHC. The lower dye loading exemplars in Class 1 were separated from Classes 2 and 3 exemplars based on the absorbance maximum. Class 2 and 3 exemplars were separated based on the sides of their respective

spectral curves, not the absorbance maximum as expected because their spectral curves plateaued and became broader. This was especially seen with spectra that exhibited absorbance values greater than unity.

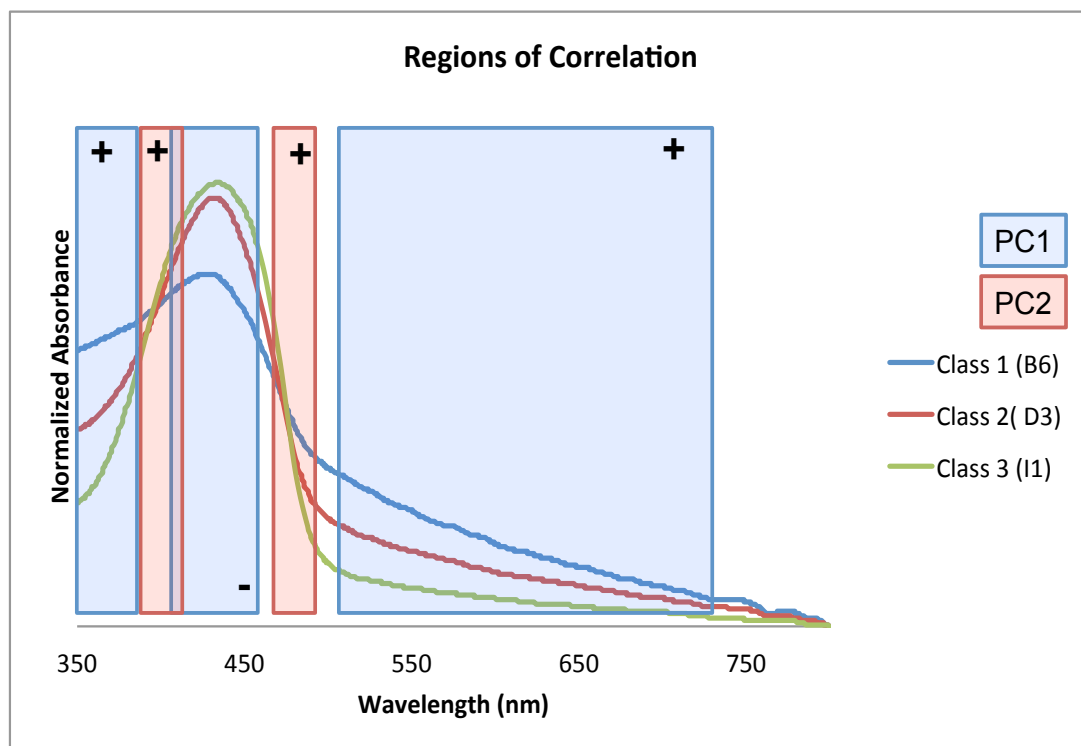


Figure 3.8. Regions of high correlation (factor loadings) superimposed over the three classes of fiber spectra from AHC

DA was performed several times on all the spectra. First, the exemplars were assigned to their own classes. The ten exemplars showed the same general trend from low to high dye loadings, as seen in Figure 3.9. This observation plot had a total captured variance of 99.82% using two canonical variates.

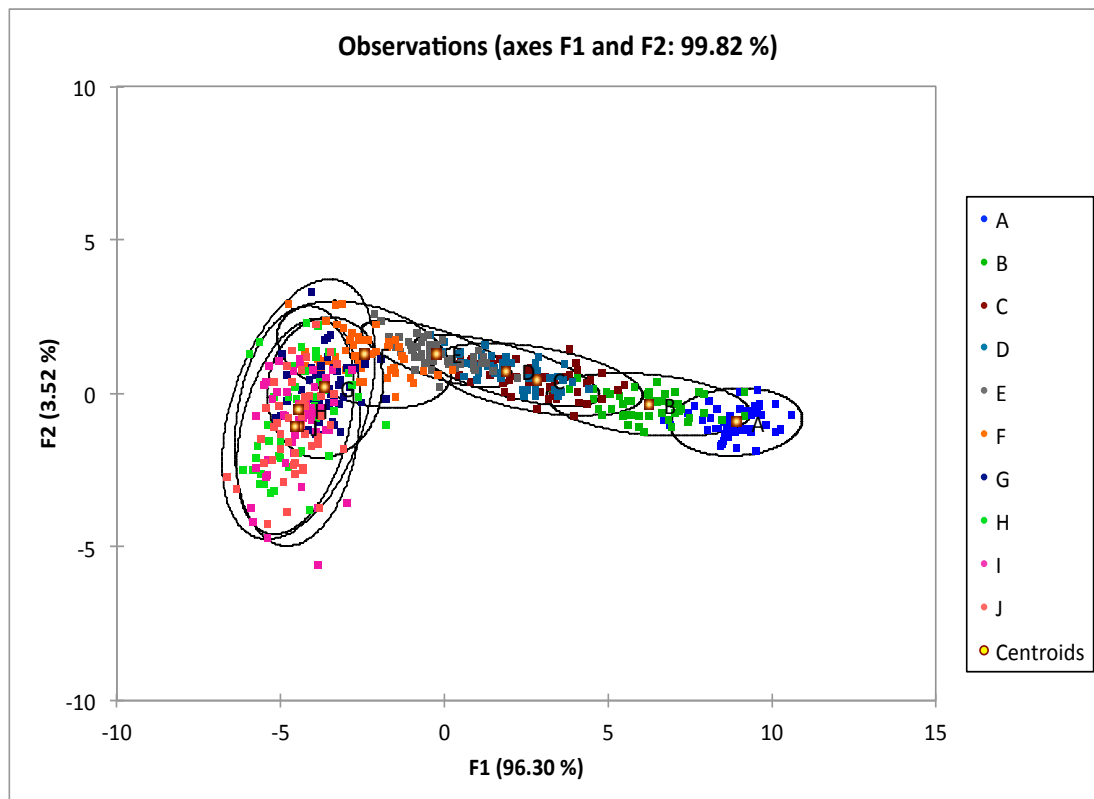
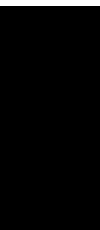


Figure 3.9. Projections of the data in the first two canonical variates

A “leave-one-out” cross validation was performed on the training set and a subsequent confusion matrix was produced. Here, DA is performed on all but one of the samples and a model is created. The left out sample is then added into the model and is assigned to a class. This step is repeated until every sample in the dataset is classified. Cross validating the training set gave a total classification accuracy of 50.80%, which can be seen in Table 3.4. The classification accuracy generally decreased as dye loading increased. This decrease can be attributed to the limitations of the instrument and/or the polychromatic light issue stated in Section 3.3.1. Higher dye loadings produced absorbance values above 1, which pushed the limits of the detector. At absorbance values of this magnitude, the detector is less sensitive to changes in concentration due to the minimal amount of light transmitted through the sample.

Table 3.4. Cross-validation confusion matrix of the training set

Predicted ----- Actual	A	B	C	D	E	F	G	H	I	J	Total	% correct
A	46	4	0	0	0	0	0	0	0	0	50	92 %
B	5	40	5	0	0	0	0	0	0	0	50	80 %
C	0	4	23	18	5	0	0	0	0	0	50	46 %
D	0	0	18	18	14	0	0	0	0	0	50	36 %
E	0	0	1	10	33	6	0	0	0	0	50	66 %
F	0	0	0	0	7	30	8	5	0	0	50	60 %
G	0	0	0	0	0	9	27	9	2	3	50	54 %
H	0	0	0	0	0	7	10	11	5	17	50	22 %
I	0	0	0	0	0	1	10	17	8	14	50	16 %
J	0	0	0	0	0	2	7	15	8	18	50	36 %
Total	51	48	47	46	59	55	62	57	23	52	500	50.80 %



Consequently, exemplars with high dye loadings had spectra with similar intensities and spectral shapes. Also, the molar absorptivity of the dye changes when polychromatic light is utilized because of the varying chemical and electrostatic interactions of the dye molecules as concentration is increased.⁶⁷ In this case, as dye loading increased to high concentrations the molar absorptivity decreased. This caused spectra of high dye loading exemplars

The exemplars were also assigned to their AHC classes and DA was performed. Three classes were formed representing low, medium, and high dye loadings. The observation plot in Figure 3.10 had a total captured variance of 100% using two canonical variates. Although the 95% ellipses did overlap, the centroids for each class were not captured by the other ellipses. A leave-one-out cross validation was performed and gave a total classification accuracy of 95.60%, which can be seen in Table 3.5. The medium dye loadings, representing Class 2, were confused the most because of the confusion with both the low and high dye loading exemplars.

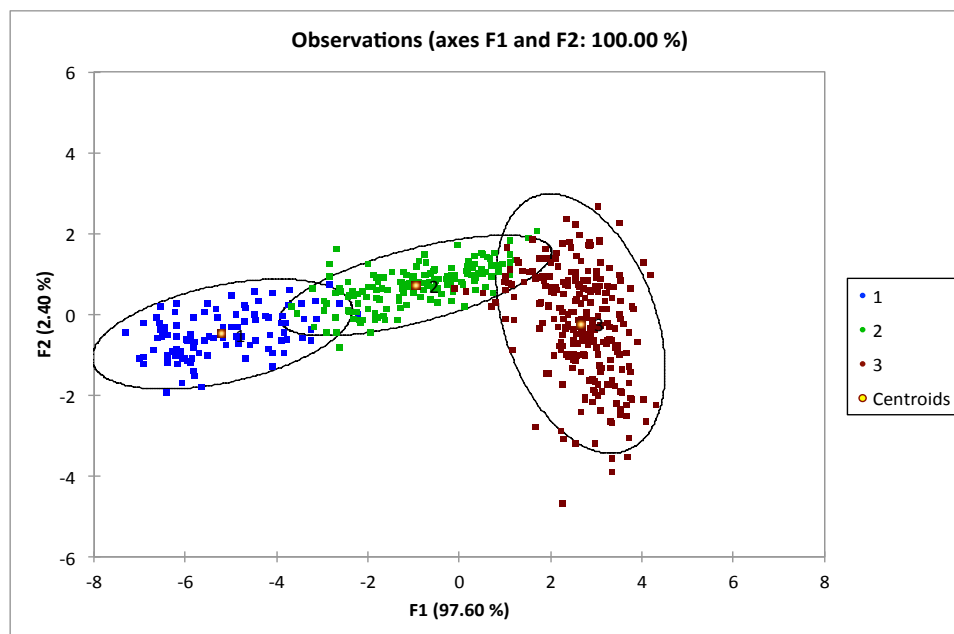


Figure 3.10. Projections of the data in the first two canonical variates of the classes generated from the dendrogram

Table 3.5. Cross-validation confusion matrix of classes generated from dendrogram

Predicted ----- Actual	1	2	3	Total	% correct
1	94	6	0	100	94.00 %
2	3	145	2	150	96.67 %
3	0	11	239	250	95.60 %
Total	97	162	241	500	95.60 %

3.3.3. External Validation

Three additional fibers from each exemplar were analyzed by PCA and DA in order to evaluate the accuracy of the model predicted in the training set. These additional samples were then placed into the class its spectra most resembled. This was based on the probabilities calculated during DA analysis. The samples were placed into one of the ten exemplar classes for the first DA analysis, and then placed into one of the three AHC classes for the second DA analysis for comparison purposes. Table 3.6 shows the classification accuracies of the external validation.

Table 3.6. External validation results

Individual Fiber Exemplars			Class Fiber Exemplars		
Additional Fiber	Class	Accuracy	Additional Fiber	Class	Accuracy
SA	A	100 %	SA	1	100 %
SB	B	93.33 %	SB	1	100 %
SC	C	46.67 %	SC	2	100 %
SD	D	46.67 %	SD	2	100 %
SE	E	60 %	SE	2	93.33 %
SF	F	86.67 %	SF	3	93.33 %
SG	G	40 %	SG	3	100%
SH	H	40 %	SH	3	100 %
SI	I	26.67 %	SI	3	100 %
SJ	J	26.67 %	SJ	3	100 %

A total classification accuracy of 56.67% was seen when the fibers were placed into the ten classes and 98.67% when the fibers were placed into the three classes. There was extensive confusion between exemplars when the fibers were placed into the individual exemplar classes. This was expected because there was significant confusion in the training set especially as dye loading increased. However, less confusion was seen when the additional fibers were predicted into the AHC class model. Exemplars E and F were the only two confused when exemplars were placed into the three classes. Both misclassified one sample into a wrong class.

3.3.4. Pair-Wise Comparisons

PCA and DA were performed on two groups of exemplars in order to determine whether the groups could be discriminated from each other. The results in Table 3.7 show that only two exemplar comparisons and both the class comparisons were discriminated from each other based on a classification accuracy of 90% or higher and a receiver operating characteristics curve score of 0.9 or higher. All other comparisons fell below this threshold and were considered to be confused with each other. The exemplar comparisons were chosen as seen because this was the closest two exemplars could be to each other and not be the same dye loading. In other words, the classification accuracies would be better if comparisons other than the ones in Table 3.7 were performed.

Table 3.7. Pair-wise comparison results

Comparison	Classification Accuracy	ROC AUC
A-B	92 %	0.992
B-C	94 %	0.992
C-D	64 %	0.720
D-E	82 %	0.936
E-F	89 %	0.982
F-G	83 %	0.918
G-H	75 %	0.845
H-I	57 %	0.721
I-J	52 %	0.644
Class 1 - 2	98 %	0.998
Class 2 - 3	97 %	0.998

3.4. Conclusions

Chemometric treatment of visible spectra from fibers with different dye loadings has been shown to be a reliable and effective way of discriminating between fibers. Higher classification accuracies were seen when fibers were placed into low, medium, and high classes. Calibration curves, based on proximity matrices, can be produced and accurately predict the dye loading of an unknown fiber. Comparisons of two groups of fibers can provide discriminating information. From this data set, absorbance differences of 0.1% between two groups of fibers could be discriminated. Overall, forensic fiber examiners can have a more objective way of comparing a known and questioned fiber using chemometric techniques on visible spectra.

CHAPTER 4. LIMITATIONS AND FUTURE WORK

4.1. Limitations

As with any research conducted, there will always be limitations. The first limitation involves the yellow polyester data set. The fibers used in this study had varying diameters ranging from 20-27.5 μm . Normalizing the spectra was required in order to compensate for these differences. Normalizing the spectra, however, can eliminate concentration effects, which would inhibit the interpretation of the data since the difference in concentration is the primary discriminating factor. Normalizing in this case is appropriate due to the sloping background of the spectra. It would have been best suited to have a data set containing fibers with the same diameter, but our collection did not contain that representative data set.

The second limitation involves both the yellow polyester study and the blue acrylic study. Both studies utilized data sets containing ten and eleven exemplars, respectively. A good representative sample is important particularly when performing discriminant analysis. A drawback to DA is that an individual sample is forced into an existing group established by the model created for the training set. This is especially important for comparisons of questioned and known samples in casework. For example, multiple fibers from different locations of a shirt with varying dye formulations and shades of color would have to be collected in order to create a model that represents every possible outcome associated with that shirt. An unknown colored fiber could mistakenly be classified into an incorrect group. Care must be taken when interpreting the

data. Other chemometric classification techniques like soft independent modeling of class analogy (SIMCA) can be used to compensate for this drawback. SIMCA is considered a soft modeling technique, meaning a sample can belong to more than one group or to no defined groups at all.³²

4.2. Future Work

The research in this thesis can be expanded upon along with other future work that can be conducted. An interlaboratory study on the yellow polyester and blue acrylic data sets should be completed. An interlaboratory study would verify the reproducibility of the chemometric procedure along with the MSP method. Performing an external validation at certain time intervals on the same set of fibers can be conducted in order to determine if drifts in the predicted samples occur as time increases. This is in response to a slight drift in the predicted samples from their correct groups of the blue acrylic data set. Finally, the fibers used in these studies were considered pristine meaning there were no other variables affecting the fiber's color other than the dye applied and the dye concentration. Data sets comprising of damaged, weathered, and/or laundered fibers should be explored in order to assess the effects these variables have on the chemometric analysis.

Future work can also be completed on different data sets using FT-IR and PLM with our chemometric procedure to classify and compare different fiber types and sub types. For FT-IR, variables (wavenumbers) contributing to the most variation within the data set can be determined along with comparing functional groups between samples. Another project that could be completed utilizes principal component regression (PCR). PCR is able to quantify the concentrations of multiple dyes applied to a fiber instead of one (i.e. yellow polyester data set). It is a combination of PCA and ordinary least squares regression. PCA is conducted in order to reduce the dimensions of the data and extract significant PCs that will be subjected to ordinary least squares regression for quantitative analysis.³² The chemometric procedure would be more

applicable to case work since the majority of fabrics in production employ multiple dyes to achieve a desired color. A final project, will quantify the dye concentration of a yellow polyester exemplar with an unknown dye concentration by LC-MS. An extraction procedure of the disperse dye and LC-MS procedure would be optimized. These procedures can also be used to validate the predicted concentrations of the yellow polyester data set found from the calibration procedure outlined in Chapter 3.

LIST OF REFERENCES

LIST OF REFERENCES

1. Eyring, M. B.; Gaudette, B. D., An Introduction to the Forensic Aspects of Textile Fiber Examination. In *Forensic Science Handbook*, Saferstein, R., Ed. Vol. II. Prentice Hall: Englewood Cliffs, NJ, 2005; pp 231-296.
2. Robertson, J.; Roux, C., From the Crime Scene to the Laboratory. In *Forensic Examination of Fibres*, 2nd ed., Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, NY, 1999.
3. Houck, M. M., Ways of identifying textile fibers and materials. In *Identification of textile fibers*, Houck, M. M., Ed. CRC Press: Boca Raton, FL, 2009; pp 6-26.
4. Tridico, S. R., Natural animal textile fibres: structure, characteristics and identification. In *Identification of textile fibers*, Houck, M. M., Ed. CRC Press: Boca Raton, FL, 2009; pp 27-67.
5. Adolf, F. P.; Dunlop, J., Microspectrophotometry/Colour Measurement. In *Forensic Examination of Fibres*, 2nd ed.; Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, NY, 1999; pp 251-290.
6. Lewis, S. W., Analysis of dyes using chromatography. In *Identification of textile fibers*, Houck, M. M., Ed. CRC Press: Boca Raton, FL, 2009; pp 203-223.
7. Goodpaster, J.V.; Liszewski, E. A., Forensic analysis of dyed textile fibers. *Analytical Bioanalytical Chemistry* 2009, 394 (8), 2009-2018.
8. Palenik, S. J., Microscopical Examination of Fibres. In *Forensic Examination of Fibres*, 2nd ed., Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, NY, 1999; pp 153-178.
9. Kirkbride, K. P.; Tungol, M. W., Infrared Microspectroscopy of Fibres. In *Forensic Examination of Fibres*, 2nd ed., Robertson, J.; Grieve, M., Eds. Taylor and Francis: New York, NY, 1999; pp 179-222.

10. Walkbridge-Jones, S., Microspectrophotometry for textile fiber color measurement. In *Identification of textile fibers*, Houck, M. M., Ed. CRC Press: Boca Raton, FL, 2009; pp 165-180.
11. S. F. Stoeffler., A Flowchart System for the Identification of Common Synthetic Fibers by Polarized Light Microscopy. *Journal of Forensic Sciences* 1996, 41 (2), 297-299.
12. Tungol, M. W.; Barick, E. G.; Montaser, A., Analysis of Single Polymer Fibers by Fourier Transform Infrared Microscopy: The Results of Case Studies. *Journal of Forensic Sciences* 1991, 36 (4), 1027-1043.
13. Cho, L.; Reffner, J. A.; Gatewood, B. M.; Wetzel, D. L., Single Fiber Analysis by Internal Reflection Infrared Microspectroscopy. *Journal of Forensic Sciences* 2001, 46 (6), 1309-1314.
14. Keen, I. P.; White, G. W.; Fredericks, P.M., Characterization of Fibers by Raman Microprobe Spectroscopy. *Journal of Forensic Sciences* 1998, 43 (1), 82-89.
15. Armitage, S.; Saywell, S.; Roux, C.; Lennard, C.; Greenwood, P., The Analysis of Forensic Samples Using Laser Micro-Pyrolysis Gas Chromatography Mass Spectrometry. *Journal of Forensic Sciences* 2001, 46 (5), 1043-1052.
16. Grieve, M. C.; Wiggins, K. G., Fibers Under Fire: Suggestions for Improving Their Use to Provide Forensic Evidence. *Journal of Forensic Sciences* 2001, 46 (4), 835-843.
17. Wiggins, K. G.; Crabtree, S. R.; March, B. M., The Importance of Thin Layer Chromatography in the Analysis of Reactive Dyes Released from Wool Fibers. *Journal of Forensic Sciences* 1996, 41 (6), 1042-1045.
18. Resua, R.; DeForest, P. R.; Harris, H., The Evaluation and Selection of Uncorrelated Paired Solvent Systems for Use in the Comparison of Textile Dyes by Thin-Layer Chromatography. *Journal of Forensic Sciences* 1981, 26 (3), 515-534.
19. Huang, M.; Yinon, J.; Sigman, M., Forensic Identification of Dyes Extracted from Textile Fibers by Liquid Chromatography Mass Spectrometry (LC-MS). *Journal of Forensic Sciences* 2004, 49 (2), 238-249.

20. Petrick, L. M.; Wilson, T.A.; Fawcett, W. R., High-Performance Liquid Chromatography-Ultraviolet-Visible Spectroscopy-Electrospray Ionization Mass Spectrometry Method for Acrylic and Polyester Forensic Fiber Dye Analysis. *Journal of Forensic Sciences* 2006, *51* (4), 771-779.
21. Dockery, C. R.; Stefan, A. R.; Nieuwland, A. A.; Roberson, S. N.; Baguley, B. M.; Hendrix, J. E.; Morgan, S. L., Automated Extraction of Direct, Reactive, and Vat Dyes from Cellulosic Fibers for Forensic Analysis by Capillary Electrophoresis. *Analytical Bioanalytical Chemistry* 2009, *394* (8), 2095-2103.
22. Zhou, C.; Li, M.; Garcia, R.; Crawford, A.; Beck, K.; Hinks, D.; Griffis, D. P., Time-of-Flight-Secondary Ion Mass Spectrometry Method Development for High-Sensitivity Analysis of Acid Dyes in Nylon Fibers. *Analytical Chemistry* 2012, *84* (22), 10085-10090.
23. Eng M.; Martin, P.; Bhagwandin, C., The Analysis of Metameric Blue Fibers and Their Forensic Significance. *Journal of Forensic Sciences* 2009, *54* (4), 841-845.
24. Grieve M. C.; Biermann, T.; Davignon, M., The Occurrence and Individuality of Orange and Green Cotton Fibres. *Science & Justice* 2003, *43* (1), 5-22.
25. Marcræ, R.; Dudley, R. J.; Smalldon, K. W., The Characterization of Dyestuffs on Wool Fibers with Special Reference to Microspectrophotometry. *Journal of Forensic Sciences* 1979, *24* (1), 117-129.
26. Wiggins, K.; Palmer, R.; Hutchinson, W.; Drummond, P., An Investigation into the Use of Calculating the First Derivative of Absorbance Spectra as a Tool for Forensic Fibre Analysis. *Science & Justice* 2007, *47* (1), 9-18.
27. Morgan, S. L.; Bartick, E. G., Discrimination of Forensic Analytical Chemical Data Using Multivariate Statistics. In *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*. Blackledge, R. D., Ed. John Wiley & Sons: New York, NY, 2007; pp 333-374.
28. William Daubert, et ux., etc., et al., Petitioners v. Merrell Dow Pharmaceuticals, Inc. U.S. Supreme Court: 1993; Vol. 509 US 579, 589.
29. National Resource Council, *Strengthening Forensic Science in the United States: A Path Forward*. National Academies Press: Washington, DC, 2009.

30. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B., *Chemometrics: A Practical Guide*. John Wiley & Sons, Inc.: New York, NY, 1998.
31. Kramer, R., *Chemometric Techniques for Quantitative Analysis*. Marcel Dekker Inc.: New York, NY, 1998. pp 173-175.
32. Varmuza, K.; Filzmoser, P., *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press: Boca Raton, FL, 2009.
33. StatSoft, I., *Electronic Statistics Textbook*. StatSoft: Tulsa, OK, 2011. <http://www.statsoft.com/textbook/> (accessed January 2013).
34. Brereton, R. G., *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. John Wiley & Sons, Ltd.: Chichester, England, 2003.
35. Meloun, M.; Militky, J.; Forina, M., *Chemometrics for Analytical Chemistry Volume 1: PC-aided statistical data analysis*. Ellis Horwood Ltd.: London, England, 1992; pp 261-262.
36. Thanasoulas, N. C.; Parisi, N. A.; Evmiridis, N. P., Multivariate Chemometrics for the Forensic Discrimination of Blue Ball-Point Pen Inks based on their Vis Spectra. *Forensic Science International* 2003, 138 (1-3), 75-84.
37. Egan, W. J.; Galipo, R. C.; Kochanowshi, B. K.; Morgan, S. L.; Bartick, E. G.; Miller, M. L.; Ward, D. C.; Mothershead II, R. F., Forensic Discrimination of Photocopy and Printer Toners. III. Multivariate Statistics Applied to Scanning Electron Microscopy and Pyrolysis Gas Chromatography/Mass Spectrometry. *Analytical Bioanalytical Chemistry* 2003, 376 (8), 1286-1297.
38. Koons, R. D.; Fiedler, C.; Rawalt, R. C., Classification and Discrimination of Sheet and Container Glasses by Inductively Coupled Plasma-Atomic Emission Spectrometry and Pattern Recognition. *Journal of Forensic Sciences* 1988, 33 (1), 49-67.
39. Chazottes, V.; Brocard, C.; Peyrot, B., Particle Size Analysis of Soils Under Simulated Scene of Crime Conditions: the Interest of Multivariate Analyses. *Forensic Science International* 2004, 140 (2-3), 159-166.
40. Thanasoulas, N. C.; Piliouris, E. T.; Kotti, M. S. E.; Evmiridis, N. P., Application of Multivariate Chemometrics in Forensic Soil Discrimination based on the UV-Vis Spectrum of the Acid Fraction of Humus. *Forensic Science International* 2002, 130 (2-3), 73-82.

41. Chalmers, J. M.; Everall, N. J., FTIR, FT-Raman and Chemometrics: Applications to the Analysis and Characterization of Polymers. Trends in *Analytical Chemistry* 1996, *15* (1), 18-25.
42. Lukasiak, B. M.; Faria, R.; Zomer, S.; Brereton, R. G.; Duncan, J. C., Pattern Recognition for the Analysis of Polymeric Materials. *Analyst* 2006, *131*, 73-80.
43. Flynn, K.; O'Leary, R.; Lennard, C.; Roux, C.; Reedy, B. J., Forensic Applications of Infrared Chemical Imaging: Multi-Layered Paint Chips. *Journal of Forensic Sciences* 2005, *50* (4), 1-10.
44. Kretschmer, K.; Helbig, W., Differentiation of Commercial Polyester Fibers Using Two-Dimensional High-Performance Liquid Chromatography (HPLC) and Multivariate Pattern Recognition Techniques. *Journal of Forensic Sciences* 1992, *37* (3), 727-737.
45. Barrett, J. A.; Siegel, J. A.; Goodpaster, J. V., Forensic Discrimination of Dyed Hair Color: II. Multivariate Statistical Analysis. *Journal of Forensic Science* 2011, *56* (1), 95-101.
46. Goodpaster, J. V.; Sturdevant, A. B.; Andrews, K. L.; Brun-Conti, L., Identification and Comparison of Electrical Tapes Using Instrumental and Statistical Techniques: I. Microscopic Surface Texture and Elemental Composition. *Journal of Forensic Sciences* 2007, *52* (3), 610-629.
47. Goodpaster, J.V.; Sturdevant, A. B.; Andrews, K. L.; Briley, E. M.; Brun-Conti, L., Identification and Comparison of Electrical Tapes Using Instrumental and Statistical Techniques: II. Organic Composition of the Tape Backing and Adhesive. *Journal of Forensic Sciences* 2009, *54* (2), 328-338.
48. Tan, B.; Hardy, J. K.; Snavely, R. E., Accelerant Classification by Gas Chromatography/Mass Spectrometry and Multivariate Pattern Recognition. *Analytica Chimica Acta* 2000, *422* (1), 37-46.
49. Turner, D. A.; Goodpaster, J. V., Comparing the Effects of Weathering and Microbial Degradation on Gasoline Using Principal Components Analysis. *Journal of Forensic Sciences* 2012, *57* (1), 64-69.

50. Enlow, E. M.; Kennedy, J. L.; Nieuwland, A. A.; Hendrix, J. E.; Morgan, S. L., Discrimination of Nylon Polymers Using Attenuated Total Reflection Mid-infrared Spectra and Multivariate Statistical Techniques. *Applied Spectroscopy* 2005, *59* (8), 986-992.
51. Kokot, S.; Crawford, K.; Rintoul, L.; Meyer, U., A DRIFTS Study of Reactive Dye States on Cotton Fabric. *Vibrational Spectroscopy* 1997, *15* (1), 103-111.
52. Ceto, X.; O'Mahony, A. M.; Samek, I. A.; Windmiller, J. R.; Valle, M.; Wang, J., Rapid Field Identification of Subjects Involved in Firearm-Related Crimes Based on Electroanalysis Coupled with Advanced Chemometric Data Treatment. *Analytical Chemistry* 2012, *84* (23), 10306-10314.
53. Kher, A.; Stewart, S.; Mulholland, M., Forensic Classification of Paper with Infrared Spectroscopy and Principal Components Analysis. *Near Infrared Spectroscopy* 2005, *13* (1), 225-229.
54. Bajic, S. J.; Aeschliman, D. B.; Saetveit, N. J.; Baldwin, D. P.; Houk, R. S., Analysis of Glass Fragments by Laser Ablation-Inductively Coupled Plasma-Mass Spectrometry and Principal Component Analysis. *Journal of Forensic Science* 2005, *50* (5), 1123-1127.
55. Kher, A. A.; Green, E. V.; Mulholland, M. I., Evaluation of Principal Components Analysis with High-Performance Liquid Chromatography and Photodiode Array Detection for the Forensic Differentiation of Ballpoint Pen Inks. *Journal of Forensic Science* 2001, *46* (4), 878-883.
56. Wilcken, H.; Schulten, H., Quality control of paints: Pyrolysis-mass spectrometry and chemometrics. *Analytica Chimica Acta* 1996, *336* (1), 201-208.
57. Liszewski, E. A.; Lewis, S. W.; Siegel, J. A.; Goodpaster, J. V., Characterization of Automotive Paint Clear Coats by Ultraviolet Absorption Microspectrophotometry with Subsequent Chemometric Analysis. *Applied Spectroscopy* 2010, *64* (10), 1122-1125.
58. Burger, F.; Dawson, M.; Roux, C.; Maynard, P.; Doble, P.; Kirkbride, P., Forensic Analysis of Condom and Personal Lubricants by Capillary Electrophoresis. *Talanta* 2005, *67* (2), 368-376.

59. Morgan, S. L.; Nieuwland, A. A.; Mubarak, C. R.; Hendrix, J. E.; Enlow, E. M.; Vasser, B. J.; Bartick, E. G., Forensic Discrimination of Dyed Textile Fibers using UV-VIS and Fluorescence Microspectrophotometry. Proceedings of the European Fibres Group Annual Meeting, 2004 May 24, Prague, Czechoslovakia.
60. Kawahara, F. K.; Santner, J. F.; Julian, E. C., Characterization of Heavy Residual Fuel Oils and Asphalts by Infrared Spectrophotometry Using Statistical Discriminant Function Analysis. *Analytical Chemistry* 1974, *46* (2), 266-273.
61. Hicks, T.; Monard Sermier, F.; Goldmann, T.; Brunelle, A.; Champod, C.; Margot, P., The Classification and Discrimination of Glass Fragments Using Non Destructive Energy Dispersive X-ray μ fluorescence. *Forensic Science International* 2003, *137* (2-3), 107-118.
62. Cheung, W.; Xu, Y.; Paul Thomas, C. L.; Goodacre, R., Discrimination of Bacteria Using Pyrolysis-Gas Chromatography-Differential Mobility Spectrometry (Py-GC-DMS) and Chemometrics. *Analyst* 2009, *134* (3), 557-563.
63. Kher, A.; Mulholland, M.; Green, E.; Reedy, B., Forensic Classification of Ballpoint Pen Inks Using High Performance Liquid Chromatography and Infrared Spectroscopy with Principal Components Analysis and Linear Discriminant Analysis. *Vibrational Spectroscopy* 2006, *40* (2), 270-277.
64. Doble, P.; Sandercock, M.; Du Pasquier, E.; Petocz, P.; Roux, C.; Dawson, M., Classification of Premium and Regular Gasoline by Gas Chromatography/Mass Spectrometry, Principal Component Analysis and Artificial Neural Networks. *Forensic Science International* 2003, *132* (1), 26-39.
65. SWGMAT, Forensic Fiber Examination Guidelines. *Forensic Science Communications* 1999, *1* (1).
66. Cochran, K. H.; Barry, J. A.; Muddiman, D. C.; Hinks D., Direct Analysis of Textile Fabrics and Dyes Using IR Matrix-Assisted Laser Desorption Electrospray Ionization (MALDESI) Mass Spectrometry. *Analytical Chemistry* 2012, *85* (2), 831-836.
67. Ingle Jr., JD; Crouch, S.R. Ultraviolet and Visible Molecular Absorption Spectrophotometry. In: *Spectrochemical Analysis*, Prentice Hall, Englewood Cliffs: NJ, 1988; pp 372-380.

68. Bouffard, S.P.; Sommer, A.J.; Katon, J.E.; Godber, S., Use of Molecular Microspectroscopy to Characterize Pigment-Loaded Polypropylene Single Fibers. *Applied Spectroscopy* 1994, 48 (11), 1387-1393

APPENDIX

APPENDIX. ADDITIONAL FIBER FIGURES

A.1. Blue Acrylic Fibers

A.1.1. Training Set Exemplar Spectra

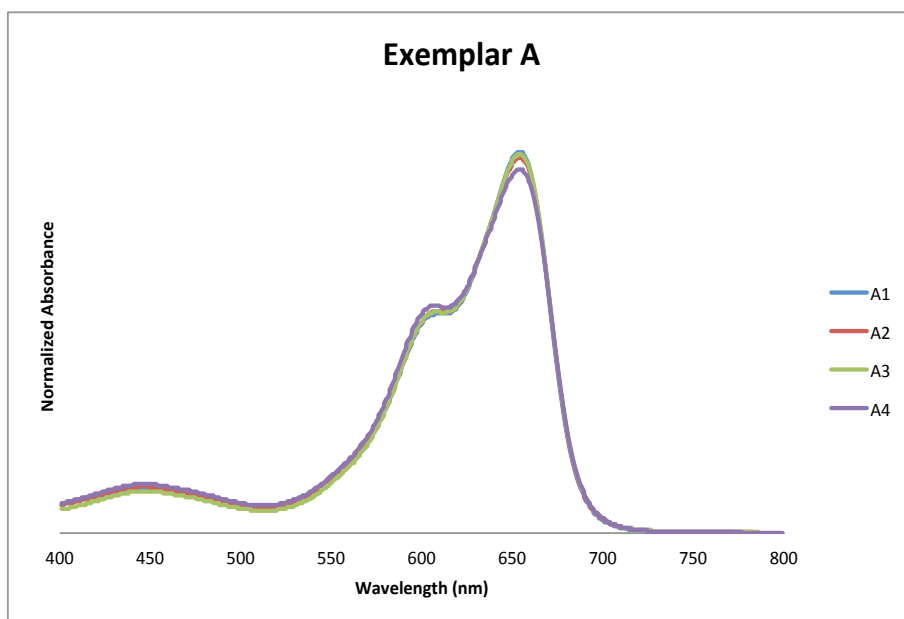


Figure A1. Averaged Spectra for Exemplar A

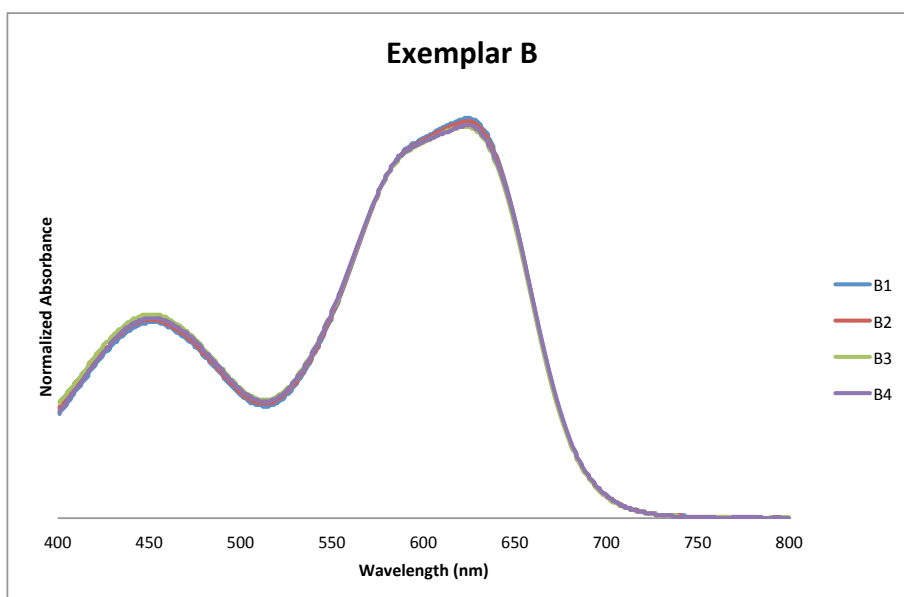


Figure A2. Averaged Spectra For Exemplar B

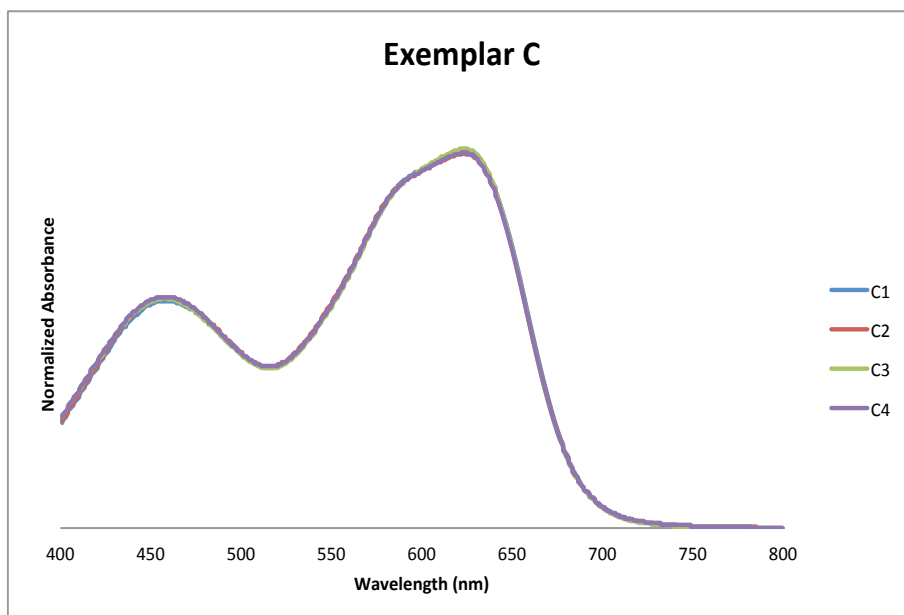


Figure A3. Averaged Spectra for Exemplar C

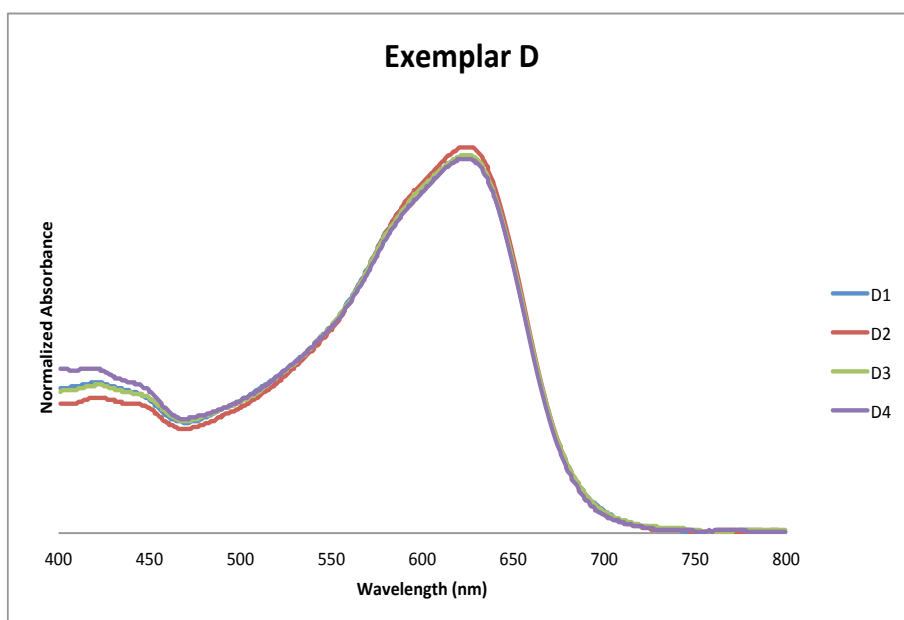


Figure A4. Averaged Spectra for Exemplar D

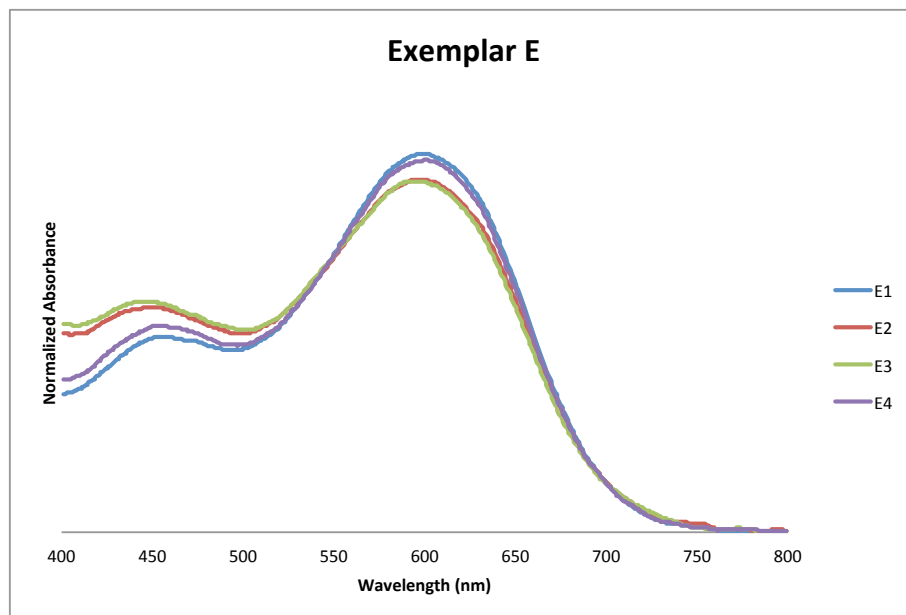


Figure A5. Averaged Spectra for Exemplar E

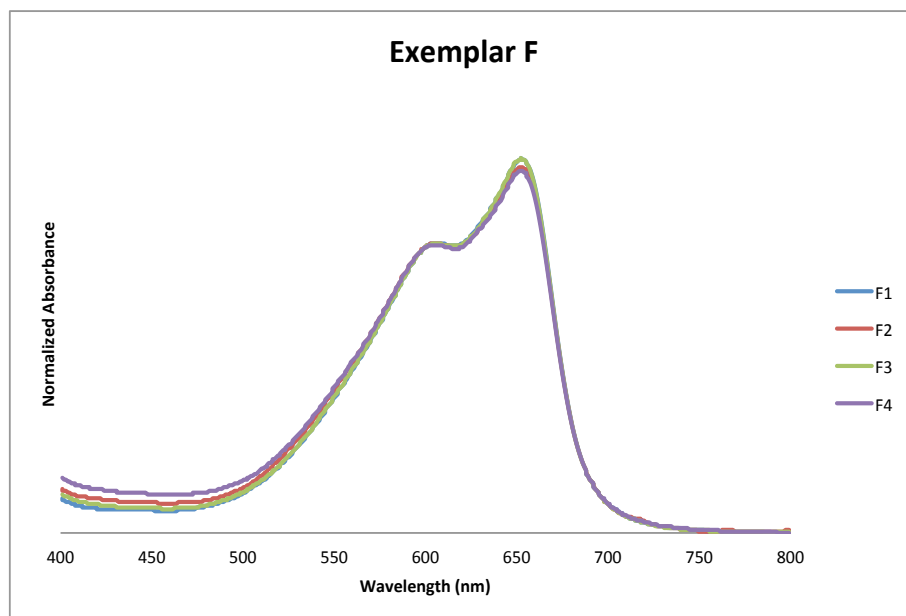


Figure A6. Averaged Spectra for Exemplar F

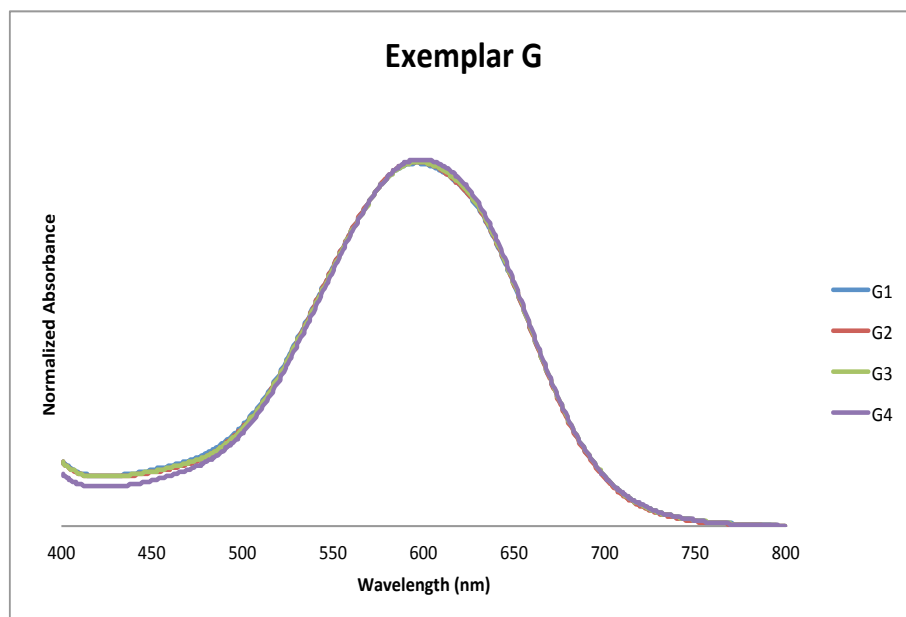


Figure A7. Averaged Spectra for Exemplar G

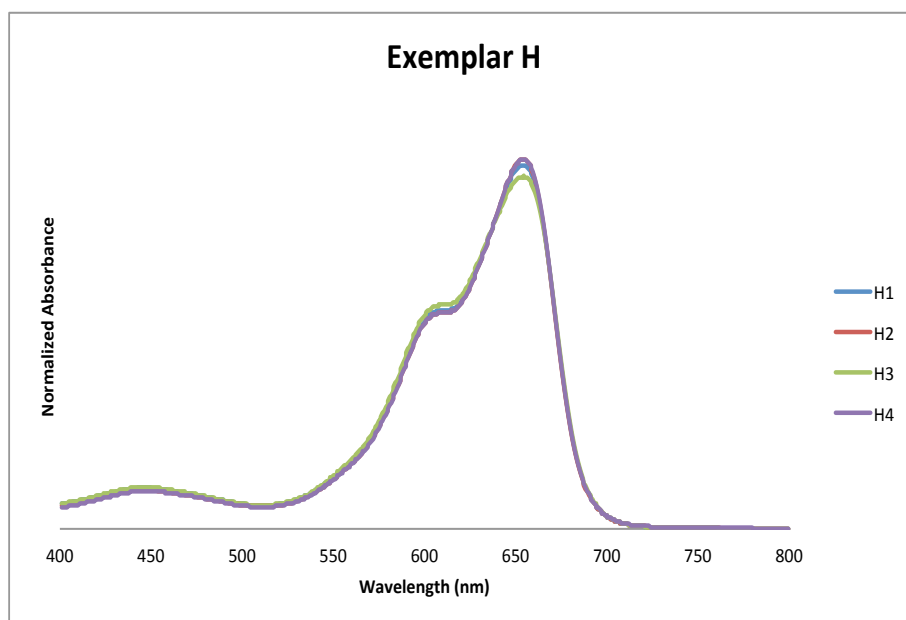


Figure A8. Averaged Spectra for Exemplar H

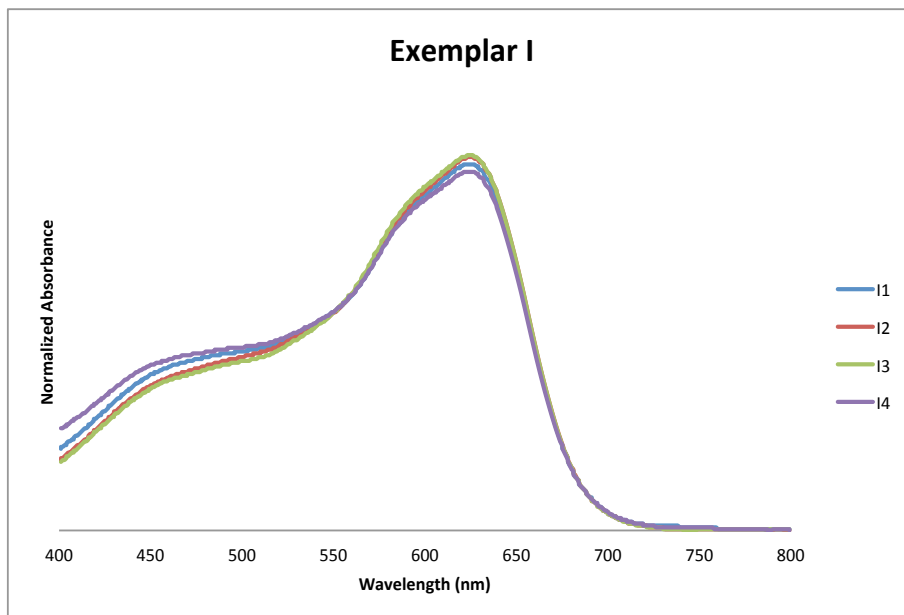


Figure A9. Averaged Spectra for Exemplar I

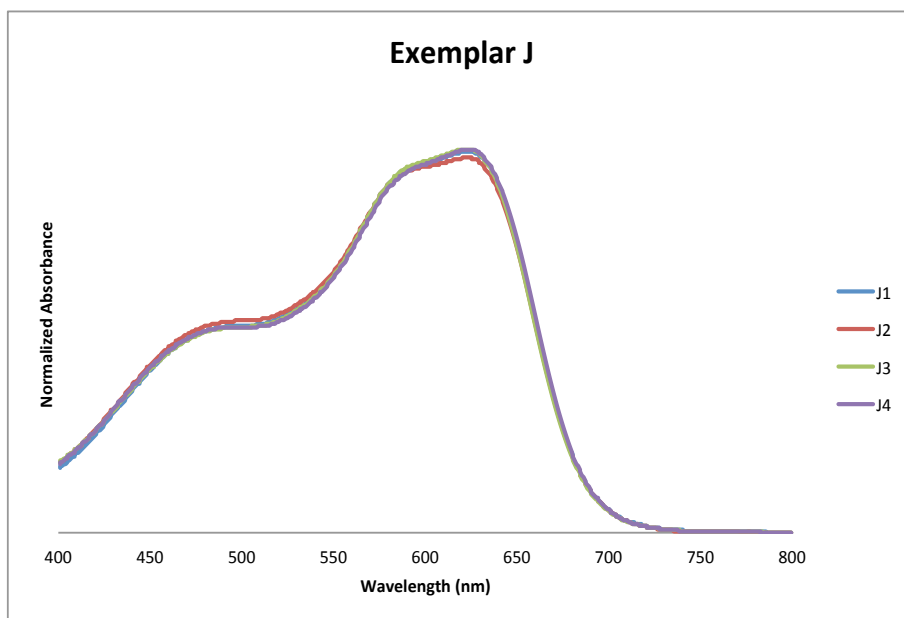


Figure A10. Averaged Spectra for Exemplar J

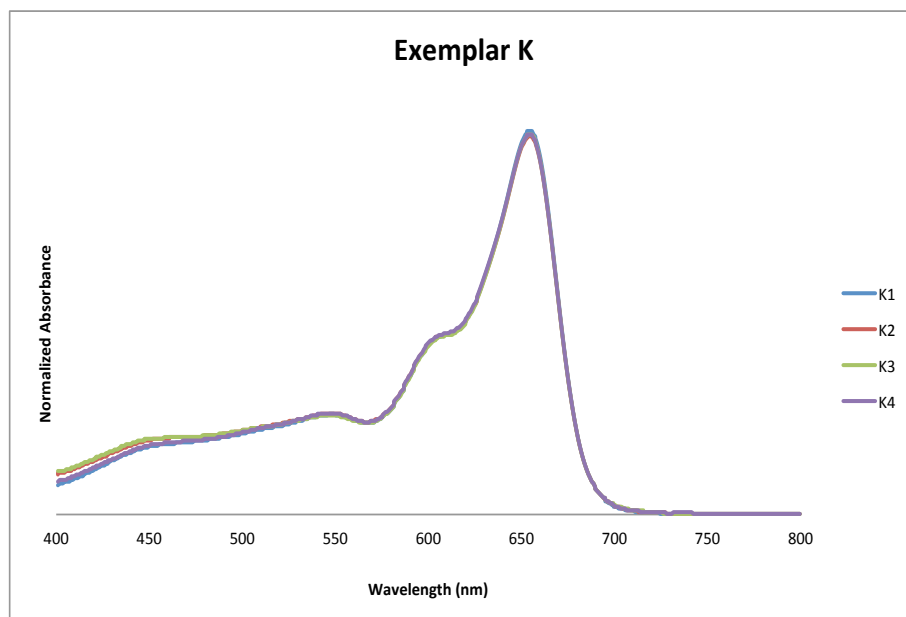


Figure A11. Averaged Spectra for Exemplar K

A.1.2. External Validation Exemplar Spectra

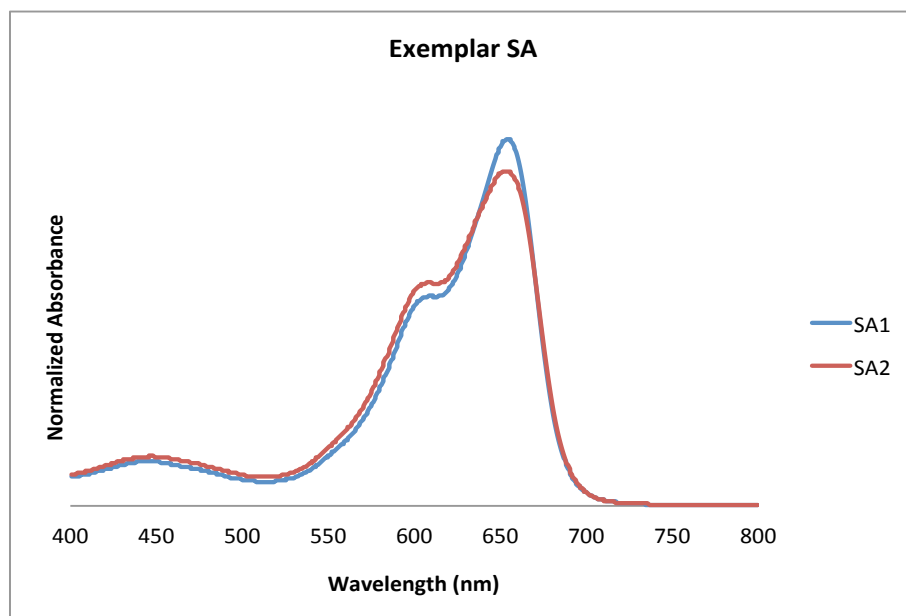


Figure A12. Averaged Spectra for Exemplar SA

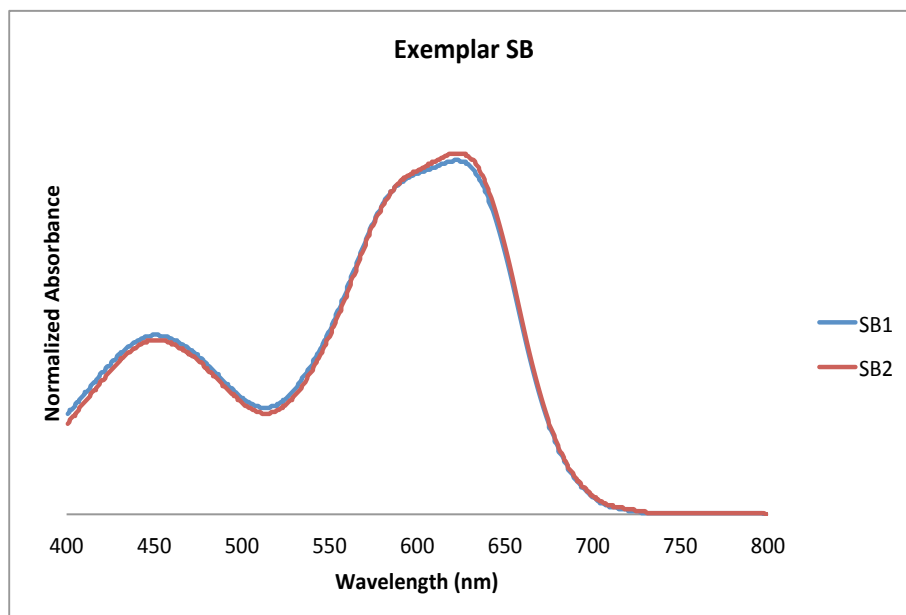


Figure A13. Averaged Spectra for Exemplar SB

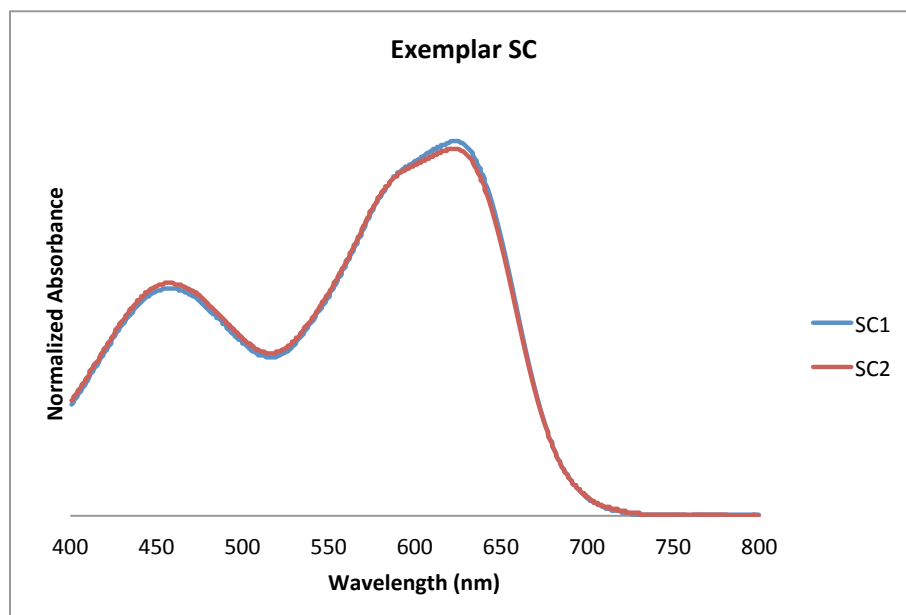


Figure A14. Averaged Spectra for Exemplar SC

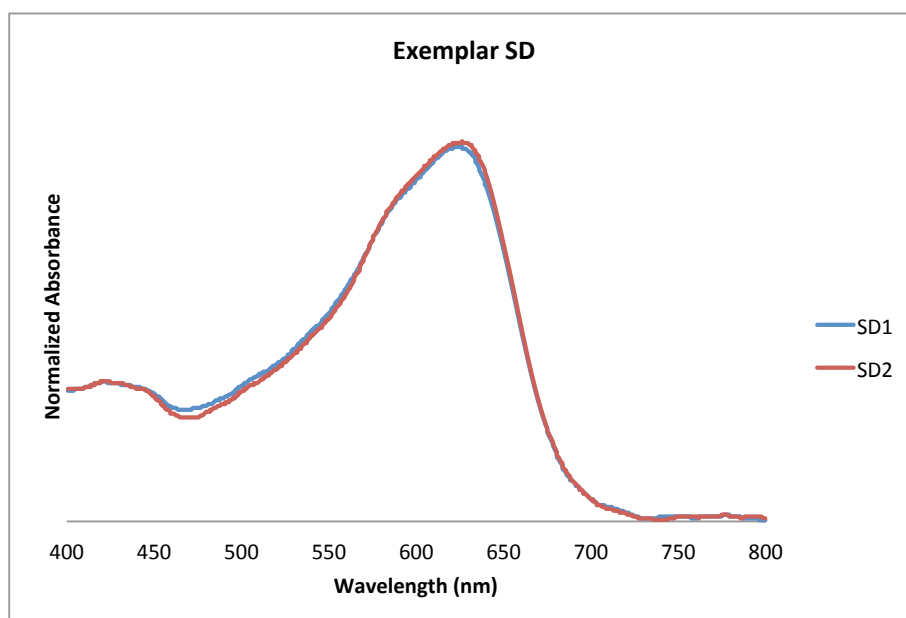


Figure A15. Averaged Spectra for Exemplar SD

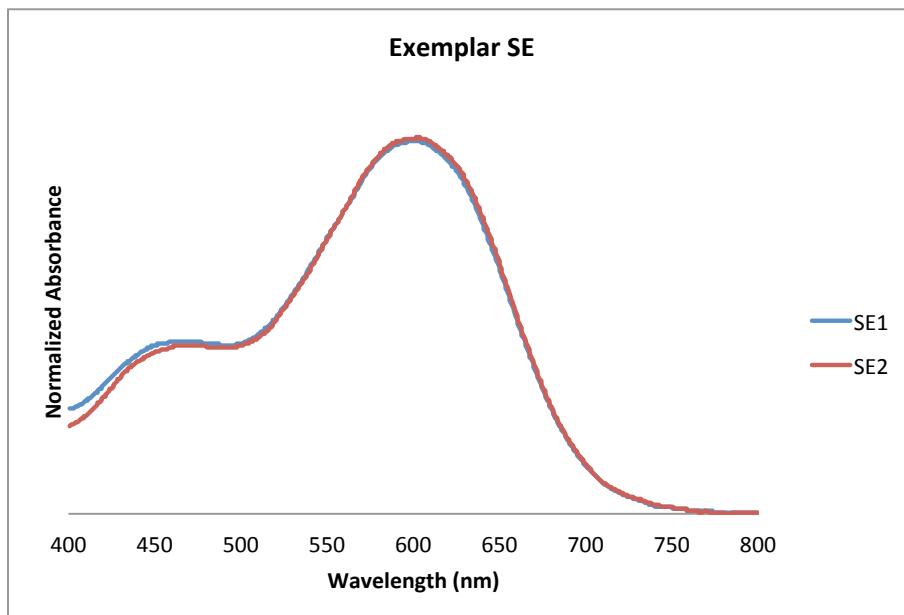


Figure A16. Averaged Spectra for Exemplar SE

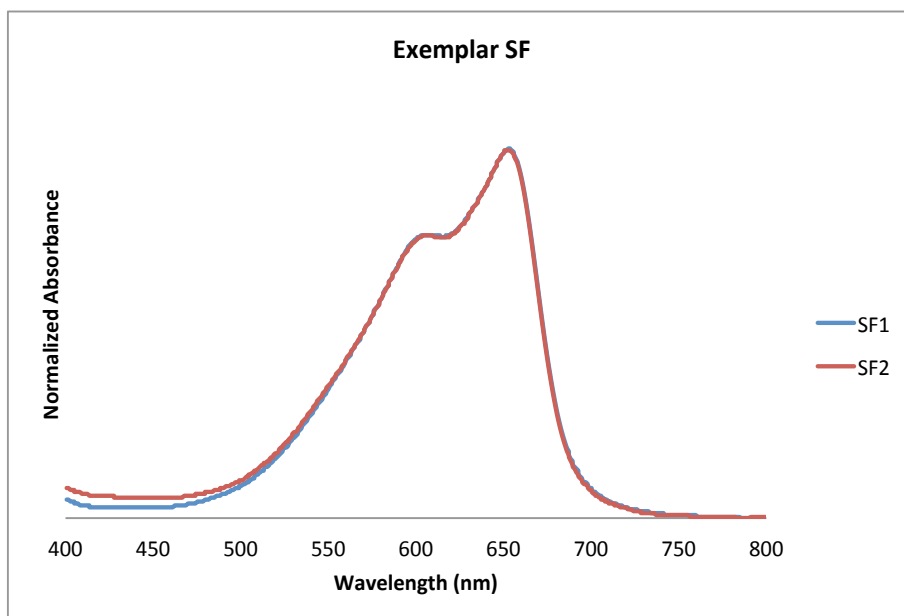


Figure A17. Averaged Spectra for Exemplar SF

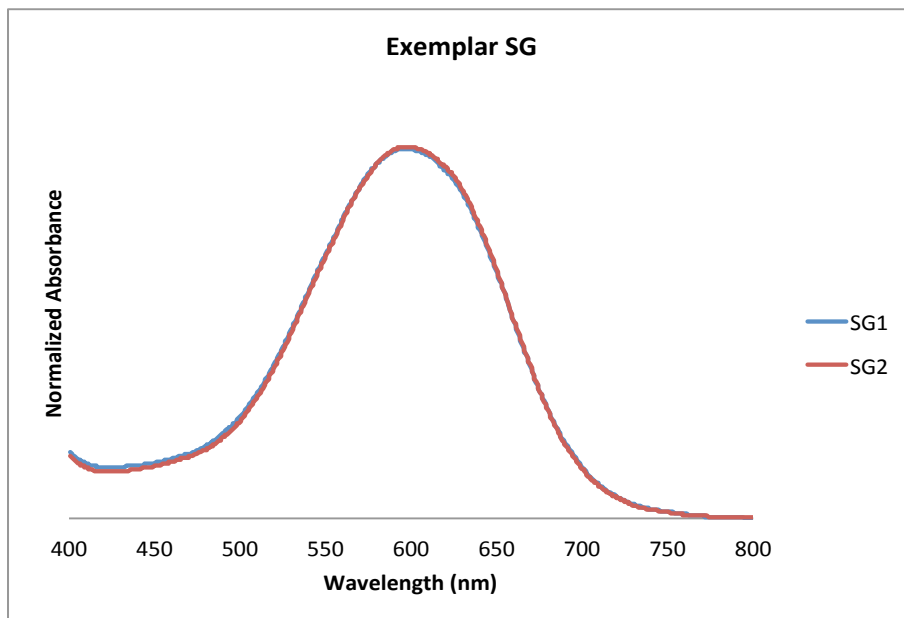


Figure A18. Averaged Spectra for Exemplar SG

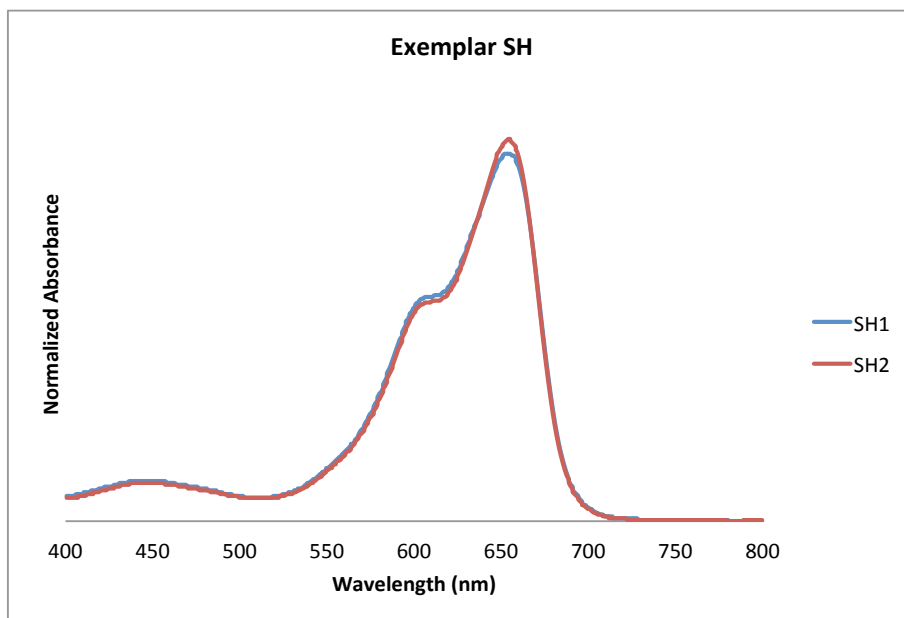


Figure A19. Averaged Spectra for Exemplar SH

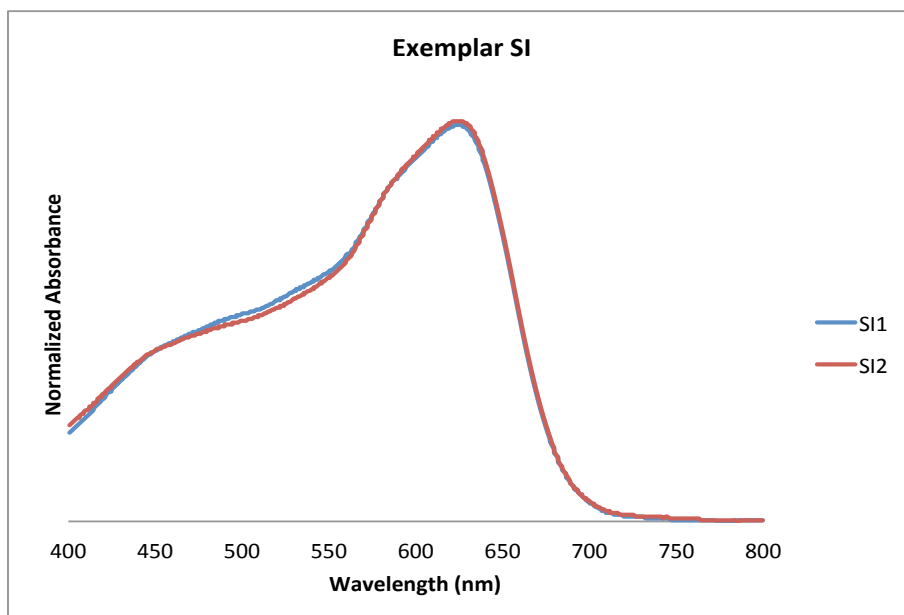


Figure A20. Averaged Spectra for Exemplar SI

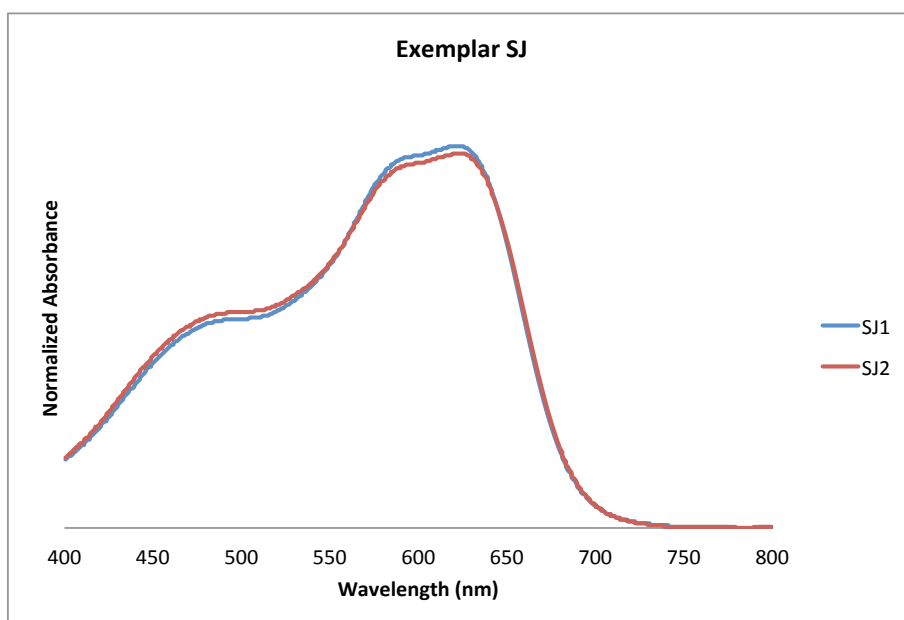


Figure A21. Averaged Spectra for Exemplar SJ

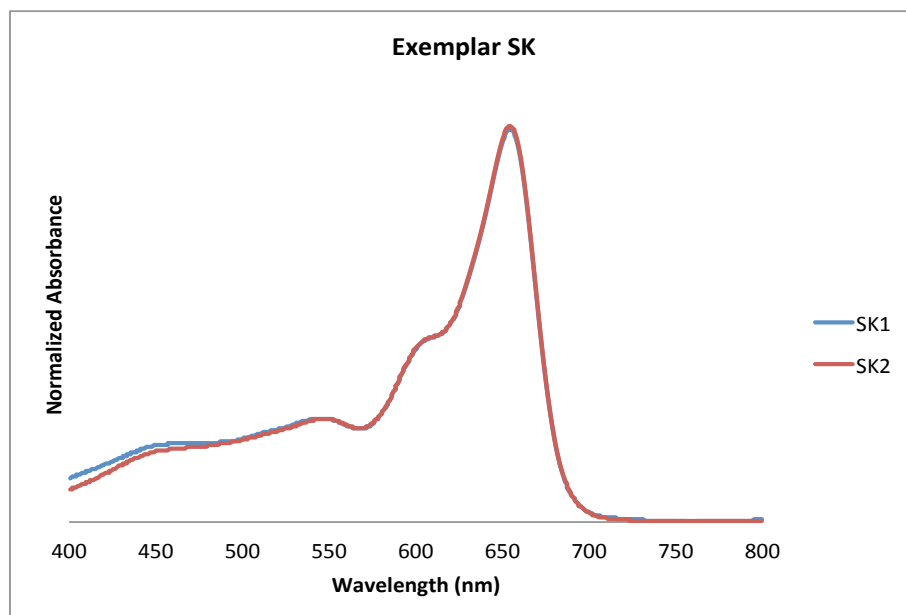


Figure A22. Averaged Spectra for Exemplar SK

A.2. Yellow Polyester Fibers

A.2.1. Calibration Plots

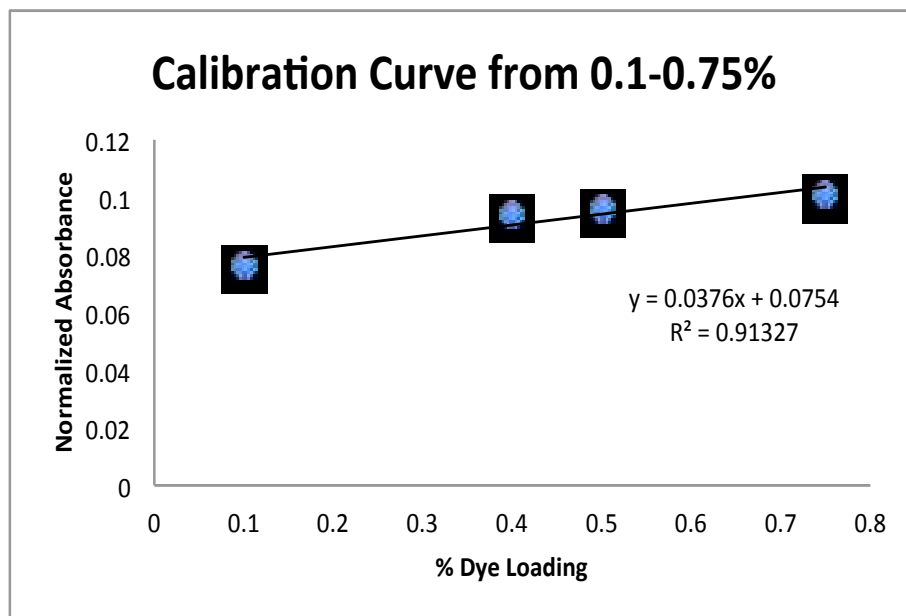


Figure A23. Calibration Plot utilizing the Max Wavelength Approach for the Exemplar with 0.2% Dye Loading

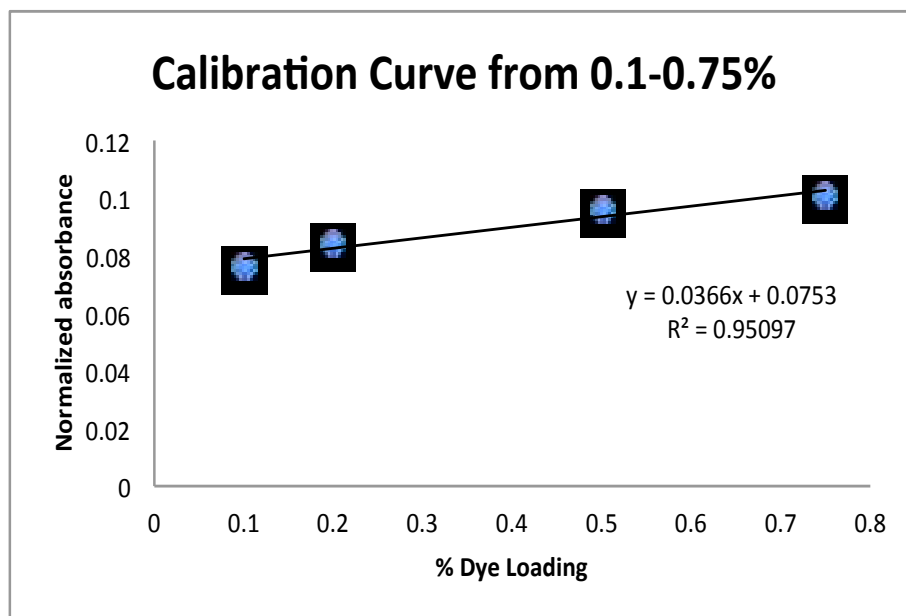


Figure A24. Calibration Plot utilizing the Max Wavelength Approach for the Exemplar with 0.4% Dye Loading

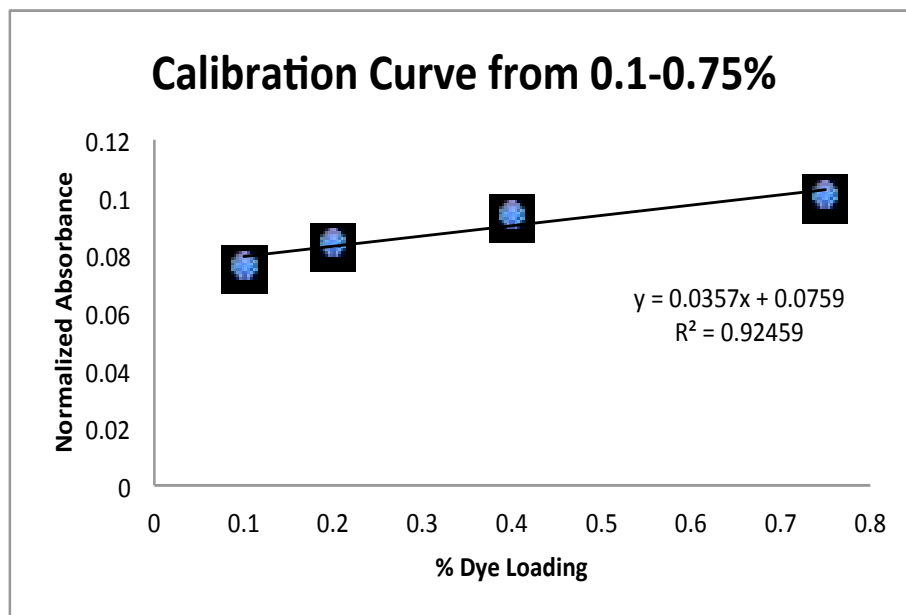


Figure A25. Calibration Plot utilizing the Max Wavelength Approach for the Exemplar with 0.5% Dye Loading

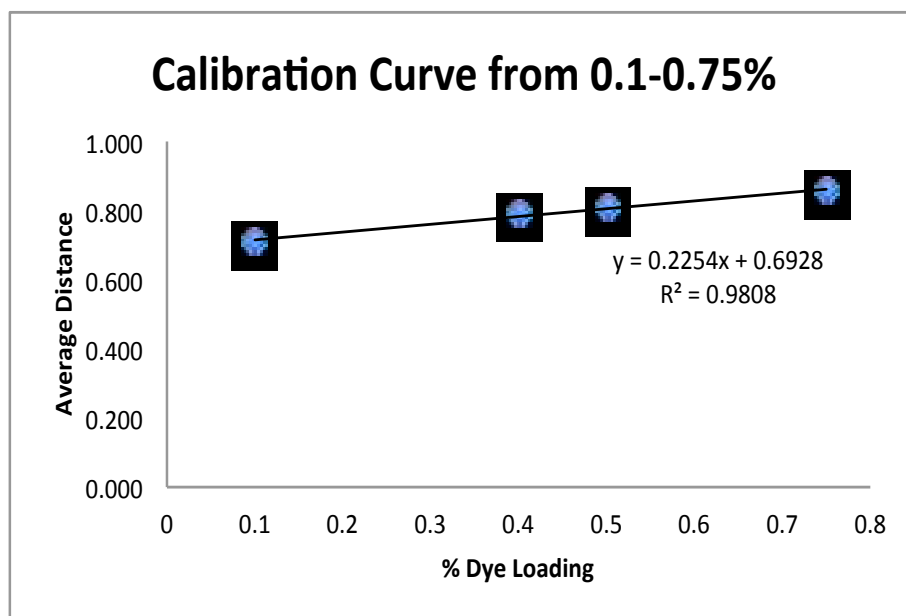


Figure A26. Calibration Plot utilizing the Proximity Matrix Approach for the Exemplar with 0.2% Dye Loading

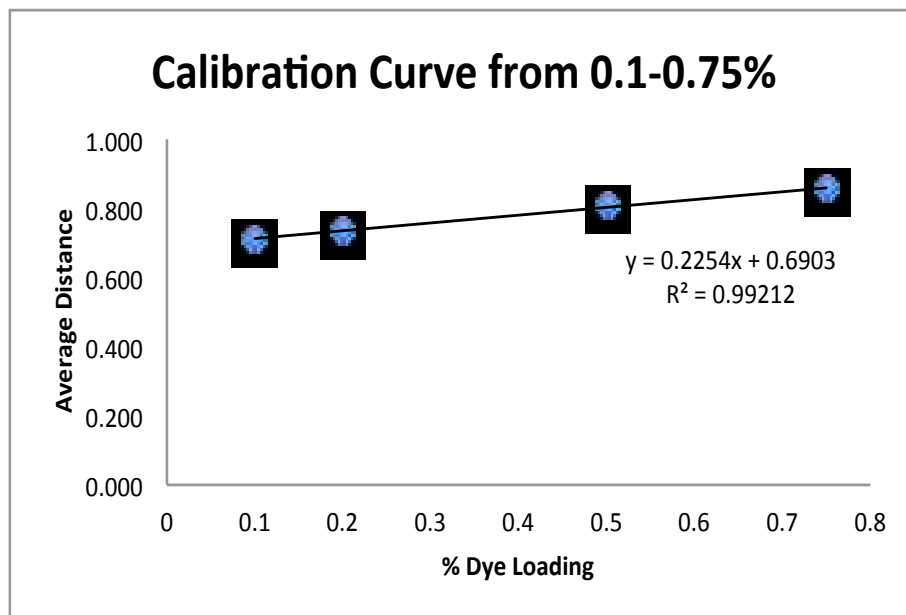


Figure A27. Calibration Plot utilizing the Proximity Matrix Approach for the Exemplar with 0.4% Dye Loading

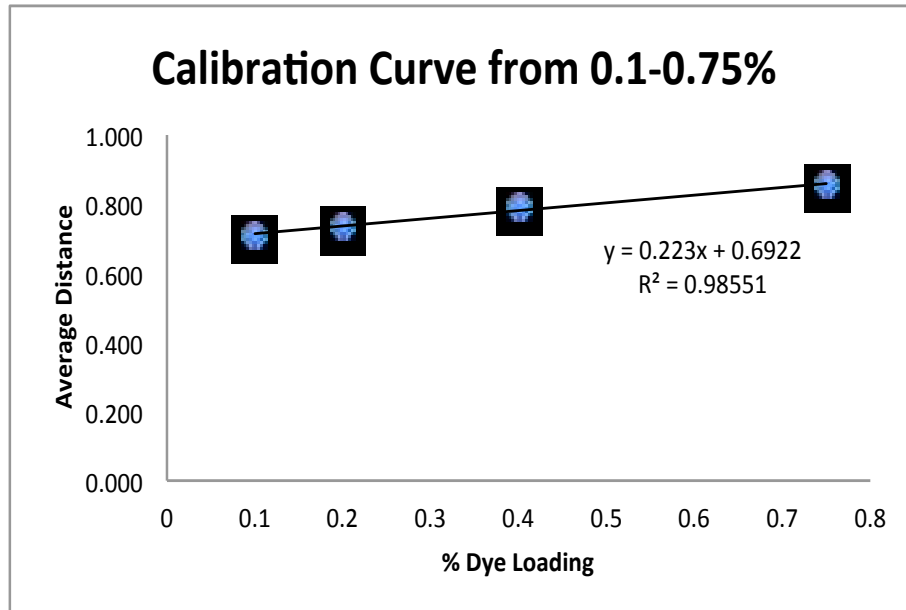


Figure A28. Calibration Plot utilizing the Proximity Matrix Approach for the Exemplar with 0.5% Dye Loading

A.2.2. Training Set Exemplar Spectra

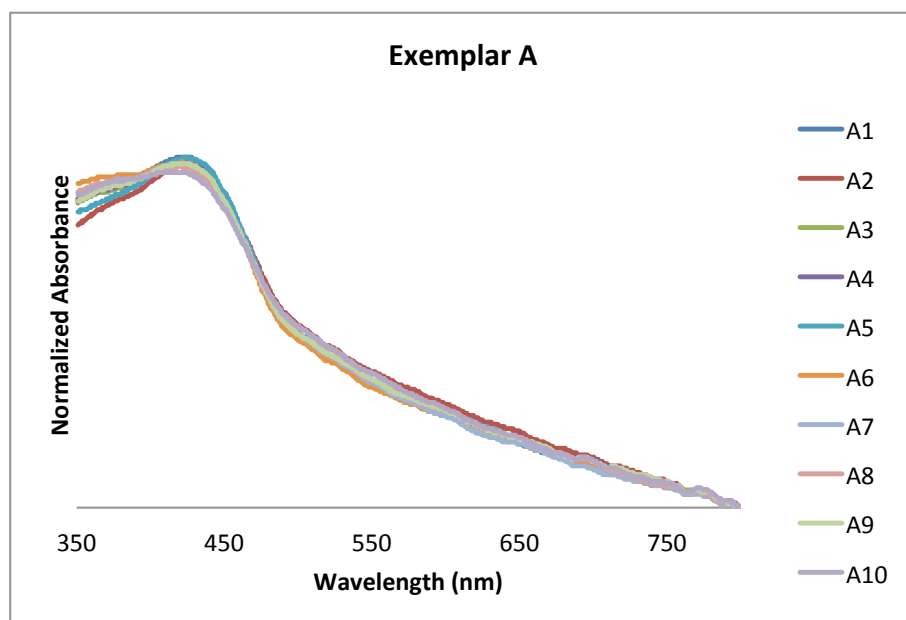


Figure A29. Averaged Spectra for Exemplar A

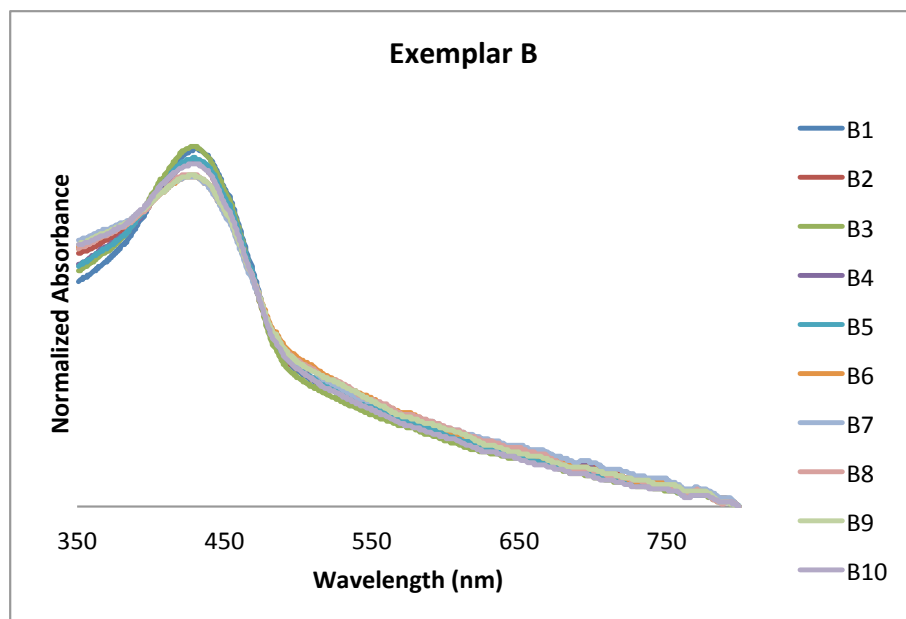


Figure A30. Averaged Spectra for Exemplar B

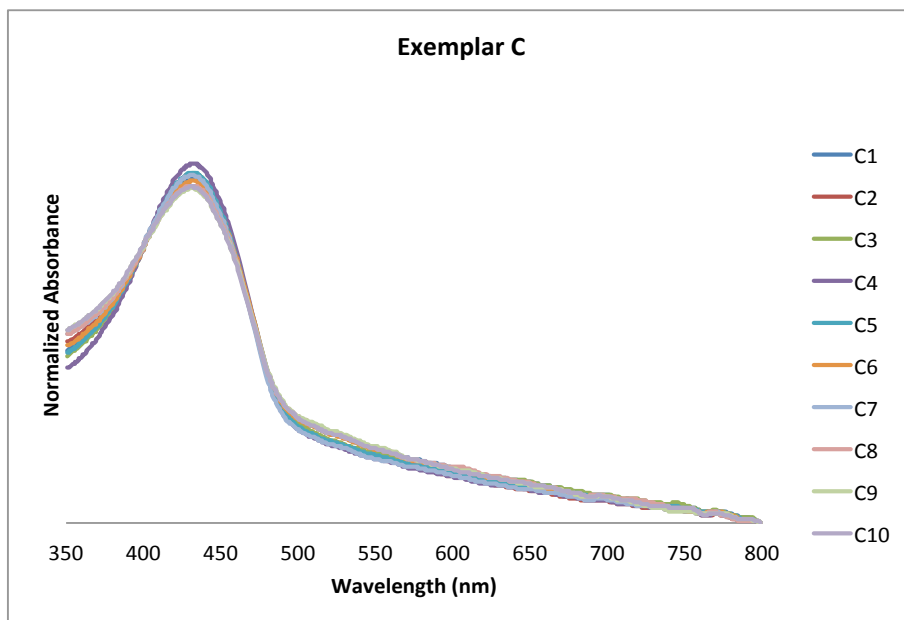


Figure A31. Averaged Spectra for Exemplar C

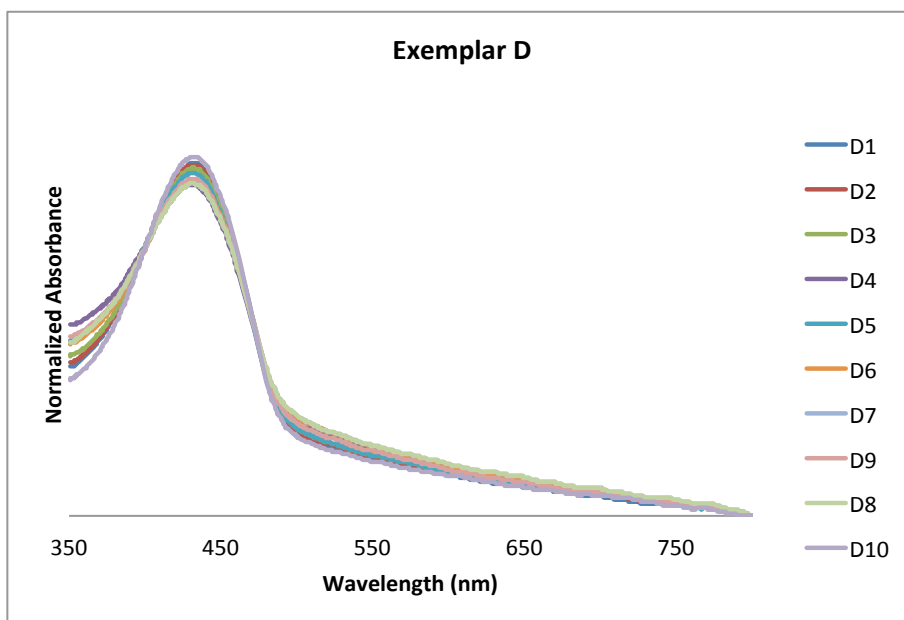


Figure A32. Averaged Spectra for Exemplar D

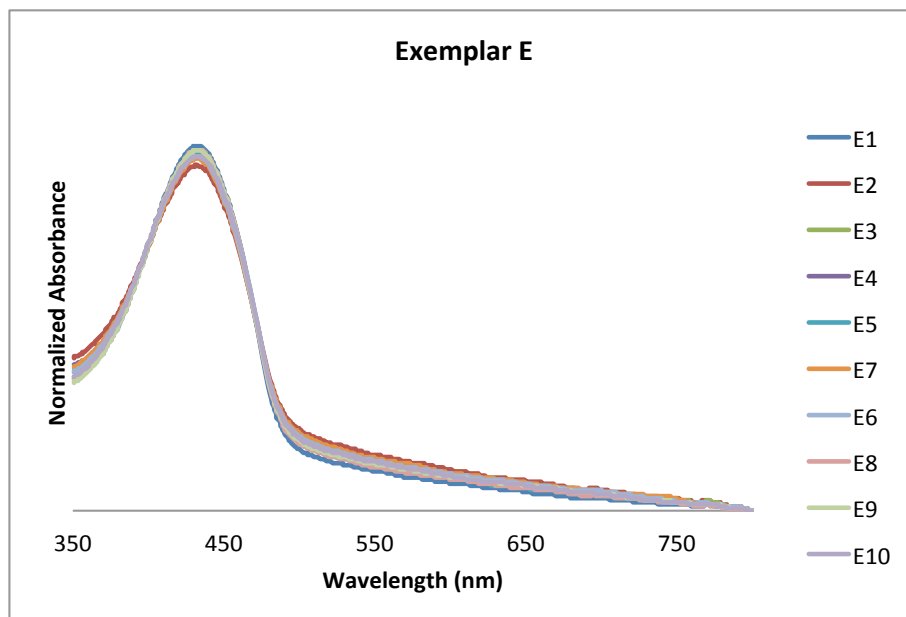


Figure A33. Averaged Spectra for Exemplar E

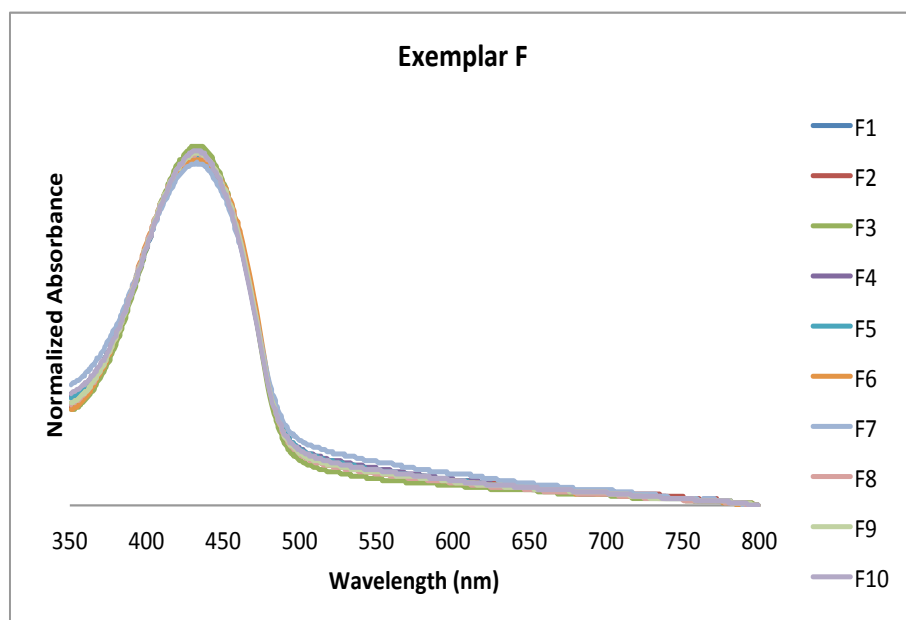


Figure A34. Averaged Spectra for Exemplar F

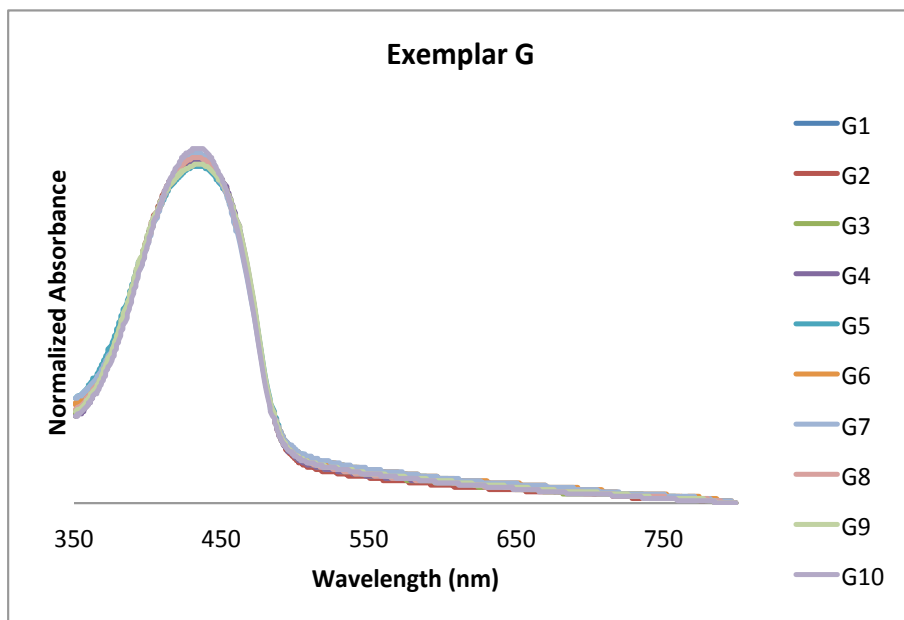


Figure A35. Averaged Spectra for Exemplar G

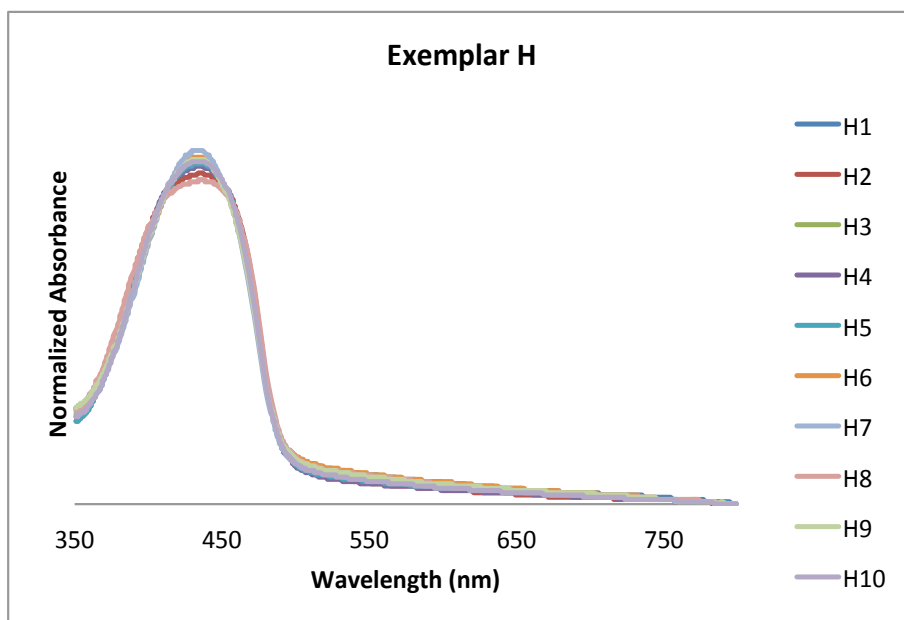


Figure A36. Averaged Spectra for Exemplar H

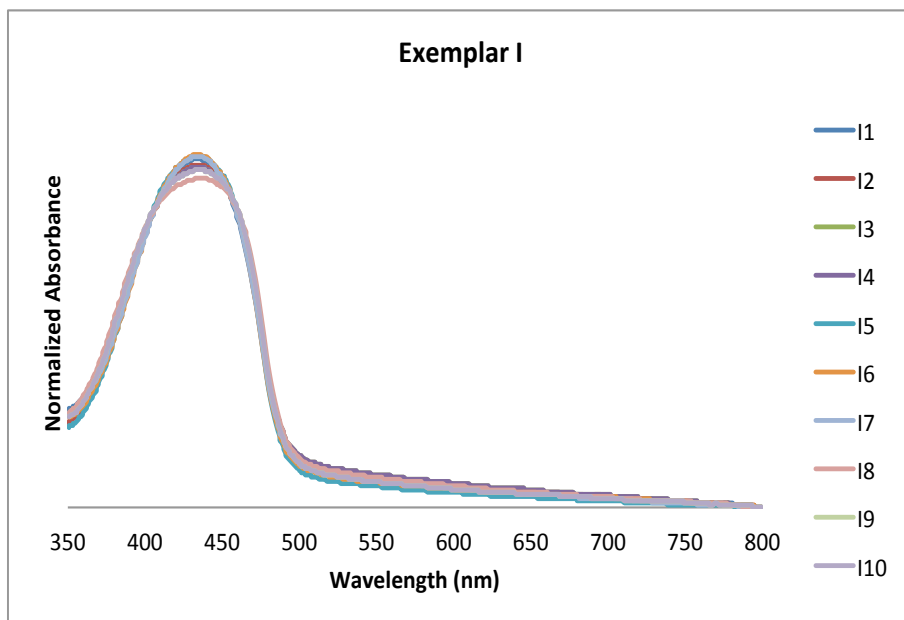


Figure A37. Averaged Spectra for Exemplar I

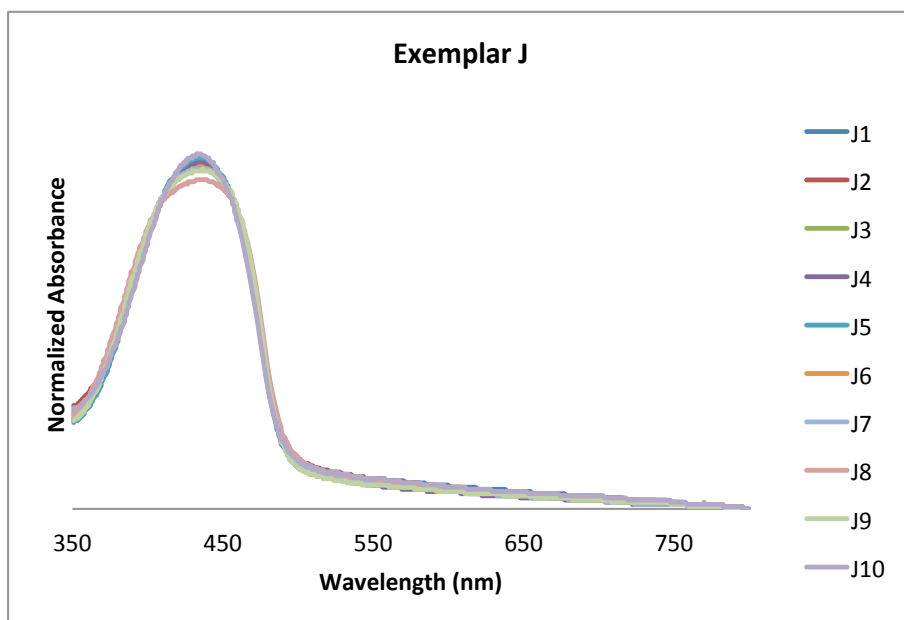


Figure A38. Averaged Spectra for Exemplar J

A.2.3. External Validation Exemplar Spectra

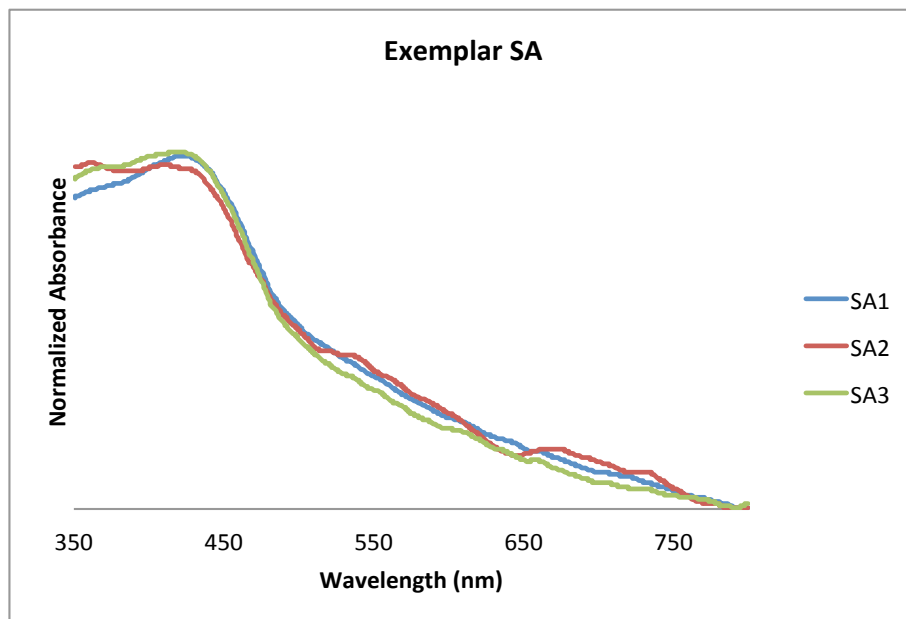


Figure A39. Averaged Spectra for Exemplar SA

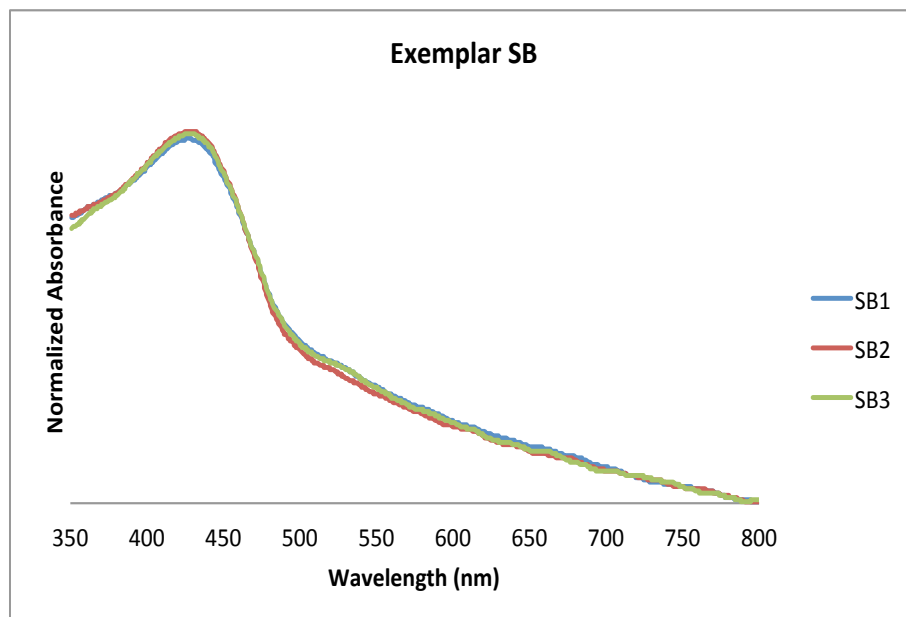


Figure A40. Averaged Spectra for Exemplar SB

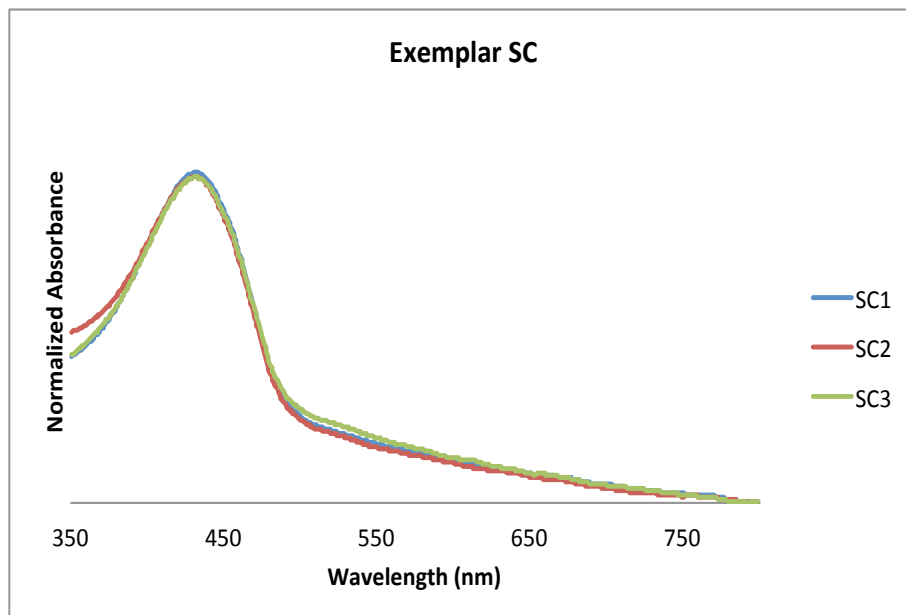


Figure A41. Averaged Spectra for Exemplar SC

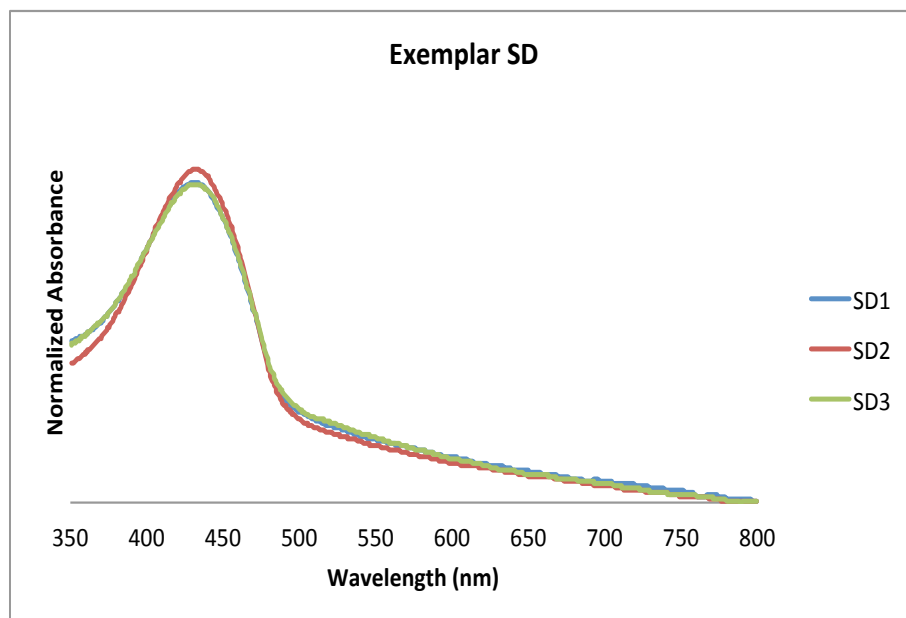


Figure A42. Averaged Spectra for Exemplar SD

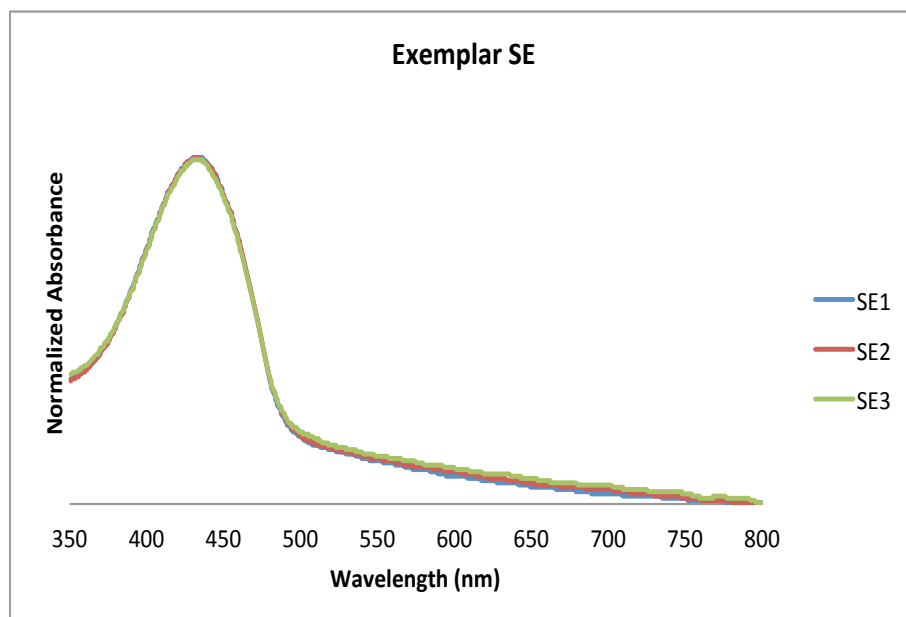


Figure A43. Averaged Spectra for Exemplar SE

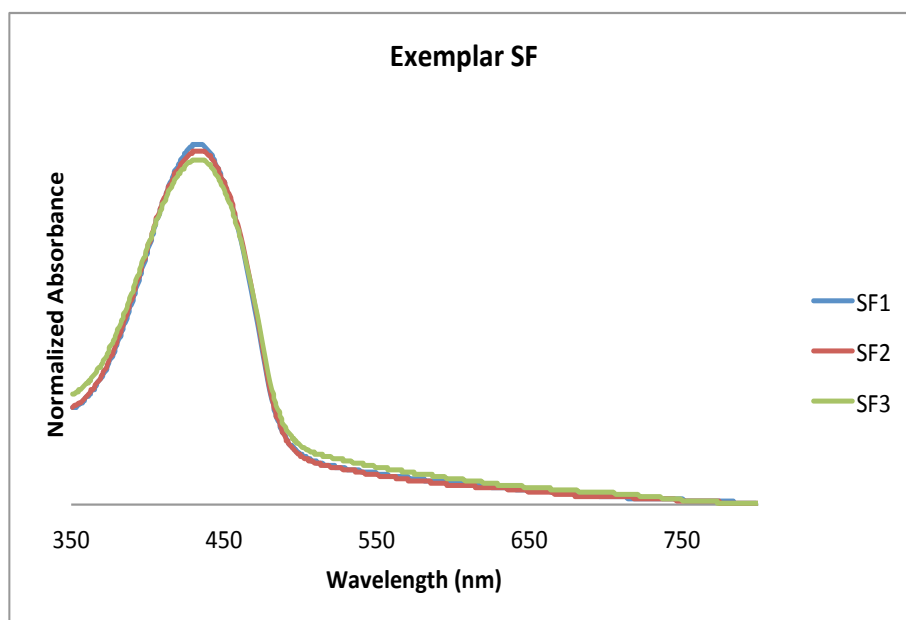


Figure A44. Averaged Spectra for Exemplar SF

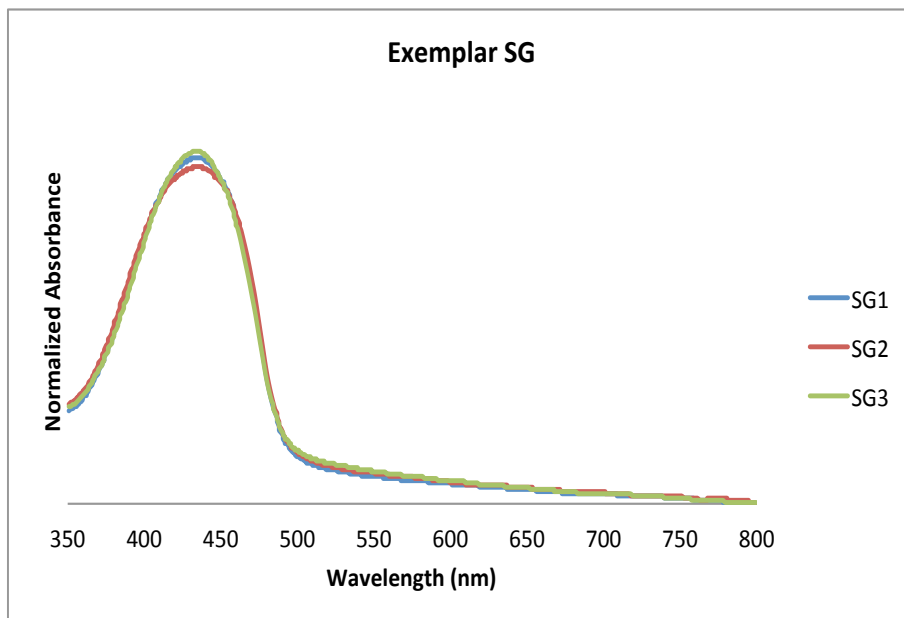


Figure A45. Averaged Spectra for Exemplar SG

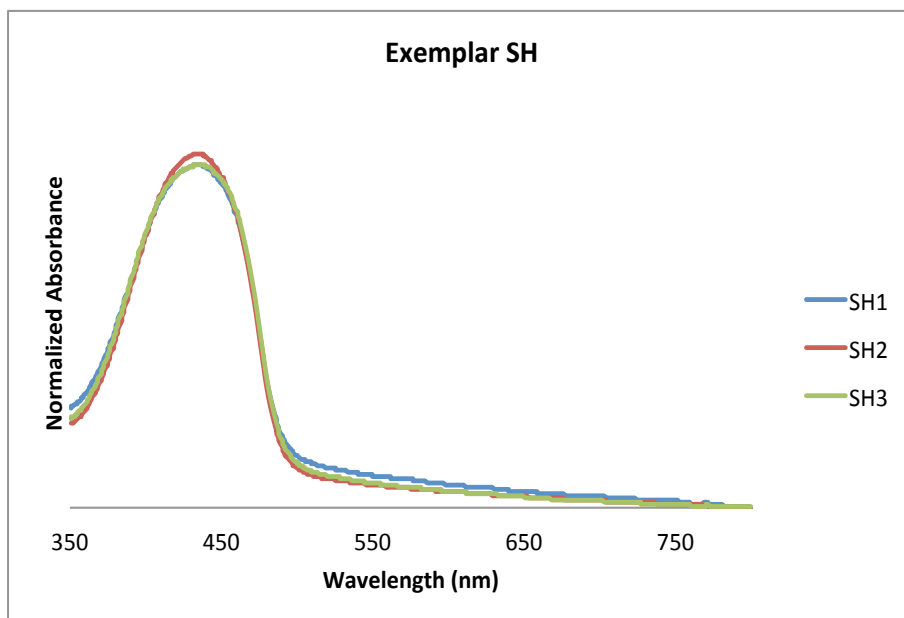


Figure A46. Averaged Spectra for Exemplar SH

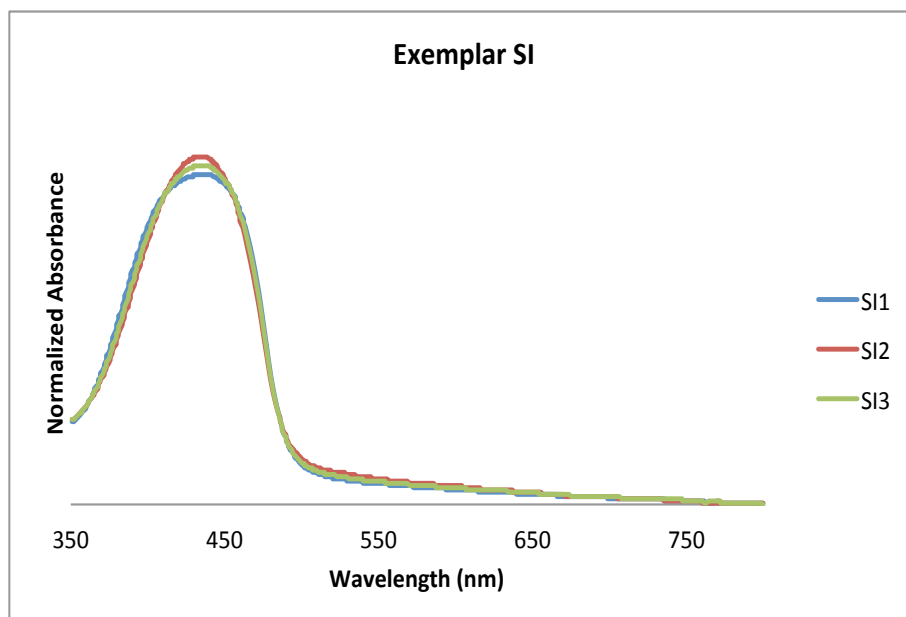


Figure A47. Averaged Spectra for Exemplar SI

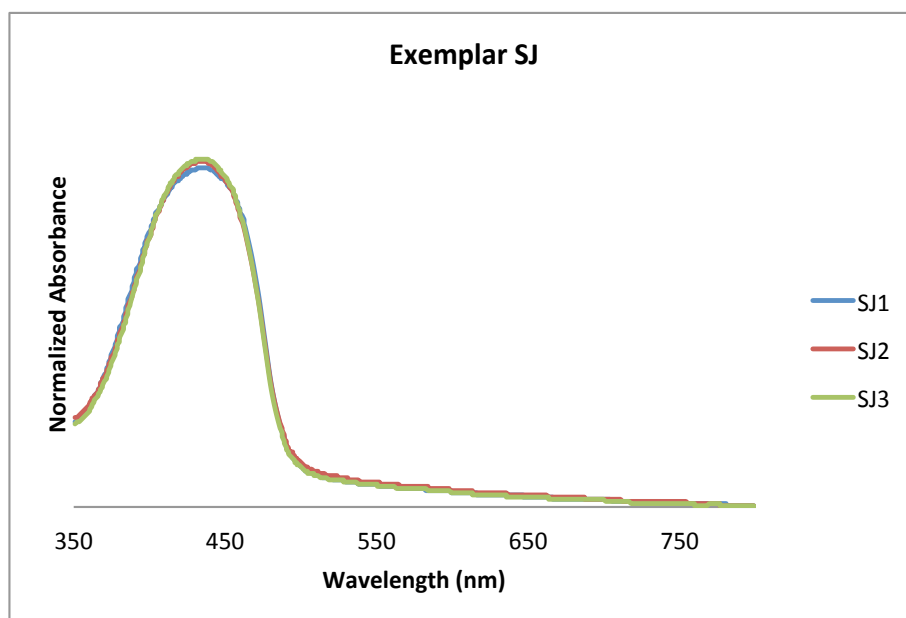


Figure A48. Averaged Spectra for Exemplar SJ

A.2.4. PCA Projections of Pair-Wise Comparison Data

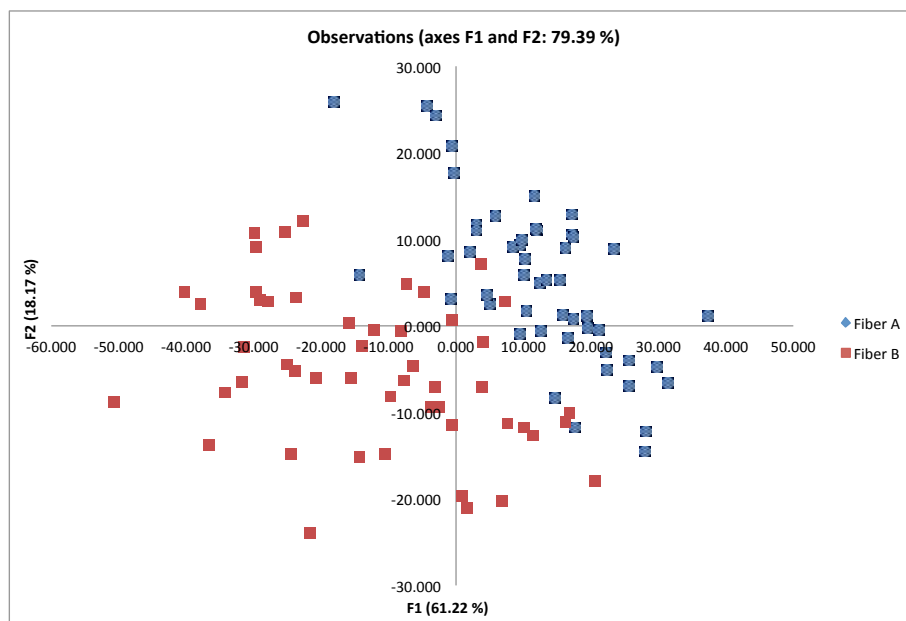


Figure A49. PCA Observations Plot of Exemplars A and B

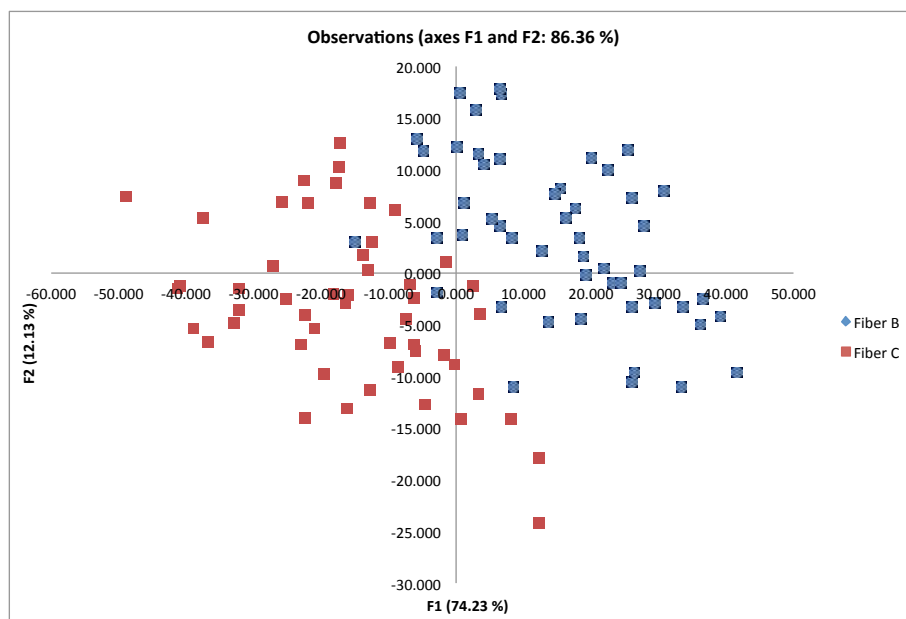


Figure A50. PCA Observations Plot of Exemplars B and C

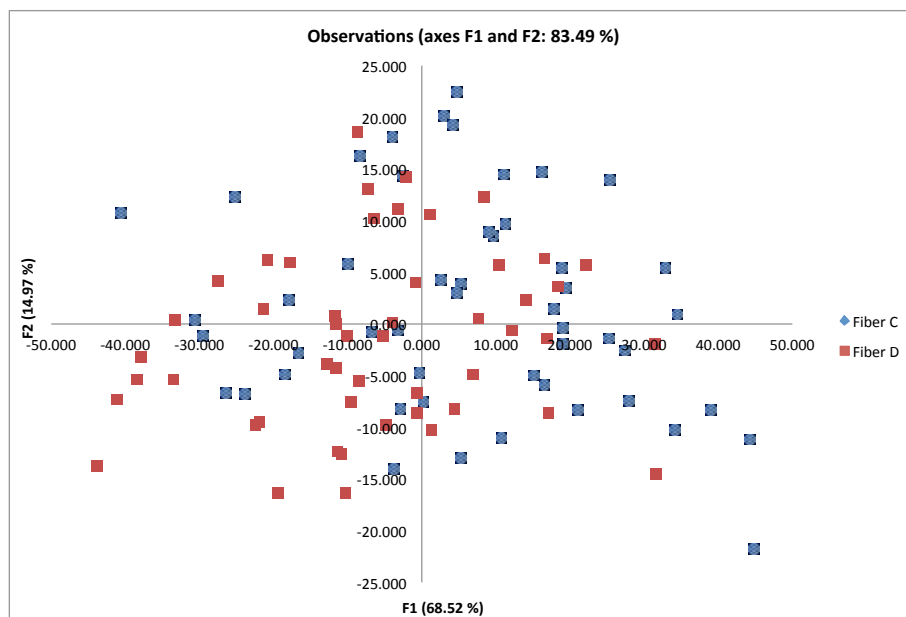


Figure A51. PCA Observations Plot of Exemplars C and D

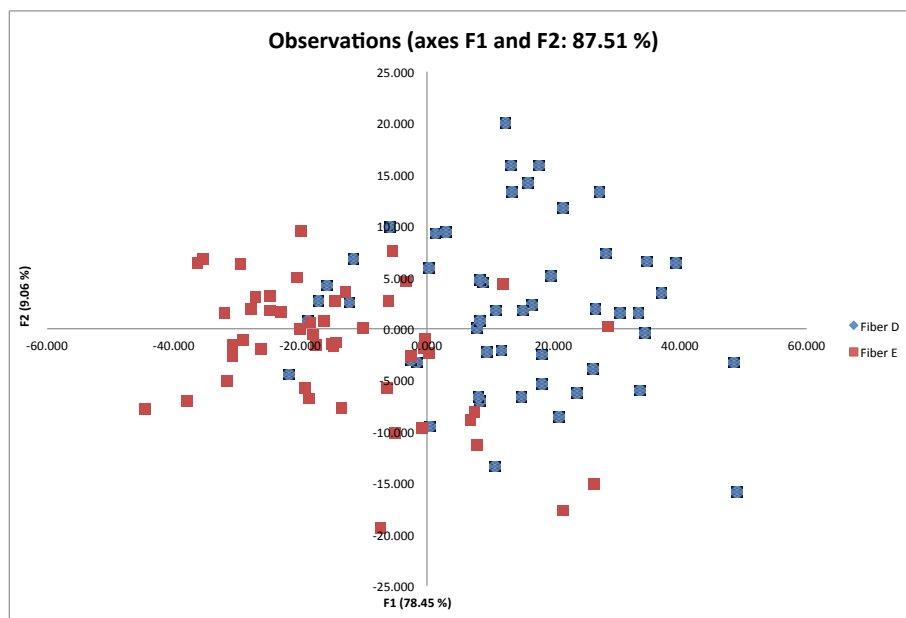


Figure A52. PCA Observations Plot of Exemplars D and E

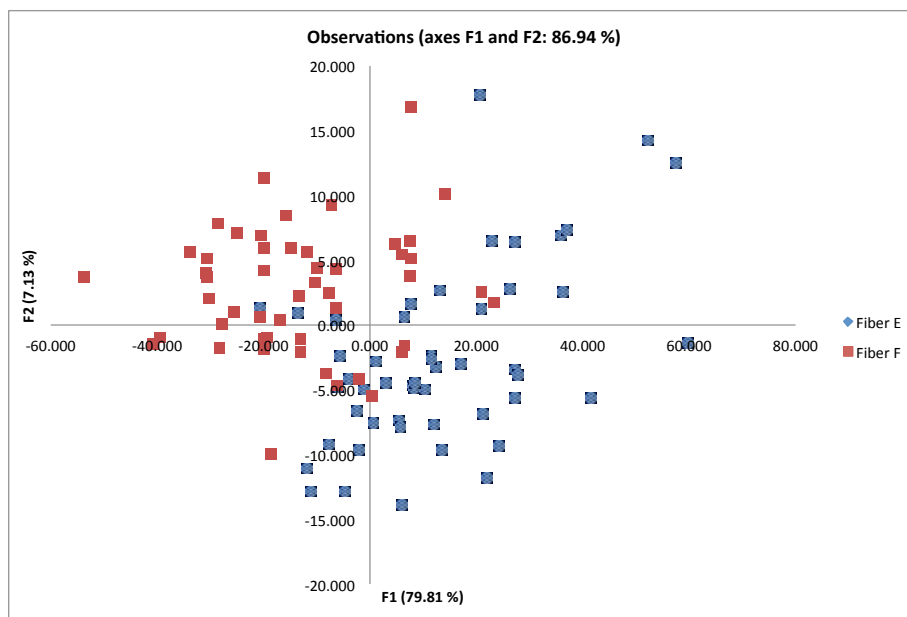


Figure A53. PCA Observations Plot of Exemplars E and F

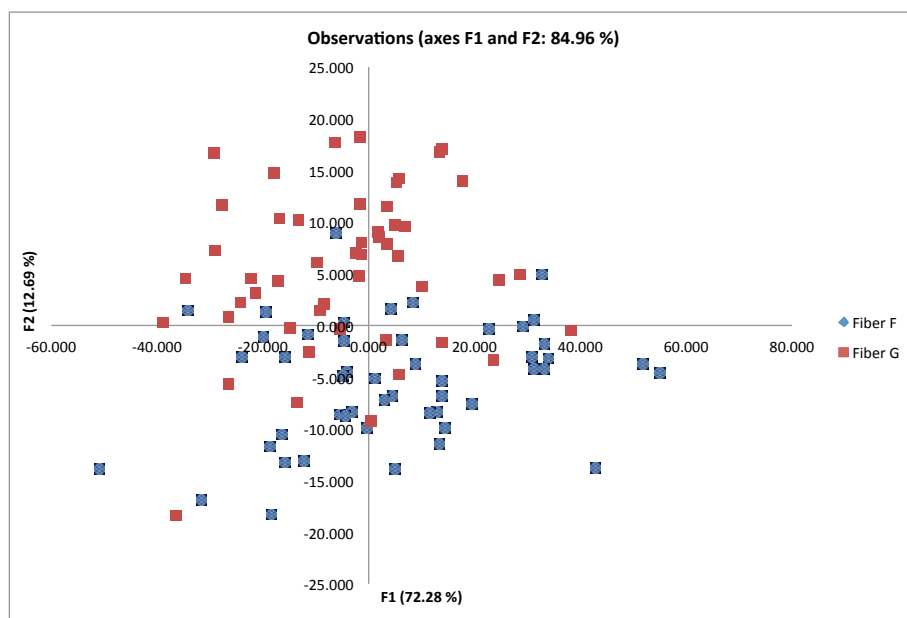


Figure A54. PCA Observations Plot of Exemplars F and G

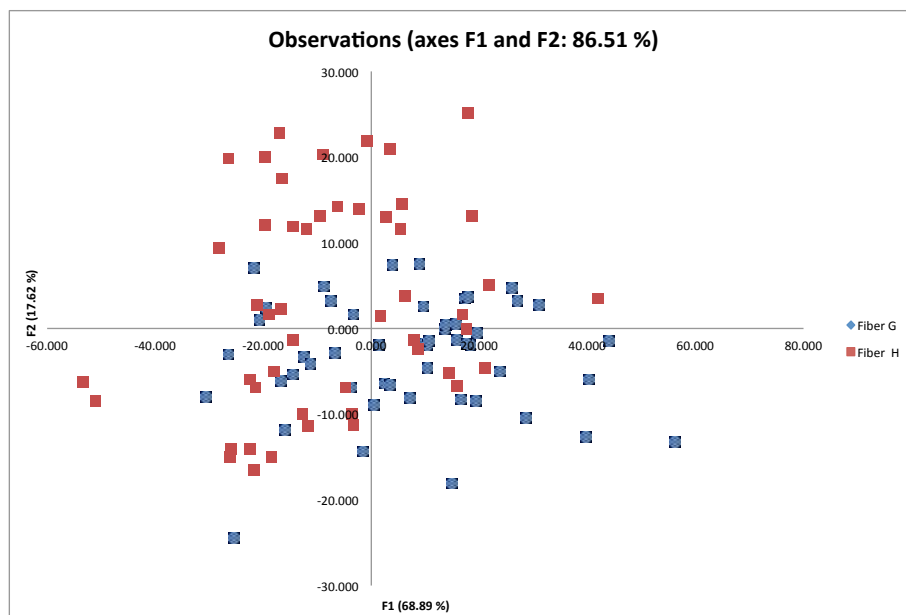


Figure A55. PCA Observations Plot of Exemplars G and H

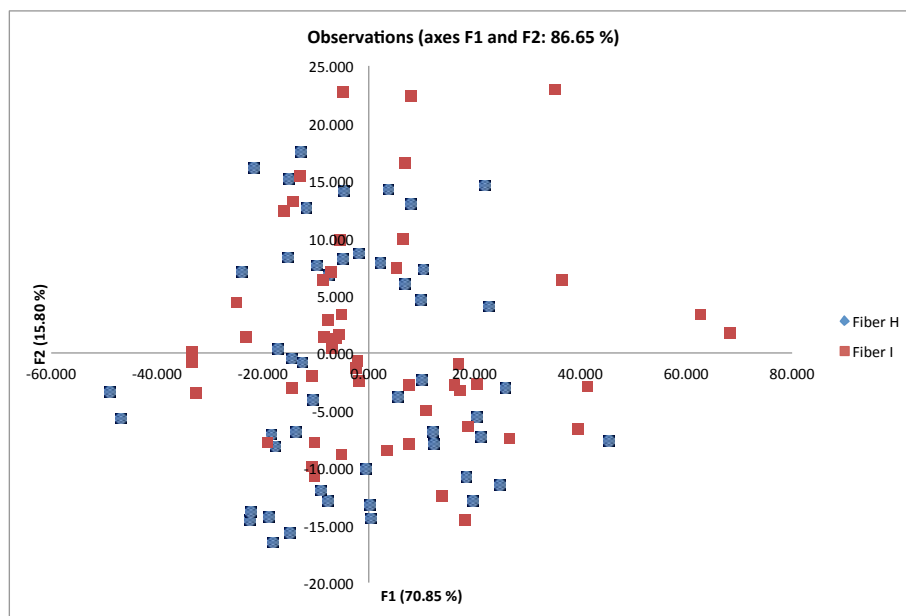


Figure A56. PCA Observations Plot of Exemplars H and I

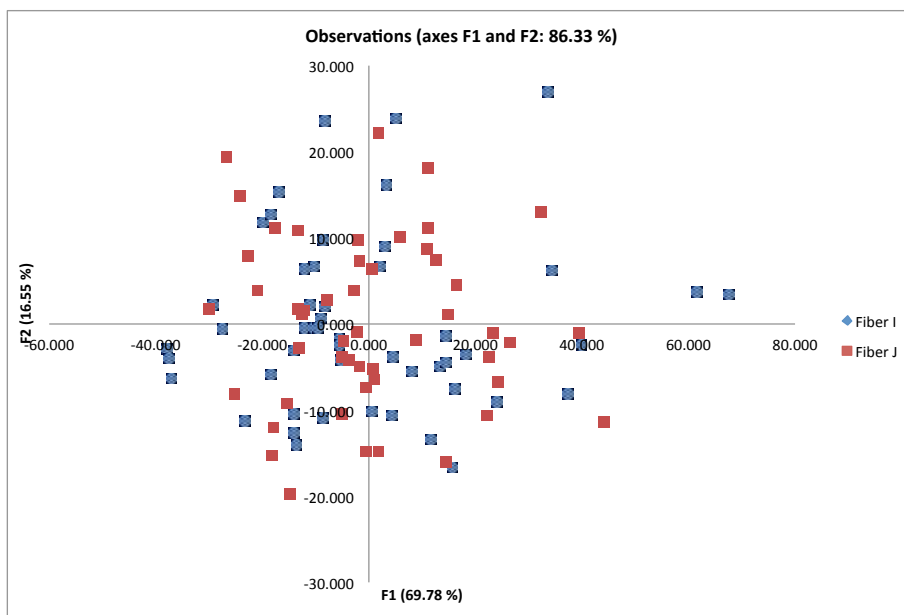


Figure A57. PCA Observations Plot of Exemplars I and J

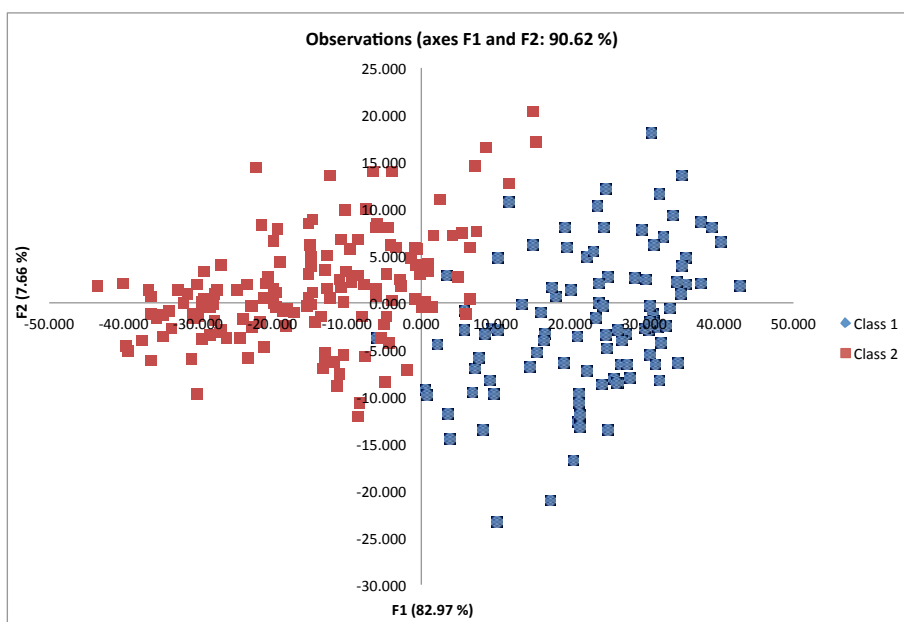


Figure A58. PCA Observations Plot of Classes 1 and 2

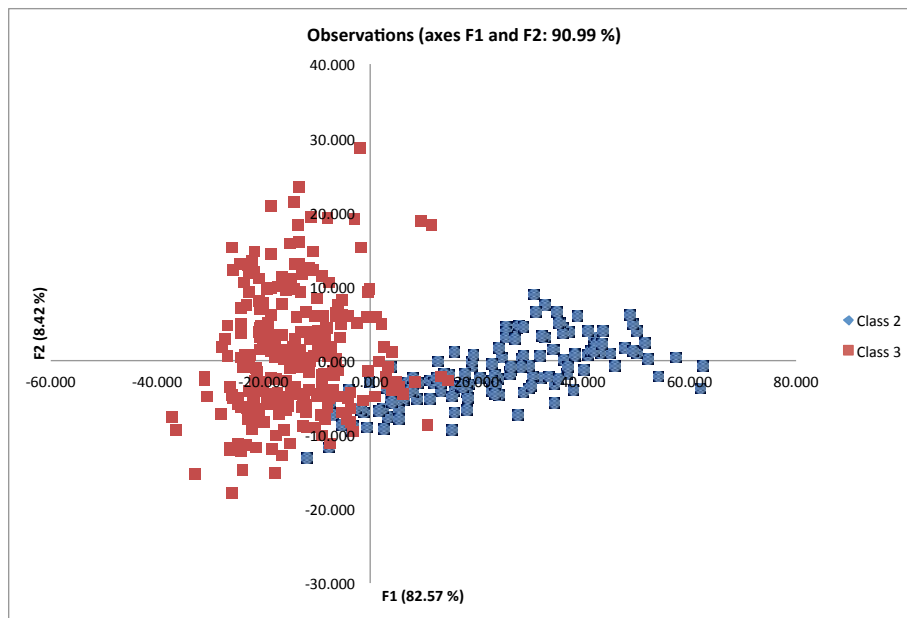


Figure A59. PCA Observations Plot of Classes 2 and 3